

# Strategies for principal component analysis in wireless sensor networks

Nisrine Ghadban<sup>1,2</sup>, Paul Honeine<sup>2</sup>, Clovis Francis<sup>1</sup>, Farah Mourad-Chehade<sup>2</sup>, Joumana Farah<sup>3</sup>

<sup>1</sup> Faculté de Génie, Université Libanaise, Lebanon

<sup>2</sup> Institut Charles Delaunay (CNRS), Université de technologie de Troyes, CNRS, France

<sup>3</sup> Telecommunications department, Faculty of Engineering, Holy-Spirit University of Kaslik, Lebanon

**Abstract**—This paper deals with the issue of monitoring physical phenomena using wireless sensor networks. It provides principal component analysis for the time series of sensors' measurements. Without the need to compute the sample covariance matrix, we derive several in-network strategies to estimate the principal axis, including noncooperative and diffusion strategies. The performance of the proposed strategies is illustrated in the issue of monitoring gas diffusion.

**Index Terms**—Principal component analysis, wireless sensor network, adaptive learning, distributed processing

## I. INTRODUCTION

A Wireless Sensor Network (WSN) involves a large number of wireless inexpensive devices, called sensor nodes [1]. Working in a stand-alone manner, a sensor node has several modules, such as a microprocessor, a radio unit for communication and a sub-system for measurement, linking it to the physical world. All these modules are powered by a system of energy supply, very often a battery. However, with its low price, a sensor node has limited amounts of memory, reduced processing capabilities and low communication capacities. Also, with a non-exchangeable battery, its autonomy imposes constraints of energy saving. In order to reduce energy consumption, sensors only communicate with their neighbors and hence, communication is focused on short distances. Many applications of WSNs exist in several domains, such as military, environmental domain, health [2], [3].

WSNs provide an inexpensive way to monitor physical phenomena such as the temperature field, the atmospheric pressure, the emission of a pollutant or a gas [4], [5]. Estimating fields measured by a WSN has received considerable attention. Using learning methods such as kernel machines, the measurements are combined leading to a regression model which links the measured quantity to the positions where measurements are made [6], [7]. The authors in [8] propose an estimation of the scalar field using adaptive networks. Diffusion strategies, i.e., adapt-then-combine and combine-then-adapt strategies [9] are used for this purpose. These strategies have been investigated to solve linear supervised learning problems, where the models are often estimated by least-squares approach. In this paper, we take advantage of these strategies to estimate the principal axis in principal component analysis (PCA) in order to provide unsupervised learning of the time series of sensors' measurements.

The PCA is one of the most popular unsupervised learning techniques, with applications in statistical analysis, data compression and feature extraction [10]. This statistical method defines a set of principal axes that transforms a number of correlated variables into uncorrelated ones, the so-called *principal components*. The most relevant principal axes retain the largest variance of the data. This technique has been investigated in many applications involving measurements of a sensor network, e.g., to extract features from noisy samples [11], or to compress and denoise high-dimensional datasets (time series of measurement) [12], [13]. See also [14] for intrusion detection and [15] for anomaly detection in network traffic with PCA.

The conventional PCA requires the eigen-decomposition of the sample covariance matrix. Sending all time series to a fusion center is not scalable, thus inappropriate for WSNs. Several techniques have been proposed to overcome these drawbacks. In [16], each node sends the estimated principal components, instead of the time series, to the fusion center. In [17], the authors use the power iteration method to estimate the most relevant principal axis. This method requires the computation of the sample covariance matrix, which is ill-suited for WSNs.

In this paper, we propose to estimate the principal axis without the need to compute the sample covariance matrix. To this end, we revisit Oja's rule, initially described in [18] and studied more recently in [19] for nonlinear PCA with kernel-based machines. Within the WSN settings, we derive several strategies in order to respect WSN constraints, including scalability, energy consumption, as well as computational and memory cost. Starting from the centralized scheme, we first provide a noncooperative strategy for PCA. Then, we study diffusion strategies, where sensors cooperate between each other to estimate the principal axis. Two diffusion strategies are derived, the so-called combine-then-adapt and adapt-then-combine strategies, have been recently investigated in adaptive filtering literature [20].

