SVD based automated dike monitoring system using DTS data
Amir Ali Khan, Valeriu Vrabie, Guy D’Urso, Jerome Mars

To cite this version:

HAL Id: hal-00326391
https://hal.archives-ouvertes.fr/hal-00326391
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.
SVD based automated dike monitoring system using DTS data

Amir A. Khan†, Valeriu Vrabie‡, Guy D’Urso‡ and Jérôme I. Mars∗
∗GIPSA-Lab—Department Image and Signal Processing (DIS)
Grenoble INP, BP 46, 38402 Saint Martin d’Hères, France
†Centre de Recherche en STIC (CReSTIC), University of Reims, BP 1039, 51687 Reims, France.
‡Direction R&D, Electricité de France (EDF), 6 quai Watier, 78401 Chatou, France
Email: amir.ali.khan@gipsa-lab.inpg.fr, Telephone: (33) 4.76.826494, Fax: (33) 4.76.82.63.84

Abstract—The detection of water leakages in dikes using distributed temperature sensors is an interesting prospect due to the commercial viability of these optical fiber based sensors. The acquired temperature data, being not directly interpretable, requires intervention of advanced signal processing techniques. In this work, we propose a system for the identification of singularities such as existing dike structures and water leakages. The distances where singularities exist show temperature variations over the course of a day which are different from the nonsingular zones. The different nonsingular zones though show a similar temperature variation trend. The proposed system estimates this reference trend as the most coherent component of the Singular Value Decomposition applied on daily data. The corresponding SVD residue subspace thus represents the deviation from the reference subspace and thus contains information on singularities. The $L_2$ norm of this residue is a good discrimination measure for identification of these singularities.

I. INTRODUCTION

Fiber optic sensors are employed 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. Now-a-days, an important issue in engineering domain is the detection of anomalies, such as leakages, (significant flow of water), in the dikes to avoid disaster at mass level. In this regard, thermometric based method employing optical fiber Distributed Temperature Sensors (DTS) provides a very efficient method. The major advantage of DTS is their commercial viability (low-cost telecommunications grade fiber), ability to multiplex large number of sensors along a single fiber and environmental robustness [1], [2]. In addition, DTS have the capability to provide long-gage temperature measurements along the optical fiber length thus providing a monitoring range that may go up to 30 km. The spatial and temperature resolutions depend on the device used (e.g., a typical scale of measurement is 1 m spatial resolution for a temperature resolution of 0.01°C achievable with Sensornet’s Sentinel). The leakage detection using DTS signals is a new research problem in the signal processing domain. The aim of our research is the development of an automated system based on the processing of DTS data capable of identifying the singularities such as water leakages in dike structures.

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 of temperature can equally be brought about by other factors such as precipitation, seasonal effects, day/night, the existing structures (e.g. drains), etc. Moreover, since the fiber optic cable is buried in ground, the temperature signals acquired by DTS are strongly influenced by the response of the near surface (ground) where the acquisitions are made. The acquired data is thus not interpretable in its present form for the leakage detection purpose and necessitates some processing to render it useful.

In this paper, we present a new system for the identification of singularities in dike structures with the eventual goal of proposing an automated functional device. The singularities include the existing dike structures, the singularities in the ground where the DTS are buried and most importantly the water leakages that may occur in dikes. The temperature variations over the course of a day at different observation distances present a common trend except for the distances having some singularities. Thus considering the temperature signal for all the distances over a 24-hour period, we propose a dissimilarity method for singularity identification based on the classical data decomposition technique of Singular Value Decomposition (SVD) [3], [4], [5]. Since temperature monitoring is a continuous process, we can take the mean of dissimilarity measures obtained over several days to remove the random effects that may occur at any particular day. We validate the proposed system on real temperature data sets under different scenarios. Particularly interesting amongst these scenarios are those of artificial leakages and the real water leakages. We present a brief introduction of the acquisition principle followed by the theory of singular value decomposition and focus on its efficient utilization for devising our system. We conclude with a discussion on various scenarios to give an insight into the practicability of the proposed system.

