

Sensing, Compression, and Recovery for WSNs: Sparse Signal Modeling and Monitoring Framework

Giorgio Quer, *Member, IEEE*, Riccardo Masiero, *Member, IEEE*, Gianluigi Pillonetto, *Member, IEEE*, Michele Rossi, *Member, IEEE*, and Michele Zorzi, *Fellow, IEEE*

Abstract—We address the problem of compressing large and distributed signals monitored by a Wireless Sensor Network (WSN) and recovering them through the collection of a small number of samples. We propose a sparsity model that allows the use of Compressive Sensing (CS) for the online recovery of large data sets in real WSN scenarios, exploiting Principal Component Analysis (PCA) to capture the spatial and temporal characteristics of real signals. Bayesian analysis is utilized to approximate the statistical distribution of the principal components and to show that the Laplacian distribution provides an accurate representation of the statistics of real data. This combined CS and PCA technique is subsequently integrated into a novel framework, namely, *SCoReI*: Sensing, Compression and Recovery through ON-line Estimation for WSNs. *SCoReI* is able to effectively self-adapt to unpredictable changes in the signal statistics thanks to a feedback control loop that estimates, in real time, the signal reconstruction error. We also propose an extensive validation of the framework used in conjunction with CS as well as with standard interpolation techniques, testing its performance for real world signals. The results in this paper have the merit of shedding new light on the performance limits of CS when used as a recovery tool in WSNs.

Index Terms—Compressive sensing, wireless sensor networks, data gathering, distributed monitoring, Bayesian estimation, principal component analysis.

I. INTRODUCTION AND RELATED WORK

THE area of communication and protocol design for Wireless Sensor Networks (WSNs) has been widely researched in the past few years. One of the first studies addressing the problem of efficiently gathering correlated data from a wide network deployment is [1], which highlights the interdependence among the bandwidth, the decoding delay

Manuscript received April 5, 2011; revised February 28, 2012; accepted May 15, 2012. The associate editor coordinating the review of this paper and approving it for publication was Q. Zhang.

Part of this work has been presented at the IEEE Global Communication Conference (GLOBECOM), Honolulu, HI, Dec. 2009, and at the IEEE International Workshop on Scalable Ad Hoc and Sensor Networks, Saint Petersburg, Russia, Oct. 2009. This work was partially supported by the MO-SAICS project, “MONitoring Sensor and Actuator networks through Integrated Compressive Sensing and data gathering,” funded by the University of Padova under grant no. CPDA094077, by the European Commission through the FP7 EU projects “Internet of Things—Architecture (IoT-A)” (G.A. no.257521, <http://www.iot-a.eu/public>) and “Symbiotic Wireless Autonomous Powered system (SWAP)” (G.A. no. 251557, <http://www.fp7-swap.eu>), and by the Center for Wireless Communications, University of California San Diego.

G. Quer is with the California Institute for Telecommunications and Information Technology, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093 (e-mail: gquer@ucsd.edu).

R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi are with the Department of Information Engineering, University of Padova, Via G. Gradenigo 6/B, 35131 Padova (PD), Italy (e-mail: masieror, giapi, rossi, zorzi@dei.unipd.it). Digital Object Identifier 10.1109/TWC.2012.081612.110612

and the routing strategy employed. Under certain assumptions of regularity of the observed process, the authors claim the feasibility of large-scale multi-hop networks from a transport capacity perspective. Classical source coding, suitable routing algorithms and re-encoding of data at relay nodes have been proposed as key ingredients for joint data gathering and compression. In fact, WSN applications often involve multiple sources which are correlated both temporally and spatially. Subsequent works such as [2], [3] proposed algorithms that involve collaboration among sensors to implement classical source coding (e.g., see [4], [5]) in a distributed fashion.

New methods for distributed sensing and compression have been developed based on the recent theory of Compressive Sensing (CS) [6]–[8]. CS was originally developed for the efficient storage and compression of digital images, which show high spatial correlation. In the very recent literature, a Bayesian approach has been used to develop efficient and auto-tunable algorithms for CS, see [9]. 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 [10] 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 [11], 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 in [12], [13], 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.

[14] also addresses the problem of gathering data in distributed WSNs through multi-hop routing: tree topologies are exploited for data gathering and routing, and the Wavelet transformation is used for data compression. In [15] an approach to distributed coding and compression in sensor networks

based on CS is presented. The authors advocate the need to exploit the data correlation both temporally and spatially. The projections of the signal measurements are performed at each source node, only taking into account the temporal correlation of the sensor readings. The spatial correlation is then exploited at the sink by means of suitable decoders through a joint sparsity model able to characterize different types of signals. 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 number of transmissions to the DCP, with a consequent reduction of the energy consumed by the sensor nodes. A preliminary work with this aim is [16], where the coherence between the routing matrix and the sparsification matrix (i.e., the matrix used to transform the signal and make it sparse) is studied to exploit CS as a reconstruction technique for WSNs. The results of this work are anyway unsatisfactory, due to the fact that exploiting only the spatial correlation of the data does not suffice to efficiently recover the signal. In a recent work [17], the authors proposed an interesting in-network aggregation technique and exploited CS to reconstruct the data at the sink. Different from our approach, the aggregation technique depends on the network topology and the design of the sparsification matrix depends on the type of data, thus it can not automatically adapt to complex spatial and temporal correlation characteristics.

In this paper 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. 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. 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) [18].

The joint CS and PCA recovery technique is integrated in a lightweight and self-adapting framework called *SCoRel* (Sensing, Compression and Recovery through ON-line Estimation for WSNs) for the accurate reconstruction, at the DCP, of large data sets through the collection of a small number (sub-sampling) of all the sensor readings. For the reconstruction of the sub-sampled signals, *SCoRel* can accommodate diverse interpolation techniques, which are all integrated into the proposed framework. The main purpose of our work is that of devising a general solution, featuring a protocol for data recovery that is able to self-adapt to the time-varying statistical characteristics of the signal of interest, without relying on their prior knowledge. This is achieved utilizing a feedback control loop that estimates, in an online fashion, the reconstruction error and acts on the recovery process in order to keep this error bounded.

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 [19], [20].

The main contributions of this paper include:

- the design of a joint CS and PCA technique that is able to capture the characteristics of real signals;
- the validation of such technique, and the proof that it is optimal, from a Bayesian point of view, under certain assumptions;
- the design of an effective and flexible framework, *SCoRel*, for distributed sampling, data gathering and recovery of signals from actual WSN deployments;
- the integration of CS as well as other standard interpolation techniques into this framework;
- the validation of our signal reconstruction framework when used in conjunction with different interpolation techniques in the presence of real world signals.

The paper is structured as follows. In Section II we present the mathematical details for the joint use of CS and PCA and we analyze a large number of WSN testbeds and signals in Section III. In Section IV we present the details of the probabilistic model, that exploits a two-level Bayesian inference to estimate the best fitting distribution for such signals. In Section V we describe our monitoring framework: the distributed sampling method, the data collection techniques and the signal recovery that jointly exploits CS and PCA. The performance of the recovery techniques is presented in Section VI for different kinds of real signals gathered from different WSNs. 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 [21], 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 $k = 1, 2, \dots, K$. 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 $\bar{\mathbf{x}}$ and the sample covariance matrix $\hat{\Sigma}$ of $\mathbf{x}^{(k)}$ are defined as:

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

Given the above equations, let us consider the orthonormal matrix \mathbf{U} whose columns are the unitary eigenvectors of $\hat{\Sigma}$, 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 $\mathbf{s}^{(k)} \stackrel{def}{=} \mathbf{U}^T(\mathbf{x}^{(k)} - \bar{\mathbf{x}})$. If the instances $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)}$ of the process \mathbf{x} are temporally correlated, then only a fraction of the elements of $\mathbf{s}^{(k)}$ may be sufficient to collect the overall energy of $\mathbf{x}^{(k)} - \bar{\mathbf{x}}$. 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:

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

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 $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(K)}\}$ 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 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)}$.

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 Ψ of size $N \times N$ such that

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

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, defined¹ as $E_s^{(k)} = \frac{1}{N} \sqrt{\langle \mathbf{s}^{(k)}, \mathbf{s}^{(k)} \rangle}$. Assuming that Ψ is known, $\mathbf{x}^{(k)}$ can be recovered from $\mathbf{s}^{(k)}$ by inverting Eq. (2), i.e., $\mathbf{s}^{(k)} = \Psi^{-1}\mathbf{x}^{(k)}$. Also, $\mathbf{s}^{(k)}$ can be obtained from 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:

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

In our framework, Φ 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 Φ 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 Eq. (2) and Eq. (3) we can write

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

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

²This selection of Φ has two advantages: 1) the matrix is orthonormal as required by CS, see [16], and 2) this type of routing matrix can be obtained through realistic routing schemes.

