

# Séparation de sources thermométriques pour la détection des fuites dans des digues en terre

Amir A. KHAN<sup>1</sup>, Valeriu VRABIE<sup>2</sup>, Jérôme MARS<sup>1</sup>, Guy D'URSO<sup>3</sup>

<sup>1</sup>GIPSA-Lab, Département Images Signal, Grenoble INP, BP 46, 38402 Saint Martin d'Hères Cedex, France

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

<sup>3</sup>R&D, Électricité de France (EDF), 6 quai Watier, 78401 Chatou Cedex, France

{amir-ali.khan, jerome.mars} @gipsa-lab.inpg.fr, valeriu.vrabie@univ-reims.fr, guy.durso@edf.fr

**Résumé** – Les mesures de température par le système DTS (Distributed Temperature Sensors) basé sur les fibres optiques est une solution économiquement efficace pour la détection des fuites d'eau dans les digues en terre. Les données de température acquises sont cependant bruitées par des facteurs dus par exemple aux variations journalières, saisonnières, à l'environnement géomécanique, au bruit. Des techniques de traitement du signal doivent être appliquées pour pouvoir extraire l'information utile des anomalies. Ainsi la détection des fuites peut être traitée comme un problème de séparation de sources. Les variations saisonnières sont tout d'abord enlevées de données. Ensuite, les techniques de séparation de sources de type SVD et/ou Analyse en Composants Indépendantes (ACI) sont appliquées sur les données prétraitées. La réponse du sol étant la partie la plus cohérente des données, elle masque les informations relatives aux anomalies. Cette réponse est estimée comme la première source de la SVD. L'ACI est ensuite appliquée sur le sous-espace résidu obtenu permettant de localiser les anomalies. Le traitement proposé est validé pour la détection des fuites simulées et réelles.

**Abstract** – DTS (Distributed Temperature Sensors) based on the optical fibers present an economical efficient solution for the detection of water leakages in the dikes of canal. However, the acquired temperature data are corrupted by factors like daily and seasonal variations, geophysical conditions, noise, etc. The signal processing techniques must thus be applied to extract the useful information linked to the anomalies. In this context, the leakage detection can be modeled as a source separation problem. The source separation techniques of SVD and Independent Component Analysis are applied to the data. The self-response of the ground in which fiber is buried is the most coherent component of the data and masks the information linked to anomalies. This response is estimated as the first source of SVD. ICA is then applied on the residual sub-space obtained thus allowing to localize the leakages. The proposed method is validated on real data for the detection of artificial percolation type and real leakages.

## 1 Introduction

De nos jours, les changements climatiques ont de répercussions importantes sur différentes infrastructures. Parmi eux, les plus exposées sont les infrastructures de génie civil qui bordent de nombreux cours de l'eau telles que les digues en terre. Le développement des méthodes efficaces pour leur surveillance et leur diagnostic devient alors inéluctable. Pour les digues en terre, l'écoulement préférentiel à travers l'ouvrage est un indicateur de destruction interne. La détection des fuites d'eau le long de l'ouvrage est alors un moyen efficace pour le diagnostic de cette infrastructure. Les méthodes conventionnelles du diagnostic comprennent l'investigation périodique, l'inspection visuelle et la mesure des différents paramètres qui sont importantes pour assurer la sécurité de l'ouvrage. Quant à des méthodes non-conventionnelles, quelques méthodes comme la méthode de résistivité, la méthode de potentiel spontané et la méthode de mesures de température sont plus privilégiées que d'autres. La principale limitation des méthodes de résistivité et de potentiel spontané est qu'elles restent manuelles, c.-à-d., qu'elles nécessitent une intervention humaine pour la collecte des mesures et requièrent une certaine prudence dans l'implémentation des capteurs sur le terrain.

Quand nous parlons des méthodes automatiques pour le diagnostic de grand linéaires comme les digues en terre, la notion de coût devient aussi importante. Dans ce cadre, les méthodes de diagnostic basées sur les mesures de température par fibre optique présentent des solutions économiquement efficaces grâce

à DTS (Distributed Temperature Sensors) [1]. Le principe des mesures thermométriques pour l'identification des anomalies est lié au fait qu'un écoulement préférentiel dans la digue provoque une anomalie thermique [2]. En l'absence d'une anomalie, c'est le phénomène de conduction qui conduit la mesure : le transport de chaleur résulte de l'interaction entre la température de l'air et de l'eau à l'amont. Un écoulement de l'eau à travers la digue apporte une chaleur supplémentaire par le phénomène de l'advection.