II. ACQUISITION PRINCIPLE AND DATA DESCRIPTION

The Distributed Temperature Sensors (DTS), based on optical fibers, were successfully used for the acquisition of temperature data [6], [7], [8]. These sensors provide temperature observations over long distances with high spatial and temperature resolutions. Moreover, their capability to integrate a large number of passive optic sensors within a
single low-cost telecommunications grade optical fiber cable significantly enhance their commercial viability. The cable does not have any moving parts and is immune to mechanical vibrations and EMC interference thus rendering it ideal for industrial sensing applications. In most commercial distributed temperature sensors, the acquisition principle of temperature profiles is based on Raman scattering using Optical Time Domain Reflectometry (OTDR) techniques [9]. The basic setup of temperature sensing based on OTDR comprises of a pulsed laser coupled to the optical fiber, the sensing element. The emitted photons interact with the molecules of the fiber material. Thermally influenced molecular vibrations in the fiber produce Raman scattering in the form of backscattering of some photons of light and can thus be used to obtain the temperature distribution along the fiber [10]. The Raman backscattered light has two components: the Stokes and Anti-Stokes scattering. The principle of temperature sensing lies in the fact that the intensity ratio between Anti-Stokes and Stokes components is temperature dependent. Measuring the travel time of probe pulse and the intensity ratio, at the fiber input, gives the temperature profile along the entire length of the fiber using one-to-one relationship between spatial resolution and traveling time.

A thermometric data monitoring setup has been installed by EDF at an experimental test site to study the leakages. The aim of this site is to extract the information pertaining to leakages (both natural and controlled) in the dike of canal. A schematic representation of this installation is given in Fig. 1. A fiber optic cable was installed in the abutment at the toe end of the canal so as to intercept the water leakages from the canal. The 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. Two distinct elevation levels (Zone1, from approximately 0.1 km to 1.25 km and Zone2, from approximately 1.25 km to 2.2 km) will be exposed with varying intensities to direct sunlight. The cable also circumvents two drains, D1 and D2, situated at 0.561 km and 0.859 km, respectively.

The temperature resolution of this device is 0.01°C with 1-meter spatial resolution which would allow us to detect very closely the occurrence of water leakages.

The temperature data is acquired with a sampling interval of 1 m along the entire cable length. Continuous temperature monitoring is important in order to observe the temporal evolution of the leakages. We have at our disposition several acquisitions from years 2005 to 2007 which would allow us to analyze different scenarios. These acquisitions were made with a sampling interval of 1-hr which gives us 24 acquisitions per day. The recorded data set for each day can be written as:

\[ Y^i = \{ y^i(t, x) \mid 1 \leq t \leq N_t, 1 \leq x \leq N_x \}, i = 1, ..., N_t \]  

where \( N_t = 24 \) represents the number of acquisitions per day and \( N_x \) represents the number of observation points. A sample real data set over a period of nonconsecutive 30 days in year 2005 with same meteorological conditions is shown in Fig. 2. The different colors represent the scale of the recorded temperature with dark blue showing lowest temperature and dark red the highest. The same meteorological conditions here refer to the days where there was no precipitation. Note here that this selection was made by consulting the meteorological data. The same can however be made using a criteria based on higher order statistics developed by authors but that does not fit into the scope of this paper and would be presented elsewhere.

We will see in the next section how we can exploit this data set to extract useful information linked to the singularities.

For all the 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. In our setup, the temperature data was recorded by a commercial device Sensornet, Sentinel DTS-MR, with a capability of covering up to 8 km range.

The aim of the proposed system is to give a measure for characterization of singularities (the leakages, existing structures, etc.) in distance. The motivation behind this idea is:
that the singularities respond differently to daily temperature variations than the homogeneous nonsingularity zones. As an example, consider the temperature signals for day 1 from Fig. 2 (see white rectangle) at 5 distances marked by arrows. Fig. 3 (solid lines) shows these temperature variations over a 24-hour period at these 5 distances. Amongst them, two distances are at the two drains, \( D_1 \) and \( D_2 \), one in a region containing a singularity of the ground (1.7 km) in which fiber is buried, whereas the remaining two are in homogeneous nonsingularity zones (0.3 km and 1.264 km). The ground singularity (1.0-1.7 km) corresponds to the zone where material composition of ground is different from other zones. It can be observed that the drains and the singularity show a different trend of variation as compared to the homogeneous nonsingularity zone. It is based on this 24-hour temperature variation that we will be devising our measure. The idea would thus be to find a suitable reference vector from the 24-hour temperature variations at all the sensing distances and then to compare this reference vector with vectors at all the distances.

