



Blind source separation techniques for percolation type leakage detection in dikes using fiber optic DTS signals

Amir Ali Khan, Valeriu Vrabie, Guy d'Urso, Jerome I. Mars

► To cite this version:

Amir Ali Khan, Valeriu Vrabie, Guy d'Urso, Jerome I. Mars. Blind source separation techniques for percolation type leakage detection in dikes using fiber optic DTS signals. 19th International Conference on Optical Fibre Sensors, Apr 2008, Perth, Australia. pp.SPIE-70042N, 10.1117/12.785995 . hal-00326386

HAL Id: hal-00326386

<https://hal.science/hal-00326386>

Submitted on 2 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blind Source Separation Techniques for Percolation type Leakage Detection in Dikes using Fiber Optic DTS Signals

Amir A. Khan^{*a}, Valeriu Vrabie^b, Guy D'Urso^c, and Jérôme I. Mars^a

^aGIPSA-Lab—Department Image Signal, Grenoble Institute of Technology, BP 46, 38402 Saint Martin d'Hères, France;

^bCentre de Recherche en STIC (CReSTIC), Université de Reims, BP 1039, 51687 Reims, France;

^cDirection R&D, Electricité de France (EDF), 6 quai Watier, 78401 Chatou, France.

ABSTRACT

Distributed temperature sensors (DTS) based on fiber optics present an efficient means for temperature data acquisition. The use of DTS data to detect leakages in dikes necessitates some processing of this data. Formulating leakage detection as a source separation problem, the goal of this paper is to compare various blind source separation techniques for percolation type leakage detection using a real data set. Singular value decomposition can be used as the first step to separate out the ground response where acquisitions are made. Two independent component analysis algorithms, JADE and FastICA, assuming independence of sources, are tested to find the best solution for leakage detection.

Keywords: Distributed Temperature Sensors, Leakage Detection, Source Separation

1. INTRODUCTION

The use of fiber optic sensors has been a common practice in diverse domains covering applications as engineering structures monitoring, fault detection in electrical circuits, fire detection systems, parameter sensing in oil and gas industry, etc.^{1,2} Now-a-days, an important issue in engineering domain is the detection of anomalies, such as leakages (significant flow of water), in dikes to avoid disaster at mass level. One of the most promising methods for this purpose is thermometric based method employing the Distributed Temperature Sensors (DTS) based on fiber optics. The major advantage of DTS is their commercial viability (low-cost telecommunications grade fiber) and ability to multiplex large number of sensors along a single fiber.³

The thermometric data acquired through DTS contains useful information, but this information is not directly interpretable for leakage identification and localization. The basic concept behind temperature acquisition is that a change of ground temperature is brought about by a significant flow of water through the structure due to leakages. However, this change can equally be brought about by other factors (or sources) such as seasonal variations, precipitation, existing structures (e.g. drains) etc. In this regard, treating the acquired data as a linear mixture of all these sources, the problem of leakage detection is formulated as that of blind source separation (BSS). The blind source separation techniques have been employed in various areas such as image processing, geophysics, biomedical engineering, neural networks, food quality assessment, etc.⁴⁻⁶ The goal of this paper is to present a comparative analysis of some source separation and data decomposition techniques from the point of view of percolation type leakage detection. The techniques employed in this paper include singular value decomposition (SVD) and two algorithms of independent component analysis (ICA).

2. SYSTEM AND DATA DESCRIPTION

The temperature data is acquired through DTS using Optical Time Domain Reflectometry (OTDR) technique based on the Raman backscattering principle. A thermometric data monitoring system has been installed, by Electricité de France (EDF), at an experimental test site in the south of France (near Oraison city) with an aim to study leakages (both natural and controlled) in the dike of canal (see Fig. 1(a)). A 2.2 km long fiber optic cable (containing 4 optic fibers, of type multimode 50/125), is buried at the downstream toe of the canal at a depth of 1 m to intercept water leakage from the canal. Two distinct zones, (Zone1, from approximately 0.2 km to 1.25 km and Zone2, from approximately 1.25 km to 2.2 km), corresponding to two different elevation levels,

*E-mail: amir-ali.khan@gipsa-lab.inpg.fr, Phone: +33.4.76.82.63.34

will be exposed with varying intensities to direct sunlight. The cable also circumvents two drains, D1 and D2, located at 0.561 km and 0.858 km, respectively. The temperature data were recorded by a commercial device SensorNet, Sentinel DTS-MR, with temperature and spatial resolutions of 0.01°C and 1-meter, respectively.