The rest of this paper is organized as follows: Next section describes the WSNs and their topologies. In Section III, we describe the proposed strategies for PCA in WSNs, and study their convergence in Section IV. Section V provides experimentation results and discussions, whereas Section VI concludes the paper.

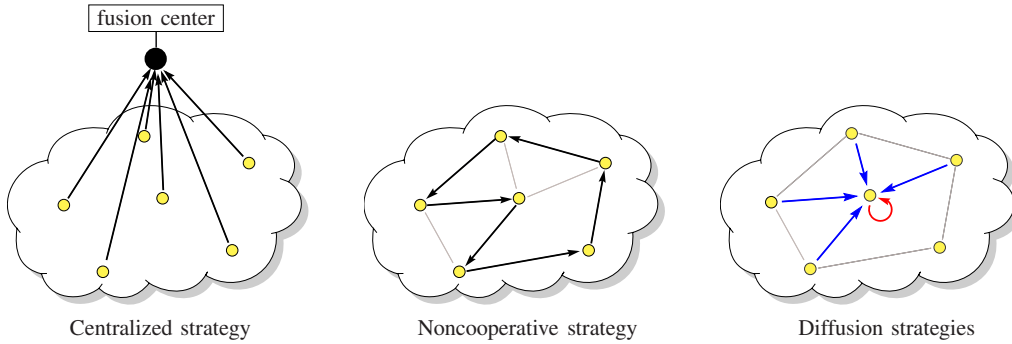


Fig. 1. Illustration of information processing within the proposed strategies for PCA in WSNs. The right-hand scheme illustrates both diffusion strategies: **combine-then-adapt** and **adapt-then-combine**.

## II. WIRELESS SENSOR NETWORKS

Let  $N$  sensors be deployed in a region  $\mathbb{X}$ , where  $\mathbb{X} \subset \mathbb{R}^2$  or  $\mathbb{X} \subset \mathbb{R}^3$ , for a two- or three-dimensional space. We denote by  $\mathbf{x}_i \in \mathbb{X}$  the position of sensor  $i$ , and  $y_{i,\theta} \in \mathbb{R}$  its measurement at time  $\theta$  of some physical quantity (e.g., temperature or gas concentration measurements). Each sensor collects measurements from  $\theta = 1$  to  $\theta = \Theta$ . Let  $\mathbb{Y} \subset \mathbb{R}^\Theta$  be the measurement vector space, with the conventional inner product  $\mathbf{y}_i^\top \mathbf{y}_j$  for any  $\mathbf{y}_i, \mathbf{y}_j \in \mathbb{Y}$ , where  $\mathbf{y}_i$  is the vector of entries  $y_{i,\theta}$ , for  $1 \leq \theta \leq \Theta$ . We denote by the real-value  $z = \mathbf{w}^\top \mathbf{y}$  the inner product associated to the orthogonal projection of any  $\mathbf{y} \in \mathbb{Y}$  onto some vector  $\mathbf{w} \in \mathbb{Y}$ .

In the following, we present several strategies to estimate the principal axis for PCA in the WSN. Potential applications include feature extraction, compression and denoising these time series (as illustrated in Section I). Let  $\mathcal{V}_i$  denote the index set of the nearest neighboring sensors to sensor  $i$ . We consider that  $\mathbf{x}_i$  is adjacent to itself, that is to say  $i \in \mathcal{V}_i$ .

### III. PCA FOR WSNs

We derive several strategies to extract the principal axis. The communication between sensors and the information interchange depend on the considered strategy. See Fig. 1.

#### A. Centralized strategy

Here, the network is assumed to be centralized, with sensors connected to a fusion center, to which they send their measured data without any local processing. Having the measurements,  $\mathbf{y}_1, \dots, \mathbf{y}_N$ , the fusion center solves the eigen-decomposition problem  $\mathbf{C}\mathbf{w} = \lambda\mathbf{w}$ , where  $\mathbf{C} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^\top$  is the covariance of the data (assumed centered). The eigenvector associated with the largest eigenvalue corresponds to the principal axis, denoted  $\mathbf{w}_*$  in the following.

