

Sensing, Compression and Recovery for Wireless Sensor Networks: Monitoring Framework Design

Giorgio Quer, Riccardo Masiero, Michele Rossi, Michele Zorzi

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

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

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 (sub-sampling) at the Data Collection Point (DCP). To this end, we propose a novel framework, namely, *SCoReI*: Sensing, Compression and Recovery through ON-line Estimation for WSNs. *SCoReI* is very general as it does not require ad-hoc parameter tuning by the user and is able to self-adapt to unpredictable changes in the signal statistics. A feedback control loop is accounted for to estimate, in an on-line fashion, the signal reconstruction error and to react accordingly in order to keep such error bounded. For the actual recovery of the sub-sampled signal, our framework accommodates diverse interpolation techniques, by framing them into the same general algorithm. As a further original contribution of this paper, we integrate a novel signal interpolation method based on Compressive Sensing (CS) into our framework. Specifically, this technique exploits Principal Component Analysis (PCA) for on-line learning of the signal statistics and CS for recovering the sub-sampled signal through convex optimization. Also, we perform an extensive validation of the proposed framework when used in conjunction with CS as well as with standard interpolation techniques, testing its performance for real world signals. Besides validating our framework, these results have the merit of shedding new light on the performance limits of CS when used as a recovery tool in WSNs. Finally, we note that although *SCoReI* is proposed for WSNs, it can be readily applied to other types of network infrastructures that require the approximation of large and distributed signals datasets showing spatial and/or temporal correlation.

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 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 work such as [2]–[6] proposed algorithms that involve collaboration among sensors to implement classical source coding (e.g., see [7]–[9]) in a distributed fashion. New methods for distributed sensing and compression have been developed, instead, based on the recent theory of Compressive Sensing (CS) [10]–[12]. CS is a novel data compression technique that exploits the inherent structure of some input data set to compress it by means of quasi-random matrices; recovery of the original data is achieved solving a convex optimization problem, i.e., an L_1 norm minimization. An early application of CS to wireless sensor networking is [13], where CS is used in a distributed communication scheme for the energy efficient estimation of sensed data in a WSN. In this approach, data packets are directly transmitted by each node to the Data Collection Point (DCP), requiring synchronization among nodes. [14] proposes an interesting application involving CS for fault detection, using a pre-distribution phase (via simple gossiping algorithms), which is however very expensive in terms of number of transmissions. [15] 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. An interesting application for network monitoring exploiting CS is presented in [16], where the aim is to efficiently monitor communication metrics, such as loss or delay, over a set of end-to-end network paths by observing only a subset of them. In [17] 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 that well characterizes different types of signals. In [18] the network topology and the routing used to transport the random projections of the data to the sink have been taken into account to evaluate the possible benefits of CS in realistic multi-hop WSNs. In addition, different transformations have been evaluated in order to meet the sparsity requirements of CS. In a recent work [19], the authors proposed an interesting in-network aggregation technique and exploited CS to reconstruct the data at the sink. Differently to 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 present 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 we all integrate 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. Also, we study the performance achievable by means of a joint use of CS and Principal Component Analysis (PCA) [20] as an interpolation technique. This investigation is based on the results presented in the companion paper [21], where we explain the sparse signal modeling underneath our framework and show that the Laplacian distribution provides an accurate representation of the statistics of the data measured from real WSN testbeds. The main contributions of this paper, instead, are:

- the design of an effective and flexible framework 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 present work is also related to the literature on signal recovery and on Bayesian theory, e.g., see [22]–[24]. Further, we stress that *SCoRel* is proposed for WSNs, but it can be readily applied to other types of network infrastructures that require the approximation of large and distributed datasets with spatial or temporal correlation.

The paper is structured as follows. In Section II we describe our monitoring framework: the distributed sampling method, the data collection techniques and the signal recovery that exploits the joint use of CS and PCA. In Section III we give a mathematical overview of alternative approaches to recover data from an incomplete measurement set and explain how these can be implemented within our monitoring framework for a direct comparison with our solution. This comparison is presented in Section IV for different kinds of real signals gathered from different WSNs. Section V concludes the paper.

II. 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 [21] into an actual monitoring framework for a WSN with N sensor nodes. A diagram showing the logic blocks of this framework is presented in Fig. 1. 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 DCP¹ 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.² Thus, the elements in $\mathbf{y}^{(k)}$ are a subset of those 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

¹The 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.

²The elements equal to 1 indicate which nodes transmit their data sample to the DCP at time k .

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 *SCoReI* framework and are illustrated in Fig. 1.

Wireless Sensor Network (WSN). In this paper we consider data collected from five different WSN deployments, whose sensor reading are available on-line. A brief technical overview of each of these WSNs can be found in [21]. The actual topologies of these deployments can be neglected within our framework, since we capture the correlation among each couple of sensors independently of their relative positions.³ The only requirement is that the sensor nodes can be ordered, e.g., based on their IDs. Multi-hops paths are taken into account for transmission energy computation by assigning a weight to each node proportionally to its distance from the DCP in terms of number of hops.

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 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 DCP is threefold: 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 $\hat{\mathcal{T}}_K = \{\hat{\mathbf{x}}^{(k-K)}, \dots, \hat{\mathbf{x}}^{(k-1)}\}$.

³With alternative recovery schemes, instead, we may need to know the physical coordinates of each node, as e.g., with the Biharmonic Spline discussed in Sec. III.

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 much $\widehat{\mathbf{x}}^{(k)}$ is close 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 a quantity given by

$$\xi_R^{(k)} = \frac{\|\mathbf{x}^{(k)} - \widehat{\mathbf{x}}^{(k)}\|_2}{\|\mathbf{x}^{(k)}\|_2}, \quad (1)$$

where $\widehat{\mathbf{x}}^{(k)}$ is the signal reconstructed at time k by the Recovery block and $\|\cdot\|_2$ is the L_2 norm function for 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 $\widehat{\mathbf{x}}^{(j)}$, for $j \leq k$. Since the quantity $\xi_0^{(k)} = \|\mathbf{y}^{(k)} - \Phi^{(k)}\widehat{\mathbf{x}}^{(k)}\|_2 / \|\mathbf{y}^{(k)}\|_2$ is always zero, due to the fact that the received samples are reconstructed perfectly, i.e., $\Phi^{(k)}\widehat{\mathbf{x}}^{(k)} = \Phi^{(k)}\mathbf{x}^{(k)}$, one might use some heuristics to calculate the error from the past samples. In this paper we use the following formula:⁴

$$\xi^{(k)} = \left\| \left[\begin{array}{c} \mathbf{y}^{(k)} \\ \mathbf{y}^{(k-1)} \end{array} \right] - \left[\begin{array}{c} \Phi^{(k)}\widehat{\mathbf{x}}^{(k-1)} \\ \Phi^{(k-1)}\widehat{\mathbf{x}}^{(k)} \end{array} \right] \right\|_2 \cdot \left(\left\| \left[\begin{array}{c} \mathbf{y}^{(k)} \\ \mathbf{y}^{(k-1)} \end{array} \right] \right\|_2 \right)^{-1}, \quad (2)$$

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., $\widehat{\mathbf{x}}^{(k-1)}$, sampled in the corresponding points, i.e., $\Phi^{(k)}\widehat{\mathbf{x}}^{(k-1)}$. 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

⁴We tried other heuristics and verified through extensive simulation that they perform similarly but slightly worse than the one in Eq. (2). These are therefore not listed here, as they do not provide additional insights.

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:

$$p_{tx}^{(k+1)} = \begin{cases} \min \{ p_{tx}^{(k)} C_1, 1 \} & \text{if } \xi^{(k)} \geq \tau \\ \max \{ p_{tx}^{(k)} - C_2/N, p_{tx}^{\min} \} & \text{if } \xi^{(k)} < \tau . \end{cases} \quad (3)$$

In Section II-A, we provide some insights on the choice of the parameters in Eq. (3).

Recovery: for the description of the recovery scheme, we first consider the joint use of CS and PCA as the interpolation technique, hereafter referred to as CS-PCA. Next, we briefly review the main aspects of CS-PCA which are needed to understand our framework. Further technical details on this can be found in [21]. As anticipated in the introduction, classical interpolation techniques can alternatively be plugged into our framework; this is covered in detail in Section III.

Specifically, CS is the mathematical tool exploited at each time k to perform centralized recovery at the DCP of the signal $\mathbf{x}^{(k)}$ from its sampled version $\mathbf{y}^{(k)} = \Phi^{(k)} \mathbf{x}^{(k)}$. We assume that there exists an invertible $N \times N$ sparsifying matrix Ψ such that $\mathbf{x}^{(k)} = \Psi \mathbf{s}^{(k)}$, so we can write:

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

which is an ill-posed and ill-conditioned system, with $\tilde{\Phi}^{(k)} = \Phi^{(k)} \Psi$ of size $L \times N$, since the number of variables N is larger than the number of equations L and a small variation in $\mathbf{y}^{(k)}$ can cause a large variation in $\mathbf{s}^{(k)}$. However, in the companion paper [21] we verified that, when Ψ is obtained through PCA, $\mathbf{s}^{(k)}$ is a sparse vector for many signals of interest; therefore, we can invert Eq. (4) solving a convex optimization problem, e.g., see [25]. The idea of iteratively exploiting PCA to compute the sparsifying matrix Ψ for CS is the key point of CS-PCA. In standard CS [11], Ψ is assumed to be given and fixed with time, but this is not the case for a realistic WSN scenario, where the statistical characteristics of $\mathbf{x}^{(k)}$, the signal of interest, can vary with time. To cope with this problem, we store the past K reconstructed samples of the signal, at times $j = k - K, \dots, k - 1$, i.e., $\hat{\mathcal{T}}_K = \{\hat{\mathbf{x}}^{(k-K)}, \dots, \hat{\mathbf{x}}^{(k-1)}\}$, and from these samples we estimate the mean vector $\bar{\mathbf{x}}$ and the covariance matrix $\hat{\Sigma}$. According to PCA, we consider the orthonormal matrix $\mathbf{U}^{(k)}$ whose columns are the eigenvectors of $\hat{\Sigma}$, placed in decreasing order with respect to the corresponding eigenvalues. Thus, we define the vector $\mathbf{s}^{(k)}$ as:

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

In [21] we have given empirical evidence of the fact that the vector $\mathbf{s}^{(k)}$ is actually sparse, or equivalently that it is well described by a Laplacian prior. This allows us to recover the monitored vector $\mathbf{x}^{(k)}$ through convex optimization (CS) by choosing Ψ in Eq. (4) as $\mathbf{U}^{(k)}$. To conclude, as recovery mechanism of *SCoReI* we can use CS and PCA in combination. The original signal $\mathbf{x}^{(k)}$ is approximated as follows: 1) find a good estimate⁵ of $\mathbf{s}^{(k)}$, namely $\widehat{\mathbf{s}}^{(k)}$, e.g., using the algorithms in [12], [25] or [26], and 2) apply the following calculation:

$$\widehat{\mathbf{x}}^{(k)} = \overline{\mathbf{x}}^{(k)} + \mathbf{U}^{(k)}\widehat{\mathbf{s}}^{(k)}. \quad (6)$$

A. SCoReI Framework Validation

In order to illustrate the choices made in the design of *SCoReI*, 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. 1.