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 [22]. However, if $\mathbf{s}^{(k)}$ is sparse and the matrix product $\Phi\Psi$ satisfies the RIP condition [8], it has been shown that Eq. (4) can be inverted with high probability through the use of specialized optimization techniques [23]. These allow to retrieve $\mathbf{s}^{(k)}$, from which the original signal $\mathbf{x}^{(k)}$ is found through Eq. (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 also make the following assumptions, that will be justified in the next sections: **(1.)** at each time k we have perfect knowledge of the previous K process samples, namely we perfectly know the set $\mathcal{X}^{(k)} = \{\mathbf{x}^{(k-1)}, \mathbf{x}^{(k-2)}, \dots, \mathbf{x}^{(k-K)}\}$, 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 section 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 $\bar{\mathbf{x}}$ 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 $\bar{\mathbf{x}}^{(k)}$ and $\hat{\Sigma}^{(k)}$, respectively. Hence, we can write:

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

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

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

where with the symbol $\Phi^{(k)}$ we make explicit that also the routing matrix Φ can change over time. The form of Eq. (6) is similar to that of Eq. (4) with $\tilde{\Phi} = \Phi^{(k)}\mathbf{U}^{(k)}$. The original signal $\mathbf{x}^{(k)}$ is approximated as follows: 1) finding a good estimate⁴ of $\mathbf{s}^{(k)}$, namely $\hat{\mathbf{s}}^{(k)}$, using the techniques in [8] or [24] and 2) applying the following calculation:

$$\hat{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}^{(k)}\hat{\mathbf{s}}^{(k)}. \quad (7)$$

³In Section V we present a practical scheme that does not need this assumption in order to work.

⁴In this paper we refer to a good estimate of $\mathbf{s}^{(k)}$ as $\hat{\mathbf{s}}^{(k)}$ such that $\|\mathbf{s}^{(k)} - \hat{\mathbf{s}}^{(k)}\|_2 \leq \epsilon$. Note that by keeping ϵ arbitrarily small, assumption (1.) above is very accurate.

III. 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, heart beats, or environmental signals like the values of temperature and humidity, indoor or outdoor. Very often both the density of sensor network deployments and the sampling rate are very high, and therefore sensor observations are strongly correlated in space and time.

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 accuracy. 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 VI to test, using the Bayesian model presented in Section IV-A, whether the SCoRe1 technique, that integrates CS and PCA, is effective for compression and recovery of real signals.

A. WSN deployments

We consider five different WSN deployments, whose sensor reading datasets were kindly provided to the authors. A brief technical overview of each of these five experimental networks follows.

T1 WSN testbed of the Department of Information Engineering (DEI) at the University of Padova, collecting data from 68 TmoteSky wireless sensor nodes [25]. 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), [26]. This measurement system exploits 100 SensorScope weather sensors 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, [27]. 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, [28]. CitySense is an urban scale

TABLE I
DETAILS OF THE CONSIDERED WSN AND GATHERED SIGNALS FOR EACH CAMPAIGN CONSIDERED (A, B, C).

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

deployment that will consist of 100 wireless sensor nodes equipped with an ALIX 2d2 single-board computer. The transmitting interface operates in 802.11b/g ad hoc mode at 2.4 GHz, and is reconfigurable by the user. Nowadays this WSN deployment counts about 25 nodes;

T5 The Sense&Sensitivity [29] 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).

B. Signals

From the above WSNs, we obtained six different types of signals: **S1**) Temperature; **S2**) Humidity; **S3**) Luminosity in the range 320 – 730 nm; **S4**) Solar Radiation; **S5**) Wind Direction; and **S6**) Voltage. A subset of the sensor reading datasets have been rearranged in a Matlab structure to be processed. These signals can be downloaded from [30]. 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

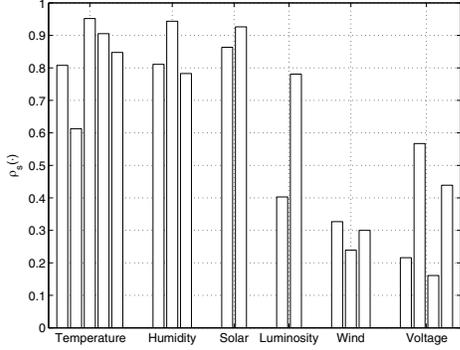


Fig. 1. Inter-node correlation for the signals (S1)–(S6) gathered from testbeds T1–T5. Not all the signals are present in each testbed, e.g., Humidity (S2) is present in only three testbeds. If we have the same signal in multiple campaigns with the same testbed, we average the value of ρ_s among them.

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 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.

C. Correlation properties

From each signal described above, named $\mathbf{x}^{(k)} \in \mathbb{R}^N$ (where N is the total number of sensors in the testbed) we calculated the average inter-node (spatial) correlation $\rho_s(\mathbf{x}^{(\cdot)})$, defined as the average correlation between the one dimensional signal sensed by node i , i.e., $x_i(k)$, and the one sensed by node j , i.e., $x_j(k)$, for all the node pairs i, j , formally:

$$\rho_s(\mathbf{x}^{(\cdot)}) = \sum_{k=1}^{K_T} \frac{1}{K_T} \sum_{i=1}^N \sum_{j>i}^N \frac{(x_i^{(k)} - E[x_i]) (x_j^{(k)} - E[x_j])}{((N^2 - N)/2) \sigma_{x_i} \sigma_{x_j}}, \quad (8)$$

where the time varies with $k = 1, \dots, K_T$. $\rho_s(\mathbf{x}^{(\cdot)})$ gives us a measure of the expected sparsity of the principal components $\mathbf{s}^{(k)} \in \mathbb{R}^N$. In a real scenario (with realistic measurements), if we calculate the principal components of a signal with maximum inter-node correlation, i.e., $\rho_s(\mathbf{x}^{(\cdot)}) = 1$, we will obtain a signal $\mathbf{s}^{(k)}$ with only the first component different from zero. Conversely, if we calculate the principal components of a signal with minimum inter-node correlation $\rho_s(\mathbf{x}^{(\cdot)}) = 0$, we will obtain a signal $\mathbf{s}^{(k)}$ with no negligible components (with respect to the overall energy of the signal). In Fig. 1 we depict the inter-node correlation for all the signals considered (S1–S6). We note that Temperature (S1), Humidity (S2) and Solar Radiation (S4) have, on average, a high inter-node correlation ($\rho_s(\mathbf{x}^{(\cdot)}) \simeq 0.7$), while Luminosity (S3), Wind Direction (S5) and Voltage (S6) have a lower inter-node correlation ($\rho_s(\mathbf{x}^{(\cdot)}) \simeq 0.4$). To further analyze these signals, we consider

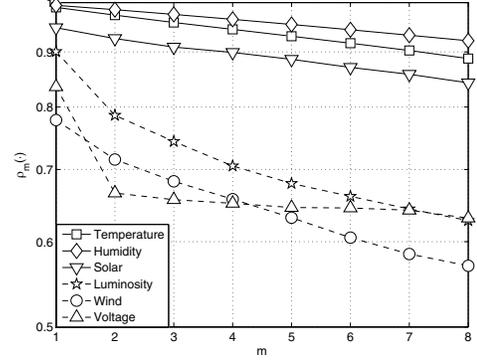


Fig. 2. Intra-node correlation for the signals chosen among the signals considered in Fig. 1 (one signal per type).

the intra-node (temporal) correlation $\rho_m(\mathbf{x}^{(\cdot)})$, that is the correlation of the one dimensional signal $x_i^{(k)}$ sensed by a single node with the same signal shifted by m time samples, i.e., $x_i^{(k+m)}$, averaged for all the N signals of $\mathbf{x}^{(k)} \in \mathbb{R}^N$. It is defined as

$$\rho_m(\mathbf{x}^{(\cdot)}) = \sum_{i=1}^N \frac{1}{N} \frac{\sum_{k=1}^{K_T} (x_i^{(k)} - E[x_i]) (x_i^{(k+m)} - E[x_i])}{K_T \sigma_{x_i}^2}. \quad (9)$$

For representation purposes, we choose one signal for each type, within the signals depicted in Fig. 1, and we represent for each chosen signal the temporal correlation $\rho_m(\mathbf{x}^{(\cdot)})$, for $m = 1, \dots, 8$ in Fig. 2. We notice that Temperature (S1), Humidity (S2) and Solar Radiation (S4) show a high intra-node correlation even for $m = 8$ ($\rho_s(\mathbf{x}^{(\cdot)}) \geq 0.85$), while for Luminosity (S3) and Wind Direction (S5) the temporal correlation quickly decreases ($\rho_s(\mathbf{x}^{(\cdot)}) \leq 0.65$). The Voltage (S6) signal, instead, has different characteristics, since even though it has inter-node and intra-node correlation similar to Luminosity and Wind Direction, it is a nearly constant signal.

Among all considered data sets, for our experimental analysis in Sections V and VI we picked a subset of the signals that is representative of the different statistical characteristics. To this end, our final choice has been to use the signals gathered from the WSN testbed deployed on the ground floor of the Department of Information Engineering at the University of Padova [25], from $N = 68$ TmoteSky wireless nodes equipped with IEEE 802.15.4 compliant radio transceivers. We have chosen these signals because: 1) they are representative of the entire data set in terms of signal statistics, and 2) we have full control on the WSN from which they have been gathered, which allowed the collection of meaningful traces for the performance evaluation of *SCoRe1*. Specifically, we considered 5 signals divided in classes according to their statistical characteristics: **C1** two signals with high temporal and spatial correlation, i.e., (S1) ambient temperature [°C] and (S2) ambient humidity [%]; **C2** a signal with lower correlation, i.e., (S3) luminosity [A/W] in the range 320–730 nm; and **C3** the battery level [V] of the sensor nodes (S6) during all the signal collection campaigns. Over time, each signal has been collected every 5 minutes. The results have

been obtained from 100 independent simulation runs over these traces and by averaging the data collection performance over all signals in each class.