Since the covariance matrix  $\mathbf{C}$  is a  $\Theta$ -by- $\Theta$  matrix, the computational complexity of such operation is  $\mathcal{O}(\Theta^3)$ . Furthermore, the communication of  $N$  sensors with a fusion center costs  $\mathcal{O}(N\Theta)$  on an average distance  $\mathcal{O}(1)$ . In order to overcome these communication and computational complexities, we propose next strategies that satisfy the scalability constraint in WSNs, namely without the need of a fusion center and without resorting to the covariance matrix.

#### B. Noncooperative strategy

Considering sensor-to-sensor scheme [21], we propose to adaptively learn the principal axis, drawing inspiration from Oja's rule [18]. According to a routing process, a sensor  $i$  receives from another sensor the estimation  $\mathbf{w}_{t-1}$ , and adjusts it with the following update rule:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \eta_t (\mathbf{y}_i z_i - z_i^2 \mathbf{w}_{t-1}), \quad (1)$$

where  $\eta_t$  is the learning rate which can also change with iteration and  $z_i = \mathbf{w}_{t-1}^\top \mathbf{y}_i$ . This learning rule converges to the first principal axis  $\mathbf{w}_*$ . In fact, when  $\mathbf{w}_t$  converges to some state  $\mathbf{w}$ , we have  $\mathbf{y}_i z_i = z_i^2 \mathbf{w}$ , or equivalently  $\mathbf{y}_i \mathbf{y}_i^\top \mathbf{w} = \mathbf{w}^\top \mathbf{y}_i \mathbf{y}_i^\top \mathbf{w} \mathbf{w}$ . Averaging over the whole data, we get the well-known eigen-decomposition problem of the covariance matrix  $\mathbf{C}\mathbf{w} = \mathbf{w}^\top \mathbf{C} \mathbf{w} \mathbf{w}$ , where the eigenvalue  $\mathbf{w}^\top \mathbf{C} \mathbf{w}$  corresponds to the squared output  $z_i$  that one wishes to maximize. Therefore, the update rule (1) converges to the largest eigenvector of the covariance matrix, without the need to compute  $\mathbf{C}$ . Communication between sensors in this strategy costs  $\mathcal{O}(N)$  on an average distance  $\mathcal{O}(1/\sqrt{N})$ .

#### C. Cooperative strategies

In a cooperative scheme, each sensor  $k$  has access to the information of its neighborhood  $\mathcal{V}_k$ , which is the subset of nodes currently connected to node  $k$ , including itself. In the following, we study two diffusion strategies.

*Combine-then-adapt strategy:* At every iteration  $t$ , each sensor  $k$  combines into  $\phi_{k,t-1}$  its estimate with the estimates from its neighborhood, namely  $\{\mathbf{w}_{l,t-1}\}_{l \in \mathcal{V}_k}$ . Then it updates its own estimate using an update rule similar to (1) where  $\mathbf{w}_{k,t-1}$  is substituted by  $\phi_{k,t-1}$ . These two steps are described as follows:

$$\begin{aligned} \phi_{k,t-1} &= \sum_{l \in \mathcal{V}_k} a_{kl} \mathbf{w}_{l,t-1} \\ \mathbf{w}_{k,t} &= \phi_{k,t-1} + \mu_k (\mathbf{y}_k z_k - z_k^2 \phi_{k,t-1}). \end{aligned}$$

Here,  $\mu_k$  is the stepsize for node  $k$ , and  $a_{kl}$  are convex combiner coefficients satisfying the conditions:

$$a_{kl} \geq 0, \quad \sum_{l \in \mathcal{V}_k} a_{kl} = 1. \quad (2)$$

*Adapt-then-combine strategy:* In this strategy, at every iteration  $t$ , each sensor  $k$  updates its own estimate and then fuses it with those from its neighbors into  $\phi_{k,t-1}$ . These two steps are described by the following update rule:

$$\begin{aligned}\phi_{k,t} &= \mathbf{w}_{k,t-1} + \mu_k (\mathbf{y}_k z_k - z_k^2 \mathbf{w}_{k,t-1}) \\ \mathbf{w}_{k,t} &= \sum_{l \in \mathcal{V}_k} a_{kl} \phi_{l,t}.\end{aligned}$$

Here,  $\mu_k$  denotes the stepsize for node  $k$ , and  $a_{kl}$  are convex combiner coefficients satisfying the conditions (2).