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 II. Note that for the *2 Phases* technique the training set \mathcal{T}_{K_1} does not contain approximations (reconstructions) of the past signals (whilst for *SCoReI* the training set is $\widehat{\mathcal{T}}_K$, see Section II), but rather contains the entire signal that is collected during the training phase. 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

⁵We can refer to a good estimate of $\mathbf{s}^{(k)}$ as $\widehat{\mathbf{s}}^{(k)}$ such that $\|\mathbf{s}^{(k)} - \widehat{\mathbf{s}}^{(k)}\|_2 \leq \epsilon$, where $\epsilon > 0$ and $\epsilon \ll 1$.

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 eliminate the training phase, so the nodes at each time k transmit with a fixed probability p_{tx} : this is the *Fixed p_{tx}* technique. In the absence of a training phase, we consider as in *SCoRel* the training set formed by the K previously reconstructed signals, i.e., $\widehat{\mathcal{T}}_K$.

To compare *SCoRel* with the above schemes, we use different signals classified according to their statistical characteristics, as detailed in Section IV: **S1)** signals with high temporal and spatial correlation, e.g., ambient temperature [°C] or ambient humidity [%]; **S2)** signals with lower correlation, e.g., luminosity [A/W]; and **S3)** the battery level [V] of the sensor nodes during the signal collection campaign.

The x-axis of Figs. 2 and 3 represents the normalized cost expressed as the average fraction of packet transmissions in the network per time sample, formally:

$$\text{Cost} = \frac{1}{K D_{\text{TOT}}} \sum_{k=1}^K \sum_{n=1}^N D_n I_n(k), \quad (7)$$

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,⁶ $D_{\text{TOT}} = \sum_{n=1}^N D_n$ 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. Conversely, the normalized cost is zero when all nodes remain silent during all time instants. The cost of the feedback transmitted by the DCP is neglected here, since we can reasonably assume that the DCP has no stringent power constraints. The y-axis, instead, shows the signal reconstruction error at the end of the recovery process, calculated according to Eq. (1).

⁶Note that D_n is considered to keep into account the multi-hop structure of the network in the performance analysis.

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 *SCoRel*, we vary the error threshold τ of Eq. (3) 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 *SCoRel* we set the training set length to $K = 2$. Further, for *SCoRel* 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. 4 we show the impact of the choice of K_1 on the performance of *2 Phases*, for the recovery of signal (S1).⁷ 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} . 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. 4 the relative distance between the curve with $K_1 = 2$ and its lower bound is the smallest. In Fig. 5, instead, we show the performance of *Fixed p_{tx}* , that are also representative for *SCoRel*, 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 are not shown here because they do not bring further insights into the understanding of the technique. Fig. 2 shows that with the slowly varying signals of type (S1) 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 (S1), (S2) or (S3), as depicted in Fig. 3. Here, the error for *Fixed*

⁷A similar analysis has been performed also for signals (S2) and (S3).

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. *SCoRel*, 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 motivate the design choices that are at the basis of *SCoRel*.

III. DATA RECOVERY FROM AN INCOMPLETE MEASUREMENT SET

The recovery algorithm is executed at the DCP (see Fig. 1) 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 $L \leq N$. To this end, in Section II 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. 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 what follows, we first review state-of-the-art interpolation techniques that formally solve the following problem:

Problem 3.1 (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 $\mathbf{y}^{(k)} = \Phi^{(k)} \mathbf{x}^{(k)}$, where $\mathbf{y}^{(k)} \in \mathbb{R}^L$, $L \leq N$ and $\Phi^{(k)}$ is an $[L \times N]$ sampling matrix, i.e., all rows of $\Phi^{(k)}$ contain exactly one element equal to 1 and all columns of $\Phi^{(k)}$ contain at most one element equal to 1, whilst all the remaining elements are zero.

At the end of this section, we detail how different interpolation techniques can be implemented within *SCoRel* and we compare their performance for real world WSN signals in Section IV.

A. Signal Models and Interpolation Techniques

The prior knowledge that we have about the signal of interest $\mathbf{x}^{(k)}$ helps us build a model for such signal. This knowledge 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 $\mathbf{x}^{(k)}$. In summary, to compute $\widehat{\mathbf{x}}^{(k)}$ from $\mathbf{y}^{(k)}$, we need a model of $\mathbf{x}^{(k)}$ that can be built according to two different approaches, *deterministic* or *probabilistic*, detailed in Sections III-B and III-C, respectively. The Deterministic approach allows us to define two recovery methods, the Biharmonic Spline (Spline) and the Deterministic Ordinary Least Square (DOLS). The recovery methods that we consider for the probabilistic approach are: Probabilistic Ordinary Least Square (POLS), which assumes a Gaussian distribution of the principal components of the signal and the CS-PCA method, which assumes a Laplacian distribution. The implementation steps for the four recovery methods are detailed in Section III-D.

B. Recovery Methods based on Deterministic Signal Models

Biharmonic Spline Interpolation (Spline): A possible way to think of $\mathbf{x}^{(k)} \in \mathbb{R}^N$ is as a signal whose elements depend on d -dimensional coordinates. To be more concrete, we can think of an environmental monitored signal collected from a WSN of N nodes that we order according to their IDs. Each node i , with $i = \{1, \dots, N\}$, at time k senses a value which is represented by element $x_i^{(k)}$ of vector $\mathbf{x}^{(k)}$. Since the considered node i is deployed in a specific location of the network, it is also linked to a set of geographical coordinates (e.g., latitude and longitude, which can be represented with a $d = 2$ dimensional coordinate vector $\mathbf{c}^{(i)}$). $x_i^{(k)}$ represents the reading of the i -th network node, which in turn is associated with a vector of d coordinates $\mathbf{c}^{(i)}$, and therefore we can express $x_i^{(k)}$ as a function of $\mathbf{c}^{(i)}$, i.e., $x_i^{(k)}(\mathbf{c}^{(i)})$. A straightforward way to model $\mathbf{x}^{(k)}$ is by defining a proper function of the d -dimensional coordinate \mathbf{c} , $\phi(\mathbf{c})$ (e.g., the Green function) that satisfies regularity conditions (e.g., smoothness) inferred by “typical” realizations of the signal of interest $\mathbf{x}^{(k)}$ [27]. Thus, we can write each element i of $\mathbf{x}^{(k)}$ as

$$x_i^{(k)}(\mathbf{c}^{(i)}) \simeq \sum_{j=1}^L \alpha_j \phi(\mathbf{c}^{(i)} - \mathbf{c}^{(j)}), \quad (8)$$

where the function $\phi(\cdot)$ is used as a sort of basis function for $\mathbf{x}^{(k)}$ and α_j is the weight associated with $\phi(\cdot)$ centered in $\mathbf{c}^{(j)}$, with $j = 1, \dots, L$, that is the d -dimensional coordinate corresponding to the physical placement of the node from which we received the j -th measurement.

The Spline method solves Problem 3.1 exploiting the deterministic model in Eq. (8), see [27]; the

objective is to find a biharmonic function that passes through L data points stored in the L -dimensional vector $\mathbf{y}^{(k)}$. In this context, with each element j of the L -dimensional vector $\mathbf{y}^{(k)}$ is associated a d -dimensional index $\mathbf{c}^{(j)} = [c_1^{(j)}, \dots, c_d^{(j)}]^T$. Similarly, with each element i of the N -dimensional vector $\mathbf{x}^{(k)}$ is associated the d -dimensional index $\mathbf{c}^{(i)}$. In order to interpolate the L points in $\mathbf{y}^{(k)}$ we require to satisfy for each element of $\mathbf{x}^{(k)}$ the smoothness condition⁸ $\nabla^4 \widehat{x}^{(k)}(\mathbf{c}) = \sum_{j=1}^L \alpha_j \delta(\mathbf{c} - \mathbf{c}^{(j)})$, given that, if $\mathbf{c} = \mathbf{c}^{(j)}$ then $\widehat{x}^{(k)}(\mathbf{c}^{(j)}) = y_j^{(k)}$, where $\mathbf{c}^{(j)}$ is the coordinate vector $\mathbf{c}^{(j)} = [c_1^{(j)}, \dots, c_d^{(j)}]^T \in \mathbb{R}^d$ related to the reading $y_j^{(k)}$ (e.g., the geographical location of the reading $y_j^{(k)}$). The solution is proved to be:

$$\widehat{x}^{(k)}(\mathbf{c}) = \sum_{j=1}^L \alpha_j \phi_d(\mathbf{c} - \mathbf{c}^{(j)}), \quad (9)$$

where $\phi_d(\cdot)$ is the Green function for the d -dimensional problem.⁹ The constants $\alpha_1, \dots, \alpha_L$ are found by solving the linear system $y_i^{(k)} = \sum_{l=1}^L \alpha_l \phi_d(\mathbf{c}^{(j)} - \mathbf{c}^{(l)}), \forall j \in \{1, \dots, L\}$. To conclude, the solution $\widehat{\mathbf{x}}^{(k)} \in \mathbb{R}^N$ is the vector whose element i is equal to $\widehat{x}^{(k)}(\mathbf{c}^{(i)})$, namely, the recovered value associated with the d -dimensional index $\mathbf{c}^{(i)}$.

Deterministic Ordinary Least Square (DOLS): an alternative way to determine a model for $\mathbf{x}^{(k)}$ allows us to abstract from the knowledge of where the signal sources are placed. Further, this second method is adaptable to the spatio-temporal correlation and structure of the signal. Observing that generally a physical phenomenon is correlated in time and that its spatial correlation can be considered as stationary over a given time period (e.g., from $k - K$ until k), a natural way to proceed is by assuming that $\mathbf{x}^{(k)}$ lies in the vector space spanned by the K previous samples contained in the training set $\mathcal{T}_K^{(k)}$ (or in the approximate training set $\widehat{\mathcal{T}}_K^{(k)}$), i.e., in $\text{span} \langle \mathcal{T}_K^{(k)} \rangle$. According to the formalism introduced in [21], let us refer to the temporal mean and the covariance matrix of the elements in $\mathcal{T}_K^{(k)}$ as $\bar{\mathbf{x}}^{(k)}$ and $\widehat{\Sigma}^{(k)}$, respectively. Let us consider also the ordered set $\mathcal{U}^{(k)} = \{\mathbf{u}_1^{(k)}, \dots, \mathbf{u}_N^{(k)}\}$ of unitary eigenvectors of $\widehat{\Sigma}^{(k)}$, placed according to the decreasing order of the corresponding eigenvalues. Let $\mathbf{U}_M^{(k)}$ be the $[N \times M]$ matrix whose columns are the first M elements of $\mathcal{U}^{(k)}$. A model of $\mathbf{x}^{(k)}$, assuming that this one lies

⁸Here, ∇^4 is the biharmonic operator which allows to formalize regularity conditions on the fourth-order derivatives; $\delta(\cdot)$ is defined as $\delta(x) = 1$ if $x = 0$, $\delta(x) = 0$ otherwise.

