Subspace Energy Monitoring for Anomaly Detection @Sensor or @Edge

(Article begins on next page)

Subspace Energy Monitoring for Anomaly Detection @Sensor or @Edge

Alex Marchioni, *Member, IEEE,* Mauro Mangia, *Member, IEEE,* Fabio Pareschi, *Senior Member, IEEE,* Riccardo Rovatti, *Fellow, IEEE,* and Gianluca Setti, *Fellow, IEEE*

***Abstract*—The amount of data generated by distributed monitoring systems that can be exploited for anomaly detection, along with real-time, bandwidth, and scalability requirements leads to the abandonment of centralized approaches in favor of processing closer to where data is generated. This increases the interest in algorithms coping with the limited computational resources of gateways or sensor nodes.**

We here propose two dual and lightweight methods for anomaly detection based on generalized spectral analysis. We monitor the signal energy laying along with the principal and *anti*-principal signal subspaces and call for an anomaly when such energy changes significantly with respect to normal conditions. A streaming approach for the online estimation of the needed subspaces is also proposed.

The methods are tested by applying them to synthetic data and real-world sensor readings. The synthetic setting is used for design space exploration and highlights the trade-off between accuracy and computational cost. The real-world example deals with structural health monitoring and shows how, despite the extremely low computations costs, our methods are able to detect permanent and transient anomalies that would classically be detected by full spectral analysis.

***Index Terms*—Anomaly detection, Principal Component Analysis, Spectral Analysis, Edge-of-the Cloud, Structural Health Monitoring.**

I. Introduction

The world in which we are currently living and more so the one we are designing for tomorrow is based on an interweaving of physical systems and information flows [1]. One of the most prominent and useful applications of such an interweaving is *monitoring*, i.e., the capability of continuously gathering fine-grain information from ever-larger portions of the physical world [2]–[6], let it be a smart city, a wild protected area, an infrastructure, etc. The general architecture of a monitoring infrastructure, as previously proposed in literature [5]–[7], is reported in Fig. 1. Sensors are deployed close to the physical phenomena to monitor. They are small nodes, in which computational resources are constrained by the requirement of being as non-intrusive as possible, that implies limits both on geometric dimensions and energy budget.

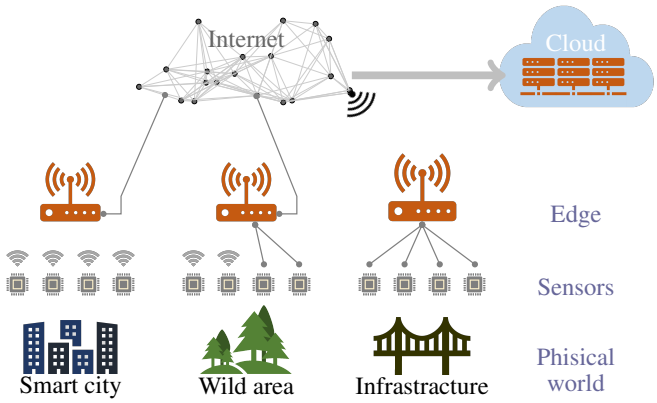


Fig. 1. Monitoring a portion of the physical world by flowing sensor readings to the cloud.

Clusters of sensors send their readings either by radio or wired links to the middle layer. Although represented in Fig. 1 with a unique device, this layer may be organized as a hierarchical structure in case of complex systems [7]. However implemented, this is the edge, i.e., a physically localized layer of processing elements that interposes between the application and the cloud abstraction. The computation resources available at the edge are typically much larger than those on each sensor node [8], providing the opportunity for local processing. This is an aspect that can be fundamental in case of wireless communication between the edge and the cloud, which is necessary, for instance, in case of monitoring systems located in remote areas. However, once in the cloud, computational constraints are virtually removed and translated into monetary cost constraints. With enough time and money, almost any amount of calculations can be performed.

Nevertheless, before reaching the cloud, the computational resources in the edge can be exploited not only for data dispatch but also for early data processing. Anomaly detection, conceived as the generation of alerts, is a task that can be performed in the edge. An alert, i.e., a deviation from normal behavior resulting from the readings of the sensors, may be caused by events in the monitored system but also by sensor fault, intrusion, tampering, damage, etc.

The effectiveness of detection may critically depend on the promptness with which the alerts are generated once anomalies

Manuscript received Mmmm dd, yyyy; revised Mmmm dd, yyyy.
This work was partially supported by the Italian Ministry for Education, University and Research (MIUR) under the program "Dipartimenti di Eccellenza (2018-2022)".

A. Marchioni, M. Mangia and R. Rovatti are with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: alex.marchioni@unibo.it, mauro.mangia2@unibo.it, riccardo.rovatti@unibo.it).

F. Pareschi and G. Setti are with the Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Torino, Italy, and also with the Advanced Research Center on Electronic Systems (ARCES), University of Bologna, 40125 Bologna, Italy (e-mail: fabio.pareschi@polito.it, gianluca.setti@polito.it).

manifest. Moreover, though detection can be easier when multiple sensors are considered [9], anomalies are usually local or semi-local phenomena that may become evident even when not all the available data is collected. This other aspect clearly hints at the opportunity of implementing anomaly detection at the edge, or even at sensors, rather than waiting for the whole data to reach the cloud to be processed. Furthermore, local early computation helps to reduce the traffic to central servers that, especially in the case of large-bandwidth sensors, would be mainly made of redundant data that becomes significant only in the presence of unusual behaviors to analyze. For this to be possible, one needs to reduce the computational complexity of anomaly detection enough to fit the constraint of the supporting computing hardware.

A. Related work

In this paper, the primary target of anomaly detection is to recognize changes in the monitored physical phenomenon. In principle, this may not be an easy task. An intuitive reason is that anomalies are deviations from behaviors that repeat in normal conditions. This *looking for repeating patterns* may call, for example, for analyses in some transformed domain (Fourier, Wavelet, adaptive, etc.) in which repeating patterns are represented by easily identifiable features (e.g., peaks). This is why spectral analysis and machine learning [10]–[12] often appear among the tools commonly called into play. Yet, when implementing anomaly detection at sensors or the edge, transformation, extraction of features in the transformed domain, and their tracking over time may result in a computational complexity exceeding the available resources.

For example, this is the case of [13]–[16], in which anomaly detection based on spectral analysis is proposed. These works entail a substantial computational effort and thus call for centralized processing. Analogously, in [17], the authors propose a supervised machine learning method to identify the occurrence of damages, whose implementation requires resources that do not fit into edge devices.

As an alternative, [18]–[20] propose detectors based on the Principal Component Analysis (PCA) have also been proposed, a concept that is quite common in many signal processing tasks and usually employed to reduce signal dimensionality. When it is employed as detector, the weighted energy collected by the principal components and the corresponding reconstruction error are the two scores that characterize this class of detectors. The approach in [20] proposes a cloud-based ecosystem where part of the processing is distributed among the nodes composing the monitoring system. Alternatively, in [18], the PCA is a building block that follows a centralized genetic algorithm employed to select which subset of sensors is more suitable to detect a specific system fault.

Another class of anomalies is related to faults of the monitoring systems. There are several possible causes for system faults that correspond to different abnormal waveforms in the sensor readings. A common practice consists of adopting a separate detector for each type of fault [21], [22], ranging from very low-complexity approaches [22], [23] to more sophisticated methods [24]. A further declination of the anomaly identification term is represented by intrusion detection [25], [26], which regards the attacks that infect the sensor nodes without any explicit network damage. As before, this class of anomalies includes a plethora of possible attacks along with a wide variety of detectors [25]. Several methods detect an intrusion when a node readings differ from the ones of its neighborhood, as for [27]–[29], while the approach in [29] can also be adapted to local intrusion detection.

B. Our contribution

The method we propose focuses on detecting changes in the observed physical phenomena, and it aims at maintaining detection capabilities close to those based on spectral analysis, with a limited computational effort to fit device resources available at sensor node or at the edge. As a working principle, we investigate how the energy of the signal is distributed over the signal space. An anomaly is detected when the energy along a suitably defined subspace of the signal space differs significantly from what expected in normal conditions.

Two different detectors are proposed, exploiting two different subspaces. The first measures the energy along the principal subspace to get a quantity able to discriminate the typical behavior from abnormal ones. Moreover, we also notice that, under some circumstances, the projections along the anti-principal subspace may yield information about the state of the system. We take advantages from this by relaying the second detector on the observation of the energy on the anti-principal subspace. With respect to previously proposed PCA-based detectors in [18]–[20], the concept of principal components is extended to principal subspace, and the anti-principal subspace is a generalization of the residual error.

We exploit the fact that our method deals with subspaces and not with eigenvalue-eigenvector pairs to propose suitably adapted, low-complexity estimation procedures for the second-order feature of the normal signals. As a result, the proposed detectors are novel in the sense that they generalize some PCA-based methods and, even more important, are able to compound classical transform-based analysis with a lightweight and implicitly adaptive procedure fitting tight computational constraints.

Additionally, we describe a method to estimate the energy on the subspaces mentioned above in a *streaming* fashion. The procedure can be derived by suitably tweaking an existing procedure for the on-line estimation of principal components. The derivation of a streaming procedure to estimate the anti-principal subspace is novel since classical approaches concentrate on principal components.

The proposed procedure has been tested on sensor readings coming from a system monitoring the vibration of a highway viaduct in Italy. Obtained results highlight that different kinds of structural anomalies are detectable: i) a massively destructive event that permanently changes the elastic structure; ii) a nondestructive event that causes only slight alterations; iii) an earthquake with no consequences in the monitored viaduct. Moreover, the detection capability has also been tested in the case of two kinds of faults affecting the sensing system.

The paper is organized as follows. In Section II we lay down the notation, define the detection method, and develop some

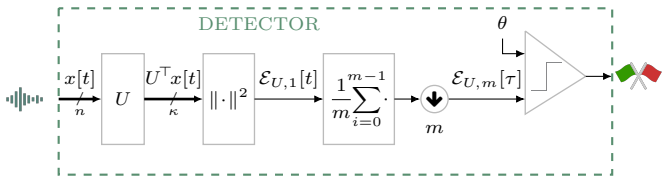


Fig. 2. Block scheme of the proposed method, where for each incoming vector x the energy of the projections Ux is computed and the average of m successive windows is adopted to establish if an anomalous event is occurring.

general considerations on its applicability and effectiveness. In Section III, we develop the online estimation of the subspaces used to characterize the distribution of energy in the signal space. In Section IV, we define the performance figures used to assess detectors' performance. In Section V, we define a Gaussian theoretical framework that allows computing semi-analytically the performance of the proposed detectors. In Section VI, we use such a framework to explore the detectors' design space and derive some qualitative behaviors. In Section VII, we show how the proposed methods behave in a real-world case. The conclusion is finally drawn.

II. Problem statement and definition of detectors

The incoming signal is modelled by a sequence of n-dimensional vectors $x[t] \in \mathbb{R}^n$ such that

$$x[t] = \begin{cases} x^{\mathrm{ok}}[t] & \text{if ok} \\ x^{\mathrm{ko}}[t] & \text{if ko} \end{cases}$$

where ok and ko are the two complementary events that define normal (ok) and anomalous (ko) behavior, while $x^{\mathrm{ok}}[t]$ and $x^{\mathrm{ko}}[t]$ are realizations of ergodic (and thus stationary) stochastic vector processes characterized by constant correlation matrices $\mathbf{K}^{\mathrm{ok}} \neq \mathbf{K}^{\mathrm{ko}}$ defined as

$$\mathbf{K}^{\times} = \mathbf{E}_x\left[x^{\times}[t]x^{\times}[t]^\top\right] = \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1} x^{\times}[t]x^{\times}[t]^\top \quad (1)$$

where $\times$ stands for either ok or ko, $\mathbf{E}$ stands for expectation and $^\top$ is for vector transpose.