#### IV. CONVERGENCE ANALYSIS AND PARAMETER STUDY

Both adapt-then-combine and combine-then-adapt strategies are “diffusion” strategies since the combination step allows information to diffuse through the network. These strategies have fundamentally the same structure. The difference lies in which variable we choose to correspond to the updated estimate. In the adapt-then-combine case, the estimate is the result of the combination step, whereas in the combine-then-adapt case the estimate is the result of the adaptation.

In the following, we provide a convergence analysis and study the choice of the parameters for the strategies proposed in the previous section.

##### A. Stepsize in the noncooperative strategy

In the noncooperative strategy, the stepsize parameter  $\eta_t$  is chosen to be sufficiently small to ensure convergence. It should be smaller than the inverse of the largest eigenvalue [20]. Unfortunately, the eigenvalues are usually unknown. For appropriate convergence, the stepsize  $\eta_t$  should decrease at each iteration such as  $\eta_t = \eta_0/t$ , where  $\eta_0$  is a positive constant parameter. An alternative approach is the “search-then-converge” scheme [19], with  $\eta_t = \frac{\eta_0}{1+t/\tau}$ . In this case, the delay parameter  $\tau$  determines the duration of the initial search phase, with  $\eta_t \simeq \eta_0$  when  $t \ll \tau$ , before a converge phase where  $\eta_t$  decreases as  $\eta_0/t$  when  $t \gg \tau$ .

##### B. Stepsize for the adaptation in diffusion strategies

In the diffusion strategies, the stepsize parameter  $\mu_k$  is chosen small to ensure the convergence. It can also decrease with iterations, such as with  $\mu_{k,t} = \frac{\mu_{k,0}}{1+t/\tau}$ . We notice that this stepsize converges toward zero as  $t \rightarrow \infty$ . Therefore it turns off the adaptation. This choice is not suitable for continuous learning. For this reason, we choose a constant stepsize to ensure continuous adaptation.

##### C. Coefficients for the combination in diffusion strategies

The combiner coefficients  $a_{kl}$  used in the diffusion strategies determine the weight that each node  $k$  assigns to the estimates from its neighbors  $l \in \mathcal{V}_k$ . These coefficients need to satisfy the convexity conditions (2). By revisiting the suggestions given in the adaptive filtering literature [22], we consider the following rules to define these coefficients:

1) Averaging rule:

$$a_{kl} = \begin{cases} 1/n_k, & \text{when } l \in \mathcal{V}_k; \\ 0, & \text{otherwise.} \end{cases}$$

2) Laplacian rule:

$$a_{kl} = \begin{cases} 1/n_{\max}, & \text{if } l \in \mathcal{V}_k \setminus \{k\}; \\ 1 - (n_k - 1)/n_{\max}, & \text{if } k = l; \\ 0, & \text{otherwise.} \end{cases}$$

3) Metropolis rule:

$$a_{kl} = \begin{cases} 1/\max\{n_k, n_l\}, & \text{if } l \in \mathcal{V}_k \setminus \{k\}; \\ 1 - \sum_{j \in \mathcal{V}_k \setminus \{k\}} a_{kj}, & \text{if } k = l; \\ 0, & \text{otherwise.} \end{cases}$$

In these expressions,  $n_k$  denotes the degree of node  $k$ , i.e., the size of its neighborhood, and  $n_{\max}$  denotes the maximum degree across the network, i.e.,  $n_{\max} = \max_{1 \leq k \leq N} n_k$ .

#### V. EXPERIMENTS

In order to illustrate the results of this work, we consider the framework proposed in [23], where we have the diffusion of a gas in a two-dimensional space  $\mathbb{X} = [-0.5, 0.5] \times [-0.5, 0.5]$ . This distribution is governed by the differential equation

$$\frac{\partial G(\mathbf{x}, \theta)}{\partial \theta} - c \nabla_{\mathbf{x}}^2 G(\mathbf{x}, \theta) = Q(\mathbf{x}, \theta),$$

where  $G(\mathbf{x}, \theta)$  is the density of gas depending on the position  $\mathbf{x}$  and time  $\theta$ ,  $\nabla_{\mathbf{x}}^2$  is the Laplace operator, and  $Q(\mathbf{x}, \theta)$  corresponds to the added quantity of gas and  $c$  the conductivity of the medium. A gas source placed at position  $(-0.3, 0.3)$  is activated from  $\theta = 1$  to  $\theta = 15$ . We use  $N = 100$  sensors deployed uniformly in the region under scrutiny  $\mathbb{X}$ , each acquiring a time series of 15 measurements, from  $\theta = 1$  to  $\theta = 15$ .