IV. SPARSITY ANALYSIS OF REAL SIGNAL PRINCIPAL COMPONENTS

In this section we first introduce a model to represent a broad range of environmental signals. Then, we infer the statistical distribution of the vector random process \mathbf{s} from the samples $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(T)}\}$, which are obtained from the WSN signals presented in Section III; the parameter T is the duration (number of time samples) of each monitoring campaign in Table I. Finally, we use the obtained statistical distribution to legitimate the use of CS in WSNs when it is exploited according to our framework.

A. Sparse Signal Model

In the following, we propose a graphical model which links together all the variables involved in our analysis, i.e., those required to define the monitoring framework, and those involved in the stochastic model for the signal \mathbf{s} . We have chosen to represent such variables with a Bayesian Network (BN) [18], i.e., a Directed Acyclic Graph (DAG) where nodes represent random variables and arrows represent conditional dependencies among them. With this approach it is possible to determine the conditional independence between two variables, applying a set of rules known as *d-separation* rules, e.g., see [31] for a detailed description about BN properties.

In detail, in the *sparse signal model* box of Fig. 3 we introduce a Bayesian model to describe the statistical properties of the elements of $\mathbf{s}^{(k)}$. In the *monitoring framework* box of Fig. 3, instead, we depict 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 $\Phi^{(k)}$ as in Eq. (3), the invertible matrix $\mathbf{U}^{(k)}$, obtained through PCA, and the sparse representation $\mathbf{s}^{(k)}$, introduced in Eq. (1). Analyzing the DAG in Fig. 3 according to the *d-separation* rules, we can describe our signal model as follows:

- **data gathering:** the WSN signal $\mathbf{x}^{(k)}$ is independent of the stochastic sampling matrix $\Phi^{(k)}$, whose nature is detailed in Section V, but the observation of the measurements in $\mathbf{y}^{(k)}$ 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 $\mathbf{x}^{(k)}$. $\mathbf{U}^{(k)}$ can be seen as the state of a dynamic system, since it summarizes at each instant k all the past history of the system, represented by the set $\mathcal{X}^{(k)}$. The system input is the signal $\mathbf{s}^{(k)}$, that can be seen as a Laplacian or Gaussian innovation process. These types of priors on the signal induce estimators that use, respectively, the L_1 and L_2 norm of the signal as regularization terms. We consider such priors because they are often used in the literature in view of their 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

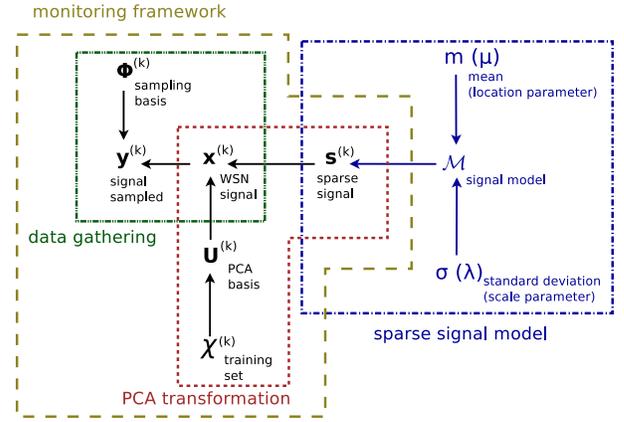


Fig. 3. Bayesian network used to model the considered real signals. In the scheme we highlight the monitoring framework at each time sample k .

in [22] for a thorough discussion. We 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 $\Phi^{(k)}$;

- **sparse signal model:** we observe that the priors assigned to the variable \mathcal{M} and to the corresponding parameters (e.g., for a Gaussian model the mean value m of each component and the standard deviation σ , whereas for a Laplacian model the location parameter μ and the scale parameter λ) are non informative priors, i.e., a uniform prior in \mathbb{R} for m and μ , and a uniform prior in $\mathbb{R}_{\geq 0}$ of the variance, where $\mathbb{R}_{\geq 0}$ is the set of positive real numbers. Here the observation of the sparse signal $\mathbf{s}^{(k)}$ makes the variable \mathcal{M} and its corresponding parameters no longer dependent on the variables of the monitoring framework, so they can be analyzed separately as we do in the following.

B. Sparsity Analysis

From the theory [21] 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 $i, j \in \{1, \dots, N\}$ and $i \neq j$. For the purpose of our analysis, we make a stronger assumption, i.e., we build our model of \mathbf{s} considering statistical independence among its components, i.e., $p(s_1, \dots, s_N) = \prod_{i=1}^N p(s_i)$. A further assumption that we make is to consider the components of \mathbf{s} as stationary over the entire monitoring period.⁵

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

$$s_i = \sum_{j=1}^N u_{ji}(x_j - \bar{x}_j), \quad i = 1, \dots, N, \quad (10)$$

⁵Note that if this assumption does not hold, i.e., the signal statistics change over time, the model is anyway able to follow such signals, since the signal basis adapts periodically if the signal statistics change, see Section V.

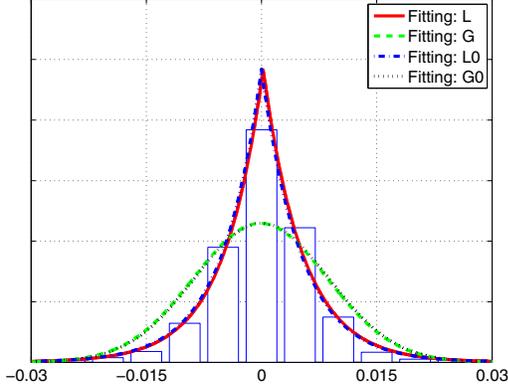


Fig. 4. Empirical distribution and model fitting for a principal component of signal S3, luminosity in the range 320 – 730 nm.

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 [32], we rely upon two levels of inference.

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

$$p(\theta|\mathcal{D}, \mathcal{M}_i) = \frac{p(\mathcal{D}|\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_i)}, \quad (11)$$

i.e.,

$$\theta_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}} p(\theta|\mathcal{D}, \mathcal{M}_i), \quad (12)$$

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. 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 equally likely, 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) [33], that we define as:

$$\text{BIC}(\mathcal{M}_i) \stackrel{\text{def}}{=} \ln [p(\mathcal{D}|\theta_{\text{MAP}}, \mathcal{M}_i)p(\theta_{\text{MAP}}|\mathcal{M}_i)] - \frac{\ell_i}{2} \ln(T), \quad (13)$$

where θ_{MAP} is defined in Eq. (12), 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 insight in the selection of the best fitting model, penalizing those models requiring more parameters.

According to the introduced formalism we consider $\{s^{(1)}, s^{(2)}, \dots, s^{(T)}\}$ as the set of collected data \mathcal{D} ; further, the observation of the experimental data gives empirical evidence

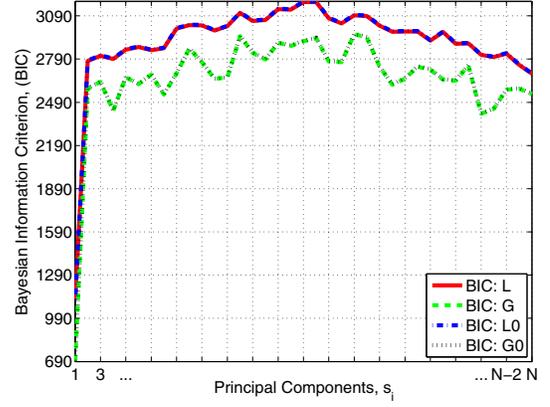


Fig. 5. Bayesian Information Criterion (BIC) per Principal Component, for each model \mathcal{M}_1 – \mathcal{M}_4 , WSN T1 (DEI), campaign A and signal S3, luminosity.

for the selection of four statistical models \mathcal{M}_i and corresponding parameter vectors θ : \mathcal{M}_1) a Laplacian distribution with $\theta = [\mu, \lambda]$, that we call \mathcal{L} ; \mathcal{M}_2) a Gaussian distribution with $\theta = [m, \sigma^2]$, 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 s_i is therefore described by the set $\{\mathcal{L}, \mathcal{G}, \mathcal{L}_0, \mathcal{G}_0\}$. In detail, for each signal $S1$ – $S6$ in the corresponding WSNs and campaigns of Table I, we collected the $T + K$ signal samples $\{\mathbf{x}^{(1-K)}, \dots, \mathbf{x}^{(-1)}, \mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}\}$ from which we computed $\{s^{(1)}, s^{(2)}, \dots, s^{(T)}\}$ according to what explained in Section II-C. Then, for each component $s_i, i = 1, \dots, N$, and for each model $\mathcal{M}_i, i = 1, \dots, 4$, we estimated the parameters (i.e., the most probable *a posteriori*, *MAP*) that best fit the data according to Eq. (12). These estimations are related to the BN in Fig. 3 (*sparse signal model* box) and since we deal with Gaussian and Laplacian distributions, they have well known and closed form solutions [20]. In detail, for each component s_i we compute:

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

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

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

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

Fig. 4 shows an example of data fitting according to the aforementioned models; in this figure 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 luminosity (S3). This signal has been observed during the data collection of the campaign A, in the WSN testbed T1 (DEI). From the graph, 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. (13). Since we assigned non informative priors to the model parameters, $p(\theta_{\text{MAP}}|\mathcal{M}_i)$ is a constant for