Such matrices are symmetric and positive-semidefinite and thus admit a spectral decomposition $\mathbf{K}^{\times} = Q^{\times}\Lambda^{\times}Q^{\times\top}$ with $Q^{\times} = \begin{pmatrix} q_0^{\times} & q_1^{\times} & \dots & q_{n-1}^{\times}\end{pmatrix}$ an orthonormal matrix with eigenvector columns $q_j^{\times}$ and $\Lambda^{\times}$ a diagonal eigenvalue matrix $\Lambda^{\times} = \mathrm{diag}\left(\lambda_0^{\times}, \lambda_1^{\times}, \dots, \lambda_{n-1}^{\times}\right)$ such that $\lambda_j^{\times} q_j^{\times} = \mathbf{K}^{\times} q_j^{\times}$ and $\lambda_0^{\times} \geqslant \lambda_1^{\times} \geqslant \dots \geqslant \lambda_{n-1}^{\times} \geqslant 0$.

It is reasonable to assume that anomalous events happen independently at each time step with a probability $p_{\mathrm{ko}} = \Pr\{\mathrm{ko}\} \ll 1$. Under these assumptions, the correlation of the sequence of signals $x[t]$ is $\mathbf{K} = (1-p_{\mathrm{ko}})\mathbf{K}^{\mathrm{ok}} + p_{\mathrm{ko}}\mathbf{K}^{\mathrm{ko}} \simeq \mathbf{K}^{\mathrm{ok}}$.

Note that this model fits both an embodiment in which the entries of $x[t]$ are the subsequent samples of a signal in the t-th window, and one in which they are the simultaneous readings of different sensors at time step t, as well as those intermediate cases in which the vector $x[t]$ is made of readings from different sources in a set of subsequent time steps.

The observable we consider is the energy of the signal along a predefined subspace. To formalize this concept, let $U = \begin{pmatrix} u_0 & u_1 & \dots & u_{\kappa-1}\end{pmatrix}$ be an $n \times \kappa$ (with $\kappa \leqslant n$) matrix with orthonormal columns u_j. For each integer τ and for a given number m of subsequent instances, we observe the value

$$\mathcal{E}_{U,m}[\tau] = \frac{1}{m}\sum_{t=m\tau}^{m(\tau+1)-1}\left\|U^\top x[t]\right\|^2 = \frac{1}{m}\sum_{t=m\tau}^{m(\tau+1)-1}\mathcal{E}_{U,1}[t] \quad (2)$$

where $\|\cdot\|$ is the standard ℓ_2 norm of a vector. We identify anomalous instances as those yielding anomalous values of $\mathcal{E}_{U,m}[\tau]$, i.e., of the average over m subsequent energies of the projection of $x[t]$ on the κ-dimension linear subspace spanned by the columns of U.

Since we want to keep the computational complexity of the detector as low as possible, we limit ourselves to methods that declare an anomaly when either $\mathcal{E}_{U,m}[\tau] \leqslant \theta$ or $\mathcal{E}_{U,m}[\tau] \geqslant \theta$, for a fixed threshold θ.

Fig. 2 summarizes the mechanism behind the proposed detector showing its main blocks. From the scheme we get that the number of multiply and accumulate (MAC) operations is determined by the projection of the signal x onto U ($n\kappa$ MACs), the computation of the energy (κ MACs) and the computation of the average (1 MAC for each signal instance). As a result, every nm elements $m(\kappa(n+1)+1)$ operations are required, i.e., at most $\kappa+1$ MACs for each incoming sample.

To see that $\mathcal{E}_{U,m}$ is linked to the correlation between entries of $x[t]$, note that its average conditioned to either ok or ko is, by direct computation from (1),

$$\mu_{\mathcal{E}_{U,m}|\times} = \mathbf{E}_x\left[\mathcal{E}_{U,m}[\tau]|\times\right] = \mathrm{tr}\left(U^\top\mathbf{K}^{\times}U\right) \quad (3)$$

where the assumption of stationarity allows dropping time indications from the statistics of $\mathcal{E}_{U,m}[\tau]$, and where $\mathrm{tr}(\cdot)$ indicates the trace of a matrix.

A possible choice for U is $u_j = q_j^{\mathrm{ok}}$ for $j = 0, \dots, \kappa-1$. Since $\lambda_0^{\mathrm{ok}} \geqslant \lambda_1^{\mathrm{ok}} \geqslant \dots \geqslant \lambda_{n-1}^{\mathrm{ok}} \geqslant 0$ this amounts to take U as the principal κ-dimensional subspace of the process $x^{\mathrm{ok}}[t]$ and implies $\mu_{\mathcal{E}_{U,m}|\mathrm{ok}} = \sum_{j=0}^{\kappa-1}\lambda_j^{\mathrm{ok}}$, that is the largest possible average energy collected by projecting $x^{\mathrm{ok}}[t]$ onto any κ-dimensional subspace. By the very definition of principal subspace the instantaneous energy of the corresponding projection $\mathcal{E}_{U,m}[\tau]$ is expected to be large in normal cases, whereas anomalies can be revealed by the fact that $\mathcal{E}_{U,m}[\tau]$ falls below a certain threshold θ. We will indicate such a method as *Lack of Energy Detection* (LoED).

From a dual point of view, we may think of choosing $u_j = q_{n-j-1}^{\mathrm{ok}}$ for $j = 0, \dots, \kappa-1$ so that $\mu_{\mathcal{E}_{U,m}|\mathrm{ok}} = \sum_{j=0}^{\kappa-1}\lambda_{n-j-1}^{\mathrm{ok}}$, implying that U is the *anti*-principal κ-dimensional subspace, i.e., the subspace along which projections of normal instances have the least possible average energy. By the very definition of *anti*-principal subspace the instantaneous energy of the corresponding projection $\mathcal{E}_{U,m}[\tau]$ is expected to be small in normal cases, and anomalies can be revealed by the fact the $\mathcal{E}_{U,m}[\tau]$ exceeds a certain threshold θ. We will indicate such a method as *Excess of Energy Detection* (EoED).

A. The effect of signal localization

Both LoED and EoED rely on the fact that some subspaces exhibit distinctive features from an energetic point of view, i.e., that the distribution of the energy of the signal is not

uniform over the signal space. For this reason, we expect them to behave poorly for close-to-white signals with almost equal eigenvalues $\lambda_0^{\mathrm{ok}} \simeq \lambda_1^{\mathrm{ok}} \simeq \cdots \simeq \lambda_{n-1}^{\mathrm{ok}}$. In fact, in that case, the average energy collected by the projection on any κ-dimensional subspace is close to κ/n times the total energy of the signal and little discrimination is possible.

On the contrary, when the eigenvalues are substantially unbalanced, principal and *anti*-principal subspaces are distinguishing features of the process that may help to identify deviations. The unbalancing of the eigenvalues implying concentration of energy along a preferential direction in the signal space is often quantified by *localization* [30] defined as

$$\mathcal{L}_{x^{\mathrm{ok}}} = \frac{\sum_{j=0}^{n-1} {\lambda^{\mathrm{ok}}}_j^2}{\left(\sum_{j=0}^{n-1} \lambda_j^{\mathrm{ok}}\right)^2} - \frac{1}{n} = \frac{\mathrm{tr}\left(\mathbf{K}^{\mathrm{ok}^2}\right)}{\mathrm{tr}^2\left(\mathbf{K}^{\mathrm{ok}}\right)} - \frac{1}{n}$$

that is such that $\mathcal{L}_{x^{\mathrm{ok}}} = 0$ for white signals with $\lambda_0^{\mathrm{ok}} = \lambda_1^{\mathrm{ok}} = \cdots = \lambda_{n-1}^{\mathrm{ok}}$, while it reaches its maximum $\mathcal{L}_{x^{\mathrm{ok}}} = 1 - 1/n$ when all the energy of the signal is concentrated along the first eigenvector q_0^{ok}, i.e., $\lambda_0^{\mathrm{ok}} > 0$ and $\lambda_1^{\mathrm{ok}} = \lambda_2^{\mathrm{ok}} = \cdots = \lambda_{n-1}^{\mathrm{ok}} = 0$. In Section VI, by means of numerical evidences derived on a Gaussian framework, we will show that the larger $\mathcal{L}_{x^{\mathrm{ok}}}$, the better the performance of both LoED and EoED.

B. The effect of uncorrelated additive white noise

We assume that each signal instance is affected by additive white noise, uncorrelated with the signal, and with average energy per sample equal to σ^2. With this, each possible observation can be expressed as $x^{\times}[t] = \bar{x}^{\times}[t] + \nu[t]$ such that $\bar{\mathbf{K}}^{\times} = \mathbf{E}_{\bar{x}^{\times}}[\bar{x}^{\times}[t]\bar{x}^{\times}[t]^{\top}]$ and $\mathbf{E}_{\nu}[\nu[t]\nu[t]^{\top}] = \sigma^2 I$, where I is the identity matrix and the overbar indicates the noiseless quantities. Since noise and signal are assumed to be uncorrelated, we concentrate on a signal without anomalies and write

$$\mathbf{K}^{\mathrm{ok}} = \bar{\mathbf{K}}^{\mathrm{ok}} + \sigma^2 I$$

in which $\mathrm{tr}(\bar{\mathbf{K}}^{\mathrm{ok}})/n\sigma^2$ represents the Signal-to-Noise Ratio (SNR).

Note that the eigenvectors of $\mathbf{K}^{\mathrm{ok}}$ coincide with those of $\bar{\mathbf{K}}^{\mathrm{ok}}$ while the eigenvalues are such that $\lambda_j^{\mathrm{ok}} = \bar{\lambda}_j^{\mathrm{ok}} + \sigma^2$. Due to the offset, the observed normal signal is less localized compared to the noiseless normal signal, and detector performance is expected to decrease in small-SNR scenarios.

III. Streaming estimation of principal and anti-principal components

As seen in Section II, LoED and EoED detectors depend on the availability of the κ-dimensional principal or *anti*-principal subspace of the normal signal. They can be estimated offline by accumulating $\mathbf{K}^{\mathrm{ok}}$ as in (1) using a finite value of N, and then extracting from that the κ eigenvectors corresponding to the largest (LoED) or smallest (EoED) eigenvalues. This task is commonly executed on the cloud. As an alternative, the subspace represented by the same eigenvectors can be obtained in a streaming fashion with a resource budget that meets the capabilities of the edge devices. For this reason, a plethora of methods for the estimation of principal components from streaming data has been investigated (see, e.g., [31], for a recent review).

A common feature of streaming methods is that, when applied for $\kappa > 1$, they provide a basis for $\mathrm{span}(q_0, \ldots, q_{\kappa-1})$ that may not coincide with $q_0, \ldots, q_{\kappa-1}$. Further to that, most of them build local approximations of $\mathbf{K}^{\mathrm{ok}}$ and rely on the power method to extract such subspaces [31].

