



**HAL**  
open science

## Automatic monitoring system for singularities detection in dikes by DTS data measurement

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

► **To cite this version:**

Amir Ali Khan, Valeriu Vrabie, Jerome I. Mars, Alexandre Girard, Guy d'Urso. Automatic monitoring system for singularities detection in dikes by DTS data measurement. *IEEE Transactions on Instrumentation and Measurement*, Institute of Electrical and Electronics Engineers, 2010, 59 (8), pp.2167-2175. 10.1109/TIM.2009.2032880 . hal-00420908

**HAL Id: hal-00420908**

**<https://hal.archives-ouvertes.fr/hal-00420908>**

Submitted on 30 Sep 2009

**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.

# Automatic monitoring system for singularities detection in dikes by DTS data measurement

Amir A. Khan<sup>\*1</sup>, Valeriu Vrabie<sup>2</sup>, Jérôme I. Mars<sup>1</sup>, Alexandre Girard<sup>3</sup>, and Guy D'Urso<sup>3</sup>

<sup>1</sup>GIPSA-Lab, Department Image and Signal Processing, Grenoble Institute of Technology,  
BP 46, 38402 Saint Martin d'Hères, France.

<sup>2</sup>Centre de Recherche en STIC (CReSTIC), Université de Reims, BP 1039, 51687 Reims, France.

<sup>3</sup>Direction R&D, Electricité de France (EDF), 6 quai Watier, 78401 Chatou, France.

Tel: +33 (0)4 76 82 71 07 Fax: +33 (0)4 76 57 47 90

e-mail: amir-ali.khan@gipsa-lab.inpg.fr

## Abstract

The development of automated monitoring systems for detection of singularities such as leakages in dikes is indispensable to avoid mass disaster. An efficient solution for dikes survey is the use of Distributed Temperature Sensors (DTS) based on optical fiber, offering multitude of advantages such as low-cost, extreme robustness, long range measurement, etc. However, the temperature data acquired with DTS, being not directly interpretable, require intervention of signal processing techniques. The current work addresses this signal processing aspect, exploiting the key idea that the temperature variations over the course of a day for singular zones are quite different from those for nonsingular zones. A daily reference temperature variation, representative of the nonsingular zones, is estimated using the Singular Value Decomposition (SVD). The residue subspace of SVD contains information linked to the deviations from this reference; thus allowing the degree of singularity to be quantified by a dissimilarity measure such as the L2 norm. In order to detect only the singularities in dikes, such as leakages or drains, a Constant False Alarm Rate (CFAR) detector is proposed by modeling each daily dissimilarity measure with a mixture of Gamma and Uniform distributions. The proposed automatic singularity detection system was validated under different scenarios on real data over periods from 2005 to 2007. The first scenario depicted the detection of percolation type artificial leakages with their detection strength depending on their flow rates. Another scenario allowed to detect the presence of a real water leakage at the site, previously unobserved during manual inspections. The repeatability of the system was also verified by periodic analysis.

## Index Terms

Distributed temperature sensors, Dikes, Automated system, Leakage detection, Probability of false

alarm, Singular Value Decomposition, Thermometric data.

## I. INTRODUCTION

The recent advances in development and deployment of fiber optic sensors as measurement technology has resulted in a multitude of interesting applications. The sectors that have greatly benefited include monitoring of engineering structures, parameter sensing in oil and gas industry, chemical sensing for environmental control and safety, early fire detection systems, biomechanical engineering, etc [1]–[5]. Many aging engineering structures like dams and dikes become vulnerable due to degradations caused by internal erosion, adverse climatic conditions and natural damage. It is therefore imperative to identify the anomalies, like the water leakages in dikes, well in advance in order to avert disaster at mass level. The conventional methods for leakage detection in dikes were based on manual inspections at the site. The measurement of different parameters like pressure, flow, temperature, deformation, etc. are also used as indicators for anomaly detection. The systems currently employed are based on measurement of self-potential, resistivity and temperature [6]–[9]. While the self-potential and the resistivity methods provide efficient measurements, the drawback lies in the fact that these methods are still manual. Moreover, they require careful placement and utilization of electronic equipments at site. On the contrary, the temperature based methods present a semi-automatic solution through the use of fiber optic based distributed temperature sensors (DTS). The fact that DTS use low-cost telecommunication grade optical fiber with an ability to multiplex large number of sensors enhance their economic viability while at the same time their robustness and immunity to electromagnetic interference allows their deployment in harsh environments. DTS are capable of long range measurements of up to 30 km with different spatial and temperature resolutions (a typical scale of measurement is 1 m for spatial resolution and  $0.01^{\circ}\text{C}$  for temperature resolution achievable with Sensornet's Sentinel device).

The method adopted for data acquisition by DTS is *passive method*, a natural measure of temperature. The method takes its name from the fact that in the absence of any anomaly, the measured temperature is driven by the phenomenon of conduction : the transfer of heat is attributed to the interaction between the air temperature and the temperature of water present naturally in the ground. The basic idea for anomaly identification using DTS systems is that a significant flow of water through the dike due to leakage brings along additional heat by the phenomenon of advection [7]. Thus, when advection superposes conduction, thermometric measurements can help in detecting leakages. The leakage detection depends not only on the flow rate but also on the temperature difference between ground (which in turn depends on the air temperature depending on the depth) and water. The acquired temperature at certain distances can

also be influenced by the existing structures (drains) as well as by the self-response of the ground in which the optical fiber cable is buried. In addition, precipitations, seasonal effects, day/night variations, etc. can also bring some change in ground temperature. The thermometric data acquired through DTS is therefore not directly interpretable for anomaly detection thus posing a new research problem for the signal processing community. The leakages, the drains and the singularity of the ground response constitute the possible singularities along the length of the dike. The contribution of this work is to propose a new system for singularity detection in dikes with the eventual goal of having an automated functional device. The aim is to localize the singularities in distance for further investigation with detailed physical inspections. The proposed system exploited the temperature variations over the course of a day at all the sensing distances. While, most of the distances showed a common trend of these variations, the singularity zones presented a deviation from this common trend. Thus, considering the temperature signal for all the distances over a 24-hour period, a dissimilarity method for singularity identification based on the classical data decomposition technique like Singular Value Decomposition (SVD) [10]–[12] is proposed. The resulting dissimilarity measure is then thresholded using constant false alarm rate criteria by modeling it as a mixture of Gamma and Uniform distributions.