TABLE II

BAYESIAN INFORMATION CRITERION (BIC) AVERAGED OVER ALL PRINCIPAL COMPONENTS AND RELATIVE CAMPAIGNS, FOR EACH MODEL \mathcal{M}_1 – \mathcal{M}_4 , EACH TESTBED **T1**–**T5** AND EACH CORRESPONDING PROVIDED SIGNAL AMONG **S1** (TEMPERATURE), **S2** (HUMIDITY), **S3** (LUMINOSITY), **S4** (SOLAR RADIATION), **S5** (WIND DIRECTION), AND **S6** (VOLTAGE).

| WSN Testbed T1 (DEI) | | | | | |
|------------------------------------|---------------|----------------|----------------|----------------|---------------|
| | S1 | S2 | S3 | S6 | |
| \mathcal{L} | 1382.8 | 1059.8 | 2191.7 | 4656.9 | |
| \mathcal{G} | 1042.1 | 804.9 | 1690 | 3814.1 | |
| \mathcal{L}_0 | 1385.5 | 1062.4 | 2194.9 | 4660.1 | |
| \mathcal{G}_0 | 1044.9 | 807.60 | 5078.3 | 3816.9 | |
| WSN Testbed T2 (EPFL LUCE) | | | | | |
| | S1 | S2 | S4 | S5 | S6 |
| \mathcal{L} | -36.1 | -992.3 | -2973.9 | -3694.9 | 1854.1 |
| \mathcal{G} | -195.3 | -1163.7 | -3628.1 | -4026.5 | 1191.4 |
| \mathcal{L}_0 | -33.3 | -989.5 | -2970.7 | -3691.5 | 1856.3 |
| \mathcal{G}_0 | -192.5 | -1160.9 | -3625.3 | -4023.6 | 1194.2 |
| WSN Testbed T3 (EPFL St Bernard) | | | | | |
| | S1 | S2 | S4 | S5 | S6 |
| \mathcal{L} | -82.3 | -1473 | -2972.6 | -3700.2 | 1617.8 |
| \mathcal{G} | -487.4 | -1700.7 | -3615.9 | -3850.3 | 1087.9 |
| \mathcal{L}_0 | -79.3 | -1469.9 | -2969.4 | -3697.3 | 1619 |
| \mathcal{G}_0 | -484.7 | -1697.8 | -3613.3 | -3847.5 | 1090.7 |
| WSN Testbed T4 (CitySense) | | | | | |
| | S1 | | S5 | | |
| \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 | S3 | S6 | | |
| \mathcal{L} | -127.7 | -196.2 | 110 | | |
| \mathcal{G} | -176.1 | -232.1 | 70.2 | | |
| \mathcal{L}_0 | -125.7 | -194.2 | 111.9 | | |
| \mathcal{G}_0 | -174.7 | -230.6 | 71.8 | | |

each \mathcal{M}_i and therefore the BIC can be redefined as:

$$\text{BIC}(\mathcal{M}_i) \stackrel{\text{def}}{=} \ln p(\mathcal{D}|\boldsymbol{\theta}_{MAP}, \mathcal{M}_i) - \frac{\ell_i}{2} \ln(T). \quad (14)$$

Fig. 5 shows the BIC for the aforementioned luminosity signal, for all its principal components and for all the considered models⁶. From this figure we see that the Laplacian models better fit the data for all principal components s_i , $i = 2, \dots, N$. 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. (14)). Based on the above results, we can conclude that the Laplacian model describes the principal components of all the real signals that we considered slightly better than the Gaussian model. Furthermore, it is worth noting that the first principal components (to be more precise, the

⁶The first component, s_1 , does not have a Laplacian or Gaussian prior, by construction of PCA.

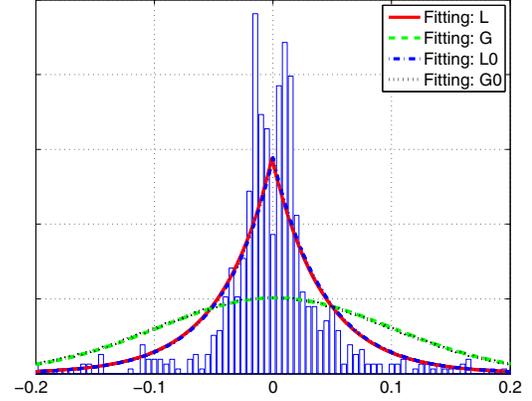


Fig. 6. Empirical distribution and model fitting for the first principal component of signal S3, luminosity in the range 320 – 730 nm.

first $K - 1$ principal components⁷ of the signal, where K is the length of the training set) 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, while 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 [34], so that only the first principal component shows a behavior different from the one illustrated in Fig. 4 as reported in Fig. 6. In any case, the Laplacian model still fits the observed data better than the Gaussian model.

C. Bayesian MAP condition and CS recovery

We have just 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, we now 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. (11).

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. Eq. (5) shows 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 $\mathbf{y}^{(k)} = \Phi \mathbf{x}^{(k)} \in \mathbb{R}^M$, we are interested in recovering, via interpolation, the signal $\mathbf{x}^{(k)} \in \mathbb{R}^N$. Typically (e.g., see

⁷Note 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 $\{\mathbf{x}^{(k-1)} - \bar{\mathbf{x}}^{(k)}, \mathbf{x}^{(k-2)} - \bar{\mathbf{x}}^{(k)}, \dots, \mathbf{x}^{(k-K)} - \bar{\mathbf{x}}^{(k)}\}$ which spans a vector space of dimension at most $K - 1$.

[32]) this problem can be solved through a linear interpolation, where the interpolated function in our case has the form (see Eq. (5))

$$\mathbf{x}^{(k)} - \bar{\mathbf{x}}^{(k)} = \sum_{i=1}^N s_i^{(k)} \mathbf{u}_i^{(k)}. \quad (15)$$

A Bayesian approach would estimate the most probable value of $\mathbf{s}^{(k)} = (s_1^{(k)}, \dots, s_N^{(k)})^T$ by maximizing a posterior pdf of the form $p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \mathbf{U}^{(k)}, \mathcal{M})$, 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 IV-B. In detail, now the vector $\mathbf{s}^{(k)}$ is seen as the parameter vector $\boldsymbol{\theta}$ in Eq. (11), 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 the matrix $\mathbf{U}^{(k)}$ (i.e., a set of basis functions) and a model \mathcal{M} for the parameter vector $\mathbf{s}^{(k)}$, according to the BN in Fig. 3. In Eq. (11) we indicated with the symbol \mathcal{M}_i a possible model for the observed phenomenon: here that symbol is replaced with the couple $(\mathbf{U}^{(k)}, \mathcal{M})$, 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 Section IV-B. This correspondence will become clearer in the following.

As in [32], we also assume that \mathcal{M} can be specified by a further parameter set α (called hyper-prior) related to $\mathbf{s}^{(k)}$, so that the posterior can be written as

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

If the hyper-prior can be inferred from the data and has non zero values $\hat{\alpha}$, maximizing the posterior corresponds to maximizing $p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \hat{\alpha}, \mathbf{U}^{(k)}, \mathcal{M})$, that as shown in [32] can be written as

$$p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \hat{\alpha}, \mathbf{U}^{(k)}, \mathcal{M}) = \frac{p(\mathbf{y}^{(k)} | \mathbf{s}^{(k)}, \mathbf{U}^{(k)}) p(\mathbf{s}^{(k)} | \hat{\alpha}, \mathcal{M})}{p(\mathbf{y}^{(k)} | \hat{\alpha}, \mathbf{U}^{(k)}, \mathcal{M})}, \quad (16)$$

where $p(\mathbf{y}^{(k)} | \mathbf{s}^{(k)}, \mathbf{U}^{(k)})$ is the likelihood function, $p(\mathbf{s}^{(k)} | \hat{\alpha}, \mathcal{M})$ is the prior and $p(\mathbf{y}^{(k)} | \hat{\alpha}, \mathbf{U}^{(k)}, \mathcal{M})$ is a normalization factor. The parameters $\hat{\alpha}$ are estimated maximizing the evidence $p(\mathbf{y}^{(k)} | \alpha, \mathbf{U}^{(k)}, \mathcal{M})$, which is a function of α . 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. (15), without loss of generality we can assume that $\bar{\mathbf{x}}^{(k)} = 0$, 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)):

$$p(\mathbf{y}^{(k)} | \mathbf{s}^{(k)}, \mathbf{U}^{(k)}) = \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}^{(k)} \mathbf{U}^{(k)} \mathbf{s}^{(k)}), \quad (17)$$

where $\delta(x, y)$ is 1 if $x = y$ and zero otherwise. In Section IV-B, we have seen that the statistics of vector $\mathbf{s}^{(k)}$ is

well described by a Laplacian density function with location parameter μ equal to 0 (\mathcal{L}_0). This pdf is widely used in the literature [9], [24] 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:

$$p(\mathbf{s}^{(k)} | \hat{\alpha}, \mathcal{M} = \mathcal{L}_0) = \frac{e^{-\hat{\alpha} \sum_{i=1}^N |s_i^{(k)}|}}{(2/\hat{\alpha})^N}. \quad (18)$$

In this equation, all the components of $\mathbf{s}^{(k)}$ are assumed to be equally distributed. As shown in [32], using Eq. (16), we can consider the following posterior:

$$p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \mathbf{U}^{(k)}, \mathcal{L}_0) \propto p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \hat{\alpha}, \mathbf{U}^{(k)}, \mathcal{L}_0) \propto p(\mathbf{y}^{(k)} | \mathbf{s}^{(k)}, \mathbf{U}^{(k)}) p(\mathbf{s}^{(k)} | \hat{\alpha}, \mathcal{L}_0). \quad (19)$$