The methods we adopt here are those exploiting the extremal properties of principal and *anti*-principal subspaces. In particular we use a common adjusted formulation of the method in [32] for principal components, a simplification of [33] as suggested by [34] in a non-stochastic setting. Starting from the case of principal subspaces, we want to find a column-orthonormal matrix $U_{\uparrow} \in \mathbb{R}^{n\times\kappa}$ maximizing the expectation of the collected energy $\mathbf{E}_x[\mathcal{E}_{U,m}[t]] = \mathbf{E}_x[\mathcal{E}_{U,1}[t]]$. Assuming to have a total of N instances, by considering the corresponding empirical average

$$E_U = \frac{1}{N}\sum_{t=0}^{N-1} \mathcal{E}_{U,1}[t] \tag{4}$$

we can translate such a problem into one that may be tackled by classical Stochastic Gradient Ascent/Descent (SGA/D) [35]–[37]. The gradient of $\mathcal{E}_U[t]$ with respect to U is

$$\nabla_U \mathcal{E}_{U,1} = 2x[t]x[t]^{\top}U$$

Since following the gradient may spoil the orthonormality between the columns of U, we insert an orthonormalization procedure every T gradient steps. Overall, we estimate the principal κ-dimensional subspace by initializing $U_{\uparrow}^{(0)} \in \mathbb{R}^{n\times\kappa}$ at random and iterating

$$U_{\uparrow}^{(t+1)} = \begin{cases} U_{\uparrow}^{(t)} + 2\eta_t x[t]x[t]^{\top}U_{\uparrow}^{(t)} & \text{if } t \neq 0 \pmod T \\ \perp\left(U_{\uparrow}^{(t)} + 2\eta_t x[t]x[t]^{\top}U_{\uparrow}^{(t)}\right) & \text{if } t = 0 \pmod T \end{cases}$$

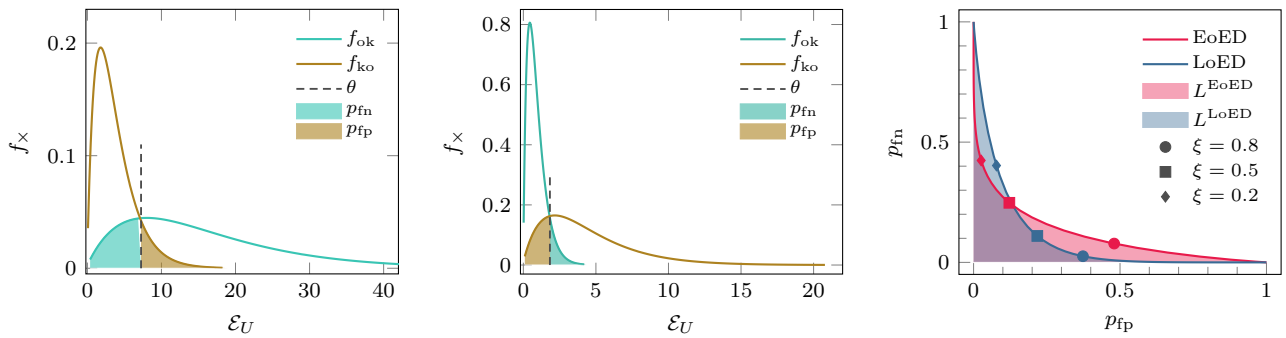
where $\perp(\cdot)$ is any column-orthonormalization procedure and $\eta_t > 0$ is a learning rate sequence that may be decided within the classical requirements $\sum_{t=0}^{\infty}\eta_t = +\infty$ and $\sum_{t=0}^{\infty}\eta_t^2 < +\infty$ [37, Theorem 4.7].

Though not mentioned in the Literature, starting from the same considerations, we may easily devise an algorithm to estimate the *anti*-principal subspace spanned by a matrix $U_{\downarrow}$. In fact, since every iteration follows the direction of the gradient of the energy, each step can increase or decrease such energy depending on the sign of the applied adjustment. Hence, starting from a random $U^{(0)} \in \mathbb{R}^{n\times\kappa}$ the iterations

$$U_{\downarrow}^{(t+1)} = \begin{cases} U_{\downarrow}^{(t)} - 2\eta_t x[t]x[t]^{\top}U_{\downarrow}^{(t)} & \text{if } t \neq 0 \pmod T \\ \perp\left(U_{\downarrow}^{(t)} - 2\eta_t x[t]x[t]^{\top}U_{\downarrow}^{(t)}\right) & \text{if } t = 0 \pmod T \end{cases}$$

produce an estimation of the κ-dimensional *anti*-principal subspace.

In both cases, the η_t and T should be administered to address the trade-off between computational complexity and convergence speed. As far as η_t is concerned, we adopt $\eta_t = \eta_0/\sqrt{t}$ that is the slowest possible decaying trend that is close enough to meeting the theoretical requirements without impairing convergence.



(a) (b) (c)

Fig. 3. Typical profiles of the probability density function of energy collected by the principal subspace (a) and *anti*-principal subspace (b) from two distinct processes, where the highlighted areas define the probabilities of false negatives and false positives for LoED (a) and EoED (b). Besides, two sample trends of false negative - false positive curves (c) with corresponding performance figures: the area values are equal to L and the highlighted points correspond to the minimal Λ_ξ.

IV. Performance figures

To quantify detection performance, we may indicate with $f_{\text{ok}}(\cdot)$ and $f_{\text{ko}}(\cdot)$ the two probability density functions of the energy $\mathcal{E}_U$ conditioned to ok and ko, respectively, and with $F_{\text{ok}}(\cdot)$ and $F_{\text{ko}}(\cdot)$ the corresponding cumulative distributions. With this, we can write the probabilities of *false positives* and *false negatives* for the LoED approach as

$$\begin{aligned} p_{\text{fp}}^{\text{LoED}}(\theta) &= \Pr\{\mathcal{E}_{U,m} \leqslant \theta | \text{ok}\} = F_{\text{ok}}(\theta) \\ p_{\text{fn}}^{\text{LoED}}(\theta) &= \Pr\{\mathcal{E}_{U,m} > \theta | \text{ko}\} = 1 - F_{\text{ko}}(\theta) \end{aligned}$$

and for the EoED approach as

$$\begin{aligned} p_{\text{fp}}^{\text{EoED}}(\theta) &= \Pr\{\mathcal{E}_{U,m} > \theta | \text{ok}\} = 1 - F_{\text{ok}}(\theta) \\ p_{\text{fn}}^{\text{EoED}}(\theta) &= \Pr\{\mathcal{E}_{U,m} \leqslant \theta | \text{ko}\} = F_{\text{ko}}(\theta) \end{aligned}$$

Fig. 3 shows the typical trends for probability density functions of the energy collected by the principal and *anti*-principal subspaces of a process (the blue ones) along with those of the energies collected by the same subspaces from an anomalous process (the orange ones). Fig. 3(b) refers to LoED, while Fig. 3(a) to EoED. The areas defining the false-positive and false-negative probabilities for both cases are highlighted.

Since we are observing energies, $\theta \geqslant 0$, and, as θ goes from 0 to ∞, one between p_{fp} and p_{fn} increases monotonically from 0 to 1, while the other decreases monotonically from 1 to 0. This means that the curve $(p_{\text{fp}}(\theta), p_{\text{fn}}(\theta))$ in the p_{fp}-p_{fn} plane has the trend reported in Fig. 3(c) where each point refers to a value of $\theta \in [0, \infty[$. Note that such a curve, sometimes indicated as Detection Error Tradeoff (DET) curve [38], is the complement of common Receiver Operating Characteristic (ROC) [39] that lives in the p_{fp}-p_{tp} plane, where the true-positive probability $p_{\text{tp}} = 1 - p_{\text{fn}}$ is the probability that an anomaly is correctly detected. We prefer to analyze performance in the p_{fp}-p_{fn} plane and the synthetic performance indexes we derive from it as losses that must be minimized. This choice is coherent with the anomaly detection setting in which devices are deployed to detect events that may cause real-world losses. Moreover, it has the likable side-effect that allows logarithmic plots to explore values very close to the optimum zero-loss situation.

In this setting, as a first quantitative assessment of detector performance, we identify the detector loss L as the area under such a curve, as shown in Fig. 3(c). L represents a cost to be minimized. Its upper bound $L = 0.5$ is for a trivial anomaly detector that randomly marks an event as anomalous with a probability 0.5, while the perfect detector would feature a DET curve such that $p_{\text{fn}} = 0$ for any target p_{fp} and vice-versa $p_{\text{fp}} = 0$ for any target p_{fn}, thus yielding $L = 0$.

Adapting [39], [40] we may give a further intuition for L. To do so, let us concentrate on the EoED that declares an anomaly when our statistic $\mathcal{E}_U$ exceeds the threshold and compute the area under p_{fp}-p_{fn} curve as

$$\begin{aligned} L &= \int_0^1 p_{\text{fn}}(p_{\text{fp}})\, dp_{\text{fp}} \\ &= \int_0^1 F_{\text{ko}}\left(F_{\text{ok}}^{-1}(1 - p_{\text{fp}})\right) dp_{\text{fp}} \\ &= \int_0^1 \int_0^{F_{\text{ok}}^{-1}(1-p_{\text{fp}})} f_{\text{ko}}(\alpha)\, d\alpha\, dp_{\text{fp}} \\ &= \int_0^{\infty} \int_0^{\beta} f_{\text{ko}}(\alpha) f_{\text{ok}}(\beta) d\alpha d\beta \end{aligned}$$

The final expression for L has a probabilistic interpretation[1]. Given U, a randomly chosen normal signal x', and a randomly chosen anomalous one x'', L is the probability that the observed statistics are in the wrong order, i.e., that $\mathcal{E}'_U$ (β in the above expression) is larger than $\mathcal{E}''_U$ (α in the above expression). When this happens, the detector makes a mistake, though we do not know of which kind. In fact, if x' does not give rise to a false positive then we must have $\theta > \mathcal{E}'_U > \mathcal{E}''_U$ and thus x'' will result in a false negative. On the contrary, if x'' does not give rise to a false negative then we must have $\theta < \mathcal{E}''_U < \mathcal{E}'_U$ and thus x' will result in a false positive. Hence, the lower the L, the smaller the chance that the detector makes an error of some kind.

[1]The last passage is due to the change of variable $\beta = F_{\text{ok}}^{-1}(1 - p_{\text{fp}})$.

A more specific quantitative assessment can be given by means of a loss depending on a parameter $0 \leqslant \xi \leqslant 1$

$$\Lambda_\xi(\theta) = \xi p_{\text{fn}}(\theta) + (1-\xi) p_{\text{fp}}(\theta)$$

that models the different relative cost that a false alarm may have compared to an overlooked anomaly. Each detector is then characterized by means of $\min_\theta \Lambda_\xi(\theta)$ for different values of ξ. As an example, the rightmost plot of Fig. 3(c) shows how the optimum point changes when ξ changes. Note that $\Lambda_{0.5}(\theta)$ is the complement to 1 of the classical balanced accuracy measured in classification tasks.

Besides, if the monitored physical phenomena changes in a way that requires a new estimation of the matrix U, also the value of θ that minimizes the adopted loss must be updated.

V. A Gaussian framework for performance evaluation

To derive analytical and semianalytical guidelines, we consider the classical theoretical setting in which the process produces vectors $x[t]$ that are independent realizations of a zero-mean Gaussian vector whose correlation matrix is either $\mathbf{K}^{\text{ok}}$ (in the normal cases) or $\mathbf{K}^{\text{ko}}$ (in the anomalous cases).

Since detectors rely on spotting significant deviations from most common behaviors, their performance is qualitatively related to the variance of the observable in the non-anomalous case. The smaller such a variance, the closer the observable to being an *invariant* of the process, and deviations from invariant behaviors are reliable indicators of anomalies.

This guideline leads us to consider EoED along with LoED. In fact, we may recognize the following property whose proof is in the Appendix.

Property 1. *Let $x \in \mathbb{R}^n$ a jointly Gaussian random vector with zero average and covariance/correlation matrix $\mathbf{K}$ with eigenvalues $\lambda_0 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_{n-1} \geqslant 0$ and corresponding orthonormal eigenvectors $q_0, q_1, \ldots, q_{n-1}$. Let $0 \leqslant j_0 < j_1 < \cdots < j_{\kappa-1} < n$ be any choice of κ indices, $U = \left(q_{j_0} | q_{j_1} | \ldots | q_{j_{\kappa-1}}\right)$ and $\mathcal{E}_{U,1} = \|U^\top x\|^2$. Setting $\mu_{\mathcal{E}_{U,1}} = \mathbf{E}_x[\mathcal{E}_{U,1}] = \sum_{j=0}^{\kappa-1} \lambda_{j_k}$, we have*

$$\sigma^2_{\mathcal{E}_{U,1}} = \mathbf{E}_x\left[\left(\mathcal{E}_{U,1} - \mu_{\mathcal{E}_{U,1}}\right)^2\right] = 2\sum_{k=0}^{\kappa-1} \lambda^2_{j_k}$$

With this, since the $x[t]$ are independent and equally distributed Gaussian vectors, from (2) we have

$$\sigma^2_{\mathcal{E}_{U,m}} = \mathbf{E}_x\left[\left(\mathcal{E}_{U,m} - \mu_{\mathcal{E}_{U,m}}\right)^2\right] = \frac{2}{m}\sum_{k=0}^{\kappa-1} \lambda^2_{j_k}$$

where no time indication is needed as we deal with statistics of stationary quantities.

