



**HAL**  
open science

## Séparation de sources thermométriques.

Amir Ali Khan

► **To cite this version:**

Amir Ali Khan. Séparation de sources thermométriques.. Traitement du signal et de l'image [eess.SP]. Institut National Polytechnique de Grenoble - INPG, 2009. Français. NNT: . tel-00477455

**HAL Id: tel-00477455**

**<https://theses.hal.science/tel-00477455>**

Submitted on 29 Apr 2010

**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.

**INSTITUT POLYTECHNIQUE DE GRENOBLE**

N° attribué par la bibliothèque

| | | | | | | | | |

**THESE**

pour obtenir le grade de

DOCTEUR DE L'Institut polytechnique de Grenoble

**Spécialité : « Signal Image Parole et Télécom »**

préparée au Gipsa-lab Grenoble Images Parole Signal Automatique

dans le cadre de l'**École Doctorale « Electronique , Electrotechnique, Automatique, Traitement du Signal »**

présentée et soutenue publiquement

par

**Amir Ali KHAN**

le 27 novembre 2009

*Titre :*

**SÉPARATION DE SOURCES THERMOMÉTRIQUES**

**Directeurs de thèse : Jérôme I. Mars**

**Co-directeur de thèse : Valeriu Vrabie**

**JURY**

|          |                            |                       |
|----------|----------------------------|-----------------------|
| Monsieur | Olivier Métais,            | Président             |
| Monsieur | Jean-Jacques Fry,          | Rapporteur            |
| Monsieur | François Guillet,          | Rapporteur            |
| Monsieur | David Brie,                | Examineur             |
| Monsieur | Jean-François Giovannelli, | Examineur             |
| Monsieur | Jérôme I. Mars,            | Directeur de thèse    |
| Monsieur | Valeriu Vrabie,            | Co-directeur de thèse |
| Monsieur | Guy d'Urso,                | Examineur             |



*à mes parents  
à Sana et Shanzeh*



# Acknowledgements

It is an immense pleasure to thank all those who have made this thesis possible. Being in the company of people who support and inspire you simultaneously allows you to come through the difficult phases of your thesis.

First of all, I would like to express my deepest gratitude to all the thesis committee members.

Mr. Olivier Métais for giving me the honor of accepting to be the president of the committee.

Mr. Jean-Jacques Fry for being the reviewer of my thesis, for his expert opinions regarding fiber optic sensors in the surveillance system and for invoking fruitful discussion.

Mr. François Guillet for being the reviewer of my thesis, for his critical reviews and for his useful suggestions in improving the final manuscript.

Mr. David Brie and Mr. Jean-François Giovannelli for accepting to be the examiners of my thesis and for their critical analysis of my work.

Mr. Guy d'Urso for his suggestions regarding the industrial aspects of the work, for periodic technical discussions, for provision of data, and for making possible my visits to the experimental sites.

Mr. Jérôme Mars, for being the coolest of the supervisors and for his guidance in professional and personal capacity. I really appreciate his work ethic, always taking out time for discussions on the progress, giving me autonomy to present my work in diverse circles, and for his critical eye regarding the methods and the results. His guidance made my research life smooth and rewarding.

Mr. Valeriu Vrabie, more of a friend to me than an advisor. My work falls in the context of remote sensing, not only in its thematic aspect but also in the sense that i had a remote advisor off-site but never off-sight. His guidance and encouragement through good times and bad was the highlight of my doctorate. He always had that precious little to tell you when you thought you have mastered your art. He took the big task on himself of correcting the French of my manuscript, not to mention his enlightenment of the technical details.

All the colleagues at GIPSA-lab, who made it a convivial place to work, to all my ever changing office buddies, thanks for your support. Not to forget my Khan connection at the lab, Murtaza, was great to be in his company for the last four years, thanks for all those coffee breaks and for technical chit chats. I am grateful to all the colleagues who intervened at different stages through their physical insight, especially Cyril Guidoux, the guys at EDF, Alexandre Girard, Pierre Cunat and Yves-Laurent Beck. I would take a moment to thank my friends from the Pakistani community here in Grenoble for all those lovely gatherings and company.

## *ACKNOWLEDGEMENTS*

---

My deepest gratitude goes to my family for their unflagging support and love throughout my life. This dissertation was made possible through my father's inspirations and my mother's prayers. I owe it to my parents who left no stone unturned to allow us to grow up in the best possible environment. To the most important lady in my life, Sana, for bearing with me over all these years, for having sacrificed a part of her career to always be on my side, for her indefatigable perseverance, for keeping up with all my relations when I could not, for all those lovely evenings and weekends, I thank you for standing by me in all the tough phases of my life, for your patience even when we did not have enough time to sit together specially during the last 6 months and for single handedly taking care of our princess Shanzeh. Not to forget, my sage daughter Shanzeh who grew up ever so quickly, who always gave me a laughing moment through her activities and who has now taken upon herself to properly utilize all my scrap papers through her abstract art. Last but not the least, praise be to God the most merciful and magnanimous, for directing me through thick and thin.

# Table des matières

|   |           |
|---|-----------|
| <b>Table des matières</b>   | <b>i</b>  |
| <b>Introduction</b>   | <b>3</b>  |
| <b>1 Etat de l'art des techniques de détection des fuites</b>   | <b>9</b>  |
| 1.1 Introduction . . . . .  | 11        |
| 1.2 Principaux mécanismes d'endommagement des digues en terre . . . . .   | 11        |
| 1.3 Diagnostic et surveillance des ouvrages . . . . .   | 14        |
| 1.3.1 Méthodes courantes de diagnostic . . . . .  | 14        |
| 1.3.2 Mesures de grandeurs physiques . . . . .  | 15        |
| 1.3.3 Méthodes de surveillance . . . . .  | 15        |
| 1.3.4 Méthodes non-conventionnelles de surveillance . . . . .   | 16        |
| 1.4 Mesures de température par capteurs à fibre optique . . . . .   | 21        |
| 1.4.1 Capteurs à fibre optique . . . . .  | 21        |
| 1.4.2 Capteurs distribués à fibre optique . . . . .   | 25        |
| 1.4.3 Télémétrie optique : réflectométrie optique dans le domaine temporel . . .  | 28        |
| 1.4.4 Mesure de température par rétrodiffusion Raman . . . . .  | 29        |
| 1.4.5 Spécification des paramètres et mesures des performances . . . . .  | 32        |
| 1.4.6 Avantages de capteurs distribués à fibre optique . . . . .  | 34        |
| 1.4.7 Applications des capteurs distribués à fibre optique . . . . .  | 35        |
| 1.5 Principe de détection des fuites en digues par mesures de température à l'aide de capteurs distribués à fibre optique . . . . . | 37        |
| 1.6 Conclusion . . . . .  | 38        |
| <b>2 Description des mesures thermométriques et Prétraitements</b>  | <b>41</b> |
| 2.1 Introduction . . . . .  | 43        |
| 2.2 Données simulées . . . . .  | 43        |
| 2.2.1 Modélisation des températures d'une digue . . . . .   | 43        |
| 2.2.2 Génération d'une digue . . . . .  | 47        |
| 2.2.3 Signaux de température simulés . . . . .  | 49        |
| 2.3 Le site expérimental d'Oraison . . . . .  | 51        |
| 2.3.1 Généralités . . . . .   | 52        |



|          |   |           |
|----------|---|-----------|
| 2.3.2    | Caractéristiques des données enregistrées à Oraison . . . . .           | 54        |
| 2.3.3    | Les fuites localisées en temps (fuites ponctuelles) . . . . .           | 57        |
| 2.4      | Le site expérimental de Kembs . . . . .                                 | 60        |
| 2.4.1    | Généralités . . . . .   | 60        |
| 2.4.2    | Caractéristiques des données enregistrées . . . . .                     | 62        |
| 2.5      | Modélisation des signaux de température . . . . .                       | 67        |
| 2.6      | Prétraitements . . . . .  | 69        |
| 2.6.1    | Désaisonnalisation . . . . .  | 69        |
| 2.6.2    | Normalisation des données . . . . .                                     | 70        |
| 2.6.3    | Identification des phénomènes éphémères énergétiques . . . . .          | 71        |
| 2.7      | Discussion . . . . .  | 75        |
| <b>3</b> | <b>Séparation de sources pour la détection des fuites</b>               | <b>79</b> |
| 3.1      | Introduction . . . . .  | 81        |
| 3.2      | Décomposition en valeurs singulières . . . . .                          | 82        |
| 3.2.1    | Définitions et propriétés . . . . .                                     | 82        |
| 3.2.2    | Interprétation de la SVD d'un jeu de données thermométriques . . . . .  | 84        |
| 3.3      | L'analyse en composantes indépendantes . . . . .                        | 87        |
| 3.3.1    | Indépendance statistique . . . . .                                      | 87        |
| 3.3.2    | Définition de l'ICA et ses indéterminations . . . . .                   | 88        |
| 3.3.3    | Prétraitements . . . . .  | 89        |
| 3.3.4    | Mesures d'indépendance . . . . .  | 90        |
| 3.3.5    | Séparation de sources en mode symétrique et en mode déflation . . . . . | 94        |
| 3.3.6    | Application aux données réelles . . . . .                               | 95        |
| 3.4      | Méthodologie proposée pour la détection des fuites . . . . .            | 96        |
| 3.5      | Application aux données simulées . . . . .                              | 99        |
| 3.5.1    | Analyse de données entières . . . . .                                   | 102       |
| 3.5.2    | Analyse par fenêtrage glissant en temps . . . . .                       | 103       |
| 3.5.3    | Interprétation des résidus . . . . .                                    | 108       |
| 3.5.4    | Étude des résidus par rapport aux différents paramètres . . . . .       | 109       |
| 3.6      | Application aux données enregistrées sur le site d'Oraison . . . . .    | 112       |
| 3.6.1    | Analyse de données entières . . . . .                                   | 114       |
| 3.6.2    | Analyse par fenêtrage glissant en temps . . . . .                       | 115       |
| 3.7      | Application aux données enregistrées sur le site de Kembs . . . . .     | 122       |
| 3.7.1    | Analyse de données entières . . . . .                                   | 124       |
| 3.7.2    | Analyse par fenêtrage glissant en temps . . . . .                       | 124       |
| 3.8      | Discussion . . . . .  | 129       |

|          |  |            |
|----------|--|------------|
| <b>4</b> | <b>Détecteur des singularités par approche journalière</b>   | <b>135</b> |
| 4.1      | Introduction . . . . .   | 137        |
| 4.2      | Méthodologie pour la détection des singularités . . . . .  | 137        |
| 4.2.1    | Définition de singularités . . . . .   | 138        |
| 4.2.2    | Principe d'identification de singularités . . . . .  | 138        |
| 4.2.3    | Mesure de dissimilarité . . . . .  | 139        |
| 4.3      | Détection de fuites par seuillage d'histogramme . . . . .  | 141        |
| 4.3.1    | Seuillage en fixant une probabilité de fausse alarme . . . . .   | 145        |
| 4.4      | Système proposé pour la détection de singularités . . . . .  | 148        |
| 4.5      | Application aux données simulées . . . . .   | 150        |
| 4.6      | Application aux données enregistrées sur le site d'Oraison . . . . .                                   | 154        |
| 4.7      | Discussion . . . . .   | 159        |
|          | <b>Conclusion et perspectives</b>  | <b>165</b> |
|          | <b>Annexes</b>   | <b>169</b> |
| <b>A</b> | <b>Autres méthodes de détection de fuites de type percolation</b>                                      | <b>171</b> |
| A.1      | Filtrage Multicomposante . . . . .   | 171        |
| A.1.1    | Introduction . . . . .   | 171        |
| A.1.2    | Extension multicomposante (3C-2D) des données monocomposantes (1C-2D) . . . . .                        | 171        |
| A.1.3    | Filtrage multicomposante . . . . .   | 173        |
| A.1.4    | Application sur les données d'Oraison . . . . .  | 173        |
| A.1.5    | Conclusion . . . . .   | 176        |
| A.2      | Détection des fuites de type percolation (JADE vs FastICA) . . . . .                                   | 178        |
| <b>B</b> | <b>Annexe relative à la réponse du sol</b>   | <b>181</b> |
| B.1      | Extraction de la réponse du sol par critère HOS en utilisant une approche de moindres carrés . . . . . | 181        |
| B.1.1    | Estimation d'une source moyenne . . . . .  | 181        |
| B.1.2    | Décomposition en sous-espace par approche de moindres carrés . . . . .                                 | 183        |
| B.1.3    | Résultats et discussion . . . . .  | 183        |
| <b>C</b> | <b>Annexes relatives au chapitre 2</b>   | <b>187</b> |
| C.1      | Définition des Cumulants . . . . .   | 187        |
| C.2      | Position des points de mesure pour les données thermométriques simulées. . . . .                       | 188        |

|          |   |            |
|----------|---|------------|
| <b>D</b> | <b>Annexes relatives au chapitre 3</b>  | <b>191</b> |
| D.1      | Analyse des données simulées . . . . .  | 191        |
| D.1.1    | Analyse sans fenêtrage glissant en temps des données simulées avec des fuites d'un jour. . . . .                      | 191        |
| D.1.2    | Analyse des fuites simulées d'une durée d'un jour par fenêtrage glissant de 15 jours . . . . .                        | 193        |
| D.1.3    | Étude par rapport à la durée des fuites : Exemple de fuites simulées d'une durée de 5 jours . . . . .                 | 195        |
| D.2      | Analyse des données d'Oraison . . . . .   | 197        |
| D.2.1    | Sources ICA pour $m = 1$ , $p = 6$ , pour l'analyse globale des données enregistrées sur le site d'Oraison . . . . .  | 197        |
| D.2.2    | Résultats de l'algorithme de séparation de sources pour les données d'Oraison avec différents prétraitements. . . . . | 198        |
| D.3      | Analyse des données de Kembs . . . . .  | 199        |
| D.3.1    | Sources ICA pour l'analyse globale avec $m = 0$ , $p = 7$ . . . . .   | 199        |
| D.3.2    | Sources ICA pour l'analyse par fenêtrage glissant de 14 jours pour la période entre 29 avril et 13 mai 2007. . . . .  | 200        |
| D.3.3    | Analyse avec un fenêtrage glissant de 40 jours. . . . .   | 202        |
| <b>E</b> | <b>Annexe relative au chapitre 4</b>  | <b>205</b> |
| E.1      | Analyse des données du contre-canal de Kembs . . . . .  | 205        |
|          | <b>Bibliographie</b>  | <b>208</b> |

# Introduction



# Introduction

Les ouvrages qui bordent les cours d'eau, comme les barrages et les digues en terre, sont "vulnérables" à cause de différents facteurs. L'ancienneté, l'environnement sévère, l'érosion interne, les mouvements de terrains, etc., se conjuguent, contribuant à l'accélération de la défaillance de ces ouvrages. La surveillance de ces ouvrages à risque est alors essentiel afin de prévenir et de se prémunir des éventuelles catastrophes.

L'érosion interne souvent mise en évidence par la présence de fuites est une cause très importante de désordre pouvant aller jusqu'à la rupture des ouvrages hydrauliques comme les digues en terre [Fell 07]. La détection de ces fuites devient alors un indicateur essentiel de l'état des ouvrages. Le monitoring consiste à mesurer différentes grandeurs physiques afin de mettre en évidence les possibles sources d'anomalie.

L'écoulement préférentiel d'eau à travers une digue induisant une anomalie thermique, la détection de ces écoulements est révélée par l'intermédiaire de mesures de température ponctuelle. Cette procédure est ainsi utilisée depuis plusieurs années, sans qu'elle soit réellement automatique.

Une problématique importante de ce genre de mesure est liée à la taille des ouvrages qui sont généralement étendus sur plusieurs kilomètres. Les méthodes d'investigation privilégiées pour de telles structures sont les méthodes dites à grand rendement. Un axe important et essentiel des travaux de recherche actuels est dirigé vers le développement de systèmes d'auscultation, sans intervention humaine sur les sites mais avec une expertise centralisée à partir d'alertes automatiques.

Depuis peu et grâce à l'utilisation de capteurs à fibre optique, nous pouvons envisager d'obtenir des mesures thermométriques à grande échelle. Cet enrichissement récent des mesures de température pour le monitoring des ouvrages hydrauliques est lié à l'utilisation de capteurs à fibre optique depuis une quinzaine d'années pour des applications environnementales très diverses. Nous les retrouvons dans des domaines tels que : la surveillance des structures de génie civil, la détection des fautes, le comptage en génie électrique, le contrôle et le monitoring de l'environnement, le contrôle de pollution par détecteurs chimiques (H<sub>2</sub>, CO<sub>2</sub>, détecteurs d'explosifs), dans l'industrie du pétrole et du gaz pour le contrôle des paramètres en conditions extrêmes (température, pression, viscosité, débit, etc.), pour la détection d'incendies (ouvrages sous terrains et tunnels), etc. [Henderson 97, Kersey 99, Hartog 00, Grosswig 01, Hartog 02, Fry 04, Wade 04, Lewis 07].

Dans ce cadre, le développement d'un système de diagnostic automatique basé sur l'auscultation humaine de grands linéaires de digues en terre deviendrait parfaitement prohibitif en termes de coût. L'utilisation des fibres optiques de télécommunication, très peu coûteuses, est donc une alternative très intéressante et économiquement viable. Les capteurs à fibre optique représentent maintenant une technologie éprouvée. La possibilité technique de multiplexer un grand nombre de capteurs sur une seule fibre optique et la grande liberté possible sur le déploiement et l'utilisation de ces capteurs sont des avantages indéniables de ces capteurs à fibre optique. Electricité de France (EDF), acteur majeur dans le secteur d'énergie, de l'eau et de l'environnement, dispose de plusieurs dizaines de centrales hydrauliques de tailles variées. Les différents canaux d'amenée d'eau vers ces centrales sont généralement bordés par des digues en terre. L'auscultation de ces digues est une priorité très importante non-seulement pour s'assurer de la stabilité des ouvrages, mais également pour s'assurer du fonctionnement optimal de la centrale hydraulique. Une digue bordant un canal pouvant avoir une longueur très importante (plusieurs, voire une centaine de kilomètres), les capteurs à fibre optique se présentent donc comme un moyen économiquement efficace pour son diagnostic.

Afin de tester l'efficacité d'un tel système de diagnostic, EDF a installé sur plusieurs sites expérimentaux des capteurs à fibre optique le long de ces digues. Les deux sites principaux sont ceux d'Oraison (Alpes de Hautes Provence) et de Kembs en Alsace. Afin de réaliser des mesures de température, des appareils de type Distributed Temperature Sensors (DTS) sont utilisés. Ce type d'instrument permet de réaliser des mesures de thermométrie sur des longues distances (jusqu'à 30 km) et avec de très bonnes résolutions en distance autour d'un mètre et en température entre  $0.01^{\circ}\text{C}$  et  $0.1^{\circ}\text{C}$  selon le type d'enregistrement.

L'idée principale d'utiliser des données de température acquises par des DTS est qu'un écoulement préférentiel de l'eau du fait de fuites peut être mis en évidence par une mesure de différence de température entre l'eau de la fuite et une température du terrain (dans lequel la fibre optique est enterrée). Cependant, il est classiquement admis et montré que cet écart de température mesuré entre l'eau et le terrain peut être également dû à d'autres facteurs additifs comme :

- les variations saisonnières et journalières de température,  
    les variations journalières d'ensoleillement,
- les précipitations sur le terrain ausculté,
- les structures existantes (drains, tunnels, regards, canalisations etc.),
- les hétérogénéités du terrain, etc.

Les données thermométriques acquises par les DTS sont donc difficilement interprétables directement en termes d'identification et de localisation de fuites. Il est impératif de traiter ces données afin d'en extraire l'information utile (la position en distance et en temps des fuites dans notre cas). Ce travail de thèse s'attaque donc à cette problématique de séparation des informations utiles liées aux fuites, des autres facteurs "non utiles". L'objet de ce travail est de proposer des techniques avancées de séparation de sources répondant à la problématique sociétale posée.

Ce mémoire est organisé de la manière suivante :

Le premier chapitre est consacré à l'introduction de la problématique du diagnostic des ouvrages pour la détection de fuites. Afin de faciliter la compréhension des mesures acquises, la première partie se focalise sur une présentation des principales sources d'endommagement d'une digue. La deuxième partie énumère les différentes grandeurs physiques qui peuvent être des indicateurs utiles de ces endommagements, puis présente des méthodes classiques de diagnostic. Du fait que ce travail est basé sur les mesures de température par des capteurs à fibre optique, la troisième partie de ce chapitre illustre en détail le processus de mesures de température à l'aide de fibres optiques. Ainsi dans cette partie nous nous intéressons aux capteurs distribués à fibre optique en présentant leur principe de mesure, leurs différents paramètres qui peuvent influencer ces mesures, avant d'envisager, de manière générale les principaux avantages et des domaines d'application de ces capteurs.

Après avoir présenté l'état de l'art en termes de monitoring de digues, le deuxième chapitre de ce mémoire nous permettra de se pencher sur la partie terrain. Dans ce travail, trois jeux de données seront analysés : des données issues d'un simulateur de données expérimentales, des données enregistrées sur les sites expérimentaux choisis à savoir celui d'Oraison (milieu méditerranéen) et celui de Kembs (milieu continental). Tout d'abord, un simulateur de données thermométriques permettant la génération des digues fictives et simulées est présenté, suivi d'une description des différents aspects du signal de température simulé. Ensuite, les installations expérimentales permettant d'enregistrer des données réelles seront présentées, suivi d'une description de données acquises ainsi que des phénomènes permettant d'influencer ou perturber la mesure recherchée. Les données acquises étant fonction de la distance et du temps, un modèle matriciel de mélange linéaire est proposé pour poser le problème. Il faut mentionner que l'hypothèse de mélange linéaire est vérifiée, parce que nous considérons uniquement les valeurs de températures induites par les différents facteurs sans vouloir quantifier les facteurs qui les ont produits. La dernière partie de ce chapitre est consacrée aux prétraitements utiles pour ce type de données. Dans cette partie nous porterons une attention particulière aux phénomènes éphémères énergétiques en temps (tels que les précipitations) qui influencent un grand nombre de capteurs et peuvent provoquer des fausses alarmes. Nous proposons ainsi un critère de détection basé sur les statistiques d'ordre supérieur (le skewness et le kurtosis) qui permettront d'identifier les périodes temporelles de ces phénomènes éphémères.

Dans le troisième chapitre, la problématique de la détection des fuites sera traitée comme un problème de séparation de sources, où les sources correspondent aux différents facteurs agissant sur les données. Ces différents facteurs sous certaines conditions peuvent être considérés comme indépendants. Une particularité dans notre problème est que les sources les plus utiles à récupérer, liées aux fuites, sont souvent très faibles en termes de contribution énergétique. Aussi, il s'agira de proposer des techniques de séparation prenant en compte cette spécificité. Les techniques de séparation de sources telles que la décomposition en valeur singulières (SVD) et l'analyse en composantes indépendantes (ICA) seront utilisées dans ce travail. L'aspect théorique de ces techniques et leur interprétation pour des données thermométriques sont présentés



dans la première partie de ce chapitre. Un algorithme basé sur ces techniques est ensuite proposé pour la détection des fuites. La dernière partie de ce chapitre sera consacrée à l'application de la méthode proposée sur les trois jeux de données, permettant une validation de la méthode en termes de détecteur de fuites artificielle et réelles.

Le quatrième chapitre est consacré à la détection des zones présentant des anomalies thermométriques ayant un comportement différent au cours d'une journée (par rapport aux zones sans anomalies). Ce principe de détection peut être utile quand nous ne disposons pas d'un grand nombre d'acquisitions ou quand nous voulons analyser la digue sur une seule journée (problématique du temps réel). Les zones singulières comme les structures existantes, les fuites et les singularités du terrain vont avoir des variations journalières de température différentes de celles des zones non-singulières. Une mesure de dissimilarité sera proposée pour quantifier les singularités. La distribution de cette dissimilarité sera ensuite modélisée par un mélange de distributions de lois gamma et uniforme. Une approche basée sur le maximum de vraisemblance permettra d'estimer les paramètres de ce mélange. En fixant une probabilité de fausse alarme, un seuil dépendant du jour analysé peut être ainsi choisi, ce dernier permettant d'éviter des erreurs de détection dues aux conditions climatiques journalières. La méthode sera au final validée sur les données simulées ainsi que sur les données enregistrées sur le site d'Oraison.

# Chapitre 1



# Chapitre 1

## Etat de l'art des techniques de détection des fuites

### Sommaire

---

|            |  |           |
|------------|--|-----------|
| <b>1.1</b> | <b>Introduction</b>  | <b>11</b> |
| <b>1.2</b> | <b>Principaux mécanismes d'endommagement des digues en terre</b>   | <b>11</b> |
| <b>1.3</b> | <b>Diagnostic et surveillance des ouvrages</b>   | <b>14</b> |
| 1.3.1      | Méthodes courantes de diagnostic   | 14        |
| 1.3.2      | Mesures de grandeurs physiques   | 15        |
| 1.3.3      | Méthodes de surveillance   | 15        |
| 1.3.4      | Méthodes non-conventionnelles de surveillance  | 16        |
| <b>1.4</b> | <b>Mesures de température par capteurs à fibre optique</b>   | <b>21</b> |
| 1.4.1      | Capteurs à fibre optique   | 21        |
| 1.4.2      | Capteurs distribués à fibre optique  | 25        |
| 1.4.3      | Téléométrie optique : réflectométrie optique dans le domaine temporel  | 28        |
| 1.4.4      | Mesure de température par rétrodiffusion Raman   | 29        |
| 1.4.5      | Spécification des paramètres et mesures des performances   | 32        |
| 1.4.6      | Avantages de capteurs distribués à fibre optique   | 34        |
| 1.4.7      | Applications des capteurs distribués à fibre optique   | 35        |
| <b>1.5</b> | <b>Principe de détection des fuites en digues par mesures de température à l'aide de capteurs distribués à fibre optique</b> | <b>37</b> |
| <b>1.6</b> | <b>Conclusion</b>  | <b>38</b> |

---



## 1.1 Introduction

Les infrastructures de génie civil qui bordent de nombreux cours d'eau telles que les barrages, les digues en terre, etc sont mises en danger par des phénomènes d'érosion interne et de glissement. Ces phénomènes détériorent progressivement ces infrastructures, des conditions climatiques sévères peuvent accélérer leur vieillissement et leur destruction. De nos jours, les changements climatiques ont de répercussions sur les ressources aquatiques, l'agriculture, sur les écosystèmes naturels, la santé, etc. Ils peuvent provoquer également des dommages matériels au niveau de digues en terre. Le développement de méthodes efficaces pour la surveillance et le diagnostic de ces structures devient alors inéluctable. Par exemple, pour les digues en terre, l'écoulement préférentiel à travers l'ouvrage est un indicateur d'érosion interne. La détection des fuites d'eau le long de l'ouvrage est alors un moyen efficace pour le diagnostic de cette infrastructure. Le but de ce chapitre est de brièvement présenter les principaux mécanismes d'endommagement des digues en terre ainsi que de justifier le choix de la méthode d'auscultation par mesure de température par rapport aux autres méthodes géophysiques généralement utilisées.

## 1.2 Principaux mécanismes d'endommagement des digues en terre

La digue désigne en général un barrage en terre de grande longueur [Fauchard 04]. Les digues servent plusieurs buts selon leur type : protection contre les inondations, irrigation, canaux d'eau pour les centrales hydrauliques, etc. Elles constituent des ouvrages linéaires étendus de plusieurs centaines de mètres à plusieurs kilomètres. Dépendant de leur ancienneté et de l'environnement dans lequel elles se trouvent, les digues sont toujours soumises aux risques de dommages. Ces dommages peuvent affecter les performances présentes ou futures de ces ouvrages avec des conséquences parfois catastrophiques. Il devient alors impératif d'identifier tout comportement anormal le plutôt possible afin de prendre des mesures préventives et/ou correctives.

Les études récentes sur les crues et les ruptures de digues ont relevé plusieurs causes de dédommagement primaires et secondaires comme [Fauchard 04] :

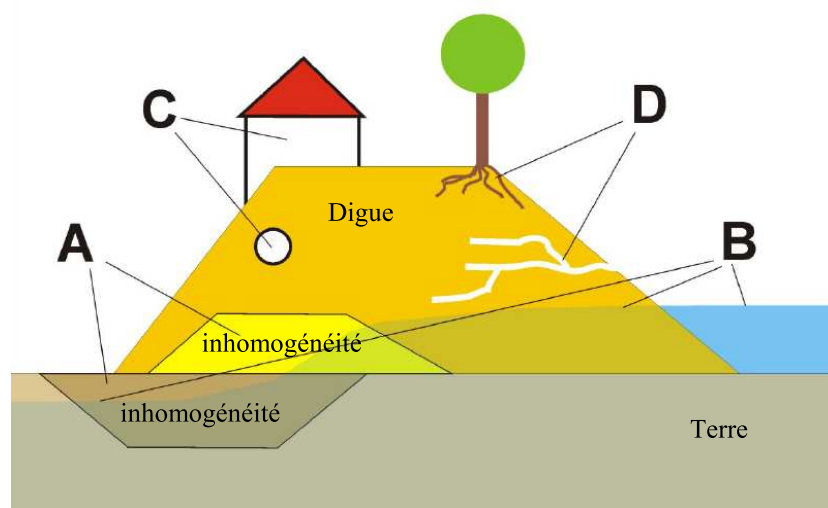
- l'érosion externe de la surface de la digue à cause de débordement de l'eau au-dessus de la digue, ou de son affaiblissement en pied,
- l'érosion interne liée à l'entraînement des particules du sol sous l'effet de l'écoulement interne qui traverse la digue,
- le glissement d'un des deux talus, principalement du talus coté aval à cause souvent de la modification de son équilibre mécanique par la présence de l'eau sous pression,
- les problèmes d'hétérogénéité de matériaux lors de la construction et/ou lors de réparations.

Le schéma de la figure 1.1 montre les quatre catégories principales de sources de problèmes dans les digues [Niederleithinger 07].

- Les problèmes dus à des propriétés hétérogènes des matériaux, de la digue ou de fondation

constituent la catégorie A. Le cisaillement du sol ainsi que la stratification des couches de sol jouent un rôle très important sur la stabilité de la digue. La perméabilité de la digue détermine la vitesse de percolation de l'eau à travers la digue et, par conséquent, sa stabilité.

- La catégorie B comprend l'effet de l'eau qui percole à travers le corps de la digue et cause des instabilités et de l'érosion. Si la surface de la nappe phréatique est artésienne, l'érosion pourrait entraîner une rupture par boulangerie de la digue. Les hautes pressions enregistrées parfois dans les levées sont responsables de la rupture générale. En outre, la vitesse élevée de l'eau de la rivière accompagnée d'un flux direct incident sur la face amont de la digue est une autre cause de rupture par affouillement.
- Les activités humaines autour la digue, c'est-à-dire, l'agriculture, le trafic, les constructions de bâtiments sur la digue ou de tuyaux dans la digue sont classées dans la catégorie C. Ces activités anthropiques touchent principalement la sécurité des digues entraînant des problèmes d'érosion interne. Les constructions sur la digue ou dans son corps sont des moyens préférentiels d'introduction de fort gradient d'écoulement.
- Les digues s'intègrent dans la nature et sont alors influencées par les processus biologiques (catégorie D). Des racines de plantes, d'arbres et les creusements des animaux ont des effets négatifs sur la stabilité de la digue en créant des hétérogénéités (vides, etc.).



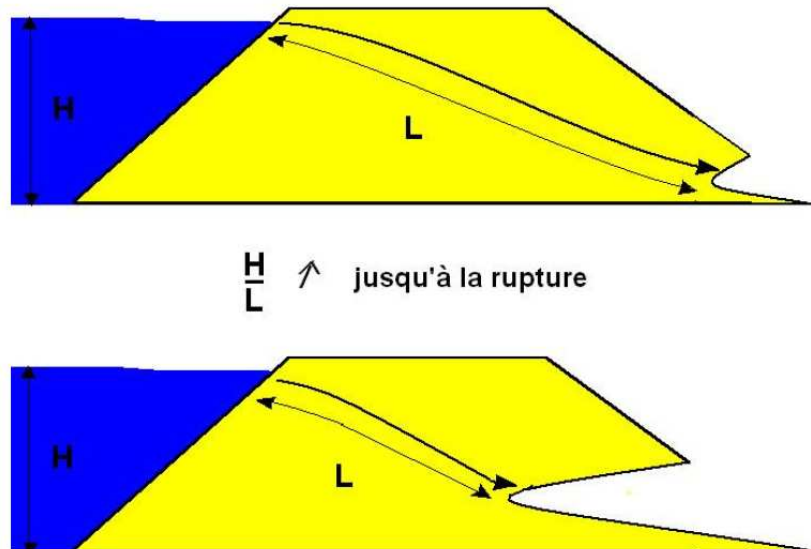
**Fig. 1.1:** Principales catégories de sources de rupture d'une digue liées aux : propriétés du sol (A), propriétés d'écoulement de l'eau (B), activité anthropique (C) et activité biologique (D).

Les brèches sont en général issues d'un mélange de plusieurs facteurs cités ci-dessus et il n'est pas toujours facile d'identifier le ou les facteurs qui les initialisent. Cependant, la plupart d'entre eux sont liées à l'érosion interne [Fry 97, Dornstädter 97, Johansson 97, Fell 07] qui correspond à un écoulement préférentiel dans l'ouvrage.

Le renard hydraulique, ou l'érosion régressive, est l'érosion interne du sol créant progressivement une galerie à travers la digue [Charlot 96]. Avec l'augmentation du niveau d'eau ( $H$ ), le gradient hydraulique ( $H/L$ ) augmente. Le long des lignes de courant préférentiel, un écoulement se crée, générant une petite fuite côté aval de la digue. Les matériaux peu cohésifs du remblai sont entraînés par l'écoulement. La fuite s'agrandit, la cavité s'élargit et se propage vers la rivière. Progressivement, le chemin hydraulique ( $L$ ) se raccourcit, le gradient hydraulique augmente et accentue le phénomène. La galerie ainsi formée peut traverser entièrement l'ouvrage et conduire à sa ruine. La chronologie du phénomène est expliquée dans la figure 1.2.

Pour pouvoir caractériser l'érosion interne d'une digue, nous pouvons nous servir de plusieurs indices parmi les quelles :

- la composition de l'ouvrage,
- la présence des excavations dans le corps ou dans la fondation de la digue,
- la qualité d'étanchéité dans le corps de digue.



**Fig. 1.2:** Érosion régressive : chronologie des évènements [Charlot 96].

Les brèches observées sur le terrain sont souvent réparées à la hâte, pas toujours avec les mêmes matériaux utilisés lors de la construction de la digue, ce qui introduit des hétérogénéités de matériaux. N'ayant toujours pas accès à ces informations, principalement à cause de l'ancienneté de l'ouvrage, il faut utiliser d'autres moyens pour caractériser l'apparition des brèches. Dans ce cadre, l'observation des fuites sur le terrain, leur localisation et leur débit sont des indicateurs potentiels pour caractériser l'état de l'ouvrage. Le but de cette thèse est la détection et la localisation des fuites d'eau le long de la digue. La détection au cours du temps de ces fuites permettra de suivre l'évolution de la fuite au cours des différentes saisons et/ou des années.



Dans la suite de ce chapitre, nous présentons des méthodes de détection des fuites avec leurs avantages et leurs inconvénients.

## 1.3 Diagnostic et surveillance des ouvrages

Le diagnostic et la surveillance sont les deux composants de la maintenance des ouvrages. Le diagnostic constitue la maintenance curative : l'incident est traité après qu'il soit apparu. Au contraire, la surveillance est une maintenance préventive en intervenant avant que l'incident ne se déclare.

### 1.3.1 Méthodes courantes de diagnostic

Le diagnostic d'un ouvrage peut être vu comme la mise en place d'une procédure d'identification des endommagements et/ou des défauts. Il existe plusieurs méthodes de diagnostic. Les méthodes parfois classées par rapport à une dimension spatiale [Fauchard 04] permettent de proposer des :

- méthodes dites locales,
- méthodes dites à grand rendement.

Les méthodes locales, comme leur nom le suggère, caractérisent localement l'ouvrage. Parmi ces méthodes nous pouvons énumérer les méthodes microgravimétriques, acoustiques, sismiques, piézométriques, ou encore traçage chimique [Fauchard 04, Niederleithinger 07]. Les digues étant des ouvrages linéaires étendus sur plusieurs kilomètres, les méthodes locales ne sont pas toujours pratiques et adaptées aux dimensions de l'ouvrage. Les méthodes à grand rendement [Fauchard 04, Morawetz 07, Niederleithinger 07] quant à elles renseignent sur l'ensemble du corps de la digue. Ces méthodes feront le sujet de la discussion présentée par la suite.

Une autre classification est basée sur la nature des méthodes eux mêmes : méthodes destructives et méthodes non-destructives. La reconnaissance géotechnique fait partie des méthodes destructives ou invasives. Elles comprennent essentiellement des forages destructifs, des sondages à la pelle mécanique, des essais pénétrométriques, de perméabilité et de cisaillement [Fauchard 04, Cemagref 05]. Assez naturellement, les méthodes non-destructives sont les plus rapides et les plus économiques parce qu'elles sont exploitables sur l'ouvrage dans son état opérationnel, sans que l'ouvrage en soit affecté. Fauchard et al. [Fauchard 04] présentent un guide des méthodes non-destructives pour le diagnostic des digues sèches (destinées à la protection contre les inondations, qui sont comme leur nom l'indique hors d'eau en dehors des épisodes de crue). Ces méthodes sont également exploitables pour les digues en eau (destinées au transport fluvial, à l'irrigation ou à l'emmenée d'eau sur des centrales hydroélectriques) [Niederleithinger 07, Guidoux 08].

Ces méthodes ont été analysées dans différents ouvrages [Cemagref 05, CSB 05, Bièvre 05, Morawetz 07]. Nous n'allons pas détailler toutes les méthodes, nous allons juste montrer les

notions de base applicables à ces méthodes pour pouvoir les comparer avec la méthode de mesures de température par fibre optique que nous mettons en œuvre.

### 1.3.2 Mesures de grandeurs physiques

Les méthodes de type “prospection géophysique” d’une digue sont basées sur le principe d’extraction et d’analyse des propriétés physiques de la digue. Ces propriétés sont extraites des variations d’une grandeur physique mesurée par des profils le long ou à travers la digue. Les grandeurs à mesurer dépendent des matériaux constituant la digue. Par exemple, le limon et l’argile sont des matériaux utilisés couramment dans le corps d’une digue. Ces deux matériaux facilitent la circulation ou la diffusion de courant grâce à leur caractère conducteur. La conductivité  $\sigma$ , est alors une grandeur qui pourra nous permettre d’accéder et d’identifier la structure interne de la digue. Les méthodes géophysiques mesurant la résistivité  $\rho$ , donneront non-seulement une idée de la composition de la digue mais aussi sur son niveau de saturation. Les méthodes électromagnétiques basses fréquences sont également utilisées avec les méthodes électriques pour retrouver la composition de la digue [Fauchard 04].

D’autres méthodes comme le sondage radar géologique (Ground Penetrating Radar) ou les méthodes électromagnétiques hautes fréquences sont basées sur des mesures de permittivité du sol, mettant en évidence l’aspect polarisation en même temps que la conductivité [Morawetz 07]. Le sens de polarisation peut éventuellement renseigner sur les fractures.

Des méthodes de type sismiques permettent d’identifier les couches de roches altérées et saines par des mesures d’impédance mécanique.

Il convient de mentionner ici que l’utilisation des différentes méthodes dépend de la profondeur d’investigation voulue, ainsi certaines méthodes seront mieux adaptées que d’autres pour un objectif donné. Par exemple, la méthode dite du radar géologique est utilisable pour une investigation des premiers mètres de profondeur à cause de la difficulté de propagation des ondes radar dans les milieux conducteurs des digues.

### 1.3.3 Méthodes de surveillance

La surveillance comprend l’auscultation et l’inspection visuelle. L’auscultation concerne la mesure des différents paramètres qui sont importants pour assurer la sécurité des digues. L’ouvrage [CSB 05] inclut une liste exhaustive des méthodes conventionnelles de diagnostic. Les intervalles de surveillance sont souvent définis par l’institut responsable de la sécurité de l’ouvrage. Les intervalles typiques sont des inspections journalières effectuées par le personnel spécialisé exploitant le site et des inspections plus approfondies toutes les années. Ces inspections sont souvent complétées par des évaluations de sécurité de l’ouvrage effectuées par des experts tous les 4 ans à tous les 15 ans. Les méthodes conventionnelles d’auscultation consistent à mesurer le débit des suintements, la pression interstitielle, les mesures de déformation, etc.

Le suintement par exemple est lié à l’érosion interne et une augmentation rapide de son débit est

un signe de fuite. La pression interstitielle est un indicateur des zones de fuites. Les mesures dans des filtres avals permettent de vérifier l'intégrité de la capacité de drainage des filtres. Même si ces méthodes conventionnelles sont importantes, elles ne sont pas adaptées à l'auscultation automatique de l'ouvrage (car souvent effectuées ponctuellement et manuellement).

### 1.3.4 Méthodes non-conventionnelles de surveillance

Pour la détection des fuites, quelques méthodes sont plus privilégiées que d'autres, plus particulièrement : la méthode de mesure des températures, la méthode de résistivité et la méthode de potentiel spontané.

Ce travail étant basé sur les mesures de température, les autres méthodes seront brièvement discutées dans les sections suivantes afin de pouvoir développer l'intérêt de l'utilisation des mesures de température.

#### La méthode de résistivité

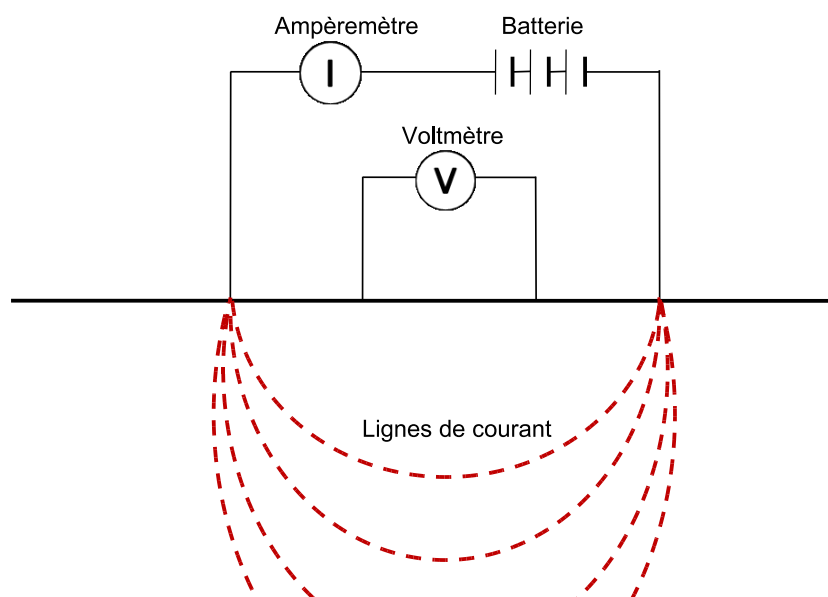
Les mesures de résistivité par courant électrique sont connues dans le domaine du diagnostic de digues et barrages sous le nom de "méthode de résistivité". Les propriétés du sol sont étudiées en termes de résistivité mesurée par sondage électrique. Depuis sa première utilisation par Conrad Schlumberger [Schlumberger 20] pour la prospection électrique, au début du 20<sup>ème</sup> siècle, cette méthode a beaucoup évolué avec les différentes applications [Johansson 96, Johansson 98, Loke 00, Dahlin 01, Sjö Dahl 08].

Le principe de la mesure par quadripôle est basé sur la mesure de la distribution de potentiel après l'introduction d'un courant dans le sol [Fauchard 04, Dahlin 01]. Le schéma sur la figure 1.3 montre une installation typique où une paire d'électrodes introduit le courant alors qu'une deuxième paire d'électrodes mesure la différence de potentiel. Le courant utilisé est un courant continu ou alternatif de très basse fréquence. Le courant mesuré peut être approximé comme une moyenne pondérée des courants de toutes les lignes de courant entre les électrodes. Cette mesure donne une information sur la résistivité moyenne d'un certain volume sous-surface.

En supposant un sol homogène, la profondeur de pénétration sera proportionnelle à la distance entre les électrodes. Cependant, une interprétation directe de la résistivité d'une couche géologique particulière est impossible à partir de ces mesures. Ceci est dû au fait que le courant sera acheminé dans les zones de basse résistivité et dévié par les zones de forte résistivité. Pour pouvoir récupérer les vraies valeurs de résistivité, les techniques d'inversion basées sur des méthodes de différences finies ou des éléments finis sont généralement utilisées [Dahlin 93, Sjö Dahl 06].

Cette méthode de résistivité est assez efficace pour représenter une tranche transversale ou longitudinale de la digue. Les mesures de résistivité permettent alors de voir les hétérogénéités en fonction de la profondeur d'investigation. En général, cette profondeur est de l'ordre de 1/8<sup>e</sup> à 1/6<sup>e</sup> de la longueur du dispositif.

Les différentes configurations possibles avec le schéma présenté sur la figure 1.3 sont [Loke 00,



**Fig. 1.3:** Méthode de résistivité : Principe de prospection électrique en courant continu.

Fauchard 04] :

- le sondage électrique : le centre est fixe par rapport aux déplacements des électrodes pour mesurer la résistivité apparente en fonction de la longueur du dispositif,
- la traînée électrique : les dimensions du dispositif sont fixes avec le déplacement du dispositif le long d'un profil,
- le panneau électrique : plusieurs électrodes sont implantées le long d'un profil rectiligne (acquisitions 2D) ou sur une grille (acquisitions 3D). Géré par ordinateur, le système comporte l'appareil de mesure de résistivité, le relais de commutation et les électrodes.

Parmi ces configurations, le panneau électrique est la plus intéressante avec une meilleure utilisation des électrodes servant alternativement comme électrode d'injection et de mesure de potentiel. Le résultat obtenu après l'application des techniques d'inversion des données est une carte de résistivité. Pour un rendement plus grand, une configuration dite "Towed Array" peut être adoptée, configuration dans laquelle le panneau peut être traîné derrière un véhicule. Ceci permet une auscultation sur plusieurs kilomètres par jour.

Pour le diagnostic des digues, l'idée fondamentale est l'existence d'un couplage physique entre la résistivité mesurée et l'infiltration (la fuite). Les deux cas principaux sont [Sjödahl 06] :

1. l'écoulement et l'arrachement des particules du sol à cause de l'érosion interne aura une incidence sur la résistivité dans l'ouvrage,
2. les variations saisonnières de la résistivité dans le réservoir, issues des variations de la température, auront un effet sur la résistivité quand l'eau d'infiltration traverse la digue.

### La méthode de potentiel spontané

La deuxième méthode couramment utilisée pour le diagnostic des ouvrages hydrauliques est celle du potentiel spontané (PS), en anglais Self-Potential Method [Ogilvy 69, Corwin 90, Revil 06]. Elle consiste à mesurer la distribution statique du potentiel électrique à la surface d'un ouvrage ou dans les forages. Le but est d'avoir une carte de potentiel électrique afin d'identifier les mécanismes de polarisation dans le sol. Les deux contributions les plus importantes aux signaux mesurés par la méthode de potentiel spontané sont :

1. le potentiel d'écoulement (streaming potential) ou le potentiel électrocinétique,
2. les processus électrochimiques associés aux gradients de potentiels chimiques des espèces ioniques dans l'eau.

C'est principalement le premier cas qui nous intéresse, où le potentiel spontané correspond au champ électrique induit par le mouvement des fluides dans les milieux poreux. Plus particulièrement, lors de son écoulement dans un système capillaire, l'eau collecte et transporte les ions positifs des matériaux environnants. Ces ions s'accumulant à la sortie du capillaire créent une nette charge positive. Au contraire, les ions négatifs non-transportés s'accumulent à l'entrée du capillaire résultant à une nette charge négative. Si les potentiels électrocinétiques ainsi développés ont des amplitudes suffisantes à mesurer, les points d'entrées et de sorties des zones de concentration d'infiltration peuvent être déterminés grâce aux anomalies négatives et positives de potentiels spontanés.

Les variations du potentiel électrocinétique,  $V_{ec}$  sont gouvernées par l'équation [Revil 03] :

$$V_{ec} = \frac{\rho_f \epsilon_f \zeta}{4\pi \mu_s} \Delta P, \quad (1.1)$$

où

$\rho_f$  est la résistivité du fluide et de son soluté éventuel,

$\epsilon_f$  est la constante diélectrique,

$\zeta$  est la différence de potentiel électrique à travers la double couche de Helmholtz établie à l'interface entre les capillaires d'un milieu et le fluide y circulant,

$\mu_s$  est la viscosité dynamique de la solution,

et  $\Delta P$  est la différence de pression le long du chemin parcouru par le fluide.

Les mesures sont réalisées avec des électrodes impolarisables enfouies à quelques centimètres de profondeur. Ce type d'électrodes permet d'éviter les potentiels propres d'électrodes qui peuvent masquer les potentiels spontanés [Lu 98]. Le fait d'être enfouies les rend moins sensible au bruit électromagnétique externe et aux variations diurnes de température qui peuvent faire dériver les tensions mesurées de 0.1 à 0.2 mV par degré Celsius. L'installation basique consiste à fixer une électrode et à déplacer l'autre électrode aux intervalles présélectionnés le long d'un profil. Dans les zones où le potentiel spontané dépasse un seuil (typiquement 10 mV), des mesures plus

détaillées sont effectuées avec des intervalles plus petits pour mieux identifier les zones d'anomalies potentielles. Les magnitudes et les signes des potentiels dépendent du processus interne. Par exemple, une fuite souterraine générera une anomalie de polarisation négative, un suintement une anomalie positive.

Comme pour la méthode de résistivité, la méthode de potentiel spontané peut être utilisée pour les investigations simples ou en mode auscultation avec plusieurs électrodes distribuées sur une grille. Cette dernière permet de réaliser une cartographie du potentiel en surface. Cependant, les mesures ne sont pas directement interprétables et doivent être combinées avec d'autres mesures ou des méthodes complémentaires pour un diagnostic précis. Plusieurs travaux ont déjà utilisé la méthode de potentiel spontané pour différentes applications hydrologiques [Titov 00, Panthulu 01, Rozycki 05, Revil 06]. Récemment, certains auteurs ont essayé d'estimer les débits des écoulements soit en utilisant les modélisations numériques soit en couplant ces mesures avec des mesures données par d'autres méthodes [Perriera 99, Sheffer 04, Bolève 09].

### **Les méthodes électromagnétiques basses fréquences**

Les méthodes électromagnétiques basses fréquences peuvent être de deux types : en champ proche (Slingram) et champ lointain (RMT). La méthode en champ proche consiste à capter le champ magnétique induit par des zones de plus ou moins grande conductivité sous l'excitation d'un champ magnétique basse fréquence. Le champ d'excitation est produit par une bobine (une boucle de courant) située à une distance fixe du récepteur. La longueur de propagation de l'onde est très petite par rapport à longueur d'onde du champ magnétique (d'où son nom champ proche). La conductivité apparente du sol est fonction du rapport entre les composantes quadratiques du champ secondaire et du champ primaire et la profondeur d'investigation dépendra de la distance entre les deux bobines.

La méthode radio magnétotellurique (RMT) considère que les émetteurs des ondes électromagnétiques sont si loin que les ondes arrivant sur les récepteurs sont planes. Les ondes primaires induisent des courants dans le sol, d'où la production d'un champ secondaire qui est modifié ensuite par les hétérogénéités du sol. Cette fois-ci le rapport entre la composante horizontale du champ électrique et la composante verticale du champ magnétique renseigne sur la résistivité apparente (leur déphasage révélant la présence des anomalies conductrices). Les électrodes capacitatives utilisées pour cette méthode ont une tendance à se charger en fonction du temps et elles sont plus sensibles à la variation de la rugosité du sol. Une autre limitation de cette méthode est qu'elle est trop sensible aux hétérogénéités de surface.

### **La méthode du radar géologique**

Un signal électromagnétique émis par une antenne se propage à travers la subsurface avec une vitesse dépendant de la permittivité relative des matériaux constituant le sol. A l'interface entre deux matériaux, une partie du signal est réfléchi. La mesure de ce signal par une deuxième antenne permet de localiser les zones d'anomalies [Fauchard 04]. La résolution verticale et la

profondeur de pénétration dépendent de la fréquence d'émission et des conditions de subsurfaces. La limitation principale de cette technique est que la perte du signal dans le milieu est proportionnelle à la conductivité du milieu investigué.

Le principe des méthodes sismiques de réfraction étant assez proche de celui des méthodes radar géologique, et ne permettant pas la détection des fuites, nous ne le présenterons pas dans ce rapport.

### **Les méthodes dans le domaine de l'InfraRouge**

Des pistolets et des caméras InfraRouge mettent en évidence des contrastes de température de surface respectivement en 1D et 2D [Niederleithinger 07]. Ces méthodes ne sont pas encore trop développées pour leur exploitation dans le diagnostic des digues. L'avantage associé est bien sûr la rapidité de la prise de mesures soit par une voiture soit par un avion tout en couvrant des distances assez longues. Cependant, ces méthodes ont deux inconvénients principaux : premièrement, comme elles prennent des mesures de surface, les phénomènes intervenant en profondeur seront détectés seulement s'ils sont relativement importants et, deuxièmement, la nature de surface de la digue pourra influencer les mesures.

Pour le diagnostic des digues en eau, la méthode de résistivité en utilisant des panneaux électriques et la méthode de potentiel spontané ont été utilisées dans plusieurs installations avec succès [Johansson 98, Bièvre 05, Niederleithinger 07, Morawetz 07, Bolève 09]. Cependant, la principale limitation de ces méthodes est qu'elles restent manuelles, c'est-à-dire, qu'elles nécessitent une intervention humaine pour la collecte des mesures. En outre, la prudence requise dans l'emplacement des capteurs et d'autres équipements électroniques sur le terrain rend ces méthodes loin d'être automatiques.

Alors que les grandeurs de type résistivité et perméabilité sont reliées implicitement avec les phénomènes d'érosion interne et avec les fuites qui en résultent, la température a un lien plus direct avec l'érosion [Johansson 97]. Initialement utilisées par les géothermiciens, les méthodes thermométriques ont été utilisées depuis plusieurs années pour le diagnostic d'ouvrages hydrauliques [Kappelmeyer 57, Armbruster 89, Johansson 91, Johansson 97, Dornstädter 97, Fry 04] surtout en Allemagne, Suède et France. Ces études ont montré la sensibilité et la fiabilité des mesures thermométriques pour l'identification des écoulements préférentiels dans les barrages. Leur application pour le diagnostic des digues est alors naturelle [Fry 04].

Les digues sont des grands linéaires et pour leur surveillance, la notion de coût devient très importante. Dans ce cadre, les méthodes de diagnostic basées sur des mesures de température par fibre optique présentent des solutions économiquement efficaces. Le grand intérêt de la thermométrie par la fibre optique est qu'elle est une technique d'auscultation tout en assurant une inspection quasi visuelle du linéaire de l'ouvrage.

## 1.4 Mesures de température par capteurs à fibre optique

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 [Johansson 04, Fry 04]. Les mesures de température sont basées sur le transport thermique. 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 d'advection. En résumé, la thermométrie est un outil naturel pour la détection des fuites en captant les modifications apportées par l'advection sur le champ de température par rapport à la conduction. En cas de fuite à haut débit, c'est la température de l'eau du canal qui détermine complètement la température captée.

Jusqu'à présent, la plupart des mesures ont été faites dans des forages en utilisant des piézomètres. Il a proposé une méthode d'estimation de débits des fuites basée sur les modèles physiques d'échange thermique. Guidoux [Guidoux 08] a récemment évoqué les limitations de cette méthode en proposant la modélisation numérique des digues avec la méthode des éléments finis utilisant des conditions hydrauliques et thermiques. Dans le cadre de ce travail, nous nous focalisons sur l'aspect détection des fuites et non sur la détermination des débits. Pour plus de détails sur la détermination des débits le lecteur pourra consulter les références énumérées ci-dessus.

Dans les systèmes d'auscultation continue et pour bien couvrir toute la longueur de la digue, la prise de mesure dans des piézomètres n'est pas facilement envisageable à cause de problèmes dus à l'échantillonnage spatial local et à l'échantillonnage temporel variable. Les capteurs distribués de température (DTS pour Distributed Temperature Sensors) basés sur des fibres optiques permettent de répondre à ces exigences.

### 1.4.1 Capteurs à fibre optique

L'industrie de l'optoélectronique et celle des communications à fibre optique ont eu une croissance importante lors des dernières décennies. L'industrie optoélectronique a apporté différents produits comme les lecteurs de disques optiques, les imprimantes lasers, les pointeurs lasers, les lecteurs de codes à barres, etc. L'industrie de communications à fibre optique a révolutionné le secteur des télécommunications en fournissant des liaisons plus performantes et plus fiables tout en étant moins chères. Parallèlement à ces développements, la technologie de capteurs à fibre optique a très bien exploité les technologies associées à l'optoélectronique et à l'industrie de communications à fibre optique. L'amélioration des performances et la réduction de prix des composants ont renforcé la capacité des capteurs à fibre optique à remplacer progressivement les capteurs traditionnels de rotation, de température, de contrainte, de vibration, de pression, de mesures chimiques, etc. dans de nombreux domaines d'application. Au début, le marché visé par les capteurs à fibre optique était orienté dans des secteurs où l'utilisation des capteurs était très peu ou complètement non-existante. Les avantages des capteurs à fibre optique tels que



[Udd 06] :

1. leur poids légers et leur petite taille,
2. leur passivité et résistance à l'interférence électromagnétique,
3. leur haute sensibilité,
4. la bande passante,
5. la robustesse à l'environnement,

ont été utilisés pour contre balancer leurs désavantages principaux à savoir le coût élevé de fabrication et la méconnaissance des utilisateurs finaux. Cependant, la décroissance rapide de coûts des composants électroniques durant ces dernières années a ouvert considérablement le marché et le potentiel de ces outils pour une utilisation croissante et massive dans de nombreux domaines.

### **Généralités sur la fibre optique**

Une fibre optique est un guide d'onde optique qui sert à guider l'onde lumineuse le long de son axe [DeCusatis 06]. Elle est constituée des couches diélectriques en verre ou en plastique avec différents indices de réfraction pour confiner la lumière au voisinage du centre. Dans sa plus simple construction, elle a une géométrie cylindrique constituée de deux éléments principaux :

#### ***Le cœur :***

Un matériel diélectrique (silice) d'indice de réfraction  $n_c$  dans lequel se propage principalement l'onde optique,

#### ***La gaine optique :***

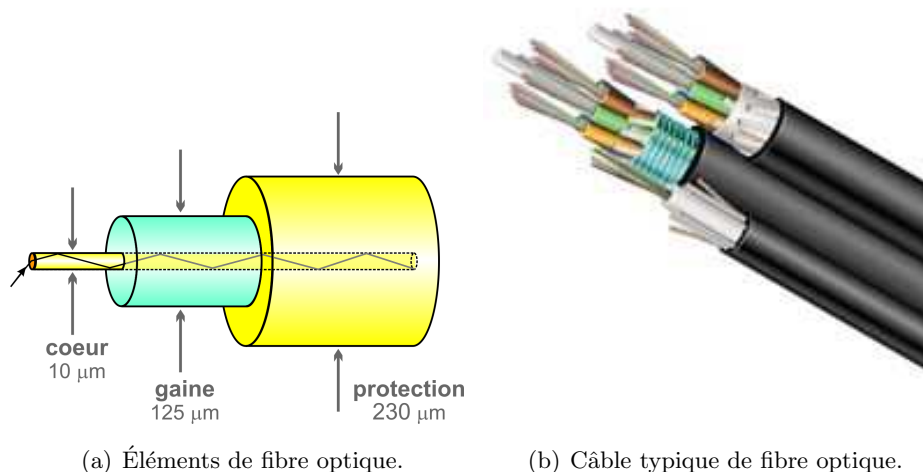
Un milieu qui entoure le cœur avec l'indice de réfraction  $n_g$  inférieur à  $n_c$  et qui peut porter une partie de l'onde optique. La différence des indices de réfraction assure que la plus grande partie de la lumière est confinée au cœur.

Ces deux couches sont généralement protégées par une enveloppe protectrice qui assure la robustesse de la fibre en lui donnant une haute résistance mécanique (cf. figure 1.4(a)).

Nous distinguons deux types de fibres selon la dimension du cœur : les fibres multimodes et les fibres monomodes.

*Les fibres multimodes* ont un cœur de diamètre de l'ordre de quelques dizaines de micromètres à quelques millimètres. Comme ils existent plusieurs modes de propagation pour l'onde optique ceci se traduit par différents parcours et différentes vitesses de phase provoquant ainsi une distorsion de phase. Les fibres multimodes ont la capacité de transmettre plus de puissance.

*Les fibres monomodes* ont un cœur de diamètre de l'ordre de  $10\mu\text{m}$ . Il existe qu'un seul mode de propagation ce qui évite la distorsion de phase. Ce sont des fibres standard de télécommunication



**Fig. 1.4:** Fibre optique en général.

avec une capacité de transfert en puissance plus petite que les fibres multimodes.

Dans le plupart des fibres optiques, le cœur et la gaine optique sont en silice. L'enveloppe protectrice est souvent composée d'un polymère. Les paramètres qualitatifs importants d'une fibre optique sont les pertes et l'atténuation de l'onde lumineuse dans la fibre. L'atténuation caractérise l'affaiblissement du signal au cours de la propagation le long de la fibre et souvent suit une forme exponentielle avec la distance. L'atténuation dépend de la longueur d'onde et est liée aux phénomènes de diffusion et d'absorption. Les fibres en silice offrent un minimum d'atténuation vers une longueur d'onde de 1550 nm. Les autres sources de pertes dans la fibre sont : les courbures et les micro-courbures de la fibre où la condition de réflexion totale risque de ne pas être satisfaite résultant ainsi en une perte dans la gaine optique par simple réfraction ; la diffusion, la réflexion aux épissures, etc. Dans la plupart des cas, les fibres utilisées pour les capteurs sont des fibres de télécommunication, cependant quelques producteurs ont commencé à réaliser des fibres plus adaptées aux applications de capteurs. Un câble typique de fibre optique est montré dans la figure 1.4(b).

### Classification générale des capteurs à fibre optique

Les capteurs à fibre optique sont généralement classés dans deux grandes classes : *extrinsèque* ou *hybride* et *intrinsèque* ou *all-fiber* [Udd 06].

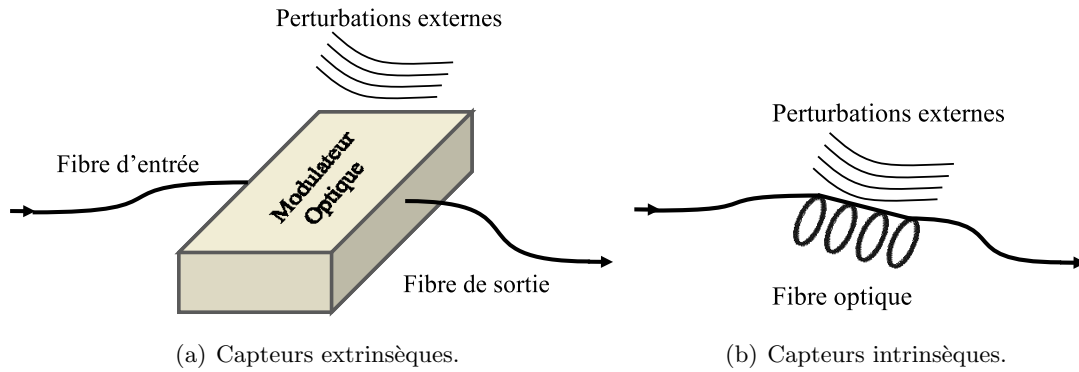
#### *Capteurs extrinsèques*

La figure 1.5(a) montre le principe des capteurs extrinsèques, où la fibre reçoit la lumière à travers une "boîte noire" qui traduit l'information sur la lumière en fonction des effets environnementaux. Cette information peut être sous la forme d'intensité, de phase, de fréquence, de polarisation, de contenu spectral, etc. La fibre optique transmet la lumière avec cette information à un processeur

optique et/ou électronique. La fibre ne joue pas, elle même, le rôle de capteur mais fait partie d'un système de capteurs.

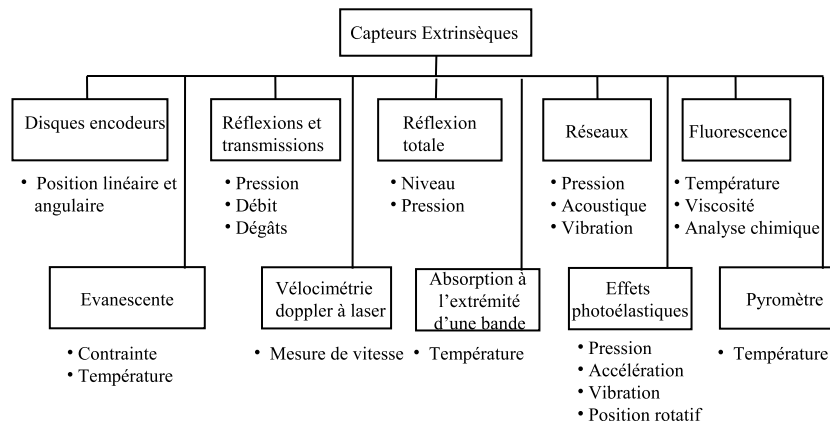
*Capteurs intrinsèques*

Dans cette configuration montrée sur la figure 1.5(b), la fibre, elle même, est le capteur. La fibre transmet toujours la lumière mais cette fois ci les facteurs externes influencent la lumière dans la fibre.