We consider a predetermined range of communication in the WSN. In this case, two nodes are connected when they are less than  $s$  units of distance apart, that is

$$\mathcal{V}_k = \{l : \|\mathbf{x}_k - \mathbf{x}_l\| < s\}.$$

In our experiments, we set this threshold to  $s = 0.15$ . For the stepsize parameters (see Section IV), we consider as a reference the constant stepsize  $\eta_0 = 0.005$ . We compare the results with the stepsize given by  $\eta_t = \frac{\eta_0}{1+t/\tau}$ , where the delay parameter  $\tau$  is given by the third of the total number of nodes. We also use the same stepsize value for the cooperative strategies, namely  $\mu_k = 0.005$ .

In order to provide a comparative study, we use the same initial random estimate for all strategies. The performance is measured with the angle between the principal axis  $\mathbf{w}_*$ , obtained from the centralized strategy with the eigen-decomposition of the covariance matrix, and the estimate  $\mathbf{w}_l$  at node  $l$ , namely

$$\arccos\left(\frac{\mathbf{w}_l^\top \mathbf{w}_*}{\|\mathbf{w}_l\| \|\mathbf{w}_*\|}\right). \quad (3)$$

Fig. 2 shows the convergence of each strategy, with the angles averaged over the  $N$  nodes. These learning curves show that the noncooperative strategy is more suitable when a variable stepsize is used, as opposed to the constant one where convergence is not guaranteed. This strategy is outperformed by any diffusion strategy, independently of the combination

rule. The analysis of the diffusion strategies shows that the adapt-then-combine strategy performs slightly better than the combine-then-adapt strategy. In terms of combination rules, the averaging rule provides a slight refinement. The overall efficiency of the diffusion strategies is shown in terms of stability.

## VI. CONCLUSION

In this paper, we studied the issue of estimating the principal axis from PCA in WSNs. Under constraints imposed in WSNs, we proposed several strategies including noncooperative and diffusion strategies. Experimental results showed the relevance of these strategies, and illustrated that the adapt-then-combine strategy with the averaging rule outperforms the other strategies. As a future work, we will extend this technique to the case of multiple principal axis by using a generalized Hebbian approach. We will also study the use of the spatial information.

## ACKNOWLEDGMENT

This work is supported by Université de technologie de Troyes, Université Libanaise, and Région Champagne-Ardenne (grant "WiDiD").

## REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, pp. 102–114, 2002.
- [2] S. W. Conner, J. Heidemann, L. Krishnamurthy, X. Wang, and M. Yarvis, "Workplace applications of sensor networks," in *Wireless Sensor Networks: A Systems Perspective*, N. Bulusu and S. Jha, Eds. Artech House Publishers, 2004.
- [3] A. Czubak and J. Wojtanowski, "On applications of wireless sensor networks," in *Internet - Technical Development and Applications*, S. B. Heidelberg, Ed., vol. 64, 2009, pp. 91–99.
- [4] J. Polastre, R. Szewczyk, A. Mainwaring, D. Culler, and J. Anderson, "Analysis of wireless sensor networks for habitat monitoring," in *Wireless Sensor Networks*, C. Raghavendra, K. Sivalingam, and T. Znati, Eds. Springer US, 2004, pp. 399–423.
- [5] Y. Ma, M. Richards, M. Ghanem, Y. Guo, and J. Hassard, "Air pollution monitoring and mining based on sensor grid in london," *Sensors*, vol. 8, no. 6, pp. 3601–3623, 2008.
- [6] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed kernel regression: An algorithm for training collaboratively," in *IEEE Proc. Information Theory Workshop*, 2006.
- [7] P. Honeine, M. Essoloh, C. Richard, and H. Snoussi, "Distributed regression in sensor networks with a reduced-order kernel model," in *Proc. 51st IEEE GLOBECOM Global Communications Conference*, New Orleans, LA, USA, 2008.
- [8] Y. Bergamo and C. Lopes, "Scalar field estimation using adaptive networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3565–3568.
- [9] J. Chen and A. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4289–4305, Aug 2012.
- [10] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [11] N. Chitradevi, K. Baskaran, V. Palanisamy, and D. Aswini, "Designing an efficient PCA based data model for wireless sensor networks," in *Proceedings of the 1st International Conference on Wireless Technologies for Humanitarian Relief*, ser. ACWR '11. New York, NY, USA: ACM, 2011, pp. 147–154.
- [12] F. Chen, F. Wen, and H. Jia, "Algorithm of data compression based on multiple principal component analysis over the wsn," in *Wireless Communications Networking and Mobile Computing (WiCOM), 6th International Conference on*, Sept 2010, pp. 1–4.