We start in Section II with the working principle of the DTS acquisition systems along with a representation of the data set acquired at an experimental site of Electricité de France (EDF). The problem is formulated in Section III, with a brief theoretical description of SVD followed by the proposed methodology used for devising the dissimilarity measure and the proposed detector based on Constant False Alarm Rate (CFAR). Section IV deals with the application and validation of the proposed system on real temperature data sets. Particularly interesting amongst the tested scenarios are those of artificial leakages and the real water leakages. A comprehensive discussion is proposed to conclude the paper.

## II. ACQUISITION PRINCIPLE AND DATA DESCRIPTION

The temperature data were acquired using the optical fiber based Distributed Temperature Sensors which have proved to be highly efficient in a large number of applications [2], [13], [14]. Most of the commercial distributed temperature sensors are based on Raman scattering using Optical Time Domain Reflectometry (OTDR) technique [15]. The OTDR setup has three major components : a pulsed laser source, a directional coupler and an optical fiber cable which forms the temperature sensing element. The photons emitted by the laser source interact with the molecules of the fiber material. The thermally induced molecular vibrations result in backscattering of some photons attributed to Raman scattering [16]. The intensity ratio  $R(T)$ , between the two components of Raman backscattered light: the temperature

dependent Anti-Stokes and temperature independent Stokes scattering, serves as a measure of temperature:

$$R(T) = \left( \frac{\nu_0 + \nu_k}{\nu_0 - \nu_k} \right)^4 \exp \left( -\frac{h\nu_k}{kT} \right), \quad (1)$$

with  $\nu_0$  the frequency of the input laser,  $\nu_k$  the frequency-shift of Raman scattering,  $h$  the Planck's constant and  $k$  the Boltzmann's constant. Measuring this ratio and pulse travel time at the fiber input, the relationship between travel time and distance gives the temperature profile along the entire fiber length. DTS offer several advantages [13], [14] over conventional sensors, the major one being their commercial viability owing to the use of low-cost optical fiber cable. Moreover, they are capable of long-range measurements with high spatial (1 m) and temperature resolutions ( $0.01^\circ C$ ). The electrical isolation, immunity to electromagnetic interference and ruggedness associated with the optical fiber render these sensors extremely useful for harsh environments like oil and gas industry and engineering structures. The *a priori* knowledge about number and placement of sensors also becomes irrelevant. For all DTS devices, there is a trade off between temperature resolution, spatial resolution, range and speed of measurement. Allowing the DTS device more time to acquire data results in a higher temperature resolution at the cost of reduced measurement speed. Likewise, for a given acquisition time, the measurement range varies inversely with the temperature resolution.

A thermometric data monitoring system has been installed by Electricité de France (EDF) at an experimental test site to study the leakages. The aim of this site is to extract the information (location in distance, flow rate, etc.) pertaining to leakages (both natural and controlled) in the dike of a canal. A schematic representation of this installation is given in Fig. 1. A fiber optic cable containing 4 optical fibers (type multimode 50/125) was buried at the downstream toe of the canal at a depth of 1 m to intercept water leakages from the canal. The two distinct elevation levels present at the site (Zone 1, from approximately 0.1 km to 1.25 km and Zone 2, from approximately 1.25 km to 2.2 km) are exposed with varying intensities to direct sunlight. The cable also circumvents two drains, D1 and D2, situated at 0.561 km and 0.858 km, respectively. The temperature data was recorded by a commercial device Sensornet, Sentinel DTS-MR, a medium range device capable of covering up to 8 km. The temperature resolution of this device is  $0.01^\circ C$  with 1-meter spatial resolution which would allow us to detect very closely the occurrence of water leakages.

Each acquisition of the temperature signal is a function of displacement along the fiber. Continuous temperature monitoring is important in order to track the evolution of the leakages over time. Several acquisitions were therefore made during different periods from years 2005 to 2007 which allowed us to analyze different scenarios. A sampling interval of 1-hour gave us 24 acquisitions per day. The recorded

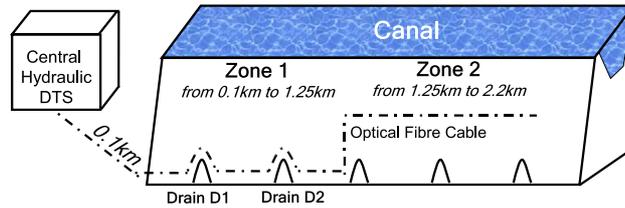


Fig. 1. Schematic of the data acquisition system using DTS with a temperature resolution of  $0.01^\circ C$  and a spatial resolution of 1 m at the experimental site. The 2.2 km long optical fiber cable circumvents two drains, whereas two different elevation levels are exposed with varying intensities to sunlight.

data set for a day “ $i$ ” can be written as:

$$\mathbf{Y}^i = \{y^i(t, x) \mid 1 \leq t \leq N_t, 1 \leq x \leq N_x\}, i = 1, \dots, N_i \quad (2)$$

where  $N_t = 24$  represents the number of acquisitions per day,  $N_x$  the number of observation points and  $N_i$  the number of days analyzed.