**Fig. 3.** (Solid Lines): The 24-hour temperature profiles on day 1 at 5 different distances of image in Fig. 2 (see white rectangle and arrows). The singular zones (\( D_1, D_2 \) and 1.7 km) present a different trend as compared to nonsingular zones (0.3 km and 1.264 km). (Dashed Lines): \( Y_{\text{sig}}^i \) 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 is defined as [4], [5]:

\[
Y^i = U_N^i \Delta_N^i V_N^i, \quad \Delta_N = \sum_{j=1}^{N} \beta_j^i u_j^i v_j^i, \quad (2)
\]

where \( N = \min(N_t, N_x) \), \( \Delta_N \in \mathbb{R}^{N \times N} \) is a matrix containing on its diagonal the singular values \( \beta_j^i \geq 0 \) arranged in a descending order and \( U_N^i \in \mathbb{R}^{N_t \times N} \) and \( V_N^i \in \mathbb{R}^{N_x \times N} \) are orthogonal matrices, containing the left and right singular vectors, \( u_j^i \in \mathbb{R}^{N_t} \) and \( v_j^i \in \mathbb{R}^{N_x} \), respectively. The left singular vectors \( u_j^i \) are identified as estimators of reference vectors and are orthogonal to each other. In the present case, these vectors are a function of time and differ from one day to another. The singular values of SVD are sorted in a descending order giving the degree of coherency of estimated vectors from the most coherent to the least coherent. The right singular vectors, \( v_j^i \), 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 [5]:

\[
Y^i = Y_{\text{sig}}^i + Y_{\text{residue}}^i = \sum_{j=1}^{P} \beta_j^i u_j^i v_j^i + \sum_{j=P+1}^{N} \beta_j^i u_j^i v_j^i \quad (3)
\]

B. **Singularity Detection**

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 nonsingular zones. SVD can be used in 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 obtained 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, \( u_1^i \), being the most energetic, is the most coherent component of the processed signal and can therefore be taken as the reference vector. The vector \( v_1^i \) represents the spatial variation of the estimated reference vector. The corresponding subspace for this reference vector, \( Y_{\text{sig}}^i \), can be constructed using \( P = 1 \) in (3). The signal subspace, \( Y_{\text{sig}}^i \), represents an estimate of the ground response for each day [11]. It can be interpreted in our case as a subspace constructed by the reference vector, \( u_1^i \), extracted from the 24-hour temperature variations at all the distances.

The singularities can be identified by using a measure of dissimilarity between this reference vector and the recorded data at all distances. This information linked to the dissimilarity can be found in the residue subspace of the SVD, \( Y_{\text{residue}}^i \). One of the possible measures for this dissimilarity is the \( L_2 \) norm for each column, \( \| Y_{\text{residue}}(x) \|_2 \), of the residue subspace, \( Y_{\text{residue}}^i \), as each column of this subspace represents the deviation from the reference vector. This can be formulated as follows:

\[
d(x, i) = \| Y_{\text{residue}}(x) \|_2, \quad x = 1, \ldots, N_x; \quad i = 1, \ldots, N_i \quad (4)
\]

Thus we will have a dissimilarity matrix, \( D = \{ d(x, i) \mid 1 \leq x \leq N_x, 1 \leq i \leq N_i \} \). Repeating the procedure over \( N_i \) days, we can take the mean of \( D \) over \( N_i \) to get an average dissimilarity vector, \( d(x) \). This averaging would allow to reduce the effects of random external factors that may change from one day to another. The above described system can be implemented with the help of schematic as shown in Fig. 4. In the next section, we will present different scenarios on real temperature data set in order to show the efficacy of the proposed system under different conditions.
The reference subspace, $Y$, for each day is constructed using the remaining singular values. This subspace contains the useful information from the anomaly detection perspective as it represents the deviation from the monitoring days is retained. Analyzing the two elevation zones (see Fig. 1), we first focus our attention to Zone 1 and Zone 2. Otherwise, this zone does not contain any useful information from the anomaly detection point of view. Considering the Zone 2, it should be mentioned that this zone is more non-homogeneous in terms of material composition than Zone 1. This zone is typified by the presence of a singularity around 1.6 – 1.7 km which is identified by two peaks at its start and end in the mean vector.

A. Case of Continuous Analysis

The temperature monitoring at the dike using the DTS can be employed as a continuous process as, once installed, the acquisition device does not require any human intervention. At the present site, the data was acquired continuously over the period of 1 year from spring 2005 to spring 2006 and intermittently over certain other periods till 2007. So, we use the continuously acquired data to validate the proposed system on two scenarios to show the repeatability and the leakage detection for artificially introduced leakages.