Using Eq. (17) through Eq. (19), maximizing the posterior corresponds to solving the problem

$$\begin{aligned} & \arg \max_{\mathbf{s}^{(k)}} p(\mathbf{s}^{(k)} | \mathbf{y}^{(k)}, \mathbf{U}^{(k)}, \mathcal{L}_0) \\ &= \arg \max_{\mathbf{s}^{(k)}} p(\mathbf{y}^{(k)} | \mathbf{s}^{(k)}, \mathbf{U}^{(k)}) p(\mathbf{s}^{(k)} | \hat{\alpha}, \mathcal{L}_0) \\ &= \arg \max_{\mathbf{s}^{(k)}} \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}^{(k)} \mathbf{U}^{(k)} \mathbf{s}^{(k)}) \frac{e^{-\hat{\alpha} \sum_{i=1}^N |s_i^{(k)}|}}{(2/\hat{\alpha})^N} \\ &= \arg \min_{\mathbf{s}^{(k)}} \sum_{i=1}^N |s_i^{(k)}|, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)} \mathbf{U}^{(k)} \mathbf{s}^{(k)} \\ &= \arg \min_{\mathbf{s}^{(k)}} \|\mathbf{s}^{(k)}\|_1, \text{ given that } \mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)} \mathbf{U}^{(k)} \mathbf{s}^{(k)}, \end{aligned} \quad (20)$$

which is the optimization problem solved by the CS reconstruction algorithms (see [8] and [23]) as we wanted to show. Note, however, that 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.

V. ITERATIVE MONITORING FRAMEWORK

In this section we present our monitoring framework called *SCoRel* for distributed compression and centralized recovery of a multi dimensional signal. We integrate the mathematical techniques proposed in Section II into an actual monitoring framework for a WSN with N sensor nodes, that exploits the model described in Section IV.

A. Logic Blocks of the Monitoring Framework

A diagram showing the logic blocks of this framework is presented in Fig. 7. Let $\mathbf{x}^{(k)} \in \mathbb{R}^N$ be the N -dimensional signal (one reading per sensor node) sampled at discrete times $k = 1, 2, \dots$. At each time k the DCP⁸ collects a compressed version $\mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)} \mathbf{x}^{(k)}$, of the original signal $\mathbf{x}^{(k)} \in \mathbb{R}^N$, with $\mathbf{y}^{(k)} \in \mathbb{R}^L$ and $L \leq N$. The *sampling* matrix $\mathbf{\Phi}^{(k)} \in \mathbb{R}^{L \times N}$, has one element equal to 1 per row and at most one element equal to 1 per column, while all the other elements are equal to zero.⁹ Thus, the elements in $\mathbf{y}^{(k)}$ are a subset of those

⁸The DCP can be the sink of the WSN or a remote server that is not battery powered, so it does not have stringent energy requirements and has enough computational resources to execute the signal recovery algorithms.

⁹The elements equal to 1 indicate which nodes transmit their data sample to the DCP at time k .

in $\mathbf{x}^{(k)}$ (spatial sampling). Note that reducing the number of nodes that transmit to the DCP is a key aspect as each sensor is supposed to be a tiny battery powered sensing unit with a finite amount of energy that determines its lifetime. At each time k the transmitting nodes are chosen in a distributed way according to a simple Random Sampling (RS) technique to be executed in each node of the WSN, as we detail shortly. The DCP is responsible for collecting the compressed data $\mathbf{y}^{(k)}$, sending a feedback to the WSN and recovering the original signal from $\mathbf{y}^{(k)}$. Next, we detail the blocks and sub-blocks which compose the *SCoRel* framework and are illustrated in Fig. 7.

Random Sampling (RS): the RS scheme is used to decide in a fully distributed way which sensors transmit their data to the DCP and which remain silent, at any given time k . This method has been chosen because it translates into a simple and general data gathering solution that is easy to implement and has a low communication overhead for the synchronization of the nodes that transmit. In detail, at each time k each sensor node decides, with probability $p_{tx}^{(k)}$, whether or not to transmit its measurement to the DCP. This decision is made independently of the past and of the behavior of the other nodes. $p_{tx}^{(k)}$ can be fixed beforehand and kept constant, or can be varied as a function of the reconstruction error and broadcast by the DCP to all the sensor nodes.

Data Collection Point (DCP). The role of the DCP is three-fold: 1) it receives as input $\mathbf{y}^{(k)}$ and returns the reconstructed signal $\hat{\mathbf{x}}^{(k)}$; 2) it adapts $p_{tx}^{(k)}$ and sends its new value to the sensor nodes, in order to reduce the number of transmissions in the network while bounding the reconstruction error; 3) it provides the recovery block with a training set $\hat{\mathcal{T}}_K$, that is used to infer the structure of the signal, which is then exploited by the signal recovery algorithm. $\hat{\mathcal{T}}_K$ is formed by the K previously reconstructed signals $\hat{\mathbf{x}}^{(k)}$, so it can be written as $\hat{\mathcal{T}}_K = \{\hat{\mathbf{x}}^{(k-K)}, \dots, \hat{\mathbf{x}}^{(k-1)}\}$.

Controller: this super-block is responsible for the estimation of the signal reconstruction's quality at the DCP and for the feedback process. It is made of the following two blocks: 1) the Error Estimation block, which computes the reconstruction quality of $\mathbf{x}^{(k)} \in \mathbb{R}^N$ from $\mathbf{y}^{(k)} \in \mathbb{R}^L$, with $L < N$ (i.e., this block evaluates how close $\hat{\mathbf{x}}^{(k)}$ is to $\mathbf{x}^{(k)}$); 2) the Feedback Control, which tunes the transmission probability $p_{tx}^{(k)}$ to reach the desired reconstruction quality, whilst saving transmissions when possible.

Error Estimation: the reconstruction error that we want to estimate is given by

$$\xi_R^{(k)} = \frac{\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\|_2}{\|\mathbf{x}^{(k)}\|_2}, \quad (21)$$

where $\hat{\mathbf{x}}^{(k)}$ is the signal reconstructed at time k by the Recovery block and $\|\cdot\|_2$ is the L_2 norm of a vector. Note that at the DCP we do not have $\mathbf{x}^{(k)}$, but only $\mathbf{y}^{(j)} = \Phi^{(j)}\mathbf{x}^{(j)}$ and $\hat{\mathbf{x}}^{(j)}$, for $j \leq k$. Since the quantity $\xi_0^{(k)} = \|\mathbf{y}^{(k)} - \Phi^{(k)}\hat{\mathbf{x}}^{(k)}\|_2 / \|\mathbf{y}^{(k)}\|_2$ is always zero, due to the fact that the received samples are reconstructed perfectly, i.e., $\Phi^{(k)}\hat{\mathbf{x}}^{(k)} = \Phi^{(k)}\mathbf{x}^{(k)}$, one might use some heuristics to calculate the error from the past samples. In this paper we use

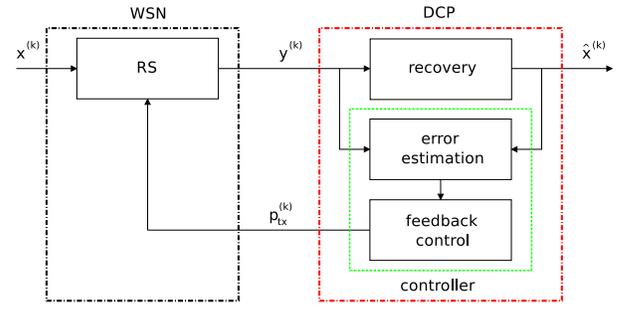


Fig. 7. Diagram of the proposed sensing, compression and recovery scheme. Note that the *Controller*, which includes the *Error estimator* and the *Feedback Control* blocks, represents the core of *SCoRel*.

the following formula:¹⁰

$$\xi^{(k)} = \frac{\left\| \begin{bmatrix} \mathbf{y}^{(k)} \\ \mathbf{y}^{(k-1)} \end{bmatrix} - \begin{bmatrix} \Phi^{(k)}\hat{\mathbf{x}}^{(k-1)} \\ \Phi^{(k-1)}\hat{\mathbf{x}}^{(k)} \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} \mathbf{y}^{(k)} \\ \mathbf{y}^{(k-1)} \end{bmatrix} \right\|_2}, \quad (22)$$

With this heuristic we compare the spatial samples collected at time k , i.e., $\mathbf{y}^{(k)}$, with the reconstructed values at time $k-1$, i.e., $\hat{\mathbf{x}}^{(k-1)}$, sampled in the corresponding points, i.e., $\Phi^{(k)}\hat{\mathbf{x}}^{(k-1)}$. Then we compare the same signals switching the roles of k and $k-1$. Note that $\xi^{(k)}$ accounts not only for the reconstruction error but also for the signal variability. This introduces a further approximation to the error estimate, but on the other hand allows the protocol to react faster if the signal changes abruptly, which is a desirable feature. In fact, if the signal significantly differs from time $k-1$ to time k , $\xi^{(k)}$ will be large and this will translate into a higher $p_{tx}^{(k+1)}$, as detailed below.