⁹E.g., $\phi_1(\mathbf{c}) = |\mathbf{c}|^3$, $\phi_2(\mathbf{c}) = |\mathbf{c}|^2(\ln |\mathbf{c}| - 1)$ and $\phi_3(\mathbf{c}) = |\mathbf{c}|$.

in $\text{span} \langle \mathcal{T}_K^{(k)} \rangle$, can be written as:

$$\mathbf{x}^{(k)} \simeq \bar{\mathbf{x}}^{(k)} + \mathbf{V}^{(k)} \mathbf{s}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}_M^{(k)} \mathbf{s}^{(k)}, \quad (10)$$

where, in general, $\mathbf{V}^{(k)}$ can be any $[N \times M]$ matrix of orthonormal columns (obtained at time k from the set $\{\mathbf{x}^{(k-K)} - \bar{\mathbf{x}}^{(k)}, \dots, \mathbf{x}^{(k-1)} - \bar{\mathbf{x}}^{(k)}\}$, e.g., through the Gram-Schmidt process [28]), with $M \leq N$; here we set $\mathbf{V}^{(k)} = \mathbf{U}_M^{(k)}$ because given $M \leq N$, the best way to represent with M components each element out of a set of N -dimensional elements is through PCA. In order to show that PCA is the solution to this representation problem, we should look at it from a geometric point of view. We can consider each sample $\mathbf{x}^{(k)}$, for all k , as a point in \mathbb{R}^N and look for the M -dimensional plane (with $M \leq N$) which provides the best fit to all the elements in $\mathcal{T}_K^{(k)}$, and therefore for all the vectors that lie in $\text{span} \langle \mathcal{T}_K^{(k)} \rangle$, in terms of minimum Euclidean distance. The key point of PCA is the Ky Fan theorem [29], here reported to better illustrate the considered deterministic approach.

Theorem 3.1 (Ky Fan Theorem): Let $\Sigma \in \mathbb{R}^{N \times N}$ be a symmetric matrix, let $\lambda_1 \geq \dots \geq \lambda_N$ be its eigenvalues and $\mathbf{u}_1, \dots, \mathbf{u}_N$ the corresponding eigenvectors (assumed to be orthonormal, without loss of generality). Given M orthonormal vectors $\mathbf{b}_1, \dots, \mathbf{b}_M$ in \mathbb{R}^N , with $M \leq N$, it holds that

$$\max_{\mathbf{b}_1, \dots, \mathbf{b}_M} \sum_{j=1}^M \mathbf{b}_j^T \Sigma \mathbf{b}_j = \sum_{j=1}^M \lambda_j, \quad (11)$$

and the maximum is attained for $\mathbf{b}_i = \mathbf{u}_i, \forall i$.

According to the Ky Fan Theorem, maximizing $\sum_{j=1}^M \mathbf{b}_j^T \widehat{\Sigma}^{(k)} \mathbf{b}_j$ corresponds to finding the linear transformation $\mathcal{F}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ that maximally preserves the information contained in the training set $\mathcal{T}_K^{(k)}$. In other words, this corresponds to maximizing the variance of the M -dimensional (linear) approximation of each element in $\text{span} \langle \mathcal{T}_K^{(k)} \rangle$ that is strictly related to the information content of each signal in $\mathcal{T}_K^{(k)}$. Because of Theorem 3.1, the best M -dimensional approximation of each $\mathbf{x} \in \text{span} \langle \mathcal{T}_K^{(k)} \rangle$ is [20]

$$\widehat{\mathbf{x}} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}_M^{(k)} \left(\mathbf{U}_M^{(k)} \right)^T (\mathbf{x} - \bar{\mathbf{x}}^{(k)}),$$

where $\left(\mathbf{U}_M^{(k)} \right)^T (\mathbf{x} - \bar{\mathbf{x}}^{(k)})$ is the projection of $\mathbf{x} - \bar{\mathbf{x}}^{(k)}$ onto its best fitting M -dimensional plane. To sum up, if the point of interest $\mathbf{x}^{(k)} \in \text{span} \langle \mathcal{T}_K^{(k)} \rangle$, we can transform it into a point $\mathbf{s}^{(k)} \in \mathbb{R}^M$ as follows:

$$\mathbf{s}^{(k)} \stackrel{\text{def}}{=} \left(\mathbf{U}_M^{(k)} \right)^T (\mathbf{x}^{(k)} - \bar{\mathbf{x}}^{(k)}). \quad (12)$$

Multiplication of Eq. (12) by $\mathbf{U}_M^{(k)}$ and summation with the sample mean return the best approximation of the original vector, in accordance to Eq. (10).

To solve Problem 3.1 exploiting the model in Eq. (10) we can simply use the Ordinary Least Square (OLS) method [22], thus we refer to this recovery solution as Deterministic Ordinary Least Square. From $\mathbf{y}^{(k)} = \Phi^{(k)}\mathbf{x}^{(k)}$ and the assumption that Eq. (10) holds, we can write

$$\mathbf{y}^{(k)} = \Phi^{(k)}(\bar{\mathbf{x}}^{(k)} + \mathbf{U}_M^{(k)}\mathbf{s}^{(k)}) . \quad (13)$$

The ordinary least square solution of Eq. (13) is given by

$$\hat{\mathbf{s}}^{(k)} = (\Phi^{(k)}\mathbf{U}_M^{(k)})^\dagger(\mathbf{y}^{(k)} - \Phi^{(k)}\bar{\mathbf{x}}^{(k)}) \quad (14)$$

and allows us to estimate the signal $\mathbf{x}^{(k)}$ as $\hat{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}_M^{(k)}\hat{\mathbf{s}}^{(k)}$. In the above expression the symbol \dagger indicates the Moore-Penrose pseudo-inverse matrix. Recalling that in Eq. (13) $\mathbf{y}^{(k)}$ is an $[L \times 1]$ vector whilst $\mathbf{s}^{(k)}$ is an $[M \times 1]$ with $M \leq L$, the system in Eq. (13) is in general overdetermined and may have no solutions (e.g., when all the L measurements are linearly independent). In this case Eq. (14) minimizes $\|(\mathbf{y}^{(k)} - \Phi^{(k)}\bar{\mathbf{x}}^{(k)}) - \Phi^{(k)}\mathbf{U}_M^{(k)}\mathbf{s}^{(k)}\|_2$, obtaining $\hat{\mathbf{s}}^{(k)}$ as the nearest (according to the Euclidean norm) possible vector to all the L collected measurements. If $L = M$, instead, the Moore-Penrose pseudo-inverse coincides with the inverse matrix and $\hat{\mathbf{s}}^{(k)}$ is uniquely determined.

C. Recovery Methods based on Probabilistic Signal Models

The probabilistic approach allows us to introduce an uncertainty in the model of $\mathbf{x}^{(k)}$ and, as we show in Section IV, this translates into a significant improvement in the recovery performance.

Considering Eq. (10), this can be reformulated as:

$$\mathbf{x}^{(k)} \simeq \bar{\mathbf{x}}^{(k)} + \mathbf{V}^{(k)}\mathbf{s}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}^{(k)}\mathbf{s}^{(k)} , \quad (15)$$

where $\mathbf{V}^{(k)}$ is now an $[N \times N]$ matrix of orthonormal columns set equal to the PCA matrix $\mathbf{U}^{(k)}$ following the same rationale as above. Here, the cardinality of the model's parameters is N (i.e., the size of vector $\mathbf{s}^{(k)}$), which is surely larger than or equal to the dimension of $\text{span}\langle \mathcal{T}_K^{(k)} \rangle$. The model in Eq. (15), therefore, allows us to account for the fact that $\mathbf{x}^{(k)}$ could not perfectly lie in $\text{span}\langle \mathcal{T}_K^{(k)} \rangle$. This kind of approach has been implicitly adopted also in [21] and, as shown in that paper, we need

further assumptions on the system input $\mathbf{s}^{(k)}$ to fully characterize the model in Eq. (15), i.e., we have to assign a prior to $\mathbf{s}^{(k)}$. In practice, $\mathbf{s}^{(k)}$ is a vector random process that we can assume to be, e.g., a Gaussian multivariate process or a Laplacian vector process with i.i.d. components, see [21].

CS-PCA: when we assign a Laplacian prior to $\mathbf{s}^{(k)}$, we can solve Problem 3.1 through our CS-PCA scheme that corresponds to minimizing $\|\mathbf{s}^{(k)}\|_1$, given that $\mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)}\mathbf{\Psi}\mathbf{s}^{(k)}$, as shown in [21].

Probabilistic Ordinary Least Square Method (POLLS): alternatively, when we assign a Gaussian prior to $\mathbf{s}^{(k)}$, we can solve Problem 3.1 again via the OLS. In this case, we just have to rewrite Eq. (14) as

$$\widehat{\mathbf{s}}^{(k)} = (\mathbf{\Phi}^{(k)}\mathbf{U}^{(k)})^\dagger(\mathbf{y}^{(k)} - \mathbf{\Phi}^{(k)}\bar{\mathbf{x}}^{(k)}). \quad (16)$$

In this equation, the size of $\mathbf{y}^{(k)}$ is smaller than that of $\mathbf{s}^{(k)}$, i.e., $L < N$. Therefore, Eq. (16) is the solution of an ill-posed system, which theoretically allows an infinite number of solutions. Nevertheless, a multivariate Gaussian prior on $\mathbf{s}^{(k)}$ with zero mean and independent components¹⁰, i.e., $p(\mathbf{s}^{(k)}) \sim \mathcal{N}(0, \mathbf{\Sigma}_s)$ where $\mathbf{\Sigma}_s$ is a diagonal matrix, makes it possible to choose, among all the possible solutions, the one estimated as¹¹

$$\begin{aligned} \widehat{\mathbf{s}}^{(k)} &= \arg \max_{\mathbf{s}^{(k)}} p(\mathbf{s}^{(k)}|\mathbf{y}^{(k)}) = \arg \max_{\mathbf{s}^{(k)}} p(\mathbf{y}^{(k)}|\mathbf{s}^{(k)})p(\mathbf{s}^{(k)}) \\ &= \arg \max_{\mathbf{s}^{(k)}} \delta(\mathbf{y}^{(k)}, \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)}) \frac{1}{(2\pi)^{\frac{L}{2}} \det(\mathbf{\Sigma}_s)^{\frac{L}{2}}} \exp \left\{ -\frac{\|\mathbf{\Sigma}_s\mathbf{s}^{(k)}\|_2^2}{2} \right\} \\ &= \arg \min_{\mathbf{s}^{(k)}} \|\mathbf{\Sigma}_s\mathbf{s}^{(k)}\|_2^2, \quad \text{given that } \mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)} \end{aligned} \quad (17)$$

that corresponds to the solution in Eq. (16), namely $\min \|\mathbf{s}^{(k)}\|_2$ given that $\mathbf{y}^{(k)} = \mathbf{\Phi}^{(k)}\mathbf{U}^{(k)}\mathbf{s}^{(k)}$.