1) Repeatability of Measure: The proposed system was tested on another data set under the same meteorological conditions as those of the previous case and with no occurrence of artificial or real leakages but in a different monitoring interval in spring, 2006. The results obtained for the average dissimilarity matrix on $N_i = 10$ days are shown in Fig. 7 and

in the top window of Fig. 6. For a given distance, the variations corresponding to different days result from several factors which may vary over days. These may 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. But despite these minor variations, there are certain distances where the behavior is different from others. This is the case with the drains and the singularity of the ground surface.

In order to remove the influence of the random variations that may occur over the studied period, we calculate the mean of $d(x, i)$ over a monitored time zone. For example, in the present case, it is calculated for $N_i = 30$ days. The resultant average dissimilarity measure is given in the bottom window of Fig. 6 where only the information that is coherent between the monitoring days is retained. Analyzing the two elevation zones (see Fig. 1), we first focus our attention to Zone 1 (from 0.2 – 1.25 km). This mean vector mainly identifies the two drains for Zone 1. Otherwise, this zone does not contain any useful information from the anomaly detection point of view. Considering the Zone 2, it should be mentioned that this zone is more non-homogeneous in terms of material composition than Zone 1. This zone is typified by the presence of a singularity around 1.6 – 1.7 km which is identified by two peaks at its start and end in the mean vector.

The real temperature data in Fig. 2 is initially used for the analysis. The data is over a period where there are neither any artificially introduced leakages nor any real leakages as confirmed by the experts. This can thus serve as a reference data set for other scenarios. Moreover, the monitoring interval was chosen here by consulting the meteorological data with similar meteorological conditions (e.g., no precipitation instances). The application of SVD on each day data, $Y^i$, reveals an extremely energetic first singular value, $\beta_1^i$, of about 99%. The first SVD source vector, $u_1^i$, is thus a very good representative of each 24-hour temperature response of the site under study and can thus be selected as the reference vector. The reference subspace, $Y^{sig}_i$, is then constructed using the spatial variation of this source, with $P = 1$. A quick analysis of this reference subspace instantly reveals the underlying idea. The dotted lines in Fig. 3 show this reference subspace on day 1 of the current data set at the 5 distances previously referred to in Sec. III. Three of these distances (drains, $D_1$, $D_2$, and at 1.7 km) correspond to the singularities, whereas the other two to reference nonsingularity zones (0.3 km and 1.264 km). It can be observed from this figure 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 is indeed a good representative of the nonsingularity zones. The residue subspace, $Y^{residue}_i$, is constructed for each day using the remaining singular values. This subspace contains the useful information from the singularity detection perspective as it represents the deviation from the reference subspace. The residue subspace for day 1 is shown in Fig. 5 and it can be observed that it is predominated by the drains, $D_1$ and $D_2$.

Calculating the $L_2$-norm of (4), we obtain the dissimilarity measure, $d(x, i)$, for each day at all the distances as shown in the top window of Fig. 6. For a given distance, the variations corresponding to different days result from several factors which may vary over days. These may 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. But despite these minor variations, there are certain distances where the behavior is different from others. This is the case with the drains and the singularity of the ground surface.

In order to remove the influence of the random variations that may occur over the studied period, we calculate the mean of $d(x, i)$ over a monitored time zone. For example, in the present case, it is calculated for $N_i = 30$ days. The resultant average dissimilarity measure is given in the bottom window of Fig. 6 where only the information that is coherent between the monitoring days is retained. Analyzing the two elevation zones (see Fig. 1), we first focus our attention to Zone 1 (from 0.2 – 1.25 km). This mean vector mainly identifies the two drains for Zone 1. Otherwise, this zone does not contain any useful information from the anomaly detection point of view. Considering the Zone 2, it should be mentioned that this zone is more non-homogeneous in terms of material composition than Zone 1. This zone is typified by the presence of a singularity around 1.6 – 1.7 km which is identified by two peaks at its start and end in the mean vector.
a comparison with the results obtained previously for summer, 2005 (Fig. 6) validates the repeatability of the proposed system with drains and the singularity of the ground being once again identified as the major singularities.