Feedback Control: this block calculates the new p_{tx} and broadcasts it to the network nodes. The calculation of the new p_{tx} is made according to a technique similar to TCP's congestion window adaptation, where p_{tx} is exponentially increased in case the error is above a defined error threshold τ (to quickly bound the error) and is linearly decreased otherwise. In detail, for some constants $C_1 \in [1, +\infty]$, $C_2 \in \{1, 2, \dots, N\}$ and p_{tx}^{\min} , we update the probability of transmission as:

$$p_{tx}^{(k+1)} = \begin{cases} \min \left\{ p_{tx}^{(k)} C_1, 1 \right\} & \text{if } \xi^{(k)} \geq \tau \\ \max \left\{ p_{tx}^{(k)} - C_2/N, p_{tx}^{\min} \right\} & \text{if } \xi^{(k)} < \tau. \end{cases} \quad (23)$$

In Section V-B, we provide some insight on the choice of the parameters in Eq. (23).

Recovery: the recovery technique adopted in our approach is the joint CS and PCA technique described in Section II-C. Specifically, the set $\hat{\mathcal{T}}_K$ is used at each time step k to compute the sample mean vector $\bar{\mathbf{x}}$ and the sample covariance matrix $\hat{\Sigma}$. Thus, $\mathbf{U}^{(k)}$ is obtained from $\hat{\Sigma}$ and Eqs. (6)-(7) are used to retrieve $\hat{\mathbf{x}}^{(k)}$, using the techniques of [8] or [24]. The joint CS and PCA technique, hereafter referred to as CS-PCA, will be compared in Section VI with other classical

¹⁰We tried other heuristics and verified through extensive simulation that they perform similarly but slightly worse than the one in Eq. (22). These are therefore not listed here, as they do not provide any additional insight.

interpolation techniques that can alternatively be plugged into our framework. These techniques are described in detail in the technical report [30].

B. Validation of the Monitoring Framework

In order to illustrate the choices made in the design of *SCoRe1*, we consider two simple strategies for iteratively sensing and recovering a given signal. In particular, we aim to explain the reasons for: 1) the adoption of an approximate training set $\widehat{\mathcal{T}}_K$ and 2) the definition of the Controller block in Fig. 7.

The first strategy we consider aims to adapt to the possible variable statistics of the observed signals and is referred to as *2 Phases*, since it alternates two phases of fixed length. The former is a *training phase* lasting K_1 time samples, during which the DCP collects the readings from all N sensors and uses them to estimate the statistics needed by the recovery algorithm. During this phase, *each sensor transmits* its data, i.e., the signal received at the DCP at time k is $\mathbf{y}^{(k)} = \mathbf{x}^{(k)}$, and at the end of this phase the DCP has stored a training set $\mathcal{T}_{K_1} = \{\mathbf{x}^{(k-K_1)}, \dots, \mathbf{x}^{(k-1)}\}$ that will be used to infer the relevant statistics. The latter is a *monitoring phase* of K_2 time samples, with $K_2 \geq K_1$, during which (on average) only $L \leq N$ nodes transmit, according to the adopted RS scheme with $p_{tx} = L/N$. The signal of interest is thus reconstructed from this data set by the Recovery block exploiting the statistics computed in the training phase, as detailed in Section V-A.

A major drawback of this technique is that it is very sensitive to the choice of the parameters that govern the compression and the recovery phases. The parameters to be set are p_{tx} and the lengths of the training phase K_1 and the monitoring phase K_2 . These parameters must be set at the beginning of the transmission and can only be tuned manually. Hence, even though the initial choice is optimal for the specific signal monitored, *2 Phases* is not able to adapt to sudden changes in the signal statistics. Moreover, the training phase accounts for the biggest part of the total cost in terms of number of transmissions, as we show shortly. A solution to the latter problem is to obtain the training set, that is necessary for the reconstruction algorithm, from the K previously reconstructed signals, i.e., $\widehat{\mathcal{T}}_K$. In this way, we eliminate the training phase and the nodes at each time k transmit with a fixed probability p_{tx} : this is the *Fixed p_{tx}* technique. This way of obtaining the training set from the previously reconstructed signals is the same as in *SCoRe1*, with the difference that in *SCoRe1* there is a feedback mechanism to bound the reconstruction error.

To compare *SCoRe1* with the above schemes, we use different signals classified according to their statistical characteristics, as detailed in Section III: **C1**) signals with high temporal and spatial correlation, e.g., ambient temperature [°C] or ambient humidity [%]; **C2**) signals with lower correlation, e.g., luminosity [A/W]; and **C3**) the battery level [V] of the sensor nodes during the signal collection campaign.

The x-axis of the performance plots represents the normalized cost expressed as the average fraction of packet

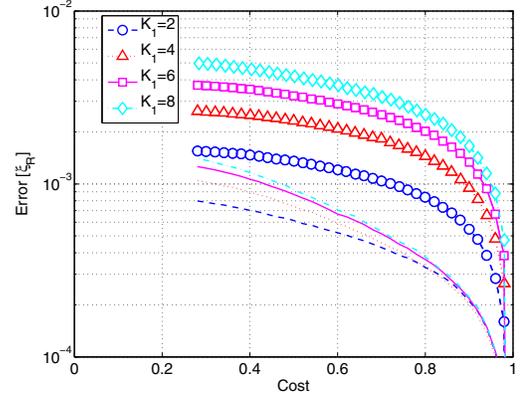


Fig. 8. Impact of the choice of K_1 on the performance of *2 Phases*, for signal (C1).

transmissions in the network per time sample, formally:

$$\text{Cost} = \frac{1}{K D_{\text{TOT}}} \sum_{k=1}^K \sum_{n=1}^N D_n I_n(k), \quad (24)$$

where K is the number of considered time instants (i.e., the overall duration of the data collection), N is the total number of nodes in the WSN, D_n is the distance in terms of number of hops from node n to the DCP,¹¹ $D_{\text{TOT}} = \sum_{n=1}^N D_n$ and $I_n(k)$ is an indicator function, with $I_n(k) = 1$ if node n transmits and $I_n(k) = 0$ if node n remains silent at time k . Note that a normalized cost of 1 corresponds to the case where all nodes transmit during all time instants $1, 2, \dots, K$, which accounts for the maximum energy consumption for the network. The cost of the feedback transmitted by the DCP is neglected here, since the feedback packets are assumed to be very short compared to the data packets.

The y-axis, instead, shows the signal reconstruction error at the end of the recovery process, calculated according to Eq. (21). In order to vary the cost (x-axis) for the three techniques we modify the following parameters: for *2 Phases* and *Fixed p_{tx}* we vary the probability of transmission p_{tx} that is set at the beginning of the data gathering in the range $]0, 1[$; for *SCoRe1*, we vary the error threshold τ of Eq. (23) in the range $]0, 1[$. Moreover, we set the training phase and the monitoring phase lengths for *2 Phases* to $K_1 = 2$ and $K_2 = 4$, respectively, whilst for *Fixed p_{tx}* and *SCoRe1* we set the training set length to $K = 2$. Further, for *SCoRe1* we also set $C_1 = 1.3$, $C_2 = 3$ and $p_{\min} = 0.05$. These parameter choices have been made after extensive simulations.

In particular, in Fig. 8 we show the impact of the choice of K_1 on the performance of *2 Phases*, for the recovery of signal (C1).¹² We see that the performance decreases with an increasing value of K_1 , so in our case the best choice for this parameter is $K_1 = 2$. In the same figure, the solid and dotted lines without marks represent lower bounds on the recovery error performance, which are obtained assuming that at each instant k the recovery algorithm can use a genie to retrieve at no cost an updated version of the training set, i.e., \mathcal{T}_{K_1} .

¹¹Note that D_n is considered to take into account the multi-hop structure of the network in the performance analysis.

¹²A similar analysis has been performed also for signals (C2) and (C3).

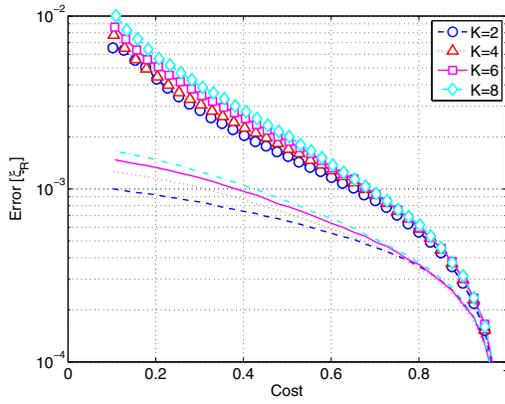


Fig. 9. Impact of the choice of K on the performance of *Fixed p_{tx}* , for signal (C1).

These bounds represent the achievable performance in the idealized case when all the previous signal samples are known at the sink. The gap between the actual recovery error and the corresponding lower bound serves as a further indication for the parameter selection and performance comparison; as an example, note that in Fig. 8 the relative distance between the curve with $K_1 = 2$ and its lower bound is the smallest.

In Fig. 9, instead, we show the performance of *Fixed p_{tx}* , that is also representative for *SCoRe1*, varying the length of the training set K . Also in this case, the best choice for the training set length is $K = 2$. Other simulation results concerning the setting of parameters have been obtained, but they are not shown here because they do not bring further insight into the understanding of the technique.

Fig. 10 shows that with the slowly varying signals of type (C1) we achieve very good performance in all cases. In some applications, the same network can be exploited to collect different signals (as a matter of fact, most currently available nodes are equipped with more than one sensor, each measuring a different signal). In this case, we would like to fix a priori the parameters of the framework in each node, and see how the network is able to reconstruct the different signals. In order to study the achievable performance, we consider the average performance obtained for signals (C1), (C2) or (C3), as depicted in Fig. 11. Here, the error for *Fixed p_{tx}* increases dramatically for small p_{tx} and this in turn leads to a more sensitive trade-off between energy reduction and recovery accuracy. This is due to the fact that, with quickly variable signals, the error propagates and increases over time. *SCoRe1*, instead, is able to iteratively adapt its parameters to the specific characteristics of the observed signals and, in turn, to significantly outperform the other schemes. This is a favorable aspect, especially when the signal statistics is not known. To conclude, the results shown in this section allow us to validate the design choices of *SCoRe1*.