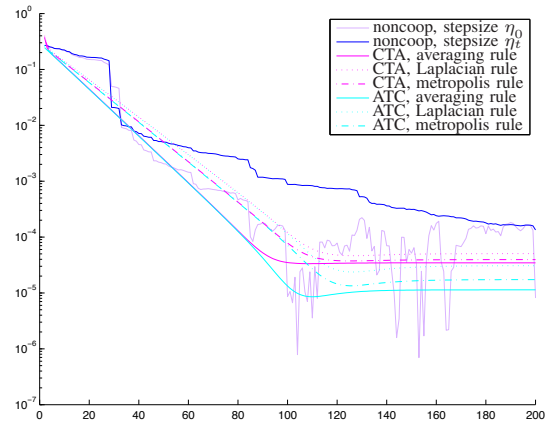


Fig. 2. Convergence over 200 iterations of the proposed strategies, measured by the angle given in (3) and averaged over all the nodes. The confronted strategies are the noncooperative strategy, with either a constant stepsize  $\eta_0$  or a decreasing one  $\eta_t = \frac{\eta_0}{1+t/\tau}$ , and both combine-the-adapt (CTA) and adapt-then-combine (ATC) diffusion strategies with several rules for combination defined in Section IV-C.

- [13] A. Rooshenas, H. Rabiee, A. Movaghar, and M. Naderi, "Reducing the data transmission in wireless sensor networks using the principal component analysis," in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sixth International Conference on*, Dec 2010, pp. 133–138.
- [14] M. Ahmadi Livani and M. Abadi, "A PCA-based distributed approach for intrusion detection in wireless sensor networks," in *Computer Networks and Distributed Systems (CNDS), 2011 International Symposium on*, Feb 2011, pp. 55–60.
- [15] L. Huang, M. I. Jordan, A. Joseph, M. Garofalakis, and N. Taft, "In-network pca and anomaly detection," in *NIPS*. MIT Press, 2006, pp. 617–624.
- [16] H. Kargupta, W. Huang, K. Sivakumar, B.-H. Park, and S. Wang, "Collective principal component analysis from distributed, heterogeneous data," in *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*. London, UK, UK: Springer-Verlag, 2000, pp. 452–457.
- [17] Y.-A. Le Borgne, S. Raybaud, and G. Bontempi, "Distributed principal component analysis for wireless sensor networks," *Sensors*, vol. 8, no. 8, pp. 4821–4850, 2008.
- [18] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, November 1982.
- [19] P. Honeine, "Online kernel principal component analysis: a reduced-order model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1814–1826, September 2012.
- [20] L.-H. Chen and S. Chang, "An adaptive learning algorithm for principal component analysis," *IEEE Transactions on Neural Networks*, vol. 6, no. 5, pp. 1255–1263, 1995.
- [21] P. Honeine, C. Richard, H. Snoussi, J. C. M. Bermudez, and J. Chen, "A decentralized approach for non-linear prediction of time series data in sensor networks," *Journal on Wireless Communications and Networking*, 2010.
- [22] A. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 155–171, May 2013.
- [23] N. Ghadban, P. Honeine, F. Mourad-Chehade, C. Francis, and J. Farah, "Mobility using first and second derivatives for kernel-based regression in wireless sensor networks," in *Proc. 21st International Conference on Systems, Signals and Image Processing*, Dubrovnik, Croatia, 12–15 May 2014.