Les données de températures acquises sont cependant affectées non seulement par les fuites, mais aussi par d'autres facteurs comme la réponse du sol dans lequel la fibre est enterrée, les structures existantes (drains), etc. De plus, les précipitations, les variations saisonnières, les variations jours/nuits peuvent également changer la température du sol. La détection des fuites peut être traitée comme un problème de séparation de sources où les sources sont composées des phénomènes ci-dessus. Un schéma basé sur les techniques de séparation de sources de type Décomposition en Valeurs Singulières (SVD) [3] et Analyse en Composants Indépendantes (ACI) [4] est proposé. La réponse du sol étant la partie la plus cohérente des données, masque les informations relatives aux anomalies. Cette réponse est estimée comme la première source de la SVD. L'ACI est ensuite appliquée sur le sous-espace résidu obtenu en enlevant cette réponse du sol [5]. Un deuxième sous-espace résidu est obtenu par l'ACI, permettant de localiser les anomalies. Le schéma proposé est validé pour la détection des fuites simulées et réelles.

## 2 Système d'acquisition et données thermométriques

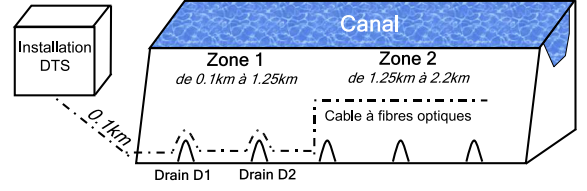
Le système d'acquisition est composé d'une fibre optique utilisée comme capteur de température et d'un appareil de mesure, le DTS (Distributed Temperature Sensing). Le système est basé sur la principe de réflectométrie optique dans le domaine temporel (OTDR pour Optical Time Domain Reflectometry en anglais). Un laser génère une pulsation de courte durée et est couplé à une fibre optique à l'aide d'un coupleur directif, ce qui permet également d'envoyer vers un détecteur la lumière rétrodiffusée retournée par la fibre. Le DTS est fondé sur la diffusion de Raman : une impulsion laser envoyée dans la fibre induit des vibrations moléculaires résultant en une diffusion spontanée Raman avec deux composantes : Stokes et Anti-Stokes, émises aux deux longueurs d'onde différentes autour de la longueur d'onde incidente. Le rapport des intensités de la composante Anti-Stokes (fortement corrélée à la température) et de la composante Stokes (indépendante de la température) permet de retrouver la température en fonction de la longueur de fibre [5]. Parmi les avantages offerts par ces capteurs distribués de la température sont : leur capacité d'intégrer un grand nombre de capteurs sur une seule fibre optique ; utilisation d'une fibre optique de télécommunication classique pas chère ; une très bonne isolation électrique ; très immun aux effets d'interférence électromagnétique, etc.

Un système thermométrique a été installé par l'Électricité de France (EDF) avec le but d'extraire les informations liées aux fuites (localisation en distance, débits, etc.) dans la digue du canal d'Oraison. Le schéma de la Fig. 1(a) montre cette installation. Une fibre optique est enterrée en aval du canal à une profondeur de 1 m pour intercepter les fuites d'eau. Deux niveaux distincts d'élévations (Zone 1, de 0,1 km à 1,25 km et Zone 2, de 1,25 km à 2,2 km) sont exposés de façon différente à l'ensoleillement. Le câble passe autour de deux drains, D1 et D2, situés à 0,561 km et à 0,858 km. Une acquisition de température correspond à des mesures le long de l'ouvrage avec un pas dépendant de la résolution spatiale de l'appareil de mesure, 1 m dans notre cas. Pour pouvoir suivre l'évolution des anomalies au cours du temps, plusieurs enregistrements sont effectués avec une résolution temporelle de 2 heures. Les données enregistrées peuvent s'exprimer sous la forme matricielle  $\mathbf{Y} = \{y(x, t) \mid 1 \leq x \leq N_x, 1 \leq t \leq N_t\}$  en fonction de la distance de mesure  $x$  et du temps  $t$ , avec  $N_x$  et  $N_t$  représentant respectivement le nombre de capteurs et le temps total d'acquisition.