Hence, the variance of the energy observed along the *anti*-principal subspace is smaller than the variance of the energy observed along the principal subspace. This difference indicates that EoED has the potential of performing better than LoED. In both cases, increasing m decreases the variance (presumably increasing performance) at the expense of a lower time resolution.

To assess the capabilities of EoED and LoED, we shall explore the detectors' design space exploiting the Gaussian assumption further to derive precise distributions for $\mathcal{E}_{U,m}$.

Since x is Gaussian, also $U^\top x$ is Gaussian with zero average and correlation/covariance matrix $U^\top \mathbf{K}^\times U$, where $\times$ is either ok or ko. Such a matrix can be given a spectral decomposition as in $U^\top \mathbf{K}^\times U = RDR^\top$ with $R \in \mathbb{R}^{\kappa\times\kappa}$ orthonormal and $D = \operatorname{diag}(d_0, \ldots, d_{\kappa-1})$. We may now consider the κ-dimensional vector $y = D^{-1/2}R^\top U^\top x$ that is also a zero-mean Gaussian vector with correlation/covariance $D^{-1/2}R^\top U^\top \mathbf{K}^\times URD^{-1/2} = I_\kappa$, i.e., its entries are independent normals with zero average and unit variance.

Since $U^\top x = RD^{1/2}y$, the observed energy can be recast in terms of y obtaining

$$\mathcal{E}_{U,1} = y^\top D^{1/2}R^\top RD^{1/2}y = y^\top Dy = \sum_{j=0}^{\kappa-1} d_j \chi^2_j(1) \qquad (5)$$

that is a linear combination with non-negative coefficients of 1-degree-of-freedom, independent chi-square random variables $\chi^2_j(1)$.

For $m > 1$, we may consider (2) in which, since the $x[t]$ are independent and equally distributed Gaussian vectors, the summands $\mathcal{E}_{U,1}[t]$ are independent and equally distributed. Hence

$$\mathcal{E}_{U,m}[\tau] = \frac{1}{m}\sum_{j=0}^{\kappa-1} d_j \chi^2_j(m) \qquad (6)$$

that is the average of a linear combination of m-degree-of-freedom independent random variables $\chi^2_j(m)$ with the same coefficients as in (5).

A plethora of analytical and numerical results are available for linear combinations in (6) (see [41], [42] and references therein) allowing the numerical evaluation of p_{fn} and p_{fp} in different test cases. In particular, it is known that the cumulative distribution function of $\mathcal{E}_{U,m}$ can be written as

$$\Pr\{\mathcal{E}_{U,m} \leqslant \theta\} = \sum_{j=0}^{\infty} c_j \frac{\Gamma\left(m\kappa + j, \frac{\theta}{2d_0}\right)}{\Gamma(m\kappa + j)} \qquad (7)$$

where $\Gamma(a,b) = \int_0^b \xi^{a-1}e^{-\xi}d\xi$ is the lower incomplete Gamma function, whose complete version is $\Gamma(a) = \Gamma(a,\infty)$, and the sequence of coefficients c_j for $j = 0, 1, \ldots$ is computed from the coefficients $d_0, \ldots, d_{\kappa-1}$ following [41].

Clearly, given any θ and letting $\mathbf{K}^\times$ be either the correlation matrix of the normal or of the anomalous cases, the method yielding (7) allows to compute $p_{\text{fp}}(\theta)$ and $p_{\text{fn}}(\theta)$ defined in Section II.

The semi-analytical assessment of the method we propose assumes that the normal process is made of independent zero-mean Gaussian vectors with $\mathbf{K}^{\text{ok}}_{j,k} = \omega^{|j-k|}$ for $j,k = 0, \ldots, n-1$. A rather straightforward computation [43, Chapter 2] allows to obtain the localization $\mathcal{L}_x = \frac{2\omega^2}{n}\frac{n(1-\omega^2)+\omega^{2n}-1}{n(1-\omega^2)^2}$ that we use to choose values for ω implying normal signals with three different levels of localization, i.e., $\mathcal{L}_x = 0.02$ (indicated as Low Localization - LL), $\mathcal{L}_x = 0.05$ (indicated as Medium Localization - ML), and $\mathcal{L}_x = 0.1$ (indicated as High Localization - HL). We expect higher localization to benefit the identifiability of anomalies.

The correlation of the anomalous process $\mathbf{K}^{\text{ko}}$ is randomly built as a perturbation of $\mathbf{K}^{\text{ok}}$. In formulas, we extract the square root $S = \sqrt{\mathbf{K}^{\text{ok}}}$ such as the matrix that $S^\top S = \mathbf{K}^{\text{ok}}$ and consider its columns s_j for $j = 0, 1, \ldots, n-1$ that are such that $s_j^\top s_j = \mathbf{K}^{\text{ok}}_{j,j} = 1$. We then rotate each s_j in a random direction by a certain angle α to obtain the columns of a matrix S' from which we set $\mathbf{K}^{\text{ko}} = S'^\top S'$, that can be seen as a perturbation of $\mathbf{K}^{\text{ok}}$ preserving the average energy of each component of $x[t]$, but gradually departing from its second-order statistics as α increases. We explore configurations with $\alpha = 0.1\pi, 0.25\pi, 0.5\pi$.

To understand the effect of such a perturbation, we may consider pairs of $\mathbf{K}^{\text{ok}}$ and $\mathbf{K}^{\text{ko}}$ and evaluate the Kullback-Leibler divergence of the two implied Gaussian distributions [44] that depends on the two correlation matrices as in

$$\mathcal{D}(\text{ok}\|\text{ko}) = \frac{1}{2}\log_2 \det\left(\mathbf{K}^{\text{ok}^{-1}}\mathbf{K}^{\text{ko}} e^{\mathbf{K}^{\text{ok}^{-1}}\mathbf{K}^{\text{ko}} - I}\right)$$

Such a divergence measures the amount of information (in bit) that each anomalous sample $x[t]$ gives to a detector that tries to discard the ok hypothesis in favor of the ko one. Since $\mathbf{K}^{\text{ko}}$ is randomly built starting from $\mathbf{K}^{\text{ok}}$, the divergence itself is a random variable. Yet, we may estimate its average in the ML case to obtain $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 0.8\,\text{bit}$ for $\alpha = 0.1\pi$, $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 12.1\,\text{bit}$ for $\alpha = 0.25\pi$, and $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 203.5\,\text{bit}$ for $\alpha = 0.5\pi$. This indicates that the performance of an ideal detector is largely affected by the value of α.

The Kullback-Leibler divergence allows us to assess also the effect of localization discussed in Section II from a more intuitive point of view. As an example, if we set $\alpha = 0.25\pi$ and consider the three localization levels, then average divergences are $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 10.8\,\text{bit}$ in the LL case, $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 12.1\,\text{bit}$ in the ML case, and $\mathbf{E}_{x^{\text{ok}},x^{\text{ko}}}[\mathcal{D}(\text{ok}|\text{ko})] \simeq 14.6\,\text{bit}$ in the HL case. Hence, though less than the difference between normal and abnormal behavior measured by α, increasing the localization of the normal behavior itself also eases detection in the ideal case.

As the first application of our Gaussian framework, we test the effectiveness of the streaming estimation of principal and *anti*-principal components. We consider $n = 64$ and a LL $\mathbf{K}^{\text{ok}}$, and we analyze 10^5 independent windows to look for κ-dimensional principal and *anti*-principal subspaces, with $\kappa = 3$. The learning rate is controlled by $\eta_0 = 0.1$. Through the singular value decomposition of the data matrix collecting all the 10^5 windows, we derive the targets $U_\uparrow^*$ and $U_\downarrow^*$, i.e., the two κ-dimensional subspaces that collect the maximum possible empirical average energy $E_\uparrow^*$ and the minimum possible empirical average energy $E_\downarrow^*$.

To evaluate the estimation quality we consider the empirical averages $E_\uparrow^{(t)}$ and $E_\downarrow^{(t)}$ of the energy (4) collected by the two estimations $U_\uparrow^{(t)}$ and $U_\downarrow^{(t)}$, and match them with $E_\uparrow^*$ and $E_\downarrow^*$.

Fig. 4 plots the relative estimation errors $\left|E_\uparrow^{(t)}/E_\uparrow^* - 1\right|$ and $\left|E_\downarrow^{(t)}/E_\downarrow^* - 1\right|$ against the number t of windows considered. The two continuous tracks correspond to $T = 1$, i.e., to a procedure in which $U_\uparrow^{(t)}$ and $U_\downarrow^{(t)}$ are re-orthonormalized after each application of a gradient step.

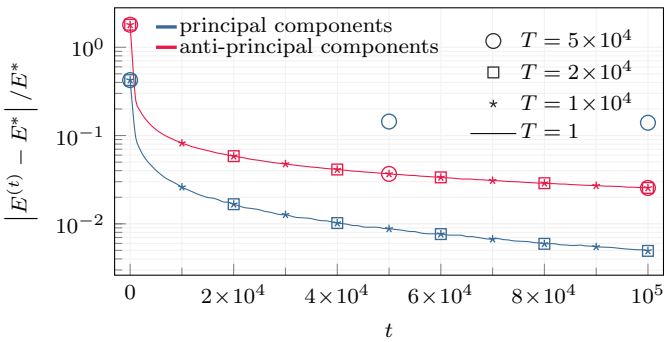


Fig. 4. Relative difference between the energy collected by the principal and *anti*-principal subspaces and their estimation by means of the streaming procedures in Section III.

Such a strategy implies a large computational overhead as orthonormalizing an $n \times \kappa$ matrix has complexity $\mathcal{O}(n\kappa^2)$ while the application of a gradient step is only $\mathcal{O}(n\kappa)$.

Fig. 4 shows what happens when $T = 1 \times 10^4$ (asterisks), $T = 2 \times 10^4$ (squares), and $T = 5 \times 10^4$ (circles). Estimation performance is the same at least up to $T = 2 \times 10^4$, while for larger values, the estimation of the principal subspace fails. In any case, orthonormalization overhead can be made negligible by choosing $T \gg \kappa$ while not suffering any performance degradation.

VI. Exploration of Detectors' Design Space

Let us remark that, in the proposed Gaussian framework, n and the localization level identify a matrix $\mathbf{K}^{\text{ok}}$ while $\mathbf{K}^{\text{ko}}$ is randomly generated depending on α. We already know that $\mu_{\mathcal{E}_{U^{\text{LoED}},m}|\text{ok}} = \sum_{j=0}^{\kappa-1} \lambda_j^{\text{ok}}$ and $\mu_{\mathcal{E}_{U^{\text{EoED}},m}|\text{ok}} = \sum_{j=0}^{\kappa-1} \lambda_{n-j-1}^{\text{ok}}$ with $\mu_{\mathcal{E}_{U^{\text{EoED}},m}} \leqslant \mu_{\mathcal{E}_{U^{\text{LoED}},m}}$.

After that, the detector depends on the couple of parameters κ and m from which it is possible to obtain the cumulative distribution function of $\mathcal{E}_{U,m}$ and then evaluate the loss L.

Monte Carlo simulations are needed to average over the possible $\mathbf{K}^{\text{ko}}$ for each given value of α, and performance is assessed by averaging the loss L over 1000 trials. Moreover, due to the complexity of the design space, we assume $n = 64$ that allows running simulations in an amount of time that is reasonable but sufficiently large to resemble a real-world signal window.

Fig. 5(a) and Fig. 5(b) show what happens to the detector loss L when we sweep the dimensionality κ of the projection subspace from its minimum $\kappa = 1$ to its maximum $\kappa = 64$, with $m = 1$. In both plots, lines correspond to median values, while the shaded areas contain 99% of the values.