D. Implementation of Signal Recovery Methods

Each of the interpolation techniques explained above can be implemented at the DCP in the Recovery block shown in Fig. 1. As previously remarked, at each time sample k , we can think of $\mathbf{x}^{(k)}$ as an N -dimensional signal whose elements depend on coordinates in d dimensions. If we measure $\mathbf{x}^{(k)}$ in

¹⁰Note that, when obtained through Eq. (12), $\mathbf{s}^{(k)}$ is the vector of principal components of $\mathcal{T}_K^{(k)}$: these are known to be uncorrelated and therefore, under the assumption of gaussianity and zero mean, they are also independent.

¹¹We recall here that, in Eq. (17), $\delta(\cdot)$ is a function defined as: $\delta(\mathbf{x}, \mathbf{y}) = 1$ if $\mathbf{x} = \mathbf{y}$, $\delta(\mathbf{x}, \mathbf{y}) = 0$ otherwise.

L different coordinate points, collecting the measurement set $\{y_1^{(k)}, \dots, y_L^{(k)}\}$, for the recovery stage we can proceed as follows, using a deterministic approach:

1) Biharmonic Spline (**Spline**)

- a) compute $\alpha_1, \dots, \alpha_M$ solving $y_j^{(k)}(\mathbf{c}^{(j)}) = \sum_{l=1}^M \alpha_l \phi_d(\mathbf{c}^{(j)} - \mathbf{c}^{(l)}) \quad \forall j \in 1, \dots, M$;
- b) estimate $x_i^{(k)}$ as $\hat{x}_i^{(k)}(\mathbf{c}^{(i)}) = \sum_{j=1}^M \alpha_j \phi_d(\mathbf{c}^{(i)} - \mathbf{c}^{(j)}) \quad \forall i \in 1, \dots, N$.

If we assume to know the K previous samples $\mathcal{T}_K^{(k)}$ or the approximate training set $\hat{\mathcal{T}}_K^{(k)}$, with $K \leq N$, we can abstract from the knowledge of the physical coordinates associated to $\mathbf{x}^{(k)}$. In this case we need to compute the PCA matrix $\mathbf{U}^{(k)}$ from $\mathcal{T}_K^{(k)}$ (or $\hat{\mathcal{T}}_K^{(k)}$). Then, knowing the matrix $\mathbf{U}^{(k)}$ and knowing also the sampled signal $\mathbf{y}^{(k)} = \Phi^{(k)} \mathbf{x}^{(k)}$, we can use an alternative deterministic approach and set $M = K - 1$. At each time k we can estimate $\mathbf{x}^{(k)}$ according to:

2) Deterministic Ordinary Least Square (**DOLS**)

- a) estimate $\mathbf{s}^{(k)}$ as $\hat{\mathbf{s}}^{(k)} = (\Phi^{(k)} \mathbf{U}_{K-1}^{(k)})^\dagger (\mathbf{y}^{(k)} - \Phi^{(k)} \bar{\mathbf{x}}^{(k)})$;
- b) estimate $\mathbf{x}^{(k)}$ as $\hat{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}_{K-1}^{(k)} \hat{\mathbf{s}}^{(k)}$.

If we adopt a probabilistic approach, we can implement one of the two methods described above, depending on the statistical distribution that we assume for the principal components of signal $\mathbf{x}^{(k)}$:

3) Probabilistic Ordinary Least Square (**POLS**)

- a) estimate $\mathbf{s}^{(k)}$ as $\hat{\mathbf{s}}^{(k)} = (\Phi^{(k)} \mathbf{U}^{(k)})^\dagger (\mathbf{y}^{(k)} - \Phi^{(k)} \bar{\mathbf{x}}^{(k)})$;
- b) estimate $\mathbf{x}^{(k)}$ as $\hat{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}^{(k)} \hat{\mathbf{s}}^{(k)}$.

4) Joint CS and PCA (**CS-PCA**)

- a) estimate $\mathbf{s}^{(k)}$ as $\hat{\mathbf{s}}^{(k)} = \arg \min_{\mathbf{s}^{(k)}} \|\mathbf{s}^{(k)}\|_1$, given that $\mathbf{y}^{(k)} = \Phi^{(k)} \mathbf{U}^{(k)} \mathbf{s}^{(k)}$;
- b) estimate $\mathbf{x}^{(k)}$ as $\hat{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}^{(k)} + \mathbf{U}^{(k)} \hat{\mathbf{s}}^{(k)}$.

The performance of the four different reconstruction techniques is compared in the next section.

IV. PERFORMANCE ANALYSIS

In this section we analyze the performance of *S-CoRel* when used in conjunction with the interpolation methods of Section III. First, we analyze the statistics of all the signals gathered from the WSN deployments described in [21] and we choose a relevant subset of them for our performance analysis. Then, we investigate the performance of the considered interpolation techniques.

Signals: We considered five different types of WSNs, described in [21], each sensing different types of signals for a total of 24 signals. For each signal $\mathbf{x}^{(k)} \in \mathbb{R}^N$, we calculated the average inter-node (spatial) correlation $\rho_s(\mathbf{x}^{(k)})$, defined as the average correlation between the one dimensional signal sensed by node i , $x_i(k)$, and the one sensed by node j , $x_j(k)$, for all the node pairs i, j , formally

$$\rho_s(\mathbf{x}^{(k)}) = \sum_{k=1}^K \frac{1}{K} \sum_{i=1}^N \sum_{j>i} \frac{\left(x_i^{(k)} - E[x_i]\right) \left(x_j^{(k)} - E[x_j]\right)}{\left((N^2 - N)/2\right) \sigma_{x_i} \sigma_{x_j}}. \quad (18)$$

$\rho_s(\mathbf{x}^{(k)})$ gives us a measure of the expected sparsity of the principal components $\mathbf{s}^{(k)} \in \mathbb{R}^N$. If we calculate the principal components of a signal with maximum inter-node correlation, i.e., $\rho_s(\mathbf{x}^{(k)}) = 1$, 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 $\rho_s = 0$, we will obtain a signal $\mathbf{s}^{(k)}$ with no negligible components (with respect to the overall energy of the signal). In Fig. 6 we depict the inter-node correlation for all the signals considered and we group them according to the signal type, i.e, Temperature, Humidity, Solar Radiation, Luminosity (indoor), Wind and Voltage. We note that Temperature, Humidity and Solar Radiation have, on average, a high inter-node correlation ($\rho_s(\mathbf{x}^{(k)}) \simeq 0.7$), while Luminosity, Wind and Voltage have a low inter-node correlation ($\rho_s(\mathbf{x}^{(k)}) \simeq 0.25$). To further analyze these signals, we consider the intra-node (temporal) correlation $\rho_m(\mathbf{x}^{(k)})$, 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

$$\rho_m(\mathbf{x}^{(k)}) = \sum_{i=1}^N \frac{1}{N} \frac{\sum_{k=1}^K \left(x_i^{(k)} - E[x_i]\right) \left(x_i^{(k+m)} - E[x_i]\right)}{K \sigma_{x_i}^2}. \quad (19)$$

For representation purposes, we choose one signal for each type, within the 24 signals depicted in Fig. 6, and we represent for each chosen signal the temporal correlation $\rho_m(\mathbf{x}^{(k)})$, for $m = 1, \dots, 8$ in Fig. 7. We notice that Temperature, Humidity and Solar Radiation show a high intra-node correlation even for $m = 8$ ($\rho_8(\mathbf{x}^{(k)}) \geq 0.85$), while for Luminosity and Wind the temporal correlation quickly decreases ($\rho_8(\mathbf{x}^{(k)}) \leq 0.65$). The Voltage signal, instead, has different characteristics, since even if it has inter-node and intra-node correlation similar to Luminosity and Wind, it is a nearly constant signal.

Among all considered datasets, for our experimental campaign 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 [30], 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 dataset 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 *SCoRel*. Specifically, we considered 5 signals divided according to their statistical characteristics: **S1**) two signals with high temporal and spatial correlation, i.e., ambient temperature [$^{\circ}\text{C}$] and ambient humidity [%]; **S2**) two signals with lower correlation, i.e., luminosity [A/W] in the range 320 – 730 nm and in the range 320 – 1100 nm; **S3**) the battery level [V] of the sensor nodes during the signal collection campaign. 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.

Performance of the Signal Recovery Methods: In the following, we show performance curves for the different interpolation techniques illustrated in Section III: Biharmonic Spline (Spline), Deterministic Ordinary Least Square (DOLS), Probabilistic Ordinary Least Square (POLS) and joint CS and PCA (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 cost expressed as the average fraction of packet transmissions in the network per time sample, again computed according to Eq. (7); the y-axis shows the signal reconstruction error at the end of the recovery process, calculated according to Eq. (1). To vary the cost (x-axis) we modify the parameters of *SCoRel* as explained in Section II. Solid and dotted lines represent lower bounds on the error recovery performance, which are obtained by assuming to have 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 $\widehat{\mathcal{T}}_K$.

In Figs. 8 and 9 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. It is also interesting to note that using $\widehat{\mathcal{T}}_K$ POLS outperforms CS-PCA, whilst with \mathcal{T}_K CS-PCA and POLS

perform equally good, also for highly variable signals, see Fig. 9, signal S2. In fact 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 [31]). Nevertheless, both POLS and CS-PCA remain valid solutions for a monitoring application framework, since the performance loss from the ideal case, which assumes perfect knowledge of \mathcal{T}_K , to the one that exploits $\widehat{\mathcal{T}}_K$ is sufficiently small. Concerning Spline, this method allows to reach good performance only above a transmission probability of 0.8; furthermore, the use of Spline as the interpolation technique in *SCoReI* 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. 10 we show similar performance curves using the signals gathered from the EPFL WSN deployment LUCE, see [21] and [32]. The signals considered in this figure are of the class S1, i.e., temperature and humidity. Also in this case, the performance are similar to Fig. 8 and all the above observations remain valid. This provides further evidence that *SCoReI* 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.