The temperature data are obtained along the entire length of the fiber with a sampling distance of 1m. To monitor temporal evolution of the anomalies, acquisitions were made over a period of 14 days with a sampling interval of 2 hours. The recorded data set is thus a two-dimensional temperature signal, $\mathbf{Y} = \{y(t, x) \mid 1 \leq t \leq N_t, 1 \leq x \leq N_x\}$ where $N_t = 168$ and $N_x = 2200$ are the total acquisition time and the number of observation points, respectively. Three artificial leakages, $L1$ (on day 10), $L2$ and $L3$ (on day 12), were introduced at the site with different flow rates of 5, 1 and 1 lit/min and at different positions, 1.562, 1.547 and 1.569 km, respectively. A hot point (HP) was also introduced on day 10 at 0.674 km.

3. SOURCE SEPARATION TECHNIQUES

The acquired temperature data can be considered as a linear mixture of various factors with a source signal, as a function of displacement, assigned for each factor. Certain factors such as drains or leakages can be modeled by sparse sources, which means that these sources are non-Gaussian. Since the sources originate from non-correlated phenomena, they are supposed to be independent of each other. The acquired data can thus be written as:

$$\mathbf{Y} = \mathbf{M}\mathbf{S}^T, \quad (1)$$

where $\mathbf{M} \in \mathbb{R}^{N_t \times p}$ is the unknown mixing matrix, $\mathbf{S} \in \mathbb{R}^{N_x \times p}$ designates the matrix made up of the p independent sources (mentioned above) to be estimated, and T denotes matrix transposition. Blind source separation consists in recovering these unknown source signals without any a priori information about the sources or their mixing coefficients. This lack of a priori information is partially compensated by physically plausible assumptions on the sources such as decorrelation and independence. Here, we investigate some of the classical data decomposition and blind source separation techniques like the Singular Value Decomposition (SVD) and Independent Component Analysis (ICA) from leakage detection point of view. We will only briefly discuss each of the following techniques and the interested reader may refer to the more detailed references. The SVD of a signal can be defined as:^{6,7}

$$\mathbf{Y} = \mathbf{Y}_{signal} + \mathbf{Y}_{noise} = \sum_{j=1}^m \sigma_j \mathbf{u}_j \mathbf{v}_j^T + \sum_{j=m+1}^N \sigma_j \mathbf{u}_j \mathbf{v}_j^T. \quad (2)$$

where $N = \min(N_t, N_x)$, $\sigma_j \geq 0$ are the singular values arranged in a descending order, $\mathbf{v}_j \in \mathbb{R}^{N_x}$ are the right singular vectors (identified as estimators of the sources) and $\mathbf{u}_j \in \mathbb{R}^{N_t}$ are left singular vectors (temporal variations of estimated sources). Note that SVD can be linked to Principal Component Analysis (PCA) when the principal components are calculated from the covariance matrix of the data, \mathbf{Y} . SVD imposes the criteria of orthogonalization of sources that is not sufficient to extract factors such as drains or leakages, which can be modeled by sparse and so non-Gaussian sources. SVD, nevertheless, allows to decompose the initial data into two complementary subspaces, namely, the signal subspace, \mathbf{Y}_{signal} , and the noise subspace, \mathbf{Y}_{noise} . The parameter m designates the number of greatest singular values used to construct the signal subspace with SVD.

ICA is a BSS computational technique which consists of finding a linear transform $\mathbf{S}^T = \mathbf{W}\mathbf{Y}$ such that the components s_i , i.e. columns of \mathbf{S} , are as independent as possible in the sense of maximizing some measure of independence, where \mathbf{W} is the unmixing matrix. The fundamental criteria of formulating ICA is maximization of non-Gaussianity of sources, s_i . ICA model has two indeterminacies: the energies (variance) and the order of estimated sources can not be determined. In our application, the goal is to detect the leakages and localize them in distance, thus these indeterminacies do not pose any restriction. Two popular ICA algorithms usually employed are JADE⁸ and FastICA.⁴ JADE exploits the fact that for Gaussian random variables, all the cumulants higher than second order are zero. In particular, it uses the 4th order cumulants which are considered suitable for instantaneous mixtures. The 4th order cumulants of the decorrelated sources, \mathbf{v}_j , given by SVD in Eq. (2) are first calculated to obtain a 4-dimensional tensor. The joint diagonalization of cumulant matrices obtained from this tensor gives the independent sources, s_i . FastICA is based on a fixed-point iteration scheme employing the maximization of negentropy. This algorithm offers not only a very fast convergence but also finds directly

independent components of any non-Gaussian distribution using a nonlinearity and thus requires no estimate of probability distribution function of the sources. Moreover, the independent sources can be estimated one by one in the deflation mode, which is equivalent to a projection pursuit. A detailed description of these algorithms can be found in Hyvärinen.⁴