A sample real data set over a period of nonconsecutive 11 days in year 2005 with similar meteorological conditions is shown in Fig. 2. The gray scale represents the intensity of the recorded temperature. The

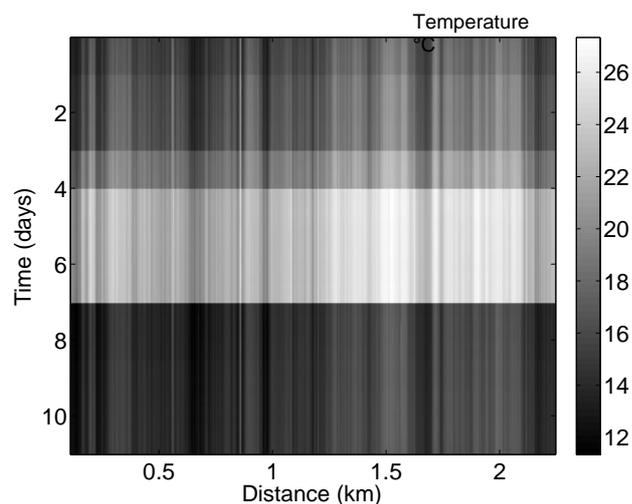


Fig. 2. Real DTS temperature data set for 11 nonconsecutive days, having no precipitation instances, during summer 2005. The gray scale map shows different observed temperatures with increasing temperatures from black to white. White rectangle represents the temperature signals acquired on the 1<sup>st</sup> day and the arrows correspond to two singular ( $x_2$  and  $x_3$ ) and two nonsingular ( $x_1$  and  $x_4$ ) zones referred to in Fig. 3.

similar meteorological trend refers to the days where there are no instances of significant precipitations or other phenomena which could greatly alter the daily temperature profile. If the meteorological conditions are also recorded along with the data, they can be used to extract from the data different time zones sharing the same meteorological conditions. Otherwise, in case of long period acquisitions or non-homogeneous meteorological conditions, it is possible to select similar zones by using a criterion based on the measurement of change in the recorded data statistics. Here, a criterion based on the higher order statistics of skewness and kurtosis, previously proposed by the authors [17] was used to identify such phenomena.

The methodology adopted to exploit this kind of data set to localize the singularities would be presented in the next section.

### III. SINGULARITY DETECTION METHODOLOGY

The acquired temperature data are a mixture of various factors amongst them different singularities (leakages, existing structures, etc). The current work focuses on the detection of these singularities, specially the leakages, and their localization in distance. A singularity detector is proposed here by exploiting the fact that the singularities respond differently to daily temperature variations than the homogeneous nonsingularity zones. The detector uses a dissimilarity measure to identify the singularities based on the deviation of their daily temperature profile from the nonsingularity zones.

Consider for example the temperature signals for day 1 from Fig. 2 (see white rectangle) at 4 distances marked by arrows. Fig. 3 (solid lines) shows these temperature variations over a 24-hour period at these 4 distances. Two of these distances ( $x_2$  and  $x_3$ ) correspond to the two drains,  $D1$  and  $D2$ , whereas the remaining two ( $x_1$  and  $x_4$ ) to the homogeneous nonsingularity zones (0.3 km and 1.264 km). It can be observed that the drains (which are assimilated with singularities) show a different trend of variation compared to the homogeneous nonsingularity zones. This 24-hour temperature variation thus contains useful information regarding the singularities. The idea for singularity detection is thus based on the estimation of a reference vector from the 24-hour temperature variations at all the sensing distances which could then be compared to vectors at all the distances. The resulting deviation from the reference would be smaller for nonsingular zones and larger for singular zones. Among the different approaches that could be used for estimation of the reference vector, the Singular Value Decomposition (SVD) offers a very efficient solution [18] as would be explained next.

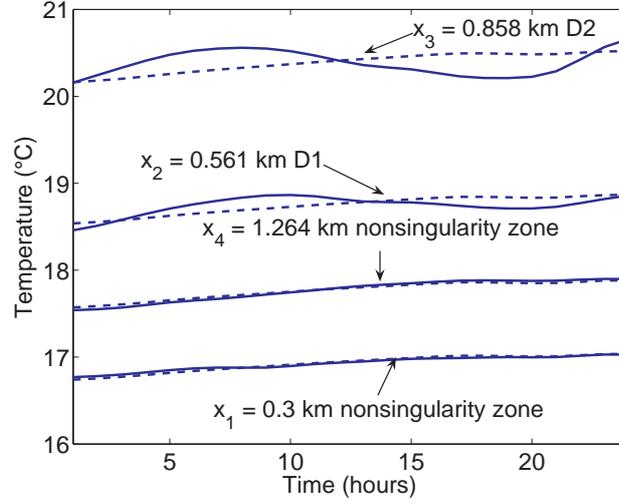


Fig. 3. (Solid Lines): The 24-hour temperature profiles on day 1 at 4 different distances of image in Fig. 2 (see white rectangle and arrows). The singular zones ( $D1$  and  $D2$ ) present a different trend as compared to nonsingular zones (0.3 km and 1.264 km). (Dashed Lines):  $\mathbf{Y}_{sig}^1$  at same day and distances as solid lines, showing that estimated reference vectors for singular zones have large deviations as compared to nonsingular zones.

### A. Singular Value Decomposition

The singular value decomposition (SVD) of a signal,  $\mathbf{Y}^i \in \mathbf{R}^{N_t \times N_x}$  is defined as [11], [12]:

$$\mathbf{Y}^i = \mathbf{U}_N^i \Delta_N^i \mathbf{V}_N^{iT} = \sum_{j=1}^N \sigma_j^i \mathbf{u}_j^i \mathbf{v}_j^{iT}, \quad (3)$$

where  $N = \min(N_t, N_x)$ ,  $\Delta_N^i \in \mathbf{R}^{N \times N}$  is a matrix containing on its diagonal the singular values,  $\sigma_j^i \geq 0$ , arranged in a descending order and  $\mathbf{U}_N^i \in \mathbf{R}^{N_t \times N}$  and  $\mathbf{V}_N^i \in \mathbf{R}^{N_x \times N}$  are orthogonal matrices, containing the left and right singular vectors,  $\mathbf{u}_j^i \in \mathbf{R}^{N_t}$  and  $\mathbf{v}_j^i \in \mathbf{R}^{N_x}$ , respectively. The left singular vectors  $\mathbf{u}_j^i$  are identified as estimators of reference vectors and are orthogonal to each other. These vectors are a function of time and may differ from one day to another. The singular values given by SVD are sorted in a descending order with their magnitude indicating the degree of coherence of estimated vectors from the most coherent to the least coherent. The right singular vectors,  $\mathbf{v}_j^i$ , represent the spatial variations of the estimated reference vectors. SVD allows to decompose the initial data space into complementary subspaces. More specifically, it can be used to achieve separation between signal and noise subspaces [12]:

$$\mathbf{Y}^i = \mathbf{Y}_{sig}^i + \mathbf{Y}_{residue}^i = \sum_{j=1}^P \sigma_j^i \mathbf{u}_j^i \mathbf{v}_j^{iT} + \sum_{j=P+1}^N \sigma_j^i \mathbf{u}_j^i \mathbf{v}_j^{iT} \quad (4)$$

where the signal subspace,  $\mathbf{Y}_{sig}^i$ , is formed with  $P$  most energetic singular values and the noise or residue subspace as we call it here,  $\mathbf{Y}_{residue}^i$ , with remaining  $N - P$  singular values. The singularities at the data site can be identified with their peculiar behavior of giving temperature variations over a period of a day that are different from those of non-singular zones. SVD was thus used as the first step to identify from amongst the 24-hour temperature profiles, at different observation distances, the one which is a representative of the nonsingularities. This reference vector can be estimated by observing the singular values obtained by application of SVD as in (3). Generally, the first singular value contains most of the signal energy and the first vector,  $\mathbf{u}_1^i$ , being the most energetic, is the most coherent component of the processed signal. It can therefore serve as a suitable estimate of the reference vector with the vector  $\mathbf{v}_1^i$  representing its spatial variation. The corresponding signal subspace for this reference vector,  $\mathbf{Y}_{sig}^i$ , can be constructed using  $P = 1$  in (4). The signal subspace,  $\mathbf{Y}_{sig}^i$ , represents an estimate of the ground response for each day [19]. It can be interpreted in our case as a subspace constructed by the reference vector,  $\mathbf{u}_1^i$ , extracted from the 24-hour temperature variations at all the distances.

### B. Dissimilarity Measure Using SVD

A quick analysis of the signal or so called reference subspace,  $\mathbf{Y}_{sig}^i$ , instantly reveals the underlying idea. The dotted lines in Fig. 3 show this reference subspace,  $\mathbf{Y}_{sig}^1$ , on day  $i = 1$  of the current data set at the 4 distances referred to earlier in the section. Two of these distances (drains,  $D1$  and  $D2$ ) correspond to singularities, whereas the other two to reference nonsingularity zones (0.3 km and 1.264 km). It can be observed that the estimated reference vectors (dashed lines) are close to the original temperature variations in the initial data (solid lines) only for the nonsingularity zones, whereas they show large deviations in case of singularity zones. This validates the fact that the reference vector  $\mathbf{u}_1^i$  is indeed a good representative of the nonsingularity zones. The residue subspace of the SVD,  $\mathbf{Y}_{residue}^i$ , contains the information linked to the deviation from the reference vector. The singularities can thus be identified by using a measure of dissimilarity between the estimated reference vector and the recorded data at all distances. One possible measure for this dissimilarity is the  $L_2$  norm,  $\|\cdot\|$ , for each column,  $\mathbf{y}_{residue}^i(x)$ , of the residue subspace,  $\mathbf{Y}_{residue}^i$ , as each column of this subspace represents the deviation from the reference vector. This can be formulated as follows:

$$d^i(x) = \|\mathbf{y}_{residue}^i(x)\|_2, x = 1, \dots, N_x; i = 1, \dots, N_i. \quad (5)$$

It should be mentioned that although euclidean distance was employed here, tests with other distances resulted in no significant difference in the dissimilarity.

Considering the ideal case as the one with identical nonsingularity zones and a few singularity zones, not necessarily identical, the dissimilarity measure obtained using the proposed method should only contain information related to the singularities. In other words, the histogram of the dissimilarity measure for any given day,  $d^i(x)$ , should be like the one in Fig. 4(a), indicating that the dissimilarity value is zero for most of the distances (no singularity). The dissimilarity measure should ideally be non zero only for distances corresponding to singularities being distributed uniformly with the range,  $u^i = ]0, \max(d^i(x)) ]$ . However, owing to the fact that different nonsingularity zones show a similar response that may not be identical and that the reference vector is only an estimation of the daily temperature variations in the nonsingular zones, the histogram would present a spread around the zero. Fig. 4(b) shows such a histogram for the dissimilarity measure,  $d^1(x)$ , obtained on day 1 of the data in Fig. 2. We propose an approach for thresholding the dissimilarity measure based on this histogram to complete the singularity detection.