V. CONCLUSIONS

In this paper we proposed a novel framework, called *SCoReI*, for the accurate approximation of large real world WSN signals through the collection of a small fraction of data points. *SCoReI* 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. As an original contribution of the paper, we considered an interpolation technique based on Compressive Sensing (CS), utilizing Principal Component Analysis (PCA) to learn the data statistics and CS to recover the signal through convex optimization. Using real world signals, we have shown that our technique achieves good performance in terms of reconstruction accuracy *vs* network cost (i.e., number of transmissions required). 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 *ACM MOBICOM*, Atlanta, GA, USA, Sep. 2002.
- [2] S. Servetto, "Distributed Signal Processing Algorithms for the Sensor Broadcast Problem," in *Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, Mar. 2003.
- [3] —, "Sensing Lena - Massively Distributed Compression of Sensor Images," in *IEEE Conference on Image Processing (ICIP)*, Ithaca, NY, USA, Sep. 2003.
- [4] S. Pradhan and K. Ramchandran, "Distributed Source Coding Using Syndromes (DISCUS): Design and Construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.
- [5] H. Luo and G. Pottie, "Routing Explicit Side Information for Data Compression in Wireless Sensor Networks," in *Distributed Computing in Sensor Systems (DCOSS)*, Marina del Rey, CA, USA, Jun. 2005.
- [6] M. Gastpar, P. Dragotti, and M. Vetterli, "The Distributed Karhunen-Loeve Transform," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 5177–5196, Dec. 2006.
- [7] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley, 1991.
- [9] Z. Xiong, A. Liveris, and S. Cheng, "Distributed Source Coding for Sensor Networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–92, Sep. 2004.
- [10] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 4036–4048, Apr. 2006.
- [11] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [12] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [13] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3629–3653, Oct. 2007.
- [14] M. Rabbat, J. Haupt, A. Singh, and R. Novak, "Decentralized Compression and Predistribution via Randomized Gossiping," in *Information Processing in Sensor Networks (IPSN)*, Nashville, Tennessee, USA, Apr. 2006.
- [15] 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 *NASA Earth Science Technology Conference (ESTC2008)*, University of Maryland, MD, USA, Jun. 2008.
- [16] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed Network Monitoring for IP and All-optical Networks," in *Internet Measurement Conference (IMC)*, San Diego, CA, USA, Oct. 2007.
- [17] M. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal Distributed Sensing via Random Projections," in *Information Processing in Sensor Networks (IPSN)*, Nashville, TN, USA, Apr. 2006.
- [18] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the Interplay Between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks," in *Information Theory and Applications Workshop (ITA 2009)*, San Diego, CA, US, Feb. 2009.
- [19] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Efficient Measurement Generation and Pervasive Sparsity for Compressive Data Gathering," *IEEE Trans. on Wireless Communications*, vol. 9, no. 12, pp. 3728–3738, Dec. 2010.
- [20] C. R. Rao, "The Use and Interpretation of Principal Component Analysis in Applied Research," *Sankhya: The Indian Journal of Statistics*, vol. 26, pp. 329–358, 1964.
- [21] R. Masiero, G. Quer, G. Pillonetto, M. Rossi, and M. Zorzi, "Sampling and Recovery with Compressive Sensing in Real Wireless Sensor Networks," *Submitted to IEEE Trans. on Wireless Communication*, 2011.
- [22] K.M. Hanson, *Bayesian and Related Methods in Image Reconstruction from Incomplete Data*. H. Stark, ed., 1987.
- [23] 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.
- [24] J. Skilling, *Maximum Entropy and Bayesian Methods*. Kluwer academic, 1989.
- [25] S. Bercker, J. Bobin, and E. J. Candès, "NESTA: a fast and accurate first order method for sparse recovery." *Submitted for publication*. [Online]. Available: <http://www-stat.stanford.edu/~candes/papers/NESTA.pdf>
- [26] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
- [27] D. T. Sandwell, "Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data," *Geophysical Research Letters*, vol. 14, no. 2, pp. 139–142, 1987.
- [28] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
- [29] K. Fan, "On a theorem of Weil concerning eigenvalues of linear transformation I," *Proc. of the National Academy of Sciences*, vol. 35, pp. 652–655, 1949.
- [30] 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.
- [31] A. Barron, "Entropy and the Central Limit Theorem," *The Annals of Probability*, vol. 14, no. 1, pp. 336–342, Jan. 1986.
- [32] "EPFL LUCE SensorScope WSN," Last time accessed: March 2011. [Online]. Available: <http://sensorscope.epfl.ch/>

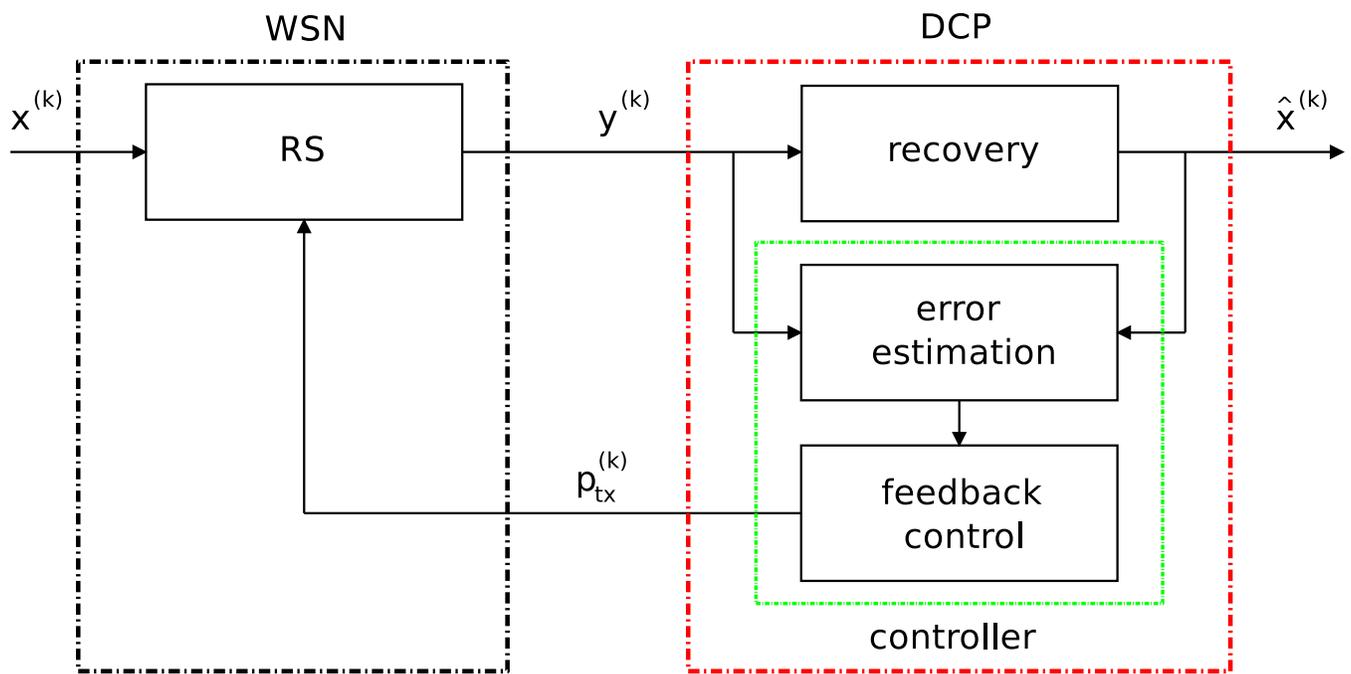


Fig. 1. Diagram of the proposed sensing, compression and recovery scheme. Note that the *Controller*, which includes the *Error estimator* and the *Feedback Control* blocks, is a characteristic of *SCoRel* and is not present in the other DC techniques.

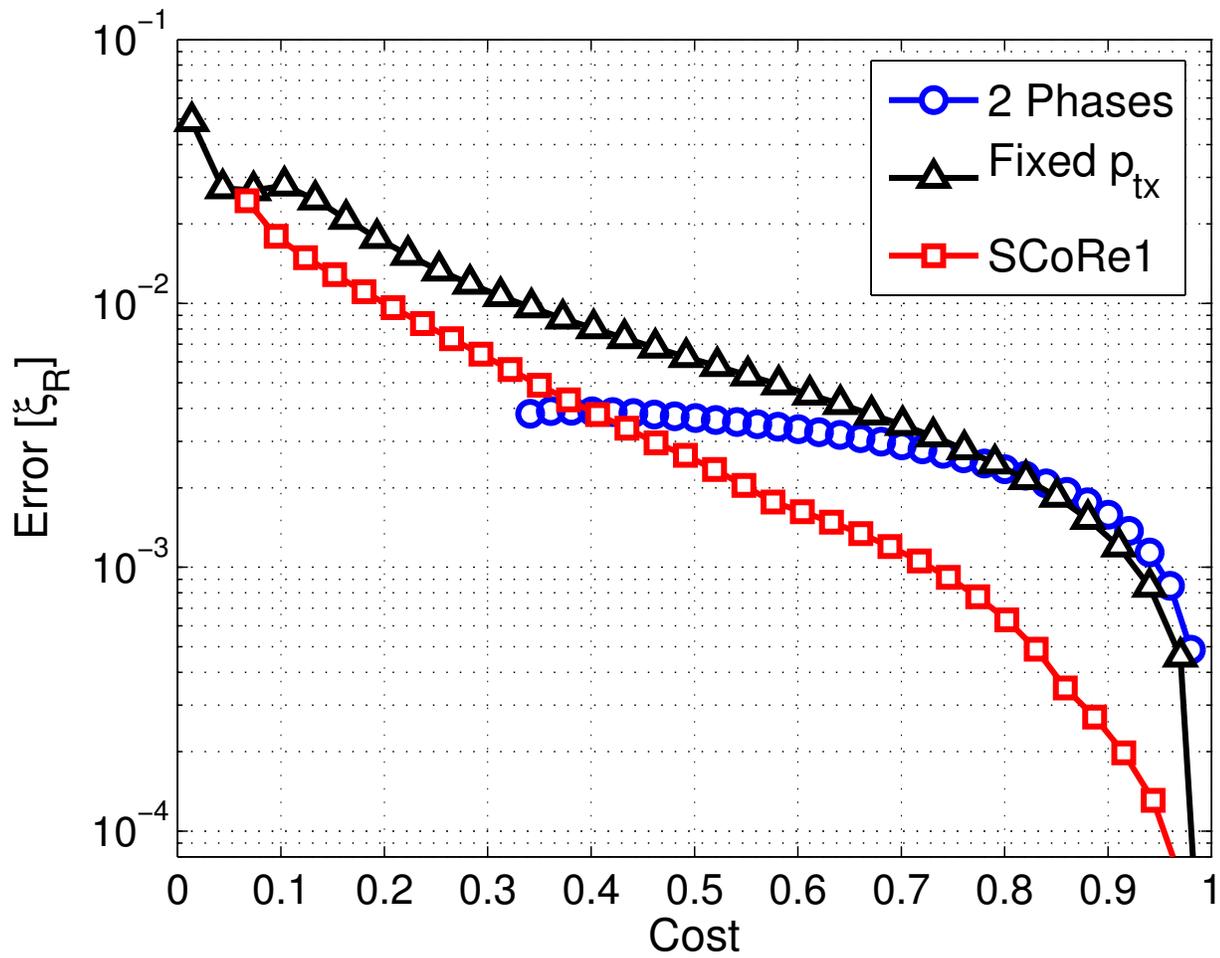


Fig. 2. Performance comparison of three iterative monitoring schemes for signals (S1), temperature and humidity.