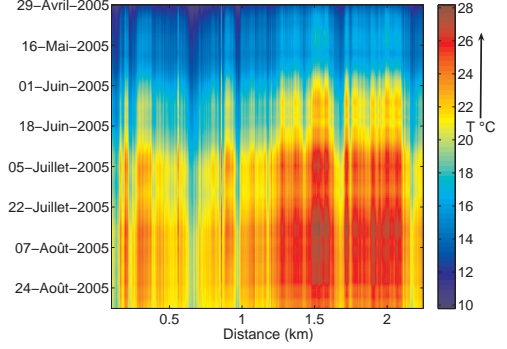
Un jeu de données enregistré en 2005 est montré sur la Fig. 1(b). En plus des structures permanentes, les drains  $D1$  et  $D2$ , trois essais de fuites de type percolation,  $L1$ ,  $L2$  et  $L3$  avec différents débits et un essai de point chaud ( $HP$ ) font partie de cette période d'analyse. Nous disposons d'autres jeux de données, dont un acquis en 2007, qui comporte des fuites réelles.

## 3 Méthodologie

Une fuite résulte dans le changement de la température mesurée mais ce changement peut être également dû à d'autres facteurs comme les structures existantes (drains, tunnels, re-



(a) Système d'acquisition



(b) Données Températures sur en 2005 avec des essais de fuites artificielles

FIG. 1 – Installation du système d'auscultation à base de DTS avec une résolution en température de  $0.01^\circ C$  et une résolution spatiale d'1 m.

gards), les variations saisonnières de température, le rayonnement, les inhomogénéités du terrain, la réponse du sol dans lequel la fibre est enterrée, etc. Le signal enregistré peut être modélisé comme un mélange des facteurs ci-dessus. Même si ces différents facteurs peuvent avoir un effet non-linéaire sur les températures mesurées, nous considérons ici uniquement les valeurs de température induites par ces facteurs sans vouloir quantifier les facteurs qui les ont produits. A chaque facteur, nous pouvons affecter une source de température (comme fonction du déplacement le long la fibre) assumant que les températures induites par chaque facteur vient pondérer linéairement la température mesurée par la fibre optique. De plus, comme ces différents facteurs sont physiquement indépendants, nous pouvons assumer que les sources des températures associées sont indépendantes. La matrice de température peut alors s'exprimer comme :

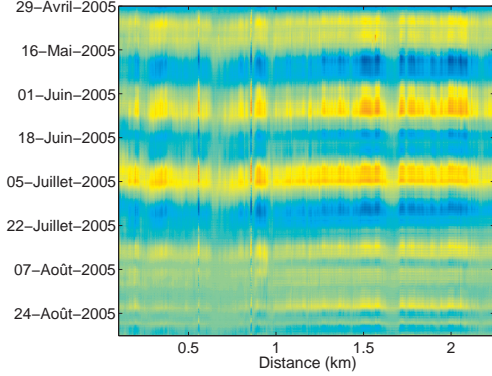
$$\mathbf{Y}^T = \mathbf{M}\mathbf{F}^T, \quad (1)$$

où  $\mathbf{M} \in \mathbb{R}^{N_t \times i}$  la matrice de mélange,  $\mathbf{F} \in \mathbb{R}^{N_x \times i}$  la matrice de sources indépendantes associées au différents facteurs. Les températures captées par la fibre optique montre une dépendance sinusoïdale au cours du temps. Dans un premier temps, une désaisonnalisation qui consiste à retirer la composante saisonnière du signal brut  $\mathbf{Y}$  de température, peut être effectuée. Une approche possible est de modéliser les variations annuelles ainsi que les variations saisonnières par des sinusoides et de prendre en compte les éventuels décalages par rapport à un modèle idéal par un mélange de sinus et de cosinus. Ces variations sont estimées à chaque distance  $x$  en posant le modèle :