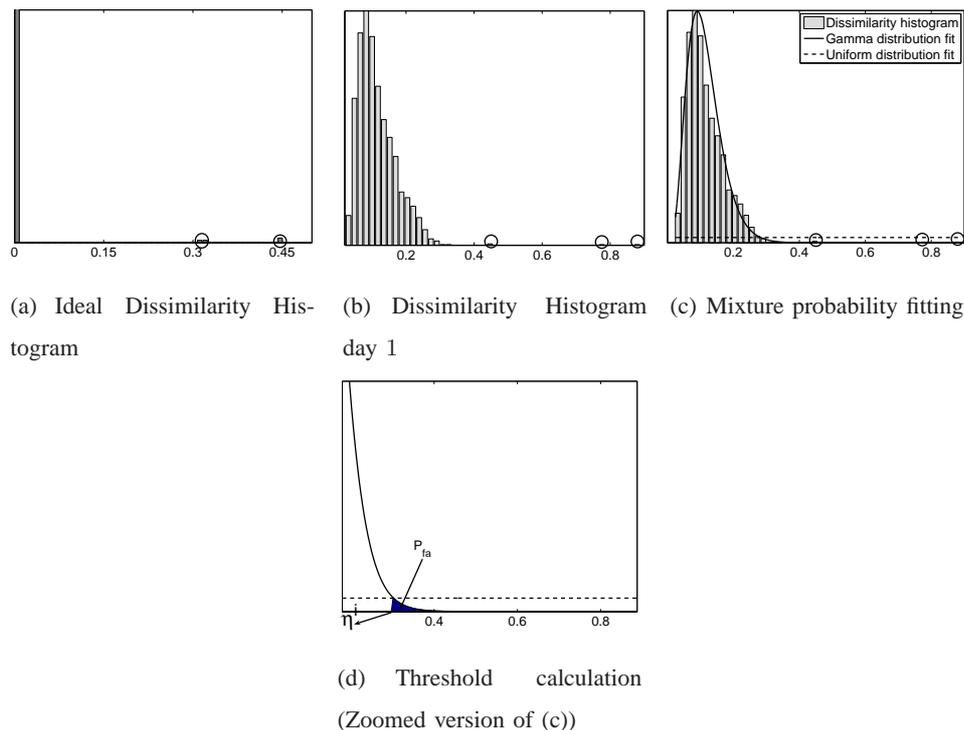


Fig. 4. Thresholding using constant probability of false alarm approach. Ideally the dissimilarity should be zero for most of the zones except singularities (a) but practically this is not the case (b). The resulting histogram for any given day could be modeled by a Gamma and Uniform distribution mixture with their parameters estimated using EM-algorithm (c). Fixing the probability of false alarm ( $P_{fa}$ , shaded region) allows to find a unique threshold,  $\eta^i$ , for each day thus accounting for effects of daily climatic changes (d). Notation “O” highlights the values of dissimilarity measures for singularities.

### C. Singularity Detection Using CFAR

Considering Fig. 4(b), the part of histogram corresponding to nonsingular distances can be modeled by a Gamma distribution. The choice of Gamma distribution is inspired by the fact that its two parameters,  $\alpha$  and  $\beta$  allow adaptability of the distribution shape to adjust to different situations. Moreover, Gamma distribution explicitly takes into account positivity of the modeled dissimilarity measure. The singular zones (marked by circles in Fig. 4) would not be captured by the Gamma model and instead a Uniform distribution is assumed for the singularities. The dissimilarity measure can thus be modeled with a mixture of Gamma and Uniform distributions whose parameters can be estimated using the maximum likelihood approach. In this regard, the Expectation-Maximization (EM) algorithm is an efficient tool that maximizes the likelihood with respect to the distribution parameters [20]. Tan et al. [21] recently proposed a formulation for parameter estimation for such a mixture based on the EM algorithm, which was employed for the present work. The log-likelihood function of the mixture for any given day “ $i$ ” can be written as:

$$\begin{aligned} \mathcal{L}(d^i | \alpha^i, \beta^i, u^i) = & \sum_{x=1}^{N_x} \gamma_x^i \left( -\alpha^i \log \beta^i + (\alpha^i - 1) \log d^i(x) \right. \\ & \left. - \log \Gamma(\alpha^i) - \frac{d^i(x)}{\beta^i} + \log \pi^i \right) \\ & + \sum_{x=1}^{N_x} (1 - \gamma_x^i) \log(1 - \pi^i), \end{aligned} \quad (6)$$

with  $\pi^i$  the prior probability of Gamma distribution in the mixture,  $\Gamma(\cdot)$  the Gamma function and  $\gamma_x^i$  the responsibility vector given by [21]:

$$\gamma_x^i = \frac{\pi^i f_{\Gamma}(d^i(x) | \alpha^i, \beta^i)}{\pi^i f_{\Gamma}(d^i(x) | \alpha^i, \beta^i) + (1 - \pi^i) (u^i)^{-1}}, \quad (7)$$

where  $f_{\Gamma}(\cdot)$  is the probability density function of Gamma distribution and  $u^i$  the range of Uniform distribution. The mixture parameters and the prior probabilities are estimated using a recursive procedure by maximizing Eq. (6) with respect to the parameters as presented in [21]. The histogram of Fig. 4(b) fitted with the estimated Gamma and Uniform distributions is shown in the Fig. 4(c). Having estimated the fitting distributions, the next step is to establish a threshold so as to detect only the singularities. At any given distance, the obtained dissimilarity measure may show variations over different days owing to several factors. These include the duration and intensity of the sunlight on a particular day, the air temperature, the wind speed, etc. The resolution of the acquisition material could also be one of the factors. Since the estimated reference temperature variations are not necessarily the same for each day,

preselecting a constant threshold may not be a good idea. A popular approach for variable thresholding applications is based on the Constant False Alarm Rate (CFAR) whereby a probability of false alarm ( $P_{fa}$ ) is fixed rather than a threshold. The probability of false alarm is usually system dependent and fixed in terms of the system deployment by experts. In our case, false alarm is raised when the detector detects a singularity while there actually was none. Thus assuming the desired  $P_{fa}$  as known, the threshold,  $\eta^i$ , can be computed for each day “ $i$ ” as shown in Fig. 4(d), using the expression:

$$P_{fa} = \int_{\eta^i}^{\infty} \frac{z^{(\alpha^i-1)} e^{-z/\beta^i}}{\beta^i \alpha^i \Gamma(\alpha^i)} dz \quad (8)$$

The detector thus takes the decision based on this threshold as:

$$d_{th}^i(x) = \begin{cases} d^i(x) & \text{if } d^i(x) > \eta^i \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

#### D. Automated Singularity Detection System

The proposed detector gives a map of the singularities in two dimensional spatio-temporal space  $(x, i)$ , with a spatial resolution of 1 m and a temporal resolution of 24 hours. The system can be implemented as shown in Fig. 5. In the next section, the proposed system is analyzed on real temperature data under different scenarios to show its efficiency.