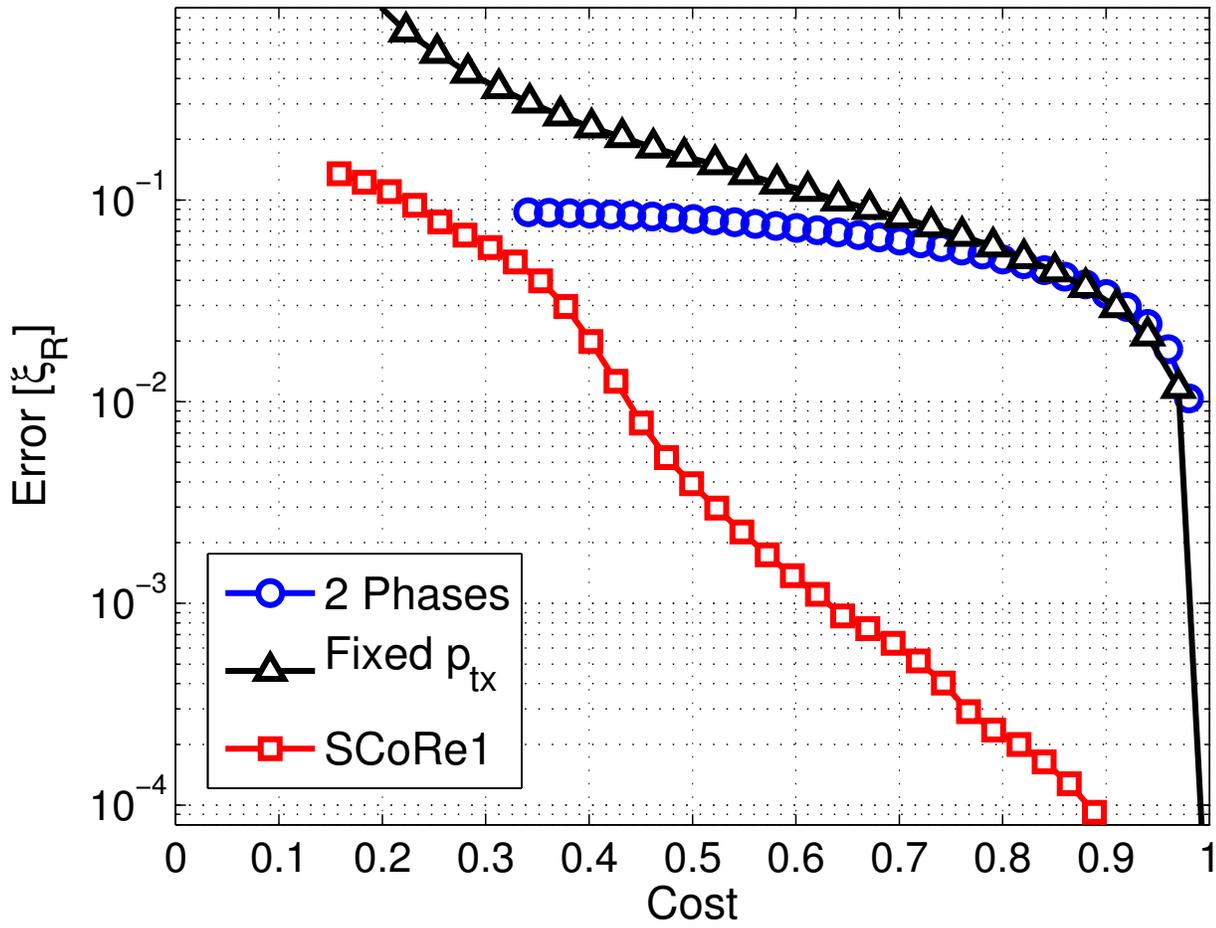


Fig. 3. Average recovery performance of the three iterative monitoring schemes for signals (S1), (S2) and (S3).

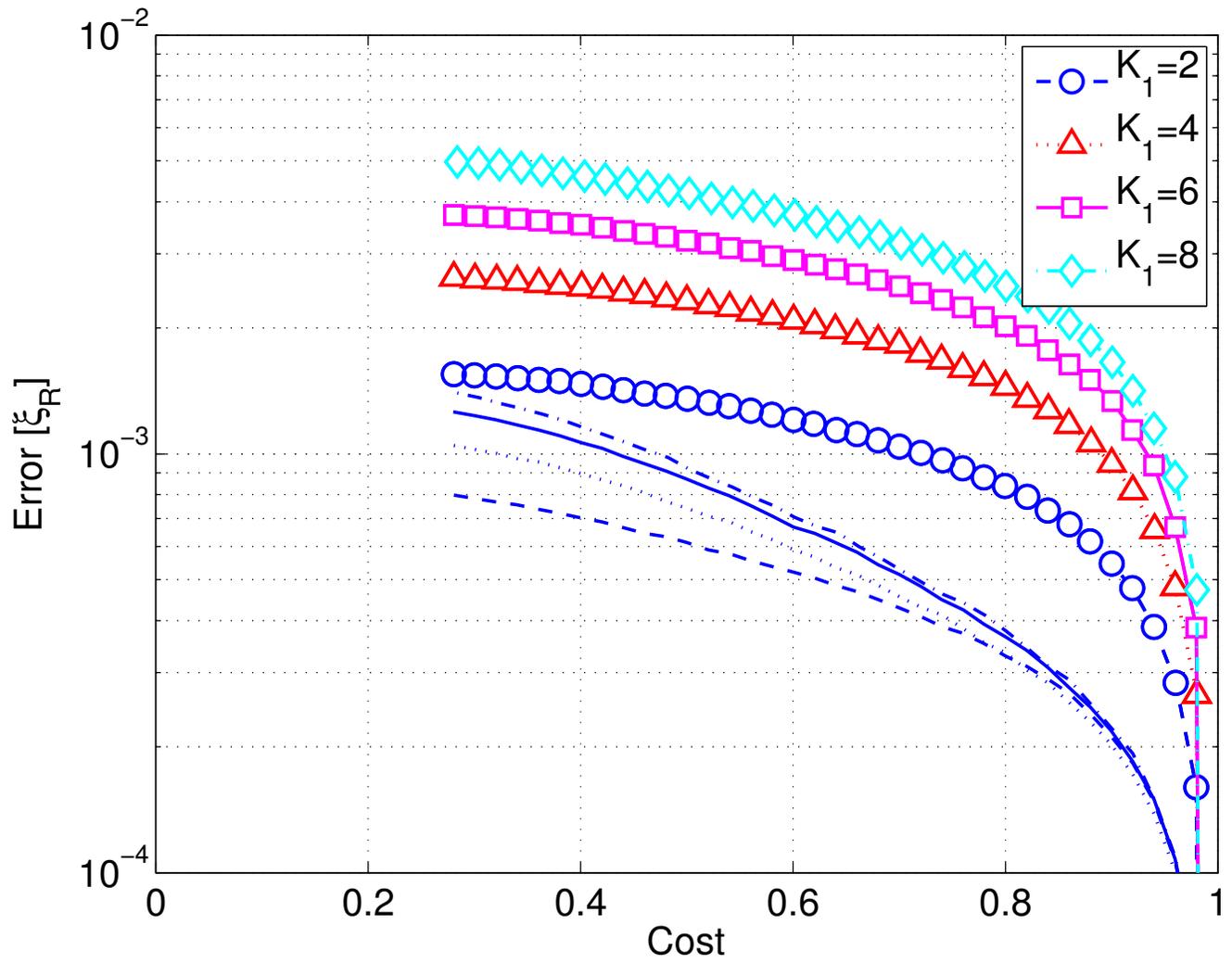


Fig. 4. Impact of the choice of K_1 on the performance of 2 Phases, for signal (S1).