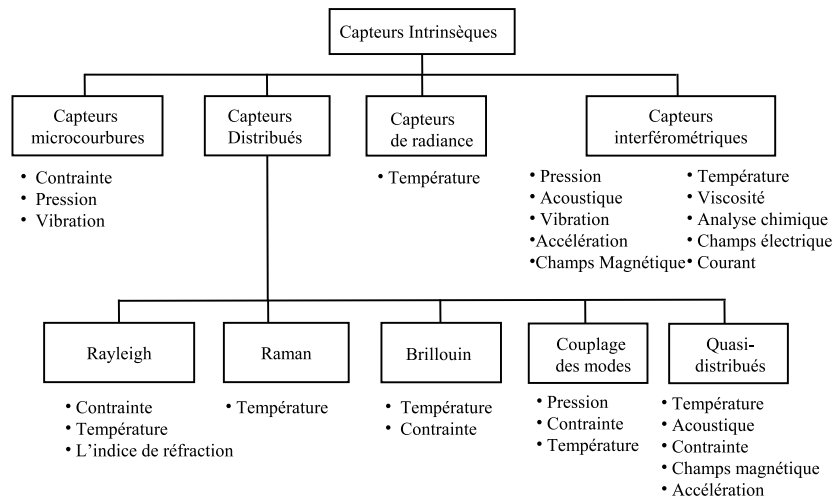


**Fig. 1.5:** Classification générale des capteurs à fibre optique selon le rôle de la fibre optique.

Pour chacune de ces classes, il existe plusieurs sous-classes. Sans rentrer dans le détail, les différentes classes sont résumées sur les figures 1.6-1.7 [Udd 91]. Dans le contexte de mesures de températures par fibre optique le long d'une digue, nous allons détailler par la suite la sous-classe dite de capteurs distribués.



**Fig. 1.6:** Classification des capteurs extrinsèques et les différentes grandeurs mesurées.



**Fig. 1.7:** Classification des capteurs intrinsèques et les différentes grandeurs mesurées.

### 1.4.2 Capteurs distribués à fibre optique

Un avantage important des capteurs à fibre optique est leur capacité d'intégrer dans une seule fibre plusieurs capteurs passifs ou régions de détection, ce qui les rend économiquement viables. Plus particulièrement, ils permettent non seulement l'auscultation de l'ampleur d'un paramètre physique mais également de suivre sa variation le long d'une fibre continue grâce à la possibilité de multiplexage [Dakin 90]. Le résultat consiste en une détection dite distribuée qui signifie que ces capteurs sont plus efficaces économiquement en ouvrant la porte à l'auscultation de la fatigue des structures critiques de génie civil.

#### Multiplexage des capteurs

Le multiplexage est l'incorporation de plusieurs capteurs dans une seule fibre optique. Il existe différents types de multiplexage, ce qui contrôle le nombre de capteurs à incorporer, la réponse fréquentielle de capteurs, l'étendue dynamique, etc.

En général, les techniques de multiplexage sont classées de la manière suivante :

1. par répartition dans le temps (TDM, Time division multiplexing)
2. par répartition en fréquence (FDM, Frequency division multiplexing)
3. par répartition en longueur d'onde (WDM, Wavelength division multiplexing)
4. par répartition en code (CDM, Code division multiplexing)
5. spatial (SDM, Space division multiplexing)
6. hybride (Hybrid Multiplexing)

***Multiplexage par répartition dans le temps :***

Ce type de multiplexage utilise une source de lumière pulsée qui est envoyée dans la fibre optique. Une analyse des temps de retard permet de distinguer entre les réponses des différents capteurs. Cette technique est souvent utilisée pour les capteurs distribués mesurant la contrainte, la température, etc. Comme nous allons voir plus tard, les phénomènes de diffusion de type Rayleigh, Raman et Brillouin sont souvent exploités dans cette configuration pour effectuer différentes mesures physiques.

***Multiplexage par répartition en fréquence :***

L'information de chaque capteur est attribuée à un intervalle particulier de fréquences, c'est-à-dire les données acquises par les capteurs sont codées à des fréquences porteuses différentes. Un exemple est la modulation linéaire d'une diode laser par une source de courant en dents de scie.

***Multiplexage par répartition en longueur d'onde :***

L'information est attribuée à une longueur d'onde particulière, c'est-à-dire, les données sont codées sur les porteurs optiques de différentes longueurs d'ondes. Cette méthode permet l'utilisation très efficace du flux énergétique optique.

***Multiplexage par répartition en code :***

L'information est codée différentiellement sur les composants de la porteuse optique selon leur cohérence mutuelle par rapport à une porteuse de référence. Cette technique est proche de celle de spectre à étalement.

***Multiplexage spatial :***

Cette approche est basée sur l'utilisation de plusieurs fibres optiques pour communiquer avec différents éléments de détection. Cette technique a plusieurs désavantages importants comme, par exemple, un coût élevé de câblage et de connecteurs.

***Le multiplexage hybride :***

Comme son nom l'indique, ce multiplexage combine différentes techniques dans le but d'incorporer les avantages de chacune.

Le choix d'une technique particulière de multiplexage est conduit par divers facteurs en fonction de chaque application.

**Principe de mesure des capteurs distribués à fibre optique**

Nous avons vu que les deux types de capteurs distribués à fibre optique sont : (1) les capteurs intrinsèquement distribués (intrinsic distributed fiber optic sensors) et (2) les capteurs

quasi-distribués [Hartog 00, Rogers 02]. Les capteurs intrinsèquement distribués sont capables de suivre une mesure d'une façon continue le long de la fibre. Dans le cas contraire, quand il est difficile de faire la détection complètement distribuée, les capteurs quasi-distribués sont utilisés pour la surveillance des positions définies par multiplexage des capteurs non-distribués à fibres optiques (point sensors). Dans les deux cas, les variations des signaux optiques se propageant le long la fibre sont observées. En général, un champ optique monochromatique,  $\mathbf{E}(\mathbf{r}, t)$ , peut être écrit comme [Yin 00] :

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{A}(\mathbf{r}, t)e^{i(\omega t + \phi(\mathbf{r}, t))} \quad (1.2)$$

avec  $\mathbf{A}(\mathbf{r}, t)$  l'amplitude du champ optique complexe,  $\omega$  la pulsation et  $\phi(\mathbf{r}, t)$  la phase du champ complexe. Il faut remarquer ici qu'une source à large bande peut être considérée comme une sommation des plusieurs sources monochromatiques des fréquences différentes. Les différents paramètres d'un champ lumineux sont alors :

- l'amplitude ou l'intensité,  $\mathbf{I}(\mathbf{r}, t) = |\mathbf{A}(\mathbf{r}, t)|^2$ ,
- la fréquence ou longueur d'onde de la lumière,
- la polarisation,
- la phase  $\phi(\mathbf{r}, t)$ .

Un ou plusieurs paramètres peuvent être sensibles aux perturbations extérieures. En mesurant alors ces paramètres ainsi que leurs changements, les perturbations extérieures peuvent être détectées. Une classification des capteurs à fibre optiques basée sur la capacité à exploiter le changement de ces paramètres peut être donnée :

1. Capteurs basés sur l'intensité,
2. Capteurs basés sur la polarisation,
3. Capteurs basés sur la fréquence ou la longueur d'onde,
4. Capteurs basés sur la phase.

Des exemples de différents mécanismes qui peuvent apporter un changement d'intensité sont : la perte de la microcourbure, la rupture, le couplage entre les fibres, le facteur de réflexion, l'absorption, l'atténuation, la diffusion moléculaire. L'indice de réfraction d'une fibre peut changer sous l'application d'une contrainte ou d'une déformation. Cette contrainte ou cette déformation peuvent varier par rapport à la direction de la fibre optique, ce qui induit un déphasage entre les différentes directions de polarisation. En suivant alors les variations de polarisation de la lumière, les différentes perturbations peuvent être identifiées. La phase de la lumière dépend du chemin parcouru par la lumière, un petit changement dans ce chemin résulte en un changement significatif dans la phase, ce qui rend le suivi de la phase très intéressant. Cependant, pour l'implémentation des capteurs basés sur la phase, l'utilisation des techniques d'interférométrie est nécessaire.

De la même manière, les perturbations extérieures peuvent jouer sur la fréquence de la lumière émise. Une des perturbations extérieures qui peut être mise en évidence avec une très bonne

résolution est la température.

### 1.4.3 Télémétrie optique : réflectométrie optique dans le domaine temporel

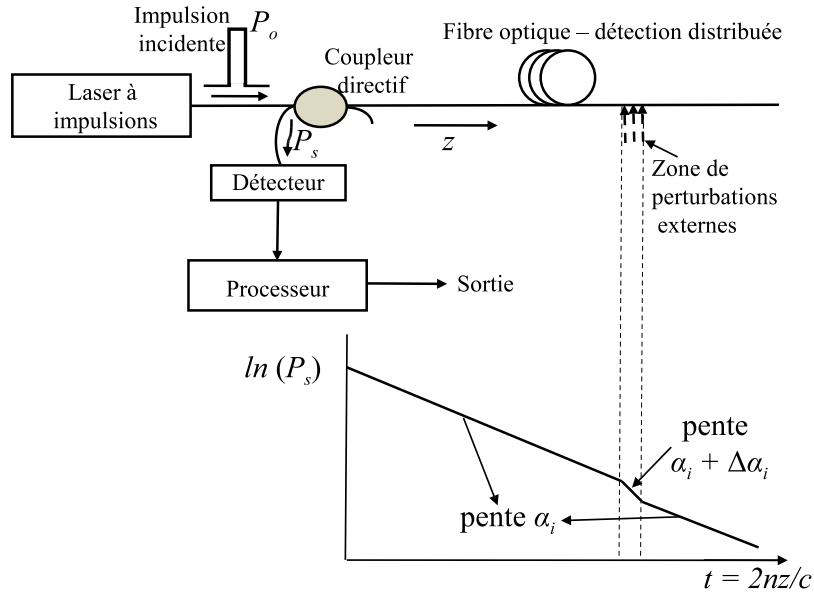
Les techniques de détection de type télémétrie optique sont l'équivalent des techniques radar dans le domaine optique. Différentes configurations de la réflectométrie existent pour les fibres optiques :

1. réflectométrie optique dans le domaine temporel (OTDR pour Optical Time Domain Reflectometry en anglais),
2. réflectométrie optique dans le domaine fréquentielle (OFDR pour Optical Frequency Domain Reflectometry en anglais),
3. la réflectométrie optique cohérente (OCDR pour Optical Coherence Domain Reflectometry en anglais).

La technique la plus connue dans le contexte des systèmes commerciaux étant l'OTDR, nous allons détailler les principes de cette technique. Le lecteur pourra se référer aux ouvrages [Sorin 92, Yu 94, Wegmüller 01] pour les autres techniques.

La technique OTDR exploite la rétrodiffusion de la même façon qu'un radar, lidar ou un sonar. La lumière envoyée dans la fibre rencontre des pertes en raison de la diffusion de Rayleigh issue des variations microscopiques aléatoires dans l'indice de réfraction du cœur de la fibre. Une partie de cette lumière est diffusée dans le contre sens de la propagation de la lumière incidente. Elle est reprise par l'ouverture (aperture) de la fibre et renvoyée vers la source. En surveillant l'intensité du signal Rayleigh rétrodiffusé en réponse d'une impulsion étroite envoyée dans la fibre, les variations spatiales dans le coefficient de diffusion de la fibre et/ou l'atténuation peuvent être déterminées. Le coefficient de diffusion à une position donnée révèle le statut local de la fibre, ce qui permet de localiser les perturbations. Le fait que cette technique est basée sur la détection de l'intensité du signal rétrodiffusé en fonction du temps lui donne son appellation de *réflectométrie optique dans le domaine temporel*. Cette technique est très courante pour la détection des fautes/imperfections et pour le diagnostic dans les applications de télécommunication à fibre optique. Son utilisation dans le domaine des capteurs peut alors être très efficace pour détecter les variations locales du coefficient de diffusion induites par des perturbations extérieures.

La méthode OTDR, développée par Barnoski et al [Barnoski 78], a donné naissance au premier type de capteur distribué à fibre optique. Le schéma de la figure 1.8 montre la configuration de base pour l'OTDR. Un laser génère une impulsion de courte durée. Ce dispositif 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 niveau du signal rétrodiffusé est mesuré en fonction du temps par rapport à l'impulsion incidente. Dans le cas idéal où la fibre homogène est soumise à un environnement uniforme, l'intensité décroît exponentiellement avec le temps à cause des pertes intrinsèques de la fibre. Cependant, si à une position donnée il existe un environnement non-uniforme, l'intensité rétrodiffusée ne va plus suivre la loi exponentielle.



**Fig. 1.8:** La configuration de mesure par réflectométrie optique dans le domaine temporel OTDR.

L'intensité de la lumière rétrodiffusée,  $P_s(t)$ , arrivant sur le détecteur après un temps  $t$  peut être écrite comme :

$$P_s(t) = P_o r(z) \exp \left\{ - \int_0^z 2\alpha(z) dz \right\} \quad (1.3)$$

avec  $P_o$  une constante dépendant de l'énergie de l'impulsion incidente et du rapport de division de puissance du coupleur (fiber coupler power splitting ratio en anglais),  $r(z)$  le coefficient effectif de rétrodiffusion par unité de longueur prenant en compte le coefficient de diffusion de Rayleigh et l'ouverture numérique de la fibre,  $\alpha(z)$  le coefficient d'atténuation de la fibre et  $z$  la distance par rapport au début de la fibre donnée par l'équation :

$$z = ct/2n \quad (1.4)$$

avec  $c = 2,99792 \times 10^8 \text{ m/s}$  la vitesse de la lumière dans le vide et  $n$  l'indice de réfraction de groupe. La pente de  $\ln P_s$  est proportionnelle à  $\alpha(z)$ . En cas d'anomalie, la pente de  $\ln P_s$  augmente proportionnellement à  $\alpha(z)$ .

En 1983, Hartog [Hartog 83] a été le premier à mettre en évidence la liaison entre la mesure de température et le changement du coefficient de diffusion de Rayleigh, idée qui sera développée dans la section suivante.

#### 1.4.4 Mesure de température par rétrodiffusion Raman

La diffusion de Rayleigh est causée par des fluctuations dans la composition des matériaux lors de la phase de construction et elle reste relativement indépendante de la température ambiante. Cependant, la diffusion de Raman est sensible à la température ambiante et les premiers capteurs



distribués de température ont été proposés dans les années 1980 [Hartog 83], ce qui a conduit au développement des systèmes commerciaux. Le premier système commercial qui s'appelait DART pour Distributed Anti-Stokes Raman Thermometry était le résultat d'un brevet déposé par Dakin en 1985 [Dakin 85].

La diffusion de Raman représente une diffusion inélastique des photons. 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 longueurs d'onde  $\lambda_s$  respectivement  $\lambda_a$  autour de la longueur d'onde incidente,  $\lambda_0$ . Le spectre de la lumière rétrodiffusée (voir la figure 1.9) montre ces composantes qui sont équidistantes par rapport à la longueur d'onde incidente  $\lambda_0$ . La diffusion de Rayleigh se situe à la longueur d'onde incidente. L'intensité de la composante Anti-Stokes est fortement corrélée à la température alors que celle de la composante Stokes n'en dépend pas. Le rapport  $R(T)$  des intensités des composante Anti-Stokes et Stokes dépend alors de la température. Ce rapport est donné par :

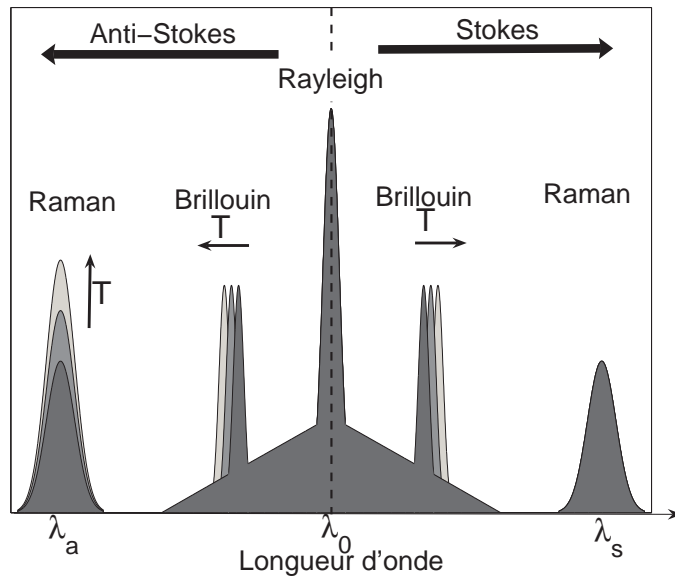
$$R(T) = \left( \frac{\lambda_s}{\lambda_a} \right)^4 \exp \left( - \frac{h\nu_0}{kT} \right), \quad (1.5)$$

où  $h = 6,626068 \times 10^{-34} m^2 kg/s$  est la constante de Planck,  $\nu_0 = \frac{c}{\lambda_0}$  est la fréquence de l'onde incidente (de longueur d'onde  $\lambda_0$ ),  $k = 1,3806503 \times 10^{-23} m^2 kg s^{-2} K^{-1}$  la constante de Boltzmann et  $T$  est la température absolue.

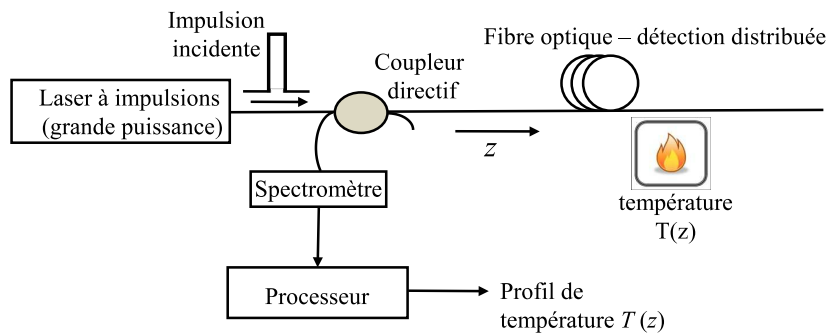
Le système OTDR basé sur la diffusion Raman est montré sur la figure 1.10. Ce système, assez similaire au système de Rayleigh, réalise de plus un filtrage spectral pour séparer les composantes Anti-Stokes et Stokes du spectre Raman. Les intensités et les fréquences des composantes sont enregistrées en fonction du temps pour calculer le rapport  $R(T)$  à partir duquel nous pouvons calculer la température  $T(t)$  en fonction du temps  $t$ . En utilisant la relation entre la distance  $z$  et le temps  $t$  (voir l'équation (1.4)) nous obtenons la température en fonction du déplacement  $z$  le long la fibre.

Un problème de la diffusion Raman est la faible amplitude du signal rétrodiffusé. Ceci exige l'utilisation d'un laser puissant à l'entrée. Les fibres multimodes permettant de propager une puissance plus grande que les fibres monomodes seront privilégiées. Cependant, les effets d'atténuation sont plus significatifs pour les fibres multimodes, ce qui limite la longueur d'auscultation. Les systèmes OTDR standard basé sur la diffusion Raman sont capables d'ausculter jusqu'à 10 km avec une résolution en température de  $0.01^\circ C$  et une résolution spatiale de 1 m.

Il paraît évident ici de dire quelques mots sur les autres composantes du spectre de la figure 1.9. C'est le cas de la diffusion Brillouin qui résulte de l'interaction entre le signal se propageant le long de la fibre et les ondes acoustiques thermiquement excitées. Le phénomène peut être envisagé comme la diffraction de la lumière par un réseau dynamique créée par une onde acoustique. Cette onde acoustique en fait simule une modulation de l'indice de réfraction par effet élasto-acoustique. La lumière diffractée suit un effet doppler à cause du fait que le réseau propage les ondes à la vitesse acoustique. Cette vitesse dépend de la densité du milieu qui elle



**Fig. 1.9:** Le schéma représentatif du spectre de la lumière rétrodiffusée à une position quelconque. La diffusion Rayleigh à la longueur d'onde incidente  $\lambda_0$  n'a aucune dépendance de la température. Le spectre Raman consiste en deux composantes aux longueurs d'onde  $\lambda_s$  et  $\lambda_a$ ; Un changement local de la température de la fibre fait varier l'amplitude de la composante Anti-Stokes mais n'influence pas celle de la composante Stokes (Capteur basé sur l'intensité). Pour le phénomène Brillouin, un changement de la température fait varier la longueur d'onde du spectre (Capteur basé sur la longueur d'onde).



**Fig. 1.10:** Principe de mesure de température par OTDR basé sur la diffusion Raman.

même est une fonction de la contrainte et de la température. Ceci signifie que la longueur d'onde de la diffusion Brillouin est à la fois une fonction de la contrainte et de la température [Bao 94]. L'avantage de cette diffusion est que les systèmes peuvent utiliser la fibre monomode permettant d'augmenter la longueur d'auscultation. En plus, les capteurs utilisant cette technique sont basés sur la mesure d'une fréquence ce qui est plus robuste par rapport à la technique Raman basée sur la mesure d'une intensité et alors plus sensible. Les capteurs de température exploitant le phénomène Brillouin ont été testés dans des environnements contrôlés où il est plus facile d'annuler la dépendance de contrainte. La plupart des systèmes commerciaux utilisent encore la thermométrie Raman mais avec les avantages d'effet Brillouin, des capteurs commerciaux de mesure distribuée de température Brillouin sont à venir. Une installation typique de DTS Sentinel Sensornet est montrée dans la figure 1.11.



**Fig. 1.11:** Installation typique de DTS Sensornet pour acquisition des mesures thermométriques.

#### 1.4.5 Spécification des paramètres et mesures des performances

La spécification d'un système distribué d'auscultation est liée aux paramètres qui contrôlent sa performance. Les différentes composantes du système, dont la fibre elle même fait partie, jouent un rôle sur la performance. L'installation et les différentes connexions doivent être vérifiées comme une première étape d'un système d'auscultation. Les différents paramètres sont souvent interdépendants et une compréhension de ces paramètres et de leurs effets sur la performance du système est importante pour bien exploiter le système. En général, les paramètres d'intérêt sont : la précision et la résolution des mesures, la résolution spatiale, le temps d'acquisition, la

longueur d'auscultation, la gamme dynamique, etc. [Niklès 07a].

### Résolution spatiale

La résolution spatiale d'un système de détection est sa capacité à distinguer deux positions adjacentes soumises à des températures différentes. Considérant un changement abrupt dans la température  $T$  introduit à une position donnée de la fibre, la distance entre les points à 10% et 90% de ce changement de température représente la résolution spatiale. Cette résolution est liée à la largeur d'impulsion, donc à la distance éclairée  $L = \tau v_g/2$ , avec  $\tau$  la largeur d'impulsion incidente et  $v_g$  sa vitesse de groupe. Dans le cas de la fibre optique en silice avec  $v_g = 2 \times 10^8 m/s$  et une largeur d'impulsion de  $10ns$ , la résolution spatiale d'un système de mesure de température par des capteurs distribués à fibre optique est de  $1m$ . Ceci signifie qu'un changement de température sur une distance plus grande que la résolution spatiale sera détecté avec un taux de réussite de 100%. Cependant, un changement de température local sur une distance inférieure à la résolution spatiale pourra être détectée mais avec une précision inférieure à 100%.

### Résolution en température et précision de mesure

Nous définissons l'exactitude d'une mesure comme la concordance entre la valeur mesurée et une valeur de référence. Elle dépend d'une erreur aléatoire (précision) et d'une erreur systématique (biais). La précision est liée aux déviations spontanées dans l'enregistrement autour de la valeur moyenne; elle n'est pas liée aux changements de mesure. La précision est établie en fonction de l'occurrence d'une mesure répétée sous des conditions identiques ou quasiment similaires. La précision est définie comme deux fois  $\pm$  l'écart type du bruit comprenant 94.5% des mesures, avec des mesures répétées suivant une loi normale. La résolution en température est définie comme une, deux ou trois fois  $\pm$  l'écart type du bruit de mesure selon le niveau de confiance désiré, avec une fois l'écart type souvent considéré comme une valeur optimale [Niklès 07a]. Une valeur typique de résolution en température des capteurs distribués à fibre optique est  $0.01^\circ C$ .

### Longueur d'auscultation

La longueur d'auscultation correspond à la distance maximale sur laquelle nous pouvons faire des mesures sans dépasser les valeurs nominales de la résolution spatiale et de la précision de mesure. Nous définissons souvent la gamme dynamique comme la perte optique maximale cumulée jusqu'au point où nous ne pouvons plus satisfaire les spécifications nominales de la résolution spatiale et de la précision. Une valeur typique de longueur d'auscultation d'un système de mesure de température par des capteurs distribués à fibre optique basés sur la diffusion Raman est de 10 km.

### Durée de mesures

Le temps d'acquisition joue sur le rapport signal sur bruit (RSB). Il faut avoir plusieurs mesures indépendantes pour améliorer ce rapport par l'intégration ou moyennage de ces mesures. Ceci en

revanche à un effet sur la dynamique de mesure en temps car nous n'arriverons plus à détecter une transition de température qui a une durée plus petite que le temps d'acquisition.

Les différents paramètres cités ci-dessus ont une interdépendance, ce qui signifie que la performance d'un système de mesure de température par des capteurs distribués à fibre optique ne reste pas constante sur toute la longueur d'auscultation. Le RSB est lié à la résolution spatiale, à la distance, à l'atténuation de la fibre et à l'intensité de l'impulsion. Même si nous fixons la résolution spatiale, l'exactitude n'est pas constante sur toute la longueur de la fibre parce que l'intensité est affectée par l'atténuation et le bruit augmente avec la distance. Il faut donc quantifier l'exactitude d'une mesure en fonction de la distance par rapport à l'atténuation nominale de la fibre. Avec les fibres multimodes utilisées par les systèmes Raman, la dispersion des ondes conduit à l'élargissement des impulsions avec la distance ce qui affecte la résolution spatiale. L'amélioration du RSB en augmentant la puissance à l'entrée est limitée par des effets non-linéaires qui affectent la longueur d'auscultation et l'exactitude d'une mesure. Il existe alors plusieurs compromis possibles, surtout entre la résolution spatiale, l'exactitude des mesures et la longueur d'auscultation. Des tests doivent être faits pour calibrer le système de mesure afin d'établir le meilleur compromis pour une application donnée [Henault 06].

#### 1.4.6 Avantages de capteurs distribués à fibre optique

Nous avons présenté une vue globale des capteurs de température à fibre optique en évoquant la motivation liée à l'utilisation des mesures de température et plus précisément les mesures avec les capteurs distribués. Les avantages offerts par ces capteurs peuvent être résumés ci-dessous [Dakin 90, Rogers 02, Yin 00] :

1. Un des plus grands avantages des capteurs à fibre optique est leur capacité d'intégrer un grand nombre de capteurs sur une seule fibre optique, ce qui n'est pas le cas avec des capteurs classiques. Cette notion est très utile pour l'auscultation des grandes structures comme les digues, les barrages, etc. et permet d'ausculter plusieurs dizaines de kilomètres avec une résolution spatiale de l'ordre de 1 m et une précision de mesure de l'ordre de  $0.01^{\circ}C$ ,
2. La fibre optique utilisée est une fibre optique de télécommunication classique, ce qui représente une solution peu onéreuse et qui permet en même temps d'utiliser des matériaux électroniques classiquement utilisés en télécommunication tout en rendant le système économiquement efficace,
3. Les capteurs à fibre optique permettent de détecter passivement une large gamme de paramètres physique et offrent une très bonne isolation électrique,
4. Le système est très robuste aux effets d'interférence électromagnétique grâce à l'utilisation des fibres optiques,
5. Les capteurs sont relativement solides et résistent très bien à la corrosion,

6. Ils ont la capacité d'être déployés dans un environnement sévère, ce qui les rend très utiles dans des secteurs particuliers comme le secteur pétrolier et gazier (où les températures peuvent monter à des valeurs très élevées) ou dans les centrales nucléaires (où les risques de radiation sont élevés),
7. Un autre avantage important par rapport aux méthodes de résistivité et de potentiel spontané est que nous n'avons pas besoin de connaissances a priori sur l'emplacement et le nombre de capteurs à utiliser,
8. La vie d'une fibre optique étant raisonnablement longue, les interventions de maintenance fréquentes ne sont pas nécessaires une fois le système bien installé.

Pour résumer, les capteurs distribués de température à fibre optique offrent une solution convenable pour l'auscultation des ouvrages. Il existe cependant quelques inconvénients qui empêchent l'explosion du marché des capteurs à fibre optique. Parmi les critiques, nous pouvons énumérer : les connectiques des fibres optiques requièrent une expertise, le fait de justifier le coût d'un système distribué de mesure par rapport au coût des capteurs traditionnels. Les critiques étant justes, elles sont biaisées par le fait que c'est une technologie relativement récente. Son ouverture à diverses applications aidera le développement de son marché et fera baisser les coûts.

### 1.4.7 Applications des capteurs distribués à fibre optique

#### L'industrie pétrolière

L'industrie pétrolière utilise une grande variété des capteurs. Les conditions souvent imposées sur ces capteurs sont des très hautes pressions et hautes températures (de l'ordre de 200° C). Les capteurs à fibre optique deviennent alors un choix naturel avec leur capacité de supporter des environnements sévères et de détection distribuée.

Un système de mesure utilisant des capteurs distribués à fibre optique offre de nombreux avantages pour l'auscultation de circulation des fluides à l'échelle des grands réservoirs. Yamate [Yamate 08] présente l'exploitation de ces capteurs à l'auscultation des profils de température dans un système de récupération de pétrole lourd par l'injection de vapeur chez Schlumberger. De la même manière, la distribution de la température dans un système de stockage de fond et sa variation au cours du temps sont des paramètres importants pour analyser les conditions d'opération [Vogel 01]. Les mesures de température sur toute la longueur d'un tubage de production sous différentes conditions d'opération fournissent des informations sur les différentes anomalies. Ces anomalies pourraient être liées aux fuites dans le tubage de production et/ou dans les cuvelages ou liées à des écoulements extérieurs. Les techniques conventionnelles de sondage ne permettent pas de mesurer la température simultanément en fonction de la profondeur et du temps sur toute la longueur de tubage de production. En plus les sondes standard de température dans des forages sont souvent invasives.

### **Contrôle des câbles électriques**

Un domaine d'application intéressant pour les capteurs distribués à fibre optique est le contrôle des câbles électriques [Hartog 02]. L'immunité électromagnétique des fibres optiques est une propriété importante dans cette application. Le but est de déterminer le courant admissible en temps réel sur un câble. Normalement, la valeur nominale dépend du milieu autour du câble qui est souvent le remblayage avec une résistivité thermique maximale garantie. La résistivité du sol autour de ce remblayage dépend de sa composition et surtout de sa teneur en eau. L'assèchement du sol dépendra de la charge nominale dans le câble et des conditions climatiques. Les techniques conventionnelles utilisent des modèles avec des valeurs de la résistivité thermique et de la teneur en eau prédéterminées et non pas les valeurs actuelles sur le terrain. Les capteurs distribués permettent d'enregistrer la température actuelle autour du câble afin de la comparer avec des caractéristiques nominales. Ces données de température sont ensuite analysées avec des modèles adaptés [Hartog 02]. En outre, les capteurs distribués permettent aussi de détecter les points locaux de température élevée (hot-spots) pour identifier leurs causes.

### **Détection d'incendies**

La détection des incendies dès leur début est essentiel dans les infrastructures comme les tunnels, les bâtiments, etc. L'adaptation de capteurs distribués comme des détecteurs linéaires de chaleur a été empêchée initialement à cause de la notion de coût mais aussi à cause de la faible connaissance sur cette nouvelle technologie. Mais récemment, ce système a été utilisé dans plusieurs installations pour la détection d'incendies [Orrell 98, Grosswig 01, Hartog 02, Sensornet 09]. Des systèmes utilisant des capteurs distribués à fibre optique ont été installés dans de nombreux tunnels. Les données acquises sont analysées pour éviter des fausses alarmes à cause des divers effets comme la chaleur émise par les véhicules, les conditions de ventilation, etc. Ces systèmes ont été également déployés dans le réseau de chemin de fer métropolitain, couvrant des longues distances, en utilisant différentes stratégies pour activer des alarmes.

### **Contrôle des conduites**

Le dégagement du pétrole ou du gaz dans les conduites peut avoir des influences graves à la fois sur l'environnement et sur l'économie. Les fuites de gaz peuvent être une source d'explosion avec une gravité dépendant de la nature de la fuite ; les fuites pétrolières peuvent résulter en pollution et/ou en incendies. Un système basé sur des capteurs distribués peut localiser ces fuites avec la fibre installée en parallèle avec la conduite. Pour les fuites de gaz, les anomalies dues aux fuites sont détectées sur la température en utilisant l'effet Joule-Thomson [Vogel 01, Niklès 07b, Grosswig 05].

### **Suivi de la nappe phréatique à long terme**

Le suivi de nappe phréatique est une activité assez importante parce que de temps en temps, de l'eau contaminée, l'eau des rivières et d'autres fluides s'infiltrent dans la nappe. Ces fluides

ont souvent une température différente de celle de la nappe normale et changent la température locale du terrain le long du chemin d'écoulement. La distribution de température ainsi que sa variation temporelle peuvent être des moyens efficaces pour qualifier l'eau [Vogel 01]. Les variations de la nappe phréatique pourraient également être surveillées avec des capteurs distribués de température. Un tel système pourra renseigner sur les processus qui se déroulent en souterrain, processus liés soit à la nappe soit à la saturation de substrat.

Il existe de nombreuses autres applications pour les systèmes basés sur des capteurs distribués à fibre optique comme la détection d'infiltrations dans les barrages ; la surveillance à long terme des décharges et des déchets solides ; les structures civiles et celles sensibles comme les ponts, les centrales nucléaires, etc. [Ferdinand 90, Vogel 01, Fernandez 05].

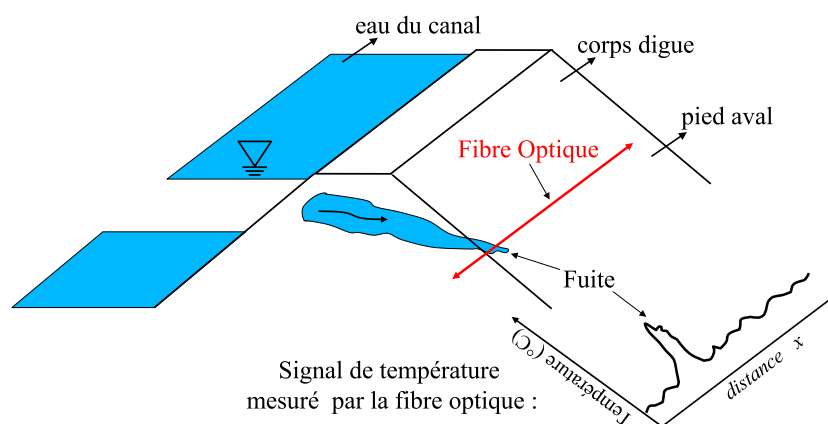
## 1.5 Principe de détection des fuites en digues par mesures de température à l'aide de capteurs distribués à fibre optique

Le principe de localisation d'une fuite dans une digue en terre par des mesures de température utilisant des capteurs distribués à fibre optique est montré sur la figure 1.12. 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 d'advection qui modifie le champ de température. Enterrant la fibre en butée de pied aval du canal, nous pourrions intercepter ces changements de température. Une fuite a donc comme conséquence un changement des conditions thermiques locales ce qui peut être détecté par le système de mesure de température à l'aide de capteurs distribués à fibre optique. Cependant, sur le terrain quand nous parlons d'un système d'auscultation en continu, la signature d'une fuite n'est pas toujours si nette que cela. D'autres phénomènes environnementaux affectent également la température captée par la fibre. Afin d'éliminer leur influence, des mesures de température dites "actives" peuvent être effectuées. La mesure active de température consiste à observer la dynamique de température le long de la fibre optique sous l'effet d'un dispositif de chauffe, typiquement un câble électrique de résistivité élevée relié à une alimentation électrique, et dissipant la puissance électrique par effet Joule.

Pour un système de détection, le but est de séparer les signatures des fuites par rapport à d'autres signatures provoquées par d'autres phénomènes existants et en même temps de limiter le nombre de fausse alarme lors d'une détection.

Dans le cadre de ce travail, nous proposons des méthodes exploitant seulement les données acquises par la fibre optique en mode passif. Le but est de réaliser un système automatique exploitant les variations de données de température sans aucun a priori quelconque. Ces méthodes feront le sujet des chapitres suivants.





**Fig. 1.12:** Principe de la détection d'une fuite en mesurant la température par fibre optique.

## 1.6 Conclusion

Après avoir défini ce que constitue une fuite, nous avons étudié les méthodes classiques de diagnostic des digues en terre. Nous avons évoqué l'importance d'avoir des systèmes non-intrusifs à grand rendement qui ne nécessitent pas l'intervention humaine pour le diagnostic. Nous avons constaté que les mesures thermométriques contiennent de l'information utile concernant les fuites d'eau dans les digues. Pour mesurer ces températures, un système utilisant les capteurs distribués qui fait appel à une fibre optique et basé sur le principe du ODTR a été présenté. Ce système offre de nombreux avantages comme le multiplexage des capteurs, un grand rendement d'auscultation, une bonne résolution de température de l'ordre de  $0.01^{\circ}\text{C}$  et une bonne résolution spatiale de l'ordre de 1 m tout en offrant une bonne précision des mesures. La robustesse et l'efficacité économique de la fibre optique rendent ces systèmes très pratiques dans les environnements sévères. Ce type de système est de plus en plus utilisé dans différentes applications pour la détection des incendies, la détection des fuites dans les conduites, l'auscultation des ouvrages, etc. Dans le chapitre suivant nous allons présenter plusieurs sites expérimentaux installés par EDF, les systèmes d'acquisition et les données acquises. Les diverses prétraitements appliquées aux données feront également le sujet du chapitre suivant.

# Chapitre 2



## Chapitre 2

# Description des mesures thermométriques et Prétraitements

### Sommaire

---

|            |  |           |
|------------|--|-----------|
| <b>2.1</b> | <b>Introduction</b>                                  | <b>43</b> |
| <b>2.2</b> | <b>Données simulées</b>                              | <b>43</b> |
| 2.2.1      | Modélisation des températures d'une digue            | 43        |
| 2.2.2      | Génération d'une digue                               | 47        |
| 2.2.3      | Signaux de température simulés                       | 49        |
| <b>2.3</b> | <b>Le site expérimental d'Oraison</b>                | <b>51</b> |
| 2.3.1      | Généralités  | 52        |
| 2.3.2      | Caractéristiques des données enregistrées à Oraison  | 54        |
| 2.3.3      | Les fuites localisées en temps (fuites ponctuelles)  | 57        |
| <b>2.4</b> | <b>Le site expérimental de Kembs</b>                 | <b>60</b> |
| 2.4.1      | Généralités  | 60        |
| 2.4.2      | Caractéristiques des données enregistrées            | 62        |
| <b>2.5</b> | <b>Modélisation des signaux de température</b>       | <b>67</b> |
| <b>2.6</b> | <b>Prétraitements</b>                                | <b>69</b> |
| 2.6.1      | Désaisonnalisation                                   | 69        |
| 2.6.2      | Normalisation des données                            | 70        |
| 2.6.3      | Identification des phénomènes éphémères énergétiques | 71        |
| <b>2.7</b> | <b>Discussion</b>                                    | <b>75</b> |

---



## 2.1 Introduction

Les mesures de température par capteurs à fibre optique représentent une solution efficace pour le diagnostic de digues en terre. Cependant, avant de présenter l'analyse des données thermométriques pour la détection des fuites, nous allons nous intéresser à la chaîne d'acquisition et les facteurs qui peuvent influencer ces données, car, les données enregistrées par les différents systèmes peuvent être très variables selon la composition géologique du terrain dans lequel la fibre optique est placée et selon les conditions d'acquisitions qui peuvent varier d'un environnement complètement contrôlé à un environnement réel.

Ce chapitre a pour but de présenter les caractéristiques des signaux de température simulés ou enregistrés sous différentes conditions, la modélisation de données de température et leur prétraitement. Basé sur un modèle physique de digue [Guidoux 08], nous nous intéressons dans un premier temps aux données de température simulées dans un environnement complètement contrôlé. Nous présentons ensuite deux sites expérimentaux permettant d'acquérir des données en présence de fuites simulées et de fuites réelles ainsi que les caractéristiques des données enregistrées sur ces deux sites. Les données de température sont ensuite modélisées par un modèle linéaire qui intègre les différents facteurs caractérisant les données acquises. La dernière partie de ce chapitre est consacrée à la présentation de prétraitements qui peuvent être utiles dans l'analyse par la suite.

## 2.2 Données simulées

La modélisation numérique des couplages physiques au sein des digues représente un des sujets actuels de recherche. Dans ce contexte, un des principaux objectifs a été de modéliser les températures aux différents points d'une digue soumise à une charge hydraulique [Guidoux 08]. Dans cette thèse, nous discutons brièvement l'aspect physique de cette modélisation. Pour plus de détails sur ce sujet le lecteur pourra consulter la thèse de Guidoux [Guidoux 08]. Nous utiliserons ensuite cette modélisation pour simuler le comportement de nos propres digues traversées par différents types des fuites. Le but est de valider par la suite les méthodes d'analyse proposées dans ce travail dans le cas d'un environnement contrôlé.

### 2.2.1 Modélisation des températures d'une digue

#### Notions de base

Afin de formaliser la modélisation d'une digue, nous reprendrons ici quelques notions introduites par Guidoux [Guidoux 08]. En général, un sol peut être vu comme un assemblage de trois phases : solide, fluide et gaz, contenus dans un volume total  $V_T$ . Notons  $V_s$  le volume occupé par la partie solide et  $V_v$  le reste du volume constitué des pores remplis de fluide et/ou de gaz. La fraction fluide, en l'occurrence l'eau, occupe le volume  $V_w$ . La porosité du sol  $\epsilon$  caractérisant le volume

non occupé par la fraction solide dans un volume de sol est donnée par :

$$\epsilon = \frac{V_v}{V_T}. \quad (2.1)$$

La quantité d'eau présente dans le sol caractérisée par la teneur en eau volumique  $\theta$  et le degré de saturation  $S$  sont définis comme :

$$\theta = \frac{V_w}{V_T}, \quad (2.2)$$

et

$$S = \frac{V_w}{V_v} = \frac{\theta}{\epsilon}. \quad (2.3)$$

A l'état de saturation, la teneur en eau  $\theta_s$  est égale à la porosité, ce qui implique  $S_s = 1$ , où l'index "s" dénote "saturation". Par ailleurs, la teneur en eau ne descend pas en dessous d'une valeur résiduelle  $\theta_r$  propre à chaque site. De même, la saturation ne descend pas en dessous d'une valeur résiduelle  $S_r = \theta_r/\epsilon$ .

La facilité avec laquelle l'eau peut être transportée dans un sol est caractérisée par la perméabilité intrinsèque du sol  $k_{int}$ , indépendante du fluide transporté. La conductivité hydraulique (ou perméabilité à l'eau)  $K_h[m/s]$ , un élément essentiel de la modélisation du sol, est définie par :

$$K_h = k_{int} \frac{\rho_w g}{\mu_w}, \quad (2.4)$$

avec  $g$  l'accélération de la pesanteur,  $\rho_w[Kg/m^3]$  et  $\mu_w[Pa \cdot s]$  respectivement la densité volumique et la viscosité dynamique de l'eau. La conductivité hydraulique varie avec la teneur en eau du sol, de sorte que  $K_h$  augmente avec  $\theta$ . La conductivité hydraulique à saturation,  $K_s$ , dépend de la dimension caractéristique des grains constitutifs du sol : plus le sol est fin, moins il sera conducteur. La conductivité hydraulique intervient dans la loi de Darcy qui relie la vitesse d'écoulement à la charge hydraulique  $H[m]$  [Guidoux 08]. En un point donné de l'espace de coordonnées  $(x, y, z)$ , la charge hydraulique est donnée par :

$$H = z + \frac{P}{\rho_w g} + \frac{v^2}{2g}, \quad (2.5)$$

avec  $z$  l'altitude du point par rapport à un repère,  $P[Pa]$  la pression d'eau dans le sol exprimée relativement à la pression atmosphérique et  $v[m/s]$  la vitesse du volume élémentaire de fluide autour du point considéré. Au lieu de charge et de pression, les notions de potentiel hydraulique,  $\psi_h$ , et de potentiel de pression,  $\psi_p$ , sont souvent utilisés [Guidoux 08]. Le potentiel hydraulique pour les milieux poreux est donné par :

$$\psi_h = z + \psi_p \quad (2.6)$$

Le potentiel de pression, considéré comme nul à pression atmosphérique, devient positif lorsque le fluide du sol est soumis à une pression hydrostatique, c'est-à-dire lorsque le fluide est situé en dessous d'une surface d'eau libre (par exemple une nappe phréatique) et devient négatif lorsque le fluide est dans un état de sous-pression (fluide soumis à des forces de capillarité ou d'absorption) [Soutter 07].

|           | Type de sol    | Perméabilité $K_h$<br>[m/s] | Débit total<br>[lit/min] | Vitesse maximale<br>[m/s] |
|-----------|----------------|-----------------------------|--------------------------|---------------------------|
|           | argile         | $10^{-8}$                   | $1.1 \times 10^{-4}$     | $1.74 \times 10^{-9}$     |
| alluvions | grave          | $10^{-3}$                   | 10.8                     | $1.74 \times 10^{-4}$     |
|           | sable          | $10^{-4}$                   | 1.1                      | $1.74 \times 10^{-5}$     |
|           | sable-limoneux | $10^{-5}$                   | 0.1                      | $1.74 \times 10^{-6}$     |
|           | limon          | $10^{-7}$                   | 0.001                    | $1.74 \times 10^{-8}$     |

**Tab. 2.1:** Débit et vitesse en fonction de la nature du sol et de sa perméabilité.

### Les équations de modélisation

Soient  $x_1 = x$ ,  $x_2 = y$  et  $x_3 = z$  les trois directions de l'espace. Utilisant la sommation des indices répétées, la conservation de la masse dans un volume élémentaire décrivant l'aspect hydraulique du problème est donné par l'équation de continuité suivante [Guidoux 08] :

$$Q = S_0 S(\psi_p) \frac{\partial \psi_p}{\partial t} + \epsilon \frac{\partial S(\psi_p)}{\partial t} + \frac{\partial v_{D,i}}{\partial x_i}, \quad (2.7)$$

avec  $i = 1, \dots, 3$ ,  $\frac{\partial}{\partial t}$  la dérivée par rapport au temps,  $\frac{\partial}{\partial x_i}$  la dérivée par rapport à la direction  $i$  de l'espace,  $S_0[m]$  la compressibilité du volume de stockage,  $Q[s^{-1}]$  le terme source défini comme un débit par unité de volume et  $v_{D,i}$  la composante dans la direction  $i$  de la vitesse de l'écoulement, donnée par la loi de Darcy.

Pour la modélisation de la température à travers une digue, trois types de sols : limon, argile et sable ont été utilisés avec leurs caractéristiques : perméabilité, débit, vitesse énumérés dans le tableau 2.1.

L'équation de transport de la chaleur permettant de modéliser les températures d'une digue est donnée par [Guidoux 08] :

$$Q_T = \frac{\partial}{\partial t} \left( (\rho C_p)^{eff} T \right) + (\rho C_p)^w v_{D,i} \frac{\partial T}{\partial x_i} - \frac{\partial}{\partial x_i} \left( \lambda_{ij} \frac{\partial T}{\partial x_j} \right) + (\rho C_p)^w (T - T_0) q, \quad (2.8)$$

avec  $Q_T[W/m^3]$  la chaleur par unité de volume,

$(\rho C_p)^{eff} [J \cdot m^{-3} \cdot K^{-1}]$  la capacité calorifique volumique effective du sol dans son ensemble,

$(\rho C_p)^w [J \cdot m^{-3} \cdot K^{-1}]$  la capacité calorifique volumique de la phase liquide,

$T[^\circ K]$  la température locale,

$\lambda_{ij}$  l'élément  $(i,j)$  du tenseur de dispersivité thermique  $[W \cdot m^{-1} \cdot K^{-1}]$ ,

$q[m/s]$  le terme source de volume fluide par unité de surface,

$T_0[^\circ K]$  la température de référence qui permet de prendre en compte l'effet de la température sur certains paramètres (densité du fluide, viscosité, etc.) par des approches polynomiales.

La modélisation a été effectuée en utilisant une méthode de modélisation par éléments finis



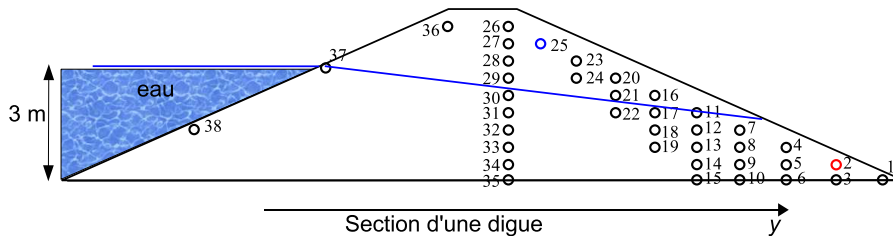
[Diersch 05]. Cette méthode nécessite des conditions initiales pour les différentes variables ainsi que des conditions aux limites du domaine de calcul. Les trois principales conditions aux limites utilisées dans la modélisation sont celles de Dirichlet ( $D$ ), Neumann ( $N$ ) et Cauchy ( $C$ ) [Guidoux 08]. La condition de Dirichlet consiste à imposer une valeur à la variable concernée (par exemple température ou charge). La condition de Neumann consiste à imposer une valeur pour la dérivée de la variable concernée, dans notre cas à imposer un flux surfacique (par exemple le flux de chaleur). La condition de Cauchy est la condition mixte combinant celles de Dirichlet et Neumann, un coefficient d'échange ainsi que les variables de référence associées sont imposées.

### Modélisation de températures en différents points d'une section de digue

Les conditions aux limites hydrauliques consistent à imposer une charge à l'amont entre 0 et 3 m, un flux nul au dessus de 3 m à l'amont jusqu'à la base de l'ouvrage passant par la crête, ainsi qu'une condition de surface libre à l'aval. Les conditions aux limites thermiques consistent à imposer la température de l'eau entre 0 et 3 m à l'amont, un flux nul à la base de l'ouvrage et une condition de Cauchy faisant intervenir la température de l'air et le coefficient d'échange  $h$  partout ailleurs.

Pour ce modèle, 51 conditions aux limites thermiques ont été utilisées avec un champ de température initial en régime sinusoïdal pour chacune de ces conditions. Les températures ont été modélisées sur deux années avec 12 points d'acquisitions par jour. En addition, 38 points de mesure ont été disposés dans le modèle aux différentes coordonnées  $x$  et  $z$ , où  $x$  et  $z$  désignent respectivement l'abscisse depuis le pied amont et la hauteur. La distribution de ces points est montrée sur la figure 2.1 avec leurs coordonnées détaillées dans l'annexe C.2.

Dans le cadre des mesures de température par fibre optique, chacun de ces points de mesures correspond à un emplacement de la fibre. Nous avons donc la notion de temps mais pas celle de distance sur la fibre. Cependant, en utilisant cette modélisation des températures d'une digue, nous pouvons générer notre propre digue comme il sera montré dans la section suivante.



**Fig. 2.1:** La surface libre pour une charge de 3 m à l'amont. Répartition des points de mesure représentés par des cercles. Le point 25 (en bleu) sera considéré pour simuler les zones homogènes alors que le point 2 (en rouge) pour simuler les fuites.

### 2.2.2 Génération d'une digue

Le modèle présenté brièvement ci-dessus permet de modéliser des températures dans différents types de sols, avec différentes perméabilités, différentes conditions aux limites et en différents points de mesure. Pour résumer, nous disposons jusqu'à ce point des températures simulées sur deux ans avec 51 conditions différentes aux limites sur 38 différents points de mesure (nous pouvons également considérer ces points de mesure comme 38 différents emplacements de fibre optique).

Afin de simuler une longueur de fibre (plusieurs points de mesure à différentes distances), nous utilisons les données décrites ci-dessus en combinant les caractéristiques des différents paramètres et points de mesure. Sous une charge hydraulique donnée, les différents types de sols réagissent de différentes façons en termes de suintement dépendant principalement de leurs perméabilités. Dans ce cadre, nous nous intéressons premièrement au tableau 2.1 qui liste les débits à travers l'ouvrage pour différentes caractéristiques des sols. Nous constatons que pour les alluvions (le matériau le plus répandu dans une digue réelle) le plus fort débit de 10.8 [lit/min] est obtenu avec une perméabilité de  $10^{-3}$  [m/s] et le plus bas débit avec une perméabilité de  $10^{-7}$  [m/s]. Nous pouvons donc considérer le sol constitué du limon avec la perméabilité de  $10^{-7}$  [m/s] comme un sol homogène sans fuite et le sol constitué de la grave ou limon érodé avec la perméabilité de  $10^{-3}$  [m/s] comme l'occurrence d'une fuite de fort débit.

La procédure adaptée pour générer une digue avec fuites est de choisir des couches de sol sans fuites et ensuite d'introduire des fuites à différentes distances sur différentes périodes de temps.

Nous avons simulé une digue avec  $N_x$  échantillons en distance pour simuler  $N_x$  mètres de digue avec une résolution de mesure de 1 m. Ce choix n'est pas strict et a été fait pour avoir un nombre important de points de mesure relativement homogènes et sans fuites. Ensuite nous pouvons introduire une variabilité sur le nombre des fuites, sur leur taille et éventuellement sur leur localisation en distance.  $N_x$  étant la longueur totale de la digue, elle sera partitionnée en zones à fuites et zones sans fuites. La longueur cumulée de toutes les zones avec fuites est notée  $N_{fuites}$  et la longueur cumulée des zones sans fuite est notée  $N_{couches} = N_x - N_{fuites}$ . La partie sans fuites est divisée en deux couches, chacune de longueur  $N_{couches}/2$ , composées respectivement de :

- limon avec une perméabilité de  $10^{-7}$  [m/s]
- limon avec une perméabilité de  $10^{-5}$  [m/s]

Parmi les points de mesure modélisés précédemment, le point 25 est choisi pour simuler des couches sans fuites. Ce point se trouve au-dessus du niveau de traversée de l'eau comme montré sur la figure 2.1 et sert de référence. Pour introduire des variations dans ces couches et tendre vers un profil géologique hétérogène, les températures de 5 différentes conditions aux limites sont tirées aléatoirement.

Pour simuler les fuites, nous sélectionnons le limon avec une perméabilité de  $10^{-3}$  [m/s] et le

point de mesure 2, ce qui nous permet de rester en-dessous du niveau d'eau. En spécifiant alors leur localisation en distance et leur longueur  $N_{fuites}$ , ces zones de fuites viennent s'ajouter aux couches sans fuites. Cette démarche permet de simuler des fuites qui vont exister tout le temps et vont apparaître comme des phénomènes ayant le même comportement dynamique en temps que les zones sans fuites. Cependant, les fuites réelles ont souvent des dynamiques temporelles différentes des zones sans fuite et sont souvent localisées en temps. Pour simuler des fuites transitoires (appelées aussi fuites ponctuelles), la première approche consiste à utiliser différentes perméabilités et conditions initiales, pour le même point de mesure 2, au cours du temps. Plus spécifiquement, pour avoir des périodes de temps sans fuite, nous privilégions les perméabilités  $10^{-7}$  [m/s], respectivement  $10^{-5}$  [m/s]. Ces périodes sont introduites en utilisant des fenêtres temporelles rectangulaires de différentes tailles. Ensuite, à l'occurrence d'une fuite, les températures simulées avec la perméabilité de  $10^{-3}$  [m/s] sont introduites de la même manière afin de simuler un débit important. Bien qu'elle permet de modéliser des fuites localisées en temps, cette approche n'est pas réaliste car dans le modèle les fuites sont modélisées par des variations temporelles en marches d'escalier, ce qui n'est pas réaliste.

Une deuxième approche consiste à ne pas utiliser des fenêtres rectangulaires, mais des fenêtres recouvrantes,  $w$ , sous la contrainte d'un recouvrement constant (Constant Overlap Add en anglais), c'est-à-dire, d'avoir des fenêtres qui à chaque instant de temps,  $t_i$ , satisfont la condition :

$$\sum_{l=1}^L w(t_i - lR) = 1, \quad (2.9)$$

avec  $L$  le nombre total des fenêtres et  $R$  le recouvrement. Une fenêtre qui satisfait cette condition pour des recouvrements  $R = (M - 1)/2$ ,  $R = (M - 1)/4$ , etc., est la fenêtre de Hamming donnée par :

$$w(t) = \frac{.54 - .46 \cos\left(\frac{2\pi t}{M-1}\right)}{1.08}, t = 0, \dots, M - 1, \quad (2.10)$$

où  $M$  est la taille de la fenêtre. Le facteur 1.08 au dénominateur et les modifications :  $w(1) = w(1)/2$  et  $w(M) = w(M)/2$  permettent de satisfaire la condition de l'équation (2.9). Un recouvrement de 50% sur ces fenêtres est un bon compromis pour minimiser le nombre de calculs. Le principe de simulation de fuites reste le même que celui décrit précédemment avec la fenêtre de l'occurrence temporelle d'une fuite utilisant la perméabilité de  $10^{-3}$  [m/s].

Pour résumer, le modèle de génération des fuites offre la possibilité de simuler des fuites sur différentes distances, avec différentes tailles en distance, soit constantes soit localisées en temps.

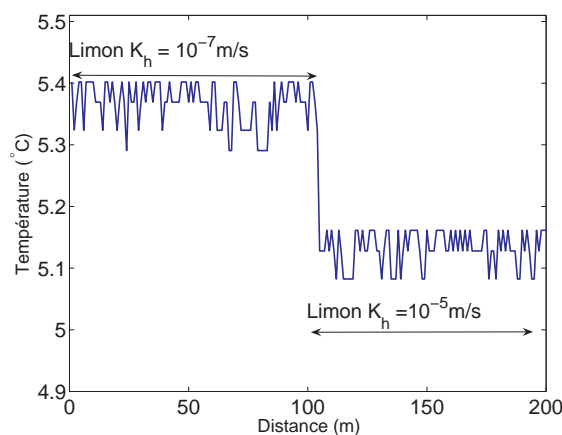
Pour avoir des signaux plus réalistes, comme ceux issus des instruments de mesure, nous ajoutons du bruit aux données simulées précédemment. Il a été démontré que le bruit dû à l'appareil de mesure DTS Sensornet est gaussien, de moyenne nulle et de faible variance [Henault 06]. L'étude de signaux réels ne mettant pas en évidence une liaison statistique ni en distance ni en temps entre les échantillons du bruit, nous pouvons admettre alors que le bruit gaussien, centré, blanc spatialement et temporellement sont un modèle suffisant. En ce qui concerne la puissance de ce

bruit, il a été trouvé que le bruit introduit par un appareil de type DTS Sensornet a un écart type de l'ordre de 0.02, alors que pour le DTS Siemens, qui donne en général des mesures plus bruitées, nous pouvons admettre un écart type plus élevé de l'ordre de 0.05.

### 2.2.3 Signaux de température simulés

Un signal simulé de température correspond à une mesure à un temps donné sur tous les capteurs. La figure 2.2 montre le signal de température sur toute la longueur de la fibre "synthétique" simulé pour  $N_x = 200$  mètres de fibre avec une résolution de 1 m sans aucune occurrence de fuite. Le signal présente deux zones principales, la première moitié (les 100 premiers mètres) correspond à un limon de perméabilité  $10^{-7}$  m/s tandis que la deuxième (les 100 derniers mètres) à un limon de perméabilité  $10^{-5}$  m/s. Pour chacune de ces zones, la variabilité de la température est due aux différentes conditions aux limites. Pour cette simulation, 5 différentes conditions aux limites ont été tirées aléatoirement sur les 51 disponibles. Les répétitions de ces 5 conditions (20 répétitions pour simuler 100m) ont été ensuite mélangées d'une façon aléatoire pour obtenir ce signal.

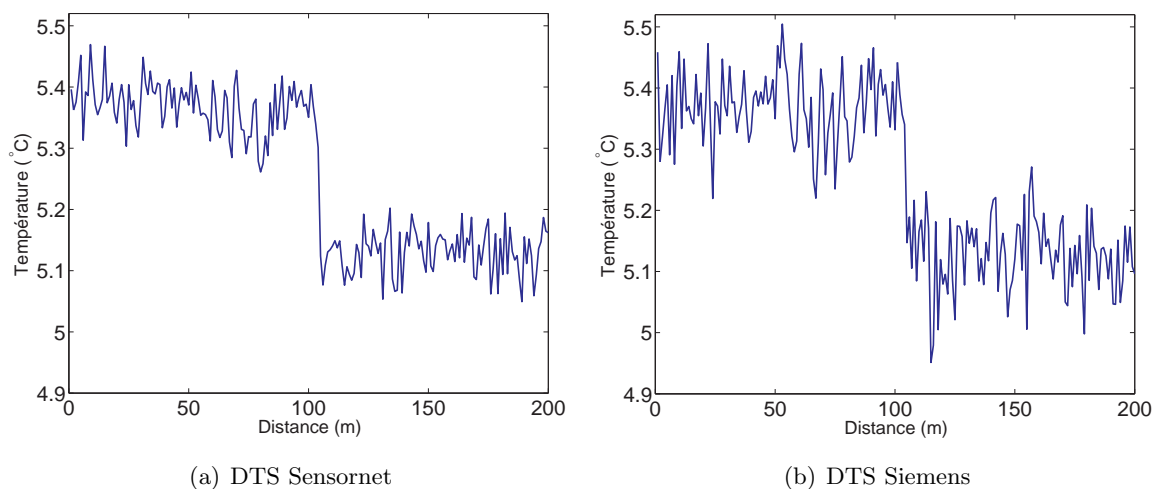
Pour simuler des signaux réalistes intégrant le bruit généré par l'appareil de mesure, nous ajou-



**Fig. 2.2:** Signal de températures simulé pour un sol composé de limon sans occurrence de fuite.

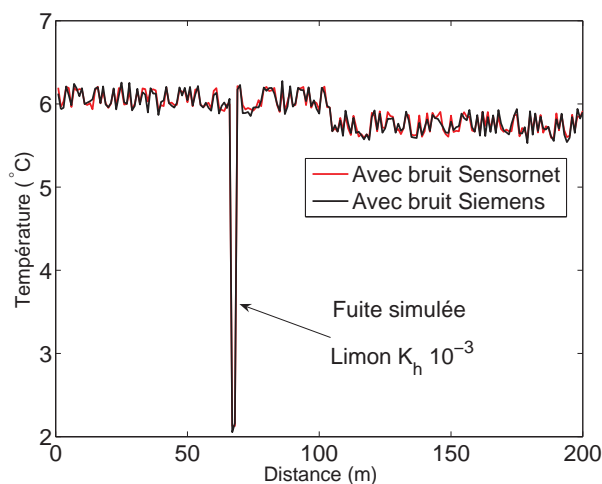
Les deux zones principales correspondent à deux perméabilités différentes. Pour chacune de ces zones, les variations de températures sont dues à un mélange aléatoire de 5 conditions aux limites différentes.

tons un bruit gaussien, centré, blanc spatialement et temporellement. La figure 2.3 montre l'effet du bruit sur le signal de température. Nous constatons que le bruit d'un appareil de type DTS Sensornet (figure 2.3(a)) n'influence pas beaucoup le signal de température simulé, le signal résultant étant assez proche de celui sans bruit. Cependant, le signal obtenu en ajoutant le bruit d'un appareil de type DTS Siemens (figure 2.3(b)) introduit des variations plus importantes qui peuvent être gênantes pour l'analyse des données issues de cet appareil.



**Fig. 2.3:** Signal synthétique en présence du bruit : a) DTS Sensornet et b) DTS Siemens. Le bruit de l'appareil de type DTS Sensornet ne change pas beaucoup le comportement du signal, mais celui de type DTS Siemens introduit des variations plus importantes.

La figure 2.4 montre un exemple d'une fuite simulée introduite aux distances 67 – 68 m et lo-

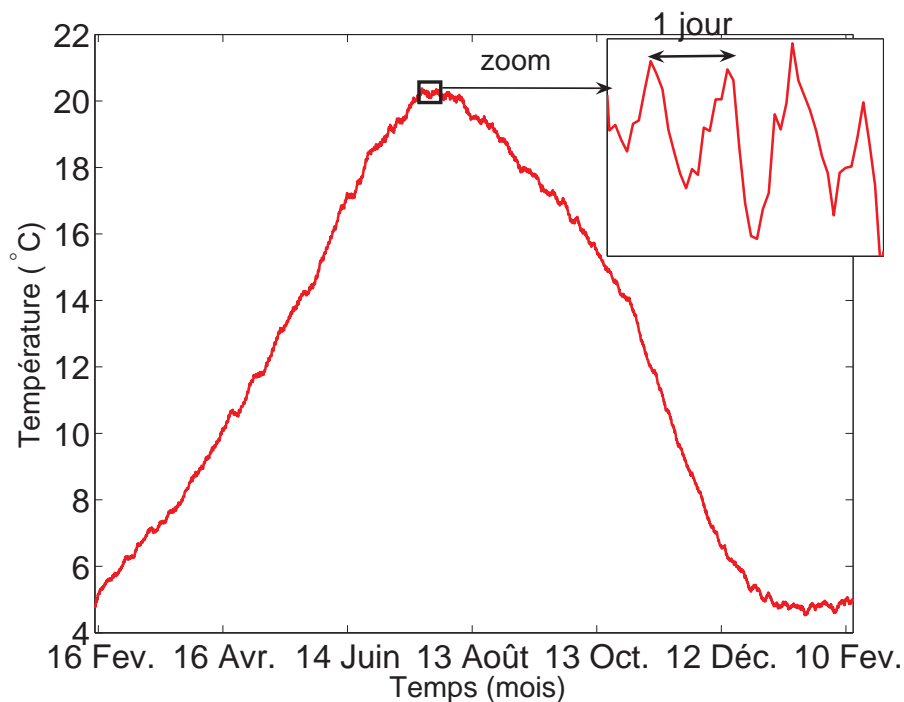


**Fig. 2.4:** Exemple d'enregistrement synthétique de température en présence du bruit avec fuite à la position 67 – 68 m simulée avec un limon de perméabilité  $10^{-3}$  m/s au point de mesure 2 avec un débit de 10.8 lit/min.

calisée en temps sur une durée de 5 jours. Cette fuite est introduite par l'approche de fenêtre recouvrante décrite dans la section précédente en utilisant limon érodé de perméabilité  $10^{-3}$  m/s au point de mesure 2 sur les 5 jours. Les 5 jours avant et après cette fuite sont simulées avec un limon de perméabilité  $10^{-4}$  m/s. Nous constatons que l'occurrence de la fuite est assez évidente, en raison de son débit très important (10.8 lit/min).

Nous disposons de plusieurs simulations espacées de 2 heures sur une période de 2 ans. Ceci nous permet de suivre l'évolution temporelle des différents phénomènes. La figure 2.5 montre l'évolution annuelle de la température à une distance donnée. La température suit en général une variation en corrélation avec la température de l'air en fonction des saisons (été, hiver). De plus, les variations jours/nuits sont également évidentes sur le signal, comme le montre le zoom affiché sur la figure 2.5.