2) Artificial Water Leakages (Percolation): At the monitoring site, three artificial leakages of percolation-type, \( L_1 \), \( L_2 \) and \( L_3 \), were introduced with different flow rates of 5, 1 and 1 lit/min and at different positions, 1.562, 1.547 and 1.569 km, respectively in the month of May, 2005. A hot point \((HP)\) was also introduced at 0.674 km. These leakages are impulsive in distance and exist momentarily in time as well. The leakages result in flow of water which would bring about a change in the ground temperature and thus would present a source of singularity. The proposed system calculates here the average dissimilarity vector over \( N_i = 10 \) days and when we would have the start of water leakages (percolations) in the analyzed time period, we will have the results as in Fig. 8. It can be observed that as for the reference case with no leakages, the drains and the singularity zone are identified. The difference in this case is the detection of the three artificially introduced leakages and the hot point, \( HP \). From the zoomed version in the vicinity of the leakages in bottom window of Fig. 8, it can be observed that the 3 leakages are uniquely identified with \( L_1 \) offering the most dissimilarity due to its highest flow rate. The other two lower flow rate leakages are also detected.

Fig. 6. (Top): The dissimilarity matrix, \( D \), obtained for data of Fig. 2 over a period of 30 days, in summer of 2005, shows variations at a given distance due to external and acquisition conditions. (Bottom): The average dissimilarity vector calculated as a mean of \( D \) over 30 days allows to bring into evidence only the coherent information over the monitoring period. In the present case, the drains, \( D_1 \) and \( D_2 \), along with the singularity of the ground surface are detected as singularities.

Fig. 7. The average dissimilarity vector for spring 2006 validates the repeatability of the proposed system by detection of the drains and singularity of the ground surface as major singularities.

B. Case of Discontinuous Analysis (Real Leakage Detection)

As a case of discontinuous analysis, we tested the proposed system on the data acquired in 2007. The resulting mean dissimilarity vector over \( N_i = 30 \) days for this case is given in Fig. 9. A comparison of this result with the one obtained in 2005 (see Fig. 6) makes an interesting reading. While the two drains, \( D_1 \) and \( D_2 \) are identified along with the singularity zone between 1.6-1.7 km, there is one marked difference. A singularity is observed just after the second drain at a distance around 0.91 km in the present case which was not there in 2005. The investigation into this singularity by physical inspection at the acquisition site revealed existence of a real water leakage in the dike structure just after the drain \( D_2 \). This real leakage has a nonimpulsive signature as opposed to the artificial leakages of the previous subsection. This leakage exists over the distance between 0.907 km and 0.916 km. 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.

V. DISCUSSION

The test of the dissimilarity measure performed for three cases validated the proposed measure. In the present analysis, the precipitation zones were avoided making use of the available meteorological data. In the absence of the availability of this data, a criteria based on higher order statistics can be used which will be presented in another work but not included here to avoid unnecessary digressions. It was observed that the differences present a significant singularity and since they form part of permanent structures, so they are always detected. Another singularity that exists always at the current acquisition site is not always detected. A continuous monitoring over the acquisition period can be exploited to detect the singularities. It was shown that the first SVD source vector can be used to extract a reference vector over this 24-hour interval. The residue subspace constructed with SVD over several days can be used to construct a dissimilarity matrix employing the $\ell_2$ norm. The resultant measure can be averaged out over several days thus minimizing the external effects during the acquisition thereby providing an efficient means of singularity detection. The tests on several real scenarios revealed the efficiency of this system. It was shown that it works well both for the artificially induced impulsive water leakages as well as the real water leakages that may occur in the dikes due to some internal deterioration phenomena in the dikes.

VI. CONCLUSION

In this paper, we presented a system for the identification of singularities in and around the dikes of a canal using real temperature data sets acquired by distributed temperature sensors. The DTS offer the practical solutions for long term, automatic temperature acquisitions with the exception advantage of using low-cost and environmentally robust optical fiber cables. The acquired data though is not immediately exploitable and thus signal processing is needed. We showed in this work how the temperature variations over the 24-hour period can be exploited to detect the singularities. It was shown that the first SVD source vector can be used to extract a reference vector over this 24-hour interval. The residue subspace constructed with SVD over several days can be used to construct a dissimilarity matrix employing the $\ell_2$ norm. The resultant measure can be averaged out over several days thus minimizing the external effects during the acquisition thereby providing an efficient means of singularity detection. The tests on several real scenarios revealed the efficiency of this system. It was shown that it works well both for the artificially induced impulsive water leakages as well as the real water leakages that may occur in the dikes due to some internal deterioration phenomena in the dikes.

REFERENCES