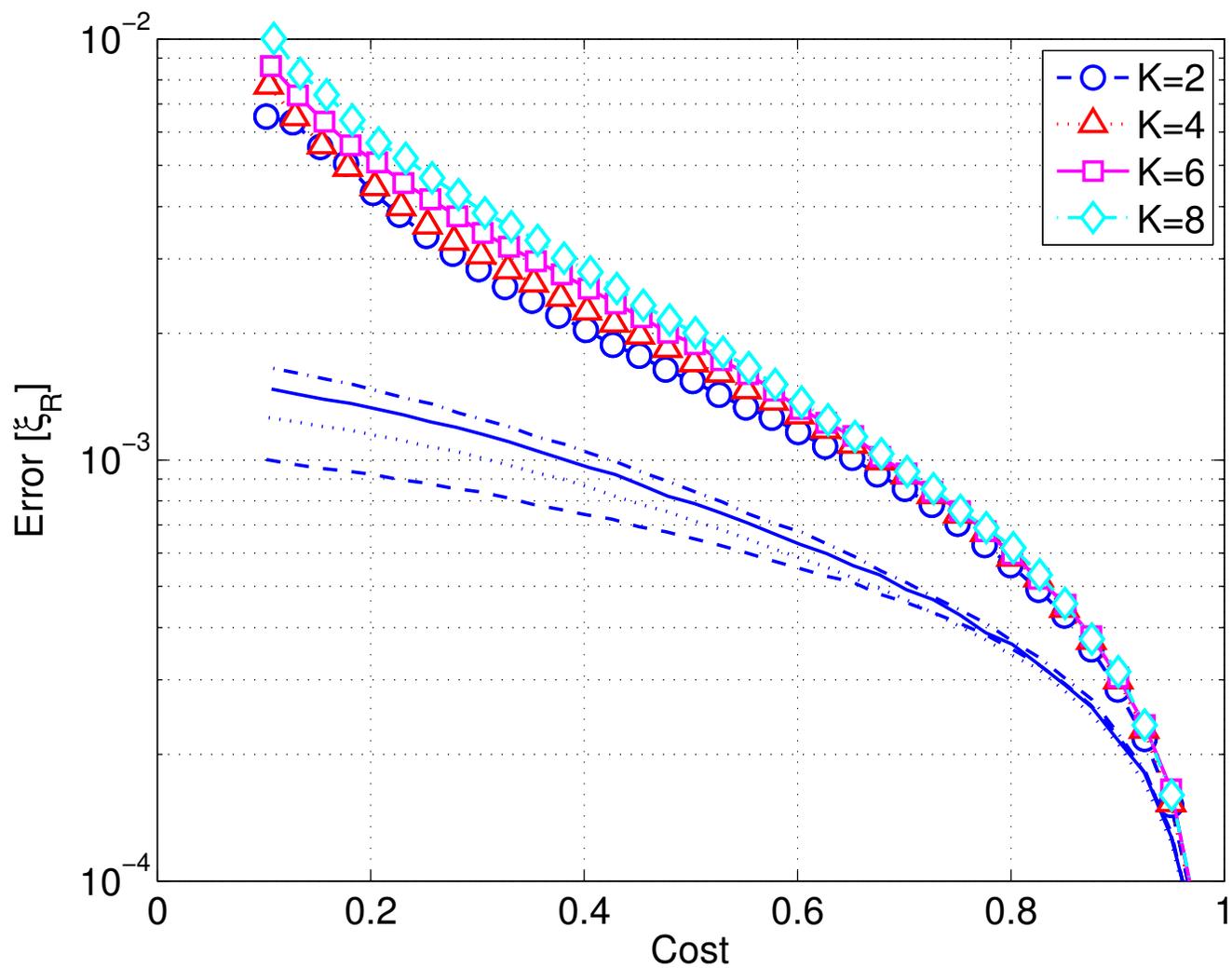


Fig. 5. Impact of the choice of K on the performance of *Fixed* p_{tx} , for signal (S1).

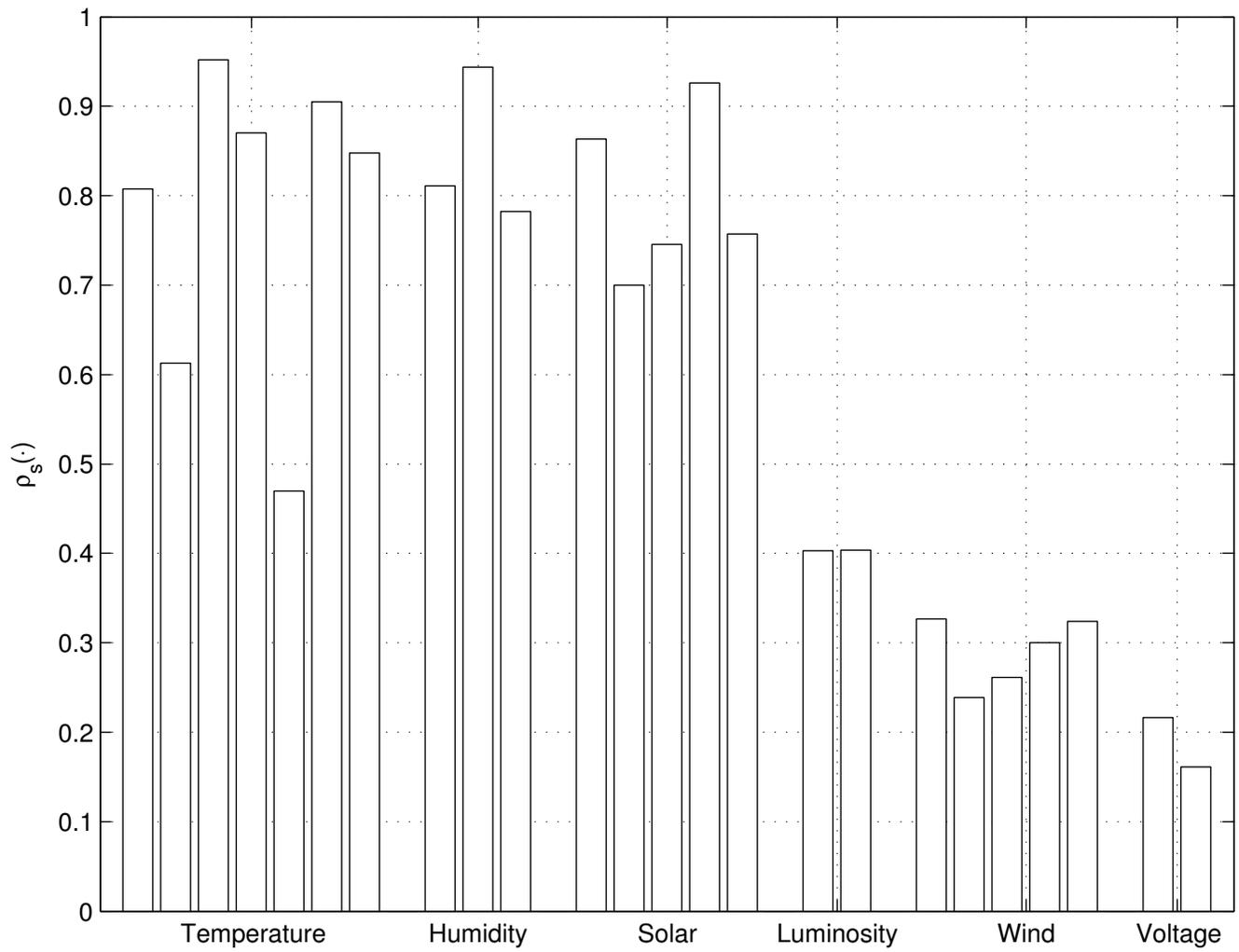


Fig. 6. Inter-node correlation for different signals gathered from the 5 different WSNs considered.

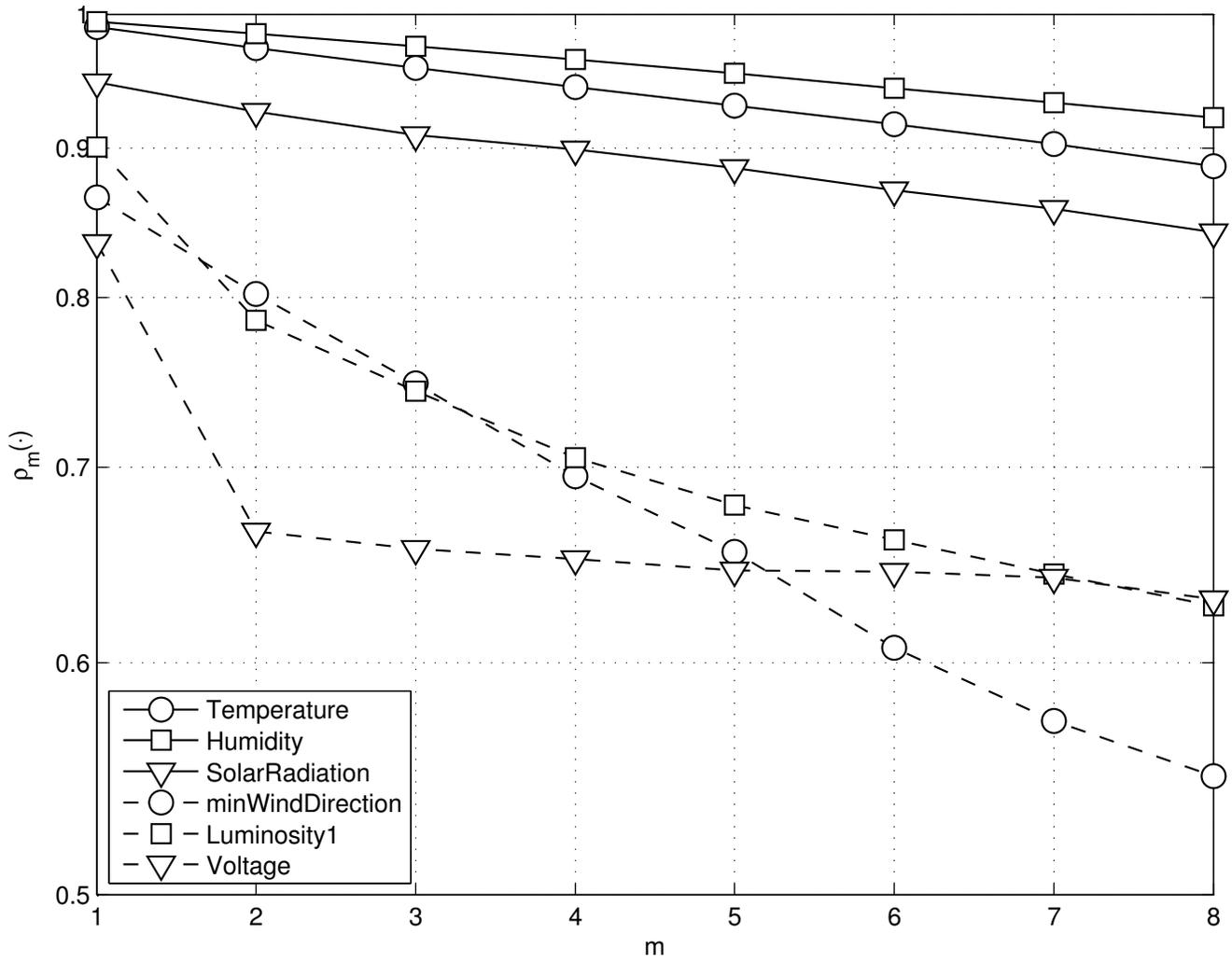


Fig. 7. Intra-node correlation for the signals chosen among all the signals considered in Fig. 6.