4. RESULTS AND DISCUSSION

These techniques are applied on a real temperature data set acquired over a period of 14 days (Fig. 1(b)). The first source vector, \mathbf{v}_1 , obtained by application of SVD (Fig. 1(c)) represents 80% of the total energy. Two elevation zones, present at the actual site, are revealed in this source thus linking it to the ground response. This source also contains partial information linked to the drains. The leakages, however, remain present mixed with the residual drain information in the remaining sources and further separation is not possible. The ground response is separated out by constructing a signal subspace with this source ($m = 1$ in Eq. (2)) and the noise subspace contains information pertaining to leakages and drains.

We next focus our attention to ICA, where the noise subspace obtained from SVD is used as initial data. Tests were performed using the two ICA algorithms with different parameters to find out an optimum separation solution for leakage detection. There are two modes for source estimation using ICA: Symmetric mode and Deflation mode. The former estimates all sources in parallel whereas the later operates on the principle of calculating one source at a time, i.e., in each recursion (step), it estimates one source working on the noise space obtained by removal of source subspace estimated in previous step.

A dimensionality reduction should be performed for the symmetric mode due to computational requirements, chosen here to take into account 6 sources, \mathbf{v}_2 to \mathbf{v}_7 , associated to the 6 expected sources ($D1$, $D2$, $L1$, $L2$, $L3$ and HP). The results obtained in the symmetric mode for both JADE and FastICA are quite similar and tests revealed that at most 3 unique sources can be identified amongst the 6 estimated sources, s_i , as shown in Fig. 2(a) for JADE algorithm. The 1st source estimates the drains, $D1$ and $D2$, the 2nd, 4th and 6th estimate some portion of drains mostly amid noise whereas the 3rd and 5th estimate the leakage $L1$ and the hot point, HP . A construction of the residue subspace subtracting the subspace corresponding to the first two sources would reveal the leakages and the hot point. It was observed that the most energetic leakage, $L1$, can be identified in the estimated sources, whereas the remaining two leakages are present in the residue but are not revealed in the estimated sources, even if we consider more than 6 sources. In order to better identify these remaining sources, we propose another method based on deflation without dimensionality reduction.

The deflation principle is not incorporated by default in JADE but we can adopt this principle for JADE by estimating one source at a time, each time working on the last obtained noise subspace. However, the dimensionality reduction constraint always exists for JADE and the results obtained with deflation mode for JADE are similar to symmetric mode. On the other hand, the computational efficiency of FastICA allows us to treat the data without first performing the dimensionality reduction. FastICA without dimensionality reduction is thus applied in the deflation mode in two time zones, one containing no leakages and other containing leakages. The first 6 sources obtained are shown in Fig. 2(b-c). Note that all plots in Fig. 2 have the same dynamics for the amplitude range. For the zone without the leakages, the estimated sources (Fig. 2(b)) reveal only the drains and noise. In the zone containing the leakages, it can be observed (Fig. 2(c)) that all phenomena present in the data, the three leakages, the hot point and the drains, are detected at their exact locations with $L1$ detected in source 1, hot point in source 2 and $L2$ and $L3$ in sources 3 and 5. The fact that the drains are not localized in distance, but to few meters, means that they will be detected on more than one source at positions differing by few meters, as observed in sources 4 and 6. This is a well known phenomenon for real data introduced by application of BSS, which tends to estimate impulsive sources as the number of estimated sources increases. The sources obtained with different nonlinearities for FastICA are quite similar.

5. CONCLUSION

In this paper, the problem of percolation type water leakage detection in dikes, using temperature data from fiber optic DTS, is treated as a source separation problem. A study of various source separation algorithms was performed in order to propose an optimum solution for percolation type leakage detection. It was shown that the