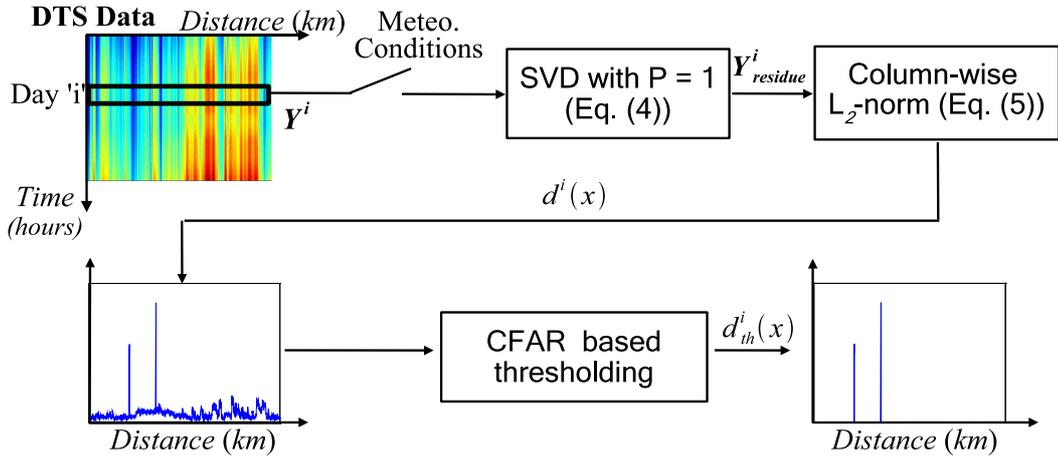


Fig. 5. SVD and CFAR based singularity detection system. The residue subspace of SVD on 24-hour data contains information on the singularities. By taking  $L_2$  norm at each distance, the resulting dissimilarity is modeled as a mixture of Gamma and Uniform distributions and thresholded by fixing a probability of false alarm.

TABLE I  
CHARACTERIZATION OF THE STRUCTURES AND INTRODUCTION OF ARTIFICIAL LEAKAGES AT THE TEST SITE

	Drains		Leakages			HotPoint
	D1	D2	L1	L2	L3	HP
Location (km)	0.561	0.858	1.566	1.551	1.573	0.674
Time (day)	-	-	8 (noon)	10 (noon)	10 (eve)	8 (noon)
Flow rate (l/min)	-	-	5	1	1	-

#### IV. APPLICATION ON DTS TEMPERATURE DATA SET

The real temperature data presented in Fig. 2 are first analyzed. These data were recorded over a period of 11 non-consecutive days in year 2005. With reference to the days shown in Fig. 2, the different phenomena present at the site are tabulated in Tab. I with regards to their position and occurrence time. Three artificial leakages of percolation-type,  $L1$ ,  $L2$ , and  $L3$ , with different flow rates, were introduced along with a hot point ( $HP$ ) simulating artificial flow of hot water on days 8 and 10. The data over the days with neither any artificial nor any real leakages serve as a reference for different scenarios. Moreover, similar meteorological conditions (e.g., no precipitation instances) were ensured here using the higher order statistics based criteria [17]. It should be highlighted that if days with intense precipitation are considered for analysis, the detector would give non-zero outputs for most of the distances along the fiber as opposed to only a few for singularities in the absence of precipitation. The first singular value,  $\sigma_1^i$ , obtained by application of SVD on daily data  $\mathbf{Y}^i$ , is extremely energetic amounting to about 99% of the total energy. Therefore, the first SVD source vector,  $\mathbf{u}_1^i$ , being a very good representative of the “ $i^{th}$ ” 24-hour temperature response of the site under study, was selected as the reference vector. The reference subspace,  $\mathbf{Y}_{sig}^i$ , is thus constructed using the spatial variation of this source with  $P = 1$  while the residue subspace,  $\mathbf{Y}_{residue}^i$ , using the remaining singular values. The residue subspace, representing deviation from the reference subspace, contains useful information from the singularity detection perspective. The dissimilarity measure,  $d^i(x)$ , was obtained by taking the  $L_2$ -norm of (see Eq. (5)) at all the distances. Fig. 6 shows the resulting measure for day 1, whose histogram was presented in Fig. 4(b)-4(c). It can be observed that drains,  $D1$  and  $D2$ , offer a strong dissimilarity depicting a behavior different from other zones. The rest of the dissimilarity measure could be thought of as background noise.

The dissimilarity measure was thresholded using the CFAR based thresholding scheme described in

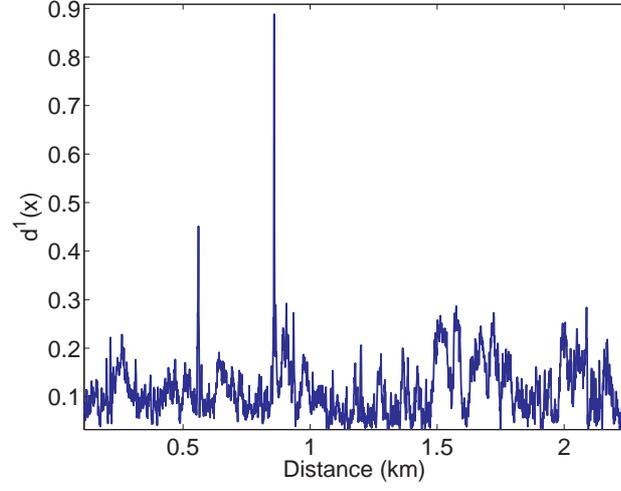


Fig. 6. The residue subspace,  $\mathbf{Y}_{residue}^i$ , obtained with SVD for day 1 of nonconsecutive 11 days data set. This subspace represents the deviation from the reference subspace,  $\mathbf{Y}_{sig}^i$ , thus putting into evidence the singularities.

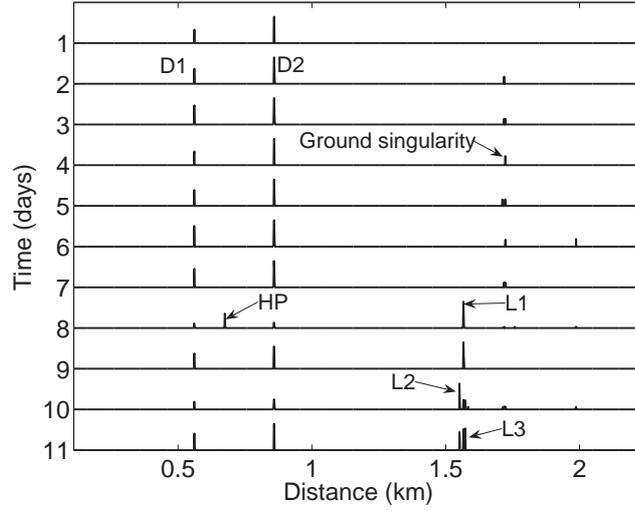


Fig. 7. Detection results for data set shown in Fig. 3. The amplitudes shown are normalized for each day with respect to the maximum for that day. The drains,  $D1$  and  $D2$  constitute omnipresent singularities and thus always detected. A singularity of ground is observed around 1.7 km on several days. Artificial percolation type leakages ( $L1$ ,  $L2$ ,  $L3$ ) are also detected from days 8 to 11 with their strength relative to the drains indicating their intensity. The hot point (HP) is also detected on day 8.