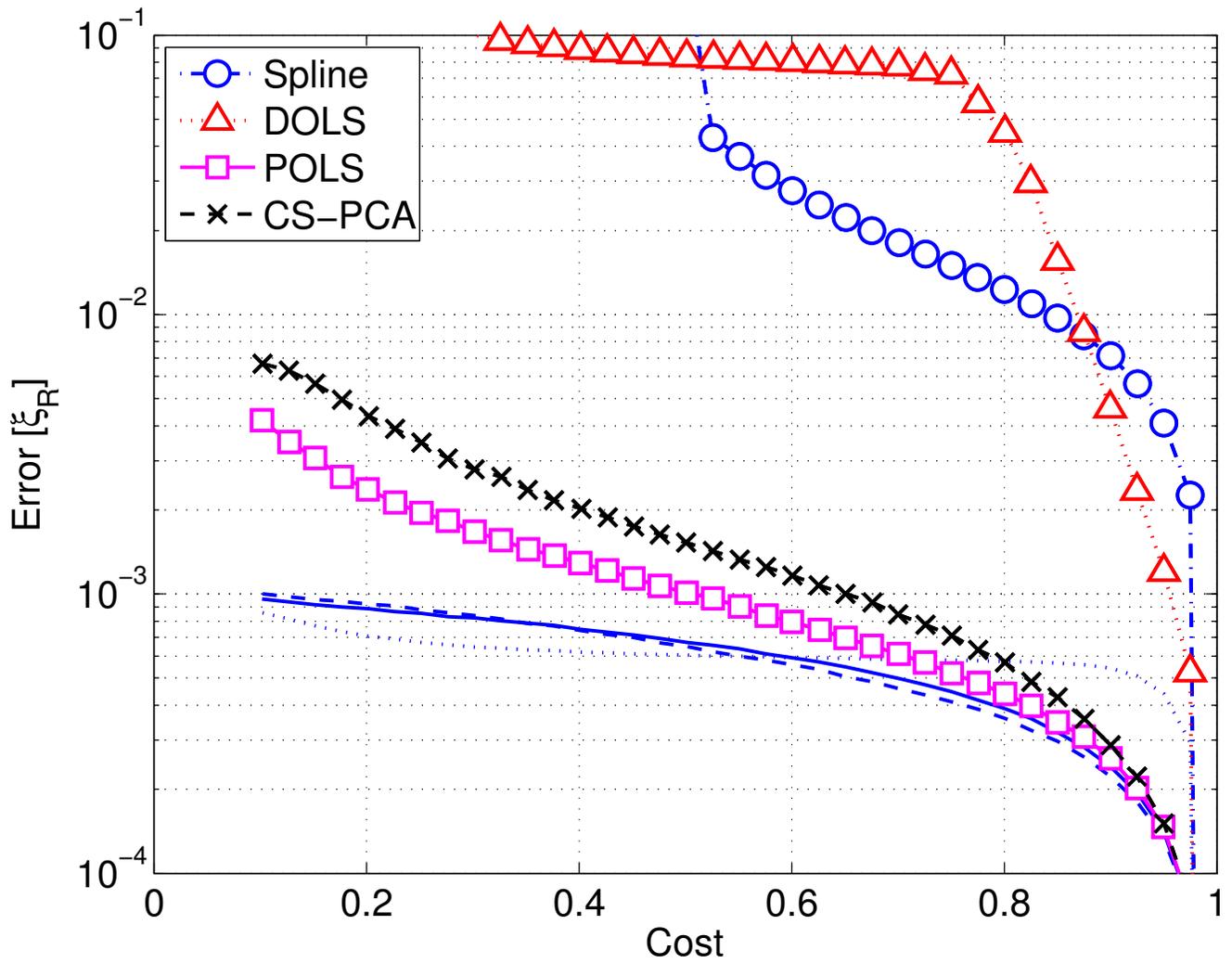


Fig. 8. Performance comparison of different interpolation techniques applied to *S*CoReI**, for signals in class S1, temperature and humidity. These performance curves are obtained with signals gathered from the DEI WSN deployment, see [30].

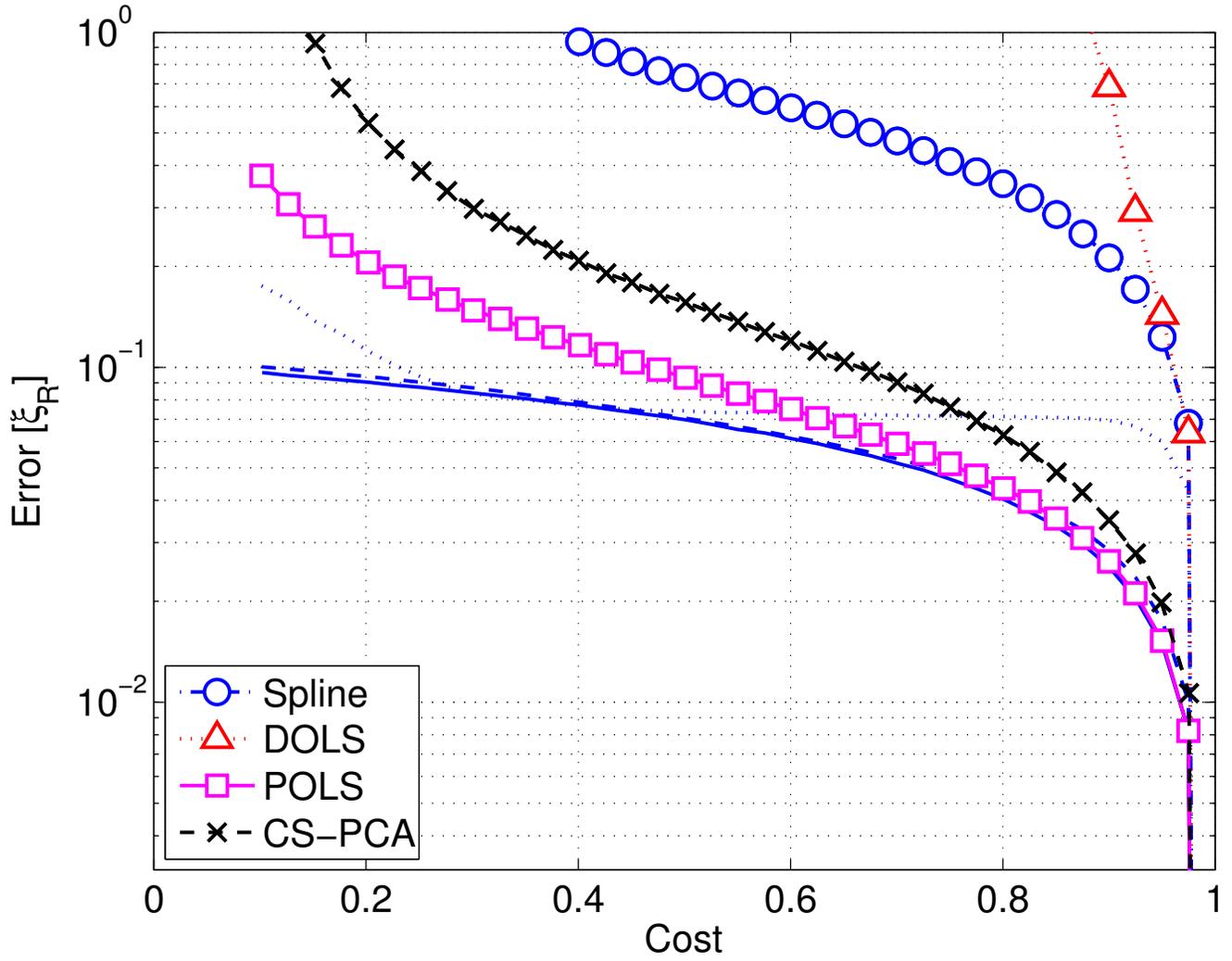


Fig. 9. Performance comparison of different interpolation techniques applied to *SCoReI*, for signals in class S2, luminosity in the range 320 – 730 nm and in the range 320 – 1100 nm. These performance curves are obtained with signals gathered from the DEI WSN deployment, see [30].

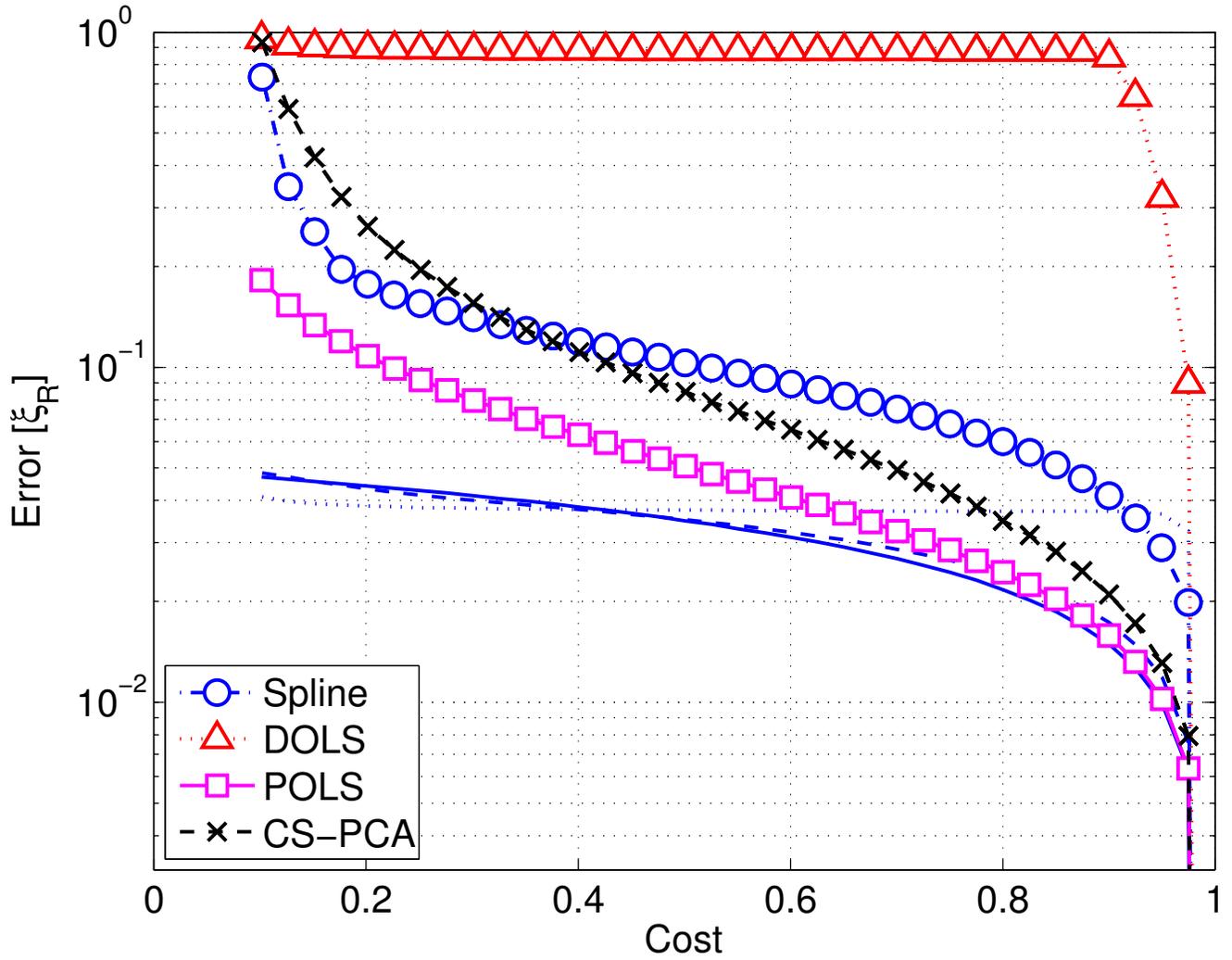


Fig. 10. Performance comparison of different interpolation techniques applied to *S-CoRel*, for signals in class S1, temperature and humidity. These performance curves are obtained with signals gathered from the EPFL WSN deployment LUCE, see [32].