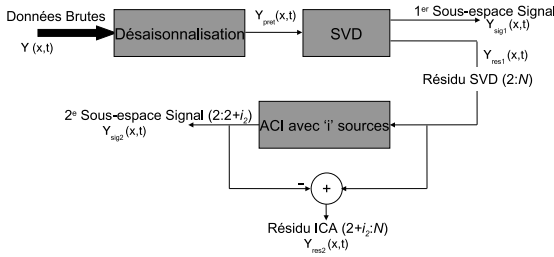
$$y_{var}(t) = a_m + \sum_{n=1}^2 [a_{1n} \sin(ns)] + [a_{2n} \cos(ns)], \quad (2)$$

avec  $s = \frac{2\pi t}{365N_j}$ ,  $t = [1 : N_t]$  et  $N_j$  le nombre d'acquisitions par jour avec les différents paramètres  $a_m$ ,  $a_{1n}$  et  $a_{2n}$  estimés en utilisant une approche des moindres carrés. Les

données prétraitées,  $\mathbf{Y}_{\text{pret}} = \mathbf{Y} - \mathbf{Y}_{\text{var}}$ , affichées sur la Fig. 2(a), montrent que les fortes variations temporelles sont enlevées. Une normalisation des données avec le but d'accorder la même importance à tous les signaux enregistrés est effectuée. Cette normalisation faite pour rendre chaque acquisition à une moyenne nulle et une variance unité permet d'atténuer les variations d'une acquisition à l'autre. Suivant les pretrait-



(a) Les données désaisonnalisées en enlevant les variations annuelles et saisonnières



(b) Technique de séparation de sources basée sur la SVD et l'ACI

FIG. 2 – Données désaisonnalisées et la synoptique pour la détection de fuites dans les digues utilisant une approche de séparation de source (SVD et ACI). La technique de séparation de sources pour la détection des fuites.

tements, nous proposons un algorithme pour la détection des fuites basé sur la Décomposition en Valeur Singulière (SVD) [3] et l'Analyse en Composants Indépendantes (ACI) [4]. La première étape de l'analyse focalise sur l'extraction de la réponse du sol à partir des données à l'aide de la SVD. Soit  $\mathbf{Y} \in \mathbb{R}^{N_x \times N_t}$  un jeu de données thermométriques, la SVD de cette matrice est :

$$\mathbf{Y}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (3)$$

où  $\mathbf{V}$  et  $\mathbf{U}$  sont des matrices orthogonales respectivement de dimension  $(N_x, N)$  et  $(N_t, N)$ , avec  $N = \min(N_x, N_t)$ . Les vecteurs  $\mathbf{v}_k \in \mathbb{R}^{N_x}$ , donnant la dépendance en fonction de la distance, peuvent être interprétés comme des estimateurs des sources [6]. Comme les vecteurs sont orthonormés par construction, les sources estimées sont décorrélées et normalisées. Les vecteurs  $\mathbf{u}_k \in \mathbb{R}^{N_t}$ , correspondant aux variations au cours du temps des sources estimées  $\mathbf{v}_j$ .  $\mathbf{\Sigma}$  est une matrice diagonale contenant les valeurs singulières ordonnées. La première valeur singulière est généralement grande, la source correspondante montre qu'elle est liée à la réponse du sol. Cependant, son estimation par une SVD appliquée sur les données globales risque d'être affectée par des phénomènes énergétiques éphémères en temps. On applique alors la SVD en fenêtrage glissant en temps

sur les données sélectionnées par un critère basé sur les statistiques d'ordre supérieur [7]. La SVD permet de décomposer les données en sous-espaces orthogonaux, le sous-espace signal et le sous-espace résidu, définis par :

$$\mathbf{Y}^T = \mathbf{Y}_{\text{sig}1}^T + \mathbf{Y}_{\text{res}1}^T = \sum_{k=1}^m \sigma_j \mathbf{u}_j \mathbf{v}_j^T + \sum_{k=m+1}^N \sigma_j \mathbf{u}_j \mathbf{v}_j^T. \quad (4)$$

Pour les données thermométriques, l'Eq. 4 avec  $m = 1$  permet d'extraire la réponse du sol et ensuite d'obtenir un sous-espace résidu  $\mathbf{Y}_{\text{res}1}$ . Pour mieux extraire les fuites et les drains, qui sont des phénomènes non-gaussiennes, la contrainte d'orthogonalité imposée aux sources estimées peut s'avérer insuffisante. En plus, nous n'avons pas aucune justification physique pour l'orthogonalité des variations temporelles de ces sources imposée par la SVD.