Il faut remarquer que même si les données présentées ici sont modélisées en utilisant un modèle



**Fig. 2.5:** Signal synthétique de température modélisé sur une année à une distance donnée sans fuite et en présence du bruit. Le zoom met en évidence les variations journalières de température.

physique, les données réelles enregistrées sur le terrain peuvent être très différentes car elles sont influencées par divers facteurs comme l'installation de la fibre, les caractéristiques du terrain, les conditions externes (précipitations, etc.), comme il sera présenté par la suite.

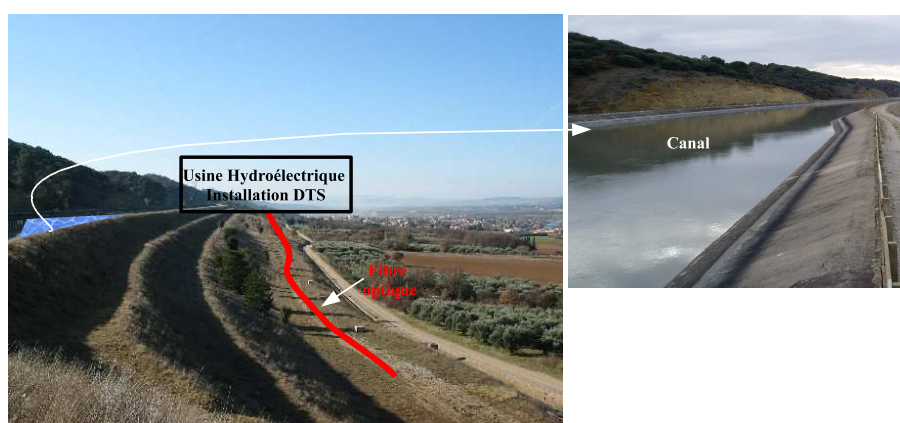
## 2.3 Le site expérimental d’Oraison

Le premier site expérimental pour l'analyse des anomalies réelles sur le terrain a été équipé par EDF au voisinage de l'usine hydroélectrique d'Oraison (Alpes de Hautes Provence).

### 2.3.1 Généralités

L'usine hydroélectrique d'Oraison est alimentée par un canal d'amenée venant du barrage de l'Escale. Un système thermométrique a été installé par EDF ayant pour but d'extraire les informations liées aux fuites (localisation en distance, débits, etc.) dans la digue du canal. La digue en rive droite (RD) de ce canal est constituée d'alluvions de la Durance avec des mélanges des déblais de type poudingues et limons. L'étanchéité de cette digue est assurée par des dalles de béton. La hauteur de la digue augmente avec le rapprochement de l'usine. Les fuites réelles pourraient se manifester par exemple lors de la rupture d'un joint entre les dalles. Les photos de la figure 2.6 montrent la digue sur le canal d'Oraison avec l'emplacement du câble contenant les fibres optiques.

Pour l'installation de la fibre optique, les deux derniers kilomètres de la digue ont été choisis



**Fig. 2.6:** Photos du site expérimental d'Oraison.

parce que la digue présente à cet endroit une hauteur maximale et que des fuites réelles ont été observées auparavant dans cette zone [Fry 04]. Les travaux effectués en 2002 ont eu comme but d'installer un câble contenant 4 fibres multimodes enterré en butée de pied aval du canal à environ 1 m de profondeur. Sur les 2238 mètres de câble, les 2213 premiers mètres sont enfouis et les fibres ont été soudées deux à deux afin de prendre des mesures en boucle. Ces mesures ont permis à EDF de réaliser une évaluation métrologique du système d'acquisition et de vérifier notamment la répétabilité de la mesure ainsi que la "justesse" des mesures [Vercoutere 03]. La fibre optique se situe à environ 30 m du canal en pied de la digue pour le premier kilomètre, puis elle passe sur la risberme pour le reste. La figure 2.7 schématise une coupe transversale de la digue avec les emplacements de la fibre optique. Cette configuration présente donc deux niveaux distincts d'élévation (Zone 1, de 0,1 km à 1,25 km et Zone 2, de 1,25 km à 2,2 km) qui seront exposés de façon différente à l'ensoleillement. La figure 2.8 schématise l'emplacement de la fibre optique le long du canal.

Sur le site expérimental, nous avons la présence d'un certain nombre de drains ovoïdes qui

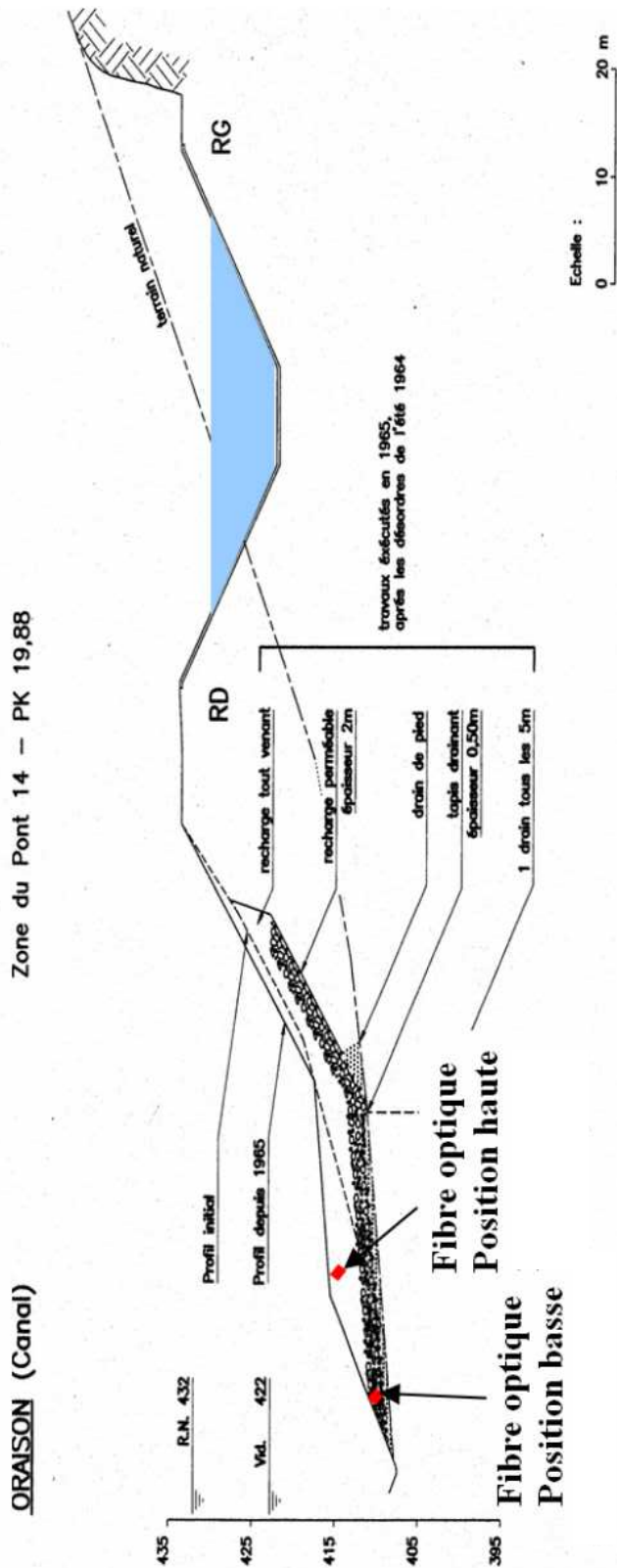
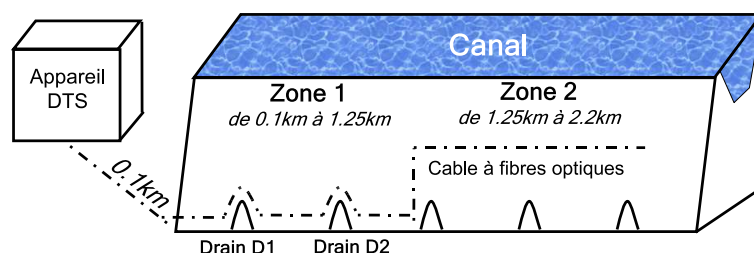


Fig. 2.7: Coupe transversale de la digue expérimentale d'Oraison montrant les niveaux d'installation des fibres optiques.





**Fig. 2.8:** L'installation de la fibre optique montre deux niveaux d'élévation différents : zone 1 pour les environ 1000 premiers mètres de la fibre et zone 2 pour les derniers mètres. La fibre passe autour de deux drains *D1* et *D2* qui sert à drainer l'eau vers le contre canal.

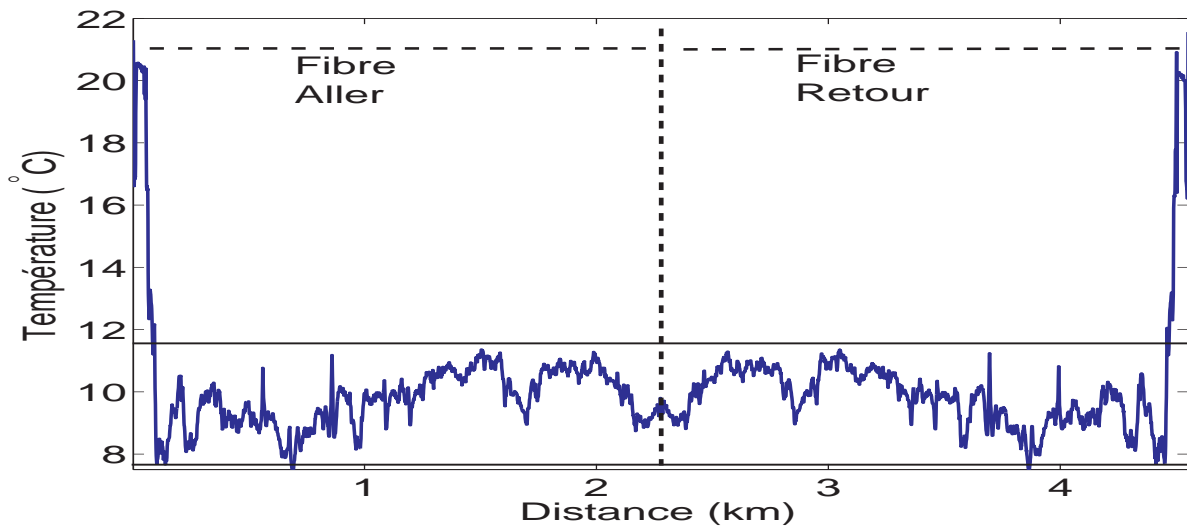
servent à drainer l'eau vers le contre canal en cas d'infiltration dans l'édifice. La fibre passe autour de deux drains qui sont creusés perpendiculairement à la fibre. Deux de ces drains, notés *D1* et *D2* dans le schéma d'installation de la figure 2.8, provoquent une réponse thermique particulière et constituent un type d'anomalie comme nous allons le détailler par la suite. En plus des zones 1 et 2, le sol n'est pas homogène tout le long de la fibre. Ceci a été confirmé par les études géotechniques effectuées par EDF qui ont révélé la présence de matériaux différents. Les fibres sont connectées à l'appareil DTS placé dans la centrale hydraulique.

### 2.3.2 Caractéristiques des données enregistrées à Oraison

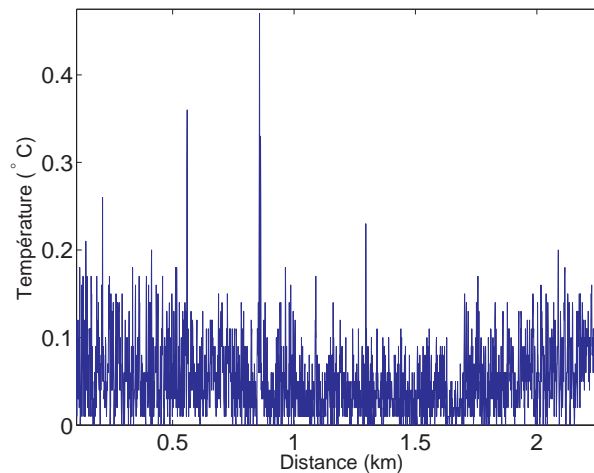
Sur le site d'Oraison, nous avons à notre disposition plusieurs jeux de données de température enregistrés sur différentes périodes entre 2005 et 2007. Le premier jeu de données a été enregistré en utilisant l'appareil DTS Sensornet entre avril 2005 et avril 2006 avec un pas d'acquisition de 1m en distance (résolution spatiale) et avec un pas temporel de 2 h. En 2006, l'appareil DTS Siemens a été utilisé sur la même configuration. Cette fois-ci, le pas spatial d'acquisition est de 0.77 m et celui en temps de 24 min. Les données en 2007 sont enregistrées avec le DTS Sensornet mais avec des acquisitions variables en temps. Nous disposons dans ce cas des acquisitions toutes les 4 heures avec un pas d'acquisition de 1m en distance. Ce jeu de données présente des données manquantes ou même des jours sans aucune acquisition. Ceci est dû à différents facteurs intervenant sur le site, la plupart du temps résultant d'un dysfonctionnement de l'appareil DTS et/ou des problèmes sur la fibre optique elle même.

La figure 2.9 montre le signal enregistré à un temps donné sur la boucle de la fibre en fonction de la distance. La température est globalement comprise entre 8 et 12°C. Les 100 premiers mètres de la fibre ne sont pas enterrés dans le sol et constituent une information inutile dans le contexte de ce travail. Les mesures de température sont symétriques autour du centre de la fibre. La première partie constitue l'aller de la fibre tandis que la deuxième partie constitue le retour de la fibre. Si nous superposons les deux mesures aller et retour, nous constaterons que les deux sont quasiment identiques, aux petites variations près liées au bruit introduit par le DTS (voir

la figure 2.10). Les variations sont visibles surtout au niveau des drains. Si nous disposons des mesures aller-retour, il est souvent conseillé de prendre une moyenne de ces deux mesures afin de réduire le bruit introduit par le DTS et donc d’améliorer le rapport signal sur bruit (RSB).



**Fig. 2.9:** Enregistrement de température à un temps donné sur la fibre en une boucle aller-retour. Les 100 premiers et derniers mètres correspondent à la fibre en aérien.

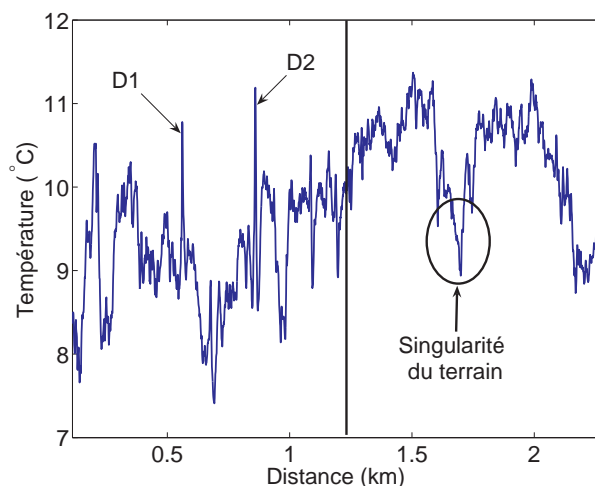


**Fig. 2.10:** Différence absolue des signaux aller-retour affichés sur la figure 2.9 montrant une bonne correspondance avec une moyenne de 0.05 et un écart type de 0.04.

### Le signal de température comme fonction de la distance

Une fois enlevés les 100 premiers mètres de données, nous nous intéressons dans cette section au signal de température obtenu comme la moyenne aller-retour à un instant donné. Comme

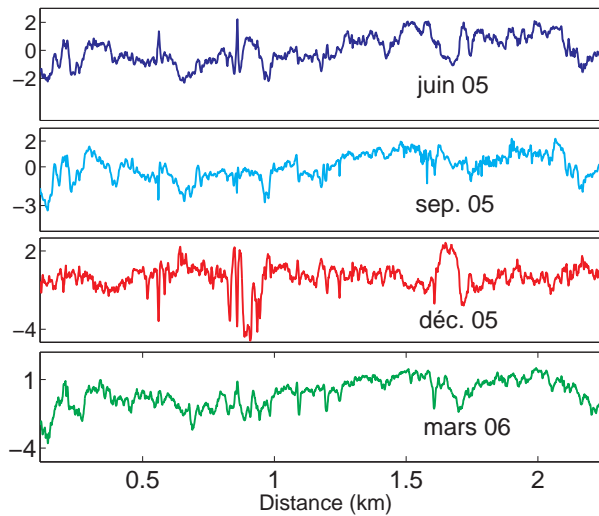
exemple l'enregistrement effectué le 10 avril 2005 à 10h00 est affiché sur la figure 2.11. Ce type de signal sera appelé par la suite *profil de température*. La température variant le long de la fibre met en évidence les deux singularités correspondantes aux deux drains,  $D1$  et  $D2$ . Même si le



**Fig. 2.11:** Profil de température enregistré le 10 avril 2005 à 10h00 comme le moyenne aller-retour. Les deux drains,  $D1$  et  $D2$ , avec une singularité de terrain sont des phénomènes marquants.

signal présente des variations relativement importantes, deux niveaux de température peuvent se distinguer, avant et après environ 1.2 km. Ces deux niveaux s'expliquent par un ensoleillement différent sur les deux niveaux d'élévation du terrain. Il faut remarquer aussi la présence d'une singularité du terrain entre 1.6 et 1.7 km. Celle-ci est liée au fait qu'à cet endroit, le terrain n'est pas constitué de la même matière et présente des hétérogénéités par rapport aux autres distances. L'origine de cette zone hétérogène n'est pas connue avec exactitude mais elle peut être liée aux travaux de réparations antérieurs qui ont fait appel à d'autres matériaux que ceux utilisés lors de la construction de la digue.

Afin de faire une idée sur la variabilité des acquisitions par rapport aux saisons, il est intéressant de comparer les enregistrements effectués à quatre périodes différentes de l'année. La figure 2.12 montre 4 signaux enregistrés le 2 juin 2005, 10 septembre 2005, 19 décembre 2005 et le 20 mars 2006. Les températures ont été normalisées à une moyenne nulle et une variance unitaire afin de les rendre comparables. La première chose que nous remarquons est que le 3<sup>e</sup> signal (décembre) est assez différent des autres et n'offre pas les mêmes variations. La singularité du terrain est plus visible pour ce signal par rapport aux autres phénomènes. Une autre chose que nous pouvons constater est que les températures enregistrées au niveau des drains ne présentent pas les mêmes intensités sur toutes ces acquisitions. Pour résumer, les signaux de températures enregistrés ne montrent pas les mêmes variations au cours des différentes saisons. Ceci reflète le fait que le sol ne se comportant pas de la même manière d'une saison à l'autre complique les profils de température acquis.



**Fig. 2.12:** Enregistrements de température sur les 4 saisons. Du haut en bas : 2 juin 2005, 10 septembre 2005, 19 décembre 2005 et 20 mars 2006. Pour faciliter la comparaison les signaux ont été normalisés à une variance unitaire.

### Le signal de température comme fonction du temps

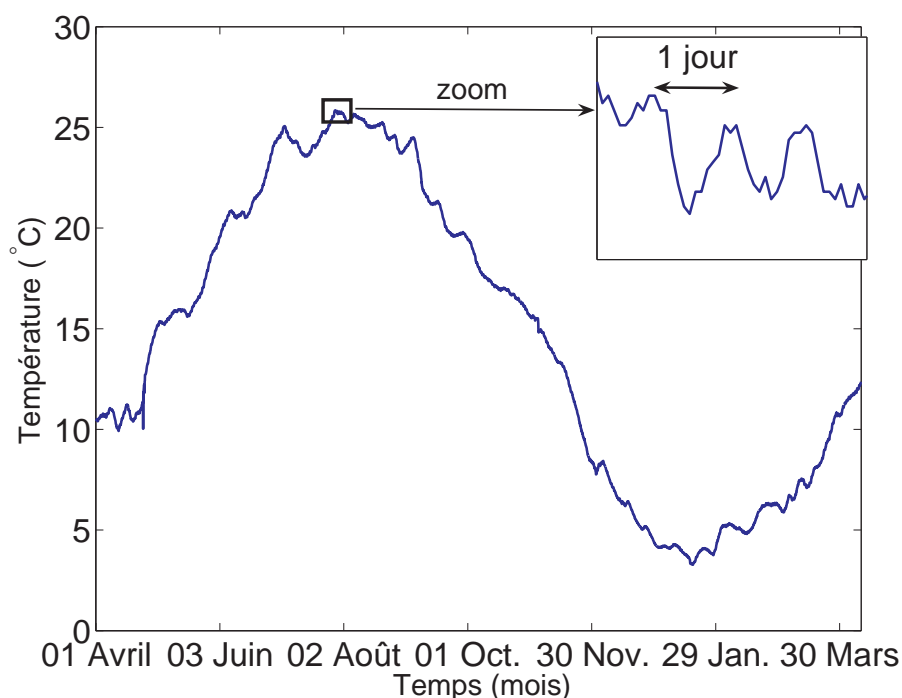
Comme nous l’avons évoqué, nous disposons de plusieurs acquisitions au cours du temps. La figure 2.13 montre un signal enregistré à la distance de 100 m pendant la période 01 avril 2005 - 10 avril 2006, soit l’évolution temporelle de la température au cours d’une année. La température suit en général une tendance “sinusoïdale” de même nature que celle de la température de l’air sur l’année avec des températures maximales en juillet et minimales fin décembre. Les variations autour de la sinusoïde sont dues aux conditions climatiques journalières comme l’humidité, le rayonnement solaire, les précipitations, etc. Nous pouvons remarquer ici la bonne concordance vis à vis des simulations (cf. figure 2.5).

### 2.3.3 Les fuites localisées en temps (fuites ponctuelles)

Afin d’évaluer le système d’acquisition et les algorithmes de traitement pour détecter des fuites au sein de l’ouvrage, des essais de fuites artificielles ont été effectués sur le site d’Oraison [Vercoutere 05]. Ces fuites étaient contrôlées en débit, en distance et en temps. La procédure de création de fuites consiste à prendre de l’eau du canal et à l’amener vers la zone où la fibre optique est enterrée à l’aide d’un tuyau. La figure 2.14 illustre la procédure utilisée pour la réalisation des fuites. Une vanne de réglage permet de contrôler le débit aval pour simuler les fuites avec des débits différents.

Trois fuites ont été simulées entre 10 et 12 mai 2005 :

- Dans un premier temps, une fuite avec un débit de 5 lit/min a été démarrée dans la matinée



**Fig. 2.13:** Évolution temporelle annuelle du signal de température à une distance donnée.

du 10 mai à la position 1.566 km. Cette fuite a eu une surface d'imprégnation d'environ  $2\text{m} \times 3\text{m} = 6\text{m}^2$ . Dans la suite de ce travail, cette fuite est labélisée *L1*.

- Une deuxième fuite avec un débit plus faible de 1 lit/min a été démarrée dans la journée du 12 mai à la position 1.551 km (à 15 m de la précédente). Cette fuite a eu une surface d'imprégnation d'environ  $2\text{m}^2$ . Cette fuite sera labélisée *L2* par la suite.
- Une troisième fuite avec un débit de 1lit/min est lancée dans la soirée du 12 mai avec l'extrémité du tuyau placée cette fois ci à la position 1.571 km. Cette fuite a également une surface d'imprégnation de  $2\text{m}^2$  et elle sera labélisée *L3*.

Il faut mentionner ici que la température de l'eau à la sortie du tuyau n'est pas la même que celle de l'eau du canal puisque l'eau est chauffé dans le tuyau sous l'influence du rayonnement solaire, surtout pour les faibles débits. Le débit a été alors augmenté au maximum pour réduire l'effet de cet échauffement dans le tuyau au démarrage de chaque fuite. L'effet de cet échauffement est moins évident pendant l'essai de la fuite *L3* qui s'est déroulé pendant la nuit.

En outre, un essai de point chaud qui consistait à faire couler de l'eau chaude à la position 0.674 km a été fait la matinée du 10 mai. Cette mesure sera labélisée *HP* pour "Hot Point" dans la suite de ce travail.



(a) Prise d'eau du canal



(b) Tuyau permettant le transport de l'eau du canal



(c) Exemple de zone de fuite

**Fig. 2.14:** Illustration des essais de fuites artificielles sur le site d'Oraison.

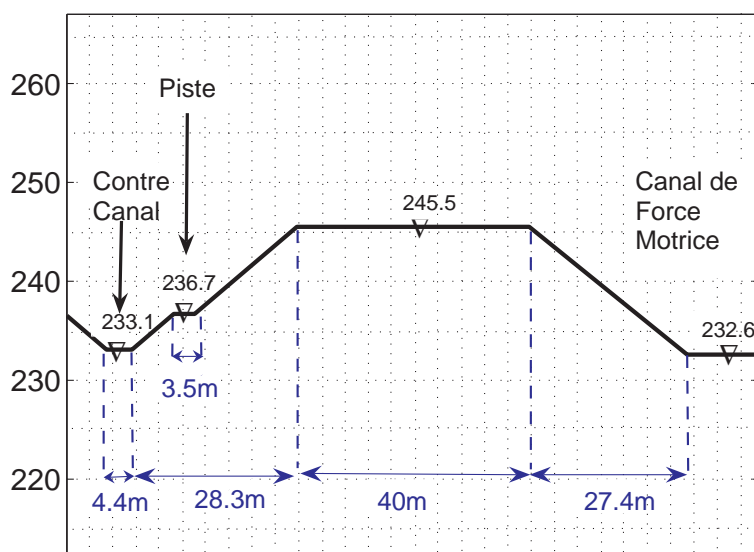
## 2.4 Le site expérimental de Kembs

Le deuxième site expérimental utilisé pour évaluer le système d'acquisition et les algorithmes de détection de fuites est la digue de rive gauche du canal d'amenée de l'usine hydroélectrique de Kembs (Alsace).

### 2.4.1 Généralités

L'usine électrique de Kembs a été construite entre 1928 et 1932 en même temps qu'était aménagé le premier bief du grand canal d'Alsace. Le corps de la digue est très homogène et il est composé de grave sableuse avec présence de galets. Le terrain de fondation est de la même nature que le corps de digue. L'étanchéité est assurée par des dalles en béton de 13 cm d'épaisseur. La fibre de mesure a été installée dans le contre canal de drainage. La coupe de la digue affichée sur la figure 2.15 montre les niveaux du contre canal et d'une piste cyclable qui passe à côté du grand canal. Le contre canal est situé à une distance de 95.7m du grand canal et a une largeur de 4.4m. La température de l'eau du contre canal n'est pas la même que celle de l'eau du grand canal.

Pour ce site expérimental, deux fibres optiques ont été installées :



**Fig. 2.15:** Coupe transversale de la digue expérimentale de Kembs montrant les niveaux d'élévation ainsi que les différentes dimensions de la digue.

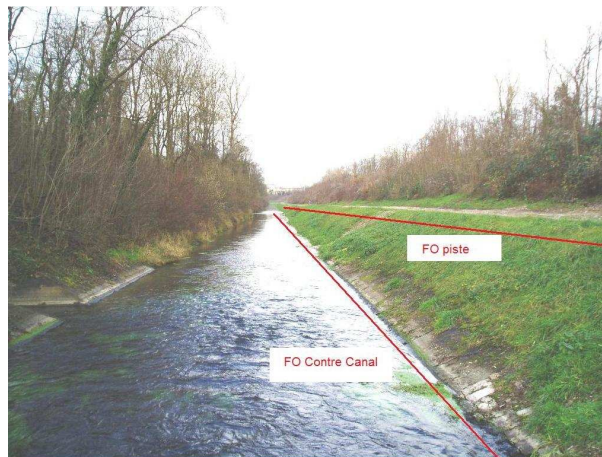
### Fibre optique contre-canal

Cet emplacement (contre canal) a été choisi en considérant qu'une fuite d'eau dans la digue va modifier la température mesurée par la fibre installée dans le contre canal. Cette fibre est située à l'intersection du talus de la berge rive droite et du plafond du contre canal. La fibre est essentiellement dans l'eau et est placée dans un complexe géosynthétique de 30 cm de largeur et

1 cm d'épaisseur. Lors des visites de terrain (en septembre 2008 et février 2009), il a été constaté que la fibre est proche de la surface de l'eau du contre canal, voir immergée, à plusieurs endroits le long de son installation.

### Fibre optique piste

Une deuxième fibre optique appelée fibre d'étalonnage d'une longueur d'environ 1 km est enterrée le long de la piste cyclable à 1m de profondeur. Cette fibre a été placée sur un lit de sable de 10 cm d'épaisseur. La figure 2.16 synthétise l'installation des fibres optiques dans le contre canal et sur la piste.



**Fig. 2.16:** Installation de la fibre optique de mesure dans le contre canal et de la fibre optique d'étalonnage sur la piste cyclable.

### Schéma de principe de l'installation optique

Une troisième partie de la fibre est utilisée afin de mesurer la température de l'eau du grand canal. L'instrumentation consiste donc en 3 boucles : eau grand canal - contre canal - piste. Une bobine de 100 m permettant de faire la calibration des mesures est installée avant la connexion de la fibre optique au DTS. Des bobines, de 20 m chacune, placées dans des boîtes de connexions sont également disponibles pour des calibrations sur une des 2 boucles contre canal respectivement piste. Le chemin parcouru par la fibre depuis le DTS est illustré sur la figure 2.17.

Il faut remarquer ici qu'à la différence du site expérimental d'Oraison où les fuites réelles sont quasiment inexistantes, il existe plusieurs fuites visibles à Kembs. Les deux sites constituent donc deux cas extrêmes par rapport à cette étude de mesures de températures.



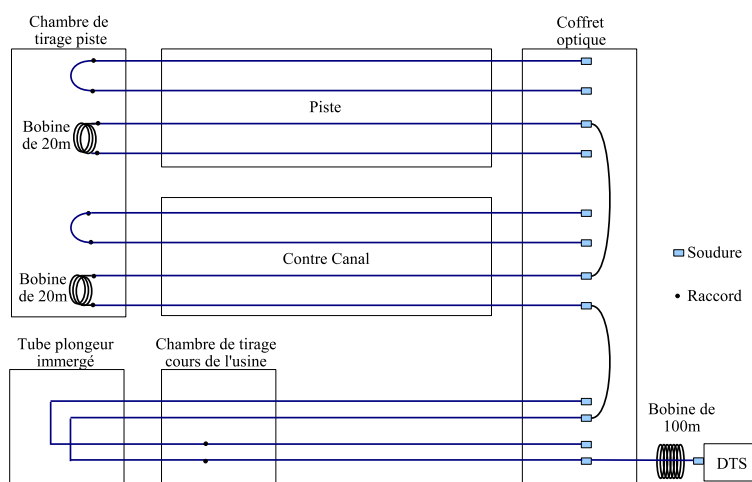


Fig. 2.17: Schéma de principe de l'installation optique du site expérimental de Kembs.

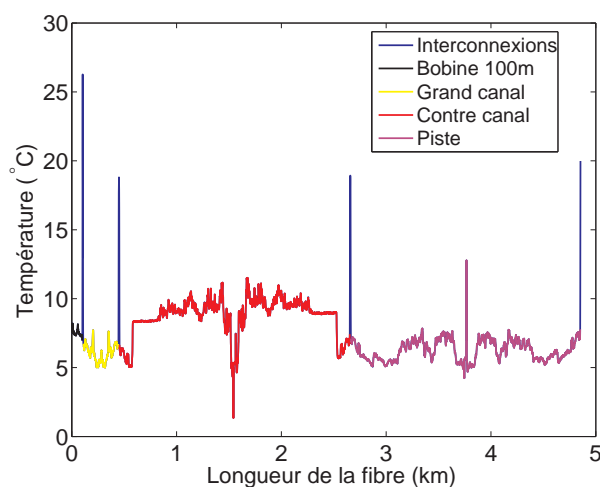
### 2.4.2 Caractéristiques des données enregistrées

Pour le site expérimental de Kembs, des données de température pour la période 30 janvier - 22 février 2007 ont été enregistrées avec l'appareil DTS Sensornet (résolution de température de  $0.01^\circ \text{C}$ ). Les données ont été acquises toutes les heures avec une résolution spatiale de 1m. Par la suite, des données ont été enregistrées entre 30 mars 2007 et 5 février 2008 avec l'appareil Siemens (résolution de température de  $0.1^\circ \text{C}$ ). Des acquisitions ont eu lieu toutes les 24 minutes, ce qui représente le temps d'intégration maximal possible avec cet appareil. La résolution spatiale est de 0.7 m.

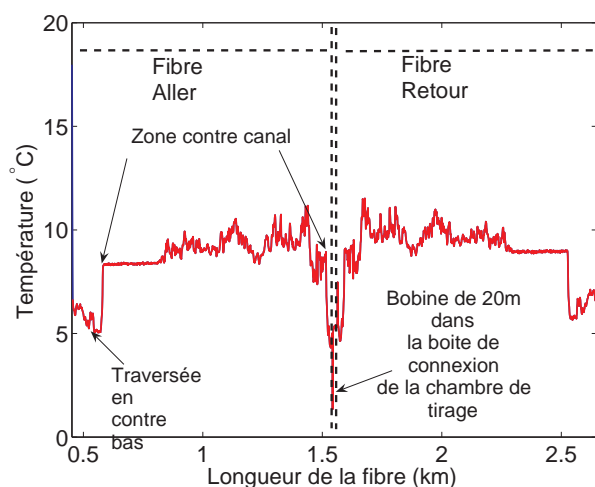
#### Le signal de température en fonction de la distance

La figure 2.18 montre une acquisition de température effectuée à l'aide de la boucle complète bobine - grand canal - contre canal - piste utilisant les parcours comportant les bobines de calibration de 20 m (voir figure 2.17) en fonction du déplacement le long de la fibre. Les différentes composantes de ce signal sont identifiées comme :

- La température de la bobine des 100 premières mètres (en noir) qui se trouve dans le local où le DTS est installé. Ceci permet de faire une calibration des mesures.
- La fibre sort ensuite du local et passe dans le grand canal, puis elle revient vers le local formant alors une boucle. Des mesures de température (en jaune) sont prises tout au long de ce trajet. La mesure de température du grand canal se fait en passant la fibre optique dans un tube immergé dans le grand canal. Cette mesure se situe au milieu du dispositif (zone jaune dans la figure 2.18).
- Le boucle continue avec la fibre enterrée au talus du contre canal (en rouge). Ce signal



**Fig. 2.18:** Acquisition de la température sur la boucle complète de mesure comprenant la bobine (noir), le grand canal (jaune), le contre canal (rouge) et la piste (violet). Les pics bleus correspondent aux soudures.



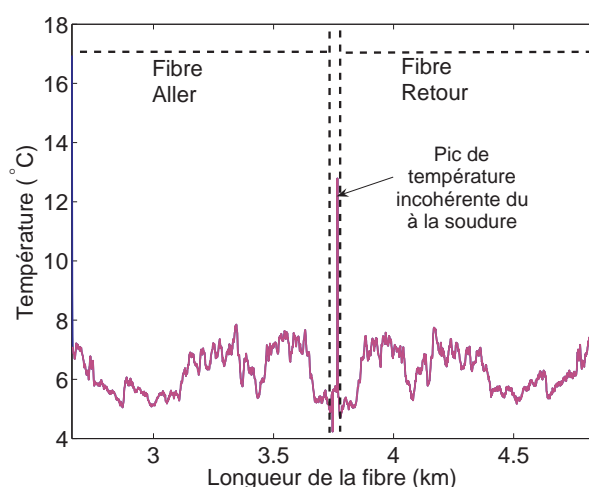
**Fig. 2.19:** Partie contre canal de l'acquisition de température de la figure 2.18.

montré également sur la figure 2.19 consiste à des mesures aller-retour sur 1km. Les mesures aller-retour sont presque identiques. Nous pouvons donc les moyenner pour lisser une partie du bruit de mesure. Nous constatons, sur le signal aller par exemple, qu'il existe une tendance générale qui est imposée par les phénomènes existants sur le terrain. Les quelques premiers mètres, correspondant à la traversée en contre bas de la fibre optique, ne seront pas pris en compte par les techniques de traitement (voir figure 2.19). Il faut mentionner ici que plusieurs fuites réelles ont été repérées lors des visites de terrain, et que ces fuites peuvent exister à des temps différents.

- La dernière partie de la fibre optique (appelée aussi fibre d'étalonnage) est utilisée pour

effectuer des mesures sur la piste cyclable. Comme précédemment, les mesures affichées également sur la figure 2.20 sont dans une boucle aller-retour. Si nous regardons la coupe de la digue (voir figure 2.15), nous constatons que la fibre se trouve sur un niveau qui est assez élevé par rapport au niveau du grand canal, ce qui signifie que les chances d'identifier une fuite sont relativement faibles.

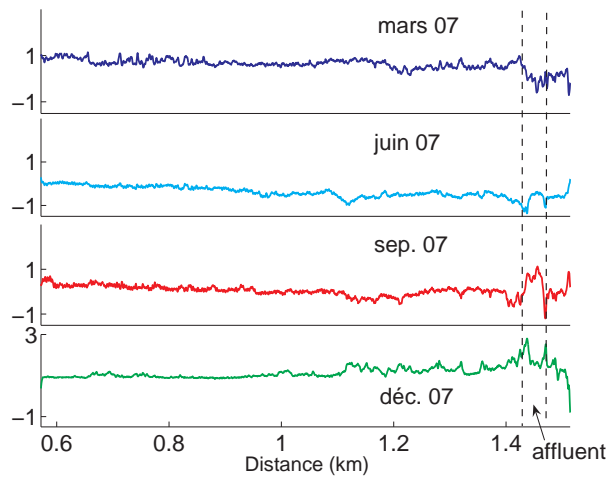
Il faut noter ici que les pics présents sur la figure 2.18 correspondent aux points de soudure de la fibre optique se situant dans le coffret optique (en bleu).



**Fig. 2.20:** Partie piste de l'acquisition de température de la figure 2.18.

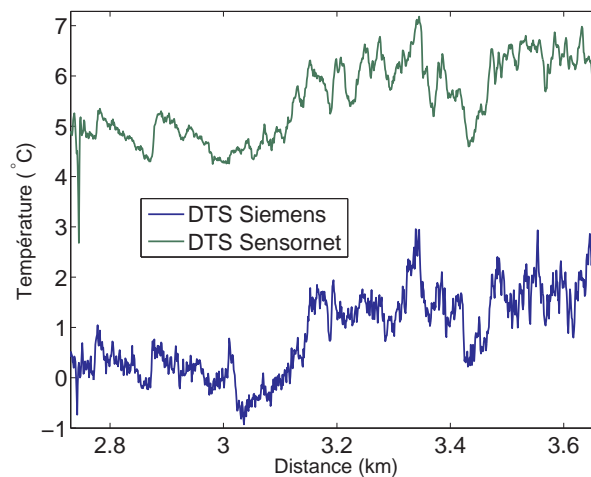
Afin de mettre en évidence les différents phénomènes qui peuvent exister d'une saison à l'autre, la figure 2.21 montre les moyennes des mesures de température aller-retour dans la zone de contre canal à 4 dates différentes en 2007 : 30 mars, 02 juin, 10 septembre et 19 décembre. Une première observation est que sur tous ces signaux la première partie est plus homogène que la deuxième et le bruit augmente avec la distance. En même temps, il existe des pics cohérents d'une saison à l'autre. Nous constatons qu'il existe des zones qui apparaissent sur une saison et disparaissent sur d'autres. Les fuites réelles ont généralement un comportement de ce type parce que leurs débits dépendent du niveau de l'eau du canal, de la température de l'eau qui agit sur l'étanchéité assurée par dalles de béton, etc.

Comme la plupart des mesures effectuées sur le site de Kembs sont réalisées avec le DTS Siemens pour lequel le RSB est plus grand que celle de DTS Sensornet, nous comparons deux signaux enregistrés par les deux appareils sur une zone de piste. Comme nous ne disposons pas de deux appareils en même temps sur le site, nous comparons les mesures prises le même jour (5 février) de deux années différentes : 2007 pour Sensornet et 2008 pour Siemens. Les signaux sont montrés dans la figure 2.22. Sachant que la résolution spatiale de DTS Siemens est de 0.76m et celui de DTS Sensornet est de 1m, nous avons fait un rééchantillonnage pour ramener les deux signaux



**Fig. 2.21:** Comparaison entre les acquisitions au cours des 4 périodes de l'année 2007 sur la fibre de mesure de contre canal. De haut à bas : 30 mars, 02 juin, 10 septembre et 19 décembre. Pour faciliter la comparaison les signaux ont été normalisés à une variance unitaire.

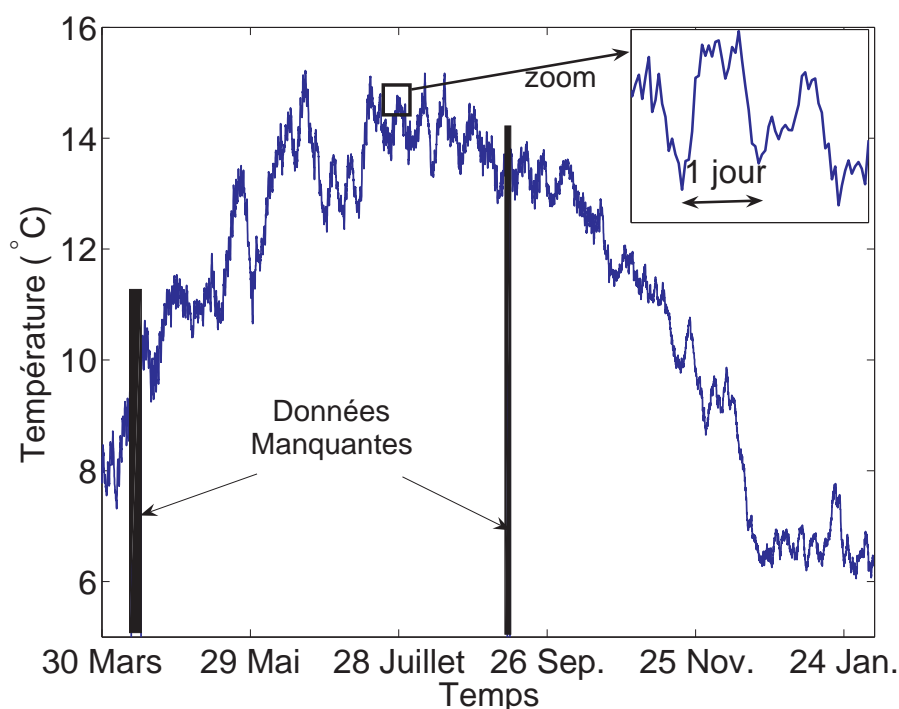
à la même résolution. Nous pouvons constater assez clairement que le signal enregistré par le DTS Siemens est plus bruité que celui enregistré par le DTS Sensornet (comme prévu).



**Fig. 2.22:** Deux mesures de température enregistrées par les deux DTS le 5 février 2007, respectivement 5 février 2008. Le signal enregistré par le DTS Siemens est plus bruité que celui enregistré par le DTS Sensornet.

### Le signal de température en fonction du temps

Sur le site de Kembs, nous disposons des enregistrements effectués courant 2007 et début 2008 en utilisant l'appareil DTS Siemens. Cependant, des acquisitions sont manquantes pendant quelques jours soit à cause des interventions sur le site, soit à cause d'un dysfonctionnement du système, soit pour prendre des mesures actives (en chauffant localement la fibre optique par un système de chauffage intégré). La figure 2.23 montre le signal enregistré sur une distance donnée dans la zone de contre canal pour la période 30 mars 2007 - 05 février 2008 (périodes manquantes hachurées en noir). La température suit une tendance sinusoidale sur l'année en correspondance avec la température de l'air et celle de l'eau. A la différence du site expérimental d'Oraison, les variations journalières sont plus nettes sur les données enregistrées. Ces variations sont présentes sur les mesures de contre canal parce que la température est fortement influencée par la température de l'eau.



**Fig. 2.23:** Évolution temporelle de la température à une distance donnée pour la fibre de mesure du contre canal. Les variations journalières sont plus visibles en raison de la forte influence de la température de l'eau du contre canal. (A comparer avec la figure 2.13).

## 2.5 Modélisation des signaux de température

Nous avons présenté dans les sections précédentes des exemples de données de température simulées et enregistrées sur deux sites expérimentaux étudiés. Une acquisition de température (un profil de température) correspond à une mesure le long de l'ouvrage avec un pas dépendant de la résolution spatiale de l'appareil de mesure. Cette résolution est typiquement de l'ordre de 1 m, ce qui signifie que toutes les anomalies ayant un support spatial supérieur à un mètre peuvent être détectables. Cependant, si plusieurs anomalies de support inférieur à 1 m sont dans le même voisinage, elles seront interprétées comme une seule anomalie.

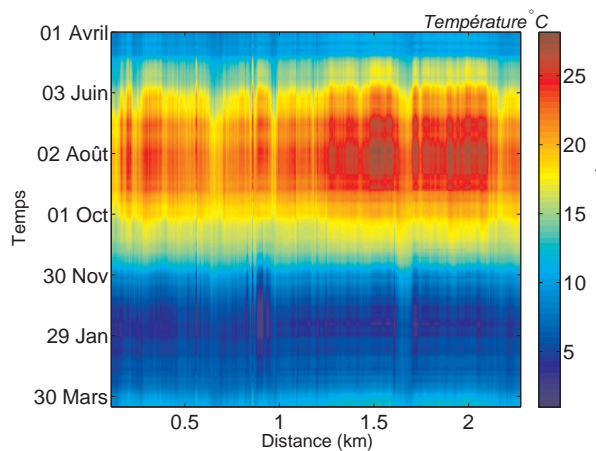
Pour pouvoir suivre l'évolution des anomalies au cours du temps, plusieurs enregistrements sont effectués avec une résolution temporelle prédéfinie. Il faut mentionner ici qu'une acquisition n'est pas instantanée mais elle résulte d'une intégration temporelle. Cette résolution temporelle est un des facteurs influençant le RSB : plus le temps d'intégration est grand plus le RSB est important.

Les données enregistrées par la fibre optique peuvent donc s'écrire comme un signal à deux dimensions,  $y(t, x)$ , en fonction du temps  $t$  (l'instant d'acquisition d'une mesure) et de la distance de mesure  $x$ . Ces données peuvent s'exprimer sous la forme matricielle suivante :

$$\mathbf{Y} = \{y(t, x) \mid 1 \leq t \leq N_t, 1 \leq x \leq N_x\}, \quad (2.11)$$

où  $N_x$  et  $N_t$  représentent respectivement le nombre de capteurs et le temps total d'acquisition. Pour des acquisitions par fibre optique,  $N_x$  est égal à la longueur de la fibre (en m) divisée par la résolution spatiale du DTS.

La figure 2.24 montre les données réelles enregistrées sur une année complète sur le site d'Oraison. Les différentes couleurs désignent les valeurs de température enregistrées en fonction de la



**Fig. 2.24:** Températures enregistrées sur une année complète sur le site d'Oraison. Les valeurs de température correspondent aux couleurs affichées sur l'échelle d'intensité.

distance et de la période d'acquisition.

Le principe des mesures de température réside dans le fait qu'un écoulement préférentiel d'eau dû à des fuites peut être mis en évidence en mesurant la différence de température entre l'eau du canal et celle du terrain dans lequel la fibre optique est enterrée.

Cependant, cet écart de température peut être également dû à d'autres facteurs comme :

- les structures existantes (drains, tunnels, regards), etc.,
- les variations saisonnières de température,
- le rayonnement solaire,
- les précipitations,
- les inhomogénéités du terrain.

En outre, le signal enregistré est fortement influencé par la réponse du sol dans lequel la fibre est enterrée. Cette réponse masque l'information pertinente due aux anomalies et doit être enlevée. D'autres facteurs secondaires comme le vent, l'humidité, etc. peuvent également avoir une influence sur le signal enregistré.

Pour résumer, les données de température enregistrées ne sont pas exploitables telles quelles et des techniques de traitement du signal doivent être développées afin de détecter les anomalies et plus particulièrement les fuites.

Une fois le signal enregistré, il peut être modélisé comme un mélange des facteurs définis 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. Par exemple, l'influence du rayonnement solaire dépend de l'angle d'incidence, du coefficient d'absorption du matériau, de la présence ou non et du type de nuages présents, etc. La relation entre la température mesurée par la fibre optique et le rayonnement solaire est donc une relation non-linéaire. Cependant, par rapport au signal de température mesuré, nous pouvons lui affecter une source de température (comme fonction du déplacement le long la fibre) qui dépendra du rayonnement solaire. Comme d'autres facteurs peuvent également modifier la température mesurée, nous pouvons assumer que les températures induites par chaque facteur viennent pondérer linéairement la température mesurée par la fibre optique. De plus, comme les différents facteurs énumérés ci-dessus sont physiquement indépendants, nous pouvons assumer que les sources des températures associées sont indépendantes.

La matrice de température peut alors s'exprimer comme :

$$\mathbf{Y} = \mathbf{MF} + \mathbf{B}, \quad (2.12)$$

où  $\mathbf{Y} \in \mathfrak{R}^{N_t \times N_x}$  est la matrice des données enregistrées,  $\mathbf{M} \in \mathfrak{R}^{N_t \times p}$  la matrice de mélange,  $\mathbf{F} \in \mathfrak{R}^{p \times N_x}$  la matrice de sources indépendantes induites par les facteurs et  $\mathbf{B} \in \mathfrak{R}^{N_t \times N_x}$  est la matrice de bruit de mesure supposé gaussien, de moyenne nulle, blanc spatialement et temporellement.

## 2.6 Prétraitements

Afin d’exploiter au mieux les données enregistrées par la fibre optique, des prétraitements peuvent être utilisés pour réduire l’influence de certains facteurs. Nous présentons dans cette section quelques prétraitements qui seront utilisés dans ce travail. Il faut cependant remarquer qu’ils ne sont pas nécessairement toujours tous appliqués et ils ne suivent pas toujours l’ordre dans lequel ils sont décrits ci-dessous.

### 2.6.1 Désaisonnalisation

Comme nous avons montré lors de la description des données simulées et réelles, le signal de température à une distance donnée suit une variation saisonnière. La désaisonnalisation consiste à retirer la composante saisonnière du signal brut  $\mathbf{Y}$  de température. Les courbes de température (voir figures 2.5, 2.13 et 2.23) montrent une dépendance “sinusoïdale” en fonction de la température. L’idée de la désaisonnalisation est alors d’enlever cette variation sinusoïdale. Une approche possible est de modéliser les variations annuelles ainsi que les variations saisonnières par des sinusoïdes 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 :

$$\mathbf{y}_{var}(t) = a_m + \sum_{n=1}^2 [a_{1n} \sin(ns) + a_{2n} \cos(ns)], \quad (2.13)$$

avec  $s = \frac{2\pi t}{365N_j}$ ,  $t = 1, \dots, N_t$  et  $N_j$  le nombre d’acquisitions par jour. Les différents paramètres  $a_m$ ,  $a_{1n}$  et  $a_{2n}$  sont estimés pour chaque distance à partir de données de température en utilisant une méthode des moindres carrés.

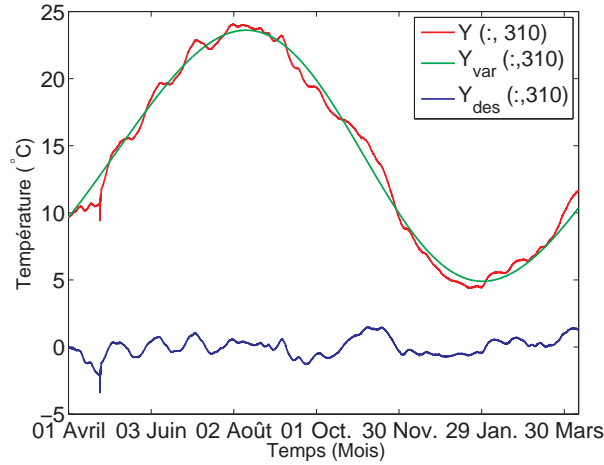
En notant par  $\mathbf{Y}_{var}$  les variations saisonnières estimées sur toutes les distances, les données désaisonnalisées peuvent s’écrire comme :

$$\mathbf{Y}_{des} = \mathbf{Y} - \mathbf{Y}_{var}. \quad (2.14)$$

La figure 2.25 montre un exemple de cette désaisonnalisation pour le site d’Oraison. Nous pouvons constater que le signal estimé  $\mathbf{y}_{var}(t)$  à la distance 310 m, noté aussi par rapport au format matriciel  $\mathbf{Y}_{var}(:, 310)$ , permet d’enlever la tendance générale sinusoïdale de la température enregistrée par la fibre,  $\mathbf{Y}(:, 310)$ . Le signal désaisonnalisé obtenu,  $\mathbf{Y}_{des}(:, 310)$ , est dépourvu de cette variation. Il faut cependant remarquer que cette désaisonnalisation permet d’enlever uniquement une tendance globale lente et n’est pas capable de modéliser les variations brusques autour de la température normale de saison. Le résultat met en évidence la dynamique du signal autour de la température normale de saison.

D’autres méthodes ont été également testées pour enlever cette variation comme : un filtrage passe-bas, des méthodes d’estimation polynomiale de la ligne de base souvent utilisées en spectroscopie [Mazet 05]. Le filtrage passe-bas ne sera pas utilisé car il peut enlever des informations utiles parmi lesquelles les fuites. En plus, il existe des paramètres à régler comme la fréquence





**Fig. 2.25:** Désaisonnalisation des données de température (en rouge) à la distance 310 m pour le site d'Oraison. La variation saisonnière (en vert) est estimée en posant un modèle sinusoïdal. Le signal désaisonné (en bleu) est dépourvu des variations annuelles et saisonnières de température.

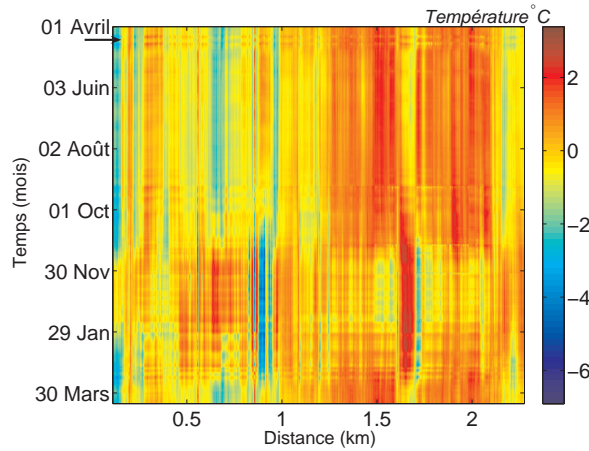
de coupure et l'ordre du filtre. L'estimation de la ligne de base fait appel à une estimation polynomiale en minimisant une fonction de coût non-quadratique. En spectroscopie Raman, où des pics énergiques et étroits sont présents, ces méthodes offrent certains avantages. En utilisant une fonction de coût symétrique, les résultats sont similaires à ceux obtenus par modélisation sinusoïdale. La fonction de coût asymétrique permet d'avoir de signaux à valeurs positives, mais la positivité n'as pas d'intérêt particulier pour ce travail. Bien d'autres méthodes, comme des méthodes basées sur des ondelettes, peuvent être utilisées, mais le choix de la modélisation sinusoïdale a été faite parce que c'est le modèle le plus proche de la réalité physique et pour la simplicité de mise en œuvre.

## 2.6.2 Normalisation des données

Un prétraitement couramment appliqué dans différentes applications est la normalisation des données, avec différentes définitions pour ce terme. Le but est souvent d'accorder la même importance à tous les signaux enregistrés. Pour les données thermométriques, cette normalisation peut être définie de manière à atténuer les variations d'une acquisition à l'autre. Cette normalisation est faite pour rendre chaque acquisition de moyenne nulle et de variance unité, c'est-à-dire pour chaque température enregistrée  $y(t, x)$  nous réalisons la transformation suivante :

$$y_{norm}(t, x) = \frac{y(t, x) - \bar{y}(t)}{\sigma_{y(t)}}, \quad (2.15)$$

où  $\bar{y}(t)$  est la moyenne du profil de température  $y(t, x)$  calculé à l'instant  $t$  et  $\sigma_{y(t)}$  la variance du profil après soustraction de la moyenne. Cette normalisation permet également d'atténuer les effets des variations saisonnières et journalières. Les données normalisées obtenues en appliquant



**Fig. 2.26:** Données de température de la figure 2.24 après normalisation.

la transformation ((2.15)) sur les données brutes de la figure 2.24 sont montrées sur la figure 2.26. Nous pouvons constater que les changements d’une journée à l’autre ou d’une saison à l’autre ont été fortement atténués par rapport aux données brutes. De plus, ce type de prétraitement permet de mieux mettre en évidence certaines particularités comme par exemple l’effet produit par des précipitations (mi avril, septembre, etc., phénomènes de rayure au cours du temps), les drains  $D1$  et  $D2$ , la singularité du terrain entre 1.6 et 1.7 km, etc.

### 2.6.3 Identification des phénomènes éphémères énergétiques

Les phénomènes éphémères énergétiques ne contiennent pas d’informations utiles pour la détection des fuites. Un exemple est la précipitation importante survenue le 10 avril (indiquée dans la figure 2.26 par une flèche) qui affecte tous les capteurs pendant approximativement deux jours. Cependant, ces phénomènes ne sont pas limités qu’aux précipitations, pouvant s’agir de tous les événements éphémères énergétiques en temps qui affectent une partie importante de la fibre. La localisation temporelle de phénomènes éphémères énergétiques est un prétraitement utile surtout pour les techniques de traitement mettant en évidence un changement de température.

Comme le montre la figure 2.26, les phénomènes éphémères, notamment les précipitations, induisent un changement dans les statistiques des données de température. L’outil le plus simple à mettre en œuvre est de mesurer les variances des profils de températures des données brutes. Cependant, cet outil peut rester inefficace lorsque les températures induites au niveau des capteurs par un phénomène ne sont pas sensiblement différentes des températures mesurées par la fibre optique avant l’apparition du phénomène. Pour cette raison, nous proposons dans ce travail d’utiliser les statistiques d’ordre supérieur (HOS - higher order statistics en anglais) [Kendall 63, Cullagh 87, Mendel 91, Lacoume 90]. Les statistiques d’ordre supérieur le plus connues sont celles d’ordre trois, le skewness, et d’ordre quatre, le kurtosis. Ces statistiques peuvent être définies en fonction des moments, mais souvent une définition faisant intervenir les

cumulants est utilisée (voir l'annexe C.1).

Le skewness et le kurtosis sont respectivement les cumulants d'ordre 3 et d'ordre 4 d'une variable aléatoire normalisée ou standardisée, où la standardisation revient à centrer la variable et à la normaliser par son écart type.

### Skewness

Le skewness,  $\mathcal{K}_{z(3)}$ , est le rapport entre le cumulant d'ordre 3,  $\kappa_{z(3)}$ , et le cumulant d'ordre 2,  $\kappa_{z(2)}$ , à la puissance 3/2 d'une variable aléatoire  $z$  :

$$\mathcal{K}_{z(3)} = \frac{\kappa_{z(3)}}{(\kappa_{z(2)})^{3/2}}. \quad (2.16)$$

Le skewness est une mesure caractérisant l'asymétrie de la distribution de probabilité de la variable aléatoire  $z$ . Ceci implique que toutes les variables avec une distribution symétrique ont un skewness nul, ce qui est le cas d'une variable gaussienne.

### Kurtosis

Le kurtosis de la variable aléatoire  $z$ ,  $\mathcal{K}_{z(4)}$ , est le rapport entre le cumulant d'ordre 4,  $\kappa_{z(4)}$ , et le cumulant d'ordre 2 au carré :

$$\mathcal{K}_{z(4)} = \frac{\kappa_{z(4)}}{(\kappa_{z(2)})^2}. \quad (2.17)$$

Le kurtosis est une mesure de l'aplatissement de la distribution de la variable aléatoire  $z$ . Pour une distribution gaussienne, le kurtosis est zéro, et dans la littérature le kurtosis est souvent référé comme une mesure de la non-gaussianité.

### Estimateurs de Skewness et de Kurtosis

Des estimateurs non biaisés des cumulants sont les  $k$ -statistiques [Kendall 63, chap. 12], [Cullagh 87, chap. 4] qui sont des fonctions symétriques caractérisant les distributions.

Dans notre application, en notant  $s_r(t) = \sum_{x=1}^{N_x} y^r(t, x)$  la somme des valeurs à la puissance  $r$  d'un profil de température à l'instant  $t$ , les premières  $k$ -statistiques sont :

$$\begin{aligned} \widehat{\kappa}_y(t)_{(2)} &= \frac{1}{N(N-1)} (Ns_2(t) - (s_1(t))^2) \\ \widehat{\kappa}_y(t)_{(3)} &= \frac{1}{N(N-1)(N-2)} (N^2s_3(t) - 3Ns_2(t)s_1(t) + 2(s_1(t))^3) \\ \widehat{\kappa}_y(t)_{(4)} &= \frac{1}{N(N-1)(N-2)(N-3)} ((N^3 + N^2)s_4(t) - 4(N^2 + N)s_3(t)s_1(t) - 3(N^2 - N)(s_2(t))^2 + \\ &\quad + 12Ns_2(t)s_1(t)^2 - 6(s_1(t))^4) \end{aligned} \quad (2.18)$$

Le skewness et le kurtosis sont estimés par les relations suivantes :

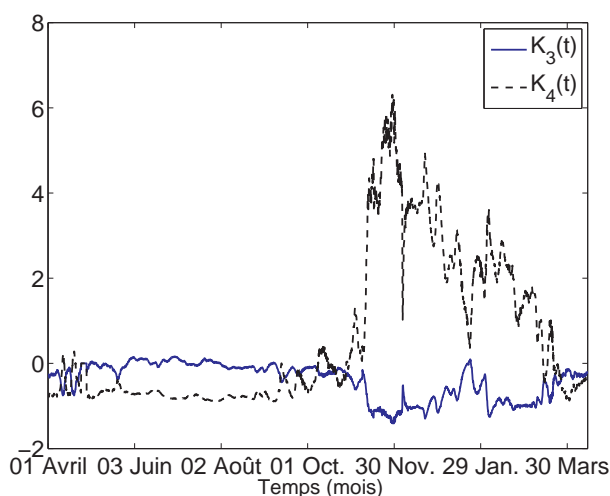
$$\widehat{\mathcal{K}}_{y(3)}(t) = \frac{\widehat{\kappa}_y(t)_{(3)}}{(\widehat{\kappa}_y(t)_{(2)})^{3/2}} \quad (2.19)$$

$$\widehat{\mathcal{K}}_{y(4)}(t) = \frac{\widehat{\kappa}_y(t)_{(4)}}{(\widehat{\kappa}_y(t)_{(2)})^2} \quad (2.20)$$

Les variances de ces estimateurs dépendent du nombre des échantillons utilisés lors du calcul de ces grandeurs, dans notre cas de  $N_x$ .

### Application sur les données thermométriques

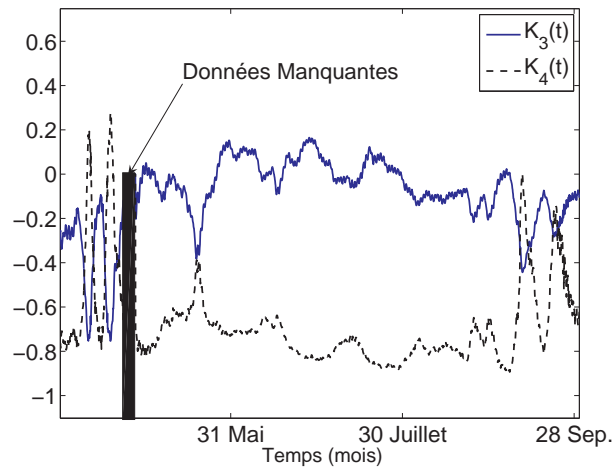
Le skewness et le kurtosis ne dépendent pas de la moyenne et de la variance des données et par conséquent nous pouvons calculer leurs estimateurs soit en fonction des données brutes  $y(t, x)$  comme présenté ci-dessus soit en fonction des données normalisées  $y_{norm}(t, x)$ . Afin d'identifier les phénomènes éphémères en temps, nous calculons le skewness et le kurtosis pour chaque acquisition de température sur tous les capteurs. Ceci nous donne le skewness et le kurtosis en fonction du temps, notés par la suite  $\mathcal{K}_3(t)$  et  $\mathcal{K}_4(t)$ . Ces deux statistiques calculées pour les données de la figure 2.26 sont affichées sur la figure 2.27 sur une année entre avril 2005 et avril 2006. Sur



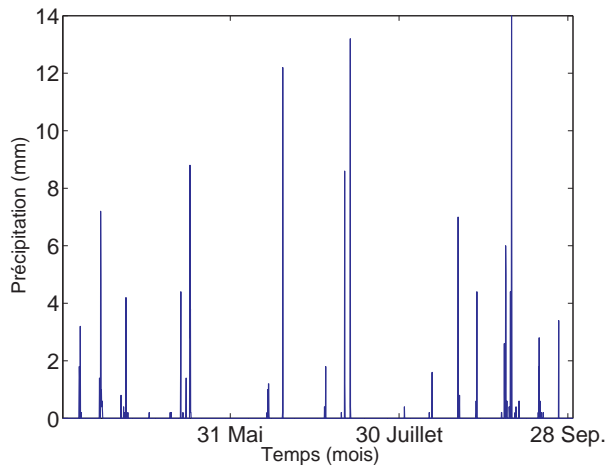
**Fig. 2.27:** Le skewness  $\mathcal{K}_3(t)$  et le kurtosis  $\mathcal{K}_4(t)$  des données de température de la figure 2.26.

ces courbes nous pouvons identifier deux zones : la première pour la période printemps-été et la deuxième pour automne-hiver.