Sec. III-C. The  $P_{fa}$  being an input parameter to the current system, we are conducting experiments in consultation with the experts to associate physical justification to  $P_{fa}$  selection. For this study, we consider a  $P_{fa}$  of  $10^{-4}$ . The detection results,  $d_{th}^i(x)$  are shown in Fig. 7 for the data set under study with

distance on abscissa and days on ordinate. The amplitudes for each day were normalized with respect to the maximum for each day, thus peaks amplitudes would give us an information about the strength of the underlying singularities on any given day. The drains,  $D1$  and  $D2$  constitute omnipresent singularities hence detected on all the days. Moreover, a singularity is detected around 1.7 km as evident on days 4 to 10. It corresponds to a singularity of the ground in a zone where the material composition is different from other zones. The different material composition results in different conduction properties owing to differences such as soil permeability, composition, etc.

At the present site, we have acquisitions over different periods from year 2005 to 2007. The system repeatability was thus verified by testing it on the data acquired in 2006 with another DTS device with different sensitivity over the days with same acquisition conditions (meteorological). The results obtained were concurrent with the ones obtained in 2005 with the two drains and the ground singularity detected as the singularities. Moreover, the results do not depend on the season when acquisitions are done. The results are not shown in order to avoid repetition.

#### A. Artificial Leakage Detection

We next focus our attention to the three artificial leakages ( $L1$ ,  $L2$ ,  $L3$ ) and the hot point ( $HP$ ). These leakages are impulsional in distance and exist momentarily in time as well. The resultant flow of water due to these leakages would bring about a change in the ground temperature thus presenting a source of singularity. The detection results for the case of artificial leakages is shown in Fig. 7 from days 8 to 11 with the corresponding zoomed version shown in Fig. 8. It can be observed that a very good localization in distance of the leakages is achieved with a very good precision (see Fig. 8). Moreover, the leakages are

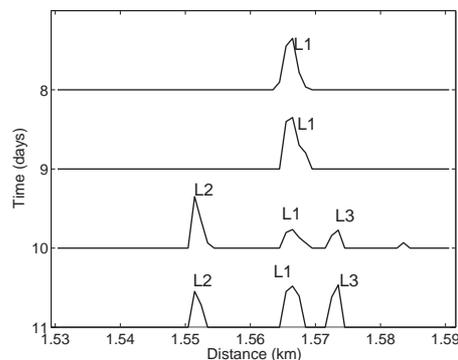


Fig. 8. Zoom on detection results for artificial leakages depicting a good localization of the leakages in distance.

identified at their original locations along with the hot point.  $L1$  is detected as the most energetic leakage

on day 8 in accordance with its highest flow rate ( $5l/min$ ) and it persists for 3 days with decreasing strength. The other two leakages are also detected albeit with lower amplitudes. Since the system was designed to work as a daily detector and normalization was carried out for each day, the comparison of the leakages strengths could be done by their amplitudes relative to the drain amplitudes for each day. The leakages sustained only over 2 – 3 days as was expected due to their controlled impulsive nature.

### B. Periodic Analysis and Real Leakage Detection

As an example of periodic analysis, the proposed system was tested on the data acquired in 2007. The detection results for 4 days in 2007 when there were no precipitations (like the case of 2005) are shown in Fig. 9. A comparison with the results obtained for year 2005 (see Fig. 7) show that while the two drains,  $D1$  and  $D2$  and the ground singularity are identified as before, an additional singularity zone is observed just after the second drain at a distance around 0.91 km. This points to an occurrence of a new event in 2007 which was not there in the first place in 2005 or 2006.

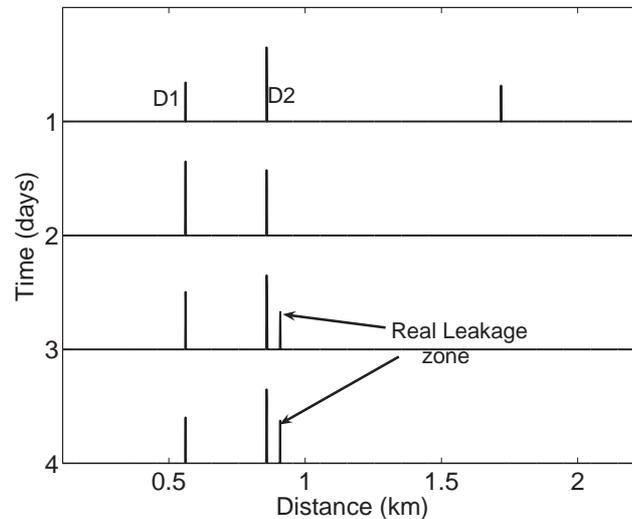


Fig. 9. Detection results for 4 days in 2007 reveal real leakage zone just after drain  $D2$  in addition to the previously observed drains and the singularity of the ground. Note that there are days when  $D1$ ,  $D2$  do not appear which could be explained by the possible climatic conditions.

The investigation into this singularity by physical inspection at the acquisition site revealed existence of a suspected real water leakage in the dike structure just after the drain  $D2$ . This real leakage has a non impulsive signature as opposed to the artificial percolation-type leakages. Its detection as a singularity is

justified by the fact that it brings additional water around the optical fiber thus providing more conduction and different temperature variation than the homogeneous nonsingular zones. Further investigations are underway to quantify this leakage by measuring its flow rate, determining its seasonal dependence, etc.