We may observe the same trends for both EoED and LoED. First, plots show that the higher the α the better the maximum performance that detectors may attain. In fact, given a localization, i.e., given a color in Fig. 5(a) and Fig. 5(b), dotted lines ($\alpha = 0.5\pi$) are able to produce lower losses with respect to dashed lines ($\alpha = 0.25\pi$) that, in turn, do the same with respect to solid lines ($\alpha = 0.1\pi$). This behavior is coherent with the general analysis of the Kullback-Liebler divergence of the distribution of the anomalous vectors compared to that of the normal vectors: as α increases, each anomalous $x[t]$

contains more information that can be used to reject the ok hypothesis in favor of the right ko hypothesis.

The same happens when the localization of the normal signal increases. Given a perturbation angle, i.e., given a line style in Fig. 5(a) and Fig. 5(b), pink tracks (HL signals) reach smaller losses with respect to yellow tracks (ML signals) that, in turn, reach smaller losses with respect to blue tracks (LL signals). As expected, more localized normal signals concentrate a larger fraction of the energy in smaller dimensional subspaces and thus are more easily distinguished from anomalies.

Beyond these common features, the vertical ranges of the two figures are different and highlight that EoED has the potentiality of performing much better than LoED. However, minimum losses localize at different values of κ, especially for high localization cases.

To explain this aspect, consider an extreme case in which $\lambda_j^{\mathrm{ok}} > 0$ only for $j = 0, \ldots, J-1$ with a certain $J \ll n$. An EoED with $\kappa = 1$ exploits the fact that the normal signal has no energy $\lambda_{n-1}^{\mathrm{ok}} = 0$ along q_{n-1}^{ok} and declares an anomaly when the instantaneous energy is larger than a certain θ. Such a detector never yields a false positive. However, an instance can be anomalous because it has energy along q_{n-2}^{ok} (that a normal signal does not have since $\lambda_{n-2}^{\mathrm{ok}} = 0$ if $n-2 \geqslant J$). Such an anomaly would go unnoticed unless we set $\kappa = 2$. This increase still causes no false positive but, assuming $n-3 \geqslant J$, leaves out anomalies that distribute their energy along q_{n-3}^{ok}. Following this path κ can be increased until $\kappa = n-J$ to yield no false positive but being able to capture all anomalies that feature energies in all the direction along which the normal signal does not. On the contrary, if we set $\kappa = n-J+1$ then a normal signal that has energy along q_{J-1}^{ok} may produce a false positive, thus increasing the detector loss. Maximum performance is therefore at $\kappa = n-J$ that, since $J \ll n$, is quite high.

LoED would behave in an exactly complementary way since increasing κ beyond J causes the detector to aggregate energy along the directions that do not give any contribution in the normal case and thus may cause false negatives when those directions contain energy from anomalous instances.

Fig. 6 shows the effect of averaging on the performance of the detectors. In particular, the loss L is plotted against κ for both LoED and EoED when trying to discriminate a LL signal from an anomaly whose second-order statistics is only slightly different from the normal one, i.e., for $\alpha = 0.1\pi$. The trends reported show how reducing the variance of the observable can be fundamental, improving detectors' performance.

Furthermore, in a safety-critical situation in which false negatives imply a much higher cost with respect to false positives, the weighted loss Λ_ξ may be preferred to the agnostic L. In addition, since κ controls the computational complexity of the detector, EoED and LoED may be useful in different settings.

To exemplify this point we may compare the two cases $\kappa = 6$ and $\kappa = 32$ by looking at the p_{fn}-p_{fp} curve as shown in Fig. 7. In particular, we compare the $\xi = 0.5$ cases (agnostic situation), and the $\xi = 0.9$ cases (false negatives imply a much larger, 10 times, loss with respect to false positives).

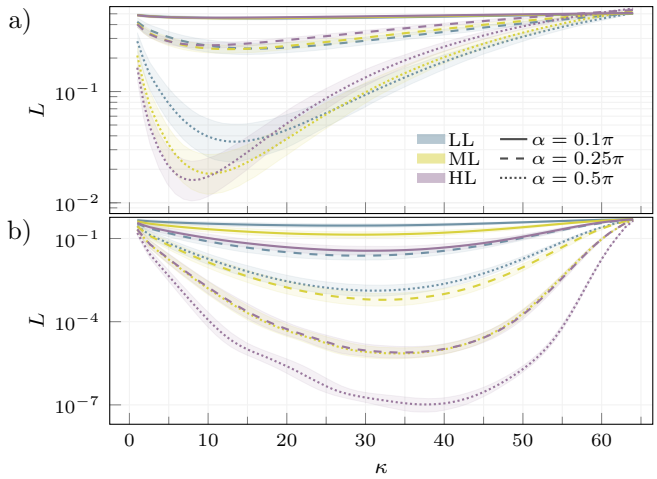


Fig. 5. The loss L of LoED (a) and EoED (b) as a function of κ of the projection subspace, for $m = 1$, for different localizations of the normal signal and for different levels of difference between the normal and anomalous statistics. Lines are median values while shaded areas show where 99% of the values fall.

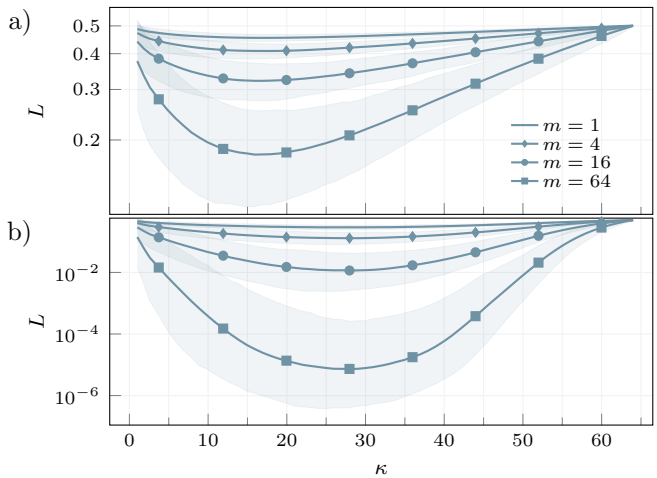


Fig. 6. The loss L of LoED (a) and EoED (b) as a function of κ of the projection subspace, for LL signals, a small difference $\alpha = 0.1\pi$ between the normal and anomalous statistics, and various level of averaging m. Lines are median values while shaded areas show where 99% of the values fall.

In the $\kappa = 32$ case, the p_{fn}-p_{fp} curve of EoED is so close to the ideal profile that, for any value of ξ, LoED is not able to deliver smaller losses. However, if the computational resources are extremely scarce, and we are forced to adopt $\kappa = 6$, the method with a lower weighted loss Λ_ξ depends on the value of ξ. As shown in Fig. 7(right), the curves for LoED and EoED feature an intersection such that for $\xi = 0.5$ EoED still outperforms LoED while in the unbalanced case, with $\xi = 0.9$, LoED reaches a lower minimum loss compared to EoED.

VII. A REAL-WORLD APPLICATION

To test the proposed method on real-world signals, we consider the data coming from the accelerometers deployed at a viaduct along an Italian motorway (see Fig. 8).

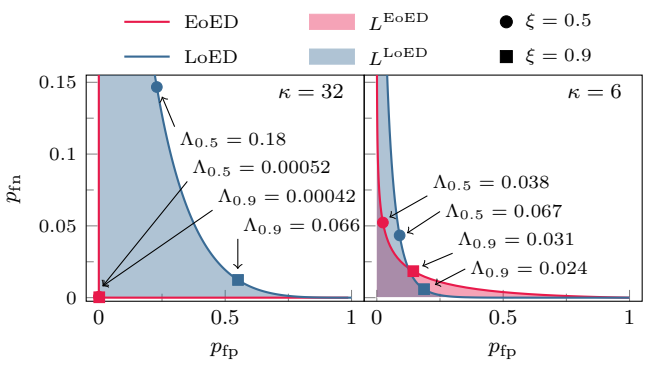


Fig. 7. p_{fn}-p_{fp} curves for both EoED and LoED, for $\kappa = 32$ (left plot) and $\kappa = 6$ (right plot) and for $m = 1$. The filled areas under the curves represent the loss L while the highlighted points correspond the the minimum loss Λ_ξ for both $\xi = 0.5$ and $\xi = 0.9$.

Fig. 8. Viaduct along an Italian motorway, where sensors deployed for structural health monitoring are attached to the internal tendons.

The structure is a composite box girder with external prestressed tendons (each of them consisting of 27 strands) used for reinforcement. The length of the viaduct is 580 m, and five concrete piers hold up the six spans, and each span counts 12 tendons. Although they are not identical, all the cables have roughly the same mass-over-length ratio, and they have been equally prestressed so that their natural vibration frequencies are about 7-8 Hz. 90 three-axis accelerometers are attached to the viaduct tendons (maximum two for each tendon) and provide a stream of 100 sample/s for each axis. The structure is stimulated by car traffic and by environmental factors such as wind and exhibits complicated elastic responses.

In this application, solutions that imply the upload of all sensor readings to the cloud are not viable, the main reason being communication costs. In fact, given the remote location, a wired connection is not available and the constant total data rate of 422 kbit/s rules out the use of LoRa technologies, whose links cannot provide more than 37.5 kbit/s and, in any case, are meant for very low duty cycles nodes. Beyond that, M2M communication through telephone-like infrastructure comes with different cost schemes: i) traffic/rate-unlimited, typically charging not less than 0.01 EUR/Mb; ii) traffic-limited, typically charging not less than 0.2 (Gb/month)/(EUR/month); iii) rate-limited, typically charging not less than 1 (kbit/s)/(EUR/month).

A constant bit rate makes iii) the most convenient option still implying not less than 400 EUR/month. Yet, by performing local caching and local anomaly detection, one may transmit data only when critical situations occur thus reverting to a situation in which high data rate connections are used unfrequently. By tuning false positives one may adopt plan i) and reduce transmission costs by almost two orders of magnitude.

The monitoring system has been active for two years with the aim of detecting anomalous responses that may indicate either a non-conventional stimulus or a compromised elastic behavior. Given a target sensor, the streams of samples corresponding to the three axes are partitioned in chunks of 100-samples = 1 s. Every second, the chunks corresponding to the three axes are arranged in a single vector, thus generating a sequence of $n = 300$-dimensional vectors $x[t]$ that account both for the time- and space- behavior of the structure as perceived by that sensor. Since the traffic on the viaduct is intermittent, we pre-filter windows to guarantee that those contributing to anomaly detection exhibit a good level of signal compared to background noise. Moreover, to focus on the correlations without the bias of the signal magnitude, each window is normalized to have zero mean and unit energy.

Using such a sieved sequence of normalized windows, we estimate both the principal and *anti*-principal subspaces of $\mathbf{K}^{\text{ok}}$ by the methods in Section III for $\kappa = 20$ and $\kappa = 70$. Fig. 9 shows what happens when we feed those methods with 4.5×10^5 windows comparing the energy collected by the resulting subspaces with that collected by the corresponding target eigenspaces of the empirical correlation matrix. We show trends for $T = 1{\times}10^4, 2{\times}10^4, 5{\times}10^4$ (in which orthonormalization gives a negligible contribution to the computational burden of the procedure) with the one with $T = 1$.

When analyzing the *anti*-principal subspace, we realize that the 30 less energetic directions collect an average energy that is negligible ($\leqslant 0.01\%$) compared to what projects on the other directions. This means that, even if the overall signal is $n = 300$-dimensional, it can be safely embedded in a 270-dimensional space. Accordingly, from now on, the *anti*-principal subspace used in EoED will span only the 40 directions obtained by discarding the 30 least energetic components from the 70 estimated by the streaming method of Section III.

The estimation of the matrix $U_\uparrow$ and of $U_\downarrow$ enables the application of LoED and EoED to reveal different anomalies in the data streams coming from the viaduct. The proposed methods have been primarily designed to detect anomalies that affect the monitored structure and are, hopefully, rare events. In fact, over more than one year of monitoring, we have access to only a few anomalies, and their limited number prevents any significant estimation of statistical metrics.