En se focalisant premièrement sur la partie printemps-été, zone qui est zoomée dans la figure 2.28(a), nous observons que les changements de statistiques mis en évidence par le skewness et le kurtosis se trouvent aux mêmes instants de temps. Considérant par exemple le cas du kurtosis, il est presque constant sauf à des endroits particuliers. Ceci signifie que peu importe les variations de température durant la saison, du fait que tous les capteurs soient affectés de la même façon, la distribution de données reste relativement la même. Cependant, si un phénomène localisé en temps vient perturber tous les capteurs, la distribution de températures enregistrées par la fibre optique va changer. Un cas idéal par exemple est le cas d'une forte précipitation qui change



(a) Le skewness (en continu) et le kurtosis (en pointillé) pour la période printemps-été 2005



(b) Quantités des précipitations enregistrées sur un site proche de la digue d'Oraison

**Fig. 2.28:** Identification des phénomènes éphémères de type précipitation pour le site d'Oraison pendant la période printemps-été 2005. Les variations locales de skewness et de kurtosis se superposent bien avec les instances de précipitation.

toutes les mesures effectuées par tous les capteurs de telle manière que les capteurs mesurent non plus la température du sol mais la température de l'eau de la précipitation. Dans ce cas idéal, si nous supposons une distribution gaussienne par rapport à une température moyenne qui est la température de l'eau de la précipitation, le kurtosis va tendre vers 0. Considérant maintenant le cas réel, la distribution ne sera pas nécessairement gaussienne puisqu'elle dépend de plusieurs facteurs comme la puissance du phénomène, la manière dont les capteurs réagissent au phénomène (par exemple, sur le site d'Oraison, les capteurs étant disposés à deux niveaux d'élévation différents, ils ne réagissent pas de la même manière), etc. Cependant, l'occurrence d'un tel phénomène va modifier la distribution de données, ce qui va se traduire par un changement de

kurtosis et éventuellement de skewness. Pour vérifier ces propos, les quantités de précipitations enregistrées sur un site proche de la digue d’Oraison sont affichées sur la figure 2.28(b). Nous constatons que les pics (les variations locales) de skewness et de kurtosis se superposent bien avec les instances de précipitation.

Il faut mentionner ici qu’en plus de la quantité et de la température d’une précipitation, plusieurs effets comme la perméabilité du sol, la durée et l’intensité de la précipitation, etc., peuvent influencer différemment les températures enregistrées par la fibre optique. Par exemple, la pluie incessante avec moins d’intensité peut avoir plus d’influence qu’une pluie plus forte mais d’une courte durée.

Imaginons maintenant un phénomène énergétique qui induit une température de même ordre de grandeur que la température mesurée juste avant le phénomène. La distribution mesurée par les capteurs va changer, ce qui va modifier principalement le kurtosis, voir le skewness, mais la variance sera moins influencée. Au contraire, considérant par exemple une augmentation constante de température pendant 1 semaine, tout le sol va répondre relativement de la même façon ce qui sera mis en évidence principalement par la variance, le kurtosis restant presque inchangé puisque la distribution va rester approximativement la même. En réalité, le sol ne répond pas de la même façon et des petites variations vont apparaître sur le kurtosis. Pour résumer, les statistiques d’ordre supérieur peuvent être très utiles pour l’identification des phénomènes éphémères énergétiques en temps (précipitations, problèmes de mesures, etc.). Cependant, l’identification automatique dans de grands volumes de données nécessite l’utilisation de seuils pour les deux grandeurs (skewness et kurtosis).

## 2.7 Discussion

Dans ce chapitre, nous avons présenté différents aspects physiques de mesures thermométriques effectuées par les capteurs à fibre optique. Dans un premier temps, la simulation d’une digue à partir des températures obtenues par un modèle physique a été présentée. Ce modèle de digue permet de simuler des fuites sur différents capteurs et de les faire évoluer au cours du temps. Même si ces simulations sont issues d’un modèle physique, les conditions réelles sur le terrain peuvent être assez variables. Dans ce cadre, la description de deux sites expérimentaux différents l’un de l’autre nous a permis de montrer les facteurs pouvant influencer le signal de température.

Les données thermométriques ont été ensuite présentées dans un format matriciel en exploitant les variations spatiales et temporelles des données. Parmi les différents facteurs qui influencent la température enregistrée nous retrouvons : la réponse du sol dans lequel la fibre optique est enterrée, les structures existantes, les singularités de terrain, les facteurs climatiques comme les précipitations et les fuites. La matrice de température peut alors s’exprimer comme un mélange linéaire des sources indépendantes de température, c’est-à-dire les températures induites par les facteurs vues comme des fonctions du déplacement le long la fibre.

Des prétraitements pouvant être utilisés pour réduire l'influence de certains facteurs ont été ensuite étudiés. Les données de température montrent des variations sinusoïdales aux cours de l'année et des saisons. Une procédure pour désaisonnaliser ces données par moindres carrées permet d'enlever les variations saisonnières et annuelles des données. Les événements énergétiques éphémères en temps comme les précipitations affectent la plus part des capteurs et risquent de perturber l'analyse et le traitement des données. Les statistiques d'ordre supérieur, le skeweness et le kurtosis, permet d'identifier ces phénomènes.

Ayant développé une compréhension de ce qui constitue les données de température et les différents facteurs pouvant les influencer, nous allons nous intéresser dans les chapitres suivants aux méthodes de traitement développées pour la détection des fuites à partir de données de température.

# Chapitre 3





## Chapitre 3

# Séparation de sources pour la détection des fuites

### Sommaire

---

|            |   |            |
|------------|---|------------|
| <b>3.1</b> | <b>Introduction</b>   | <b>81</b>  |
| <b>3.2</b> | <b>Décomposition en valeurs singulières</b>                       | <b>82</b>  |
| 3.2.1      | Définitions et propriétés   | 82         |
| 3.2.2      | Interprétation de la SVD d'un jeu de données thermométriques      | 84         |
| <b>3.3</b> | <b>L'analyse en composantes indépendantes</b>                     | <b>87</b>  |
| 3.3.1      | Indépendance statistique  | 87         |
| 3.3.2      | Définition de l'ICA et ses indéterminations                       | 88         |
| 3.3.3      | Prétraitements  | 89         |
| 3.3.4      | Mesures d'indépendance  | 90         |
| 3.3.5      | Séparation de sources en mode symétrique et en mode déflation     | 94         |
| 3.3.6      | Application aux données réelles                                   | 95         |
| <b>3.4</b> | <b>Méthodologie proposée pour la détection des fuites</b>         | <b>96</b>  |
| <b>3.5</b> | <b>Application aux données simulées</b>                           | <b>99</b>  |
| 3.5.1      | Analyse de données entières                                       | 102        |
| 3.5.2      | Analyse par fenêtrage glissant en temps                           | 103        |
| 3.5.3      | Interprétation des résidus  | 108        |
| 3.5.4      | Étude des résidus par rapport aux différents paramètres           | 109        |
| <b>3.6</b> | <b>Application aux données enregistrées sur le site d'Oraison</b> | <b>112</b> |
| 3.6.1      | Analyse de données entières                                       | 114        |
| 3.6.2      | Analyse par fenêtrage glissant en temps                           | 115        |
| <b>3.7</b> | <b>Application aux données enregistrées sur le site de Kembs</b>  | <b>122</b> |
| 3.7.1      | Analyse de données entières                                       | 124        |
| 3.7.2      | Analyse par fenêtrage glissant en temps                           | 124        |
| <b>3.8</b> | <b>Discussion</b>   | <b>129</b> |

---



### 3.1 Introduction

Dans le précédent chapitre, nous avons introduit la modélisation des données thermométriques et nous avons montré qu'une fuite d'eau dans une digue en terre génère un changement de la température mesurée par la fibre optique. Cependant, ce ne sont pas seulement les fuites qui influencent cette température, ce qui signifie que les données acquises ne seront pas directement interprétables pour la détection de fuites.

Il existe plusieurs facteurs qui peuvent influencer la température mesurée à une distance et un temps donnés, comme :

- l'influence du terrain dans lequel la fibre est installée,
- les variations annuelles et saisonnières de température ambiante,
- les précipitations et autres phénomènes éphémères enregistrés par les capteurs,
- les fuites d'eau,
- les structures existantes (drains, tunnels, regards), etc.

Les données enregistrées  $\mathbf{Y}$  ont été modélisées par l'équation (2.12) comme un mélange de sources associées à ces différents facteurs, avec les sources comme des fonctions de la distance  $x$ . Même si les différents facteurs peuvent avoir un effet non-linéaire sur les températures mesurées, nous avons choisi un modèle linéaire en considérant uniquement les valeurs de température induites par ces facteurs sans vouloir quantifier les facteurs qui les ont produits.

En outre, les différents phénomènes pouvant être considérés comme indépendants parce que l'occurrence de l'un n'affecte pas celle de l'autre, les sources associées sont considérées comme indépendantes. Le problème de détection de fuites devient alors un problème de séparation de sources indépendantes à partir d'un mélange linéaire.

A partir de l'équation (2.12), il faut donc identifier les matrices  $\mathbf{M} \in \mathfrak{R}^{N_t \times p}$  et  $\mathbf{F} \in \mathfrak{R}^{N_x \times p}$  à partir de la matrice d'observations,  $\mathbf{Y} \in \mathfrak{R}^{N_t \times N_x}$ , avec la seule hypothèse d'indépendance de sources. Par rapport à un problème classique de séparation de sources, il faut souligner que la source recherchée pour notre étude est celle liée à la fuite et que celle-ci n'est pas forcément la plus énergétique. Les techniques les plus communément utilisées pour ce type de problème comprennent la décomposition en valeurs singulières et l'analyse en composantes indépendantes.

Nous proposons un algorithme pour la détection des fuites basé sur ces deux techniques en utilisant les différents prétraitements présentés dans le chapitre antérieur. La première partie de ce chapitre est alors réservée à la présentation de ces techniques. Elle sera suivie par la description de l'algorithme proposé. L'application de ce schéma de traitement sur les différents jeux de données présentés en chapitre 2 fera l'objet de la troisième partie de ce chapitre. Une discussion compréhensive conclura ce chapitre.

## 3.2 Décomposition en valeurs singulières

Dans un premier temps, nous présentons la définition de la décomposition en valeurs singulières<sup>1</sup> (SVD) pour une matrice quelconque. Ensuite, les différentes façons d’interpréter cette décomposition seront discutées sur quelques exemples. Les données thermométriques constituant des matrices réelles, nous nous focalisons sur le cas réel, l’extension au cas complexe étant simple mais redondante ici.

### 3.2.1 Définitions et propriétés

Soit  $\mathbf{Y} \in \mathbb{R}^{N_t \times N_x}$  une matrice réelle. Il existe deux matrices orthogonales [Golub 89] :

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_N] \in \mathbb{R}^{N_t \times N} & \text{telle que} & \quad \mathbf{U}\mathbf{U}^T = \mathbf{I}_{N_t}, \\ \mathbf{V} &= [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_N] \in \mathbb{R}^{N_x \times N} & \text{telle que} & \quad \mathbf{V}\mathbf{V}^T = \mathbf{I}_{N_x}, \end{aligned} \quad (3.1)$$

pour lesquelles :

$$\mathbf{U}^T \mathbf{Y} \mathbf{V} = \mathbf{\Sigma} \in \mathbb{R}^{N \times N}, \quad (3.2)$$

avec  $\mathbf{\Sigma}$  une matrice diagonale :

$$\mathbf{\Sigma} \stackrel{\text{not}}{=} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N) \quad (3.3)$$

dont les éléments sont ordonnés en ordre décroissant, c’est-à-dire,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$ , avec  $N = \min\{N_t, N_x\}$  et  $\mathbf{I}_N$  la matrice identité de dimension  $N \times N$ .

La relation (3.2) est dénommée décomposition en valeurs singulières de la matrice  $\mathbf{Y}$ . Utilisant la propriété d’orthogonalité des matrices  $\mathbf{U}$  et  $\mathbf{V}$ , nous pouvons écrire :

$$\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (3.4)$$

Les vecteurs  $\mathbf{u}_k \in \mathbb{R}^{N_t}$ ,  $k = 1 \dots N$ , sont dénommés vecteurs singuliers gauches tandis que les vecteurs  $\mathbf{v}_k \in \mathbb{R}^{N_x}$ ,  $k = 1 \dots N$ , sont dénommés vecteurs singuliers droits. Les éléments  $\sigma_k$ ,  $k = 1 \dots N$ , de la matrice diagonale  $\mathbf{\Sigma}$  sont appelés valeurs singulières.

Le rang d’une matrice est défini en termes des valeurs singulières non nulles qu’elle possède. Si, par exemple, les valeurs singulières de la matrice  $\mathbf{Y}$  satisfassent l’inégalité :

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > (\sigma_{r+1} = \dots = \sigma_N = 0), \quad (3.5)$$

la matrice  $\mathbf{Y}$  est dite de rang  $r$ . Dans ce cas, les “ $r$ ” premiers vecteurs singuliers et les “ $r$ ” premières valeurs singulières caractérisent complètement la matrice  $\mathbf{Y}$ , c’est-à-dire,

$$\mathbf{Y} = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T. \quad (3.6)$$

---

<sup>1</sup>Singular Value Decomposition (SVD) en anglais. Par la suite, nous utiliserons l’acronyme SVD en faisant référence à la décomposition en valeurs singulières.

Les vecteurs singuliers gauches  $\mathbf{u}_k$  sont les vecteurs propres du produit extérieur  $\mathbf{Y}\mathbf{Y}^T$  alors que les vecteurs singuliers droits  $\mathbf{v}_k$  sont les vecteurs propres du produit intérieur  $\mathbf{Y}^T\mathbf{Y}$ . La SVD diagonalise donc simultanément le produit extérieur et le produit intérieur.

Notons  $\langle \mathbf{Y} \rangle$  l'espace vectoriel de la matrice  $\mathbf{Y}$ . Si les “ $r$ ” valeurs singulières non-nulles sont différentes, cet espace vectoriel peut être décrit par la somme de  $r$  sous-espaces  $\langle \mathbf{Y}_k \rangle$ ;  $k = 1, \dots, r$  orthogonaux entre eux [Le Bihan 01] :

$$\langle \mathbf{Y} \rangle = \langle \mathbf{Y}_1 \rangle \oplus \langle \mathbf{Y}_2 \rangle \oplus \dots \oplus \langle \mathbf{Y}_r \rangle \quad (3.7)$$

où  $\oplus$  est la somme directe d'espaces vectoriels. Cette propriété de la décomposition en sous-espace orthogonaux est assez utile dans le contexte de séparation de sources. Ces méthodes de séparation sont souvent nommées méthodes de séparation par sous-espaces. L'espace initial est souvent décomposé en deux sous-espaces orthogonaux entre eux donnés par la relation  $\langle \mathbf{Y} \rangle = \langle \mathbf{Y}_1 \rangle \oplus \langle \mathbf{Y}_2 \rangle$ . Le premier sous-espace est construit avec les  $m$  premières valeurs singulières et le deuxième avec les  $r - m$  valeurs singulières restantes.

Supposons que nous voulons obtenir une approximation de la matrice  $\mathbf{Y}$  de rang  $r$  par une matrice  $\mathbf{Y}_m$  de rang  $m$  inférieur à  $r$ . La meilleure approximation au sens des moindres carrés est obtenue par la troncature de la SVD de  $\mathbf{Y}$  jusqu'au  $m$  premières valeurs singulières [Scharf 91] :

$$\mathbf{Y}_m = \sum_{k=1}^m \sigma_k \mathbf{u}_k \mathbf{v}_k^T. \quad (3.8)$$

La norme de l'écart entre la matrice de départ et son approximation par une matrice de rang inférieur est :

$$\|\mathbf{Y} - \mathbf{Y}_m\| = \sum_{k=m+1}^r \sigma_k^2. \quad (3.9)$$

La SVD de  $\mathbf{Y}$  peut être aussi interprétée comme la décomposition qui produit une séquence d'approximations de rangs successifs,  $\mathbf{Y}_i = \mathbf{U}\boldsymbol{\Sigma}_i\mathbf{V}^T$ , où  $\boldsymbol{\Sigma}_i$  est la version de rang  $i$  de  $\boldsymbol{\Sigma}$  obtenue en mettant les dernières  $r - i$  valeurs singulières à zéro [Scharf 91].

### Interprétation de la SVD

A partir de la définition et des différentes propriétés associées, la SVD peut être interprétée de différentes façons comme :

- Nous pouvons considérer la SVD comme une méthode qui permet de transformer des variables corrélées en des variables non-corrélées. Ceci permet de mieux exploiter la relation entre les différentes composantes des données. Cette décorrélation est due au fait que les vecteurs singuliers sont orthogonaux.
- La SVD est une méthode efficace pour identifier et ordonner les dimensions le long desquelles les données montrent le maximum de variation. Les valeurs singulières d'une matrice renseignent quelles directions ou dimensions sont les plus importantes.

- Le troisième point de vue, relié au deuxième, est qu’une fois le maximum de variation identifié (le sens du maximum de variation), il est possible de retrouver la meilleure approximation des données par une matrice de rang inférieur. Cette propriété dite de réduction de dimensionnalité est exploitée dans différentes applications exigeant une réduction de données.

La plupart du temps, l’utilisation de la SVD a pour but de traiter des jeux de données de dimension élevée ayant une forte variabilité. La SVD permet de représenter les données dans un espace de dimension réduite qui révèle les sous-structures des données originales d’une façon plus claire et qui les ordonne par rapport à leur variabilité.

### Lien entre la SVD et l’analyse en composantes principales

Il existe un lien direct entre la SVD et l’analyse en composantes principales (ACP) quand les composantes principales sont calculées à partir de la matrice de covariance. Cette dernière méthode est connue aussi comme la transformée de Karhunen-Loeve. L’ACP d’une matrice  $\mathbf{Y}$  qui a pour matrice de corrélation (en considérant les données déjà centrées) la matrice  $\mathbf{\Gamma}_Y = E(\mathbf{Y}\mathbf{Y}^T)$ , est une paire de matrices  $\{\mathbf{U}, \mathbf{\Sigma}^2\}$  telle que la matrice de corrélation se factorise [Comon 94] :

$$\mathbf{\Gamma}_Y = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T \quad (3.10)$$

avec  $\mathbf{\Sigma}^2$  une matrice diagonale positive et  $\mathbf{U}$  une matrice orthogonale. En multipliant l’équation (3.4) par  $\mathbf{Y}^T$  à droite, nous obtenons :

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T. \quad (3.11)$$

Or, puisque  $\mathbf{Y}\mathbf{Y}^T$  est une matrice symétrique, elle est diagonalisable dans une base orthonormale de vecteurs propres, avec la décomposition écrite comme :

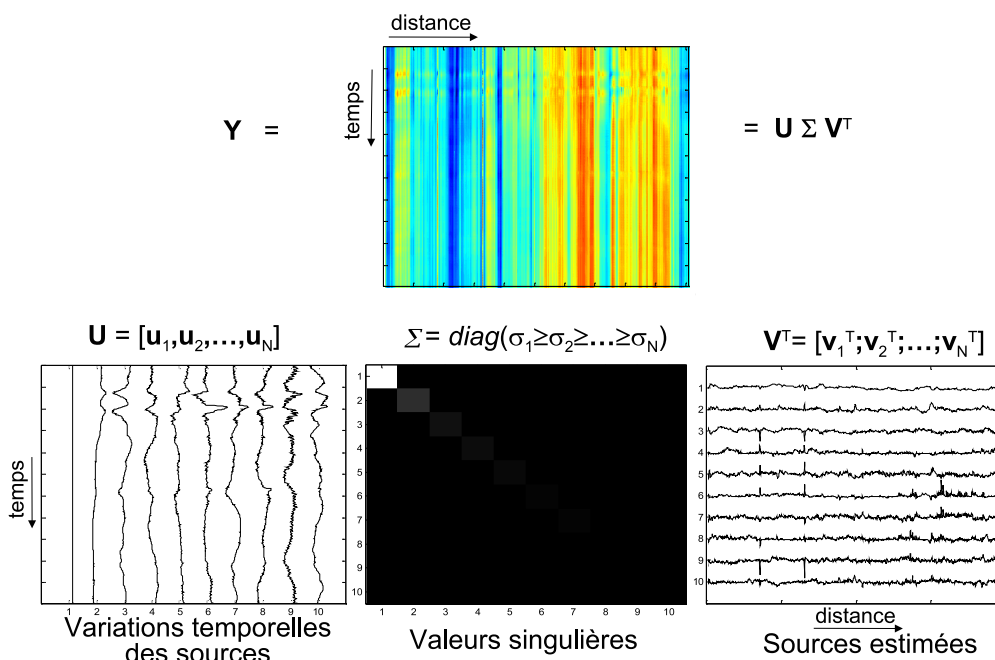
$$\mathbf{Y}\mathbf{Y}^T = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T, \quad (3.12)$$

où  $\mathbf{W}$  est une matrice orthogonale et  $\mathbf{\Lambda}$  est une matrice diagonale. Une comparaison entre (3.11) et (3.12) montre que  $\mathbf{\Sigma}^2 = \mathbf{\Lambda}$  et  $\mathbf{W} = \mathbf{U}$ . Autrement dit  $\sigma_k = \sqrt{\lambda_k}$ , où  $\lambda_k$  sont les valeurs propres et  $(\mathbf{W}, \mathbf{\Lambda})$  représente la décomposition en valeurs propres de la matrice  $\mathbf{Y}\mathbf{Y}^T$ . Les vecteurs singuliers quant à eux sont équivalents aux composantes principales. La matrice  $\mathbf{\Sigma}\mathbf{V}^T$  contient les fonctions transformées de composantes principales (principal component scores en anglais) [Diamantaras 96].

### 3.2.2 Interprétation de la SVD d’un jeu de données thermométriques

Soit  $\mathbf{Y} \in \mathbb{R}^{N_t \times N_x}$  un jeu de données thermométriques comprenant  $N_t$  acquisitions enregistrées chacune sur  $N_x$  distances. La SVD de cette matrice est donnée par l’équation (3.4).

Les matrices  $\mathbf{U}$  et  $\mathbf{V}$  peuvent être interprétées dans ce cas de la manière suivante :



**Fig. 3.1:** Les différentes composantes de la SVD appliquée sur un jeu de données thermométriques. Uniquement les 10 premières valeurs singulières et les vecteurs correspondant sont affichés ici.

- $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_N]$  est une matrice orthogonale de dimensions  $(N_x, N)$ . 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 définies par les facteurs mentionnés à la section 3.1. Comme les vecteurs sont orthonormés par construction, les sources estimées sont décorrélatées et normalisées [Vrabie 04].
- $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_N]$  est une matrice orthogonale de dimensions  $(N_t, N)$ . Les vecteurs  $\mathbf{u}_k \in \mathbb{R}^{N_t}$  donnant l'amplitude dans le cas des signaux réels, peuvent être interprétés comme les variations des sources estimées  $\mathbf{v}_k$  au cours du temps.

La décomposition des données thermométriques est représentée schématiquement sur la figure 3.1. La matrice  $\Sigma$  est représentée comme une image noir et blanc, les valeurs  $\sigma_k$  affichées sur sa diagonale étant codées entre blanc (la valeur la plus grande) et noir (zéro).

Une propriété importante de la SVD est sa capacité à décomposer les données en sous-espaces orthogonaux, (cf. équation (3.7)). Un sous-espace est construit à l'aide d'une ou plusieurs sources, ses variations temporelles et les valeurs singulières correspondantes, comme expliqué sur la figure 3.2. Nous avons choisi ici d'afficher les valeurs singulières (la diagonale de la matrice  $\Sigma$ ) sous la forme d'une courbe. Même si la SVD n'arrivera pas à bien séparer les différents phénomènes, elle permet de décomposer les données initiales en deux sous-espaces complémentaires, le sous-espace

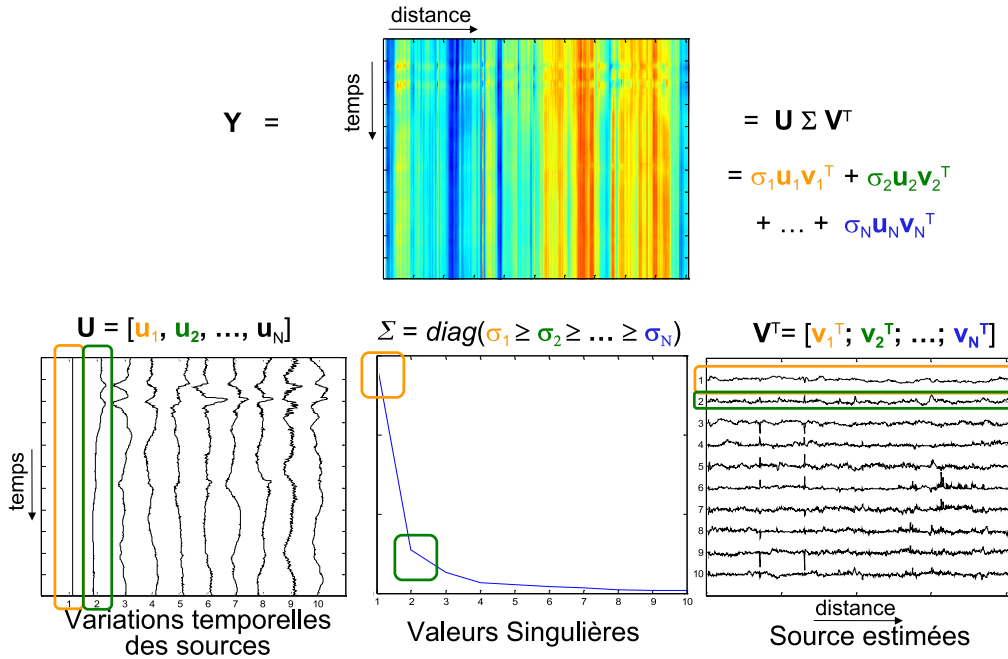


signal et le sous-espace résidu, définis par :

$$\mathbf{Y} = \mathbf{Y}_{sig} + \mathbf{Y}_{résidu} = \sum_{k=1}^m \sigma_k \mathbf{u}_k \mathbf{v}_k^T + \sum_{k=m+1}^N \sigma_k \mathbf{u}_k \mathbf{v}_k^T. \quad (3.13)$$

Ces deux sous-espaces sont orthogonaux, le premier est de dimension  $m$  et le deuxième de dimension  $N - m$ . Le paramètre critique est donc “ $m$ ”, le nombre de valeurs singulières à retenir pour construire le sous-espace signal. Cette décision est souvent basée sur l’observation de la pente de décroissance des valeurs singulières,  $\sigma_k$ , permettant de garder les valeurs singulières les plus importantes [Bihan 06]. Il existe des critères Akaike [Akaike 73], BIC [Schwarz 78], etc., qui sont très sensibles au RSB et ne donnent généralement pas de bons résultats sur les données réelles. Nous allons étudier dans la section 3.4 comment cette décomposition peut être utile pour extraire la réponse du sol dans lequel la fibre optique est enterrée.

Les sources estimées par la SVD sont décorrélées. La décorrélation permet d’extraire des sources



**Fig. 3.2:** Décomposition en sous-espaces. Un sous-espace pourra être construit en utilisant une ou plusieurs valeurs singulières et leurs vecteurs correspondants.

gaussiennes, ce qui est insuffisant dans notre cas puisque plusieurs facteurs (comme les drains ou les fuites) peuvent être modélisés par des sources parcimonieuses et donc non-gaussiennes. Dans la décomposition (3.4), les vecteurs  $\mathbf{u}_k$  sont eux aussi orthogonaux. Ces vecteurs représentent les variations temporelles des sources, mais il n’existe aucune justification physique à ce que ces vecteurs soient orthogonaux. De plus, comme expliqué ci-dessus, la contrainte d’orthogonalité imposée aux sources estimées peut s’avérer insuffisante pour estimer des sources qui caractérisent les drains et les fuites. Une technique plus réaliste basée non plus sur l’orthogonalité mais sur

l'indépendance de sources à estimer permet de s'affranchir des contraintes limitatives de la SVD. Cette technique est l'analyse en composantes indépendantes [Comon 94, Hyvärinen 99b, Hyvärinen 01, Common 07].

### 3.3 L'analyse en composantes indépendantes

Depuis son introduction dans les années 80, l'analyse en composantes indépendantes<sup>2</sup> (ICA) a été utilisée dans divers domaines d'application [Hyvärinen 01, Deville 07]. L'ICA est une technique statistique utilisée pour révéler les facteurs internes qui renseignent sur la structure des mesures ou des signaux. En traitement du signal, selon le contexte, l'ICA est utilisée pour la séparation aveugle des sources, l'extraction des caractéristiques, la déconvolution aveugle, comme une première étape dans le problème de localisation de sources, etc. [Pham 96, Lee 98, Cardoso 98].

Plaçons nous dans le cadre des variables aléatoires observées et alors connues  $\mathbf{z} = \{z_1, z_2, \dots, z_{N_t}\}$ . Le terme *connues* traduit la connaissance d'une réalisation des processus  $z_i$ . Nous n'avons nécessairement pas une connaissance de leurs densités de probabilité et seules leurs statistiques sont estimables. Selon l'hypothèse d'un mélange linéaire instantané, nous supposons que ces variables ont été générées à partir des variables latentes  $\mathbf{s} = \{s_1, s_2, \dots, s_p\}$  selon le modèle :

$$\mathbf{z} = \mathbf{A}\mathbf{s}, \quad (3.14)$$

où  $\mathbf{A} \in \mathfrak{R}^{N_t \times p}$  est une matrice de mélange.

Les variables observées  $z_i$  sont donc générées par des combinaisons linéaires de  $p$  variables aléatoires inconnues  $s_j$ , *inconnues* signifiant qu'aucune information n'est disponible sur ces variables. Ces variables sont appelées sources puisqu'elles sont à l'origine des variables aléatoires observées. L'ICA est une technique de séparation aveugle de sources (BSS pour Blind Source Separation en anglais), ce qui signifie que le mélange et les sources sont inconnus. Le but de l'ICA est d'estimer les sources ainsi que la matrice de mélange à partir des observations. Cette estimation se fait sur l'hypothèse de l'indépendance des sources. Les données thermométriques confirmant ce modèle, cette section est consacrée à une brève présentation de l'aspect théorique de l'ICA afin de pouvoir l'exploiter pour la détection des fuites.

#### 3.3.1 Indépendance statistique

L'hypothèse de base pour l'ICA est l'indépendance des sources. Ceci mérite de rappeler la notion de l'indépendance des variables aléatoires.

Pour un ensemble de variables aléatoires, la notion de l'indépendance signifie que des connaissances sur certaines variables ne renseignent pas en aucune façon sur le reste des variables de l'ensemble [Papoulis 91]. Dans une formulation plus formelle, l'indépendance des variables aléatoires  $\mathbf{s}_i$ ,  $i = 1, \dots, p$ , signifie que la densité de probabilité conjointe,  $f(\mathbf{s})$ , supposant qu'elle

<sup>2</sup>Independent Component Analysis (ICA) en anglais. Par la suite, nous utiliserons l'acronyme ICA en faisant référence à l'analyse en composantes indépendantes.

existe, peut s'écrire comme un produit de densité de probabilité marginale de chacune des variables :

$$f(\mathbf{s}) = \prod_{i=1}^p f(\mathbf{s}_i). \quad (3.15)$$

Il ne faut pas confondre l'indépendance avec la décorrélation, qui est moins exigeante que l'indépendance. La décorrélation de deux variables aléatoires, qui est une statistique d'ordre 2, s'écrit :

$$E(\mathbf{s}_i \mathbf{s}_j) - E(\mathbf{s}_i) E(\mathbf{s}_j) = 0, \quad (3.16)$$

où  $E(\cdot)$  dénote l'espérance mathématique, tandis que l'indépendance peut également s'exprimer sous la forme :

$$E(g_1(\mathbf{s}_i)g_2(\mathbf{s}_j)) - E(g_1(\mathbf{s}_i)) E(g_2(\mathbf{s}_j)) = 0, \quad (3.17)$$

pour  $i \neq j$  et pour toutes fonctions  $g_1$  et  $g_2$  [Papoulis 91]. En comparant (3.17) et (3.16), il est évident que toutes les variables indépendantes sont décorrélées, pourtant la décorrélation n'implique pas l'indépendance.

Une limitation de l'ICA est que les sources à estimer doivent être non-gaussiennes. L'indépendance n'apporte aucune information complémentaire puisque l'indépendance des variables aléatoires gaussiennes se résume à leur décorrélation et le mieux qu'on pourra faire est d'avoir une estimation du modèle ICA à une transformation orthogonale ce qui revient à un blanchiment [Hyvärinen 01].

Pour les données de température, les sources proviennent des facteurs indépendants et les sources associées aux drains et aux fuites sont non-gaussiennes. Il faut remarquer ici que l'ICA pose la contrainte de non-gaussianité des sources mais ne demande pas la connaissance exacte de la densité de probabilité des sources.

### 3.3.2 Définition de l'ICA et ses indéterminations

L'ICA constitue une famille des méthodes basées sur les mêmes principes. D'une façon générale, l'ICA peut être définie de la manière suivante [Hyvärinen 99b] : Soit un vecteur aléatoire  $\mathbf{z} \in \mathfrak{R}^{N_t}$ . L'ICA consiste à estimer le modèle génératif de l'équation (3.14) en déterminant une transformation linéaire  $\mathbf{s} = \mathbf{W}\mathbf{z}$  de telle manière que les composantes de  $\mathbf{s}$  soit les plus indépendantes possibles par maximisation d'une mesure de l'indépendance statistique  $G(\mathbf{s})$ . La matrice  $\mathbf{W} \in \mathfrak{R}^{p \times N_t}$  est dénommée matrice de séparation.

La notion de fonction  $G(\mathbf{s})$  qui maximise l'indépendance statistique est importante du point de vue pratique parce que sans connaître les densités de probabilité des sources, la définition exacte de l'indépendance est inutilisable. Dans ce cadre, différentes approches existent pour maximiser cette indépendance, ce qui sera discuté par la suite.

L'ICA permet d'estimer au plus une source gaussienne. De plus, le nombre de sources à estimer, noté ici " $p$ ", doit être inférieur aux nombres des mélanges  $N_t$  disponibles. Dans le cas contraire, le mélange est dit sous-déterminé et la matrice de mélange  $\mathbf{A}$  n'est pas inversible sans d'autres

*a priori* sur les sources. Dans notre cas, le problème est bien posé puisque nous disposons de plusieurs acquisitions dont le nombre est plus grand que le nombre de sources à estimer.

L'ICA a quelques indéterminations énumérées ci-dessous :

1. Dans le cas où aucune information *a priori* sur le mélange n'est disponible, une identification complète du modèle ICA est impossible. Considérant l'équation (3.14), la multiplication d'une source par une constante peut être annulée par la division de la colonne correspondante de la matrice de mélange, sans influencer les observations. Cette ambiguïté dite indétermination d'échelle n'a pas d'influence sur la forme des sources. Souvent elle est évitée en supposant les sources de variance unité. Elle peut être également évitée en fixant la norme de chaque colonne de la matrice de mélange à l'unité ou en supposant les éléments diagonaux de la matrice de mélange égaux à 1. Le signe de chaque source est également impossible à déterminer, mais ce signe se retrouve dans le vecteur de mélange associé.
2. L'ordre des sources n'est pas déterminable, les sources ne seront identifiables qu'à une permutation près. Une matrice de permutation avec son inverse ajouté au modèle de l'ICA n'affecte pas les données observées, d'où cette indétermination. Il est cependant possible d'ordonner les sources par post traitement associé à un critère quelconque comme, par exemple, le kurtosis de sources estimées ou l'énergie des sous-espaces engendrés par chaque source [Vrabie 04].

### 3.3.3 Prétraitements

Les deux prétraitements souvent effectués avant l'application de l'ICA sont le centrage des données pour la simplification des calculs et le blanchiment qui permet la décorrélation.

#### Centrage

Le centrage permet à simplifier les calculs sans avoir aucune incidence sur la matrice de mélange. Par exemple, lors de la résolution de l'ICA en utilisant les cumulants croisés, les relations sont beaucoup plus simples lorsque les variables sont centrées. La normalisation décrite dans le chapitre 2 prend en compte le centrage des données. Pour un vecteur aléatoire  $\mathbf{y}$ , l'étape de centrage donne le vecteur  $\mathbf{y}^c$  :

$$\mathbf{y}^c = \mathbf{y} - E(\mathbf{y}). \quad (3.18)$$

Le centrage de données de température revient à enlever la moyenne de chaque acquisition, donc de chaque ligne de la matrice  $\mathbf{Y}$  du modèle (2.12).

### Blanchiment

Le blanchiment consiste à transformer linéairement les données centrées afin d'obtenir des observations blanchies, c'est-à-dire des composantes décorrélées et de variance unitaire :

$$\mathbf{z} = \mathbf{H}\mathbf{y}^c. \quad (3.19)$$

Plusieurs méthodes existent pour effectuer ce blanchiment, la plus connue étant l'ACP, tout en sachant que d'un point de vue calculatoire nous pouvons utiliser la SVD. Dans le cas où le nombre de sources à estimer est plus petit que le nombre de signaux observés et que le modèle est supposé sans bruit, seules  $p$  valeurs propres de l'ACP sont non nulles. En présence de bruit, l'ACP permet de projeter les données dans un espace de dimension  $p$  engendré par les sources. Seules les  $p$  premières valeurs propres et les vecteurs propres associés sont conservés. Cette étape réalise donc également une réduction de dimensionnalité du problème. Il faut noter que le blanchiment des données n'est pas systématique. Certains algorithmes d'ACI comme l'algorithme Infomax de Bell et Sejnowski [Bell 95] s'appliquent directement sur les données centrées.

D'autres prétraitements, comme par exemple le filtrage ou la dérivation des observations, peuvent être appliqués avant de faire l'ICA, mais ils sont intrinsèques à d'autres applications et pas retenus dans notre cas.

#### 3.3.4 Mesures d'indépendance

L'hypothèse fondamentale de l'ICA est l'indépendance statistique des sources à estimer. Il a été remarqué également qu'une restriction de l'ICA est la non-gaussianité des sources, avec au maximum une source gaussienne dans le mélange. Le calcul strict de l'indépendance statistique par sa définition classique repose sur la connaissance des densités de probabilité des sources à estimer, qui ne sont pas connues dans notre cas. Il nous faut alors d'autres mesures qui caractérisent l'indépendance de sources.

D'un point de vue heuristique, le théorème limite central nous dit qu'une somme des variables aléatoires indépendantes de même densité de probabilité tend vers une variable gaussienne lorsque le nombre de variables aléatoires tend vers l'infini, ce qui a été considéré comme la base de la plupart des phénomènes physiques. Une interprétation de ce théorème est que la somme des variables aléatoires est plus proche de la gaussianité que chacune des variables. Dans le cas de modèle de l'ICA, les sources sont estimées par  $\mathbf{y} = \mathbf{W}\mathbf{z}$ , où la matrice de séparation  $\mathbf{W}$  doit être estimée. Dans le cas idéal, si  $\mathbf{W}$  est l'inverse de  $\mathbf{A}$ , nous pouvons retrouver les composantes indépendantes  $s_i$ . Cependant, un tel cas idéal n'existe pas parce que nous ne connaissons pas la matrice de mélange  $\mathbf{A}$ .

En notant  $\mathbf{X} = \mathbf{W}\mathbf{A}$ , nous avons  $\mathbf{y} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{X}\mathbf{s}$ . Les composantes de  $\mathbf{y}$  sont alors obtenues par des combinaisons linéaires de  $s_i$  pondérées par les éléments de  $\mathbf{X}$ . Comme une somme de deux variables est plus gaussienne que chacune des variables, chaque composante de  $\mathbf{y}$  est plus gaussienne que chaque composante de  $\mathbf{s}$ . L'estimation de  $\mathbf{w}_i$ , chaque ligne de  $\mathbf{W}$ , repose alors

sur le principe de la maximisation de la non gaussianité du produit  $\mathbf{w}_i \mathbf{z}$ .

Différentes mesures d'indépendance et de non-gaussianité existent dans la littérature et les différentes méthodes d'ICA sont issues de ces mesures [Cardoso 93, Comon 94, Hyvärinen 99a]. Les mesures les plus courantes sont, le kurtosis, la néguentropie, l'information mutuelle, l'infomax, le maximum de vraisemblance, les méthodes à l'ordre 2, etc. [Gaeta 90, Pham 92, Cardoso 93, Comon 94, Bell 95, Delfosse 95, Pham 96, Hyvärinen 01]. Certaines méthodes sont plus privilégiées que d'autres par leur efficacité, leur généralité et leur vitesse de convergence. Nous nous limitons à présenter quelques unes de ces méthodes qui ont été utilisées dans ce travail : JADE et FastICA.

### JADE

Une approche pour identifier le modèle de l'ICA est basée sur l'utilisation des tenseurs des cumulants d'ordre 4. Le tenseur est une généralisation des matrices. Les tenseurs de cumulants peuvent donc être vus comme une généralisation de la notion de matrice de covariance. Si la matrice de covariance est un tenseur de cumulants d'ordre 2, le tenseur de cumulants d'ordre 4 est défini par les cumulants d'ordre 4. Les éléments de la diagonale principale des tenseurs de cumulants d'un vecteur aléatoire sont les cumulants marginaux du vecteur aléatoire et les éléments non diagonaux sont des cumulants croisés. Souvent, nous négligeons les cumulants d'ordre 3 car pour toute distribution symétrique ils sont égaux à zéro et nous considérons ici uniquement les cumulants d'ordre 4. Les cumulants ont certaines propriétés utiles qui deviennent très intéressantes dans le contexte de l'ICA comme :

1. tous les cumulants d'ordre supérieur à deux sont nuls pour des variables aléatoires gaussiennes.
2. pour des variables mutuellement indépendantes, les cumulants croisés à tous les ordres sont nuls.
3. les cumulants possèdent la propriété de multilinéarité qui permet d'exprimer les cumulants d'un vecteur  $\mathbf{Y} = \mathbf{W}\mathbf{z}$  en fonction des cumulants du vecteur  $\mathbf{z}$  et de la transformation linéaire  $\mathbf{W}$ .

L'ICA peut être donc basée sur la diagonalisation du tenseur des cumulants à l'aide des propriétés 2 et 3 ci-dessus, car si le tenseur est diagonal, les cumulants croisés sont nuls ce qui est le cas pour les sources indépendantes. Tout comme une matrice est un opérateur linéaire de l'espace des vecteurs de dimensions  $p$ , le tenseur des cumulants d'ordre 4 est un opérateur linéaire de l'espace des matrices de dimensions  $p \times p$ . Pour diagonaliser le tenseur de cumulants d'ordre 4, Cardoso [Cardoso 93] a défini la notion de matrice cumulante  $\mathbf{N} = Q_z(\mathbf{M})$  associé à une matrice  $\mathbf{M}$  de dimension  $p \times p$  :

$$\mathbf{N} = Q_z(\mathbf{M}) \stackrel{\Delta}{\iff} \left\{ N_{ij} = \sum_{k=1}^p \sum_{l=1}^p Q_{ij}^{kl} M_{lk} \mid 1 \leq i, j, k, l \leq p \right\}, \quad \text{avec} \quad (3.20)$$

$$Q_z = \left\{ Q_{ij}^{kl} = \text{Cum}[z_i, z_j, z_k, z_l] \mid 1 \leq i, j \leq p \right\} \quad (3.21)$$

le tenseur de cumulants croisés d'ordre 4 du vecteur  $\mathbf{z}$ . Il existe une décomposition en valeurs propres pour le tenseur de cumulants comme pour tout opérateur linéaire symétrique. En analogie avec le vecteur propre pour une décomposition matricielle, nous pouvons définir  $p$  matrices propres  $\mathbf{M}^q$ ,  $q = \{1, \dots, p\}$  et  $p$  valeurs propres  $\lambda_q$  pour le tenseur  $Q_z$  selon la décomposition suivante :

$$Q_z(\mathbf{M}^q) = \lambda_q \mathbf{M}^q. \quad (3.22)$$

Il a été démontré dans [Cardoso 93] que les matrices propres  $\mathbf{M}^q$  sont définies par :

$$\mathbf{M}^q = \mathbf{w}_q \mathbf{w}_q^T, \quad (3.23)$$

avec  $\mathbf{w}_q$  une ligne de la matrice de séparation  $\mathbf{W}$ . Une connaissance des matrices propres du tenseur de cumulants du vecteur blanchi  $\mathbf{z}$  nous permettra de retrouver les composantes indépendantes. L'algorithme FOBI (Fourth Order Blind Identification en anglais) [Cardoso 89] se fonde sur la diagonalisation d'une seule matrice cumulée  $Q_z(\mathbf{M})$  pour obtenir la matrice de rotation. L'inconvénient de cette méthode est que les sources à estimer doivent avoir des kurtosis différents.

Au contraire des matrices, il n'existe pas une méthode de diagonalisation de tenseurs mais uniquement des approximations. Cardoso a proposé l'algorithme JADE (Joint Approximate Diagonalization of Eigen-matrices en anglais) pour diagonaliser conjointement plusieurs matrices cumulantes obtenues à partir du tenseur de cumulants. Un choix simple pour la matrice  $\mathbf{M}$  est  $\mathbf{M} = \mathbf{I}_p$ . Cependant, il peut arriver que plusieurs sources aient le même kurtosis et dans un tel cas la diagonalisation n'est pas possible à cause de la dégénérescence des valeurs propres.

Le plus simple est alors de choisir des tranches parallèles et donc de construire des matrices en utilisant la relation (3.20) avec  $M_{lk} = \delta(l, k)$ , où  $\delta$  est le symbole de Kronecker. Nous pouvons donc construire  $p \times p$  matrices, où  $p$  est le nombre de sources à estimer par JADE. Après l'estimation des matrices  $\mathbf{M}^q$ , il reste à estimer la matrice  $\mathbf{W}$ . Les propriétés de multilinéarité, d'additivité et de réjection gaussienne permettent facilement de montrer que la matrice  $\mathbf{W}$  diagonalise conjointement les  $p$  matrices  $Q_z(\mathbf{M}^q)$  [Cardoso 93]. La matrice  $\mathbf{W}$  peut donc être estimée de manière à rendre simultanément les matrices  $\mathbf{W}Q_z(\mathbf{M}^q)\mathbf{W}^T$  aussi diagonales que possible. Ceci peut être effectué en maximisant la fonction objectif suivante :

$$Q^{\text{JADE}}(\mathbf{W}) = \sum_{q=1}^p \sum_{k=1}^p (\mathbf{W}\mathbf{M}^q\mathbf{W}^T)_{kk}^2. \quad (3.24)$$

La diagonalisation peut être faite en utilisant des rotations successives de type Givens. Nous ne rentrons pas dans le détail de ces calculs, le lecteur pourra consulter les ouvrages proposés par Cardoso [Cardoso 93, Cardoso 99]. Un inconvénient de l'algorithme JADE est sa complexité quand le nombre de sources à estimer augmente (ce qui est dû à la décomposition en matrices propres). Dans ce cas, il faut soit utiliser une réduction de la dimensionnalité des données avant

d'appliquer JADE, soit utiliser d'autres méthodes qui offrent une convergence plus rapide, comme la méthode dite FastICA. Pour des dimensions réduites de données, JADE offre des solutions robustes, ce qui fait que cet algorithme est populaire et utilisé dans des nombreuses applications [Hyvärinen 01].

### FastICA

Comme nous avons discuté dans les sections précédentes, l'ICA est basée sur la maximisation de l'indépendance ou de la non-gaussianité. FastICA est un algorithme qui permet de maximiser des fonctions de contraste qui sont des mesures de non-gaussianité [Hyvärinen 99a, Hyvärinen 99b]. Une notion assez courante dans la théorie de l'information est l'entropie d'un variable aléatoire. Parmi tous les vecteurs ayant la même matrice de covariance, le vecteur aléatoire gaussien offre l'entropie la plus élevée. Pour une variable aléatoire  $s$  de densité de probabilité  $f(s)$ , en supposant qu'elle existe, l'entropie est définie par :

$$H(\mathbf{s}) = - \int f(\mathbf{s}) \log f(\mathbf{s}) d\mathbf{s}. \quad (3.25)$$

La néguentropie est définie comme la différence entre l'entropie d'un vecteur aléatoire,  $s$ , et l'entropie d'un vecteur gaussien  $\mathbf{u}$  ayant une matrice de covariance identique :

$$J(\mathbf{s}) = H(\mathbf{u}) - H(\mathbf{s}). \quad (3.26)$$

Le problème de son utilisation directe dans des algorithmes ICA est qu'elle dépend de la densité de probabilité qui n'est pas connue. Plusieurs approximations de néguentropie [Hyvärinen 00] existent alors, parmi lesquelles la plus générale est, dans le cas monodimensionnel donnée par :

$$J(s_i) = c [E \{G(s_i)\} - E \{G(u_i)\}]^2, \quad (3.27)$$

où  $G$  peut être toute fonction non-quadratique et  $c$  une constante.

L'idée de l'approximation des composantes indépendantes est toujours reliée à l'estimation d'une matrice de séparation,  $\mathbf{W}$ , qui permet de retrouver les composantes indépendantes  $\mathbf{Wz}$ . La fonction objectif à maximiser devient alors :

$$J_G(\mathbf{W}) = \sum_{j=1}^p [E \{G(\mathbf{w}_j^T \mathbf{z})\} - E \{G(\nu)\}]^2, \quad (3.28)$$

où  $\nu$  est une variable aléatoire gaussienne de même variance que  $\mathbf{w}_j^T \mathbf{z}$  et  $\mathbf{w}_j$  une colonne de la matrice  $\mathbf{W}$ .

Le choix de la fonction de contraste  $G$  pour l'approximation de néguentropie repose sur l'obéissance de quelques propriétés afin qu'il soit plus robuste et qu'il ait une variance asymptotique minimale. En particulier,  $G$  ne doit pas augmenter trop vite afin d'obtenir des estimateurs robustes. Dans ce cadre, les fonctions de contrastes suivantes sont les plus utilisées :

$$G_1(y) = \frac{1}{a_1} \log \cosh(a_1 y) \quad (3.29)$$



et

$$G_2(y) = -\frac{1}{a_2} \exp\left(-\frac{a_2 y^2}{2}\right) \quad (3.30)$$

avec  $1 \leq a_1 \leq 2$  et  $a_2$  des constantes souvent fixées à l'unité. La maximisation de la non-gaussianité par maximisation de la fonction (3.28) est plus rapide quand elle est implémentée par un algorithme à point fixe. L'algorithme à point fixe le plus utilisé pour retrouver une seule direction dans laquelle la non-gaussianité se maximise, se nomme *one unit FastICA* pour l'estimation d'une seule composante indépendante à la fois. Cet algorithme peut être schématisé de la manière suivante :

1. Centrer et blanchir les données.
2. Choisir un vecteur initial  $\mathbf{w}_j$  de norme unité.
3. Mettre à jour :  $\mathbf{w}_j \leftarrow E \left\{ \mathbf{z}g \left( \mathbf{w}_j^T \mathbf{z} \right) \right\} - E \left\{ g' \left( \mathbf{w}_j^T \mathbf{z} \right) \right\} \mathbf{w}_j$ , avec  $g$  et  $g'$  respectivement les dérivées de  $G$  et  $g$ .
4. Mettre à jour la contrainte de normalité :  $\mathbf{w}_j \leftarrow \mathbf{w}_j / \|\mathbf{w}_j\|$ .
5. Boucler sur 3 jusqu'à la convergence. La convergence ici signifie que les deux valeurs consécutives du vecteur  $\mathbf{w}_j$  pointent dans la même direction, c'est-à-dire, que la valeur absolue de leur produit scalaire est proche de l'unité.

La méthode d'approximation discutée ci-dessus est pour l'estimation d'une source indépendante. Cette méthode pourra être très bien adaptée pour l'estimation des plusieurs sources indépendantes. Nous profitons ici pour discuter les notions de séparation de sources en mode symétrique et en mode déflation.

### 3.3.5 Séparation de sources en mode symétrique et en mode déflation

Il existe deux approches pour estimer le modèle ICA : symétrique et par déflation.

#### Mode symétrique

L'ICA en mode symétrique consiste à estimer toutes les sources conjointement ou simultanément. C'est le mode qui a été proposé initialement pour résoudre l'ICA. L'algorithme JADE utilise naturellement ce mode d'extraction. Pour estimer plusieurs sources avec l'algorithme FastICA, nous appliquons ce qu'on appelle l'orthogonalisation symétrique. Ceci consiste à faire une itération d'algorithme présenté ci-dessus pour calculer parallèlement les vecteurs de la matrice de séparation correspondant à chaque source  $\mathbf{w}_j, j = 1, \dots, p$ . La matrice ainsi obtenue,  $\mathbf{W}$ , est ensuite orthogonalisée comme :

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-(1/2)} \mathbf{W}, \quad (3.31)$$

ou en utilisant d'autres méthodes d'orthogonalisation comme présenté en [Hyvärinen 01, pp. 141-143].

### Mode déflation

Dans le mode déflation, les sources sont estimées une par une, c'est-à-dire, nous estimons une source selon un certain critère, nous enlevons l'apport de cette source et puis, itérativement, nous calculons les sources suivantes jusqu'à ce que toutes les sources soient estimées. Pour enlever l'apport d'une source, une approche est la soustraction simple de la contribution de la source extraite,  $\hat{s}_j$ , aux observations (données), cette contribution étant estimée par une régression linéaire de type :

$$\hat{\mathbf{w}}_j = E \{ \mathbf{z}^T \hat{s}_j \} / E \{ |\hat{s}_j|^2 \}. \quad (3.32)$$

Cependant, FastICA n'utilise pas la régression et après chaque itération elle projette le vecteur extrait dans un sous-espace orthogonal aux vecteurs extraits précédemment. La procédure peut ainsi être résumée comme ci-dessous :

1. Soit  $j = 1, \dots, p$  sources indépendantes à estimer,  $j \leftarrow 1$ .
2. Initialiser  $\mathbf{w}_j$ .
3. Faire une itération de l'algorithme présenté ci-dessus sur  $\mathbf{w}_j$ .
4. Faire l'orthogonalisation suivante :

$$\mathbf{w}_j \leftarrow \mathbf{w}_j - \sum_{k=1}^{j-1} (\mathbf{w}_j^T \mathbf{w}_k) \mathbf{w}_k \quad (3.33)$$

5. Normaliser  $\mathbf{w}_j \leftarrow \mathbf{w}_j / \|\mathbf{w}_j\|$
6. Si pas de convergence, revenir à 3
7. Passer à la source suivante  $j \leftarrow j + 1$ . Si  $j \leq p$ , revenir à 2.

Les algorithmes utilisant l'approche symétrique sont souvent considérés comme étant plus performants puisque en mode déflation il y a un risque d'accumuler les erreurs lors de l'orthogonalisation ou de la régression. Au contraire, les méthodes basées sur le mode déflation sont considérées plus efficaces du point de vue temps de calcul.

#### 3.3.6 Application aux données réelles

De nombreux signaux réels sont modélisables par des réalisations de processus stationnaires et ergodiques. De nombreuses applications des techniques d'ICA sur des exemples réels ont ainsi pu être étudiées et les résultats cohérents ont permis de valider les méthodes d'ICA. Pour cette raison, les réalisations  $z(x)$  et  $s(x)$  des vecteurs aléatoires  $\mathbf{z}$  et  $\mathbf{s}$  sont supposées comme étant des réalisations de processus stationnaires et ergodiques. Pour les données de température, ces réalisations représentent les différentes lignes de la matrice  $\mathbf{Y}$ , après centrage et blanchiment et les différentes lignes de la matrice  $\mathbf{F}$  du modèle linéaire (2.12). Le bruit de mesure,  $\mathbf{B}$ , peut être estimé comme un sous-espace résidu défini par l'équation (3.13). Un estimateur consistant de la matrice de covariance de  $\mathbf{z}$  peut être donc calculé, ainsi que les signaux blanchis. De même, à partir des données blanchies, nous pouvons calculer des estimateurs consistants des cumulants d'ordre 4,  $Cum[z_i, z_j, z_k, z_l]$ . Ceci permet donc de calculer une estimée de la matrice de mélange  $\mathbf{A}$  grâce à l'estimation de la matrice  $\mathbf{W}$  selon les méthodes exposées ci-dessus.

### 3.4 Méthodologie proposée pour la détection des fuites

Après avoir décrit en détail la décomposition en valeurs singulières et l'analyse en composantes indépendantes, nous présentons la méthodologie adoptée pour la détection des fuites en utilisant ces techniques de séparation de sources.

Lors de la présentation des différents sites expérimentaux et des données thermométriques acquises sur les sites, nous avons présenté la structure générale des données. La fibre optique de mesure de température est enterrée dans le sol. Les données acquises sont influencées par ce que nous appelons la réponse du sol, qui comprend la composition du sol, les hétérogénéités du terrain, les différents niveaux d'élévation du terrain, etc. En général, elle comprend tous les éléments qui donnent une forme générale à une acquisition de température. Elle peut être interprétée comme un signal moyen qui caractérise la forme d'un signal de température enregistré.

Cette réponse du sol étant assez énergétique, elle cache les informations intéressantes concernant les autres phénomènes qui interviennent sur les données de température et donc les fuites. Il est donc impératif de bien estimer cette réponse du sol enfin de l'enlever pour mieux faire ressortir les autres phénomènes présents lors de l'acquisition, notamment les fuites.

Dans un premier temps, certains prétraitements de type désaisonnalisation et normalisation de données, discutés dans le chapitre 2, sont appliqués aux données. La désaisonnalisation a pour but d'enlever les variations annuelles et saisonnières de données. Il faut mentionner ici que cette étape de prétraitement n'est pas obligatoire, mais en règle générale elle apporte une amélioration des résultats finaux [Khan 09a]. La normalisation a pour le but de centrer chaque acquisition et de la ramener à une variance unitaire. Ceci permet d'atténuer également les effets saisonniers de température des données, surtout quand l'étape de désaisonnalisation n'est pas appliquée. De toute façon, le centrage des données facilite l'application des algorithmes de séparation.

Afin de limiter les notations, nous notons les données prétraitées ou non par la matrice  $\mathbf{Y}$ , les données prétraitées remplaçant les données originales dans cette matrice dans le cas d'un prétraitement.

La première étape de l'analyse se focalise sur l'extraction de la réponse du sol à partir des données de température. Pour estimer cette réponse, nous appliquons la SVD sur les données  $\mathbf{Y}$  pour les décomposer en deux sous-espaces complémentaires, en utilisant l'équation (3.13) :

- le sous-espace signal  $\mathbf{Y}_{sig}$  construit avec les “ $m$ ” premières valeurs singulières les plus importantes,
- le sous-espace résidu,  $\mathbf{Y}_{résidu}$  construit avec les “ $N - m$ ” valeurs singulières restantes.

Les fuites que nous cherchons à identifier montrent des variations au cours du temps et en distance. Elles sont parfois limitées en temps et les sous-espaces engendrés par les sources correspondantes sont d'énergies inférieures aux sous-espaces engendrés par des sources afférentes à d'autres facteurs, comme par exemple la réponse du sol. Ceci signifie que les fuites ne sont pas

vues comme cohérentes dans les données ou elles auront une faible énergie. Elles ne vont pas se retrouver dans les premières sources constituées par les vecteurs singuliers gauches  $\mathbf{v}_j$ , avec  $j = 1, \dots, m$ . Elles peuvent donc être considérées comme du bruit par rapport à une décomposition SVD et existent alors dans le sous-espace,  $\mathbf{Y}_{résidu}$ .

Le paramètre critique pour cette étape est le choix de “ $m$ ”, paramètre qui désigne le nombre de valeurs singulières à garder pour construire le sous-espace signal. Ce choix est généralement fait en observant la pente de décroissance des valeurs singulières,  $\sigma_j$ , données par la SVD, qui sont par défaut ordonnées par ordre décroissant.

La réponse du sol étant la composante la plus cohérente des données, elle est généralement représentée par le premier vecteur singulier de gauche,  $\mathbf{v}_1$ , comme il sera montré lors de l’application de la SVD sur des données thermométriques dans la section suivante. Le choix  $m = 1$  permettra généralement d’estimer la source correspondante à cette réponse, le sous-espace correspondant pouvant être construit en prenant en compte les variations temporelles de cette source. Cependant, l’estimation de cette source par une SVD sur les données globales (acquises pendant une longue durée) risque d’être influencée par des phénomènes énergétiques éphémères en temps comme, par exemple, des précipitations importantes. Dans le chapitre 2, lors de la discussion sur les prétraitements des données thermométriques, nous avons proposé un critère pour identifier ce type de phénomènes éphémères. Ce critère, basé sur le kurtosis et le skewness, permet de sélectionner des zones temporelles d’acquisition sans instance de phénomènes énergétiques [Khan 08d]. Ensuite, sur les données présélectionnées par ce critère, nous pouvons estimer la réponse du sol par fenêtrage glissant en temps. Le vecteur caractéristique correspondant à cette source peut être obtenu ensuite en moyennant les vecteurs obtenus par la SVD à chaque position de la fenêtre glissante. En enlevant l’apport que les phénomènes énergétiques éphémères en temps pourraient avoir, ce vecteur caractéristique représente un bon estimateur de la réponse du sol.

Dans ce cas, le sous-espace signal  $\mathbf{Y}_{sig}$  correspondant à ce vecteur caractéristique peut être estimé par une approche des moindres carrés [Khan 08d], le sous-espace résidu étant obtenu comme la différence entre les données de départ,  $\mathbf{Y}$ , et le sous-espace signal,  $\mathbf{Y}_{sig}$  :

$$\mathbf{Y}_{résidu} = \mathbf{Y} - \mathbf{Y}_{sig} \quad (3.34)$$

Il faut noter que les fenêtres glissantes ne sont pas obligatoirement recouvrantes. Dans ce dernier cas, la SVD estime, pour chaque position de la fenêtre, la réponse du sol comme le vecteur  $\mathbf{v}_1$  et le sous-espace signal associé. Le sous-espace bruit est estimé par la différence entre les données définies par une fenêtre et le sous-espace signal obtenu à cette position.

Cependant, pour extraire les fuites et dans certains cas, les drains, qui sont des phénomènes non-gaussiens, la contrainte d’orthogonalité imposée aux sources estimées peut s’avérer insuffisante. L’application de l’ICA, qui suppose l’indépendance des sources, permettra de mieux séparer les sources associées à ces phénomènes, comme il sera montré dans la section suivante.

Les algorithmes de séparation de sources en aveugle essayent d’estimer des sources impulsion-

nelles quand le nombre de source augmente. Pour éviter ce phénomène, nous devons réaliser une réduction de la dimensionnalité des données. En effet, sans faire une réduction de la dimensionnalité, il n'est pas facile de donner un sens physique aux résultats. Néanmoins, pour le cas des fuites artificielles de type percolation, qui constituent des fuites impulsionnelles en distance, une approche par déflation sans réduction de la dimensionnalité de données s'avère très efficace [Khan 08a] (cf. annexe A.2). L'ICA est un algorithme à deux étapes, une première étape de blanchiment et une deuxième étape de maximisation de l'indépendance. La première étape est donnée directement par la SVD pour obtenir des vecteurs blanchis, c'est-à-dire, décorrélés et normalisés,  $\mathbf{v}_j$ . Les premiers  $m$  vecteurs étant reliés à la réponse du sol, ils ne doivent pas être pris en compte pour l'identification des sources associées aux fuites. Nous pouvons donc considérer les  $p$  vecteurs suivants  $\mathbf{v}_j$ ,  $j = m + 1, \dots, m + p$ , parmi ceux qui définissent le sous-espace résidu  $\mathbf{Y}_{résidu}$ . L'ICA sera ensuite appliquée sur le sous-espace  $\mathbf{Z}_p$  défini par ces vecteurs  $\mathbf{v}_j$  et les vecteurs gauches correspondants  $\mathbf{u}_j$ . Il faut noter que par rapport aux équations (3.14) et (3.19), la matrice  $\mathbf{Z}_p$  contient des réalisations du processus stationnaire et ergodique  $\mathbf{z}$ . L'ICA en mode symétrique estime une matrice de séparation  $\mathbf{W}$  par le biais d'une matrice de rotation de Givens  $\mathbf{B}$  et les vecteurs indépendants  $\tilde{\mathbf{v}}_j$ . D'un point de vue calculatoire, nous pouvons écrire la relation :  $[\dots, \tilde{\mathbf{v}}_j, \dots] = [\dots, \mathbf{v}_j, \dots]\mathbf{B}$ . Cette estimation se fait soit en diagonalisant le tenseur de cumulants de l'ordre 4 (JADE) soit en maximisant un autre critère de non-gaussianité (FastICA). Une fois les vecteurs indépendants  $\tilde{\mathbf{v}}_j$  estimés, nous pouvons calculer leurs variations temporelles  $\tilde{\mathbf{u}}_j$  ainsi que leurs énergies  $\tilde{\sigma}_j$  et donc réécrire le sous-espace  $\mathbf{Z}_p$  par rapport à ces nouvelles composantes comme [Vrabie 04] :

$$\mathbf{Z}_p = \sum_{j=m+1}^{m+p} \sigma_j \mathbf{u}_j \mathbf{v}_j^T = \sum_{j=m+1}^{m+p} \tilde{\sigma}_j \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^T, \quad (3.35)$$

où la première égalité est identifiée à la 1<sup>ère</sup> étape (décomposition par la SVD) et la deuxième égalité est identifiée à la 2<sup>ème</sup> étape (application de l'ICA). Les valeurs  $\tilde{\sigma}_j$  sont appelées valeurs singulières modifiées [Vrabie 04] par l'ICA. La 2<sup>ème</sup> étape évite la contrainte d'orthogonalité sur les vecteurs  $\tilde{\mathbf{u}}_j$  et, en même temps, impose le critère d'indépendance à l'ordre 4 des vecteurs  $\tilde{\mathbf{v}}_j$ . Cet estimateur permet donc d'extraire des sources non-gaussiennes permettant d'obtenir une meilleure estimation des sources définies par des facteurs tels que les drains et/ou les fuites. Afin de représenter temporellement ces fuites, et donc d'avoir une notion de localisation temps-distance, nous pouvons extraire un 2<sup>ème</sup> sous espace signal  $\mathbf{Z}_{sig}$  défini cette fois-ci par " $i_2$ " sources indépendantes à l'ordre 4 :

$$\mathbf{Z}_{sig} = \sum_{j=m+1}^{m+i_2} \tilde{\sigma}_j \tilde{\mathbf{u}}_j \tilde{\mathbf{v}}_j^T. \quad (3.36)$$

Ce sous-espace  $\mathbf{Z}_{sig}^T$  est construit avec les " $i_2$ " sources ICA les plus importantes du point de vue de la décroissance des valeurs singulières modifiées  $\tilde{\sigma}_j$ . Ceci donne une notion de l'ordonnance des sources indépendantes qui n'est pas incorporée par défaut dans l'ICA. Ce choix est motivé par le fait que même après l'extraction de la réponse du sol, les fuites ont une énergie relativement faible par rapport aux autres phénomènes. D'autres critères permettant d'ordonner ces sources

peuvent être définis, par exemple en considérant une mesure de la non-gaussianité des sources (magnitudes de kurtosis des sources, ou informations à priori sur la nature sub-gaussienne ou super-gaussienne des sources). Nous allons évoquer cette discussion sur l'ordonnement des sources dans la section suivante consacrée à l'application de cette méthode sur les données thermométriques. Le but est la détection des fuites et leur localisation en temps et surtout en distance. Même si nous n'arrivons pas à séparer toutes les fuites sur les différentes sources, nous cherchons à mieux les extraire parmi tous les autres facteurs qui viennent influencer les données de température. Dans ce cadre, nous pouvons donc construire un nouveau sous-espace résidu appelé *résidu ICA*, obtenu comme :

$$\mathbf{Z}_{résidu} = \mathbf{Y}_{résidu} - \mathbf{Z}_{sig} \quad (3.37)$$

Ce résidu de l'ICA contient l'information liée aux fuites.