Afin de mieux extraire chaque phénomène et principalement les fuites, on fait appel à l'ACI qui est basée sur l'indépendance statistique des sources [8, 9]. L'indépendance statistique signifie que les cumulants croisés de toute l'ordre sont nulles. L'ACI peut être formulée comme :

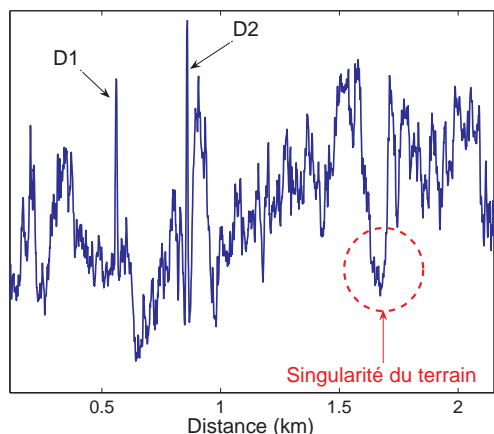
$$\mathbf{Z}^T = \mathbf{A}\mathbf{S}^T, \quad (5)$$

avec  $\mathbf{Z} \in \mathbb{R}^{N_x \times N_t}$  une matrice d'observation,  $\mathbf{S} \in \mathbb{R}^{N_x \times i}$  une matrice des sources et  $\mathbf{A} \in \mathbb{R}^{N_t \times i}$  une matrice de mélange. Le but d'ACI est d'estimer  $\mathbf{A}$  et/ou  $\mathbf{S}$  à partir des observations  $\mathbf{Z}$  avec la seule l'hypothèse de l'indépendance des sources. L'ACI est un algorithme à deux étapes, une première étape de blanchiment et une deuxième étape de maximisation de l'indépendance. La première étape est obtenue directement par la SVD pour obtenir des vecteurs blancs, c.à.d., décorrélés et normalisés,  $\mathbf{v}_j$ . A ce point, une matrice,  $\mathbf{V}_p$  peut être construite en considérant  $i$  vecteurs  $\mathbf{v}_j$  parmi ce qui définissent le sous-espace résidu  $\mathbf{Y}_{\text{res}1}$ . La matrice  $\mathbf{Z}$  représente alors le sous-espace construit avec les  $i$  vecteurs correspondants  $\mathbf{v}_j$  et  $\mathbf{u}_j$ . L'ACI ensuite estime une matrice de rotation  $\mathbf{B}_j \in \mathbb{R}^{i \times i}$  soit en diagonalisant le tenseur de cumulants de l'ordre 4 (JADE) soit en maximisant une autre critère de la non-gaussianité (par exemple FastICA) [4, 9]. A partir de ces  $i$  sources estimées, un nouveau sous-espace signal,  $\mathbf{Y}_{\text{sig}2}$ , contenant principalement les drains, peut être construit avec  $i_2$  sources indépendantes. Les fuites étant à l'énergie faible restent toujours dans le sous-espace résidu,  $\mathbf{Y}_{\text{res}2} = \mathbf{Y}_{\text{res}1} - \mathbf{Y}_{\text{sig}2}$ . Ce résidu de l'ACI contient l'information liée aux fuites avec l'influence des autres facteurs enlevée. La méthode détaillée ci-dessus peut être synthétisée par le schéma de la Fig. 2.

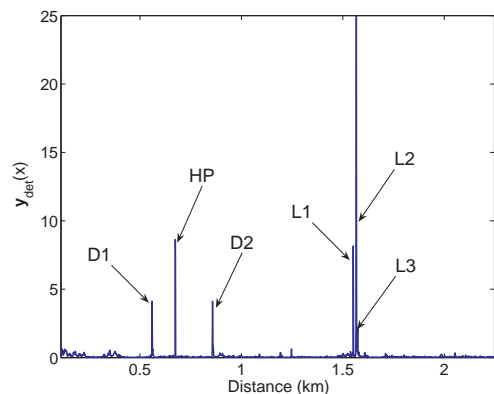
## 4 Applications sur des données thermométriques réelles

L'algorithme proposé a été validé sur plusieurs jeux de données enregistrés à différentes périodes. On présente ici les résultats pour deux cas intéressants : 1) Fuites artificielles de type percolation ; 2) Fuites réelles. On considère premièrement les données de la Fig. 1(b) où 3 fuites ont été simulées en Mai 2005 : L1 (le matin du 12 mai, à 1.551 km, débit 11/min), L2 (le midi du 10 mai, à 1.566 km, débit 51/min) et L3 (le soir du 12 mai, à 1573 km, débit 11/min). Ces fuites sont considérées comme ponctuelles en temps avec une durée d'un jour et demi