VI. PERFORMANCE COMPARISON: RECONSTRUCTION TECHNIQUES

In Section V we presented a mechanism, integrated into our monitoring framework, to jointly exploit CS and PCA for signal interpolation in WSN, which we called CS-PCA.

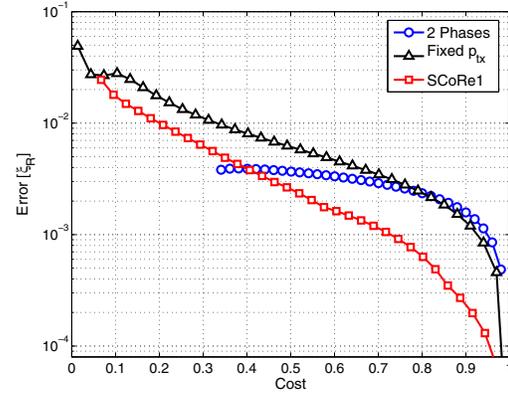


Fig. 10. Performance comparison of three iterative monitoring schemes for signals (C1), temperature and humidity.

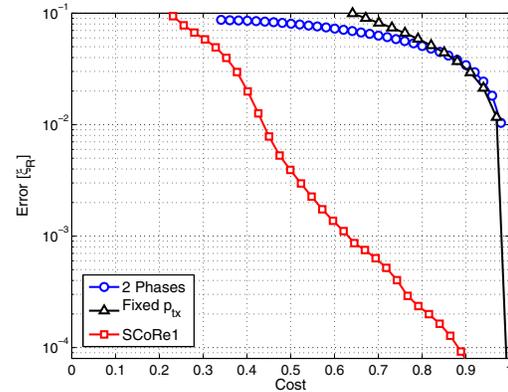


Fig. 11. Average recovery performance of the three iterative monitoring schemes for signals (C1), (C2) and (C3).

This mechanism is executed at the DCP (see Fig. 7) and, at any time k , tries to recover the original signal $\mathbf{x}^{(k)} \in \mathbb{R}^N$ from its compressed version $\mathbf{y}^{(k)} \in \mathbb{R}^L$, with $L \leq N$. Many alternatives exist in the literature, each based on a particular signal model. Given a signal model, in fact, we can determine through theoretical analysis the optimal recovery mechanism to adopt. However, the performance is strongly affected by how well the given signal model fits the real signals considered. In the first part of this section we will briefly review some state of the art interpolation techniques that, as well as CS, provide us with a solution to the following problem:

Problem (Interpolation Problem). Estimate $\hat{\mathbf{x}}^{(k)}$ (such that $\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|_2 / \|\mathbf{x}^{(k)}\|_2 \simeq 0$) knowing that $\mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)} \mathbf{x}^{(k)}$, where $\mathbf{y}^{(k)} \in \mathbb{R}^L$, $L \leq N$ and $\mathbf{\Phi}^{(k)}$ is an $[L \times N]$ sampling matrix, i.e., all rows of $\mathbf{\Phi}^{(k)}$ contain exactly one element equal to 1 and all columns of $\mathbf{\Phi}^{(k)}$ contain at most one element equal to 1, whilst all the remaining elements are zero.

In the second part of this section, instead, we analyze the performance of *SCoRe1* when used in conjunction with the interpolation methods described in the first part. This study shows that CS can be effectively exploited for WSNs.

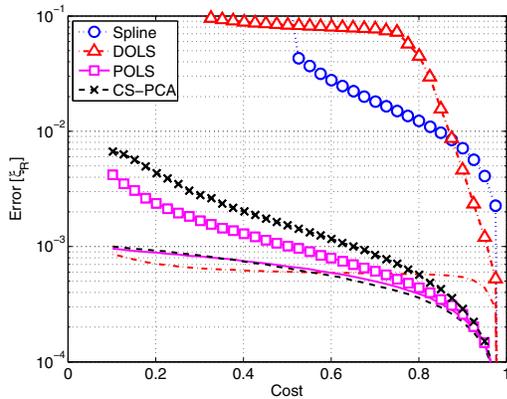


Fig. 12. Performance comparison of different interpolation techniques applied to *SCoReI*, for signals in class (C1), temperature and humidity. These performance curves are obtained with signals gathered from the DEI WSN deployment.

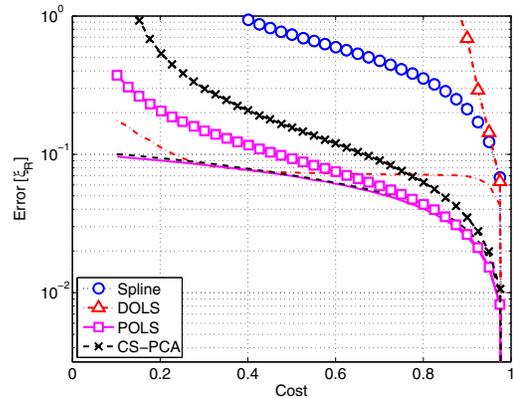


Fig. 13. Performance comparison of different interpolation techniques applied to *SCoReI*, for signals in class (C2), luminosity in the range 320–730 nm. These performance curves are obtained with signals gathered from the DEI WSN deployment.

A. Signal Models and Interpolation Techniques

The prior knowledge that we have about a given signal of interest $\mathbf{x}^{(k)}$ helps us to build a model for such signal. This knowledge¹³ can be deterministic, e.g., a description of the physical characteristics of the observed process, or probabilistic, e.g., the formulation of a probability distribution, called prior, to describe the possible realizations of $\mathbf{x}^{(k)}$. The Deterministic approach allows us to define for instance two standard recovery methods, the Biharmonic Spline (Spline) and the Deterministic Ordinary Least Square (DOLS). The first one is a standard interpolation technique that is dependent on the knowledge of where the signal sources are placed and exploits smooth functions to interpolate among different measurement points; the second, instead, assumes that the signal is stationary over a given time period, exploits a portion of the PCA matrix (i.e., the first $K - 1$ columns of $\mathbf{U}^{(k)}$ indicated with $\mathbf{U}_{K-1}^{(k)}$, where K is the length of the training set $\mathcal{T}_K^{(k)}$ or of its approximate version $\hat{\mathcal{T}}_K^{(k)}$) to determine the lower dimensional approximation of the original signal and, finally, recovers the signal with the Ordinary Least Square method (OLS). Alternatively, a probabilistic approach can be adopted as we have done in Section IV, where such solution allowed us to legitimate the use of CS-PCA when a Laplacian prior is assumed for $\mathbf{s}^{(k)}$, the principal components of the monitored signal $\mathbf{x}^{(k)}$. Assigning a Gaussian prior to $\mathbf{s}^{(k)}$, instead, leads to a classical recovery method that we call Probabilistic Ordinary Least Square (POLS). Note that all the above interpolation techniques can be implemented at the DCP in the Recovery block shown in Fig. 7.

B. Performance of the Signal Recovery Methods

In the following, we show performance curves for Spline, DOLS, POLS and CS-PCA. Note that DOLS cannot be considered as an effective solution since it is affected by a numerical stability problem. Nevertheless, we considered it in view of its simplicity and low complexity. Along the x-axis of the figures of this section we have the normalized

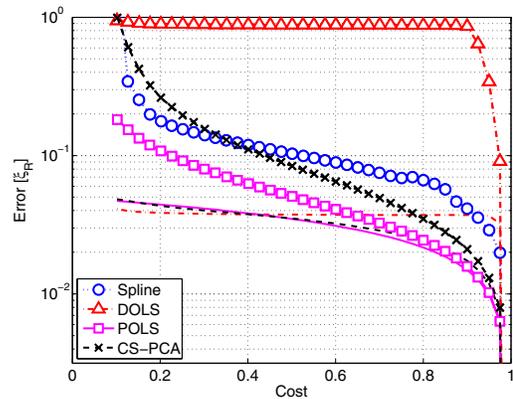


Fig. 14. Performance comparison of different interpolation techniques applied to *SCoReI*, for signals in class (C1), temperature and humidity. These performance curves are obtained with signals gathered from the EPFL WSN deployment LUCE, see [26].

cost expressed as the average fraction of packet transmissions in the network per time sample, again computed according to Eq. (24); the y-axis shows the signal reconstruction error at the end of the recovery process, calculated according to Eq. (21). To vary the cost (x-axis) we modify the parameters of *SCoReI* as explained in Section V. Solid and dotted lines represent lower bounds on the error recovery performance, which are obtained by assuming a genie that provides a perfect knowledge about past signals for the same transmission cost incurred with the actual scheme, i.e., this is implemented considering \mathcal{T}_K instead of $\hat{\mathcal{T}}_K$.