Pour résumer, cette méthodologie demande quelques décisions importantes concernant le choix des paramètres comme [Khan 08e] :

- le choix du nombre “ $m$ ”, qui signifie le nombre de valeurs singulières à garder pour construire le sous-espace signal permettant d'estimer la réponse du sol,
- le nombre de sources ICA à estimer, “ $p$ ”, qui décrivent les fuites et les autres facteurs,
- le nombre de sources ICA à éliminer “ $i_2$ ” pour construire le deuxième sous-espace bruit contenant principalement l'information liée aux fuites.

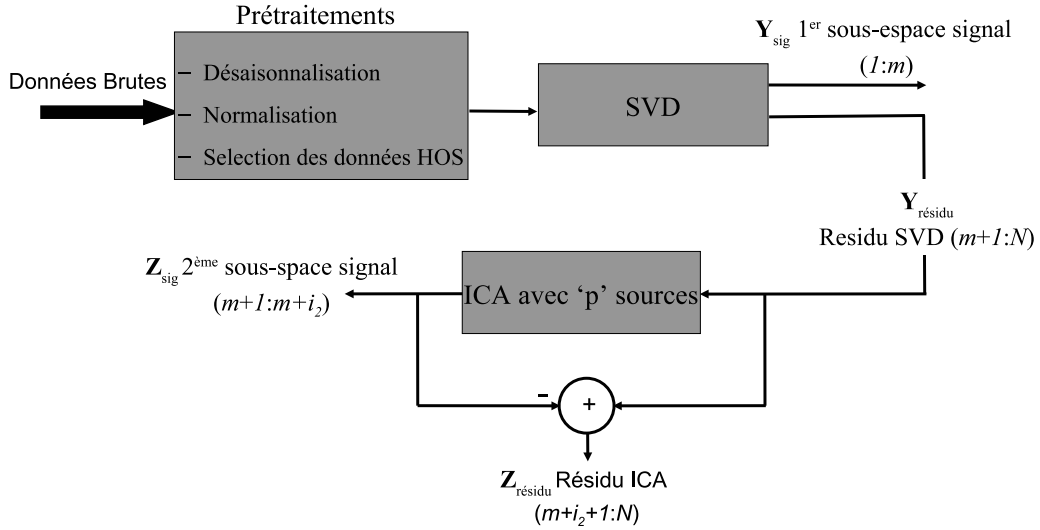
De plus, les données peuvent être traitées dans leur globalité ou, comme évoqué ci-dessus, en utilisant des fenêtres glissantes. Cette dernière méthodologie apporte plusieurs avantages surtout quand les fuites ont une énergie relativement faible par rapport à l'énergie des données entières. Une approche avec fenêtrage glissant entraîne, en plus, le choix de la taille de la fenêtre d'analyse. Cette taille est un paramètre difficile à régler parce que le résultat dépendra des types de fuites et de leur taille temporelle.

La méthode détaillée ci-dessus peut être synthétisée par le schéma de la Fig. 3.3 [Khan 08e] et sera appliquée aux données thermométriques afin de la valider et de caractériser le choix de différents paramètres.

Dans un premier temps nous considérons les données simulées, puis les données réelles acquises sur les deux sites d'Oraison et de Kembs.

### 3.5 Application aux données simulées

Un simulateur de données thermométriques a été présenté dans la section 2.2.2. Afin de réaliser l'analyse, nous simulons une digue de 1 km de longueur présentant deux couches homogènes offrant les débits les plus bas. Ces deux zones sont simulées par des sols de type limon de perméabilité  $10^{-7}$  [m/s] respectivement  $10^{-5}$  [m/s]. Afin de pouvoir considérer ces zones comme étant sans fuites, les simulations au point de mesure 25 ont été considérées (voir figure 2.1).



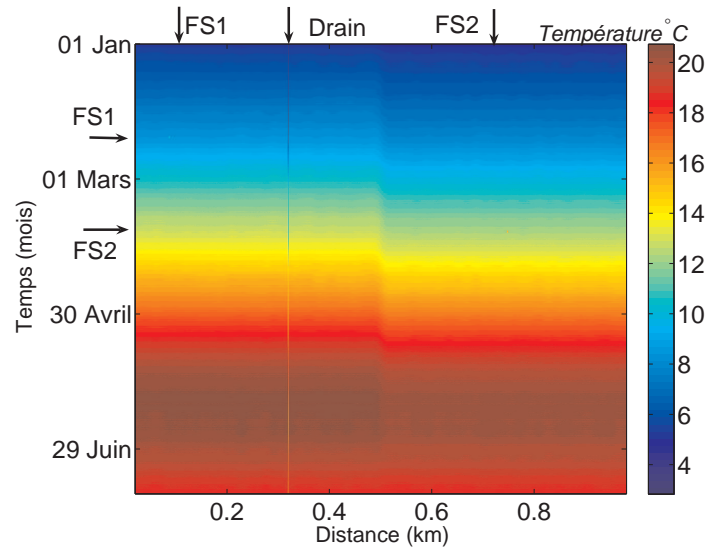
**Fig. 3.3:** Synoptique de la méthodologie proposée pour la détection de fuites dans des digues en utilisant une approche de séparation de sources (SVD et ICA).  $m$  est le nombre des valeurs singulières pour construire le premier sous-espace signal caractérisant la réponse du sol;  $p$  est le nombre de sources à estimer par l'ICA;  $i_2$  est le nombre de sources ICA à éliminer pour construire un 2<sup>ème</sup> sous-espace bruit.

L'utilisation des 5 conditions aux limites du domaine de calcul, (cf. section 2.2.2), mélangées aléatoirement pour chacune de ces zones, permet d'introduire une variabilité des données par rapport à la distance. Les températures ayant une forte variabilité par rapport à la distance entre deux capteurs successifs, un filtrage passe-bas a été effectué pour lisser les données. Les effets de bords introduits par ce filtrage sont pris en compte en enlevant les 20 m de chaque coté d'une simulation.

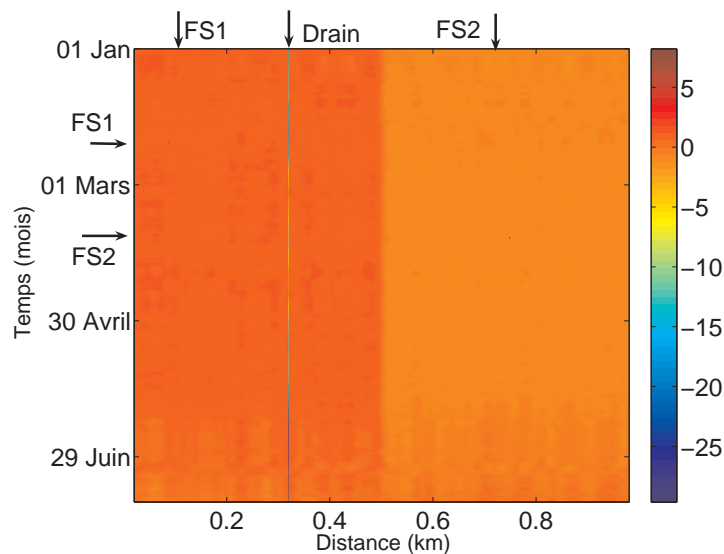
Pour simuler les différentes structures et les fuites, un drain et deux fuites ont été ajoutés au modèle. Le drain, qui est une anomalie existant à tous les instants du temps, mais qui peut varier lentement au cours du temps, a été simulé en utilisant la température simulée sur du sol de type limon de perméabilité  $10^{-4}$  [m/s] au point de mesure 3 et une condition à la limite choisie aléatoirement. Le drain est introduit à la distance de 0.32 km. Deux fuites, appelées ci-après  $FS1$  et  $FS2$ , ont été introduites à partir des simulations de sol de type limon de perméabilité  $10^{-3}$  [m/s]. Le principe de simulation des fuites est celui exposé à la section 2.2.2.

Dans un premier temps, les fuites ont été considérées comme ayant une nature impulsive en distance et en temps. Les fuites considérées sont donc de 2 m chacune et d'une durée temporelle d'une journée. Les fuites ont été introduites aux distances 87 – 88 m respectivement 747 – 748 m, la première le 11 février et la deuxième le soir du 23 mars. Ensuite, pour simuler le bruit de mesure, du bruit blanc gaussien spatialement et temporellement est ajouté aux données, avec un écart type de 0.02 pour simuler un appareil de mesure DTS Sensornet et de 0.05 pour un

DTS Siemens. Les données de température brutes pour le DTS Sensornet, montrées sur la figure 3.4(a), seront utilisées pour une première analyse. (La désaisonnalisation des données n'est pas



(a) Données thermométriques brutes.



(b) Données thermométriques normalisées.

**Fig. 3.4:** Données thermométriques simulées en fonction de la distance et du temps sur 0.96 km de longueur entre 1 janvier et 18 juillet d'une année. Un drain a été introduit à la distance 0.32 km. Deux fuites d'une durée d'un jour,  $FS1$  (à 87 – 88 m introduite le 11 février) et  $FS2$  (à 747 – 748 m introduite le 23 mars), ont été simulées également. Les différentes couleurs représentent les températures simulées.

faite.) Cependant, les données sont normalisées pour rendre chaque acquisition de moyenne nulle et variance unitaire comme montré sur la figure 3.4(b).



### 3.5.1 Analyse de données entières

Dans un premier temps, nous analysons les données en leur globalité, c'est-à-dire sur tout le temps disponible. La première étape consiste à faire une décomposition en valeurs singulières. La figure 3.5(a) présente le pourcentage de la contribution des 30 premières valeurs singulières,  $\frac{\sigma_j}{\sum_{n=1}^N \sigma_n}$ ,  $j = 1, \dots, 30$  obtenues en appliquant la SVD sur les données normalisées. Ces valeurs étaient relativement faibles, la cohérence entre les sources estimées  $\mathbf{v}_j$  par la SVD n'est pas très importante. Le rapport,  $\frac{\sigma_1}{\sum_{n=1}^N \sigma_n} = 22\%$  pour la première valeur singulière indique quand même un valeur élevée par rapport aux autres sources, la source  $\mathbf{v}_1$  étant la plus cohérente.

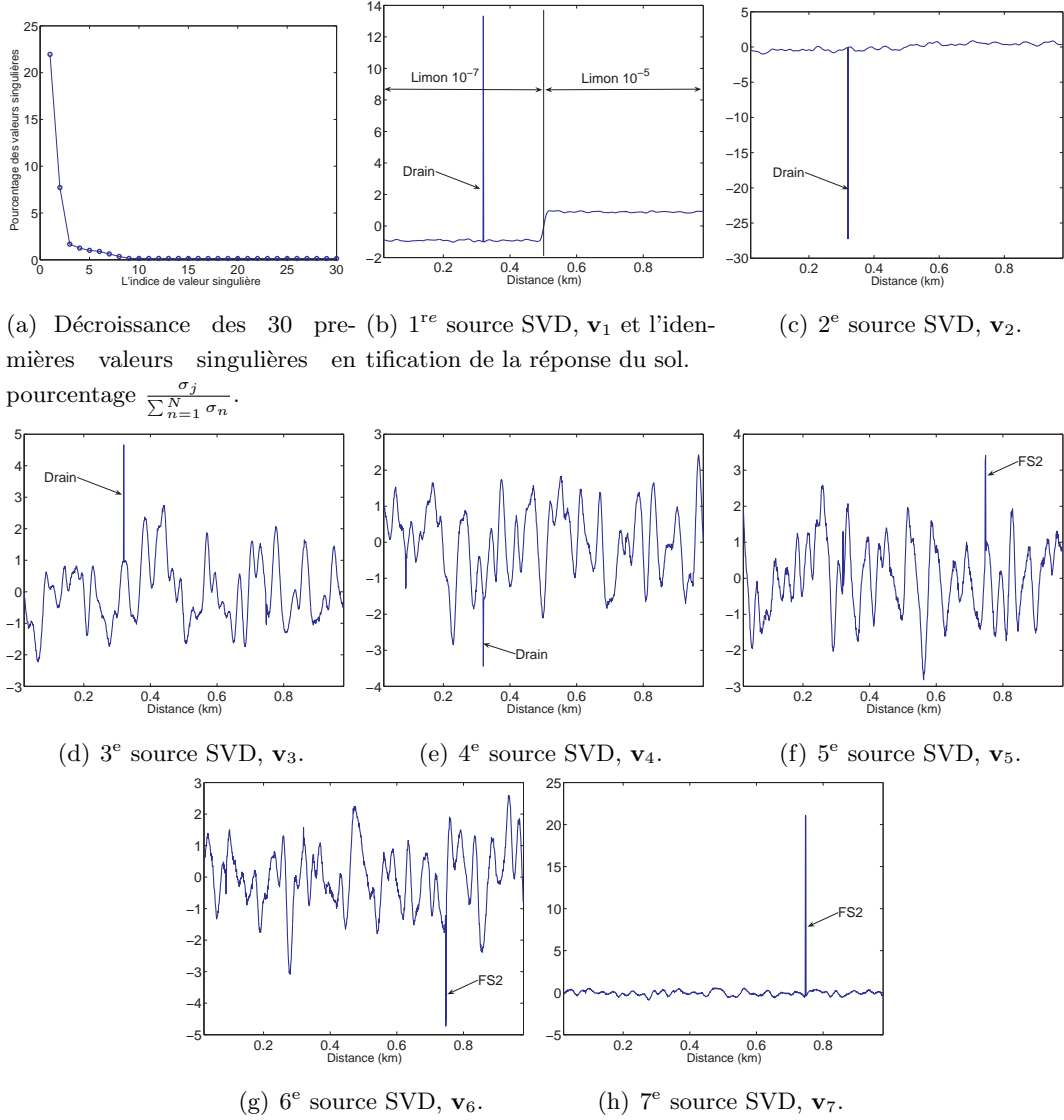
Les sept premières sources correspondantes obtenues par la SVD sont montrées sur les figures, 3.5(b)-3.5(h). Les sources ne caractérisent pas l'information intéressante du notre point de vue (les fuites) mais plutôt le bruit, les drains et la non homogénéité du sol. La première source SVD  $\mathbf{v}_1$  (figure 3.5(b)) montre des particularités liées aux phénomènes qui existent sur tout le temps formant alors la composante la plus cohérente de données. Cette source est liée à la réponse du sol. Nous pouvons constater deux zones distinctes en distance, correspondant aux deux types de sol utilisés de perméabilités différentes : limon de  $10^{-7}$  [m/s] et limon de  $10^{-5}$  [m/s]. En outre, le drain apparaît également sur cette source  $\mathbf{v}_1$  à une localisation de 0.32 km. Ceci est consistant au fait que le drain est présent à tous les instants de temps.

Sur la deuxième source SVD  $\mathbf{v}_2$  de la figure 3.5(c), c'est le drain qui présent la forte discontinuité. Ceci est dû au fait que le drain induit des variations de température au cours du temps et que la première source ne l'estime pas totalement.

Sur les sources SVD  $\mathbf{v}_3$  et  $\mathbf{v}_4$ , (figures 3.5(d) et 3.5(e)), une partie de drain est retrouvée mais également les variations issues du mélange aléatoire des conditions limites qui a été utilisé lors de la phase de simulation des données. La fuite *FS2* apparaît sur les sources  $\mathbf{v}_5$  à  $\mathbf{v}_7$  (figures 3.5(f)-3.5(h)). Comme pour le drain, la fuite est identifiée par ces sources mais elle est mélangée avec les variations issues du mélange aléatoire des conditions limites. Il a été constaté que sur les autres sources SVD les fuites restaient toujours mélangées et qu'une séparation n'est pas possible. Les fuites ont une énergie faible par rapport aux données entières et restent donc présentes dans le résidu même si nous considérons plusieurs sources pour construire le sous-espace signal avec la SVD, c'est-à-dire une valeur élevée pour "*m*".

Maintenant, si nous considérons que la première source SVD  $\mathbf{v}_1$  représente la réponse du sol et si nous construisons le sous-espace signal  $\mathbf{Y}_{sig}$  avec  $m = 1$ , nous pouvons appliquer l'ICA sur une partie du résidu  $\mathbf{Y}_{résidu}$ . Puis en choisissant  $p = 6$  sources à estimer par l'ICA, nous obtenons les sources montrées sur la figure 3.6.

La première source  $\tilde{\mathbf{v}}_1$  obtenue avec l'ICA, voir figure 3.6(a), fait ressortir principalement le drain. Les sources suivantes de  $\tilde{\mathbf{v}}_2$  à  $\tilde{\mathbf{v}}_5$  (figures 3.6(b)-3.6(e)) mettent en évidence les variations dues aux conditions limites. A la différence de la SVD, ce que nous pouvons déjà constater sur ces sources est que le drain n'est plus mélangé avec les autres variations sur les différentes sources. Cette remarque est valable également pour la fuite *FS2* qui sort sur la source  $\tilde{\mathbf{v}}_6$ , comme

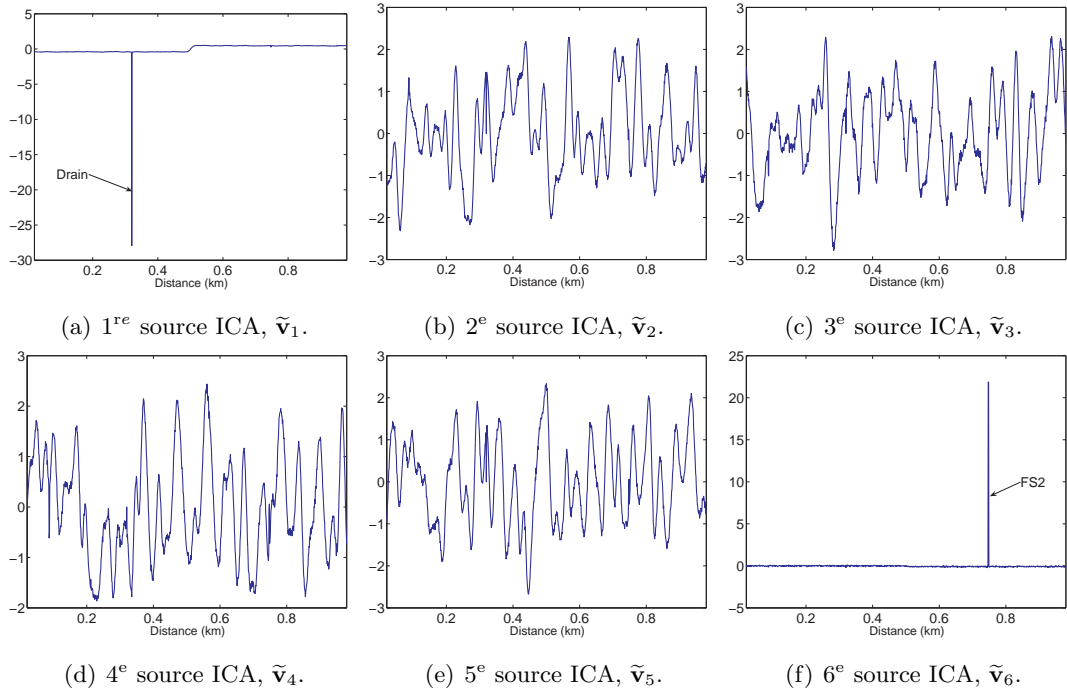


**Fig. 3.5:** Sources obtenues par l'application de la SVD sur les données simulées en leur globalité, après normalisation.

montré sur la figure 3.6(f). Cependant, l'autre fuite  $FS1$  se retrouve toujours dans le résidu. Nous pouvons construire alors un deuxième sous-espace signal avec  $i_2 = 5$  et ensuite retrouver les fuites dans le résidu final,  $\mathbf{Z}_{résidu}$ . Des tests ont été faits en augmentant le nombre de sources à estimer par l'ICA, et il a été trouvé que à partir de  $p = 7$  sources la fuite  $FS1$  est estimée par la 7<sup>e</sup> source ICA (cf. annexe D.1.1).

### 3.5.2 Analyse par fenêtrage glissant en temps

Nous avons vu que l'énergie d'une fuite est assez faible par rapport à l'énergie des données entières, les fuites apparaissent à partir de la 5<sup>e</sup> source SVD. Nous nous intéressons alors à

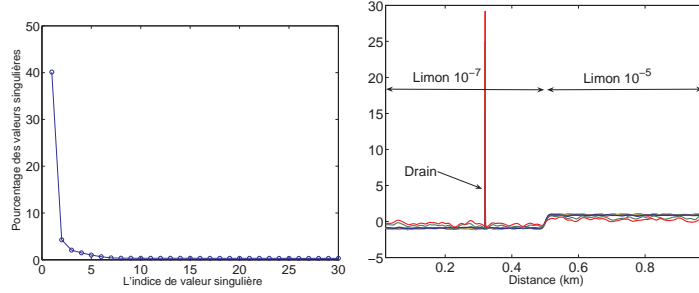


**Fig. 3.6:** Sources obtenues dans la deuxième étape de l'algorithme en appliquant l'ICA sur le résidu  $\mathbf{Y}_{résidu}$  sans fenêtrage glissant en temps.

une analyse par fenêtrage glissant en temps afin d'augmenter l'énergie relative de la fuite par rapport aux autres phénomènes. Un fenêtrage glissant permettra aussi d'exclure les zones où des phénomènes éphémères énergétiques sont présents, comme indiqué à la section 2.6.3.

Une taille de fenêtre de 20 jours a été initialement choisie pour l'analyse par fenêtrage glissant. Dans un premier temps, nous avons considéré un fenêtrage glissant sans recouvrement. Sachant que les données en temps sont définies sur  $N_t = 2400$  échantillons, nous obtenons 10 fenêtres de données à analyser. La figure 3.7 montre les 30 premières valeurs singulières en termes de pourcentage ainsi que les premières sources SVD  $\mathbf{v}_1$  obtenues sur les 10 fenêtres. Les pourcentages des valeurs singulières dans les différentes fenêtres d'analyse étant relativement constants par rapport à la position de la fenêtre, nous montrons sur la figure 3.7(a) les valeurs obtenues en moyennant les valeurs données par les SVD aux différentes positions de la fenêtre. Nous constatons que la première valeur singulière devient plus importante,  $\frac{\sigma_1}{\sum_{n=1}^N \sigma_n} = 40\%$  par rapport au cas précédent (sans fenêtrage). Les premières sources SVD,  $\mathbf{v}_1$ , pour les 10 zones temporelles d'analyse sont affichées sur la figure 3.7(b). Ces sources sont assez proches l'une de l'autre et assez similaires à la source obtenue pour l'analyse des données globales (cf. figure 3.5(b)). Les deux niveaux correspondant aux deux limons de perméabilités différentes sont décrits par cette source, ainsi qu'une grande partie du drain.

L'écart entre la première valeur singulière (40%) et les valeurs singulières suivantes (5%) est plus important maintenant (soit 35% par rapport au 16% du cas global), ce qui nous permettra



(a) Décroissance des 30 premières valeurs singulières en pourcentage  $\frac{\sigma_j}{\sum_{n=1}^N \sigma_n}$ . (b) Les 1<sup>res</sup> sources SVD,  $\mathbf{v}_1$ , pour les 10 zones temporelles d'analyse.

**Fig. 3.7:** Valeurs singulières et les premières sources SVD,  $\mathbf{v}_1$ , obtenues par l'application sur les données simulées de la SVD aux différentes positions du fenêtrage glissant de 20 jours. Les différentes couleurs en (b) correspondent aux sources obtenues pour différentes positions de la fenêtre glissante.

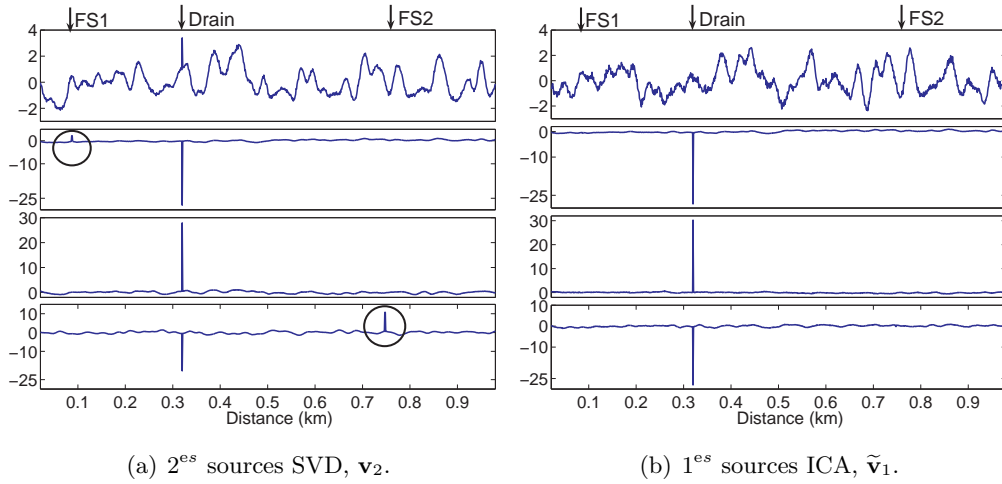
de choisir  $m = 1$  pour construire un sous-espace signal pour chaque position de la fenêtre. Ensuite, nous pourrions construire les 10 sous-espaces résidu sur lesquels l'ICA sera appliquée. Pour l'application de l'ICA,  $p = 4$  sources sont à estimer, sachant que ce choix est expérimental. Ayant déjà discuté l'avantage de l'ICA par rapport à la SVD pour la détection des fuites, nous faisons ici une comparaison entre les sources SVD et les sources ICA obtenues à partir des résidus de la SVD. Comme, pour chaque position de la fenêtre, la première source SVD est écartée, les 4 sources ICA  $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \tilde{\mathbf{v}}_3, \tilde{\mathbf{v}}_4$  seront comparées aux sources 2 à 5 de la SVD  $\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ , respectivement  $\mathbf{v}_5$ . De plus, comme nous avons une approche par fenêtrage glissant en temps, nous affichons uniquement les sources obtenues pour 4 positions de la fenêtre, sur les 10 positions possibles. Sur ces 4 positions, 2 contiennent l'occurrence des fuites et 2 sont sans fuites. Les sources obtenues pour les autres 6 positions de la fenêtre sont similaires aux positions sans fuites affichées précédemment.

La figure 3.8(a) montre les 2<sup>es</sup> sources SVD  $\mathbf{v}_2$  obtenues pour les 4 fenêtres temporelles. De haut en bas :

- zone de début de données sans fuites,
- zone contenant la fuite  $FS1$ ,
- zone sans fuites,
- zone contenant la fuite  $FS2$ .

Dans les fenêtres où il existe des fuites, les fuites sont détectées, mais elles sont mélangées avec le drain. La fuite  $FS1$  a une amplitude beaucoup plus petite que le drain. L'amplitude relative de  $FS2$  par rapport au drain est plus importante, ce qui signifie qu'une grande partie de cette fuite sera estimée par cette source. Par conséquent, elle sera moins visible dans le 2<sup>e</sup> résidu.

Au contraire, si nous considérons les premières sources  $\tilde{\mathbf{v}}_1$  estimées par l'ICA, sources affichées



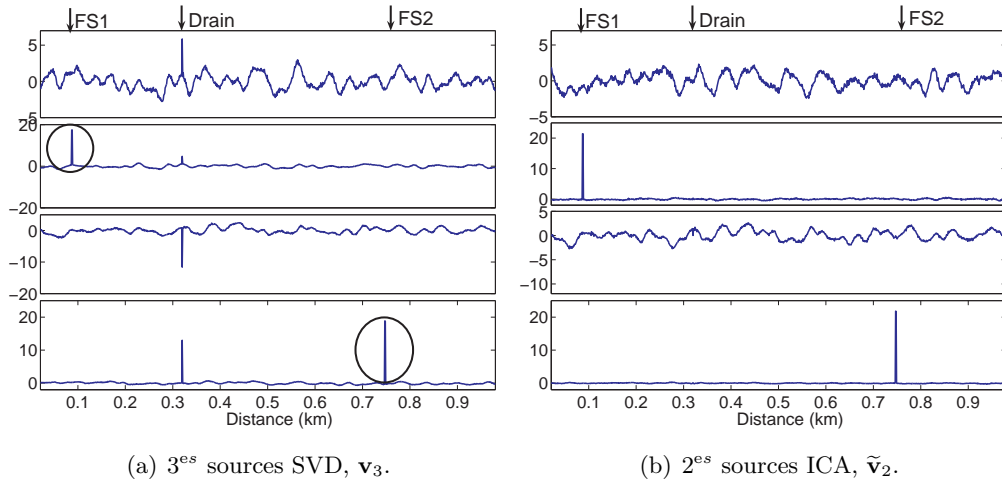
**Fig. 3.8:** Comparaison 2<sup>es</sup> sources SVD et 1<sup>es</sup> sources ICA pour les 4 différentes zones temporelles. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . Sur les sources SVD, les fuites sont mélangées avec le drain alors que sur les sources ICA, les fuites ne sont pas présentes, ces sources modélisant le drain et les variations dues aux conditions limites.

sur la figure 3.8(b), nous constatons que les fuites ne se retrouvent pas sur ces sources. Dans la première zone d'analyse, cette source montre des variations dues aux conditions limites et dans les autres zones elles estiment le drain. Le fait que le drain n'est pas estimé dans la première fenêtre temporelle peut être expliqué par sa variabilité au cours du temps, dans cette zone le drain ayant une énergie relative plus faible que dans les autres zones.

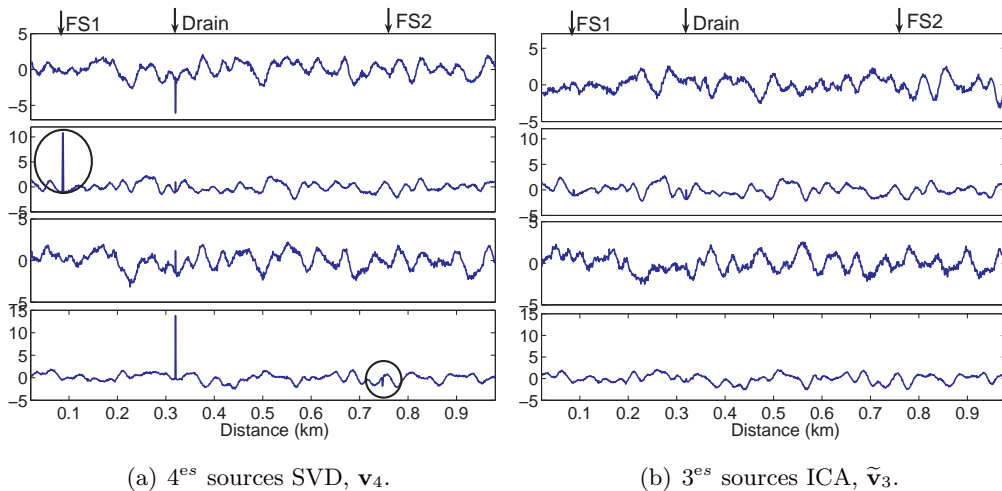
L'apport de l'ICA devient nettement plus évident sur les sources suivantes. Considérant, par exemple, les 3<sup>es</sup> sources SVD  $\mathbf{v}_3$  affichées sur la figure 3.9(a), la détection de la fuite  $FS1$  est relativement bonne, mais une partie de drain est également détectée par cette source. La fuite  $FS2$  reste également mélangée avec le drain pour la 4<sup>e</sup> zone d'analyse. Une comparaison avec les sources correspondantes  $\tilde{\mathbf{v}}_2$  de l'ICA (voir figure 3.9(b)) montre une meilleure séparation des fuites, du drain et d'autres variations dues aux conditions limites, d'où l'avantage de l'ICA pour la détection des fuites. En effet sur la source  $\tilde{\mathbf{v}}_2$  (ICA), les sources  $FS1$  et  $FS2$  ne sont jamais détectées avec les drains.

De la même façon, en analysant les 4<sup>es</sup> et les 5<sup>es</sup> sources SVD  $\mathbf{v}_4$  et  $\mathbf{v}_5$ , montrées sur les figures 3.10(a) et 3.11(a), les fuites restent toujours présentes, mais elles sont mélangées avec le drain et les autres variations. Ce qui n'est pas le cas pour les sources correspondantes  $\tilde{\mathbf{v}}_3$  et  $\tilde{\mathbf{v}}_4$  de l'ICA, affichées sur les figures 3.10(b) et 3.11(b).

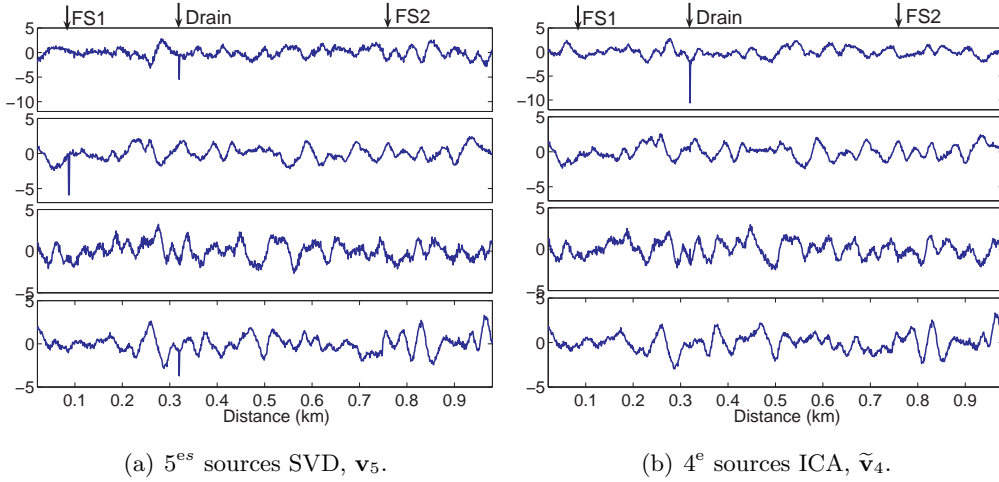
Nous pouvons alors conclure que l'ICA est plus adaptée pour la détection des fuites que la SVD. De plus, comme la SVD estime la réponse du sol sur sa première source et comme elle réalise le blanchiment des données, nous pouvons l'utiliser comme une première étape dans la méthodologie proposée. Une analyse des sources ICA montre également que nous pouvons choisir



**Fig. 3.9:** Comparaison 3<sup>es</sup> sources SVD et 2<sup>es</sup> sources ICA pour les 4 différentes zones temporelles. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . Sur les sources SVD, les fuites sont toujours mélangées avec le drain, notamment la fuite  $FS2$ , alors que pour les sources ICA les fuites sont parfaitement estimées et non mélangées aux autres phénomènes.



**Fig. 3.10:** Comparaison 4<sup>es</sup> sources SVD et 3<sup>es</sup> sources ICA pour les 4 différentes zones temporelles. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . La fuite  $FS1$  apparaît également sur les sources SVD, alors que les sources ICA ne contiennent plus les fuites.



**Fig. 3.11:** Comparaison 5<sup>es</sup> sources SVD et 4<sup>es</sup> sources ICA pour les 4 différentes zones temporelles. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . La fuite  $FS1$  reste toujours présente sur les sources SVD, mettant en évidence l'avantage de l'ICA qui ne fait pas ressortir les fuites sur ces sources.

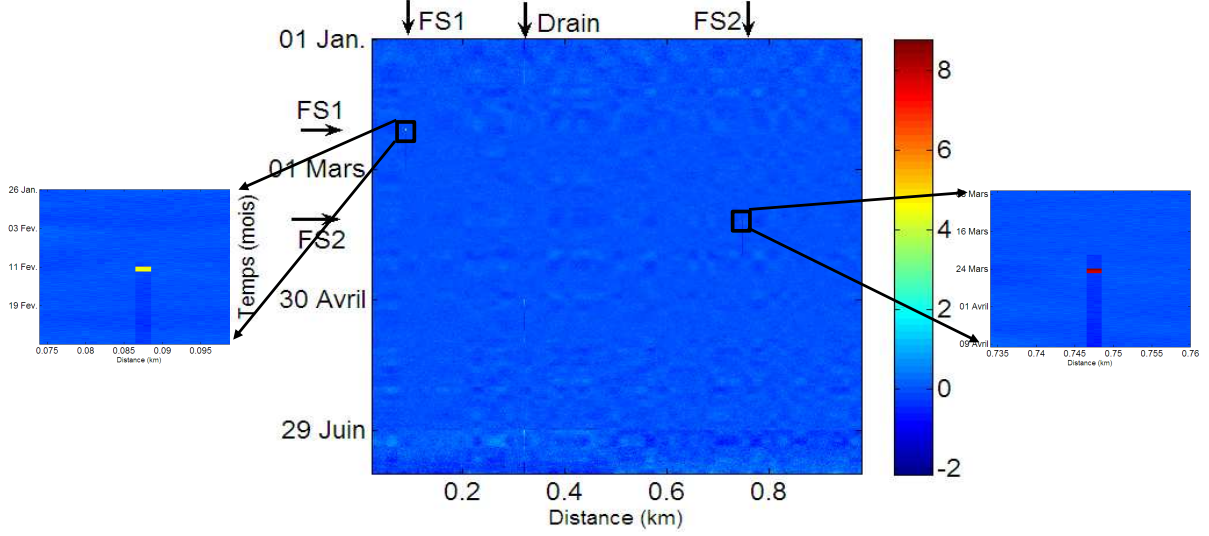
une seule source ICA pour construire le deuxième sous-espace signal, ce qui signifie  $i_2 = 1$ , afin de pouvoir retrouver les fuites dans le deuxième résidu. Il faut remarquer ici que nous pouvons très bien construire un troisième sous-espace en considérant uniquement la deuxième source  $\tilde{\mathbf{v}}_2$  de l'ICA (voir figure 3.9(b)) pour détecter des fuites, mais nous restons sur une approche générale dans laquelle les fuites seront extraites dans un espace résidu. Ceci est motivé par le fait que les fuites réelles peuvent avoir des amplitudes relatives plus faibles, ou des légères variations en distance, ce qui fera qu'elles seront décrites par plusieurs sources en même temps.

### 3.5.3 Interprétation des résidus

Une fois que nous avons choisi le nombre de sources à garder pour construire le deuxième sous-espace signal  $\mathbf{Z}_{sig}$  en utilisant l'équation (3.36), nous pouvons construire le sous-espace résidu final,  $\mathbf{Z}_{résidu}$ , pour la détection de fuites en utilisant l'équation (3.37) soit en considérant les données entières soit les données correspondantes à chaque fenêtre d'analyse.

Le résidu peut être présenté sous la forme d'une image afin de pouvoir localiser la fuite en distance et en temps. Le résidu  $\mathbf{Z}_{résidu}$  obtenu avec  $i_2 = 1$  pour une analyse par fenêtrage glissant est montré simultanément à la figure 3.12. Les fuites sont très bien localisées en temps et en distance, mais, par rapport à la taille de l'image, elles ne sont pas trop visibles.

Si nous nous limitons uniquement à une localisation en distance des fuites, nous pouvons faire une projection de l'image sur l'axe des distances. Une façon de faire la projection du résidu



**Fig. 3.12:** Résidu  $\mathbf{Z}_{résidu}$  obtenu en appliquant la méthodologie proposée en utilisant un fenêtrage glissant de 20 jours. Les fuites sont très bien localisées en temps et en distance comme montré dans les fenêtres zoomées.

$\mathbf{Z}_{résidu} = \{z_{résidu}(t, x)\}$ , à chaque distance  $x$ , est d'intégrer son énergie selon le temps :

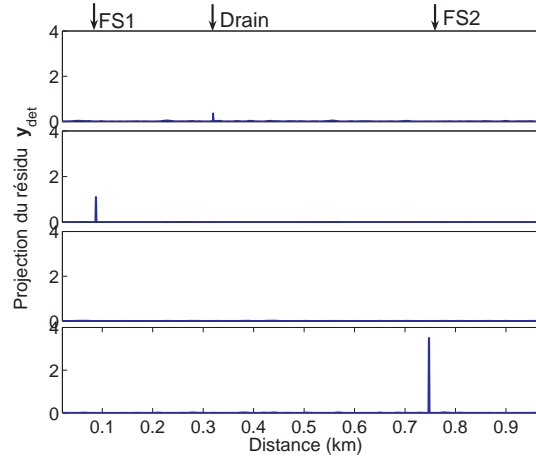
$$y_{det}(x) = \frac{1}{N_p} \sum_{t=1}^{N_p} [z_{résidu}(t, x)]^2 \quad (3.38)$$

où  $N_p$  représente le nombre d'échantillons temporels de données entières  $N_t$ , ou la taille de la fenêtre d'analyse en cas d'analyse par fenêtrage glissant. Dans le cas des fuites limitées en temps, comme c'est le cas dans cette simulation, cette projection ne serait pas utile en considérant les données entières. Nous choisissons alors  $N_p$  correspondant à la taille d'une fenêtre d'analyse, c'est-à-dire  $N_p = 20 * 12$  pour 12 acquisitions par jour. Comme précédemment, nous affichons sur la figure 3.13 le résidu dans les 4 fenêtres d'analyse. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . Les deux fuites sont détectées avec une amplitude relativement grande par rapport aux autres phénomènes. Nous constatons pour la première position de la fenêtre qu'une toute petite partie de drain est récupérée sur la projection. Ceci est du au fait que le drain montre des variations qui ne peuvent pas être extraites par la première source SVD et la première source ICA.

### 3.5.4 Étude des résidus par rapport aux différents paramètres

Nous nous intéressons par la suite à l'influence de diverses paramètres sur les résidus obtenus sur les données simulées.





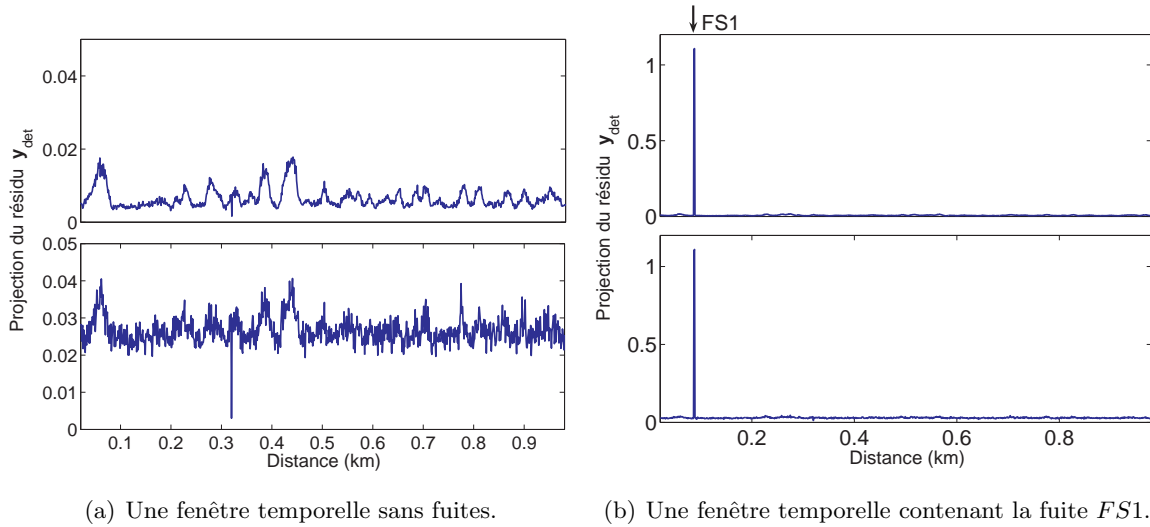
**Fig. 3.13:** Les projections,  $y_{det}(x)$ , du résidu  $\mathbf{Z}_{résidu}$  sur l'axe de distances pour les 4 différentes zones temporelles.  $N_p = 20 * 12$ . De haut en bas : zone au début de données sans fuites; zone contenant la fuite  $FS1$ ; zone sans fuites; zone contenant la fuite  $FS2$ . Les fuites  $FS1$  et  $FS2$  sont très bien détectées par rapport aux autres phénomènes.

### Étude des résidus par rapport au bruit de mesure

Pour mettre en évidence la différence entre un signal enregistré par un appareil de mesure DTS Siemens et un appareil DTS Sensornet, nous comparons les résidus obtenus à partir des données simulées sur lesquelles les bruits de mesure correspondants à chacun des appareils ont été rajoutés. Nous considérons dans les deux cas la projection des résidus pour une fenêtre temporelle d'analyse sans occurrence de fuite, puis sur la fenêtre d'analyse contenant la fuite  $FS1$ . Les résultats présentés sur la figure 3.14, montrent que l'énergie des résidus sur données simulées par DTS Siemens est approximativement deux fois plus grande que celle sur données Sensornet. De plus, les variations des résidus sont plus rapides pour les données Siemens que pour les données Sensornet. Il faut remarquer que les fuites analysées ici ont un débit important et donc une énergie assez importante par rapport au bruit. Sur les sites réels, ceci pourra nous poser un problème surtout pour identifier des fuites de faible débit.

### Étude des résidus par rapport à la taille de la fenêtre d'analyse

Le choix de la taille de la fenêtre d'analyse dépend de la durée des fuites à mettre en évidence. Par exemple, pour les fuites simulées dans cette section, si nous choisissons une taille d'analyse inférieure à 15 jours, la réponse du sol sera toujours estimée par la première source SVD, mais la fuite  $FS2$  sera détectée sur la première source ICA quand les sources sont ordonnées par rapport à l'énergie (cf. annexe D.1.2). Le sous-espace signal construit après ICA contiendra alors une partie importante de la fuite  $FS2$ . La fuite  $FS1$  par contre se retrouve toujours sur la deuxième source ICA. Il faut remarquer ici que les fuites  $FS1$  et  $FS2$  ont été simulées avec des limons de même perméabilité, mais à différentes dates, l'énergie relative de la fuite  $FS2$  étant plus grande



**Fig. 3.14:** Projection des résidus avec  $N_p = 20 \times 12$  pour deux fenêtres temporelles : (a) sans fuites et (b) contenant la fuite  $FS1$ . En haut, projection des résidus obtenus sur données simulées par DTS Sensornet ; en bas sur données simulées par DTS Siemens.

que celle de la fuite  $FS1$ . Il faut aussi noter que, pour les deux fuites, leurs énergies relatives sont plus importantes que dans des cas réels dans lesquels l'écart entre une fuite mesurée et la température du sol adjacent est relativement faible. Cet écart dépend de différents facteurs comme : l'écart entre la température de l'eau et celle du sol, la saison, la perméabilité locale du sol, le débit de la fuite, la durée de la fuite, etc. D'un point de vue pratique, la taille de la fenêtre d'analyse dépendra donc des données à analyser. De plus, les données analysées dans cette section sont des données simulées, mais la richesse de l'algorithme en termes de différents paramètres à régler ainsi qu'un manque d'expertise profonde dans la génération de ces données simulées ne nous permet que de proposer des valeurs empiriques pour les tailles de fenêtre d'analyse.

### Étude des résidus par rapport à la durée et à la localisation des fuites

La méthode proposée a été testée avec le même succès sur des données pour lesquelles les fuites sont localisées sur plusieurs mètres. Des fuites ayant des largeurs raisonnables de quelques mètres sont détectables sans aucun changement de paramètres. Cependant, nous ne pouvons pas localiser des fuites qui ont une largeur inférieure à la résolution spatiale de l'appareil de mesure.

Afin d'étudier l'influence de leur durée, des fuites ont été simulées sur 5 jours. Il a été constaté que lors d'une analyse par fenêtrage glissant de 20 jours, la fuite  $FS1$  était détectée sur la deuxième source ICA, mais la fuite  $FS2$  apparaissait sur la première source ICA. En testant différentes tailles de fenêtres d'analyse, il a été constaté qu'une taille d'analyse de 100 jours permet de retrouver les deux fuites dans le résidu ICA avec  $i_2 = 1$  (cf. annexe D.1.3.) Comme évoqué précédemment, leur extraction dans le résidu ICA dépend de leurs énergies relatives. Ayant dit cela, les fuites qui existent à tous les instants de temps peuvent apparaître pour

cet algorithme comme un drain et resteront alors non évolutives et peut-être non-déTECTABLES. Cependant, dans des cas réels, les fuites montrent des variations au cours du temps et dans le cas où une fuite montre un débit important sur une courte durée, elle sera détectable par cet algorithme. Il faut aussi mentionner que nous pouvons appliquer un autre critère permettant d'ordonner les sources après ICA, comme par exemple le kurtosis. Nous allons explorer cette possibilité d'ordonnement des sources dans le cadre d'analyse des données enregistrées sur le site de Kembs.

### 3.6 Application aux données enregistrées sur le site d'Oraison

Nous analysons dans cette section les données réelles acquises sur le site d'Oraison. Les données ont été enregistrées par un appareil DTS Sensornet. Nous allons nous focaliser sur deux cas particuliers : fuites artificielles de type percolation et fuites réelles. Les données initiales contenant les fuites artificielles sont présentées sur la figure 3.15(a). Les données ont été acquises pendant la période 12 avril 2005 - 13 octobre 2005. Les températures mesurées sont affichées (3.15(a)) par des couleurs variant du bleu pour des températures basses enregistrées pendant le printemps (avril-mai) et l'automne (octobre) au rouge pour des températures élevées pendant l'été (juin-septembre). Il faut remarquer ici que les deux mesures aller-retour ont été moyennées afin de réduire le bruit introduit par le DTS. Les données contiennent deux drains,  $D1$  et  $D2$ , situés respectivement autour des distances 0.561 km et 0.858 km (marqués avec des flèches sur la figure). Les drains sont localisés sur plusieurs mètres et montrent des variations au cours du temps. Trois fuites artificielles de type percolation,  $L1$ ,  $L2$  et  $L3$ , ont été simulées entre le 10 et 12 mai 2005 (leur localisation, leur début de démarrage et leur débit étant spécifiés dans le tableau 3.1). La zone des fuites est localisée sur l'image 3.15(a) par des flèches. Un point chaud  $HP$ , qui consistait à verser de l'eau chaude sur la fibre a été également introduit le 10 mai à la distance 0.675 km.

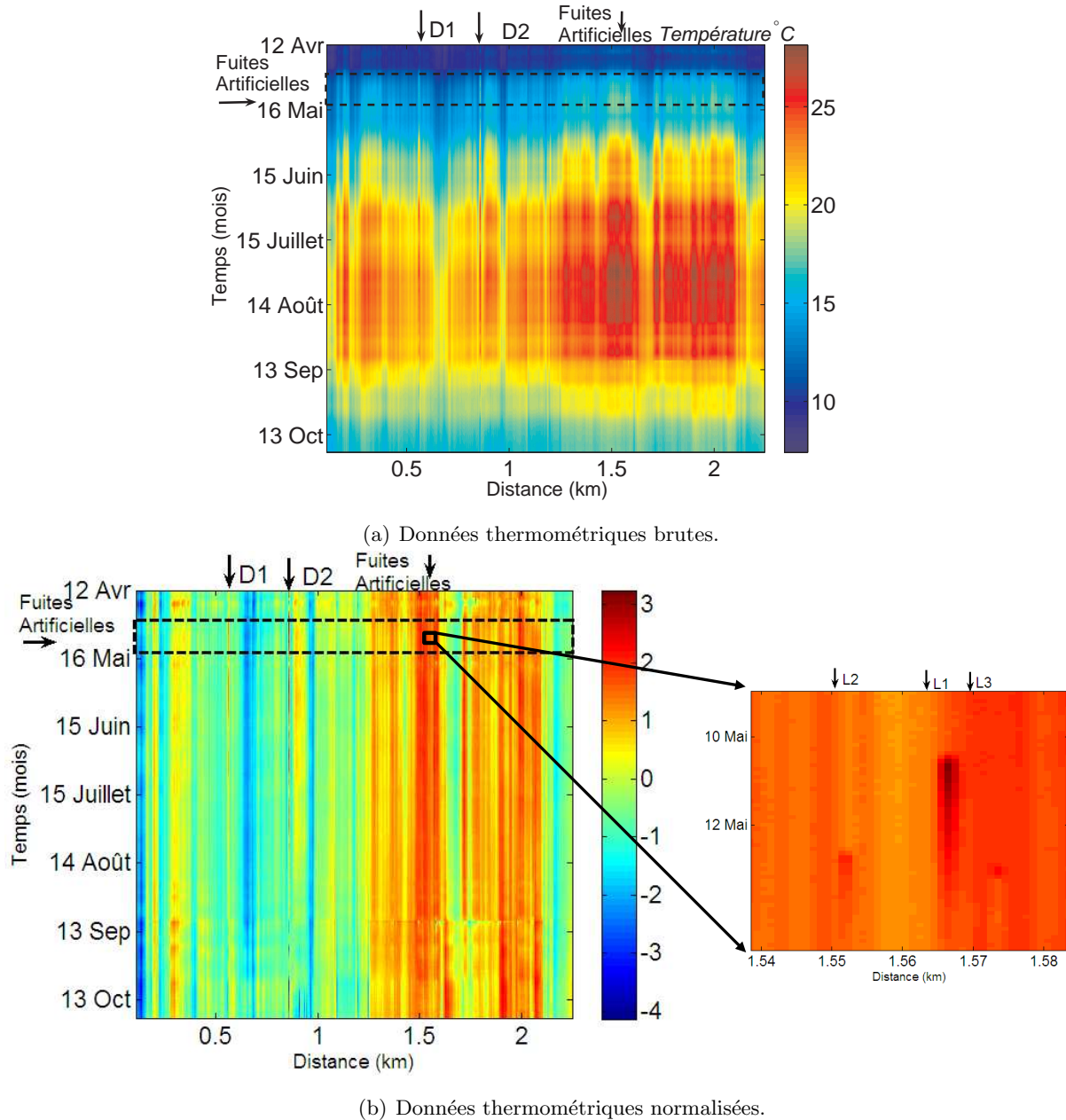
Il faut mentionner que les enregistrements sur les premiers 100m correspondant à la présence de la fibre dans les locaux de la centrale hydraulique ne sont pas pris en compte. Les fuites arti-

**Tab. 3.1:** Caractérisation des structures et fuites artificielles au site d'Oraison.

|               | Drains |       | Fuites          |                 |           | Point Chaud |
|---------------|--------|-------|-----------------|-----------------|-----------|-------------|
|               | D1     | D2    | L1              | L2              | L3        | HP          |
| Distance (km) | 0.561  | 0.858 | 1.566           | 1.551           | 1.573     | 0.674       |
| Temps (Mai)   | -      | -     | 10 (après-midi) | 12 (après-midi) | 12 (soir) | 10 (matin)  |
| Débit (l/min) | -      | -     | 5               | 1               | 1         | -           |

cielles et le point chaud ont une nature impulsionnelle en distance (car localisés sur un mètre) et également en temps (car enregistrés sur une durée de quelques heures à une dizaine d'heures).

Nous avons normalisé les données pour que chaque acquisition soit de moyenne nulle et de va-



**Fig. 3.15:** Données thermométriques enregistrées sur le site d'Oraison sur 2.128 km de longueur entre 12 avril et 13 octobre 2005. Les données contiennent deux drains,  $D1$  et  $D2$ , ainsi que trois fuites artificielles de type percolation,  $L1$ ,  $L2$  et  $L3$ . Un point chaud,  $HP$ , est également simulé le 10 mai à 0.675 km. Les différentes couleurs codent les températures enregistrées.

riance unitaire (cf. section 2.6.2). Le but est de conférer la même importance à tous les signaux enregistrés. Cette normalisation permet d'atténuer les effets des variations saisonnières et journalières par rapport aux données brutes. Elle permet aussi de mieux mettre en évidence certains

phénomènes, principalement les effets produits par les précipitations. Les données après normalisation sont montrées sur la figure 3.15(b). Nous avons constaté qu’un prétraitement de type désaisonnalisation (voir section 2.6.1) n’améliore pas les résultats obtenus sur ce type de données. En effet, lorsque nous effectuons une normalisation, celle-ci atténue également les variations annuelles et saisonnières de température.

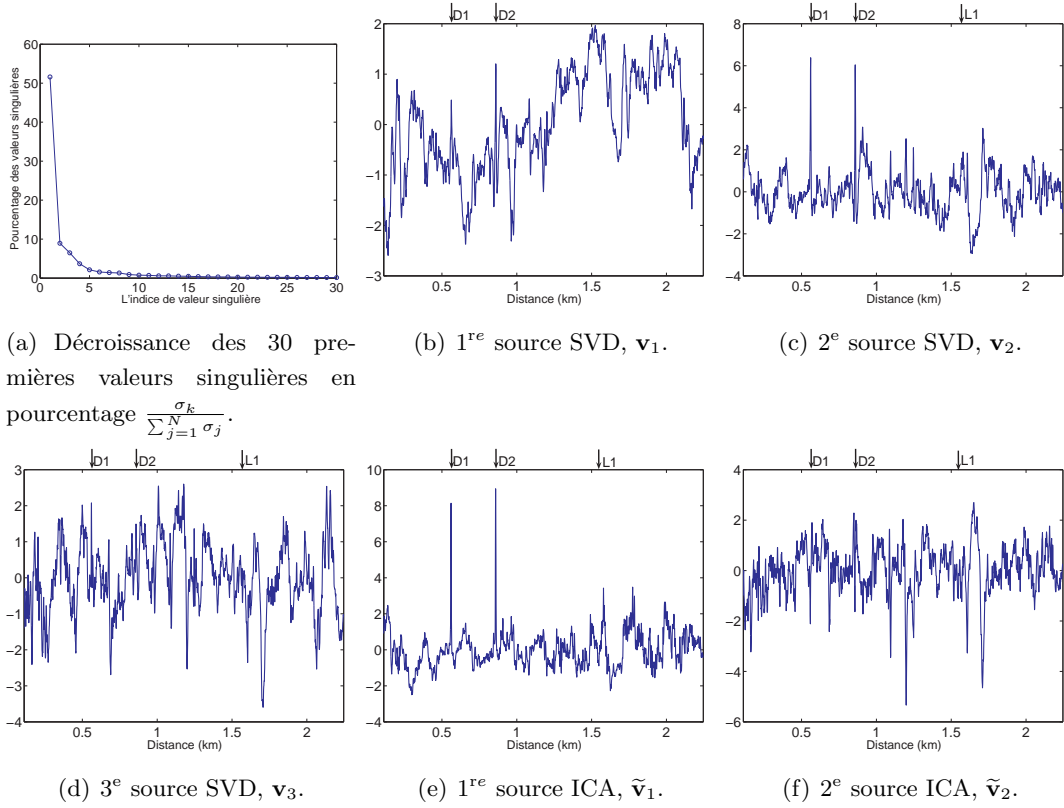
### 3.6.1 Analyse de données entières

Les données normalisées sont analysées dans leur globalité par la SVD. Les valeurs singulières étant ordonnées, nous affichons sur la figure 3.16(a) que la contribution des 30 premières valeurs singulières en termes de pourcentage de leur contribution, c’est-à-dire  $\sigma_k / \sum_{j=1}^N \sigma_j; k = 1, \dots, N$ . Pour la première valeur singulière nous obtenons une contribution de  $\sigma_1 / \sum_{j=1}^N \sigma_j = 54.8\%$ . Même si cette valeur n’est pas très grande, nous constatons sur la courbe de valeurs singulières une nette séparation par rapport aux autres valeurs singulières (la deuxième est à 10%). Ceci signifie que la source correspondante,  $\mathbf{v}_1$ , récupère un signal très cohérent entre les différentes acquisitions.

Nous pouvons choisir les paramètres “ $m$ ” et “ $p$ ” en regardant la décroissance des valeurs singulières. Pour construire le sous-espace signal  $\mathbf{Y}_{sig}$  avec la SVD, nous pouvons sélectionner  $m = 1$  parce qu’il y a un changement de pente à cette valeur dans la courbe des valeurs singulières [Bihan 06]. Ensuite, nous estimons  $p = 2$  sources par l’ICA. Parmi ces deux nouvelles sources, nous utilisons une source pour caractériser le deuxième sous-espace signal et une pour le sous-espace bruit  $\mathbf{Z}_{résidu}$ , donc  $i_2 = 1$ .

Les 3 sources estimées par la SVD sont présentées dans les figures 3.16(b)-3.16(d) et les deux sources indépendantes estimées par l’ICA dans les figures 3.16(e)-3.16(f). Nous constatons que les sources SVD ne montrent aucune discrimination au niveau des fuites artificielles, par contre les drains  $D1$  and  $D2$  peuvent être clairement identifiés sur la 2<sup>ème</sup> source issue de la SVD (à 0.561 km et 0.859 km). Nous constatons également que les sources  $\tilde{\mathbf{v}}_j$  issues de l’ICA n’apportent pas d’informations supplémentaires concernant les fuites. Il convient de rappeler que les sources extraites par la SVD sont ordonnées par rapport à l’ordonnement des valeurs singulières  $\sigma_j$ . De la même façon, les sources estimées par l’ICA sont aussi ordonnées ici en termes des valeurs singulières modifiées  $\tilde{\sigma}_j$ . L’énergie d’une fuite étant trop petite par rapport à l’énergie des données entières, les fuites restent toujours présentes dans le résidu (sous-espace bruit) même si nous gardons plusieurs valeurs singulières pour construire le sous-espace signal par la SVD.

En testant plusieurs valeurs de  $m$ ,  $p$  and  $i_2$ , nous constatons que les résultats restent similaires à ceux obtenus précédemment (cf. annexe D.2.1). L’estimation des fuites reste donc incomplète en prenant une analyse globale.



**Fig. 3.16:** Sources obtenues par l'application de la SVD suivi de l'ICA sur le résidu SVD pour les données Oraison normalisées sans fenêtrage glissant en temps.

### 3.6.2 Analyse par fenêtrage glissant en temps

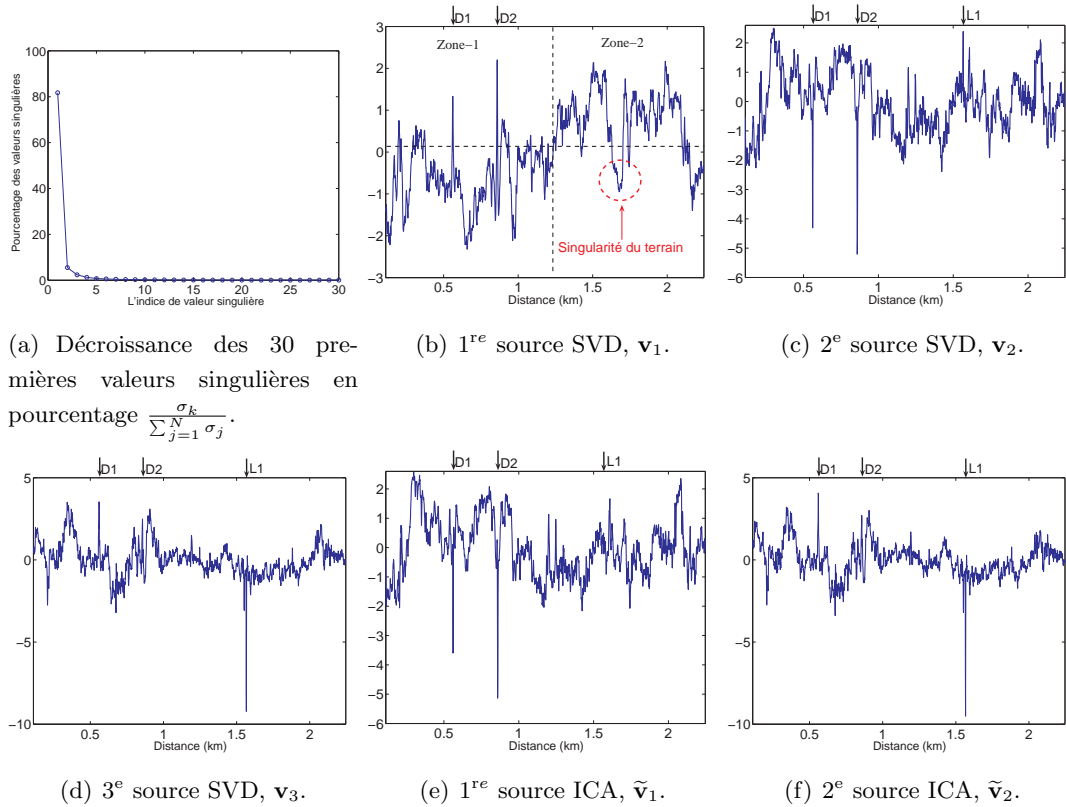
Compte tenu du fait que l'énergie d'une fuite de type percolation qui ne dure pas longtemps (la fuite  $L1$  dure environ 2 jours) est assez faible par rapport à l'énergie de données entières (6 mois), nous appliquons l'algorithme proposé en utilisant un fenêtrage glissant en temps. L'idée de cette approche est que dans une fenêtre de courte durée, l'énergie de la fuite peut devenir comparable aux énergies des phénomènes existants pendant toute la durée d'analyse. Nous avons testé plusieurs tailles de fenêtre d'analyse pour les données enregistrées sur le site d'Oraison et la fenêtre d'analyse choisie empiriquement à 14 jours donne les meilleurs résultats. Ce choix dépend de plusieurs facteurs comme la variabilité temporelle des fuites, l'énergie des fuites, etc. Nous allons revenir sur le choix de la taille de la fenêtre d'analyse dans la section discussion à la fin de ce chapitre.

Nous présentons ici uniquement les sources obtenues dans la zone de 14 jours qui contiennent les 3 fuites, les drains, ainsi que le point chaud. Cette zone correspond à la période d'analyse 30 avril - 13 mai 2005, identifiée par un carré pointillé dans la figure 3.15. Il faut mentionner ici que nous pouvons très bien analyser une autre zone temporelle de 14 jours sans rien changer.

Les 30 premières valeurs singulières sont montrées sur la figure 3.17(a) en termes de pour-

centage,  $\frac{\sigma_k}{\sum_{j=1}^N \sigma_j}$ ;  $k = 1, \dots, N$ . La première valeur singulière présente maintenant un rapport  $\sigma_1/\sum_{j=1}^N \sigma_j = 81\%$ , ce qui signifie que la 1<sup>re</sup> source issue de la SVD définit un sous-espace plus énergétique que dans le cas précédent. Cette source, montrée sur la figure 3.17(b), ne contient aucune information pertinente liée aux fuites. De plus, cette source est assez similaire à celle obtenue dans l’analyse globale des données entières (cf. figure 3.16(b)). Cette première source montre deux zones : la “Zone 1”, d’environ 0.1 km à 1.25 km et la “Zone 2”, de 1.25 km à 2.2 km.