Nevertheless, the properties of LoED and EoED allow for the detection of those anomalies that are related to faults of the monitoring system, causing data corruption. Since these are more common events, their number is statistically significant and are used for a comparative assessment. Although the

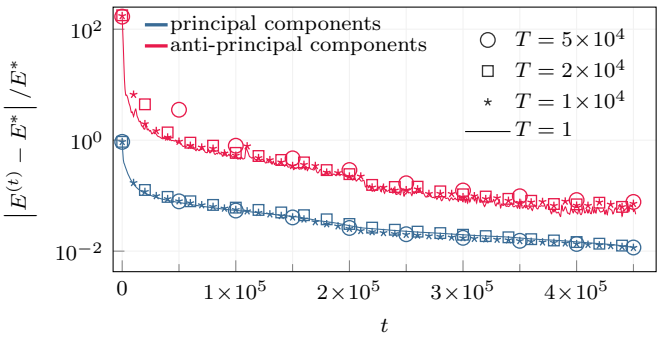


Fig. 9. The relative difference between the energy collected by the principal and *anti*-principal subspaces and their estimation by means of the streaming procedures in Section III in the case of a real world acceleration signal (for LoED $\kappa = 20$, $\eta_0 = 0.3$ and for EoED $\kappa = 70$, $\eta_0 = 40$).

selected system faults are not representative of all possible malfunctions, their analysis is presented to investigate the effectiveness of the proposed methods also on this class of anomalies.

A. Structural anomalies

In the monitoring period, the system witnesses three kinds of anomalies related to structural health: tendon break, strand break, and earthquake.

The tendon break is a massively disruptive event that significantly and permanently alters the elastic properties of the tendon. This event is not something that goes unnoticed: maintenance-on-failure is triggered to recover the structural integrity while exploiting the redundancy of the viaduct design to avoid service downtime. However, continuous monitoring is thought to be able to detect more subtle anomalies that may be prodromal to significant failures, thus enabling the pursual of predictive-maintenance approaches.

That is the case of the breakage of a strand, whose detection may be a prodrome of the tendon break. This event is much less noticeable and may cause slight and permanent changes in the tendon elastic properties. Here we consider a strand break that occurred on the tendon that broke four months after.

The last observed anomaly is an earthquake that lasted a few seconds and did not result in any severe damage to the structure, which maintained its original elastic behavior.

Fig. 10 shows how the three different anomalies are observed with an established analysis in the frequency domain [45]–[47] as well as with the proposed methods. Each column of the figure regards a different anomaly, while each row refers to a different detector. The same figure also reports the behavior of two scores defined by previous PCA-based detectors [18]–[20]: T2 and SPE. The former accounts for the energy of the weighted projection along the first principal components, while the latter account for the energy on the orthogonal subspace. Scores are such that high values hint at anomalies.

The first row, i.e., Fig. 10(a)-(c), shows the spectrum of the signal along the x-axis (the one parallel to the ground and orthogonal to the tendon length) before and after the anomalies. Spectrum is estimated by averaging the periodograms over 18 non-overlapping Hanning windows of 200 s each. Hence, the computation of each spectral profile requires 18 Fourier Transforms of 2×10^4 samples. The final minimum frequency accuracy is 5 mHz and is needed to detect small relative variations of peak frequencies, that are themselves in the order of few Hertz.

The evolution of the observable of both LoED and EoED for $\kappa = 20$ are shown in Fig. 10(d)-(f) and Fig. 10(g)-(i) respectively. Tracks of different colors correspond to different averaging, namely, $m = 1$ (no average), $m = 1800$ (i.e., average over $1/2$ h) and $m = 43200$ (i.e., average over 12 h). In each plot, dashed lines indicate the expected value of the observable under normal conditions, computed as the sum of the eigenvalues of $\mathbf{K}^{\mathrm{ok}}$ corresponding to the eigenspaces spanned by the columns of $U_\uparrow$ and $U_\downarrow$. Profiles for the reference scores T2 and SPE are reported in Fig. 10(l)-(n) and Fig. 10(o)-(q) respectively. To keep the computational complexity unaltered with respect to the LoED and EoED cases, these scores have been computed considering the first 20 principal components, for T2, and by the corresponding residue for SPE. In all cases, the trends run for 14 days centered on the day on which the event happened, and $\mathbf{K}^{\mathrm{ok}}$ is estimated by considering a week of data preceding the observation period.

Firstly, we consider the tendon brake, which is the more evident anomaly that any detector should be able to notice. The break of the tendon drastically changes its elastic properties, and that is confirmed in Fig. 10(a) where the spectrum after the event is noticeably different from the one before. The event is also immediately detectable both from EoED and LoED observables shown in Fig. 10(d) and 10(g). where the new regime of the observables is completely different from the previous one. In according with their definitions, the observable of LoED decreases while the observable of EoED is increasing. This is also the case of the score SPE in Fig. 10(o) while the profile of T2 in 10(l) shows a decrease in the score that makes T2 insensitive to the tendon break.

The second column of Fig. 10 refers to the strand break. Fig. 10(b) shows that the occurred event results in a $\simeq 2\%$ downshift for all the harmonics. Though this is a subtle change, it is still detectable from an accurate analysis of the tendon frequency response that considers suitably long data windows. The anomaly is also visible in both the EoED and LoED observables depicted in Fig. 10(e) and Fig. 10(h). In fact, a permanent deviation from the expected values of the observables can be seen in both plots even if their magnitude is smaller than those measured after the tendon break. Considering the cases of T2 and SPE, plots in Fig. 10(m) and Fig. 10(p) shown that, as before, a permanent deviation is evident in SPE only while T2 exhibits a feeble drop that, again, makes it insensitive to this anomaly. In both tendon and strand cases, the comparison between trends for different values of m shows how averaging is fundamental in reducing the variance of the observables. Averaging does not impair detection capabilities if the events to reveal either have a non-negligible duration or have long-lasting effects whose period is larger than the averaging window, while it filters out impulsive anomalies.

A representative of this last class of anomalies is an earthquake that hit the viaduct. The earthquake lasted few

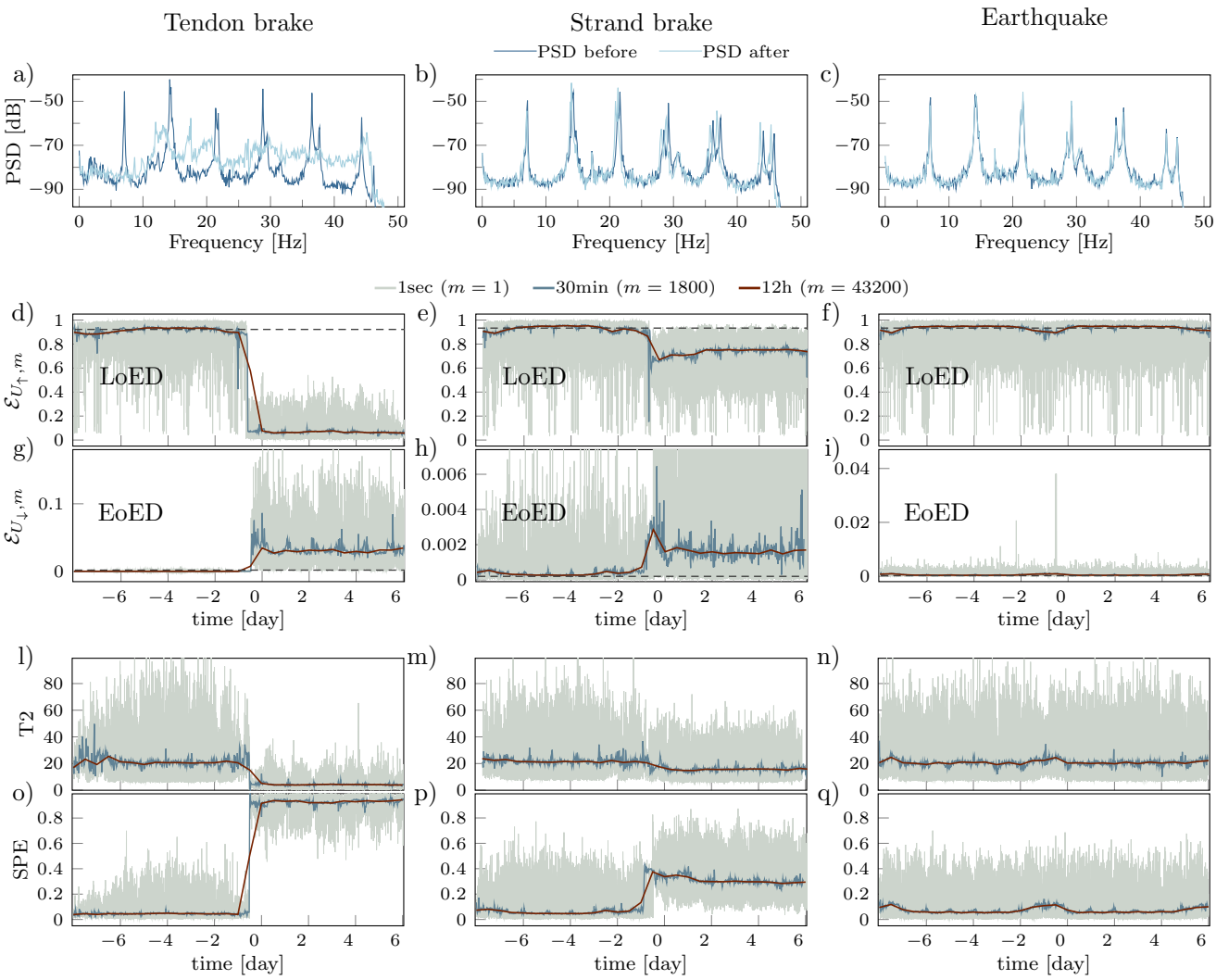


Fig. 10. The tendon break, strand break and earthquake anomalies observed by means of conventional spectral analysis as well as both with EoED and LoED ($\kappa = 20$), and with the scores T2 and SPE in [19] (20 principal components). Plots in the first row show the Power Spectral Densities of acceleration along the x-axis before and after the anomalous event. Plots in the second and third rows depict the Behavior of LoED and EoED observables for different values of m. Dashed lines indicate the expected value of the observables in normal conditions. Finally, plots in the forth row are for the T2 score while the last row is for the SPE score. Also for T2 and SPE, different values of m have been considered.

seconds and the frequency response of Fig. 10(c), as well as the averaged observables of Fig. 10(f), Fig. 10(i), Fig. 10(n) and Fig. 10(q), do not give any hint of it.

However, when $m = 1$, EoED highlights that something anomalous was happening. This is possible thanks to the fact that, even for small m, the observable of EoED features a small variance that lets the high-energy event stand out from normal conditions. On the other hand: LoED observable, T2 score and SPE score are characterized by a much higher variance and thus exhibits values comparable to those assumed during the earthquake even in normal conditions.

Focusing only on LoED and EoED, note that in general the adopted value of κ must be low enough to reduce the detector complexity and high enough to highlight changes in the structure. This change can be measured with the normalized gap G defined as

$$G_{U,m} = \frac{M^{\text{bef}}_{U,m} - M^{\text{aft}}_{U,m}}{(S^{\text{bef}}_{U,m} + S^{\text{aft}}_{U,m})/2}$$

where $M^{\text{bef}}_{U,m}$ and $M^{\text{aft}}_{U,m}$ are the empirical means and $S^{\text{bef}}_{U,m}$ and $S^{\text{aft}}_{U,m}$ are the empirical standard deviations of the projections energy $\mathcal{E}_{U,m}$ computed on the week before (bef) and the week after (aft) the event. Therefore, G gives an estimate of how large the gap is compared to the standard deviation.