In Figs. 12 and 13 we can see how an imperfect knowledge of the training set severely impacts the recovery performance of DOLS. This is however not as dramatic for CS-PCA and POLS. The uncertainty on the training set makes the Gaussian prior for $\mathbf{s}^{(k)}$ more effective than the Laplacian one, in accordance to the central limit theorem (e.g., see [35]). As a consequence, if we use $\hat{\mathcal{T}}_K$, POLS outperforms CS-PCA, whilst if we use \mathcal{T}_K , CS-PCA and POLS perform equally well, also for highly variable signals, see Fig. 13, signal class (C2). In any case, both POLS and CS-PCA remain valid solutions for a monitoring application framework, since the

¹³For a more detailed description of the argument and the pseudo-code of all the techniques introduced, we refer the reader to the technical report [30].

performance loss from the ideal case, which assumes perfect knowledge of \mathcal{T}_K , to the one that exploits $\hat{\mathcal{T}}_K$ is sufficiently small. Concerning Spline, this method is able to reach good performance only with a transmission probability above 0.8; furthermore, the use of Spline as the interpolation technique in *SCoRel* leads to large errors due to the fact that it does not exploit any prior knowledge on the statistics of the signal to recover.

Finally, in Fig. 14 we show similar performance curves using the signals gathered from the EPFL WSN deployment LUCE, see [26]. The signals considered in this figure are of the class (C1), i.e., temperature and humidity. Also in this case, the performance is similar to Fig. 12 and all the above observations remain valid. This provides further evidence that *SCoRel* is an effective solution for monitoring applications for WSNs in different scenarios. Equally important, the achieved performance shows that CS recovery can be effectively used for data gathering 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 proposed a novel framework, called *SCoRel*, for the accurate approximation of large real world WSN signals through the collection of a small fraction of data points. *SCoRel* accommodates diverse interpolation techniques, either deterministic or probabilistic, and embeds a control mechanism to automatically adapt the recovery behavior to time varying signal statistics, while bounding the reconstruction error. We remark that our approach is also robust to unpredictable changes in the signal statistics, and this makes it very appealing for a wide range of applications that require the approximation of large and distributed datasets, with time varying statistics.

REFERENCES

- [1] A. Scaglione and S. D. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *2002 ACM MOBICOM*.
- [2] S. Servetto, "Distributed signal processing algorithms for the sensor broadcast problem," in *2003 Conference on Information Sciences and Systems*.
- [3] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 5177–5196, Dec. 2006.
- [4] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [5] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–92, Sep. 2004.
- [6] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 4036–4048, Apr. 2006.
- [7] E. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [8] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [9] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.
- [10] N. Dobleigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, Sep. 2009.
- [11] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [12] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.
- [13] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3629–3653, Oct. 2007.
- [14] G. Shen, S. Y. Lee, S. Lee, S. Patten, A. Tu, B. Krishnamachari, A. Ortega, M. Cheng, S. Dolinar, A. Kiely, M. Klimesh, and H. Xie, "Novel distributed wavelet transforms and routing algorithms for efficient data gathering in sensor webs," in *2008 NASA Earth Science Technology Conference*.
- [15] M. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal distributed sensing via random projections," in *2006 Information Processing in Sensor Networks*.
- [16] 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 *2009 ITA Workshop*.
- [17] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Efficient measurement generation and pervasive sparsity for compressive data gathering," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3728–3738, Dec. 2010.
- [18] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.
- [19] 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.
- [20] J. Skilling, *Maximum Entropy and Bayesian Methods*. Kluwer Academic, 1989.
- [21] C. R. Rao, "The use and interpretation of principal component analysis in applied research," *Sankhya: The Indian J. Statistics*, vol. 26, pp. 329–358, 1964.
- [22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2008.
- [23] S. Bercker, J. Bobin, and E. J. Candès, "NESTA: a fast and accurate first order method for sparse recovery," *SIAM J. Imaging Sciences*, vol. 4, no. 1, 2011.
- [24] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
- [25] 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.
- [26] "EPFL LUCE SensorScope WSN," last time accessed: June 2012. Available: <http://sensorscope.epfl.ch/>
- [27] "EPFL Grand-St-Bernard SensorScope WSN," last time accessed: June 2012. Available: http://sensorscope.epfl.ch/index.php/Grand-St-Bernard_Deployment
- [28] "CitySense," last time accessed: March 2011. Available: <http://www.citysense.net>
- [29] 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*, vol. 21, no. 12, Oct. 2010.
- [30] R. Masiero, G. Quer, G. Pilonetto, M. Rossi, and M. Zorzi, "Sensing, compression and recovery for wireless sensor networks: sparse signal modelling (part I) and monitoring framework design (part II)," Tech. Rep., 2011. Available: www.dei.unipd.it/~rossi/publications.html
- [31] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [32] D. J. MacKay, "Bayesian interpolation," *Neural Computation J.*, vol. 4, no. 3, pp. 415–447, May 1992.
- [33] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [34] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "Data acquisition through joint compressive sensing and principal component analysis," in *2009 IEEE Globecom*.
- [35] A. Barron, "Entropy and the central limit theorem," *The Annals of Probability*, vol. 14, no. 1, pp. 336–342, Jan. 1986.



Giorgio Quer (S'06, M'12) was born in Asolo, Italy, on July 27th, 1983. He received the B.Sc. degree, the M.Sc. degree (with honors) in Telecommunications Engineering and the Ph.D. degree in Information Engineering from University of Padova, Italy, in 2005, 2007 and 2011, respectively. In 2007 he was a visiting researcher at the Centre for Wireless Communication at University of Oulu, Finland, where he performed research on forwarding and routing techniques for Wireless Sensor Networks.

During his PhD studies his research has been focused on wireless networks optimization, as part of the EU-funded SENSEI project and in collaboration with DOCOMO Euro-Labs, Munich, Germany. In 2010 he was a visiting scholar at California Institute for Telecommunications and Information Technology (Calit2), University of California San Diego (UCSD), working on learning techniques for cognitive networks within the military-funded ARO-CogNet project. Since 2011, he is a postdoc researcher at UCSD. His research interests include wireless network optimization, wireless sensor networks, learning and compression techniques, pervasive healthcare, signal processing and real-time information extraction.



Riccardo Masiero was born in Venice on March 27th, 1983. He received both the Bachelor's degree in Information Engineering and the Master's degree in Telecommunication Engineering from the University of Padova (Italy) in 2005 and 2007, respectively. In April 2011, he also received the PhD degree in Information Engineering from the same institution. During his PhD, he has been supervised by Dr. Michele Rossi and his research activity has been focused on distributed techniques for data collection in Wireless Sensor Networks (WSNs). During 2010,

for six months, he carried on his research activity at INRIA, Sophia Antipolis (France) as a visiting PhD student within the MAESTRO team, under the supervision of Dr. Giovanni Neglia. During that period he focused his activity on distributed optimization techniques for Delay Tolerant Networks (DTNs). Presently, he is a post-doc researcher at SIGNET group, DEI, University of Padova and his research activity is now focused on underwater networking. Within the research project NAUTILUS, he is one of the developer and maintainer of the DESERT Underwater libraries.



Gianluigi Pillonetto was born on January 21, 1975 in Montebelluna (TV), Italy. He received the Laurea degree in Computer Science Engineering cum laude from the University of Padova in 1998 and the PhD degree from the Polytechnic of Milan in 2002. In 2000 and 2002 he was visiting scholar and visiting scientist, respectively, at the Applied Physics Laboratory, University of Washington, Seattle. From 2002 to 2005 he was Research Associate at the Department of Information Engineering, University of Padova. Since 2005 he has been Assistant Professor of Control and Dynamic Systems at the Department of Information Engineering, University of Padova. His research interests are in the field of system identification, stochastic systems, deconvolution problems, nonparametric regularization techniques, learning theory and randomized algorithms.



Michele Rossi received the Laurea degree in Electrical Engineering and the Ph.D. degree in Information Engineering from the University of Ferrara in 2000 and 2004, respectively. From March 2000 to October 2005 he has been a Research Fellow at the Department of Engineering of the University of Ferrara. During 2003 he was on leave at the Center for Wireless Communications (CWC) at the University of California San Diego (UCSD), where he performed research on wireless sensor networks. In November 2005 he joined the Department of

Information Engineering of the University of Padova, Italy, where he is an Assistant Professor. Dr. Rossi is currently involved in the EU-funded IoT-A and SWAP projects, both on wireless sensor networking. His research interests are centered around the dissemination of data in distributed ad hoc and wireless sensor networks, including protocol design through stochastic optimization, integrated MAC/routing schemes, compression techniques for wireless sensor networks and cooperative routing policies for ad hoc networks. Dr. Rossi currently serves on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



Michele Zorzi (S'89, M'95, SM'98, F'07) was born in Venice, Italy, on December 6th, 1966. He received the Laurea and the PhD degrees in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992/93, he was on leave at the University of California, San Diego (UCSD) as a visiting PhD student, working on multiple access in mobile radio networks. In 1993, he joined the faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. After spending three

years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, where he became a Professor in 2000. Since November 2003, he has been on the faculty at the Information Engineering Department of the University of Padova. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, energy constrained communications protocols, broadband wireless access and underwater acoustic communications and networking.

Dr. Zorzi was the Editor-In-Chief of the *IEEE Wireless Communications Magazine* from 2003 to 2005 and the Editor-In-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2011, and currently serves on the Editorial Board of the *Wiley Journal of Wireless Communications and Mobile Computing*. He was also guest editor for special issues in the *IEEE Personal Communications Magazine* (Energy Management in Personal Communications Systems), *IEEE Wireless Communications Magazine* (Cognitive Wireless Networks), and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Multi-media Network Radios, and Underwater Wireless Communications Networks). He served as a Member-at-large of the Board of Governors of the IEEE Communications Society from 2009 to 2011.