Par identification avec la géographie du site d’Oraison, il apparaît que ces deux zones corres-



**Fig. 3.17:** Sources obtenues par l’application de la SVD suivi de l’ICA sur le résidu de la SVD pour les données d’Oraison avec un fenêtrage glissant en temps de 14 jours. Les sources montrées ici sont celles obtenues dans la fenêtre temporelle contenant les fuites artificielles.

pondent à deux niveaux d’élévation différents (la Zone 1 ayant un niveau d’élévation plus bas que la Zone 2). La présentation du site dans la section 2.3 indique que la Zone 2 est plus exposée au soleil que la Zone 1. De plus, une singularité du terrain est également identifiée sur cette source dans la zone 1.6 – 1.7 km. En conséquence, cette première source de la SVD peut être liée à la réponse du sol.

Une observation de la courbe de valeurs singulières montre que la première valeur singulière est

très bien séparée des autres. Cette observation étant consistante pour d'autres zones d'analyse de 14 jours, nous pouvons construire le premier sous-espace signal avec  $m = 1$  et appliquer ensuite l'ICA sur le sous-espace résidu.

Les deux sources SVD  $\mathbf{v}_2$  et  $\mathbf{v}_3$  (figures 3.17(c)-3.17(d)) contiennent des informations liées à la fuite L1 située à 1.566 km. Cela nous permet de choisir  $p = 2$  pour le nombre de sources à estimer par l'ICA. Les deux sources ICA sont montrées dans les figures 3.17(e) et 3.17(f). La première source ICA contient des informations liées aux drains,  $D1$  et  $D2$ , alors que la deuxième source ICA contient des informations principalement liées à la fuite  $L1$ . Nous pouvons donc choisir  $i_2 = 1$  pour détecter les fuites.

Nous comparons maintenant les informations extraites par les deux techniques, la SVD et l'ICA, du point de vue identification des fuites. La première source SVD ne contient pas des informations liées aux fuites et nous pouvons alors construire le premier sous-espace signal uniquement avec une source SVD. Cependant, la deuxième source SVD contient la fuite mélangée avec les drains (cf. figure 3.17(c)). A la différence de la deuxième source SVD, la première source ICA ne contient pas des informations liées aux fuites.

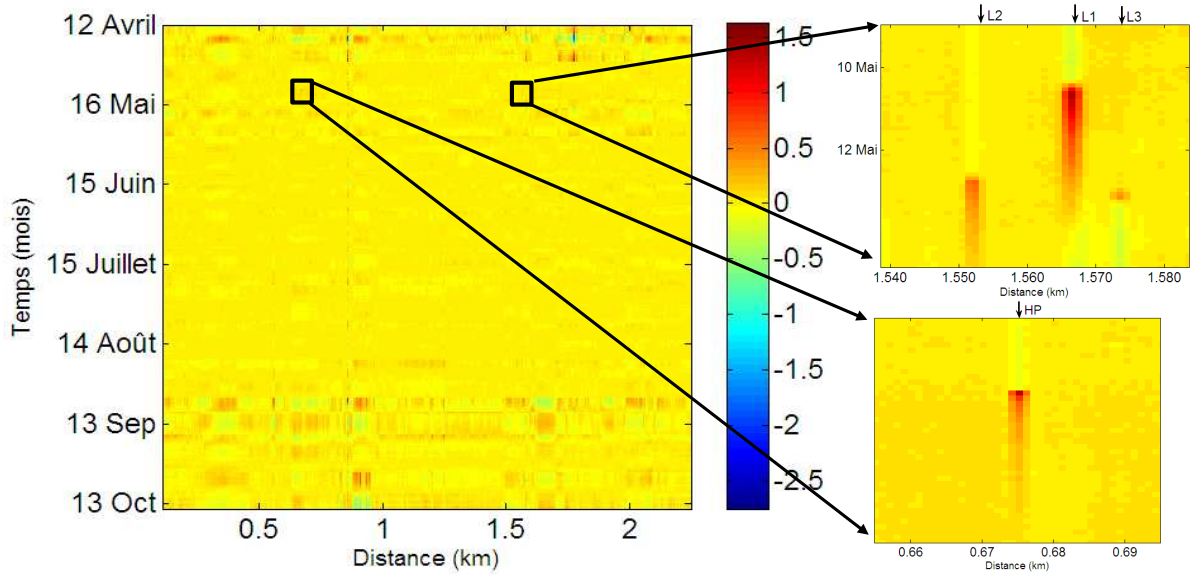
A la différence de la SVD qui extrait des sources décorréelées, l'ICA permet d'extraire des sources indépendantes qui sont associées aux fuites et aux drains. Ceci signifie qu'en utilisant la première source ICA pour construire le 2<sup>e</sup> sous-espace signal, nous pourrions obtenir une meilleure extraction des fuites dans le résidu final  $\mathbf{Z}_{\text{résidu}}$ .

### Interprétation des résidus

Les résidus obtenus dans les différentes fenêtres d'analyse peuvent être concaténés pour obtenir l'image de la figure 3.18. Un zoom autour de la zone contenant les fuites artificielles,  $L1$ ,  $L2$  et  $L3$ , montre une bonne détection et localisation de fuites artificielles. A l'aide de cette image, nous pouvons suivre également l'évolution temporelle des fuites. Pour faciliter la localisation en distance des fuites, nous considérons par la suite la projection  $y_{det}(x)$  de ce résidu sur l'axe de distances à l'aide de l'équation (3.38). Nous utilisons  $N_p = 168$ , correspondant aux 14 jours d'analyse.

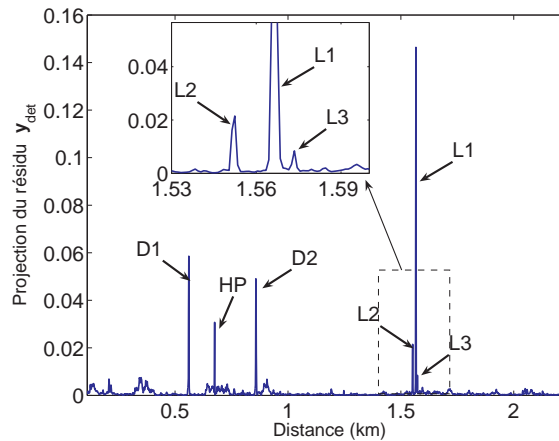
Pour les données acquises sur le site d'Oraison (voir figure 3.15) nous obtenons 13 fenêtres temporelles sans recouvrement à analyser, chacune de 14 jours. Nous considérons dans un premier temps la projection du résidu pour la période d'analyse contenant les trois fuites artificielles. Cette projection présentée sur la figure 3.19 montre la détection de phénomènes d'intérêt : les trois fuites,  $L1$ ,  $L2$ ,  $L3$  ainsi que le point chaud  $HP$ . Parmi les trois fuites,  $L1$  est détectée comme étant la plus puissante, suivi par  $L2$ . La détection de  $L3$  est cependant plus difficile comme elle se retrouve au niveau du bruit de fond. Cette faible amplitude est liée au fait qu'elle avait au départ un débit moins important et qu'elle a été lancée pendant le soir quand l'écart entre la température de l'eau et celle du sol est faible. Ce qui gêne également sa bonne détection





**Fig. 3.18:** Concaténation des résidus  $\mathbf{Z}_{\text{résidu}}$  obtenus en appliquant la méthodologie proposée sur les différentes fenêtres d'analyse de 14 jours avec  $i_2 = 1$ . Le zoom autour de la zone contenant les fuites artificielles de type percolation montre leur détection.

est le fait que la projection est une moyenne sur 14 jours alors que cette fuite existe que sur quelques heures. Nous constatons que les deux drains apparaissent aussi sur cette projection. Les drains se manifestent sur plusieurs mètres adjacents dont quelques uns existent la plupart du temps alors que les autres montrent des amplitudes importantes momentanément. Ils ont été principalement estimés par la 1<sup>re</sup> source ICA,  $\tilde{\mathbf{v}}_1$ , cependant, du à la variabilité de leurs réponses

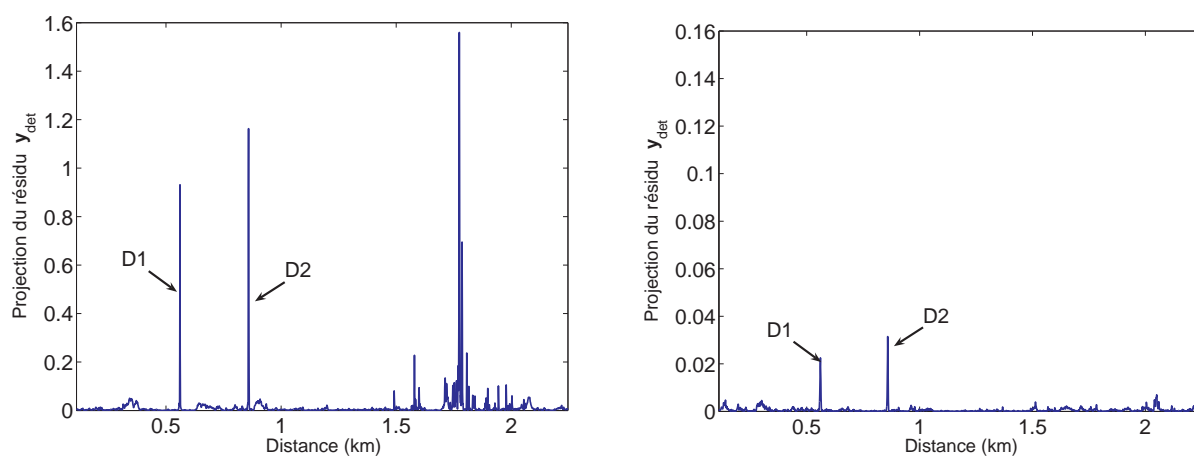


**Fig. 3.19:** Projection,  $y_{\text{det}}(x)$ , du résidu  $\mathbf{Z}_{\text{résidu}}$  sur l'axe de distance pour la fenêtre temporelle contenant les fuites artificielles.  $N_p = 14 * 12$ .

en temps et en distance, ils apparaissent aussi en partie sur la 2<sup>e</sup> source ICA, donc sur le résidu final, mais avec une énergie relativement faible.

Nous considérons ensuite une zone temporelle d'analyse avec des précipitations identifiées par les mesures météo. La projection du résidu pour cette fenêtre d'analyse, affichée sur la figure 3.20(a), révèle des amplitudes importantes sur plusieurs capteurs ainsi qu'au niveau des drains  $D1$  et  $D2$ . Nous sélectionnons ensuite une zone d'analyse qui ne contient pas de précipitation. La projection du résidu obtenu est montrée sur la figure 3.20(b). Cette projection contient les drains ayant une énergie relativement basse comme c'était le cas de l'analyse de la zone contenant les fuites, les autres capteurs ne présentant aucune information. Par conséquent, excepté les zones contenant des précipitations et les informations liées aux drains, la localisation des fuites est très bien réalisée par la méthodologie proposée.

Afin d'éviter les zones d'analyse contenant des précipitations, ce qui nous permettra de diminuer les fausses alarmes, nous pouvons considérer le critère basé sur les statistiques d'ordre supérieur, (critère présenté dans la section 2.6.3). Nous avons vu, en corroboration avec les données météorologiques, comment le skewness et le kurtosis peuvent caractériser ces phénomènes. Le kurtosis calculé à chaque instant de temps pour la période analysée, en prenant en compte tous les capteurs (toute la longueur de la fibre) est affiché sur la figure 3.21. Les nombres associés aux flèches en haut de cette image correspondent à la position de la fenêtre d'analyse. Pour la position contenant les fuites (position 2) ou pour la position 7, le kurtosis ne présente pas des informations liées à des fortes précipitations. Nous pouvons donc considérer les résultats affichés sur les figures 3.19 et 3.20(b) comme étant justes. A la position 1 à laquelle le kurtosis nous indique un changement de statistique sur plusieurs capteurs, donc la présence d'un événement



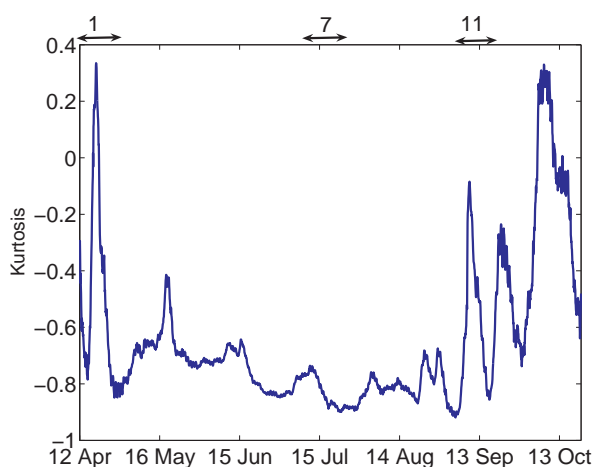
(a) Projection pour une zone contenant des précipitations risquant de générer des fausses alarmes.

(b) Projection pour une zone d'analyse sans précipitations et sans fuites.

**Fig. 3.20:** Projections,  $y_{det}(x)$ , des résidus  $\mathbf{Z}_{résidu}$  pour deux périodes d'analyse de 14 jours : avec et sans précipitations.

de type précipitation, le résultat obtenu par notre algorithme est biaisé, d'où les amplitudes importantes sur plusieurs capteurs dans la projection du résidu affichée sur l'image 3.20(a). Il faut noter que le but d'une analyse a posteriori a uniquement un sens pédagogique ici, cette analyse devrait être effectuée avant d'appliquer la méthodologie proposée.

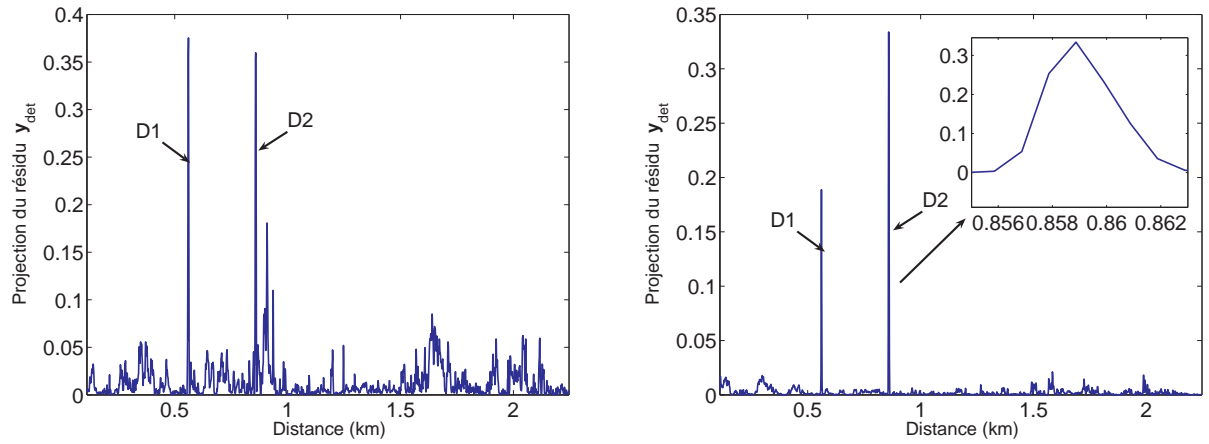
Pour montrer l'importance d'une analyse a priori des données dans le but d'éviter les zones conte-



**Fig. 3.21:** Kurtosis pour les données d'Oraison de la figure 3.15 en fonction du temps mettant en évidence des phénomènes éphémères en temps de type précipitation.

nant des précipitations, nous considérons la période d'analyse 03 - 17 septembre 2005, période indiquée par le numéro 11 sur la courbe 3.21. La projection du résidu obtenu dans cette zone d'analyse est montrée sur la figure 3.22(a). Cette projection met en évidence les drains, deux autres zones importantes tout juste après le drain *D2* ainsi que d'autres zones, principalement à partir de 1.5 km. Une étude du kurtosis montre un pic important pour cette zone d'analyse, ce qui met en évidence un phénomène énergétique sur quelques acquisitions. Une sélection de données en considérant uniquement les acquisitions pour lesquelles la valeur du kurtosis reste autour de la valeur médiane de zones précédentes  $\pm 2\sigma$ , où  $\sigma$  est la variance de l'estimation du kurtosis [Khan 08d] permet de retrouver un résidu pour lequel la projection est affichée sur la figure 3.22(b). Cette image montre que les zones mises en évidence avant la sélection de données étaient dues à des phénomènes énergétiques en non pas à des fuites.

Jusqu'au présent, nous avons montré des résultats obtenus pour des fuites simulées et nous avons montré comment nous pouvons diminuer les fausses alarmes à l'aide d'une présélection de données basée sur le kurtosis. Nous présentons sur les figures 3.23(a) et 3.23(b) les projections des résidus obtenus respectivement pour des périodes d'analyse autour de 10 octobre 2005 et de 12 décembre 2005 (cette deuxième zone n'est pas affichée sur les données de départ). Pour la première période d'analyse, les jours de précipitation ont été enlevés pour retrouver la projection du résidu de la figure 3.23(a). Cette projection met en évidence des zones des fuites réelles juste

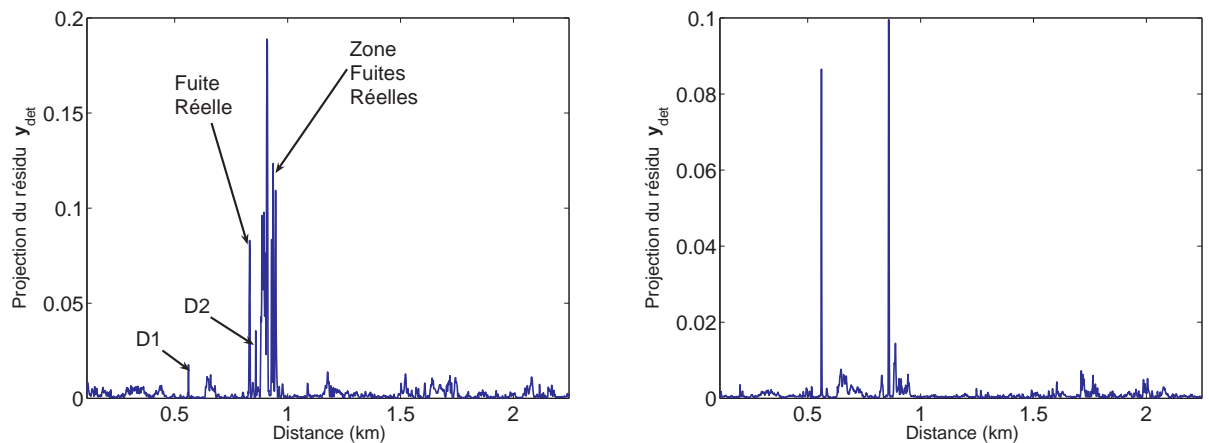


(a) Projection pour une période contenant des précipitations mettant en évidence plusieurs phénomènes.

(b) Même projection que (a) mais pour des données présélectionnées par le kurtosis. Les zones détectées précédemment sont évitées.

**Fig. 3.22:** Projections du résidu  $Z_{résidu}$  pour la période marquée par la double flèche 11 dans la figure 3.21 sans et avec sélection de données par kurtosis.

avant et après le deuxième drain. Ces fuites, qui n'apparaissent pas pendant les analyses en été, sont localisées sur plusieurs mètres. Il faut remarquer ici que c'est normal que les fuites soient détectées sur plusieurs mètres parce qu'une fuite réelle dans la digue se diffusant dans le talus, est captée par la fibre optique sur une zone étendue. Pour l'autre période d'analyse, ces fuites ne sont pas détectées.



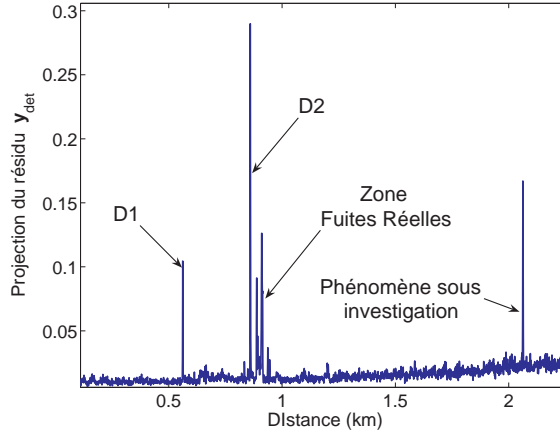
(a) Projection pour une période d'analyse autour de 10 octobre 2005 permet de détecter des fuites réelles avant et après le drain  $D2$ .

(b) Projection pour une période d'analyse autour de 12 décembre 2005 ne montre pas l'existence des fuites.

**Fig. 3.23:** Projections des résidus sur l'axe de distance,  $y_{det}(x)$  avec  $N_p = 14 * 12$ , pour deux périodes en automne et en hiver 2005.

Une autre analyse faite pour des données acquises en 2007 permet également de détecter pendant l'automne ces zones de fuites (figure 3.24), surtout celles d'après le drain *D2*. Ces zones de fuites réelles ont été confirmées par des experts EDF.

Des tests avec différentes tailles de la fenêtre d'analyse temporelle et différentes valeurs pour les



**Fig. 3.24:** Projections des résidus sur l'axe de distance,  $y_{det}(x)$  avec  $N_p = 14 * 12$ , pour les données acquises en 2007 entre le 17 et 30 octobre.

nombre de sources à estimer avec l'ICA ont été effectués. Des informations utiles complémentaires à ce qui a été montré ci-dessus n'ont pas pu être obtenues.

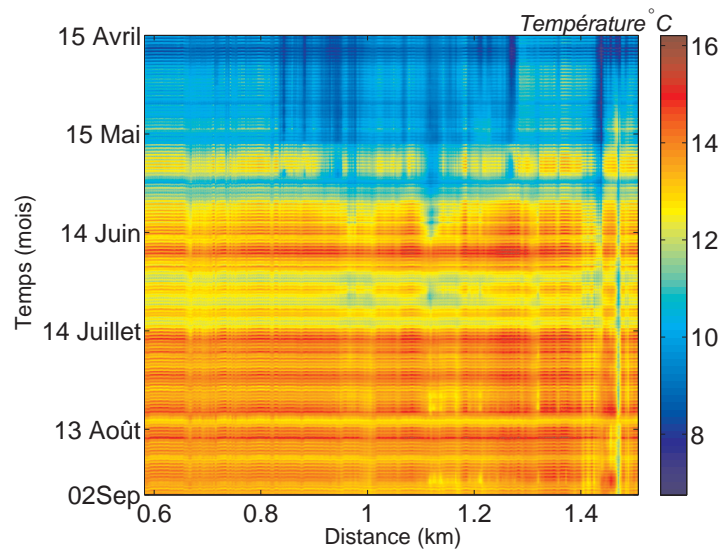
### 3.7 Application aux données enregistrées sur le site de Kembs

Le site expérimental de Kembs est un cas extrême dans le sens où il existe de nombreuses fuites et anomalies. La fibre optique de mesure située dans le contre canal sera considérée pour l'analyse dans cette section parce qu'elle est la plus susceptible de capter les fuites. Cependant, il faudra remarquer que cette fibre, complètement immergée, est fortement influencée par la température de l'eau du contre canal. En cas de fuite, la température de l'eau du grand canal vient s'ajouter aux mesures effectuées, ce qui peut résulter en un apport de chaleur (ou de froid) dépendant des conditions climatiques. Il faudra également noter qu'en fonction des saisons et du débit dans le grand canal, différentes fuites peuvent se manifester avec différentes intensités au cours du temps, ce qui a également comme conséquence un repérage biaisé de ces fuites lors des visites de terrain. Dans ce contexte, la méthodologie proposée dans cette partie utilise des paramètres différents de ceux utilisés pour l'analyse des données enregistrées sur le site d'Oraison.

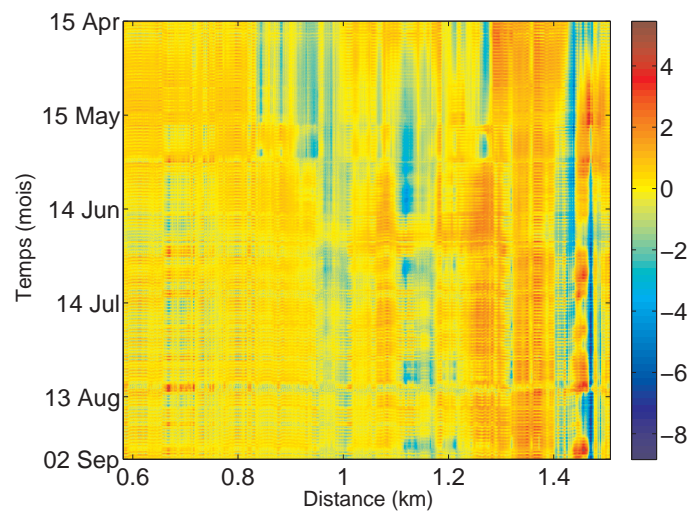
Nous présentons sur la figure 3.25(a) les données enregistrées entre 15 avril et 02 septembre 2007 sur une longueur de 0.925 km de la fibre optique du contre canal. Les données ont été enregistrées par l'appareil DTS Siemens, avec une acquisition toutes les 0.77 m et toutes les 24 minutes. Comme il y a moins d'accumulations effectuées, contrairement à l'appareil DTS Sensornet, les

données sont plus bruitées. De plus, les données montrent également une d'atténuation en fonction de la distance : plus on s'éloigne du local dans lequel le DTS est installé, plus importante est l'atténuation. Cette tendance étant linéaire, elle a été corrigée par un "detrend" des données. Il faut remarquer ici que les deux mesures aller-retour sont moyennées afin de réduire le bruit introduit par le DTS. Pour réduire d'avantage le bruit de mesure lié à un nombre moins important d'accumulations, toutes les 5 acquisitions consécutives sont moyennées, ce qui conduit à un nouveau pas d'acquisition de 2 heures.

Les données ont été ensuite normalisées afin de rendre chaque acquisition de moyenne nulle et



(a) Données thermométriques brutes.



(b) Données thermométriques normalisées.

**Fig. 3.25:** Données thermométriques enregistrées sur le site de Kembs par la fibre optique de contre canal sur 0.925 km de longueur entre 15 avril et 02 septembre 2007. Les différentes couleurs codent les températures enregistrées.

de variance unitaire. Les données obtenues après cette normalisation sont affichées sur la figure 3.25(b). Comme dans le cas précédent, un prétraitement de type désaisonnalisation n'améliore pas les résultats obtenus sur ce type de données.

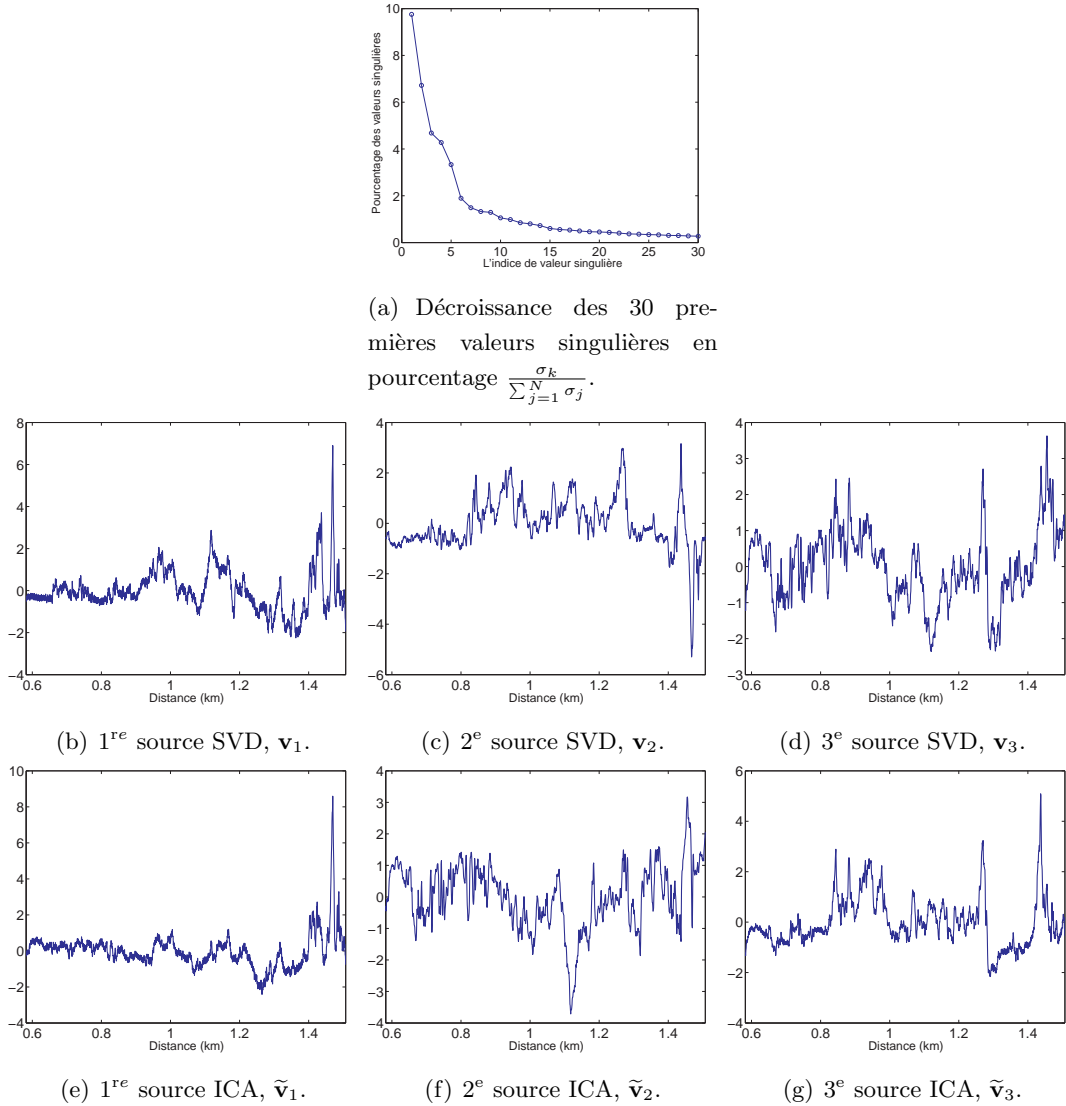
### 3.7.1 Analyse de données entières

Dans un premier temps, nous analysons les données en leur globalité. La première étape consiste à appliquer la décomposition en valeurs singulières. La figure 3.26(a) montre le pourcentage de la contribution des 30 premières valeurs singulières,  $\frac{\sigma_k}{\sum_{j=1}^N \sigma_j}$ ,  $k = 1, \dots, 30$ , obtenues en appliquant la SVD sur les données normalisées.

La faible contribution,  $\frac{\sigma_1}{\sum_{j=1}^N \sigma_j} = 9.8\%$ , de la première valeur singulière indique qu'il n'existe pas de cohérence importante entre les sources estimées par la SVD. A la différence du cas précédent (Oraison), la séparation entre la première valeur singulière et les autres n'est pas importante. Les trois premières sources SVD correspondantes sont montrées sur les figures 3.26(b)-3.26(d). La première source ne peut pas donc être considérée comme une estimation de la réponse du sol. Ceci est dû au fait que le site de Kembs est un cas extrême avec plusieurs types d'anomalies. Comme la première source SVD n'est pas attribuable à la réponse du sol, sa contribution doit être prise en compte pour la recherche des fuites. Nous pouvons alors appliquer l'ICA sur les données originales, sans enlever la première source SVD en choisissant donc  $m = 0$ . En analysant les valeurs singulières [Bihan 06], nous observons un palier après la 5<sup>e</sup> valeur. Par conséquent, 5 sources sont estimées avec ICA, dont les 3 premières sont présentées sur les figures 3.26(e)-3.26(g). Une comparaison entre les trois sources ICA et les trois sources SVD montre quelques différences. Sur les sources SVD par exemple, nous pouvons considérer que les différents phénomènes sont estimés par plusieurs sources, alors qu'avec l'ICA ils ont la tendance d'être estimés par des sources différentes du fait que l'ICA estime des sources ayant un kurtosis plus grand, donc des sources plus "impulsionnelles". Ceci est évident pour les deux pics entre 0.84-0.848 km et 0.88-0.885 km. Néanmoins, une distinction entre les différents phénomènes n'est pas évidente sur ces différentes sources. Même en augmentant le nombre de sources à estimer par ICA, les différents phénomènes ne sont pas très bien identifiés sur les sources (cf. annexe D.3.1). Pour cette raison nous appliquons la méthodologie proposée en utilisant un fenêtrage glissant.

### 3.7.2 Analyse par fenêtrage glissant en temps

Comme précédemment, la taille de la fenêtre d'analyse temporelle doit être choisie. Différentes tailles de fenêtre ont été testées, sachant qu'il existe plusieurs anomalies énergétiques susceptibles d'être décrites par les premières sources SVD. En choisissant comme précédemment une taille de 14 jours, 5 fenêtres d'analyse sont obtenues. Les pourcentages des valeurs singulières dans les différentes fenêtres d'analyse varient relativement peu en fonction de la position de la fenêtre, nous montrons sur la figure 3.27(a) les pourcentages des 30 premières valeurs obtenues en moyennant sur les 5 positions de la fenêtre. Malgré cette variabilité, il existe 4 valeurs singulières importantes pour toutes les 5 positions, c'est-à-dire, les courbes des valeurs singulières



**Fig. 3.26:** Sources obtenues par l'application de la SVD et de l'ICA sur les données de Kembs en leur globalité, après normalisation.

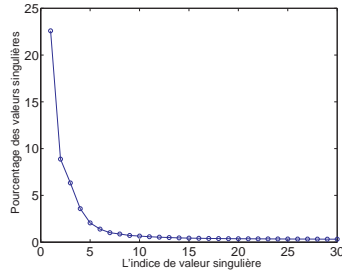
montrent un changement de pente après la 5<sup>e</sup> valeur singulière. Nous constatons que la première valeur singulière reste toujours assez faible,  $\frac{\sigma_k}{\sum_{j=1}^N \sigma_j} = 22.5\%$  en moyenne, nous indiquant que la première source SVD ne représente pas un comportement moyen significatif des données. Cette source ne peut pas être considérée comme représentative de la réponse du sol.

Construire un sous-espace signal avec cette première source peut enlever des informations importantes dans le contexte de détection des fuites. Nous choisissons alors de ne pas enlever l'information apportée par cette première source SVD, ce qui correspond à  $m = 0$ . Nous appliquons ensuite l'ICA pour estimer des sources indépendantes. Suivant la courbe de valeurs singulières, différents tests ont été effectués en choisissant  $p = 3$ ,  $p = 4$  et  $p = 5$  sources à estimer par l'ICA. Dans les figures 3.27(b)-3.27(g), les trois premières sources estimées par la

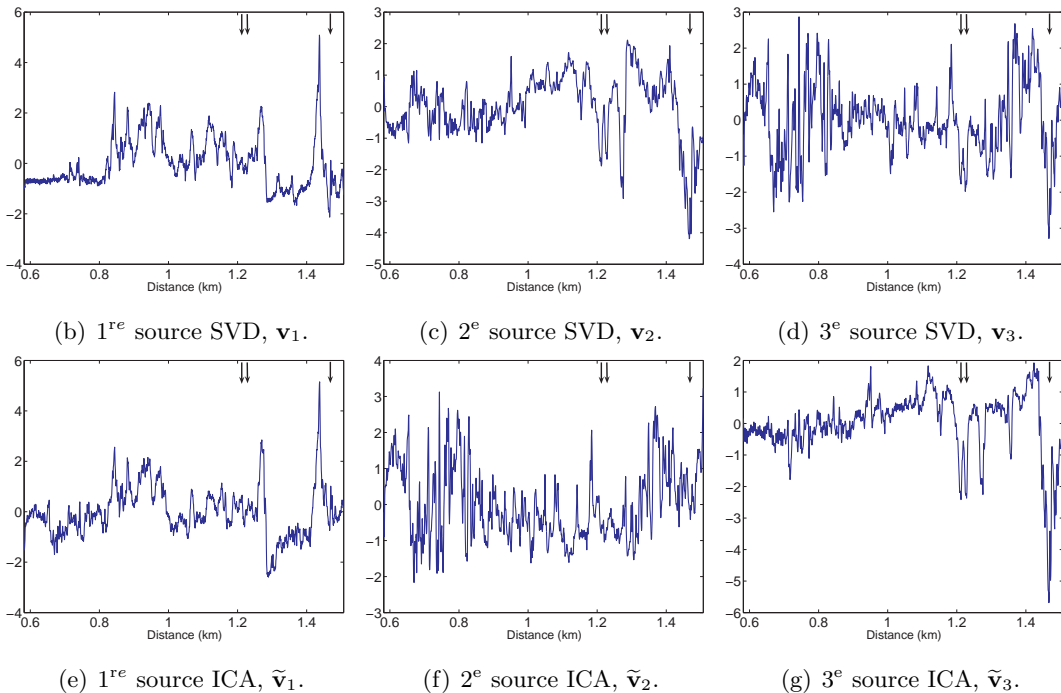


SVD sont comparées avec les trois sources estimées par l'ICA pour la deuxième position de la fenêtre glissante (29 avril - 13 mai 2007) en choisissant  $p = 3$ . Les résultats obtenus avec  $p = 4$  et  $p = 5$  sont affichés dans l'annexe D.3.2.

Les sources SVD montrent le même comportement constaté pour l'analyse de données en leur



(a) Décroissance des 30 premières valeurs singulières en pourcentage  $\frac{\sigma_k}{\sum_{j=1}^N \sigma_j}$ . Moyennage effectué sur les 5 positions de la fenêtre d'analyse.



**Fig. 3.27:** Sources obtenues par l'application de la SVD et de l'ICA pour les données de Kembs avec un fenêtrage glissant en temps de 14 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 2<sup>e</sup> fenêtre temporelle (du 29 avril au 13 mai 2007).

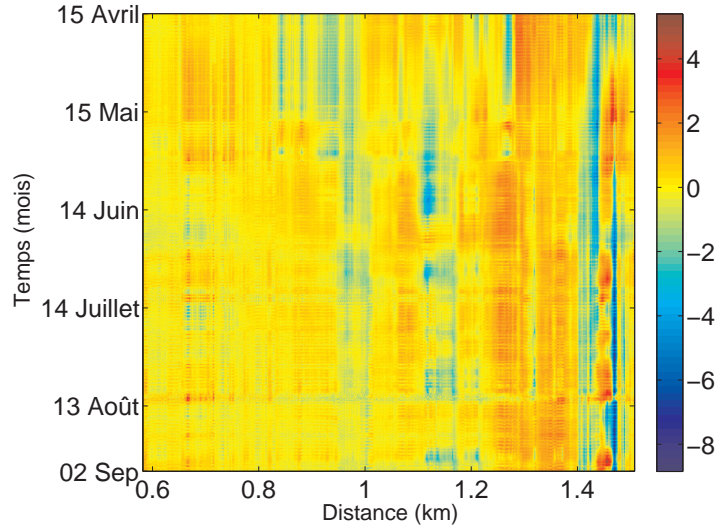
globalité, c'est-à-dire les différents phénomènes sont mélangés sur plusieurs sources. Considérant par exemple les sources SVD  $\mathbf{v}_2$  et  $\mathbf{v}_3$  (figures 3.27(c) et 3.27(d)), les mêmes pics sont identifiés sur tous les deux, comme marqué par les flèches, alors que ce n'est le cas avec les sources ICA.

Nous pouvons constater que les sources  $\tilde{\mathbf{v}}_1$  (figure 3.27(e)) et  $\tilde{\mathbf{v}}_3$  (figure 3.27(g)) contiennent plusieurs pics qui peuvent être des phénomènes intéressants du point de vue détection des fuites. Cependant, la 2<sup>e</sup> source ICA  $\tilde{\mathbf{v}}_2$  (voir figure 3.27(f)) ne contient pas des pics localisés en distance. Il faut remarquer ici que notre méthode ordonne par défaut les sources ICA par rapport à l'énergie des sous-espaces qu'elles construisent. Néanmoins, afin de ne pas enlever des phénomènes intéressants, nous devons garder la première source ICA et enlever la deuxième. Donc au lieu d'ordonner les sources en fonction d'énergie, un autre critère peut être de les ordonner en fonction de leur kurtosis, ou en fonction du kurtosis de leur variations temporelles  $\tilde{\mathbf{u}}_k$ . Quand nous ordonnons les sources en fonction des énergies de sous-espaces correspondants, il y a le risque d'écarter des fuites qui peuvent être relativement puissantes dans une fenêtre d'analyse, ce qui n'était pas le cas sur le site d'Oraison qui comportait très peu de fuites. Un ordonnancement de sources en fonction de leur kurtosis peut résoudre ce problème. Quand une source contiendra de pics localisés en distance, elle va avoir un kurtosis plus grand du fait qu'elle sera plus éloignée de la non-gaussianité. De la même façon, les variations temporelles de sources peuvent renseigner si une fuite fait son apparition à un instant donné en apportant un changement important dans la tendance générale de la variation temporelle de la source. Ce dernier critère requiert cependant de fenêtres d'analyse de taille importante afin d'avoir suffisamment de points pour estimer le kurtosis, sachant que la variance de son estimateur dépend du nombre d'échantillons sur lesquels il est calculé. En ordonnant les sources par ordre croissant de leur kurtosis, la 2<sup>e</sup> source prend le premier rang. Nous pouvons ensuite construire un sous-espace signal avec  $i_2 = 1$  afin de mettre en évidence toutes les anomalies existantes sur le site.

### Interprétation des résidus

Les résidus  $\mathbf{Z}_{résidu}$  obtenus dans les différentes fenêtres d'analyse sont concaténés pour obtenir l'image de la figure 3.28. Nous pouvons constater qu'à la différence du résidu obtenu pour les données enregistrées sur le site d'Oraison, le résidu ici n'est pas parcimonieux et contient de nombreuses zones d'anomalies, ce qui est la caractéristique du site de Kembs. Les différentes anomalies varient au cours du temps et une interprétation physique doit être accordée aux différents phénomènes avant de poursuivre l'analyse.

Pour étudier la localisation en distance des différentes anomalies, des projections  $y_{det}(x)$  de ce résidu dans les différentes fenêtres d'analyse peuvent être effectuées. Sur les figures 3.29, nous présentons ces projections pour 4 positions de la fenêtre d'analyse. Dans toutes les fenêtres, une observation intéressante est que la plupart des anomalies existent dans la zone 0.8 km - 1.5 km avec des anomalies importantes après 1.4 km. De plus, certaines anomalies sont plus évidentes dans certaines fenêtres que dans d'autres. Par exemple, pour la première période, c'est-à-dire 15 - 29 avril, les anomalies autour de 0.844 km et 0.882 km (cf. flèches dans la figure 3.29(a)) sont plus puissantes que dans les autres fenêtres d'analyses. Les différentes anomalies obtenues en utilisant cette méthode sont listées dans le tableau 3.2. Les résultats obtenus sont en cours d'investigation par EDF en les comparant à la réalité du terrain et aux résultats des méthodes



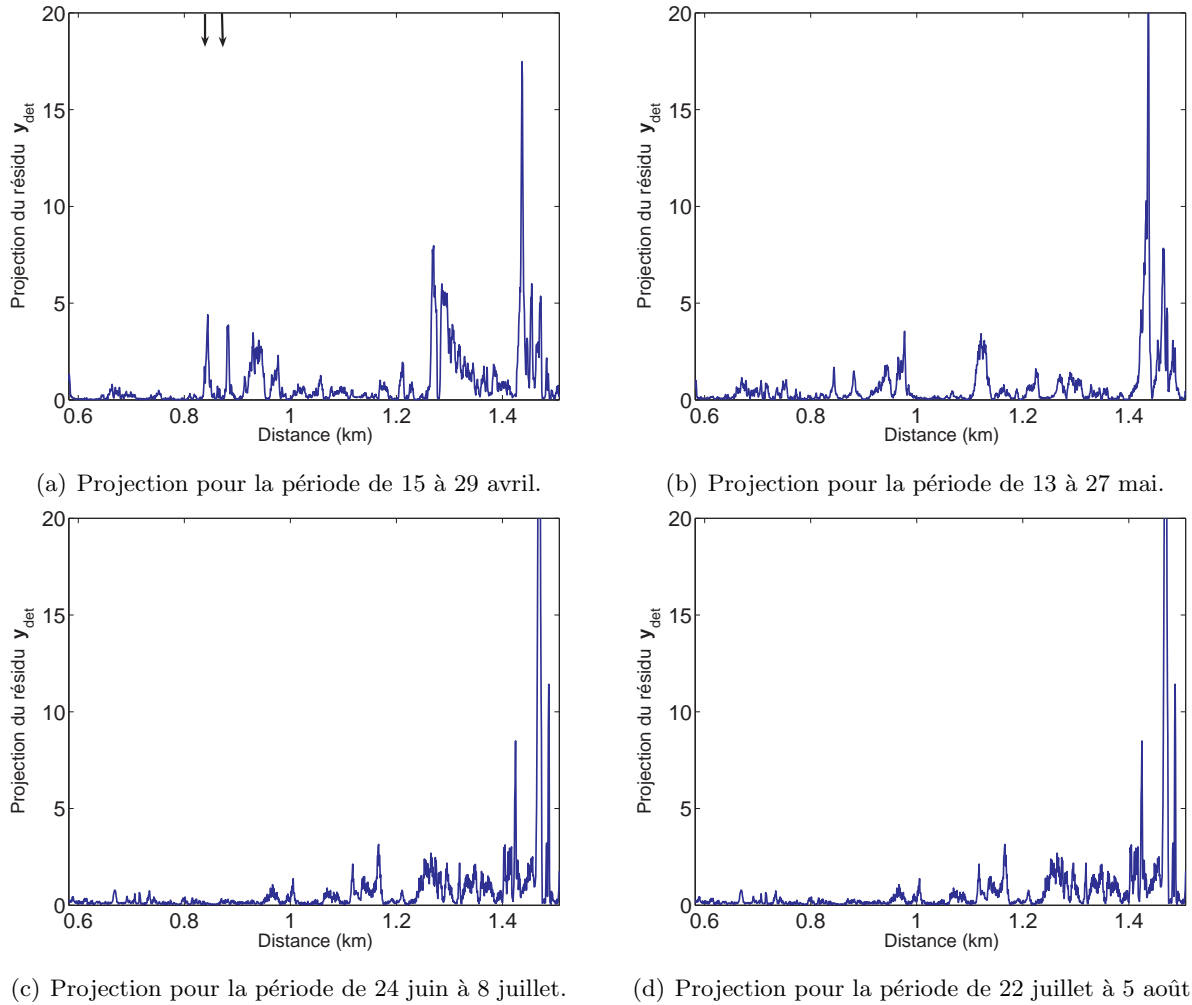
**Fig. 3.28:** Concaténation des résidus  $\mathbf{Z}_{résidu}$  obtenus en appliquant la méthodologie proposée sur les différentes fenêtres d’analyse de 14 jours. Les sources ICA sont ordonnées cette fois par rapport à leur kurtosis. Le résidu contient de nombreuses zones d’anomalies, ce qui est la caractéristique du site de Kembs. Une interprétation physique doit être accordée aux différents phénomènes avant de poursuivre l’analyse.

**Tab. 3.2:** La localisation des différentes anomalies détectées sur le site de Kembs.

|                  |                  |                  |                  |
|------------------|------------------|------------------|------------------|
| 0.840 – 0.847 km | 0.879 – 0.885 km | 0.925 – 0.945 km | 0.974 – 0.978 km |
| 1.11 – 1.135 km  | 1.206 – 1.214 km | 1.265 – 1.277 km | 1.285 – 1.298 km |
| 1.3 – 1.340 km   | 1.435 – 1.444 km | 1.452 – 1.456 km | 1.462 – 1.468 km |
| 1.469 – 1.474 km | 1.467 – 1.471 km | 1.482 – 1.488 km | 1.486 – 1.490 km |

paramétriques développées par EDF. En fonction de l’identification proposée par EDF, des analyses approfondies peuvent être effectuées. Par exemple, nous pouvons effectuer une analyse par ICA en mode déflation en utilisant la méthode FastICA [Khan 08a] pour pouvoir séparer ces différentes anomalies sur différentes sources. Nous pouvons également envisager d’appliquer l’ICA sur les projections de résidus obtenues dans les différentes fenêtres. L’étude de résidus devra également renseigner sur le seuil à choisir, par rapport à la référence du bruit dans chaque fenêtre d’analyse. Le choix d’un seuil peut conduire au fait que certains événements seront détectés à une période de temps mais pas à une autre, soit puisqu’ils n’existent pas physiquement soit parce que leur apport en température est moins important. L’analyse de données enregistrées sur le site de Kembs reste incomplète à ce stade, mais la méthode proposée permet de mettre en évidence certaines anomalies. L’analyse avec d’autres tailles de fenêtre d’analyse révèle les mêmes anomalies comme celles du tableau 3.2 (cf. annexe D.3.3).

Le critère de HOS proposé pour l’identification des phénomènes éphémères énergétiques ne marche pas pour ces données. La fibre contre canal étant principalement dans l’eau, les préci-



**Fig. 3.29:** Projections,  $y_{det}(x)$ , des résidus  $\mathbf{Z}_{résidu}$  sur l'axe de distance pour les 4 différentes zones temporelles.  $N_p = 14 * 12$ .

pitations n'ont pas une influence importante sur les températures acquises. De plus, le nombre de capteurs étant de moitié de ce qu'on avait pour le site d'Oraison, la variance des estimateurs est augmentée.

### 3.8 Discussion

Dans ce chapitre, le problème de détection des fuites a été traité comme un problème de séparation des sources. Nous avons présenté un algorithme basé sur la décomposition en valeurs singulières et l'analyse en composantes indépendantes afin de mieux séparer les fuites. Cet algorithme qui a été initialement mis en place pour la détection des fuites qui ont une durée limitée en temps a été validé sur les différents scénarios. Le schéma présenté pour l'analyse par cette approche est assez riche en termes de paramètres qui peuvent être choisis pour l'adapter aux

différentes situations. Ces paramètres sont :

- “ $m$ ”, le nombre de sources SVD à garder pour construire le premier sous-espace signal avec la SVD, correspondant à la réponse du sol,
- “ $p$ ”, le nombre de sources à estimer avec l’ICA,
- la taille de la fenêtre d’analyse
- “ $i_2$ ”, le nombre de source ICA pour reconstruire le deuxième sous-espace signal.

Nous avons montré dans le cas des données simulées et des données enregistrées sur le site d’Oraison que la première valeur singulière est assez importante en termes d’énergie. La fibre optique est enterrée dans le sol et les mesures de température mettant en évidence la réponse du sol. La concordance de la première source SVD avec la géographie du site permet de la relier à la réponse du sol. Un choix  $m = 1$  permet alors de séparer la réponse du sol pour ces deux cas d’analyse.

Les fuites sont des phénomènes qui ont une énergie faible par rapport à l’énergie de données entières. Elles ne sont pas détectables si la taille temporelle d’analyse est trop élevée. Une analyse par fenêtrage glissant en temps est alors proposée, qui signifie que dans une fenêtre d’analyse l’énergie de la fuite devient comparable par rapport aux données analysées. Pour les nombre de sources “ $p$ ” à estimer avec ICA, un choix  $p = 2$  a été donné pour séparer les drains des fuites. Les sources ICA ne sont pas ordonnées par défaut. Une façon de les ordonner est par rapport à leurs énergies en utilisant les valeurs singulières modifiées par ICA. Cet ordonnancement considère le fait que les fuites ont une énergie plus faible par rapport aux drains. Une observation de ces sources permet de sélectionner le paramètre “ $i_2$ ”, le choix de  $i_2 = 1$  donnant les meilleurs résultats.

Un deuxième paramètre à choisir est la taille de la fenêtre d’analyse pour cet algorithme. Nous avons fait des tests avec différentes tailles de fenêtre d’analyse pour les données contrôlées (les données simulées et les données d’Oraison avec fuites de type percolation) pour retrouver qu’une fenêtre d’analyse de 14 jours présente le meilleur compromis. Cependant, trouver un critère permettant le choix automatique de ce paramètre n’est pas évident puisque ce paramètre peut changer d’un site à l’autre. Plusieurs facteurs peuvent influencer ce choix, les plus importants étant le débit des fuites, la variation temporelle de la fuite recherchée, etc.

L’algorithme proposé repose surtout sur le fait qu’une fuite montrera une variation au cours du temps, sinon elle apparaîtra comme un drain. Dans le cadre de la surveillance des digues, cet algorithme sera assez efficace pour mettre en évidence de nouvelles fuites et les fuites qui montrent des variations en termes de débit. Pour des fuites qui montrent des variations au cours des saisons, soit une taille de fenêtre d’analyse très importante sera requise soit l’analyse de l’ensemble des données sur différentes saisons rééchantillonnées en temps.

Le site de Kembs a plusieurs anomalies, dont la nature n’est pas toujours attribuable aux fuites. L’application de la SVD donne une première valeur singulière qui ne compte que pour 10% de toutes les valeurs singulières. Contrairement au site d’Oraison, les différentes valeurs singulières pour Kembs ne sont pas bien séparées. La première source n’est pas alors cohérente pour toutes

les données. Une valeur de  $m = 0$  est alors posé pour ne pas construire de sous-espace signal avec la SVD. Les sources ICA sont cette fois-ci ordonnées en termes de leurs kurtosis au lieu de leurs énergies. Un choix  $i_2 = 1$  permet de mettre en évidence plusieurs anomalies sur le site de Kembs. La nature et la cause de ses anomalies ne sont pas connues avec certitude, même si certaines de ces anomalies correspondent aux fuites réelles qui ont été repérées lors des visites et d'analyses visuelles de terrain. L'étude de Kembs reste alors incomplète à ce stade.

Il faut mentionner ici que d'autres prétraitements du type gradient spatial peuvent être appliqués avant l'analyse. Le gradient spatial consiste en l'application de l'opérateur gradient à chaque profil de température. Ceci permet d'atténuer les effets des facteurs qui varient lentement au cours de la distance comme la réponse du sol ainsi que les variations saisonnières. Les données peuvent également être traitées dans l'orientation transposée en recherchant les sources en fonction du temps au lieu de la distance. Comme l'intérêt principal de ce travail est la localisation des fuites en distance, nous restons sur la recherche des sources en distance. Quelques uns de ces cas ont été présentés dans [Khan 08e]. Un sommaire des résultats obtenus avec ces différentes prétraitements pour les données d'Oraison est présenté en annexe D.2.2.



# Chapitre 4





## Chapitre 4

# Détecteur des singularités par approche journalière

### Sommaire

---

|            |   |            |
|------------|---|------------|
| <b>4.1</b> | <b>Introduction</b>   | <b>137</b> |
| <b>4.2</b> | <b>Méthodologie pour la détection des singularités</b>            | <b>137</b> |
| 4.2.1      | Définition de singularités  | 138        |
| 4.2.2      | Principe d'identification de singularités                         | 138        |
| 4.2.3      | Mesure de dissimilarité   | 139        |
| <b>4.3</b> | <b>Détection de fuites par seuillage d'histogramme</b>            | <b>141</b> |
| 4.3.1      | Seuillage en fixant une probabilité de fausse alarme              | 145        |
| <b>4.4</b> | <b>Système proposé pour la détection de singularités</b>          | <b>148</b> |
| <b>4.5</b> | <b>Application aux données simulées</b>                           | <b>150</b> |
| <b>4.6</b> | <b>Application aux données enregistrées sur le site d'Oraison</b> | <b>154</b> |
| <b>4.7</b> | <b>Discussion</b>   | <b>159</b> |

---



## 4.1 Introduction

Nous avons présenté dans le chapitre précédent une méthode pour la détection des fuites dans les digues basée sur des techniques de séparation de sources. Cette approche est utile quand nous disposons d'un grand nombre d'acquisitions, donc dans le cadre d'analyse des données à long terme. Dans ce chapitre nous développons un système de détection instantanée qui a pour but d'identifier les anomalies ayant un comportement différent au cours d'une journée. Même si la méthodologie proposée est proche de celle présentée dans le chapitre précédent, nous présentons une autre façon d'analyser les données qui repose sur l'analyse des variations journalières de température à différentes distances. Le but est d'identifier et de localiser les singularités de type :

- structures existantes (drains, regards, etc.),
- fuites,
- singularités du terrain, etc.,

à partir des acquisitions effectuées pendant une journée, en considérant que les températures enregistrées pour les zones de singularité montrent des variations différentes de celles enregistrées pour les zones non-singulières et homogènes.

Nous proposons dans un premier temps une mesure de dissimilarité qui permet de détecter les singularités en distance. Nous modélisons ensuite la distribution de probabilité de cette dissimilarité par un mélange de distributions gamma et uniforme. Les paramètres de ce mélange sont estimés par une approche de maximum de vraisemblance. À l'aide de cette formulation, un seuillage est effectué en considérant une probabilité de fausse alarme. Ceci permet de ne pas avoir un seuil fixe mais un seuil variable en fixant la probabilité de fausse alarme. La méthode est validée sur les données simulées ainsi que sur les données réelles acquises sur le site d'Oraison.

Dans le cadre de la méthode actuelle, l'idée de base repose sur l'hypothèse d'avoir un grand nombre de zones non-singulières par rapport aux zones singulières. Pour le site de Kembs, le fait d'avoir la fibre optique d'acquisition du contre-canal immergée dans l'eau ainsi qu'un grand nombre de fuites limite l'exploitation de cette méthode pour ce site. Le calcul d'un vecteur de référence à partir des données acquises conduit à des mesures de dissimilarité biaisées et alors les résultats sont inexploitablement tels quels. Nous ne présentons pas ici les résultats de l'application du système proposé sur les données de Kembs, ceux-ci étant montrés dans l'annexe E.1.

## 4.2 Méthodologie pour la détection des singularités

Des données thermométriques ont été modélisées dans le chapitre 2 par un mélange des plusieurs facteurs parmi lesquels : les fuites, les structures existantes comme les drains, la réponse du sol, les variations journalières, les précipitations, etc. Du point de vue de l'identification des anomalies, il est important de pouvoir localiser en distance ces anomalies.

Les données thermométriques peuvent être reformulées de la manière suivante :

$$\mathbf{Y}^i = \{y^i(t, x) \mid 1 \leq t \leq N_t, 1 \leq x \leq N_x\}, i = 1, \dots, N_i, \quad (4.1)$$

avec  $N_t$  le nombre d'acquisitions par jour,  $N_x$  le nombre de capteurs (la longueur de la fibre) et  $N_i$  le nombre de jours analysés.

#### 4.2.1 Définition de singularités

Les singularités sont définies ici en fonction du déplacement le long la fibre optique. Nous groupons les distances d'acquisition de température en deux types de zones :

- zones non-singulières
- zones singulières

##### Les zones non-singulières

Nous définissons ces zones comme celles représentant le comportement général du terrain sans existence d'anomalies. Ces zones sont identifiées par les acquisitions les plus cohérentes. Une notion d'homogénéité peut aussi être accordée à ces zones en considérant que le sol à différentes distances est composé d'un même matériau. Cependant, cette définition est vague puisque sur une longueur plus étendue de la fibre, ils peuvent exister des matériaux différents sans qu'aucune anomalie ne soit présente.

##### Les zones singulières

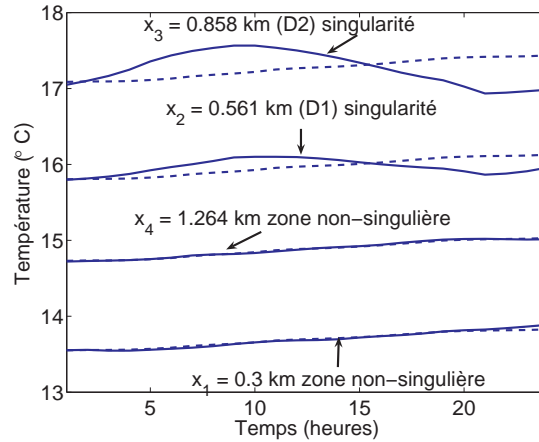
Nous définissons une singularité comme tout point d'acquisition montrant une anomalie en distance. Cette anomalie dans le contexte actuel correspond à une variation de température au cours d'une journée qui sera différente des variations de température cohérentes mesurées dans des zones non-singulières.

Il faut remarquer ici qu'il est assumé qu'il existe un plus grand pourcentage de distances sans singularités que de distances avec singularités. Basé sur la définition de singularités ci-dessus, un détecteur est proposé en exploitant le fait que les variations journalières de température pour les singularités sont différentes de celles pour les non-singularités.

#### 4.2.2 Principe d'identification de singularités

Afin d'identifier et de localiser les singularités, le détecteur proposé utilise une mesure de dissimilarité basée sur l'écart entre une acquisition journalière de température d'une singularité et les acquisitions effectuées dans des zones non-singulières [Khan 08c]. Comme la méthode est basée sur une approche heuristique, nous utiliserons des acquisitions réelles de température pour décrire cette méthode. Considérons par exemple une acquisition journalière effectuée sur le site d'Oraison à 4 distances différentes :

- $x_1 = 0.3$  km ; une distance sans singularité,



**Fig. 4.1:** *Lignes continues* : Les profils de température enregistrés sur 24 heures à 4 distances différentes sur le site d'Oraison. Les zones singulières ici sont les deux drains  $D1$  (0.561 km) et  $D2$  (0.859 km) qui présentent des variations journalières différentes par rapport aux zones non-singulières (0.3 km et 1.264 km). *Lignes pointillées* :  $\mathbf{Y}_{sig}^i$  le même jour et aux mêmes distances, montrant que les zones singulières présentent des écarts plus importants par rapport au vecteur de référence que les zones non-singulières.

- $x_2 = 0.561$  km ; une distance correspondant au drain  $D1$  (singularité),
- $x_3 = 0.859$  km ; une distance correspondant au drain  $D2$  (singularité),
- $x_4 = 1.264$  km ; une distance non-singulière.

Les variations de température pour ces 4 distances pendant 24 heures sont affichées sur la figure 4.1 par des lignes continues. Nous pouvons constater dans un premier temps que les différents signaux ont des amplitudes différentes, suggérant qu'ils ne réagissent pas de la même manière aux mêmes conditions. De façon plus importante, nous pouvons constater que les drains présentent des variations différentes par rapport aux zones non-singulières. Entre eux, les zones non-singulières montrent des variations relativement proches. Nous pouvons alors construire un critère qui mesure la distance (dans le sens d'écart) entre ces variations journalières pour les différentes distances. En considérant les variations journalières comme des vecteurs indexés par le temps, l'idée repose sur la recherche d'un vecteur de référence parmi les variations journalières enregistrées à toutes les distances. Ensuite, ce vecteur de référence pourra être comparé avec tous les vecteurs (variations journalières) enregistrés à toutes les distances. Suivant l'hypothèse d'avoir plus de zones de non-singulières que des zones singulières, ce vecteur de référence sera représentatif des zones non-singulières. Une déviation par rapport à cette référence sera alors une mesure de singularité.

### 4.2.3 Mesure de dissimilarité

Le vecteur de référence doit être estimé comme le vecteur le plus cohérent parmi tous les vecteurs enregistrés à toutes les distances. La décomposition en valeurs singulières (SVD) est

une approche classique qui permet de retrouver la composante la plus cohérente des données à décomposer. La SVD (voir la définition donnée dans la section 3.2) de la matrice de données représentant un jour d'acquisition,  $\mathbf{Y}^i \in \mathbf{R}^{N_t \times N_x}$ , est :

$$\mathbf{Y}^i = \mathbf{U}_N^i \Delta_N^i \mathbf{V}_N^{iT} = \sum_{j=1}^N \sigma_j^i \mathbf{u}_j^i \mathbf{v}_j^{iT}, \quad (4.2)$$

avec  $N = \min(N_t, N_x)$ ,  $\Delta_N^i \in \mathbf{R}^{N \times N}$  la matrice diagonale des valeurs singulières,  $\sigma_j^i \geq 0$ , et  $\mathbf{U}_N^i \in \mathbf{R}^{N_t \times N}$  et  $\mathbf{V}_N^i \in \mathbf{R}^{N_x \times N}$  les matrices orthogonales. Dans le cas actuel, les vecteurs de gauche,  $\mathbf{u}_j^i$ , qui représentent les colonnes de la matrice  $\mathbf{U}_N^i$ , sont des fonctions indexées par le temps. Une des propriétés de la SVD est qu'en la tronquant, elle donne la meilleure approximation (au sens des moindres carrés) de rang  $q < N$ , d'une matrice  $\mathbf{Y}^i$  de rang  $N$  [L.L. Scharf, 91, chap. 2]. La SVD peut alors être utilisée comme une première étape pour identifier un vecteur de référence qui donne la meilleure approximation de données journalières. Ce vecteur est donc estimé comme le premier vecteur de gauche,  $\mathbf{u}_1^i$ . Un sous-espace signal,  $\mathbf{Y}_{sig}^i$ , pourra être ensuite construit avec ce vecteur de référence et ses variations temporelles,  $\mathbf{v}_1^i$  :

$$\mathbf{Y}_{sig}^i = \sigma_1^i \mathbf{u}_1^i \mathbf{v}_1^{iT}. \quad (4.3)$$

Dans le chapitre 3, le sous-espace signal a été identifié comme le sous-espace représentant la réponse du sol. Dans le cas actuel, il peut être interprété comme un sous-espace référence construit par le vecteur de référence,  $\mathbf{u}_1^i$ , extrait à partir des variations journalières de température acquises à toutes les distances.

Nous analysons par la suite le sous-espace de référence,  $\mathbf{Y}_{sig}^i$ , obtenu pour l'acquisition journalière effectuée sur le site d'Oraison (données affichées sur la figure 4.1). Les courbes en pointillé affichées sur la même figure montrent ce sous-espace de référence sur les 4 distances  $x_1, x_2, x_3$  et  $x_4$  définies dans la section précédente. Deux de ces distances (les drains  $D1$  et  $D2$ ) correspondent aux singularités alors que les deux autres aux zones non-singulières (0.3 km et 1.264 km). Nous pouvons constater que pour les zones non-singulières, le sous-espace de référence est proche des variations journalières du signal d'origine. Au contraire, pour les deux distances singulières (drains), les écarts entre le sous-espace de référence et les variations journalières sont plus importants.