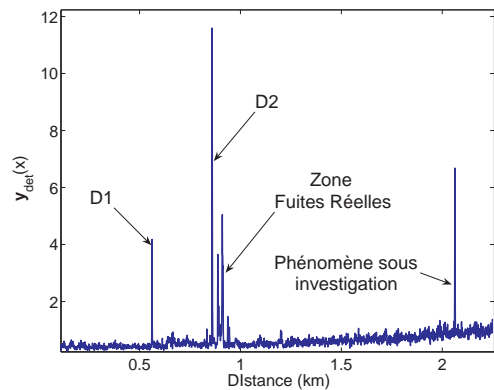
pour  $L2$  et quelques heures pour  $L1$  et  $L3$ . Après avoir désaisonnalisés les données, comme le montre la Fig. 2(a), on applique l'algorithme proposé. Une fenêtre glissante de 14 jours a été empiriquement choisie pour estimer d'une façon optimale la réponse du sol. La première valeur singulière représente 96 % de l'énergie totale de données. L'étude de la source correspondante, montrée en Fig. 3(a), confirme que cette source est liée à la réponse du sol (singularité du terrain dans la zone 1.6-1.7 km correspondant à une zone où la composition du sol est différente des autres zones, deux zones distinctes correspondant à l'ensoleillement différent, etc.). Sur les autres sources SVD,



(a) La première source obtenue avec l'application de SVD, un estimateur de la réponse du sol



(b)  $y_{det}(x)$  pour l'année 2005 montrant détection des fuites artificielles et le point chaud



(c)  $y_{det}(x)$  pour l'année 2005 montrant détection d'une zone des fuites réelles

FIG. 3 – Les résultats de détection avec la méthode basé sur la SVD et l'ACI.

les phénomènes (fuites, drains, etc.) restent mélangés. On l'applique donc l'ACI sur le résidu donnée par la SVD,  $Y_{res1}$ . Différents tests ont été effectués pour choisir le nombre optimal de sources sur lesquelles l'ACI sera appliquée. On constate qu'avec  $i = 2$  nous obtenons les meilleurs résultats. La première source après l'application de l'ACI permet d'enlever une grande partie des drains, les fuites se retrouvant dans le résidu  $Y_{res2}$ . Pour simplicité d'affichage, on fait une projection en distance de ce résidu, c.à.d.  $y_{det}(x) = \sum_{t=1}^{N_t} |y_{res2}(x, t)|^2$ ,  $x = 1, \dots, N_x$ . Les résultats obtenus pour les données de 2005 (voir Fig. 1(b)) sont affichés sur la Fig. 3(b). Cette figure indique une détection optimale des fuites, même si les drains restent présents dans le résidu, avec la fuite  $L1$  la plus puissante (débit le plus élevé). Le point chaud,  $HP$  est également détecté. Dans un deuxième temps, on applique la même procédure avec les mêmes paramètres sur les données acquises en 2007. La projection du résidu, affichée sur la Fig. 3(c), révèle la présence de drains, comme en 2005. En plus, juste après  $D2$ , des nouvelles anomalies sont identifiées. Une investigation physique sur le site a confirmé la présence des fuites réelles qui, à la différence des fuites artificielles impulsionnelles de 2005, sont localisées sur une dizaine de mètres. Cependant un autre phénomène a été identifié vers 2 km. Même si ce phénomène a été éphémère en temps (2 jours), cette zone est actuellement sous investigation.

## 5 Conclusion

Le système d'acquisition basé sur DTS et fibre optique présente une solution efficace, économe et robuste pour la détection des anomalies dans les grands ouvrages. Cependant, les températures acquises par les DTS ne sont pas directement interprétables pour la détection des fuites dans des digues en terre. Nous avons proposé une approche utilisant des techniques de séparation de source telles que SVD et ACI. La méthode proposée a été validée sur des fuites artificielles de types percolation et sur des fuites réelles.

## Références

- [1] 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.
- [2] 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.
- [3] L. L. Scharf, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*. New York : Addison-Wesley, 1991.
- [4] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.
- [5] 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.
- [6] 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.
- [7] 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 *EUSIPCO*, Lausanne, Switzerland, Aug 2008.
- [8] P. Comon, "Independent component analysis, A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [9] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York : Wiley, 2001.