Using the tendon brake as reference, Fig. 11 depicts the dependency of G with respect to κ for different values of m in the cases of both LoED and EoED. To limit the computational complexity of the detector, only values for $\kappa \leqslant 50$ are considered. The graph shows that, for LoED, the normalized gap presents a maximum at about $\kappa = 20$ for all three values of m, while for EoED, the more evident gaps are for $\kappa > 40$. Although an increase of m reduces time resolution,

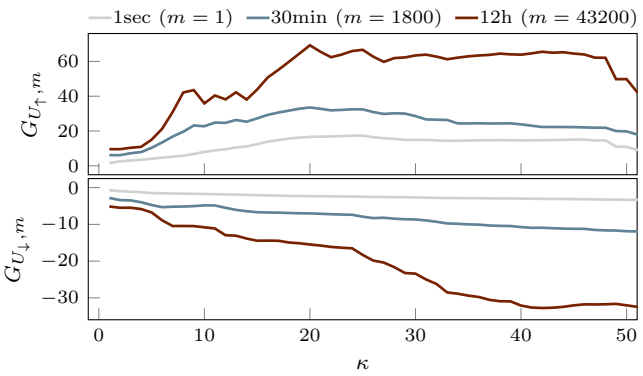


Fig. 11. Normalized gap G in LoED and EoED observables in the case of the tendon break anomaly depending on κ for different values of m.

it significantly lowers the variance of the observables while keeping nearly constant their mean. For this reason, higher values of m allow for more evident gaps. It is worth noting that the adoption of the κ values corresponding to the maxima observed in the G curves is not critical. As possible setting we suggest $\kappa = 20$ for LoED and $\kappa = 40$ EoED.

Let us now focus on the difference in computational resources needed by the method we propose and the spectral analysis. For a rough comparison of time-complexities, we consider a python implementation of the two alternatives whose running times are averaged over 10^3 Montecarlo trials when analyzing 1 h of readings from one sensor. The results are reported in the first row of Table I, where spectral analysis results to be not less than one order of magnitude more time-consuming than our methods.

As far as memory needs are concerned, note that in 1 h the accelerometer produces 3.6×10^5 samples per axis, encoded as 16 bit words. The streams are partitioned into 18 windows with $N = 2 \times 10^4$ samples. Since axes are sampled in parallel, three buffers are needed, each requiring $2N$ byte. In the most straightforward implementation, the Fourier transform of a buffer produces an equal number of frequency samples, each being a complex number with a 32 bit real part and a 32 bit imaginary part yielding further $8N$ bytes. The computation of spectrum averages in the modified periodogram implies an accumulating buffer for each axis requiring 4N byte. The overall memory footprint of the spectral analysis is therefore $\simeq 3 \cdot 2N + 8N + 3 \cdot 4N = 26N \simeq 500$ kB.

The memory requirement of EoED and LoED is dominated by the $\kappa \times n$ matrix, and we assume that each entry is encoded in a 32 bit word. To this, we should add a buffer of $n = 300$ samples, each encoded as a 16 bit word, and one of κ 32 bit words to contain the projections. As an example, considering $\kappa = 40$ the memory footprint is about 50 kB that results in one order of magnitude lower compared to the spectral analysis approach. This reduction allows for the deployment of the proposed detector on commercial devices typically devoted to sensor nodes such as ultra-low-power microcontrollers belonging to the STMicroelectronics' STM32L0 family (based on Cortex M0+) [48] and the ones based on Texas Instruments' MSP430 processor [49].

TABLE I
EMPIRICAL COMPARISON OF RESOURCES NEEDED BY THE PROPOSED DETECTORS AND BY CLASSICAL SPECTRAL ANALYSIS.

	LoED $\kappa = 20$	EoED $\kappa = 40$	spectral analysis
running time [ms]	1.85	4.43	49.7
memory footprint [Kb]	24.7	48.2	507.8

Note also that, for these class of devices a possible setting considers U stored in non-volatile memory (maximum 2 clock cycles to read each element) and 2 clock cycles for each MAC. With a typical 8 MHz clock frequency the computational time for each incoming 1 second instance is approximately equal to 6 ms.

B. System faults

The system faults that we consider are of two kinds. The first is represented by a reset of the sensing device. At each sensor boot, the digital filters used for signal pre-processing need a settling period of few samples for which the acquired data is corrupted. Following the classification, in [22] this anomaly falls within the SHORT class that describes a sharp change in the measured value between two successive data points.

The second system anomaly is a disturbance that has been observed to randomly affects some sensor devices. It manifests as an additive noise with a high correlation among the acceleration signals along the three axes (it is probably a disturbance injected during the simultaneous analog to digital conversion of the three channels). According to [22] it falls within the NOISE class.

Similarly to the previous subsection, a training set composed of one week of data without known anomalies is used to estimate the $\mathbf{K}^{\text{ok}}$ from which we get $U_\uparrow$ and $U_\downarrow$. After that, a validation phase is performed to compute the optimal threshold θ for each κ and detector. Here we refer to $m = 1$. The validation set contains 50 signal instances associated with sensor reset, one week of data corrupted by the mentioned disturbance, and finally, one week of data related to normal behavior. Performances are assessed in terms of loss $\Lambda_{0.5}$ on a test set that is composed like the validation set but referring different signal instances.

For these classes of anomalies, as a competitor, we adopt a variance-based anomaly detector that represents a common and low-complexity detector that can identify substantial changes in the signal statistics [22], [23]. This method matches the variance of the incoming vector x with a pre-trained threshold whose value is the one that minimizes $\Lambda_{0.5}$ on the validation set.

Since the two kinds of system faults affect only the principal subspace, we report in this subsection only the performance related to LoED method. Fig. 12 shows $\Lambda_{0.5}$ as a function of κ, in both validation and test set, compared to the values obtained with the variance-based method. The top plot depicts performances in case of resets. When κ is more of a few units, LoED reaches performances equal to the one observed with the

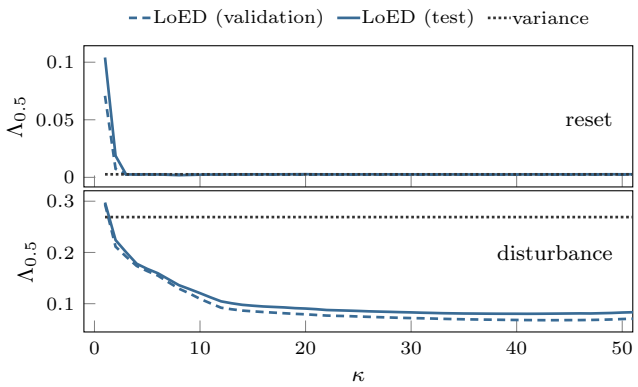


Fig. 12. Loss $\Lambda_{0.5}$ for LoED and the variance-based detector on the test set for the disturb and reset anomalies.

variance-based detector. Results related to the disturbance case are reported in the bottom plot and show that LoED clearly outperforms its competitor.

VIII. Conclusion

In this paper, we present methods for anomaly detection based on the evaluation of spectral-like features whose implementation can fit low resources devices such as sensor nodes and gateways.

Signal energy in a proper subspace is measured, and a threshold discriminates between normal and anomalous instances. The proposed methods, LoED and EoED, are associated with two different subspaces. The former identifies the place where the signal usually concentrates, while the latter refers to *anti*-principal components, i.e., those normally receiving a minimal amount of energy. Since the method targets low resources devices, we also derive a procedure for the on-line estimation of the *anti*-principal subspace by slightly modifying an existing one for principal components estimation.

The assessment of the methods is conducted firstly on synthetic data in terms of detector loss. Although EoED outperforms LoED in the investigated settings, it achieves the best performance with higher computational effort compared to what is observed for LoED.

The effectiveness of LoED and EoED is tested in a real-world scenario implying, as input signal, the vibrations of a highway viaduct in operational conditions. The methods are able to identify anomalies generated by not only destructive event causing permanent structural changes but also slight alterations in the viaduct elastic properties. Moreover, by monitoring the *anti*-principal components, EoED is even able to identify anomalies that do not permanently affect the structure. Performance assessment in the case of two kinds of system faults has also been presented, showing that LoED is still able to detect such a kind of abnormal waveforms.

Given the promising performance assessment, we will extend our trials to different signal sources and explore the implementation design space as far as sensor and gateway devices are concerned.

Appendix

Proof of Property 1. From (3), and independently of the Gaussian assumption, one immediately gets the well-known average energy $\mu_{\mathcal{E}_U} = \sum_{k=0}^{\kappa-1} \lambda_{j_k}$.

As far as the second-order statistics are concerned, note that $\mathcal{E}_U^2 = \left[\left(U^\top x\right)^\top U^\top x\right]^2 = x^\top U U^\top x x^\top U U^\top x = \operatorname{tr}\left(U U^\top x x^\top U U^\top x x^\top\right)$ that, brought down to sums over indexed quantities, gives

$$\begin{aligned}\mathbf{E}_x\left[\mathcal{E}_U^2\right] &= \mathbf{E}_x\left[\sum_{a,b,c,d,e,f} U_{a,b}U_{c,b}x_c x_d U_{d,e}U_{f,e}x_f x_a\right] \\ &= \sum_{a,b,c,d,e,f} U_{a,b}U_{c,b}U_{d,e}U_{f,e}\mathbf{E}_x\left[x_c x_d x_f x_a\right]\end{aligned}$$

where a sum in a set of indices shortens a sequence of sum each summing over a different index going from 0 to $n-1$.

Since x is Gaussian and zero-mean $\mathbf{E}_x\left[x_c x_d x_f x_a\right] = \mathbf{K}_{c,d}\mathbf{K}_{f,a} + \mathbf{K}_{c,f}\mathbf{K}_{d,a} + \mathbf{K}_{c,a}\mathbf{K}_{d,f}$. Hence,

$$\begin{aligned}\mathbf{E}_x\left[\mathcal{E}_U^2\right] = &\sum_{a,b,c,d,e,f} U_{a,b}U_{c,b}\mathbf{K}_{c,d}U_{d,e}U_{f,e}\mathbf{K}_{f,a} + \\ &\sum_{a,b,c,d,e,f} U_{a,b}U_{c,b}\mathbf{K}_{c,f}U_{f,e}U_{d,e}\mathbf{K}_{d,a} + \\ &\sum_{a,b,c} U_{a,b}U_{c,b}\mathbf{K}_{c,a}\sum_{d,e,f} U_{f,e}U_{d,e}\mathbf{K}_{d,f}\end{aligned}$$

where factors and sums have been rearranged and distributed to reconstruct high-level matrix and vector operations. By recognizing such operations we have

$$\begin{aligned}\mathbf{E}_x\left[\mathcal{E}_U^2\right] &= 2\operatorname{tr}\left(U U^\top \mathbf{K} U U^\top \mathbf{K}\right) + \operatorname{tr}\left(U U^\top \mathbf{K}\right)\operatorname{tr}\left(U U^\top \mathbf{K}\right) \\ &= 2\operatorname{tr}\left(U^\top \mathbf{K} U U^\top \mathbf{K} U\right) + \operatorname{tr}^2\left(U^\top \mathbf{K} U\right) \\ &= 2\sum_{j=0}^{\kappa-1}\lambda_{j_k}^2 + \left(\sum_{j=0}^{\kappa-1}\lambda_{j_k}\right)^2\end{aligned}$$

where we have exploited the fact that since U has eigenvectors of $\mathbf{K}$ as columns, $U^\top \mathbf{K} U = \operatorname{diag}\left(\lambda_{j_0}, \lambda_{j_1}, \ldots, \lambda_{j_{\kappa-1}}\right)$.

Since $\sigma_{\mathcal{E}_U}^2 = \mathbf{E}_x\left[\mathcal{E}_U^2\right] - \mu_{\mathcal{E}_U}^2$ the property is proved. □

Acknowledgements

This work was partially supported by the Smart-Data@PoliTO center, and by the Italian Ministry for Education, University and Research (MIUR) under the program "Dipartimenti di Eccellenza (2018-2022)".

References

[1] A. Alamri, W. S. Ansari, M. M. Hassan, M. S. Hossain, A. Alelaiwi, and M. A. Hossain, "A survey on sensor-cloud: Architecture, applications, and approaches," *International Journal of Distributed Sensor Networks*, vol. 9, no. 2, p. 917923, 2013.