## V. CONCLUSION

The use of distributed temperature sensors for anomaly detection in dikes is a very efficient solution due to their long-term and automatic monitoring capabilities with the added advantage of using low-cost and robust optical fiber cable. The acquired temperature data using DTS is not immediately exploitable in terms of anomaly detection and signal processing is inevitable to extract this information. A detector was proposed in this paper based on SVD and a fixed probability of false alarm. It was shown that the temperature variations over the course of a day do not show same variations at all the sensing distances and these daily temperature variations can be exploited to characterize a given distance as a singularity or nonsingularity. A reference vector for the temperature variation over a 24-hour period can be estimated using SVD as it is linked to the first singular value of SVD. This reference vector being a representative of the nonsingular zones, the residue subspace contains the deviation from this reference and thus serves as an indicator of singularity. A dissimilarity measure is constructed employing the  $L_2$  norm on this residue subspace. The resulting measure is then modeled by a mixture of Gamma and Uniform distributions with the Gamma distribution accounting for the small deviations from the nonsingularity due to estimation and the Uniform accounting for the anomalies or the singularities. The distribution parameters are estimated using an EM-algorithm and then a thresholding scheme is designed based on the constant probability of false alarm. Fixing the probability of false alarm allows to estimate the threshold for each day which may vary from one day to another due to factors such as duration and intensity of sunlight, air temperature, wind speed, etc. The application of the proposed system on real data showed that drains present a significant singularity and since they form part of permanent structures, they are always detected. The proposed system was validated under different scenarios and the repeatability of the system was also verified by periodic analysis. Amongst the scenarios discussed were that of the detection of percolation type artificial leakages with their detection strength depending on their flow rate. The tests on year 2007 allowed to detect the presence of a real water leakage at the site which was previously unobserved during manual inspections. The future work would focus on the testing of this system on other data sets acquired at different sites as well as the development of other measures to characterize not only a dissimilarity, but also to quantify the shapes and variability of singularity zones.

## REFERENCES

- [1] D. Sampson, S. Collins, K. Oh, and R. Yamauchi, Eds., *Proc. 19th Intl. Conf. OFS*, vol. 7004. SPIE, 2008.
- [2] A. H. Hartog, "Distributed fiber-optic temperature sensors: principles and applications," in *Optical Fiber Sensor Technology*, K. T. Grattan and B. T. Meggitt, Eds. Kluwer, 2000, pp. 241–301.
- [3] S. A. Wade, K. T. Grattan, and B. McKinley, "Incorporation of fiber-optic sensors in concrete specimens: Testing and evaluation," *IEEE Sensors Journal*, vol. 4, no. 1, pp. 127–134, Feb. 2004.
- [4] 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*, 19-21, June 2001, pp. 9–17.
- [5] E. Lewis, C. Sheridan, M. O'Farrell, D. King, C. Flanagan, W. B. Lyons, and C. Fitzpatrick, "Principal component analysis and artificial neural networks based approach to analysing optical fiber sensors signals," *Sensors and Actuators A*, vol. 136, pp. 28–38, 2007.
- [6] 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, 2001*, D. I. E. Udd, Ed., vol. 4328, pp. 23–34.
- [7] S. Johansson and P. Sjö Dahl, "Downstream seepage detection using temperature measurements and visual inspection monitoring experiences from Røsvatn field test dam and large embankment dams in Sweden," in *Proc. Intl. Seminar on Stability and Breaching of Embankment Dams*, Oslo, Norway, Oct. 2004, p. 21.
- [8] A. Rozycki, J. M. Ruiz, and A. Caudra, "Detection and evaluation of horizontal fractures in earth dams using the self potential method," *Engineering Geology*, vol. 82, no. 3, pp. 145–153, 2005.
- [9] P. Sjö Dahl, T. Dahlinb, S. Johansson, and M. H. Lokec, "Resistivity monitoring for leakage and internal erosion detection at hllby embankment dam," *Journal of Applied Geophysics*, vol. 65, pp. 155–164, 2008.
- [10] V. C. Klema and A. J. Laub, "The singular value decomposition: its computation and some applications," *IEEE. Trans on Auto. Control*, vol. 25, no. 2, pp. 164–176, 1980.
- [11] L. L. Scharf, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1991.
- [12] V. Vrabie, J. I. Mars, and J.-L. . Lacoume, "Singular value decomposition by means of independent component analysis," *Signal Processing*, vol. 84, no. 3, pp. 645–652, 2004.
- [13] S. Yin, "Distributed fiber optic sensors," in *Fiber Optic Sensors*, F. T. Yu and S. Yin, Eds. Marcel Dekker Inc., 2000, pp. 201–229.
- [14] A. Rogers, "Distributed optical fiber sensing," in *Handbook of Optical Fiber Sensing Technology*, J. M. Lopez-Higuera, Ed. Wiley, 2002, pp. 271–308.
- [15] A. H. Hartog, "A distributed temperature sensor based on liquid-core optical fibers," *IEEE Journal of Lightwave Tech.*, vol. 1, no. 3, pp. 498–509, Sept. 1983.
- [16] J. P. Dakin, D. Pratt, and G. W. Bibby, "Distributed optical fiber Raman temperature sensor using a semiconductor light source and detector," *Electron. Lett.*, vol. 21, pp. 569–570, 1985.
- [17] A. A. Khan, V. Vrabie, J. I. Mars, and A. Girard, "A least square approach for bidimensional source separation using higher order statistics criteria," in *16th European Signal Processing Conference*, Lausanne, Switzerland, aug 2008.
- [18] A. A. Khan, V. Vrabie, G. d'Urso, and J. I. Mars, "SVD based automated dike monitoring system using DTS data," in *Proc. IEEE IECON'08*, Orlando, USA, 2008.

- [19] A. A. Khan, V. Vrabie, J. Mars, A. Girard, and G. d'Urso, "A source separation technique for processing of thermometric data from fiber-optic DTS measurements for water leakage identification in dikes," *IEEE Sensors Journal*, vol. 8, pp. 1118–1129, 2008.
- [20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.
- [21] C. O. Tan, J. A. Taylor, A. S. H. Ler, and M. A. Cohen, "Detection of multifiber neuronal firings: A mixture separation model applied to sympathetic recordings," *IEEE Trans. Biomed. Eng.*, In Press.