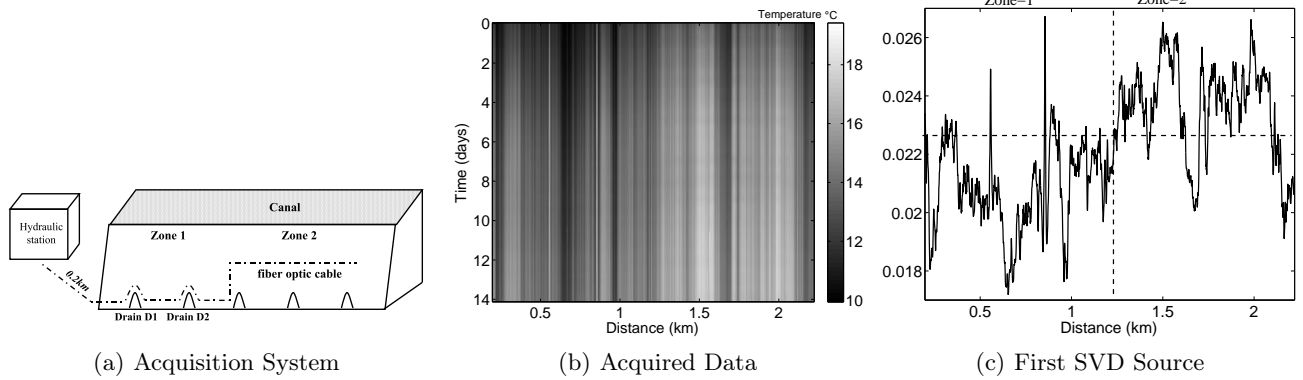


Figure 1. Data Acquisition system, temperature data set and the first SVD source.

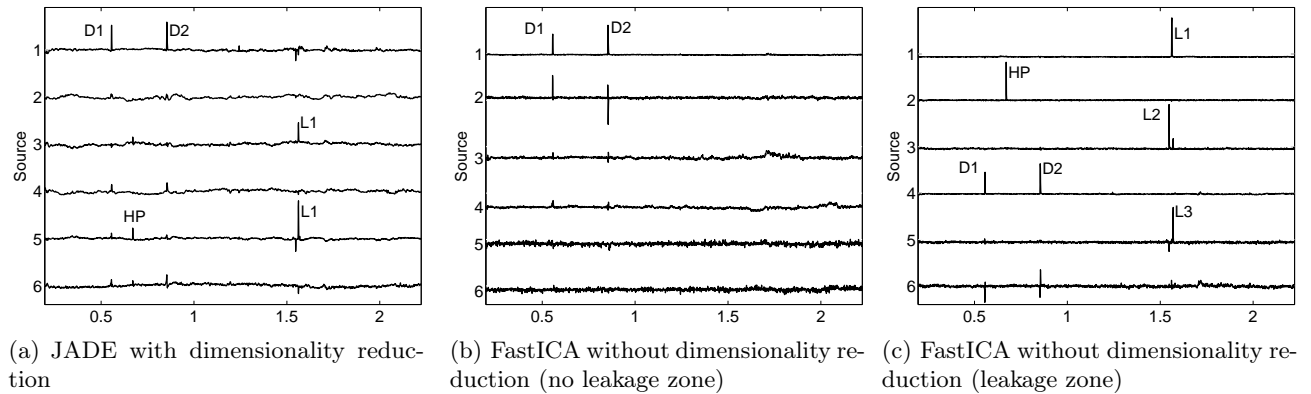


Figure 2. Estimated sources with ICA in (a) symmetric and (b-c) deflation modes as function of distance (km).

first SVD source represents ground response and thus corresponding subspace can be separated out to give better access to leakages and other phenomena. The ICA algorithms utilized in symmetric mode with dimensionality reduction reveal leakages in the residue but only the most energetic leakage is detected by the sources. FastICA, run without dimensionality reduction in deflation mode, however, clearly identifies all the leakages and the hot point on different sources and thus gives a perfect detection.

REFERENCES

1. A. H. Hartog, "Progress in distributed fiber-optic temperature sensing," in *Proc. SPIE, Fiber Optic Sensor Technology and Applications, 2001*, M. A. Marcus and B. Culshaw, eds., **4578**, pp. 43–52.
2. S. Grosswig, A. Graupner, and E. Hurtig, "Distributed fiber optical temperature sensing technique - a variable tool for monitoring tasks," in *Proc. 8th Intl. Symposium on Temperature and Thermal Measurements in Industry and Science*, pp. 9–17, 19–21, June 2001.
3. B. Vogel, C. Cassens, A. Graupner, and A. Trostel, "Leakage detection systems by using distributed fiber optical temperature measurements," in *Proc. SPIE Smart Structures and Materials*, **4328**, pp. 23–34, 2001.
4. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.
5. E. Lewis, C. Sheridan, M. O'Farrell, and D. King, "Principal component analysis and artificial neural networks based approach to analysing optical fiber sensors signals," *Sensors and Actuators A* **136**, pp. 28–38, 2007.
6. V. Vrabie, J. I. Mars, and J. L. Lacoume, "Singular value decomposition by means of independent component analysis," *Signal Processing* **84**(3), pp. 645–652, 2004.
7. L. L. Scharf, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*, Addison-Wesley, New York, 1991.
8. J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings-F* **140**(6), pp. 362–370, 1993.