[2] L. Chen, Y. Ho, H. Hsieh, S. Huang, H. Lee, and S. Mahajan, "Adf: An anomaly detection framework for large-scale pm2.5 sensing systems," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 559–570, April 2018.

[3] H. Messer, A. Zinevich, and P. Alpert, "Environmental monitoring by wireless communication networks," *Science*, vol. 312, no. 5774, pp. 713–713, 2006.

[4] J. P. Lynch, C. R. Farrar, and J. E. Michaels, "Structural health monitoring: technological advances to practical implementations [scanning the issue]," *Proceedings of the IEEE*, vol. 104, no. 8, pp. 1508–1512, Aug 2016.

[5] C. Arcadius Tokognon, B. Gao, G. Y. Tian, and Y. Yan, "Structural health monitoring framework based on internet of things: A survey," *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 619–635, June 2017.

[6] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, April 2014.

[7] J. Fu, Y. Liu, H. Chao, B. K. Bhargava, and Z. Zhang, "Secure data storage and searching for industrial iot by integrating fog computing and cloud computing," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4519–4528, Oct 2018.

[8] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, Oct 2016.

[9] T. Yu, X. Wang, and A. Shami, "Recursive principal component analysis-based data outlier detection and sensor data aggregation in iot systems," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2207–2216, Dec 2017.

[10] G. Wang and J. Jiao, "A kernel least squares based approach for nonlinear quality-related fault detection," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3195–3204, April 2017.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 823–839, May 2012.

[13] R. Jiang, H. Fei, and J. Huan, "Anomaly localization for network data streams with graph joint sparse pca," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 886–894.

[14] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2710–2720, May 2019.

[15] Y. Liu, J. Zeng, L. Xie, S. Luo, and H. Su, "Structured joint sparse principal component analysis for fault detection and isolation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2721–2731, May 2019.

[16] Y. Zhu, Y.-Q. Ni, H. Jin, D. Inaudi, and I. Laory, "A temperature-driven mpca method for structural anomaly detection," *Engineering Structures*, vol. 190, pp. 447 – 458, 2019.

[17] M. W. Hackell, R. Rolfes, M. B. Kane, and J. P. Lynch, "Three-tier modular structural health monitoring framework using environmental and operational condition clustering for data normalization: Validation on an operational wind turbine system," *Proceedings of the IEEE*, vol. 104, no. 8, pp. 1632–1646, Aug 2016.

[18] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed pca process monitoring based on fault-relevant variable selection and bayesian inference," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 1, pp. 377–386, Jan 2016.

[19] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, Nov 2014.

[20] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "A global manufacturing big data ecosystem for fault detection in predictive maintenance," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 183–192, Jan 2020.

[21] K. Ni *et al.*, "Sensor network data fault types," *ACM Trans. Sen. Netw.*, vol. 5, no. 3, pp. 25:1–25:29, Jun. 2009.

[22] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Trans. Sen. Netw.*, vol. 6, no. 3, pp. 23:1–23:39, Jun. 2010.

[23] J. Gupchup, A. Sharma, A. Terzis, R. Burns, and A. Szalay, "The perils of detecting measurement faults in environmental monitoring networks," 2019.

[24] N. E. ElHady and J. Provost, "A systematic survey on sensor failure detection and fault-tolerance in ambient assisted living," *Sensors*, vol. 18, no. 7, 2018.

[25] S. Hajiheidari, K. Wakil, M. Badri, and N. J. Navimipour, "Intrusion detection systems in the internet of things: A comprehensive investigation," *Computer Networks*, vol. 160, pp. 165 – 191, 2019.

[26] B. Hajimirzaei and N. J. Navimipour, "Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm," *ICT Express*, vol. 5, no. 1, pp. 56 – 59, 2019.

[27] J. Li, Z. Zhao, R. Li, and H. Zhang, "Ai-based two-stage intrusion detection for software defined iot networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2093–2102, April 2019.

[28] P. Chen, S. Cheng, and K. Chen, "Information fusion to defend intentional attack in internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 337–348, Aug 2014.

[29] S. Rajasegarar, A. Gluhak, M. A. Imran, M. Nati, M. Moshtaghi, C. Leckie, and M. Palaniswami, "Ellipsoidal neighbourhood outlier factor for distributed anomaly detection in resource constrained networks," *Pattern Recognition*, vol. 47, no. 9, pp. 2867 – 2879, 2014.

[30] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, "Rakeness-based design of low-complexity compressed sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, 2017.

[31] L. Balzano, Y. Chi, and Y. M. Lu, "Streaming pca and subspace tracksing: The missing data case," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1293–1310, 2018.

[32] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, 1982.

[33] T. Krasulina, "The method of stochastic approximation for the determination of the least eigenvalue of a symmetrical matrix," *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 9, no. 6, pp. 1383–1387, 1969.

[34] D. Faddeev and V. Faddeeeva, "Computational methods of linear algebra," *Journal of Soviet Mathematics*, vol. 15, no. 5, pp. 531–650, 1981.

[35] S. M. H. Robbins, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400 – 407, 1951.

[36] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 462–466, 09 1952.

[37] L. Bottou, F. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018.

[38] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The det curve in assessment of detection task performance," in *Proceedings of Eurospeech '97*, 1997, pp. 1895–1898.

[39] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982, pMID: 7063747.

[40] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006, rOC Analysis in Pattern Recognition.

[41] P. Moschopoulos and W. Canada, "The distribution function of a linear combination of chi-squares," *Computers & Mathematics with Applications*, vol. 10, no. 4, pp. 383 – 386, 1984.

[42] R. Farebrother, "Algorithm as204," *Journal of the Royal Statistical Society (Series C)*, vol. 33, no. 3, pp. 332–339, 1984.

[43] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, *Adapted Compressed Sensing for Effective Hardware Implementations: A Design Flow for Signal-Level Optimization of Compressed Sensing Stages*. Springer International Publishing, 2018.

[44] S. Kullback, *Information Theory and Statistics*. John Wiley & Sons, 1959.

[45] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, 2007.

[46] P. C. Chang, A. Flatau, and S. C. Liu, "Review paper: Health monitoring of civil infrastructure," *Structural Health Monitoring*, vol. 2, no. 3, pp. 257–267, 2003.

[47] H. Sohn, "Effects of environmental and operational variability on structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 539–560, 2007.

[48] STMicroelectronics, "Stm32l0 family, based on cortex m0+," https://www.st.com/en/microcontrollers-microprocessors/stm32l0-series.html, accessed: 2020-02-07.

[49] Texas Instruments, "Msp430 ultra low-power microcontroller," http://www.ti.com/microcontrollers/msp430-ultra-low-power-mcus/overview.html, accessed: 2010-09-30.

Alex Marchioni received the B.S. and M.S. degree (with honors) in electronic engineering from the University of Bologna, respectively in 2011 and 2015. In 2018, he joined the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI) of the University of Bologna, where he is currently working as research fellow. His research interests include compressed sensing, biomedical applications and signal processing for the Internet of Things and Big Data analytics.

Mauro Mangia (S'09-M'13) eceived the B.Sc. and M.Sc. degrees in electronic engineering and the Ph.D. degree in information technology from the University of Bologna, Bologna, Italy, in 2005, 2009, and 2013, respectively. He was a Visiting Ph.D. Student with the Ecole Polytechnique Federale de Lausanne in 2009 and 2012. He is currently a Post-Doctoral Researcher with ARCES, Statistical Signal Processing Group, University of Bologna. His research interests are in nonlinear systems, machine learning, compressed sensing, anomaly detection, Internet of Things, Big Data analytics and optimization.

He was a recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and of the 2019 IEEE BioCAS Transactions Best Paper Award. He received the Best Student Paper Award at ISCAS2011. He was the Web and Social Media Chair for ISCAS2018.

Fabio Pareschi (S'05-M'08-SM'19) received the Dr. Eng. degree (Hons.) in electronic engineering from the University of Ferrara, Italy, in 2001, and the Ph.D. degree in information technology from the University of Bologna, Italy, in 2007, under the European Doctorate Project (EDITH).

He is currently an Assistant Professor with the Department of Electronic and Telecommunication, Politecnico di Torino. He is also a Faculty Member with ARCES, University of Bologna. His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, compressed sensing, random number generation and testing, and electromagnetic compatibility.

Dr. Pareschi received the Best Paper Award at ECCTD 2005 and the Best Student Paper Award at EMC Zurich 2005 and IEEE EMCCompo 2019. He was a recipient of the 2019 IEEE BioCAS Transactions Best Paper Award. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-PART II from 2010 to 2013. He is currently Associate Editor for the IEEE OPEN JOURNAL OF CIRCUITS AND SYSTEMS.

Riccardo Rovatti (M'99-SM'02-F'12) received the M.S. degree in electronic engineering and the Ph.D. degree in electronics, computer science, and telecommunications from the University of Bologna, Italy, in 1992 and 1996, respectively. He is currently a Full Professor of electronics with the University of Bologna. He has authored approximately 300 technical contributions to international conferences and journals and two volumes. His research focuses on mathematical and applicative aspects of statistical signal processing and on the application of statistics to nonlinear dynamical systems.

He was Distinguished Lecturer of the IEEE CAS Society for the years 2017-2018. He was a recipient of the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award and the 2019 IEEE BioCAS Transactions Best Paper Award. He received the Best Paper Award at ECCTD 2005 and the Best Student Paper Award at the EMC Zurich 2005 and ISCAS 2011. He contributed to nonlinear and statistical signal processing applied to electronic systems.

Gianluca Setti (S89,M91,SM02,F06) received a Ph.D. degree in Electronic Engineering and Computer Science from the University of Bologna in 1997.From 1997 to 2017 he has been with the School of Engineering at the University of Ferrara, Italy as an Assitant, Associate and, since 2009 as a Full Professor of Circuit Theory and Analog Electronics. Since December 2017 he is a Professor of Electronics for Signal and Data Processing at the Department of Electronics and Telecommunications (DET) of Politecnico di Torino, Italy. Since 2002 is also a permanent (in kind) faculty member of ARCES, University of Bologna. His research interests include nonlinear circuits, recurrent neural networks, statistical signal processing, electromagnetic compatibility, compressive sensing, biomedical circuist and systems, power electronics, design and implementation of IoT nodes.

Dr. Setti received the 1998 Caianiello prize for the best Italian Ph.D. thesis on Neural Networks. He is also receipient of the 2013 IEEE CAS Society Meritorious Service Award and co-recipient of the 2004 IEEE CAS Society Darlington Award, of the 2013 IEEE CAS Society Guillemin-Cauer Award, the 2019 IEEE Transactions on Biomedical Circuits and Systems best paper award, as well as of the best paper award at ECCTD2005, and the best student paper award at EMCZurich2005, ISCAS2011, PRIME2019 and EMCCOMPO 2019.

He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE Transactions on Circuits and Systems - Part II (2006-2007) and of the IEEE Transactions on Circuits and Systems - Part I (2008-2009). He also served in the editorial Board of IEEE Access (2013-2015) and, since of the Proceedings of the IEEE (2015-2018). Since 2019 he served as the first non US Editor-in-Chief of the Proceedings of the IEEE, the flagship journal of the Institute.

Dr. Setti was the Technical Program Co-Chair of NDES2000 (Catania), ISCAS2007 (New Orleans), ISCAS2008 (Seattle), ICECS2012 (Seville), BioCAS2013 (Rotterdam) as well as the General Co-Chair of NOLTA2006 (Bologna) and ISCAS2018 (Florence).

He was a Distinguished Lecturer (2004-2005 and 2014-2015) of the IEEE CAS Society, as well as a member of its Board of Governors (2005-2008), and served as the 2010 CASS President. He held several other volunteer positions for the IEEE and in 2013-2014 he was the first non NorthAmerican Vice President of the IEEE for Publication Services and Products.