Nous pouvons alors considérer le sous-espace de référence,  $\mathbf{Y}_{sig}^i$ , comme une mesure de similarité qui permettra de mettre en évidence les zones non-singulières. Dans l'autre sens, nous pouvons construire une mesure de dissimilarité qui quantifie la déviation entre les variations journalières d'origine et le sous-espace de référence estimé pour chaque distance. Cette déviation est caractérisée par le sous-espace résidu

$$\mathbf{Y}_{résidu}^i = \mathbf{Y}^i - \mathbf{Y}_{sig}^i, \quad (4.4)$$

qui contient alors les informations liées aux singularités. Une mesure de dissimilarité mettant en évidence les singularités peut être la norme  $L_2$ ,  $\|\cdot\|_2$  de chaque colonne  $\mathbf{y}_{résidu}^i(x)$  du sous-espace

résidu  $\mathbf{Y}_{résidu}^i$ , parce que chaque colonne de ce sous-espace représente la déviation par rapport au vecteur de référence. La dissimilarité  $d^i(x)$  à une distance  $x$  est alors donnée par :

$$d^i(x) = \|\mathbf{y}_{résidu}^i(x)\|_2, x = 1, \dots, N_x; i = 1, \dots, N_i. \quad (4.5)$$

D'autres normes peuvent être utilisées, cependant des tests ont montré que la norme  $L_2$  donne les meilleurs résultats.

Considérons le cas idéal pour lequel toutes les zones non-singulières sont identiques et pour lequel il existe quelques zones singulières, pas nécessairement identiques. Pour un tel cas, la mesure de dissimilarité (4.5) ne contiendra que les informations liées aux singularités. Autrement dit, la courbe (l'histogramme) de dissimilarités pour un jour  $i$ ,  $d^i(x)$ , sera comme celle montrée sur la figure 4.2(a) (voir page 146), indiquant le fait que la dissimilarité tend vers 0 pour la plupart des distances. Dans un tel cas idéal, la dissimilarité doit être non-nulle seulement pour les distances correspondant aux singularités, avec une distribution uniforme dans la gamme  $m^i = ]0, \max(d^i(x)) ]$ .

Dans un cas réel, les zones non-singulières montrent une réponse similaire mais pas tout à fait identique. Ceci, couplé au fait que le vecteur de référence est une estimation de variations journalières de température pour les zones singulières, conduira à une dispersion autour de zéro de l'histogramme de dissimilarité. Par exemple, la figure 4.2(b) affichée à la page 146 montre un tel histogramme obtenu pour une acquisition effectuée sur le site d'Oraison. Afin de détecter les singularités, nous pouvons réaliser un seuillage de cet histogramme. Le seuillage est effectué en considérant une probabilité de fausse alarme. Ceci permet de ne pas avoir un seuil fixe mais un seuil variable en fixant la probabilité de fausse alarme.

### 4.3 Détection de fuites par seuillage d'histogramme

L'histogramme de dissimilarité pour les zones non-singulières peut être associé à une densité de probabilité puisque informellement, une densité de probabilité peut être vue comme la limite d'un histogramme si nous disposons d'un échantillon suffisamment important de valeurs. Sachant que les mesures de dissimilarité pour ces zones sont dominées par des valeurs positives proches de zéro et que le vecteur de référence est estimé en appliquant la SVD sur les données acquises à toutes les distances, nous pouvons considérer la loi gamma pour cette densité. La distribution de gamma dépend de deux paramètres [Papoulis 91, Schervish 01] :

$$f_{\Gamma}^i(d^i; \alpha^i, \beta^i) = \frac{(d^i)^{\alpha^i-1} \exp\{-d^i/\beta^i\}}{\beta^{\alpha^i} \Gamma(\alpha^i)}, \quad (4.6)$$

avec  $\Gamma(\cdot)$  la fonction gamma et  $\alpha$  et  $\beta$  les deux paramètres de la distribution gamma dépendant du jour analysé  $i$ . Ces deux paramètres donnent une grande flexibilité à la distribution gamma, avec  $\alpha$  le paramètre de forme et  $\beta$  le paramètre d'échelle. La densité gamma appartient à la famille d'exponentielles et englobe la densité exponentielle, la densité khi-2 et la densité d'Erlang comme



cas particuliers [SO94]. La distribution gamma prend explicitement en compte la positivité des mesures de dissimilarité à modéliser. Bien qu'aucun choix n'est totalement objectif, la densité de gamma est un choix flexible en termes d'adaptativité aux différents scénarios (dimension, variabilité, etc.) et sera exploitée par la suite pour modéliser les zones non-singulières dans les mesures de dissimilarité.

Au contraire, les zones singulières (marquées par des cercles sur la figure 4.2(b) affichée sur la page 146), comme nous l'avons précisé dans le cas idéal, auront des dissimilarités distribuées entre 0 et une valeur maximale. En supposant que les singularités ne sont pas corrélées entre elles et qu'elles peuvent apparaître à différentes distances avec différentes intensités, nous pouvons les modéliser par une distribution uniforme formulée comme [Papoulis 91, Schervish 01] :

$$f_u^i(d^i) = \begin{cases} \frac{1}{m^i} & \text{si } 0 \leq d^i \leq m^i \\ 0 & \text{sinon} \end{cases} \quad (4.7)$$

où  $m^i = \max(d^i)$ . L'histogramme de dissimilarités  $d^i(x)$ , peut alors être modélisée comme un mélange de distributions gamma et uniforme. La densité de probabilité  $f^i(d^i)$  de  $d^i(x)$  peut s'écrire comme :

$$f^i(d^i) = \pi^i f_\Gamma^i(d^i; \alpha^i, \beta^i) + (1 - \pi^i) f_u^i(d^i), \quad (4.8)$$

où  $\pi^i$  représente la proportion de la distribution gamma dans le mélange.

Les paramètres définissant ce mélange sont les paramètres des distributions constituantes ainsi que la proportion de chaque distribution. Les paramètres à estimer pour un jour  $i$  sont donc :

$$\theta^i = (\pi^i, \alpha^i, \beta^i, m^i).$$

Ces paramètres sont souvent estimés en utilisant une approche de maximum de vraisemblance [Kay 91]. Pour la distribution uniforme définie par l'équation (4.7), le paramètre à estimer  $m^i$  est tout simplement la valeur maximale de dissimilarité pour le jour correspondant. Les paramètres de la distribution gamma et sa contribution dans le mélange (4.8) doivent ensuite être estimés.

La maximisation de la vraisemblance est souvent simplifiée en maximisant son logarithme, qui est une fonction monotone et croissante. En considérant les mesures de dissimilarité effectuées aux distances  $\{1, 2, \dots, x, \dots, N_x\}$  comme un échantillon de  $N_x$  observations indépendantes tirées de la distribution  $f^i(d^i)$ , le logarithme de la fonction de vraisemblance s'écrit comme [Hastie 01] :

$$\begin{aligned} \mathcal{L}(d^i | \alpha^i, \beta^i, m^i) &= \sum_{x=1}^{N_x} \log f^i(d^i(x)) \\ &= \sum_{x=1}^{N_x} \log [\pi^i f_\Gamma^i(d^i(x); \alpha^i, \beta^i) + (1 - \pi^i) f_u^i(d^i(x))] \end{aligned} \quad (4.9)$$

La maximisation numérique directe de  $\mathcal{L}$  est difficile à cause de la somme des termes dans le logarithme. Une approche plus simple repose sur la procédure appelée Espérance-Maximisation (EM), ou Expectation Maximization en anglais [Dempster 77]. Initialement développée pour le

cas de données manquantes (non-observées), son utilisation pour simplifier la vraisemblance est devenu classique [McLachlan 08]. Dans le contexte de mélange des distributions, le problème est formulé artificiellement comme un problème de données incomplètes en augmentant les échantillons par l'ajout des paramètres latents. Considérons les variables latentes,  $\Delta_x^i \in \{0, 1\}$  avec  $Pr(\Delta_x^i = 1) = \pi^i$ , ou, autrement dit, si  $\Delta_x^i = 1$  alors  $d_x^i$  vient du modèle gamma, sinon, du modèle uniforme. Les données complètes incorporant les variables latentes sont notées par  $T = (d^i(x), \Delta_x^i)$ . En supposant que nous connaissons les valeurs de  $\Delta_x^i$ , le logarithme de vraisemblance pour ces nouvelles données complètes peut être écrit comme [Hastie 01] :

$$\begin{aligned} \mathcal{L}_0(d^i | \alpha^i, \beta^i, m^i, \Delta_x^i) &= \sum_{x=1}^{N_x} [\Delta_x^i \log f_{\Gamma}^i(d^i(x); \alpha^i, \beta^i) + (1 - \Delta_x^i) \log f_u^i(d^i(x))] \\ &+ \sum_{x=1}^{N_x} [\Delta_x^i \log \pi^i + (1 - \Delta_x^i) \log(1 - \pi^i)], \end{aligned} \quad (4.10)$$

ce qui signifie que toutes les mesures de dissimilarité  $d^i(x)$  pour lesquelles  $\Delta_x^i = 1$  permettront d'estimer les paramètres de la distribution gamma en maximisant (4.10) par rapport aux paramètres de la distribution gamma. L'estimation de  $\pi^i$  donnera la proportion des  $\Delta_x^i = 1$ . Cependant, les  $\Delta_x^i$  ne sont pas connus. Nous adoptons alors une approche itérative en remplaçant les  $\Delta_x^i$  dans (4.10) par leurs espérance mathématique [Hastie 01] :

$$\gamma_x^i = E(\Delta_x^i | \theta^i, d^i(x)) = Pr(\Delta_x^i = 1 | \theta^i, d^i(x)), \quad (4.11)$$

ces valeurs étant nommées responsabilités de la distribution gamma pour l'observation  $d^i(x)$ . Ces responsabilités et les paramètres de mélanges sont ensuite estimés par l'algorithme EM.

L'algorithme EM est une approche couramment utilisée pour le calcul itératif des estimations du maximum de vraisemblance dans le cas des variables latentes non observables. L'appellation EM a été donnée par Dempster et al. [Dempster 77] parce que chaque itération de l'algorithme consiste en deux étapes : *étape E* : évaluation de l'espérance ou Expectation-step en anglais (E-step) et *étape M* : maximisation ou Maximization-step en anglais (M-step). Dans l'étape E, les données manquantes (non observées) sont estimées à l'aide des données observées et des estimations actuelles des paramètres du modèle,  $\hat{\theta}^i(j)$ . Dans le cas de mélange, cette étape consiste à l'estimation des responsabilités de l'équation (4.11). Plus précisément, l'étape E consiste à construire le logarithme de vraisemblance de données complètes (équation (4.10)). Comme celui-ci dépend partiellement de données non observées, nous les remplaçons par leurs espérances conditionnelles par rapport aux données observées,  $Q(\gamma_x^i, \hat{\theta}^i(j))$ . Dans l'étape M, la fonction de vraisemblance,  $\mathcal{L}_0$  est maximisée sous l'hypothèse de données latentes connues, les données latentes étant remplacées par leurs estimations calculées dans l'étape E. Suivi d'une initialisation, les deux étapes sont répétées jusqu'à la convergence. La convergence est assurée parce que l'algorithme augmente la vraisemblance à chaque itération. L'algorithme EM peut être formulé généralement comme :

1. Initialiser les paramètres  $\hat{\theta}^i(0)$ .

2. *étape E* : pour la  $j^e$  itération, calculer les espérances conditionnelles des  $\gamma_x^i$  par rapport aux paramètres du mélange

$$Q\left(\gamma_x^i, \hat{\theta}^i(j)\right) = E\left(\mathcal{L}_0(\gamma_x^i; T \mid d^i(x), \hat{\theta}^i(j))\right)$$

où  $T = (d^i(x), \Delta_x^i)$  représente les données complètes incorporant les variables latentes remplacées par les responsabilités  $\gamma_x^i$ .

3. *étape M* : estimer les nouveaux paramètres  $\hat{\theta}^i(j+1)$  qui maximisent  $Q\left(\gamma_x^i, \hat{\theta}^i(j)\right)$  par rapport à  $\gamma_x^i$ .
4. Répéter 2 et 3 jusqu'à la convergence.

Le logarithme de vraisemblance de données complètes,  $\mathcal{L}_0$ , à maximiser par EM en termes de responsabilités et des paramètres à estimer sera [Hastie 01] :

$$\begin{aligned} \mathcal{L}_0(d^i \mid \alpha^i, \beta^i, m^i, \gamma_x^i) &= \sum_{x=1}^{N_x} \gamma_x^i \left( -\alpha^i \log \beta^i + (\alpha^i - 1) \log d^i(x) \right. \\ &\quad \left. - \log \Gamma(\alpha^i) - \frac{d^i(x)}{\beta^i} + \log \pi^i \right) \\ &\quad + \sum_{x=1}^{N_x} (1 - \gamma_x^i) \log(1 - \pi^i), \end{aligned} \quad (4.12)$$

Le vecteur de responsabilité calculé dans l'étape E est [Tan 09] :

$$\gamma_x^i = \frac{\pi^i f_{\Gamma}^i(d^i(x) \mid \alpha^i, \beta^i)}{\pi^i f_{\Gamma}^i(d^i(x) \mid \alpha^i, \beta^i) + (1 - \pi^i) (m^i)^{-1}} \quad (4.13)$$

Ensuite, l'estimation de la contribution de la distribution gamma dans le mélange peut s'écrire comme :

$$\hat{\pi}^i = \frac{1}{N_x} \sum_{x=1}^{N_x} \gamma_x^i \quad (4.14)$$

Les paramètres  $\alpha^i$  et  $\beta^i$  sont estimés en maximisant l'équation (4.12) sous les contraintes :

$$\begin{aligned} \frac{\partial \mathcal{L}_0}{\partial \hat{\alpha}^i} &= 0 \\ \frac{\partial \mathcal{L}_0}{\partial \hat{\beta}^i} &= 0 \end{aligned} \quad (4.15)$$

La deuxième égalité nous conduit à l'estimateur de  $\beta^i$  :

$$\frac{\partial \mathcal{L}_0}{\partial \hat{\beta}^i} = \sum_{x=1}^{N_x} \gamma_x^i \left[ \frac{d^i(x)}{\hat{\beta}^{i2}} - \frac{\hat{\alpha}^i}{\hat{\beta}^i} \right] = 0 \quad \Rightarrow \hat{\beta}^i = \frac{\overline{d^i(x) \gamma_x^i}}{\hat{\alpha}^i}, \quad (4.16)$$

où

$$\overline{d^i(x) \gamma_x^i} = \frac{\sum_{x=1}^{N_x} \gamma_x^i d^i(x)}{\sum_{x=1}^{N_x} \gamma_x^i}$$

En introduisant l'estimateur de  $\beta^i$  dans (4.12) nous obtenons :

$$\mathcal{L}_0 = \sum_{x=1}^{N_x} \gamma_x^i \left( -\alpha^i \log \overline{d^i(x)}_{\gamma_x^i} + \alpha^i \log \alpha^i + (\alpha^i - 1) \log d^i(x) \right. \\ \left. - \log \Gamma(\alpha^i) - \frac{\alpha^i d^i(x)}{\overline{d^i(x)}_{\gamma_x^i}} + \log \pi^i \right) + \sum_{x=1}^{N_x} (1 - \gamma_x^i) \log(1 - \pi^i)$$

En appliquant ensuite la première égalité de l'équation (4.15) nous obtenons :

$$\frac{\partial \mathcal{L}_0}{\partial \hat{\alpha}^i} = \sum_{x=1}^{N_x} \gamma_x^i \left[ \log d^i(x) - \psi(\hat{\alpha}^i; 1) - \frac{d^i(x)}{\overline{d^i(x)}_{\gamma_x^i}} \right] + \sum_{x=1}^{N_x} \gamma_x^i \left[ \log \hat{\alpha}^i + 1 - \log \overline{d^i(x)}_{\gamma_x^i} \right] = 0$$

où  $\psi(\alpha^i; 1)$  représente la première dérivée du logarithme de la fonction  $\Gamma(\alpha^i)$ . La simplification de cette relation conduit à l'expression suivante :

$$-\log \hat{\alpha}^i + \psi(\hat{\alpha}^i; 1) = \overline{\log d^i(x)}_{\gamma_x^i} - \log \overline{d^i(x)}_{\gamma_x^i} \quad (4.17)$$

où

$$\overline{\log d^i(x)}_{\gamma_x^i} = \frac{\sum_{x=1}^{N_x} \gamma_x^i \log d^i(x)}{\sum_{x=1}^{N_x} \gamma_x^i}$$

Les paramètres  $\pi^i$  et  $\beta^i$  ont une résolution analytique (voir équations (4.14) et (4.16)) et peuvent être directement estimés en fonction des mesures de dissimilarités et des responsabilités. Cependant, le paramètre  $\alpha^i$  n'a pas une solution analytique cf. équation (4.17). Pour l'estimer, la méthode de Newton permet une résolution rapide de cette équation. Les équations (4.13)-(4.17) permettent donc de retrouver les responsabilités,  $\gamma_x^i$ , et les paramètres  $(\pi^i, \beta^i, \alpha^i)$ .

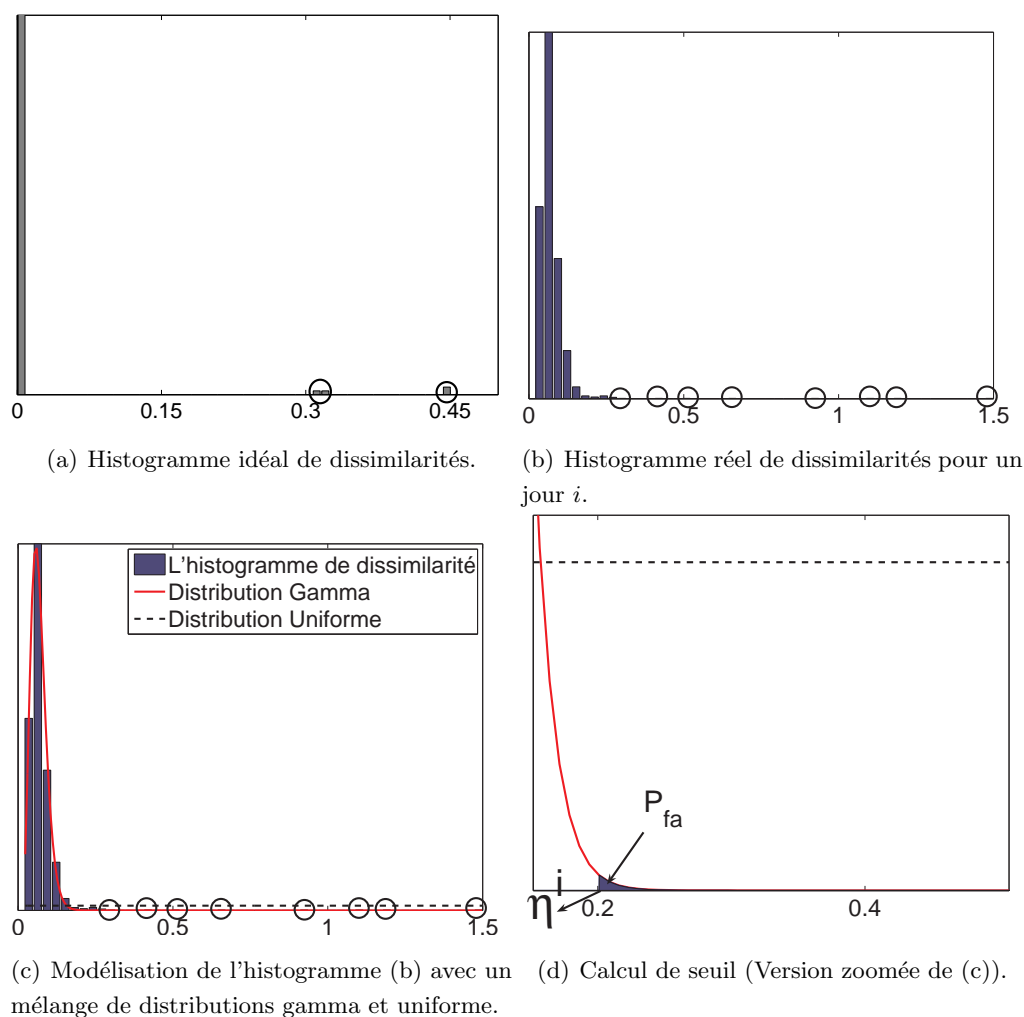
L'histogramme de dissimilarités pour un jour  $i$  pourra alors être modélisé par un mélange des distributions gamma et uniforme dont les paramètres sont estimés avec la procédure ci-dessus. Pour l'histogramme affiché à la page 146 sur la figure 4.2(b), cette approximation est montrée sur la figure 4.2(c). Une fois les mesures de dissimilarité modélisées par ce mélange, la prochaine étape est d'établir un seuil afin de réaliser une extraction optimale de zones de singularité.

### 4.3.1 Seuillage en fixant une probabilité de fausse alarme

A une distance donnée, la mesure de dissimilarité peut manifester des variations sur différents jours à cause de plusieurs facteurs comme :

- la durée et l'intensité du rayonnement solaire pour un jour donné,
- la température de l'air,
- la vitesse du vent, etc.

En plus, la résolution de l'appareil de mesure peut également influencer ces variations. Comme le vecteur de référence de variations journalières n'est pas identique pour chaque jour, le choix d'un seuil constant peut conduire à des résultats erronés. Une approche classique pour les applications



**Fig. 4.2:** Seuillage des mesures de dissimilarité,  $d^i(x)$ , en utilisant un taux de fausses alarmes constant. Dans le cas idéal, (a) la dissimilarité doit être nulle pour la plupart des zones sauf les zones singulières, mais en pratique (b) ce n'est pas le cas. L'histogramme résultant pour un jour donné peut être modélisé (c) par un mélange de distributions gamma et uniforme, dont les paramètres sont estimés avec l'algorithme EM. Une probabilité de fausse alarme  $P_{fa}$  permet l'estimation d'un seuil unique pour chaque jour,  $\eta^i$ , prenant alors en compte les effets journaliers. La notation "O" met en évidence les valeurs de dissimilarité pour les zones singulières.

à seuillage variable est basée sur le taux de fausses alarmes constant (CFAR pour Constant False Alarm Rate en anglais) où la probabilité de fausse alarme ( $P_{fa}$ ) est fixée au lieu d'avoir un seuil fixe.

Dans le cadre de la détection, nous avons un cas binaire qui correspond à l'existence ou à l'absence de la singularité. La décision est à prendre sur la mesure de dissimilarité à une distance donnée en l'attribuant à une des deux hypothèses suivantes :

- $H_0$  : correspond à l'absence de la singularité. Dans ce cas, la mesure de dissimilarité quantifie les erreurs d'estimation et/ou le bruit. La dissimilarité sera alors distribuée selon la loi gamma comme expliqué précédemment.
- $H_1$  : correspond à la présence d'une singularité. Dans ce cas, la mesure de dissimilarité compte pour les différentes singularités. Une distribution uniforme modélise ces singularités comme décrit ci-dessus.

Une fausse alarme sera produite quand le détecteur identifie une singularité alors qu'il n'y avait pas une. Autrement dit, la décision prise par le détecteur va attribuer l'hypothèse  $H_1$  alors que en vérité c'était l'hypothèse  $H_0$ . La probabilité de fausse alarme peut alors s'exprimer comme :

$$P_{fa} = \int_{\eta^i}^{\infty} \frac{z^{(\alpha^i-1)} e^{-z/\beta^i}}{\beta^i \alpha^i \Gamma(\alpha^i)} dz \quad (4.18)$$

où  $\eta^i$  est la marge de décision entre  $H_0$  et  $H_1$ , c'est-à-dire le seuil pour le jour  $i$ . La procédure de sélection de ce seuil est montrée sur la figure 4.2(d).

Pour une probabilité de fausse alarme fixe, nous pouvons facilement retrouver le seuil  $\eta^i$  à partir de l'équation (4.18). Le choix de la probabilité de fausse alarme dépend du système d'exploitation et du débit de fuites à mettre en évidence et peut s'établir pour chaque site d'acquisition en réalisant des fuites contrôlées sous différentes conditions. La valeur de la probabilité de fausse alarme n'a pas encore été quantifiée sur les deux sites d'acquisition (Oraison et Kembs). Pour cette raison nous utiliserons par la suite des valeurs de  $P_{fa}$  choisies arbitrairement.

Le détecteur proposé identifie donc les singularités en appliquant le seuil estimé par l'équation (4.18) avec la sortie finale :

$$d_{th}^i(x) = \begin{cases} d^i(x) & \text{si } d^i(x) > \eta^i \\ 0 & \text{sinon} \end{cases} \quad (4.19)$$

Si la mesure de dissimilarité à une distance donnée  $x$  est donc supérieure au seuil  $\eta^i$  calculé pour une probabilité de fausse alarme imposée  $P_{fa}$ , alors à cette distance nous considérons qu'il y a une singularité et nous gardons comme valeur la mesure de la dissimilarité à cette distance. Ceci nous permettra ensuite de quantifier les singularités obtenues à plusieurs distances et sur plusieurs jours  $i$ .

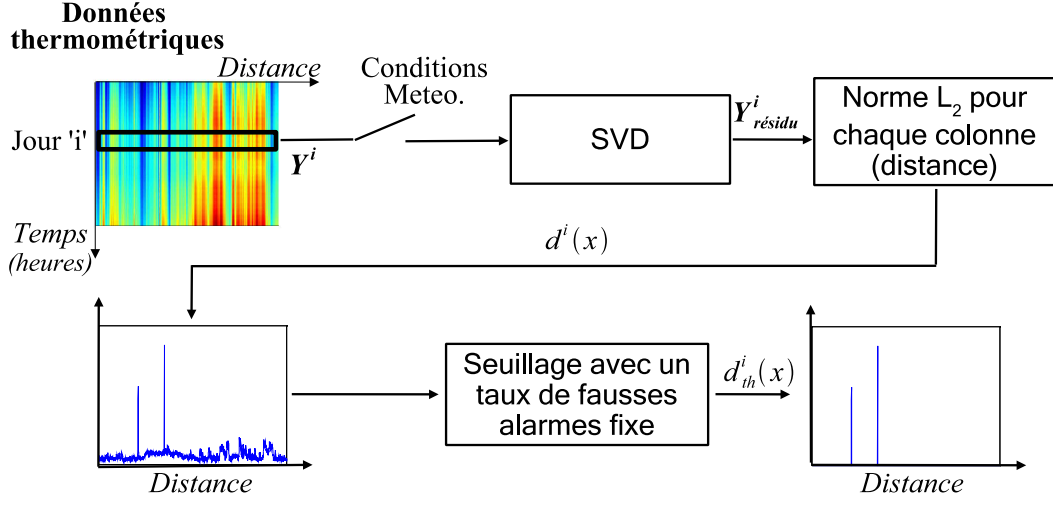
## 4.4 Système proposé pour la détection de singularités

Les différentes phases de la détection de singularités sont résumées sur le schéma de la figure 4.3. Les données thermométriques acquises sur le site sont présentées en fonction de la distance et du temps. Le système proposé se base sur les variations des températures au cours d’une journée, avec les singularités montrant des variations différentes par rapport aux zones non-singulières. Le système traite alors les données pour un jour “ $i$ ” sur tous les capteurs [Khan 09b].

1. Premièrement, les conditions météorologiques sont vérifiées pour ne pas avoir des jours avec des précipitations ou d’autres phénomènes énergétiques qui affectent plusieurs capteurs en même temps. Ceci est fait parce qu’il existe des zones non-singulières qui réagissent de façon différente face à un tel type de phénomène. Le critère basé sur le kurtosis proposé dans le chapitre 2 (section 2.6.3) peut être utilisé pour identifier ces zones [Khan 08d].
2. Un vecteur de référence, qui représente les zones non-singulières, est ensuite estimé par la première source de la SVD. Cette décomposition est appliquée sur les données acquises pendant une journée. Le sous-espace résidu construit en enlevant l’influence de ce vecteur de référence mesure les déviations par rapport au vecteur de référence et contient alors les informations liées aux singularités.
3. Une mesure de dissimilarité qui permettra de localiser les singularités en distance est calculée en prenant la norme  $L_2$  de chaque colonne du sous-espace résidu. La mesure de dissimilarité ainsi obtenue est une fonction de la distance.
4. Pour prendre en compte les erreurs d’estimation du vecteur de référence ainsi que le bruit de mesure, une approche basée sur le taux de fausse alarme constant est proposée. Ceci correspond à modéliser l’histogramme de dissimilarités pour un jour  $i$  par un mélange de distributions gamma et uniforme, dont les paramètres sont estimés par maximisation de vraisemblance.
5. Pour ne pas imposer un seuil fixe pour chaque jour, le seuillage de l’histogramme de dissimilarités est effectué en fixant une probabilité de fausse alarme.
6. En appliquant la procédure décrite ci-dessus pour différents jours, nous obtenons une projection des singularités dans l’espace 2-D spatio-temporel  $(x, i)$ , avec une résolution spatiale correspondant à la résolution spatiale de l’appareil de mesure DTS et une résolution temporelle de 24 heures.

Les résultats obtenus sur les données réelles et simulées seront présentés par la suite comme des images 2-D pour toute la durée d’analyse avec un point en distance correspondant à un capteur et un point en temps correspondant à un jour. Comme le système utilise des données journalières, l’idéal sera d’avoir un résultat pour chaque jour. Ceci sera utilisé pour les fuites artificielles de type percolation et pour les données simulées. Ces deux affichages permettront non-seulement de localiser les singularités en distance mais également de suivre leur évolution d’un jour à l’autre.

Dans le cadre de détection des anomalies, le but le plus important est d’arriver à les localiser en



**Fig. 4.3:** Système de détection de singularités. Le sous-espace résidu de la SVD obtenu sur les données acquises pendant une journée contient des informations liées aux singularités. La norme  $L_2$  des colonnes de ce sous-espace donne une mesure de dissimilarités comme une fonction de la distance. L’histogramme de dissimilarités peut être modélisé comme un mélange des distributions gamma et uniforme. Un seuillage est ensuite appliqué en fixant une probabilité de fausse alarme.

distance. Nous pouvons considérer deux cas ici :

### Cas 1 : étude en continue

Ce cas correspond à l’analyse continue pour plusieurs jours consécutifs. Comme évoqué à la section 3.5.3, afin de rendre le résultat facilement interprétable, nous pouvons faire une projection sur l’axe de distances de la mesure de dissimilarité finale. Cette projection prend la forme suivante :

$$d_{proj}(x) = \frac{1}{N_p} \sum_{i=1}^{N_p} [d_{th}^i(x)]^2 \quad (4.20)$$

où  $N_p$  représente le nombre de jours (continus ou non) pour lesquels les données sont analysées. Nous pouvons par exemple utiliser  $N_p = 30$ , correspondant à une analyse sur 1 mois, mais d’autres durées d’analyse seront également étudiées dans la section application.

### Cas 2 : étude périodique (ou discontinue)

Dans ce cas, l’intérêt est d’analyser un jour donné choisi arbitrairement sans avoir analysé les jours voisins. La projection ne sera pas alors requise pour un tel cas et le détecteur identifiera les singularités pour le jour donné.

Dans la section suivante, l’application de système proposé pour différents scénarios pour les données simulées et les données réelles sera présentée afin de montrer l’efficacité de ce système.



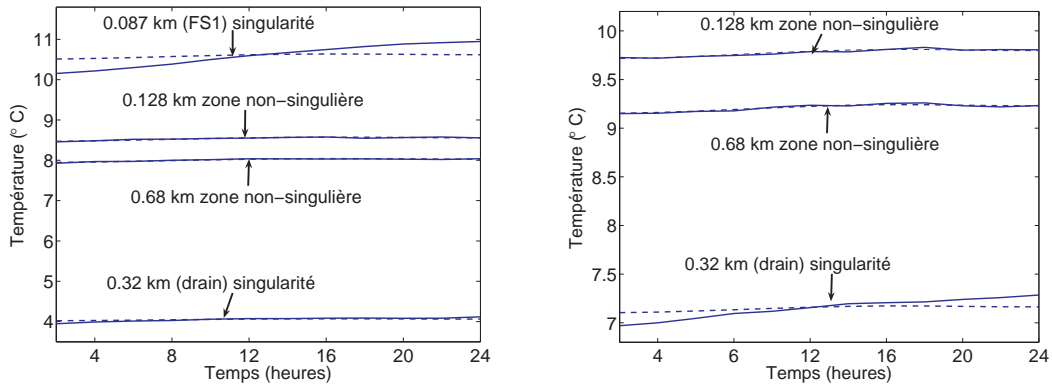
## 4.5 Application aux données simulées

Dans un premier temps, la méthode proposée est testée sur des données simulées obtenues en utilisant le simulateur décrit dans la section 2.2. Comme dans le chapitre précédent, une digue de 1 km de longueur avec deux couches homogènes et non-singulières est simulée. Les singularités sont ensuite introduites sous la forme d'un drain et de deux fuites. Un bruit gaussien blanc spatialement et temporellement avec une moyenne nulle et un écart type de 0.02 est ajouté aux données pour simuler des mesures effectuées avec un appareil DTS Sensornet. Pour plus de détails sur la construction de la digue le lecteur pourra se référer à la section 3.5.

Les deux fuites de longueur de 2 m chacune sont simulées avec :

- la fuite FS1 située aux distances 87 – 88 m existe pendant 3 jours entre le 11 et le 13 février sur la 1<sup>e</sup> couche homogène,
- la fuite FS2 située aux distances 727 – 728 m existe pendant 3 jours entre le 23 et le 25 mars sur la 2<sup>e</sup> couche homogène.

La figure 4.4(a) montre les profils de température simulés le 12 février à 4 distances différentes (lignes continues) ainsi que le sous-espace signal obtenu à ces distances (lignes pointillées) après estimation d'un vecteur de référence pour ce jour. Deux de ces distances correspondent aux zones singulières, la fuite FS1 (0.087 km) et le drain (0.32 km). Les deux autres distances correspondent



(a) Profils de température et le sous-espace signal estimés le 12 février.

(b) Profils de température et le sous-espace signal estimés le 23 février.

**Fig. 4.4:** *Lignes continues* : Profils journaliers de température à 4 distances et à deux dates différentes. Les deux zones singulières, la fuite FS1 (0.087 km) et le drain (0.32 km), présentent des variations journalières différentes par rapport aux zones non-singulières (0.128 km et 0.68 km). *Lignes pointillées* : Sous-espace signal  $\mathbf{Y}_{sig}^i$  pour les mêmes jours et aux mêmes distances, montrant que le vecteur de référence estimé présente des déviations plus importantes pour les zones singulières que pour les zones non-singulières.

aux zones non-singulières, une (0.128 km) dans la première couche homogène construite avec du limon de perméabilité  $10^{-7}$  et l'autre (0.68 km) dans la deuxième couche homogène construite avec du limon de perméabilité  $10^{-5}$ . La fuite FS1 montre une variation journalière assez différente par rapport aux zones non-singulières. Ceci est mis en évidence par le vecteur de référence

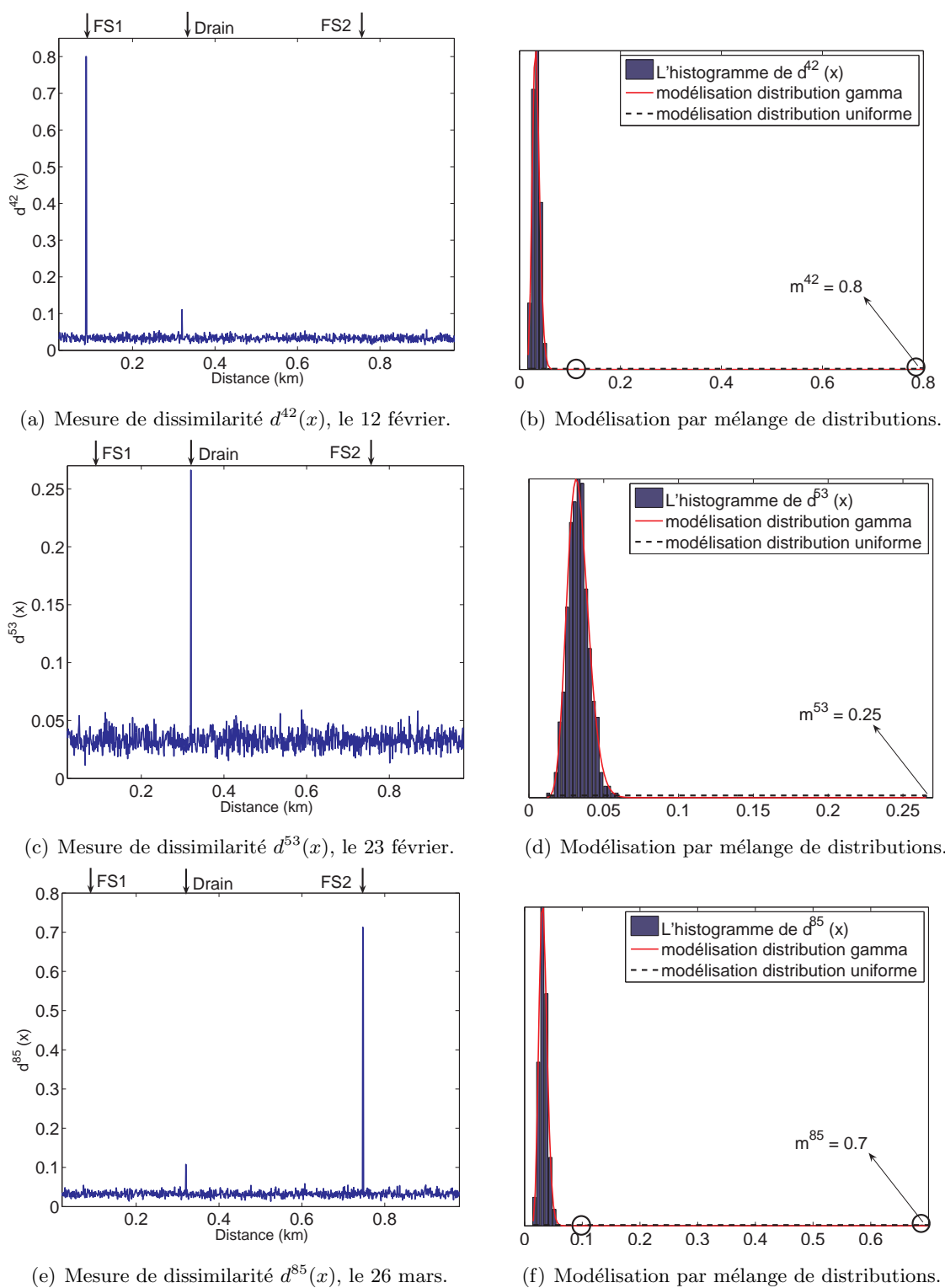
qui est représentatif de zones non-singulières et qui présente des déviations importantes par rapport au profil de température acquis pour la FS1. Au contraire, le vecteur de référence estime très bien les variations journalières pour les zones non-singulières. Il faut remarquer ici que pour le drain, même si la déviation est relativement faible, elle est visible au début et à la fin de la journée. Ceci est dû au fait que le drain, même s'il existe toujours, montre de faibles variations au cours du temps.

Nous considérons ensuite les profils journaliers simulés le 23 février, profils affichés sur la figure 4.4(b). Nous pouvons constater que la variation journalière du drain est significativement différente par rapport à celles des zones non-singulières. Ceci est mis en évidence par une déviation plus importante entre le profil et le sous-espace signal correspondant construit avec le vecteur de référence estimé pour ce jour.

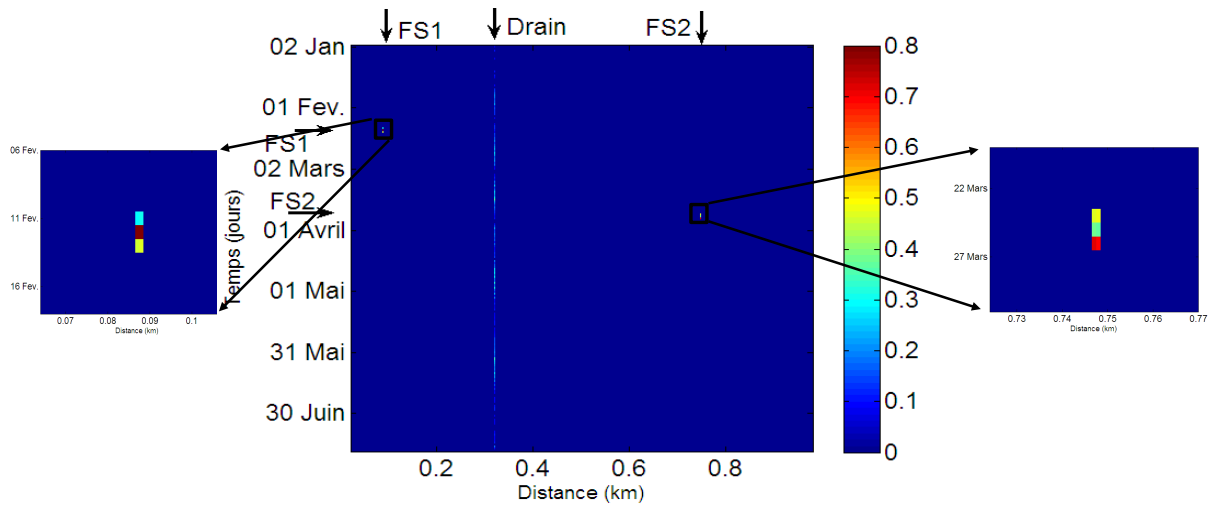
La mesure de dissimilarité  $d^i(x)$  peut ensuite être obtenue en utilisant l'équation (4.5). Les figures 4.5(a), 4.5(c) et 4.5(e) montrent ces dissimilarités pour 3 dates : le 12 février (occurrence de la fuite FS1), le 23 février (sans occurrence de fuite) et le 26 mars (occurrence de la fuite FS2). Tenant compte du fait que nous analysons ici des données simulées, les vecteurs de référence estiment très bien les zones non-singulières avec une erreur d'estimation très faible. C'est alors le bruit rajouté, bruit qui simule l'appareil de mesure, qui conduit à des mesures de dissimilarité non nulles pour les zones non-singulières. Pour ne détecter que les singularités, chaque histogramme de dissimilarité est modélisé avec un mélange de distributions gamma et uniforme. Les figures 4.5(b), 4.5(d) et 4.5(f) montrent respectivement ces histogrammes pour les trois dates (12 février, 23 février et 26 mars) ainsi que leurs ajustements par mélange de distributions. La distribution gamma permet de modéliser les zones non-singulières tandis que la distribution uniforme permet de modéliser les zones singulières. Un seuil est ensuite appliqué pour chacun de ces jours en fixant une probabilité de fausse alarme,  $P_{fa}$ . Une fausse alarme correspond à la détection d'une singularité quand en vérité ce n'est qu'une non-singularité. Pour analyser ces données, cette probabilité de fausse alarme est fixée à  $10^{-5}$ .

Un seuil variable pour chaque jour est ensuite estimé à l'aide de l'équation (4.18), ce qui permet d'effectuer le seuillage sur  $d^i(x)$  afin d'obtenir  $d_{th}^i(x)$  en utilisant l'équation (4.19). Le fait de ne pas utiliser un seuil fixe implique que le détecteur s'adapte aux conditions spécifiques du jour analysé. La figure 4.6 montre le résultat final après seuillage pour toute la période d'analyse considérée ici (du 1 janvier au 18 juillet). Le détecteur permet de localiser les singularités avec une résolution spatiale d'un mètre et une résolution temporelle d'un jour. Le résultat de la détection est très net dans le sens que seulement les trois zones singulières, c'est-à-dire les deux fuites FS1 et FS2 et le drain, sont détectées. Les zooms au voisinage des fuites FS1 et FS2 indiquent que celles-ci sont parfaitement localisées en temps et en distance.

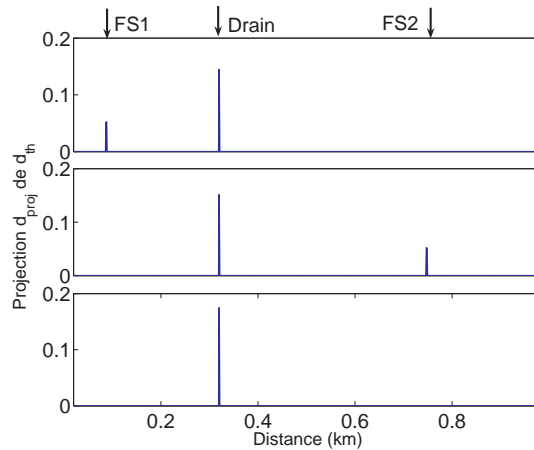
Pour faciliter l'interprétation des résultats, comme expliqué dans la section 4.4, la figure 4.7 présente la projection sur l'axe de distance en utilisant l'équation (4.20) avec  $N_p = 30$ , pour trois périodes : (de haut en bas) du 01 février au 01 mars, du 02 mars au 31 mars et du 21 mai au 19 juin. La première période contient la fuite FS1 et le drain. La projection  $d_{proj}(x)$  effec-



**Fig. 4.5:** Mesures de dissimilarités et leur modélisation par mélange de distributions gamma et uniforme pour trois dates : (de haut en bas) le 12 février à l’occurrence de  $FS1$  ; le 23 février sans occurrence de fuites ; le 26 mars à l’occurrence de  $FS2$ . Les flèches indiquent les positions des singularités. La notation “O” met en évidence les valeurs de dissimilarité pour les zones singulières.



**Fig. 4.6:** Résultat de la détection après seuillage de la mesure de dissimilarité pour les données simulées permet la détection des singularités. Les zooms au voisinage des fuites mettent en évidence la parfaite localisation des fuites  $FS1$  et  $FS2$ .



**Fig. 4.7:** Projections  $d_{proj}(x)$  sur l'axe de distances avec  $N_p = 30$  pour 3 différentes périodes. De haut en bas : du 01 février au 01 mars, période contenant la fuite  $FS1$  et le drain ; du 02 mars au 31 mars, période contenant la fuite  $FS2$  et le drain ; du 21 mai au 19 juin, période contenant que le drain. Les fuites  $FS1$  et  $FS2$  et le drain sont les seules singularités détectées.

tuée sur cette période met en évidence ces deux singularités. Il faut cependant faire attention aux amplitudes qui représentent une moyenne sur les  $N_p = 30$  jours. Pour cette raison, la fuite  $FS1$  qui existe pendant trois jours apparaît comme ayant une contribution plus petite que celle donnée par le drain qui existe pour toute la durée d'analyse. De même manière, la deuxième projection effectuée sur la période contenant la fuite  $FS2$  permet la détection de celle-ci. La troisième projection met en évidence que le drain, la seule singularité pour cette période d'analyse.

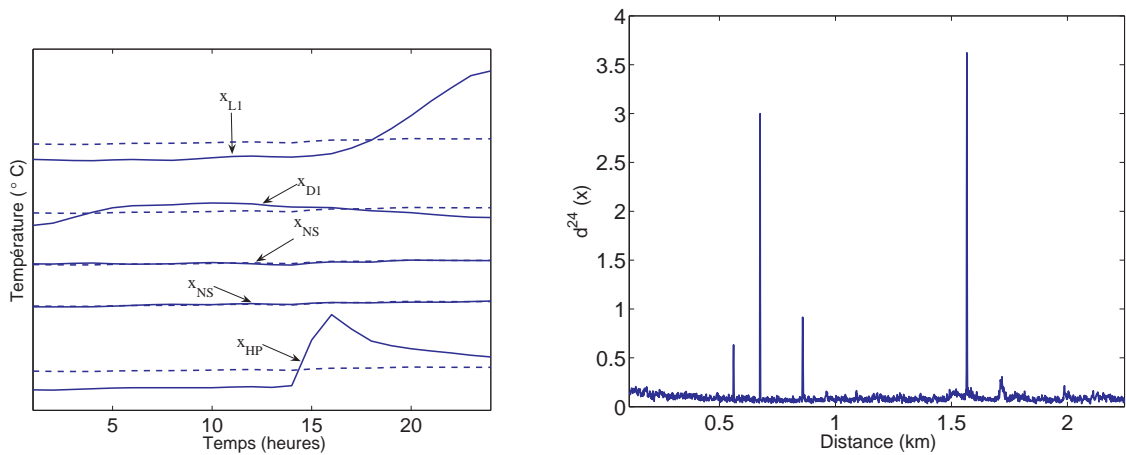
## 4.6 Application aux données enregistrées sur le site d’Oraison

Cette section est consacrée à la validation de la méthode proposée sur les données réelles acquises sur le site d’Oraison. Les données ont été enregistrées aux différentes périodes entre 2005 et 2007 par un appareil DTS Sensornet. Les deux cas particuliers d’intérêt sur ce site sont ceux de fuites artificielles et de fuites réelles. Les fuites artificielles sont des fuites de type percolation présentées dans la section 2.3.3. Trois fuites,  $L1$ ,  $L2$  et  $L3$ , de type percolation ont été introduites entre le 10 et le 12 mai 2005 avec des débits et des localisations différents, spécifiés dans le tableau 3.1 (page 112). En plus, un point chaud a été également introduit à la distance de 0.674 km la matinée du 10 mai.

Les profils de température pour 5 différentes distances enregistrés le 10 mai sont présentés sur la figure 4.8(a). Ils comprennent trois zones singulières,  $x_{L1}$  pour la fuite  $L1$ ;  $x_{HP}$  pour le point chaud et  $x_{D1}$  pour le drain  $D1$ . En plus, deux distances non-singulières choisies arbitrairement et notées  $x_{NS}$  sont affichées également sur cette figure. La différence entre les profils journaliers des zones singulières et non-singulières est évidente, surtout pour la fuite  $L1$  et le point chaud  $HP$ . Le sous-espace signal  $Y_{sig}$  construit avec le vecteur de référence estimé pour ce jour d’analyse est affiché en pointillé sur la même image pour les mêmes distances. Les déviations entre les profils enregistrés et les sous-espaces signaux correspondants permettent de quantifier la dissimilarité : une petite déviation pour les zones non-singulières va donner une mesure de dissimilarité plus petite par rapport aux zones singulières qui présentent des déviations plus importantes. La mesure de dissimilarité  $d^{24}(x)$ , où 24 représente le 24<sup>e</sup> jour d’analyse, soit le 10 mai, est montrée sur la figure 4.8(b). Nous pouvons constater à partir de cette image que les zones singulières, notamment la fuite, le point chaud et les drains, présentent des valeurs de dissimilarité très élevées. Également, autour de la distance 1. km, nous observons des valeurs élevées de dissimilarité qui sont dues à la singularité du terrain, comme expliqué dans la section 2.3.2. La prochaine étape consiste à effectuer un seuillage sur cette dissimilarité afin de ne pouvoir identifier que la fuite, le point chaud et le drain, mais sans mettre en évidence la singularité du terrain qui n’est pas intéressante dans notre analyse. L’histogramme de cette dissimilarité est affichée dans la figure 4.8(c). Cet histogramme est ensuite modélisé par un mélange de distributions gamma et uniforme. Les singularités sont marquées sur l’histogramme par des cercles. Les différents paramètres du mélange sont estimés par maximisation de vraisemblance en utilisant l’algorithme EM. La contribution de la distribution gamma dans le mélange,  $\pi^i$ , est très importante, marquant le fait qu’il existe plutôt des zones non-singulières. La valeur de la probabilité de fausse alarme,  $P_{fa}$ , a été fixé à  $10^{-5}$ , comme dans le cas de signaux simulés. Il faut rappeler ici que cette valeur est à choisir en fonction du site d’expérimentation, du débit et du type de fuites à mettre en évidence, etc. Le seuil  $\eta^i$ , avec  $i = 24$ , correspondant à cette probabilité est ensuite appliqué sur les mesures de dissimilarité afin de retrouver  $d_{th}^{24}(x)$ . Cette mesure, représentant les zones singulières, est affichée sur la figure 4.8(d). Les différentes singularités détectées sont : les deux drains  $D1$  et  $D2$ , la fuite  $L1$ , le point chaud  $HP$ , ainsi que quelques distances correspondant aux singularités du terrain autour de 1.7 km. Il faut mentionner ici que l’apparition de la singularité

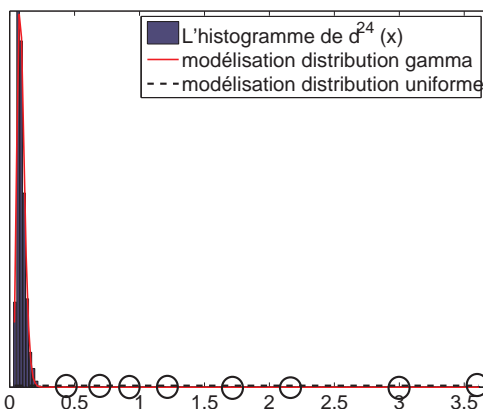
du terrain autour de 1.7 km peut être évitée en choisissant une autre valeur de probabilité de fausse alarme, par exemple  $P_{fa} = 10^{-6}$ . Dans ce cas, seulement la fuite, le point chaud et les drains ressortiront.

Le résultat final de la détection appliqué à toutes les distances et tous les acquisitions effec-

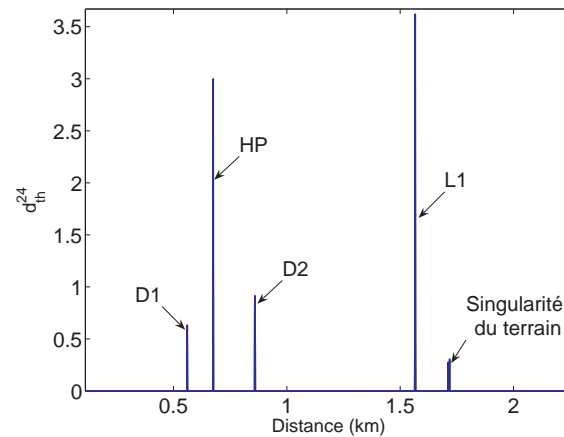


(a) Profils journaliers de température ainsi que le sous-espace signal construit à partir du vecteur de référence pour 4 distances différentes.

(b) Mesure de dissimilarité  $d^{24}(x)$ ; jour 24, le 10 mai.



(c) Modélisation du histogramme de dissimilarité par mélange de distributions gamma et uniforme.



(d) Résultat de la détection finale,  $d_{th}^{24}(x)$ , pour le 10 mai.

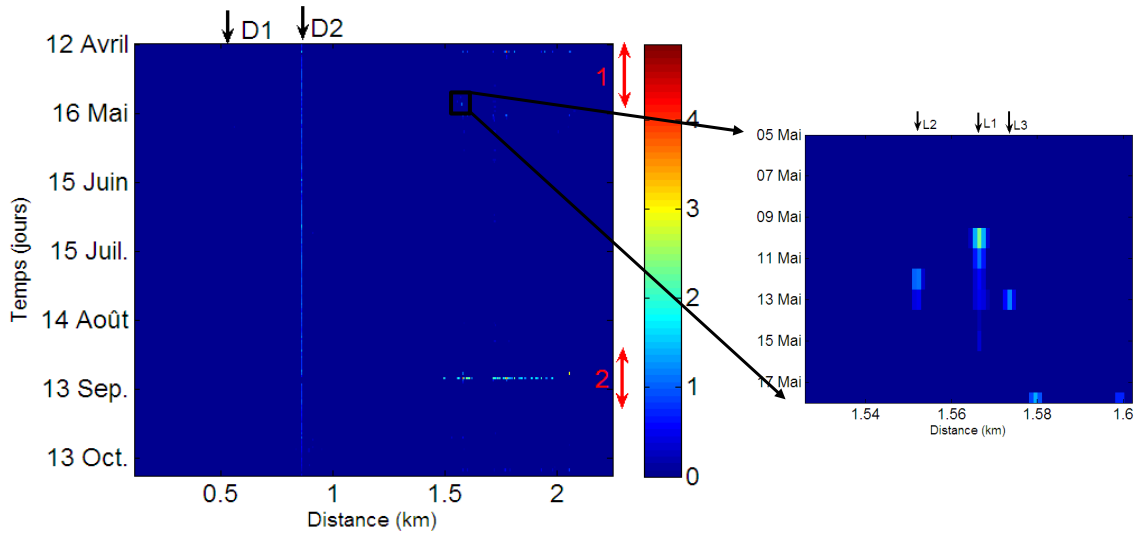
**Fig. 4.8:** Différentes étapes du système proposé pour les données acquises le 10 mai, à l'occurrence de la fuite  $L1$ , sur le site d'Oraison. Les zones singulières sont très bien séparées des zones non-singulières.

tuées entre le 12 avril et le 13 octobre 2005 (données brutes affichées sur la figure 3.15(a) page 113) est affiché sur la figure 4.9. Cette image permet de localiser les singularités dans un espace spatio-temporel :

- la localisation en distance avec une résolution d'un mètre,
- la localisation en temps avec une résolution d'un jour.

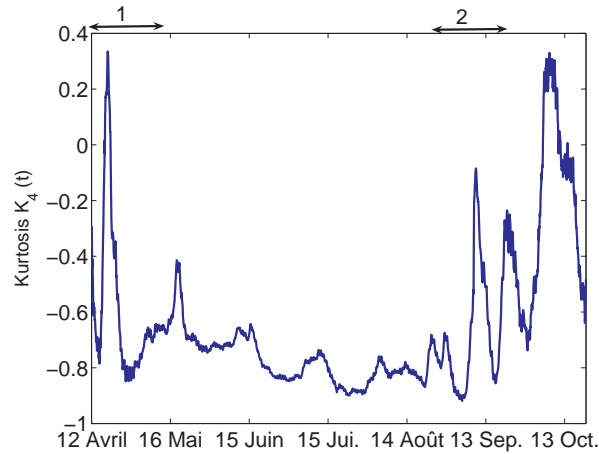
Un zoom dans la proximité des fuites permet de constater que les fuites sont détectées avec une très bonne localisation en temps et en distance. Cependant les plus fortes amplitudes dans cette image (codées par le rouge) ne sont pas au niveau des fuites. Ceci est dû au fait que pendant quelques jours, plusieurs zones d'anomalies sont mises en évidence dans la deuxième moitié de la fibre. Pour analyser ces zones plus en détail, nous allons nous intéresser à la projection  $d_{proj}(x)$  de cette image sur l'axe de distance. Mais avant cela, nous sommes intéressés à identifier ces jours en considérant le kurtosis calculé sur les données brutes. Les valeurs obtenues pour cette période d'analyse sont montrées à nouveau sur la figure 4.10. Le kurtosis, comme expliqué dans les chapitres précédents, peut être utile pour identifier les phénomènes éphémères énergétiques en temps comme les précipitations, etc. Sur la courbe de kurtosis, ceci est mis en évidence par un changement de kurtosis pour les périodes de précipitations où d'autres tels phénomènes par rapport aux périodes normales.

Dans un premier temps, nous considérons la projection  $d_{proj}(x)$  pour la période 12 avril - 15

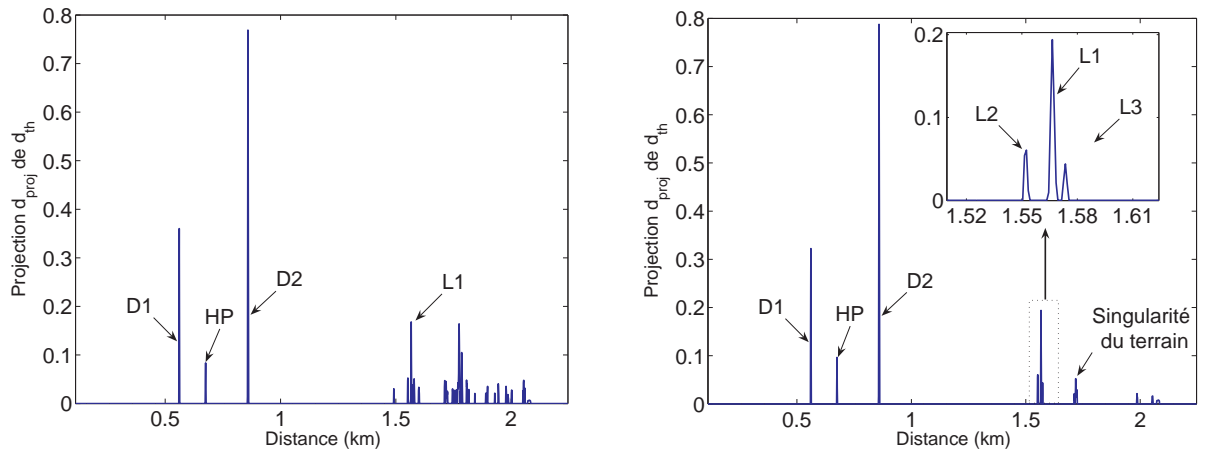


**Fig. 4.9:** Résultat de détection pour les données d’Oraison acquises entre le 12 avril et le 13 octobre 2005 après seuillage de la mesure de dissimilarité. L’algorithme proposé permet de mettre en évidence les singularités. La version zoomée montre que la détection de fuites simulées est de très bonne qualité. En outre, il existe plusieurs phénomènes (se comportant comme des singularités) sur quelques jours. Pendant ces jours, les mesures sur la plupart des capteurs sont perturbées, ce qui indique l’existence d’un phénomène de type précipitation risquant de générer des fausses alarmes.

mai 2005 (marquée par la flèche 1 sur l’image de détection, figure 4.9). Ceci correspond à une période de projection  $N_p = 30$  jours (sachant qu’il y a des données manquantes entre le 25 et le 27 avril). Cette projection, affichée sur la figure 4.11(a), permet d’extraire les fuites, les drains et le point chaud comme des singularités mais également plusieurs zones après 1.5 km. Une analyse de la courbe de kurtosis pour cette période (voir figure 4.10, période indiquée par la flèche 1) permet de constater la présence d’un phénomène anormal en temps. Les données météorolo-



**Fig. 4.10:** Kurtosis calculé sur les données d'Oraison acquises entre le 12 avril et le 13 octobre 2005 mettant en évidence des phénomènes éphémères énergétiques de type précipitation. Les flèches indiquent les zones analysées.



(a) Projection en considérant les jours de précipitations.  $N_p = 30$ .

(b) Projection en écartant les jours de précipitations.  $N_p = 26$ .

**Fig. 4.11:** Projections  $d_{proj}(x)$  sur l'axe de distance du résultat de détection affiché sur l'image 4.9 pour les données acquises entre le 12 avril et le 15 mai. Une étude de la courbe du kurtosis pour cette période (voir période 1 sur la figure 4.10) indique l'occurrence des précipitations pendant quelques jours. Sans prendre en compte ces jours de précipitation, la nouvelle projection obtenue en (b) met clairement en évidence les deux drains  $D1$  et  $D2$ , le point chaud  $HP$  ainsi que les 3 fuites  $L1$ ,  $L2$  et  $L3$  (voir le zoom). En plus, la singularité du terrain est aussi mise en évidence autour de 1.7 km.

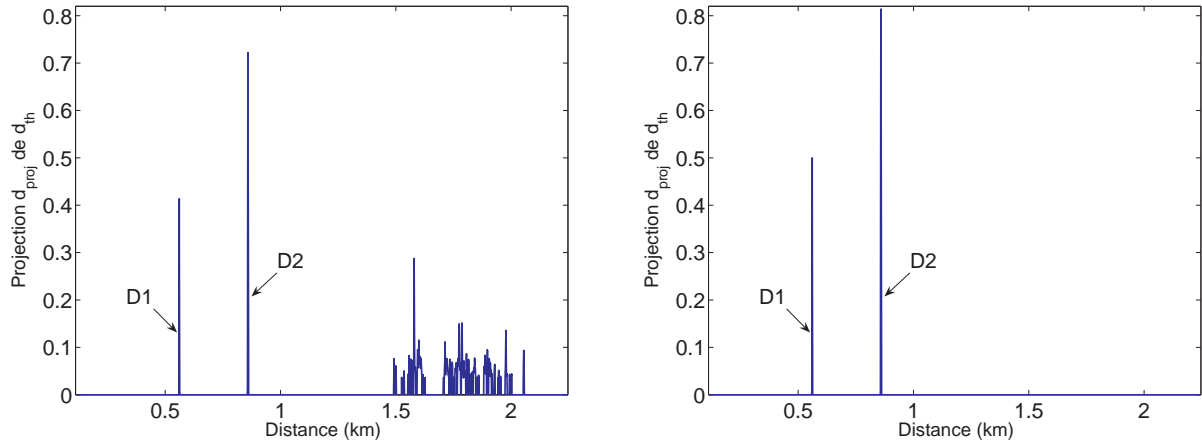
giques confirment la présence des précipitations intermittentes pendant quelques jours en début de cette période. Une nouvelle projection sans prendre en compte les résultats pour ces jours est montrée sur la figure 4.11(b). Cette projection permet de mettre en évidence les trois fuites, les drains, le point chaud et la singularité du terrain autour de 1.7 km. En plus, des zones autour



de 1.98 km, 2.052 – 2.055 km et 2.072 – 2.083 km sont également détectées. Ces zones n’ont pas été identifiées comme des zones particulières pendant les inspections visuelles effectuées sur le site. Il faut mentionner que cette projection est obtenue en utilisant une probabilité de fausse alarme de  $10^{-5}$ . En modifiant la probabilité de fausse alarme, nous pouvons écarter ces zones tout en gardant les points d’intérêt.

De même façon, la projection effectuée pour la période 25 août - 24 septembre (marqué par 2 sur l’image de détection et la courbe de kurtosis) met à nouveau en évidence plusieurs zones de singularité. Une étude de la courbe de kurtosis permet d’identifier les jours de précipitation. En évitant ces jours lors du calcul de la projection, la nouvelle projection affichée sur la figure 4.12(b) met en évidence uniquement les deux drains comme étant des singularités.

Ce type d’analyse permet également la détection des fuites réelles. Pour l’année 2005, la projec-



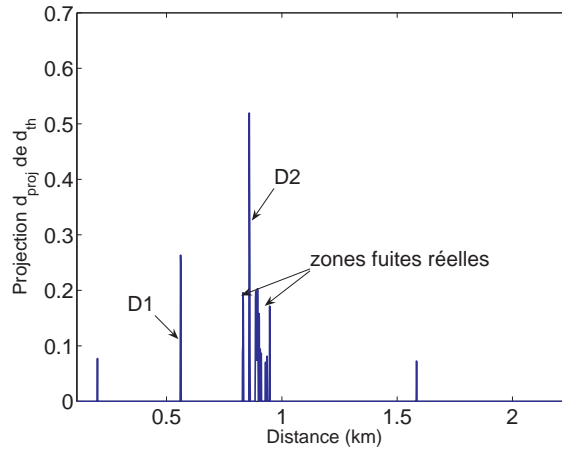
(a) Projection en considérant les jours de précipitations.  $N_p = 30$ .

(b) Projection en écartant les jours de précipitations.  $N_p = 23$ .

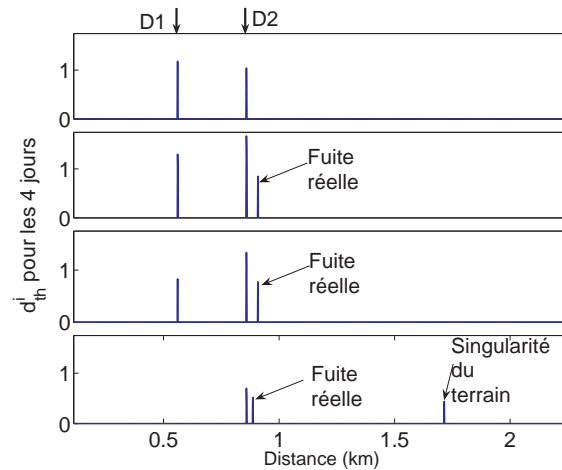
**Fig. 4.12:** Projections  $d_{proj}(x)$  sur l’axe de distance du résultat de détection affiché sur l’image 4.9 pour les données acquises entre le 25 août et le 24 septembre. Une étude de la courbe du kurtosis pour cette période (voir période 2 sur la figure 4.10) indique l’occurrence des précipitations pendant quelques jours. Ceux-ci perturbent le résultat en introduisant de fausses alarmes. En évitant ces jours, la nouvelle projection obtenue en (b) met en évidence que les deux drains  $D1$  et  $D2$ .

tion  $d_{proj}(x)$  affichée sur la figure 4.13, projection calculée pour une période d’analyse  $N_p = 5$  autour de 10 octobre, met en évidence des zones singulières autour du deuxième drain  $D2$ . A la suite de cette opération, une investigation visuelle sur le site a confirmé la présence des fuites réelles dans cette zone (fuites qui ne se manifestent pas d’une façon continue avec le temps).

Comme nous disposons de plusieurs acquisitions discontinues, nous analysons les données acquises pendant quelques jours en 2007. La figure 4.14 présente les résultats de la détection,  $d_{th}(x)$ , pour les dates (de haut en bas) 20-22 octobre et 04 novembre 2007. Les fuites réelles sont repérées après le deuxième drain  $D2$ . Ce type d’analyse, appelé étude périodique (ou discontinue) dans la section 4.4 apporte un grand confort d’utilisation car il permet à l’utilisateur



**Fig. 4.13:** Détection des fuites réelles sur la projection  $d_{proj}^i(x)$  calculée pour une période d'analyse autour de 10 octobre 2005.



**Fig. 4.14:** Fuites réelles identifiées en 2007 sur les résultats de la détection,  $d_{th}^i(x)$ , pour les dates (de haut en bas) 20-22 octobre et 04 novembre 2007.

d'analyser la digue le jour voulu, sans avoir besoin d'une référence ou d'une étude continue sur plusieurs jours.

## 4.7 Discussion

Nous avons présenté une approche pour la détection des zones singulières à partir d'analyse des données thermométriques journalières. Les zones singulières comprennent les anomalies en distance de type fuites, drains, singularités du terrain, etc. Nous avons proposé un système qui exploite les variations de température au cours d'une journée. Les variations journalières pour les zones singulières sont différentes par rapport aux zones non-singulières. Une hypothèse d'avoir plus de zones non-singulières que des zones singulières a été faite, ce qui est une hypothèse assez juste pour la plupart de sites. Cependant, le site de Kembs, de par sa spécificité et par

l'installation de la fibre optique d'acquisition au voisinage de l'eau du contre-canal, parfois immergée, ne répond pas à cette hypothèse.

A partir de toutes les variations journalières enregistrées à toutes les distances, un vecteur de référence qui représente les zones non-singulières est estimé. L'idée est de calculer ensuite une dissimilarité entre ce vecteur de référence et les variations pour chaque distance. La SVD a été utilisée pour estimer le vecteur de référence comme étant le premier vecteur singulier estimé. Les déviations entre ce vecteur de référence et les profils de température à différentes distances sont mises en évidence par le sous-espace résidu de la SVD, avec le sous-espace signal construit avec le vecteur de référence, donc avec la 1<sup>e</sup> valeur singulière. D'autres approches peuvent également être utilisées pour estimer ce vecteur. Par exemple, un vecteur médian peut être sélectionné comme le vecteur de référence. Également, nous pouvons envisager de choisir comme vecteur de référence une mesure, ou la moyenne d'un ensemble de mesures, effectué à un endroit spécifique de la digue. Les distances entre chacun des profils et le sous-espace signal correspondant peuvent être calculées pour construire une mesure de dissimilarité avec une distance faible signifiant moins de dissimilarité. Différentes types de distances peuvent être utilisées. Nous avons testé plusieurs distances (distance de Manhattan ou norme  $L1$ , distance Euclidienne ou norme  $L2$ , distance de Minkowski, etc.) et plusieurs façons à calculer le vecteur de référence sans avoir trouvé des différences notables. Pour cette raison nous nous sommes concentrés sur la norme  $L2$  calculée sur les colonnes du sous-espace résidu donné par la SVD.

La dissimilarité ainsi obtenue met en évidence les zones singulières comme les drains et les fuites. Les mesures de dissimilarité peuvent être perturbées par les erreurs d'estimation et/ou le bruit d'appareil de mesure. De plus, dans des cas réels, les réponses des zones non-singulières ne sont pas identiquement les mêmes selon la distance. Pour compléter la détection, un seuillage est alors appliqué sur la dissimilarité obtenue à chaque jour. Au lieu de choisir un seuil qui est le même peu importe le jour analysé, nous choisissons un seuil pour chaque jour. Ceci a l'avantage d'éviter des erreurs dues aux conditions climatiques qui ne sont pas obligatoirement identiques pour chaque jour. La dissimilarité pour un jour donné est ensuite modélisée par un mélange de distributions gamma et uniforme. Les singularités sont modélisées par une distribution uniforme dont les paramètres sont donnés par la gamme des valeurs de dissimilarité, avec le maximum représentant la dissimilarité la plus importante. La distribution gamma modélise les zones non-singulières et a été choisie parce que, grâce à ses deux paramètres, elle offre une grande flexibilité pour s'adapter à plusieurs situations. Une approche de taux de fausses alarmes constant est utilisée pour effectuer le seuillage en posant une probabilité de fausse alarme fixe. Le choix de la valeur de cette probabilité est souvent fait avec des tests rigoureux sous des conditions maîtrisées et pour des sites expérimentaux. N'ayant pas encore les possibilités d'effectuer ces tests, nous avons choisi une probabilité de fausses alarmes de  $10^{-5}$ . Les résultats de détection après seuillage ont permis une très bonne identification des singularités.

La méthode a été validée sur les données simulées ainsi que sur les données réelles enregistrées sur le site expérimental d'Oraison. A Oraison, deux cas intéressants de fuites artificielles de type

percolation et de fuites réelles ont été validés. Les fuites réelles, détectées autour du deuxième drain, ont été confirmées a posteriori par des inspections visuelles sur le site.



## Conclusion et Perspectives



# Conclusion et perspectives

Acteur mondial du domaine de l'énergie, de l'eau et de l'environnement, le groupe EDF gère plusieurs centrales hydrauliques pour la production d'électricité. Les différents canaux d'aménagements d'eau vers ces centrales sont bordés par des digues en terre. L'auscultation de ces digues est une priorité très importante non-seulement pour s'assurer de la stabilité des ouvrages, mais également pour s'assurer du fonctionnement optimal de la centrale hydraulique. L'une des causes majeures de la rupture des digues est l'érosion interne, résultant d'un écoulement préférentiel d'eau produisant une fuite. Il a été montré qu'une anomalie thermique étant produite lors de la génération d'une fuite, la mesure de température peut révéler des informations utiles sur la fuite. Une problématique importante de ce genre de mesure est liée à la taille des ouvrages qui sont généralement étendus sur plusieurs kilomètres. La notion de coût prend une importance élevée pour la surveillance des digues car un grand nombre de capteurs de température est requis. Dans ce cadre, les capteurs distribués de température (DTS), basés sur les fibres optiques, fournissent une solution efficace. Leurs avantages les plus importants sont leur capacité de multiplexer plusieurs capteurs sur une seule fibre, leur coût faible, leur robustesse et leur durée de vie importante. Dans le but de développer des systèmes d'auscultation de grand rendement, sans intervention humaine sur les sites mais avec une expertise centralisée à partir d'alertes automatiques, ce travail de recherche a abordé un nouveau problème de traitement des signaux thermométriques qui est celui de la détection des fuites d'eau dans des digues en terre à partir de mesures de température effectuées à l'aide des capteurs à fibre optique.

Les données de température spatio-temporelles sont enregistrées en fonction du déplacement le long de la digue et en fonction du temps. Dépendant de l'appareil DTS utilisé, les mesures ont une bonne résolution spatiale, de l'ordre de 1 m, et de température, entre  $0.01^{\circ}\text{C}$  et  $0.1^{\circ}\text{C}$ . Nous avons présenté comment l'interdépendance de ces différents paramètres affecte les données enregistrées. Les données thermométriques acquises exigent l'intervention des techniques avancées de traitement du signal pour extraire l'information relative aux fuites. Dans cette recherche, différentes méthodes ont été proposées afin de mieux exploiter les caractéristiques de données de température.

Dans l'étape de prétraitement des données, nous avons proposé un critère basé sur les statistiques d'ordre supérieur, le skewness et le kurtosis, pour identifier les phénomènes éphémères et



énergétiques en temps de type précipitation. Nous avons montré que l'occurrence d'un tel phénomène induit un changement de kurtosis et de skewness des données de température puisque la distribution de données est modifiée. La méthode a été validée pour le site d'Oraison après comparaison avec les données météorologique de précipitations. Ce critère a permis d'identifier des périodes temporelles qui peuvent générer des fausses alarmes lors de la détection des fuites par les méthodes d'analyse proposées. Cependant, ce critère n'a pas pu être validé sur les données acquises sur le site de Kembs. Comme la fibre optique de mesure de Kembs est dans le contre-canal, immergée dans l'eau du contre canal, l'influence des précipitations sur les températures acquises ne semble pas très importante. Les données enregistrées ont été modélisées par un mélange linéaire des différents facteurs influençant les mesures de température comme les structures existantes, les fuites, les variations journalières et saisonnières, les inhomogénéités du terrain, etc. La fibre optique de mesure étant enterrée dans le sol, la réponse du sol influence également les données acquises. Nous avons alors traité le problème de détection des fuites comme un problème de séparation de sources, avec les fuites constituant les sources les plus importantes à mettre en évidence.

Un algorithme fondé sur les techniques de séparation de sources, la SVD et l'ICA, a été proposé. Dans le contexte de données thermométriques, les sources estimées sont en fonction de la distance. La SVD permet l'estimation de sources orthogonales par ses vecteurs singuliers droits. Les signatures de fuites et de drains pouvant être représentées par des sources non-gaussiennes, le critère de l'indépendance de sources a été utilisé pour mieux extraire les facteurs correspondants. Cependant, la SVD permet une décomposition en sous-espaces orthogonaux de l'espace initial. La première source de la SVD a permis de caractériser la réponse du sol comme la composante la plus cohérente de données ayant la valeur singulière correspondante élevée. Une fois la réponse du sol extraite, l'ICA a été appliquée sur le résidu, ce qui a permis une meilleure identification des fuites. La méthode a été validée pour la détection des fuites simulées à partir de données synthétiques et des fuites artificielles et réelles sur le site d'Oraison. La richesse de l'algorithme en termes de différents paramètres le rend utile pour l'adaptation aux différentes situations. Différentes zones d'anomalies ont été mises en évidence sur le site de Kembs par cet algorithme, mais leur association aux phénomènes physiques requiert des études supplémentaires.

Lorsque nous ne disposons pas d'un grand nombre d'acquisitions ou quand nous voulons analyser la digue sur une seule journée, un détecteur de singularités en distance a été proposé. Les variations de températures journalières pour les zones singulières de type fuites, drains, etc., ont été montrées différentes de celles de zones non-singulières. Sous l'hypothèse d'avoir plus de zones non-singulières que de zones singulières, une mesure de dissimilarité a été proposée pour séparer les singularités de non-singularités. Cette dissimilarité étant basée sur la distance par rapport à un profil journalier de référence estimé à partir de données acquises, elle est bruitée par des erreurs d'estimation et par le bruit de mesure. Nous avons modélisé cette dissimilarité par un mélange des distributions gamma et uniforme en estimant leurs paramètres par une approche de maximum de vraisemblance à l'aide d'un algorithme EM. Un seuillage appliqué ensuite en fixant

une probabilité de fausse alarme permet la détection des singularités. La méthode a été validée sur les données simulées et sur les données réelles acquises sur le site d'Oraison. En revanche, les données acquises sur le site de Kembs contenant plusieurs zones d'anomalies qui n'ont pas été encore maîtrisées et le fait que la fibre optique soit parfois immergée dans l'eau du contre canal, perturbent l'estimation d'un profil journalier de référence. Par conséquent, les mesures de dissimilarité sont biaisées et les résultats obtenus par cette méthode sont inexploitable tels quels.

Par ailleurs, nous avons étudié empiriquement l'influence des différents paramètres dans les algorithmes d'analyse (taille de fenêtre d'analyse, choix du nombre de sources, etc.). Il sera intéressant de développer des critères analytiques pour caractériser les différents paramètres des algorithmes prenant en compte des particularités des sites analysés. Par exemple, les différents sites peuvent être classés selon leur caractère, la composition des matériaux constituant le terrain, le nombre des anomalies, etc., ce qui permettra de définir des paramètres pour chacune de ses classes.

Les méthodes présentées dans ce travail prennent en compte uniquement les données enregistrées par la fibre optique à l'exclusion de toute autre information. Il sera intéressant d'explorer l'apport d'informations supplémentaires comme la température de l'air, le rayonnement solaire, etc. Un profil journalier de température de l'eau du canal pourra, par exemple être considéré comme un vecteur de référence pour le calcul d'une mesure de dissimilarité.

Plusieurs anomalies existent sur le site de Kembs, non seulement en termes de fuites réelles mais également en termes de zones de fibre immergées dans l'eau. De plus, plusieurs fuites réelles ont des variations importantes de débit au cours du temps. Lors des inspections visuelles, certaines de ces anomalies ont été identifiées, mais, malheureusement, nous ne disposons pas de données enregistrées pendant la même période. Une analyse des enregistrements coïncidant avec les visites de terrain sera primordiale. Le bruit de mesure de l'appareil DTS Siemens jouant sur la dynamique des signaux enregistrés, il sera intéressant soit de trouver des critères supplémentaires pour enlever ce bruit, soit d'avoir des mesures avec l'appareil DTS Sensornet qui offre une meilleure résolution en température de  $0.01^{\circ}\text{C}$ . D'autres prétraitements plus efficaces doivent être mis au point afin de pouvoir distinguer le bruit de mesure et les informations utiles.

La détection de fuites par mesures thermométriques étant un domaine récent, les modèles physiques définissant la génération des fuites ne sont pas encore bien développés. EDF étant en train de développer une expertise dans ce cadre, les analyses peuvent profiter d'une meilleure compréhension de ces modèles physiques qui sont à la base de génération de fuites.

Jusqu'à présent, nous avons généré des données synthétiques de température en utilisant une modélisation numérique d'une section (2D) de digue. Un autre axe de recherche important sera la génération de modèles de digues plus réalistes en intégrant une troisième dimension (la longueur de la digue) et des notions d'emplacement (élévation) de la fibre et d'hétérogénéités du terrain. La simulation pourra profiter de la description et/ou de la classification des signatures

de fuite attendues. Cette connaissance de signatures de fuites permettra également d'améliorer les méthodes proposées en exploitant de manière optimale les statistiques de données.

Le but de ce travail était de détecter les fuites afin de les localiser en distance et en temps. La caractérisation complète de fuites consiste également à déterminer leur débit. Les méthodes présentées dans cette thèse ne sont pas exploitables dans ce sens. Cependant, la maîtrise des données synthétiques avec la possibilité de génération des fuites de débit variable (connu) couplé à des tests exhaustifs de méthodes proposées devra permettre de caractériser les débits de fuites réelles.

Dans la thématique détection/estimation des fuites par fibre optique, une thèse consacrée à l'estimation des débits de fuites a été démarrée au Laboratoire d'étude des Transferts en Hydrologie et Environnement de Grenoble INP. Nous envisageons d'analyser les méthodes développées afin de tester leur optimalité et éventuellement de les améliorer. Les mesures thermométriques offrent une solution prometteuse pour la détection des fuites, mais la meilleure exploitation de ces mesures nécessite un mariage entre les aspects physiques et mathématiques.

# Annexes



# Annexe A

## Autres méthodes de détection de fuites de type percolation

### A.1 Filtrage Multicomposante

#### A.1.1 Introduction

L'analyse multicomposante a été utilisée depuis longtemps pour diverses applications dans le domaine du traitement du signal et des images. L'existence des composantes multiples d'un même signal permet d'exploiter les redondances d'information. Une information utile pourra être distribuée sur les différentes composantes avec des compositions différentes. Le traitement conjoint de toutes les composantes peut révéler une information qui n'est pas évidente ou clairement présente dans chacune des composants indépendamment [Paulus 05, Mars 06].

Dans le cadre de notre application, la température acquise par la fibre optique dépend non seulement des anomalies (des fuites d'eau présentes sur la digue) mais également des facteurs externes comme les variations saisonnières, les précipitations, les structures existantes, etc. Tous ces facteurs ont leur caractère particulier. Alors que certains sont localisés en temps (les précipitations, les variations saisonnières), d'autres sont localisés en distance (les drains). Ces redondances spatiales et temporelles peuvent être utilisées pour la détection des fuites.

Au départ, le DTS fournit un signal monocomposante (1C-2D) à deux dimensions (temps et distance). Nous pouvons construire un signal à trois composantes (3C-2D) à partir du signal acquis (1C-2D). A l'aide des techniques de filtrage vectoriel de type vecteur médian (VMF), souvent utilisées en traitement d'image couleur, la présence des fuites de type percolation peut être alors mise en évidence. Le filtrage médian classique 1D ou 2D, appliqué directement sur les données acquises (1C-2D), ne permet pas une extraction parfaite des fuites de type percolation.

#### A.1.2 Extension multicomposante (3C-2D) des données monocomposantes (1C-2D)

Considérons les données monocomposantes,  $Y$  en fonction du temps et de la distance enregistrées sur le site d'Oraison. Les données sont influencées par les fuites de type percolation mais également par d'autres phénomènes comme les précipitations, la réponse du sol, les variations saisonnières, etc. Les caractéristiques propres de tous ces facteurs sont leur localisation soit en temps soit en distance. Par exemple, une précipitation est un facteur bien localisé en temps mais qui varie lentement en distance. Les structures comme les drains sont bien localisées en distance mais existent toujours en temps. Cependant, les facteurs que nous cherchons, à savoir les fuites de type percolation, sont aussi bien localisés en temps

## ANNEXE A. Autres méthodes de détection de fuites de type percolation

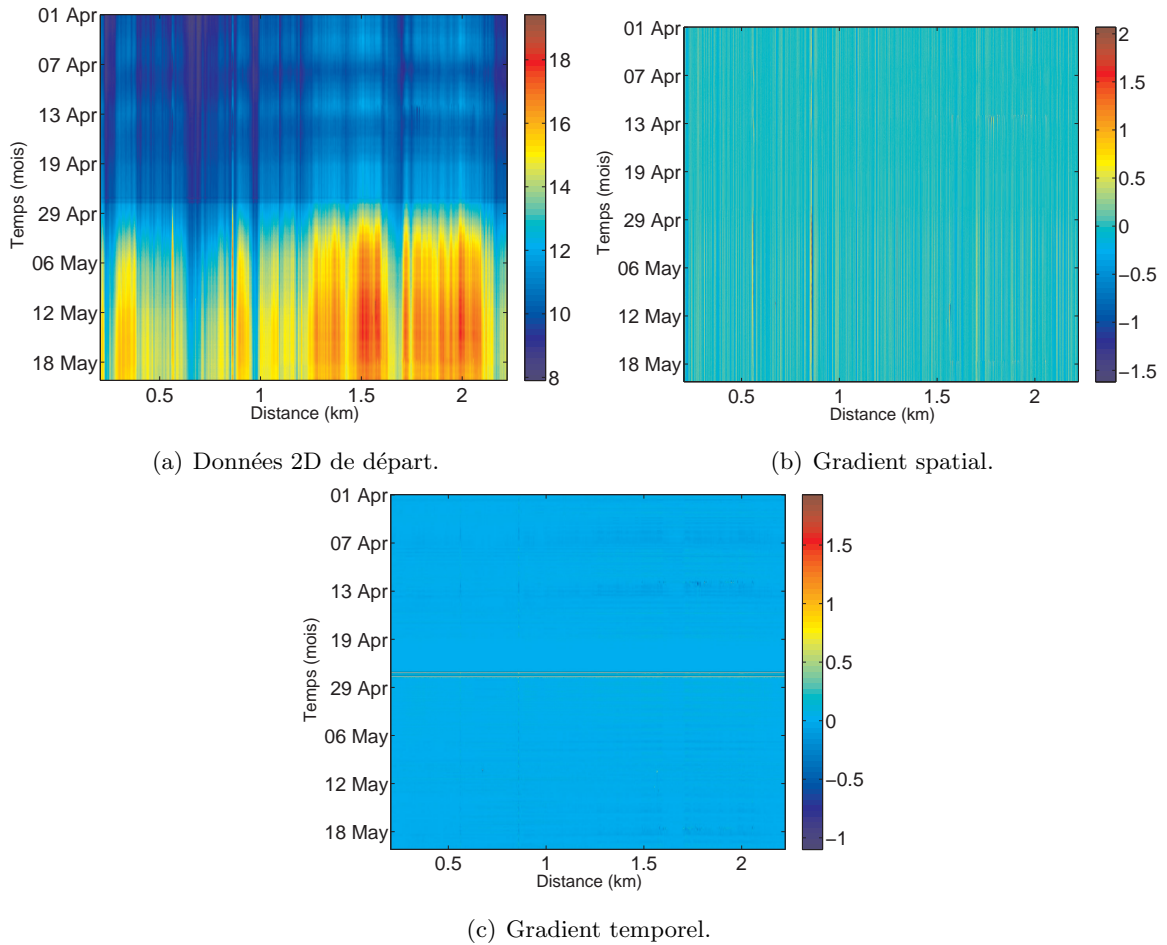
qu'en distance. L'utilisation d'un opérateur de gradient pourrait exploiter ces caractéristiques inhérentes [Khan 08b].

Pour le gradient temporel, le gradient  $\frac{\partial \mathbf{Y}}{\partial t}$  est appliqué pour chaque capteur pour toute la durée du signal alors que pour le gradient spatial, le gradient  $\frac{\partial \mathbf{Y}}{\partial x}$  est appliqué pour chaque profil sur toute la longueur de la fibre. Le gradient temporel atténue les phénomènes qui varient lentement en temps comme les structures existantes (drains) alors que le gradient spatial atténue les phénomènes variant lentement en distance comme la réponse du sol ainsi que les précipitations.

Un signal 3-C peut être alors construit sous la forme d'un cube de données :

$$\mathcal{Y} = \left\{ \mathcal{Y}(:, :, 1) = \mathbf{Y}, \mathcal{Y}(:, :, 2) = \frac{\partial \mathbf{Y}}{\partial x}, \mathcal{Y}(:, :, 3) = \frac{\partial \mathbf{Y}}{\partial t} \right\}$$

Ces données multicomposantes, présentées sur la figure A.1 pour le site d'Oraison, ont été enregistrées pendant la période entre 01 avril et 20 mai 2005, une période qui contient les fuites de type percolation. (Pour information, la ligne verticale dans la figure A.1(c) représente l'interpolation faite pour les données manquantes sur trois jours).



**Fig. A.1:** Données multicomposantes à partir de données monocomposantes.

### A.1.3 Filtrage multicomposante

A partir du cube de données construit précédemment, nous pouvons remarquer qu'à chaque distance  $x$ , et à chaque temps  $t$ , nous avons un vecteur,  $\mathbf{y}_i = \mathcal{Y}(t, x, \cdot) \in \mathbf{R}^3$ . Ces vecteurs, à différents temps et positions, vont contenir au plus deux éléments importants. Ceci est vrai pour toutes les positions et pour tous les temps sauf pour les positions et les instants correspondant à des fuites de type percolation. A ces positions, les trois éléments seront importants.

La distance calculée entre un vecteur correspondant à des fuites et d'autres vecteurs va alors être importante par rapport aux distances calculées entre les autres zones (sans fuites). Une approche basée sur le filtrage vectoriel peut être envisagée pour exploiter ce type d'information. Une des méthodes est le filtrage médian vectoriel (VMF pour Vector Median Filtering) proposé comme une extension du filtrage médian scalaire pour éliminer le bruit impulsionnel [Astola 90]. Le VMF a été introduit en utilisant des statistiques d'ordres basées sur la métrique de Minkowski afin de quantifier la distance entre deux vecteurs [Lukac 05].

Nous nous intéressons ici à un cas particulier de distance, à savoir la distance Euclidienne. Compte tenu des remarques effectuées ci-dessus et d'une étude empirique de données dont on dispose, ce type de distance permet d'obtenir de très bons résultats.

La formulation du résultat du VMF est donnée par l'équation :

$$\mathbf{y}_{min} = \arg \min_{\mathbf{y}_i \in \mathbf{W}} \sum_{j=1}^N \|\mathbf{y}_i - \mathbf{y}_j\|_2, \quad (\text{A.1})$$

où  $\mathbf{W}$  est la fenêtre de filtrage contenant les  $N$  vecteurs à comparer  $\mathbf{y}_i$  (ou  $\mathbf{y}_j$ ),  $\mathbf{y}_{min}$  le vecteur choisi parmi les  $N$  vecteurs dans la fenêtre de traitement et  $\|\cdot\|_2$  la norme  $L_2$  appelée distance Euclidienne.

Le signal acquis contient trois fuites et un point chaud qui ont tous des signatures (informations) presque impulsives en temps et en distance contrairement aux autres phénomènes mis en évidence par le signal acquis. L'effet de l'algorithme VMF sera alors d'annihiler cette information impulsionnelle. Le résidu obtenu ensuite comme la différence entre les données initiales et les données filtrées ne contiendra que des informations liées aux fuites.

### A.1.4 Application sur les données d'Oraison

La technique VMF présentée à la section précédente est appliquée sur le signal 3C-2D en considérant une fenêtre glissante,  $\mathbf{W}$ , de taille  $3 \times 3$ , ce qui donne  $N = 9$  vecteurs dans l'eq. (A.1). Ce choix a été fait parce qu'il correspond à 3 mètres de longueur et une durée de 6 heures, permettant ainsi d'incorporer les petites dynamiques en temps et en distance sauf pour les cas des fuites et peut être des précipitations. Les données filtrées,  $\mathcal{Y}_{filtered}$ , sont obtenues par glissement de la fenêtre afin de balayer toutes les positions ( $x$ ) et tous les instants ( $t$ ) en remplaçant le centre de chaque fenêtre par le résultat de l'eq. (A.1).

Le résidu obtenu sur la première composante est :

$$\mathbf{Y}_{residue} = \mathbf{Y} - \mathbf{Y}_{filtered},$$

où  $\mathbf{Y}$  représente la matrice de données acquises et  $\mathbf{Y}_{filtered}$  la première composante de  $\mathcal{Y}_{filtered}$  (c'est.à.d.  $\mathcal{Y}_{filtered}(:, :, 1)$ ).

Le résidu obtenu, affiché sur la figure A.2, contient plutôt des informations liées aux fuites. Ceci est mis en évidence par un zoom en proximité des fuites où nous pouvons constater que la séparation fuites/fond



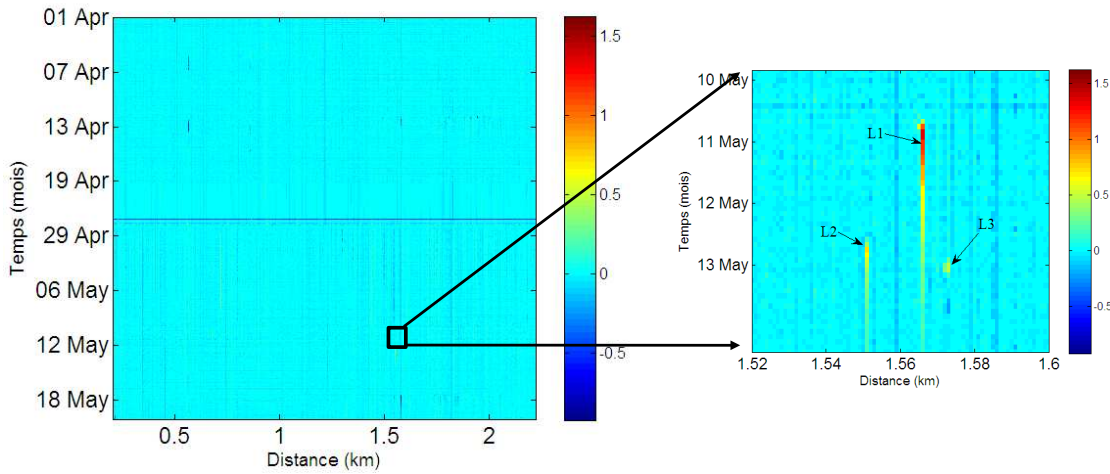
est nette. Pour améliorer la détection des fuites dans le résidu, un seuillage peut être appliqué. Le seuil a été choisi en s'appuyant sur l'aspect physique de la détection de température par DTS.

La température d'une fuite est proportionnelle à la différence entre la température de l'air et celle de l'eau du canal mais aussi proportionnelle au débit de fuite. Les effets externes comme la vitesse du vent, etc. peuvent être négligés dans le cas de fuites de type percolation en supposant que ces fuites sont proches de capteurs. Nous pouvons constater sur les données qu'une fuite de bas débit (1 lit/min), existante dans une zone où la différence entre la température de l'air est celle de l'eau est faible (le soir), apporte un changement de  $0.51^{\circ}C$ . Nous pouvons donc fixer le seuil à  $0.5^{\circ}C$ . Il faut mentionner ici que ce choix de seuil n'a pas été automatisé. Il faut mentionner ici qu'un seuil automatique pourrait être envisagé en modélisant le résidu (son module ou que sa partie positive en supposant la température de la fuite plus grande que celle du terrain) par un mélange de distributions gamme et uniforme comme présenté à la section 4.3.

Le résultat après seuillage est :

$$y_{residue}^{th}(t, x) = \begin{cases} y_{residue}(t, x) & \text{si } y_{residue}(x, t) > \text{threshold} \\ 0 & \text{sinon} \end{cases}$$

Un zoom sur le résidu après seuillage dans une zone à proximité des fuites et du point chaud montre



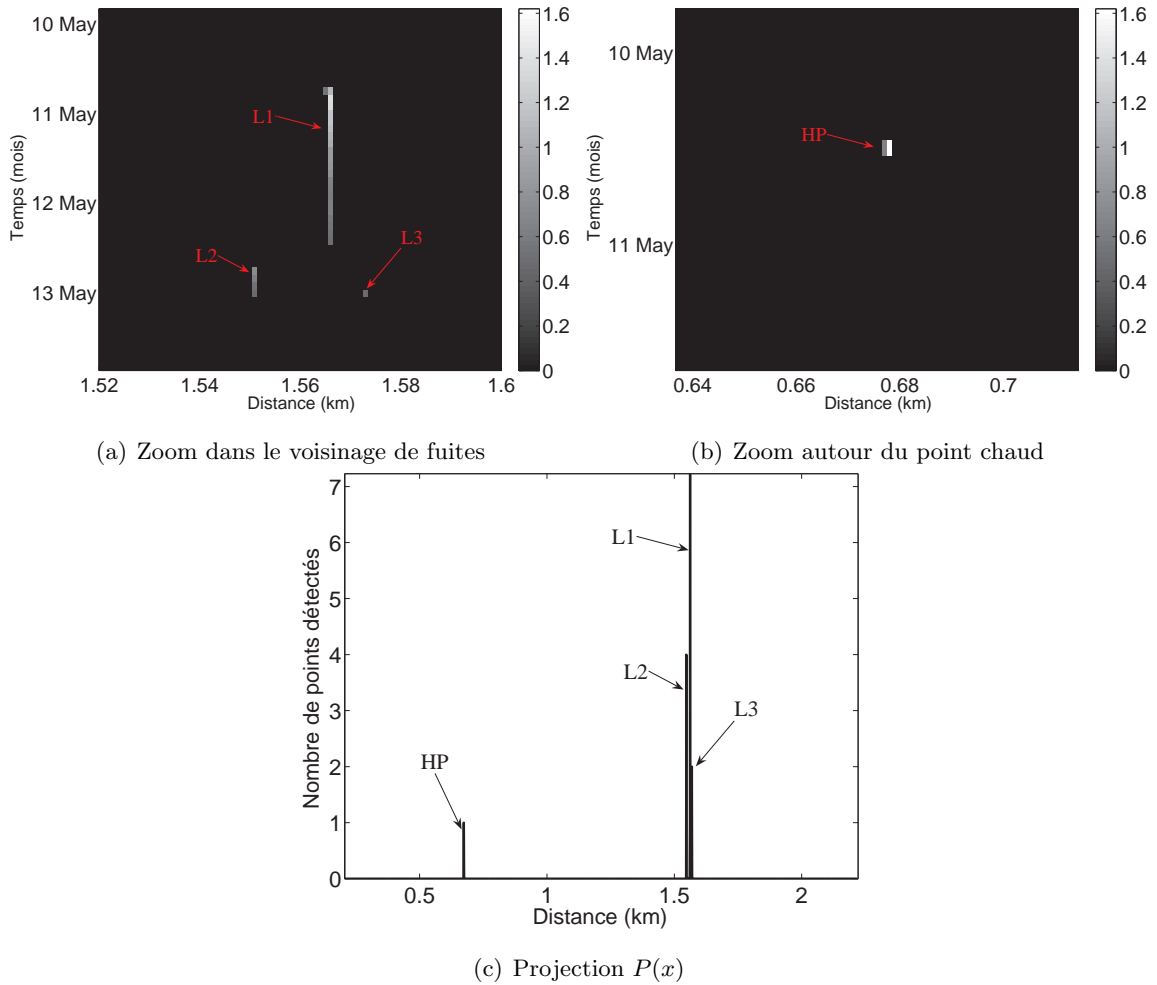
**Fig. A.2:** Résidu  $Y_{residue}$  obtenu en appliquant le VMF sur les données 3C-2D. Zoom du résidu en proximité des fuites montre une bonne séparation.

que le résidu est parfait pour les besoins de détection (figures A.3(a)-(b)). La même observation peut être faite en considérant la projection du résidu (figure A.3(c)) après seuillage sur l'axe des distance :

$$P(x) = \sum_{t=1}^{N_t} y_{residue}^{th}(t, x), \quad (\text{A.2})$$

Cette projection montre que seules les trois fuites et le point chaud sont détectés dans le résidu après seuillage. Le résidu est dépourvu de tous les autres phénomènes présents dans le signal initial.

Pour mettre en évidence les avantages de l'analyse multicomposante par rapport à l'analyse monocomposante, nous testons également le filtrage médian scalaire. Les résultats du filtrage médian 2D pour une



**Fig. A.3:** Détection de fuites dans le résidu de VMF.

fenêtre  $3 \times 3$  utilisant la même valeur pour le seuillage ne révèle ni le point chaud ni la fuite  $L3$ . La détection est alors incomplète. Si nous admettons un seuil plus bas pour garder la fuite  $L3$  dans le résidu, nous obtenons les résultats affichés sur la figure A.4. Nous pouvons constater que le résidu contient des informations liées aux drains non désirés (figure A.4(a)) dans cette étude. La détection de drains est liée au fait que les drains sont plus exposés à la température de l'air et présentent des variations jour/nuit plus importante que les autres capteurs. L'application du filtrage médian 2D sur le gradient spatial (figure A.4(b)) donne une mauvaise localisation du point chaud et des fuites et met en évidence l'apparition de faux phénomènes autour des positions 1.8km. L'application du filtrage médian 2D sur le gradient temporel donne des résultats complètement non-utilisables. Le choix d'un seuil par rapport à la fuite  $L3$  rend la séparation des fuites impossible.

Le filtrage médian 1D appliqué sur chaque capteur (colonne par colonne sur la matrice  $\mathbf{Y}$ ) ne donne pas des bons résultats. Les résultats obtenus par l'application du filtrage médian 1D pour chaque acquisition (ligne par ligne sur la matrice  $\mathbf{Y}$ ) montre que les fuites détectées sont "corrompues" par les drains qui forment une partie assez dominante du résidu. Le point chaud n'est pas du tout détecté dans ce cas. En plus, il y a certains points inconnus identifiés dans le résidu, quelques uns parmi ces points pourraient

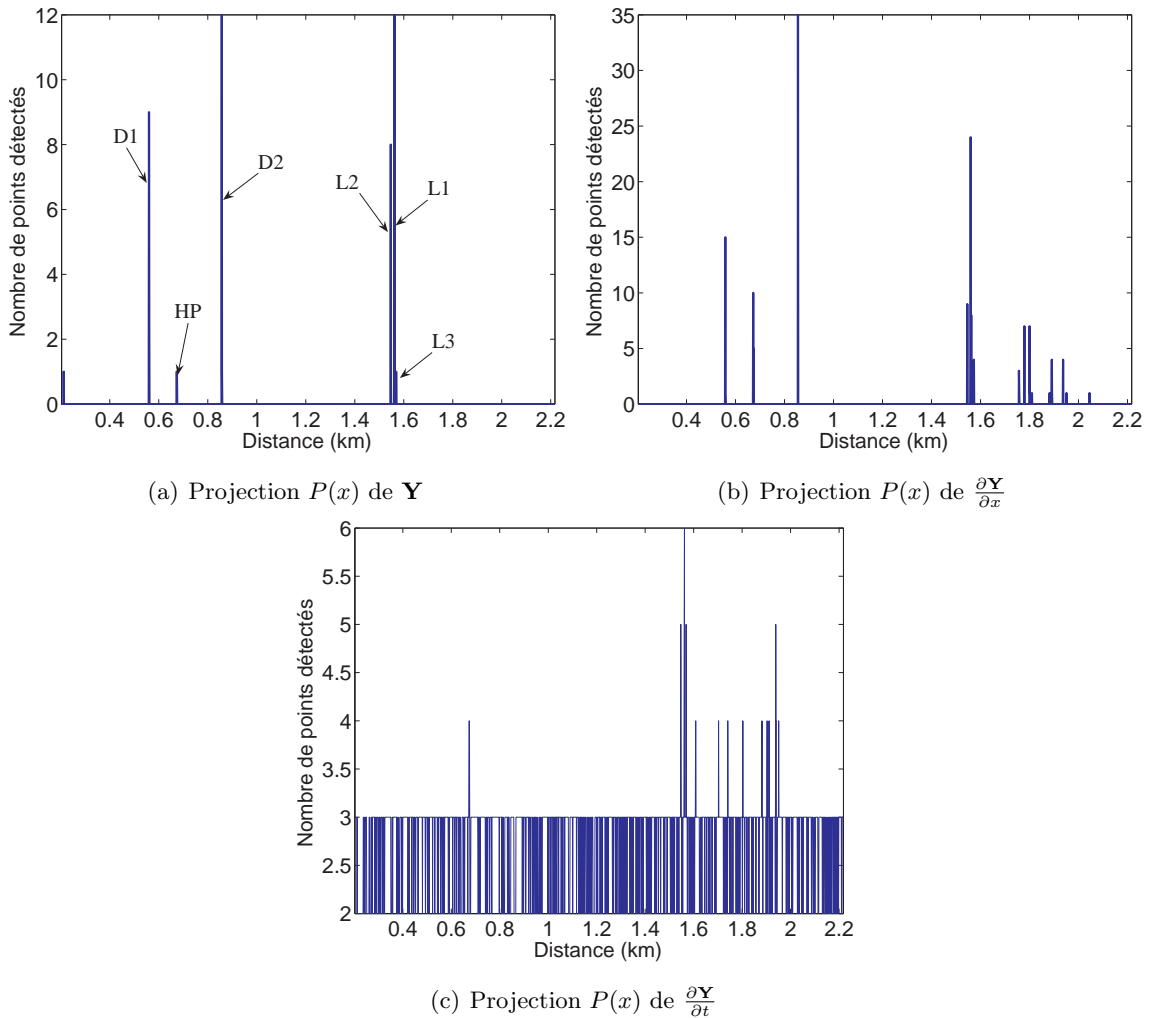


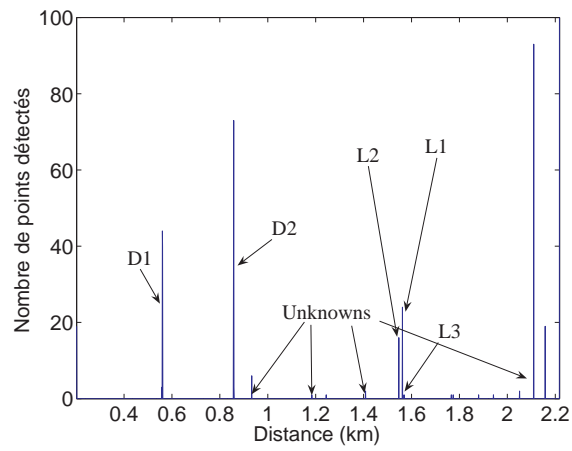
Fig. A.4: Résultats du filtrage médian 2D sur les 3 composantes pour une fenêtre  $3 \times 3$ .

être attribués aux zones spatiales plus sensibles aux précipitations. L'application du filtrage sur les deux autres composantes donne des mauvais résultats.

Les résultats avec d'autres tailles de fenêtre ne présentent pas d'amélioration par rapport à ces résultats. Le filtrage multicomposante présente une solution efficace pour la détection des fuites de type percolation et le point chaud. Il permet en même temps de définir un seuil basé sur la physique d'acquisition.

### A.1.5 Conclusion

Nous avons montré comment les gradients de type spatial et temporel permettent de construire de données multicomposantes 3C-2D. L'application du gradient spatial permet d'enlever les variations lentes en distance pendant que le gradient temporel permet d'enlever les variations lentes en temps. Les fuites étant bien localisées en temps et en distance sont ensuite enlevées par l'application d'un filtrage médian vectoriel sur ce signal 3C-2D. Un seuillage est fait sur le résidu de ce filtrage afin d'améliorer la détection de fuites. Le résidu après seuillage ne contient que des informations liées aux fuites et au point chaud, ce qui montre que la méthode proposée est très efficace. Pour l'identification des fuites réelles, la méthode



**Fig. A.5:** Projection  $P(x)$  du résultat du filtrage médian 1D appliqué sur les colonnes de  $\mathbf{Y}$  pour une fenêtre de dimension 3.

peut être efficace dans le cas où une nouvelle fuite impulsionnelle en distance vient d'apparaître, fuite se comportant comme une fuite de type percolation pour la période d'analyse.

## A.2 Détection des fuites de type percolation (JADE vs FastICA)

Nous comparons ici les résultats de deux algorithmes de séparation aveugle de sources : JADE et FastICA pour la détection des fuites de type percolation. Les deux algorithmes sont appliqués en mode symétrique (toutes les sources sont estimées en même temps) puis l'algorithme FastICA est appliqué en mode déflation (les sources sont estimées les unes après les autres). L'algorithme FastICA en mode déflation (et sans réduction de la dimensionnalité) permet d'estimer des sources impulsionnelles en distance, ce qui peut s'avérer très intéressant pour identifier des fuites de type percolation. Toutefois, il faut rappeler que même si cette approche nommée "Spikes et Bumps" [Hyvärinen 99c] permet d'estimation des fuites impulsionnelles, il n'est pas facile de conférer un sens mathématique aux résultats.

Les données considérées pour l'analyse sont celles comprenant les fuites artificielles de type percolation pour la période entre le 30 avril et le 13 avril, 2005. Dans un premier temps, le sous espace résidu est obtenu en enlevant le sous-espace signal correspondant à la première source SVD  $\mathbf{v}_1$  suivant l'approche du chapitre 3.

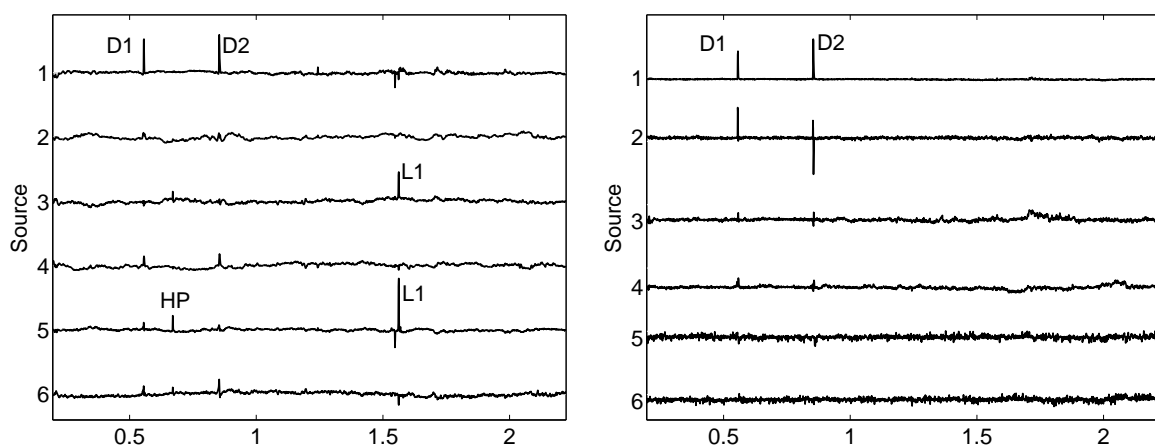
Ensuite l'ICA pourra être appliquée en deux modes : symétrique et déflation. Dans le mode symétrique, toutes les sources sont estimées en parallèle (en même temps) alors que dans le mode déflation une source est estimée à chaque pas (les unes après les autres), c'est-à-dire l'algorithme estime une source en utilisant le sous-espace bruit de l'étape précédente.

Pour appliquer les algorithmes en mode symétrique, il faut faire une réduction de dimensionnalité. Ici, nous pouvons considérer 6 sources à estimer,  $\mathbf{v}_2$  to  $\mathbf{v}_7$ , associées aux 6 phénomènes qui nous intéressent ( $D1$ ,  $D2$ ,  $L1$ ,  $L2$ ,  $L3$  and  $HP$ ). Les résultats obtenus en mode symétrique pour les deux algorithmes JADE et FastICA sont assez proches et des tests ont montré qu'au plus 3 sources uniques pouvaient être identifiées parmi les 6 sources estimées,  $s_i$ , comme présenté sur la Fig. A.6(a). La 1<sup>ère</sup> source estime les drains,  $D1$  et  $D2$ , la 2<sup>ème</sup> source estime une partie des drains noyée dans du bruit, la 3<sup>ème</sup> et la 5<sup>ème</sup> estiment la fuite  $L1$  et le point chaud,  $HP$ .

La construction d'un sous-espace résidu par soustraction du sous-espace correspondant aux drains va révéler toujours les fuites et le point chaud ce qui revient au schéma du chapitre 3. Nous pouvons constater à partir de cette image que la fuite la plus énergétique,  $L1$ , est identifiée parmi les sources estimées mais les autres fuites ne sont pas identifiées sur les 6 sources. Afin d'identifier ces fuites moins énergétiques, nous utilisons la méthode de déflation sans faire la réduction de dimensionnalité.

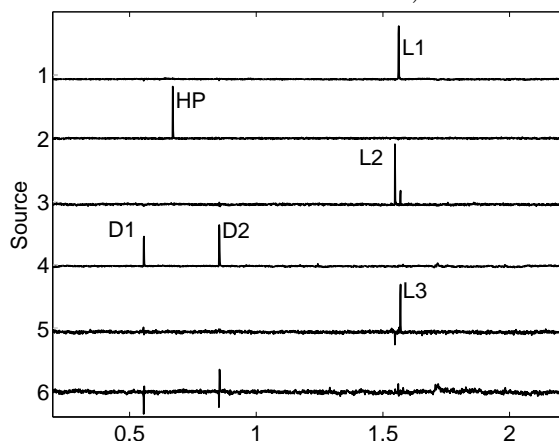
Il faut mentionner que le principe de déflation n'est pas incorporé par défaut dans l'algorithme JADE mais nous pouvons forcer l'algorithme à estimer une seule source, puis à travailler sur le résidu ainsi obtenu et ainsi de suite. Cependant, la contrainte de réduction de dimensionnalité existe toujours pour JADE et les résultats obtenus avec JADE en mode déflation sont similaires à ceux obtenus en mode symétrique. FastICA, grâce à son efficacité supérieure en temps de calcul, permet de traiter les données sans faire de réduction de la dimensionnalité, mais comme déjà mentionné un sens mathématique ne peut pas être accordé au résultat obtenu. Néanmoins, pour le cas des fuites simulées de type percolation, nous essayons d'estimer des sources impulsionnelles en distance et cette approche s'avère efficace.

FastICA en mode déflation sans réduction de dimensionnalité est appliquée dans deux zones temporelles, une sans fuite et l'autre avec fuite. Les 6 sources ainsi obtenues sont affichées sur la Fig. A.6(b-c). Toutes les courbes de la Fig. A.6 ont la même dynamique en amplitude. Pour la zone sans fuite, les sources estimées (Fig. A.6(b)) montrent seulement les drains et le bruit. Sur la zone avec fuite, tous les phénomènes



(a) JADE avec réduction de dimensionnalité réduction

(b) FastICA sans réduction de dimensionnalité (zone sans fuites)



(c) FastICA sans réduction de dimensionnalité (zone avec fuites)

**Fig. A.6:** Sources estimées par ICA en modes : (a) symétrique et (b-c) déflation comme signaux indexés par le temps (km).

contenus dans les données, c'est-à-dire les trois fuites, le point chaud et les drains, sont détectés à leurs positions exactes avec la fuite  $L1$  sur la source 1, le point chaud sur la source 2 et les fuites  $L2$  et  $L3$  sur les sources 3 et 5 (Fig. A.6(c)). Le fait que les drains soient présents sur plusieurs mètres implique qu'ils sont détectés sur quelques sources qui sont géographiquement très proche en distance. On voit le drain  $D2$  présent à la position 0.854 km sur la source 4 et à la position 0.855 km sur la source 6. Des tests avec différentes non-linéarités de FastICA ne montrent pas des différences significatives.



# Annexe B

Annexe relative à la réponse du sol

## B.1 Extraction de la réponse du sol par critère HOS en utilisant une approche de moindres carrés

Pour les données simulées ainsi que pour les données acquises au site d'Oraison, nous avons constaté que la première source SVD peut être liée à la réponse du sol. Il est donc impératif de bien estimer cette réponse afin de l'enlever pour mieux faire ressortir les autres phénomènes présents lors de l'acquisition, notamment les fuites. L'estimation de cette source par une SVD globale présente deux problèmes majeurs. Le premier est que la source estimée risque d'être influencée par des phénomènes éphémères énergétiques en temps comme par exemple des précipitations importantes. Le deuxième problème est que la SVD assume non-seulement l'orthogonalité des sources mais également l'orthogonalité de variations de ces sources ce qui n'est pas justifiable physiquement. Le sous-espace résidu peut être donc affecté par la modification de l'information existante ou par l'ajout d'une autre information compensant cette orthogonalité.

Un critère basé sur des statistiques d'ordre supérieur a été proposé pour l'identification des phénomènes éphémères, comme les précipitations, exploitant le fait qu'ils présentent un comportement statistique spécial. En sélectionnant les données à l'aide de ce critère nous évitons les zones temporelles affectées par des phénomènes éphémères. Une première source est estimée par fenêtrage glissant sur les données sélectionnées en utilisant la classique décomposition SVD. Un vecteur caractéristique peut être ensuite obtenu par moyennage de vecteurs obtenus aux différentes positions de la fenêtre d'analyse ou comme le vecteur médian calculé à partir des vecteurs obtenus dans les différentes fenêtres si les fuites à estimer sont très énergétiques dans la fenêtre d'analyse. Ce vecteur caractéristique donne un meilleur estimateur de la réponse du sol. De plus, pour construire le sous-espace correspondant à cette source, la condition physiquement non-justifiée d'orthogonalité des variations temporelles de sources est évitée en utilisant une approche de moindres carrés.

### B.1.1 Estimation d'une source moyenne

Nous considérons les données acquises pendant une durée de 5 semaines sur le site d'Oraison. Ces données contiennent les fuites de types percolation, les deux drains et des instances de précipitation.

Dans la formulation de la SVD (équation (3.4), page 82), le premier vecteur  $\mathbf{v}_1$  caractérise la réponse du sol. La condition d'orthogonalité des sources  $\mathbf{v}_j$  peut être justifiée en considérant qu'elles sont issues des phénomènes physiquement indépendants l'un de l'autre. De l'autre côté, les vecteurs  $\mathbf{u}_j$  représentant les variations temporelles de ces sources sont eux aussi orthogonaux par construction, ce qui n'est pas ici



physiquement justifiable.

L'estimation de la première source,  $\mathbf{v}_1$ , par l'application de la SVD sur toute la longueur temporelle de données risque d'être influencé par des phénomènes éphémères énergétiques comme des précipitations. Pour résoudre ce problème, la SVD est calculée sur de petites zones temporelles en utilisant une approche par fenêtre glissante. Pour cela, il faut choisir la taille ( $\Delta T$ ) de la fenêtre mais aussi l'intervalle de recouvrement. Comme les sources estimées sont des fonctions indexées par la distance, nous nous focalisons à éliminer les phénomènes éphémères temporels. Prenant en compte le  $m^{\text{th}}$  bloc de données,  $\mathbf{Y}_m = \{y(t, x) | t_m \leq t \leq t_m + \Delta T, 1 \leq x \leq N_x\}$ , on peut écrire la décomposition SVD glissante comme :

$$\mathbf{Y}_m = \mathbf{U}_N^m \Delta_N^m \mathbf{V}_N^{mT} = \sum_{j=1}^N \beta_j^m \mathbf{u}_j^m \mathbf{v}_j^{mT} \quad (\text{B.1})$$

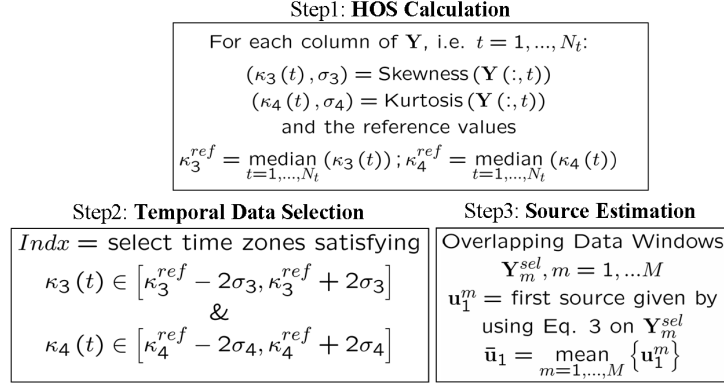
où  $N = \min(N_x, \Delta T)$  et  $t_m$  dépendant de  $\Delta T$  et du recouvrement.

Parmi les premières sources SVD,  $\mathbf{v}_1^m$ , avec  $m = 1, \dots, M$  et  $M$  étant le nombre total de blocs, il y aura des sources très proches l'une de l'autre liées à la réponse du sol pendant que les autres seront influencées par des phénomènes éphémères. Le but est de sélectionner un vecteur moyen  $\bar{\mathbf{v}}_1$  à partir de ces vecteurs  $\mathbf{v}_1^m$ . L'application d'un opérateur de moyennage pour cette sélection assure que le vecteur sélectionné sera bien adapté pour toutes les zones temporelles. Cependant, il y aura des zones où les vecteurs sont dominés par des phénomènes autres que la réponse du sol (par exemple des précipitations) et, s'ils sont importants, ces phénomènes peuvent introduire des perturbations lors de l'estimation d'un vecteur moyen.

Pour éviter une telle situation, nous utilisons un critère basé sur des statistiques d'ordre supérieur (voir section 2.6.3 page 71 pour les détails). Ce critère permettra d'identifier ces zones et donc d'enlever leur contribution dans le calcul d'un vecteur moyen. Le skewness et le kurtosis mettent en évidence les phénomènes éphémères énergétiques. Il faut noter que les variances de ces deux estimateurs (du skewness et du kurtosis) ne dépendent pas du temps,  $t$ , mais uniquement du nombre d'éléments utilisés pour les estimer, c'est-à-dire du nombre des capteurs dans notre application.

La sélection de données nécessite l'utilisation de seuils pour les deux grandeurs (skewness et kurtosis). Nous avons constaté empiriquement qu'un seuil de  $\pm 2\sigma$  autour des valeurs de référence des estimateurs est bien adapté si on considère les valeurs références pour une saison (été par exemple). Si nous diminuons le seuil, par exemple à  $\pm\sigma$ , nous n'arrivons pas à bien séparer les zones affectées par des phénomènes non désirés des zones utiles. Par contre, le résultat n'est pas trop sensible entre un seuil de  $\pm 2\sigma$  et un seuil de  $\pm 3\sigma$ . Nous pouvons donc choisir un seuil de  $\pm 2\sigma$ . Les valeurs de référence,  $\kappa_3^{ref}$  et  $\kappa_4^{ref}$ , sont calculées ici comme les médianes du skewness  $\kappa_3(t)$  et du kurtosis,  $\kappa_4(t)$  respectivement. Nous avons choisi l'opérateur médian pour éviter des facteurs qui peuvent apparaître occasionnellement comme des perturbations impulsives, des précipitations significatives, etc.

Ayant supprimé les phénomènes éphémères de données brutes,  $\mathbf{Y}$ , par le critère ci-dessus, nous pouvons construire un nouveau jeu de données réduites,  $\mathbf{Y}^{sel}$ . La SVD glissante de l'équation (B.1) est ensuite appliquée sur les nouvelles données  $\mathbf{Y}^{sel}$ . Un vecteur moyen caractérisant la réponse du sol est alors estimé par moyennage de premiers vecteurs obtenus pour chaque position de la fenêtre glissante. Il faut mentionner ici que les fuites très énergétiques peuvent influencer la moyenne, dans un tel cas, le vecteur médian pourra remplacer le vecteur moyen. Un résumé de l'approche suivie est donné dans l'algorithme à trois étapes de la figure B.1. Une fois le vecteur caractéristique calculé, nous pouvons poursuivre avec l'étape de séparation en sous-espaces.



**Fig. B.1:** Algorithme pour l'estimation d'une source moyenne caractérisant la réponse du sol.

### B.1.2 Décomposition en sous-espace par approche de moindres carrés

La construction d'un sous-espace correspondant à la source moyenne,  $\bar{\mathbf{v}}_1$ , nécessite la connaissance des variations temporelles de cette source. Une approche possible est d'utiliser la notion de "unmixing" de source comme proposé dans le domaine d'imagerie hyperspectrale [Heinz 01]. Pour cela, nous pouvons poser un modèle linéaire de données comme :

$$\mathbf{Y} = \mathbf{Y}_{sig}^{LS} + \mathbf{Y}_{residue}^{LS} = \alpha \bar{\mathbf{v}}_1 + \mathbf{Y}_{residue}^{LS}$$

où  $\alpha$  contient les variations temporelles du vecteur  $\bar{\mathbf{v}}_1$  sur les données enregistrées.  $\alpha$  et  $\bar{\mathbf{v}}_1$  définissent le sous-espace signal,  $\mathbf{Y}_{sig}^{LS}$ . Le résidu  $\mathbf{Y}_{residue}^{LS}$  est défini par des phénomènes autres que la réponse du sol.

En supposant le sous-espace résidu distribuée selon la loi normale autour de la réponse du sol, le vecteur  $\alpha$  peut être estimé en utilisant une procédure par moindres carrés (LS - least squares en anglais) :

$$\hat{\alpha}_{LS} = (\bar{\mathbf{v}}_1^T \bar{\mathbf{v}}_1)^{-1} \bar{\mathbf{v}}_1^T \mathbf{Y} \quad (\text{B.2})$$

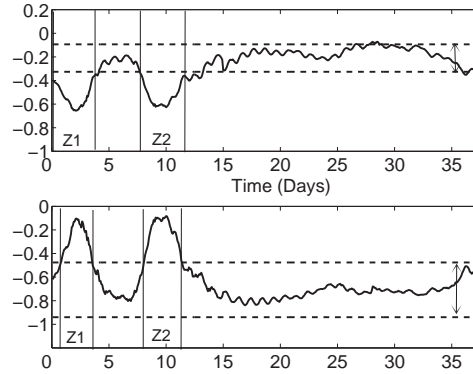
Le résidu final,  $\mathbf{Y}_{residue}^{LS}$ , ne contient pas des informations liées à la réponse du sol. Le sous-espace résidu met alors en évidence des informations liées à d'autres facteurs importants comme les précipitations, les anomalies (fuites), les drains, etc. Les différences entre cette décomposition et celle obtenue en utilisant la SVD sur les données globales sont : 1) le vecteur caractérisant la réponse du sol ne prend pas en compte les phénomènes éphémères comme les précipitations et 2) aucune condition d'orthogonalité n'est imposée sur les variations temporelles de cette source.

### B.1.3 Résultats et discussion

L'algorithme proposé est appliqué sur les données enregistrées sur le site d'Oraison dans la période entre le 09 avril et le 19 mai 2005. et normalisées. Les données contiennent des informations sur les drains, les fuites et les instances de précipitation. Toutes ces informations sont masquées par la réponse du sol dans lequel la fibre optique est enterrée. La normalisation a été faite pour rendre chaque acquisition à une moyenne nulle et une variance unitaire permettant ainsi d'atténuer les effets des variations saisonnières et journalières. Le skewness et le kurtosis pour la période d'analyse sont présentés dans la figure B.2 avec le skewness dans la partie haute et le kurtosis dans la partie basse. Les seuils,  $\pm 2\sigma$ , avec  $\sigma = \sigma_3$  étant la variance d'estimateur de skewness et  $\sigma = \sigma_4$  celle de kurtosis, sont représentés par les pointillés. On constate qu'il y a deux zones (Z1 and Z2), où les statistiques dépassent les seuils respectifs. En

consultant les données météo, nous avons trouvé que ces zones correspondent principalement aux instances de précipitation. L'application du seuil pour enlever ces zones temporelles résulte de la sélection des données pour lesquelles la tendance générale observée avec les statistiques d'ordre supérieur reste plus constante qu'avec les données entières.

Une taille de la fenêtre  $\Delta T = 1$  jour a été sélectionnée avec un recouvrement de 25%. Il faut mentionner



**Fig. B.2:** Statistiques d'ordre supérieur : Skewness (haut) et Kurtosis (bas) avec les lignes pointillés correspondant aux seuils  $\pm 2\sigma$ .

que les résultats ne sont pas très sensibles au choix de la taille de fenêtre et du recouvrement. La première source,  $\mathbf{u}_1^m$ , est alors estimée pour les différentes fenêtres avec  $m = 1, \dots, M$  et  $M = 25$  avec les données actuelles. Le fait que la première source représente une énergie très importante pour chaque fenêtre et que sa forme reste identique, signifie que ces sources sont liées aux facteurs les plus cohérents de données, c'est-à-dire, la réponse du sol. Le vecteur caractéristique,  $\bar{\mathbf{v}}_1$ , est la moyenne de toutes les vecteurs recouverts. Grâce au critère HOS, la source est dépourvue des effets dus aux précipitations. Le sous-espace résidu avec la SVD globale (c'est-à-dire sans utilisation du critère HOS),  $\mathbf{Y}_{residue}^{SVD}$ , est présenté dans la figure B.3. La variation temporelle de la source estimée avec la critère HOS,  $\bar{\mathbf{v}}_1$ , est obtenue avec une approche moindres carrés pour éviter la condition d'orthogonalité imposée par la SVD. Le résidu final,  $\mathbf{Y}_{residue}^{LS}$ , ensuite obtenu est proposé dans la figure B.4.

La suppression de la réponse du sol met en évidence d'autres phénomènes comme les précipitations, les fuites et les drains (figures B.3 et B.4). En comparant les deux résidus, on peut tout de suite constater que le sous-espace résidu obtenu avec la méthode proposée ici est plus net que celui issu d'une approche SVD classique. Cela est surtout visible dans les zones où l'information est parcimonieuse.

Les versions zoomées des deux résidus (figure B.5(a)-(b)) mettent en évidence cette différence.

Un autre défaut connu de l'utilisation d'une SVD globale est la présence d'amplitude compensée. En effet, lors de la construction d'un sous-espace, la méthode a tendance à compenser les zones temporelles plus intenses (plus énergétique) par la présence de zones ayant une intensité opposée, c'est-à-dire une intensité plus basse (figure B.5(a)). Cet effet est lié à la contrainte d'orthogonalité sur les variations temporelles. La méthode proposée ne montre pas ces effets (figure B.5(b)).

La figure B.6 montre les deux résidus sur le jour 21, avec  $\mathbf{Y}_{residue}^{SVD}$  dans la partie haute et  $\mathbf{Y}_{residue}^{LS}$  dans la partie basse. Les deux courbes sont normalisées par leur maxima relatif pour faciliter la comparaison.

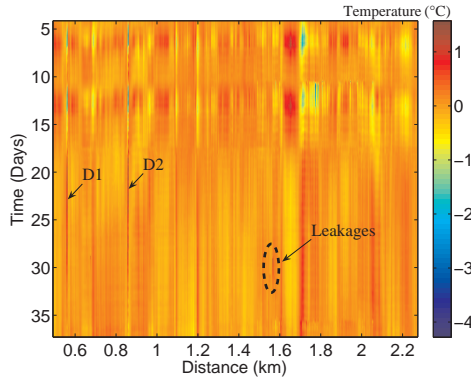


Fig. B.3: Sous-espace résidue  $\mathbf{Y}_{residue}^{SVD}$ , obtenu par SVD.

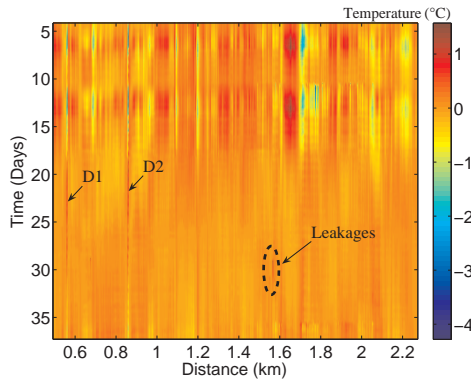
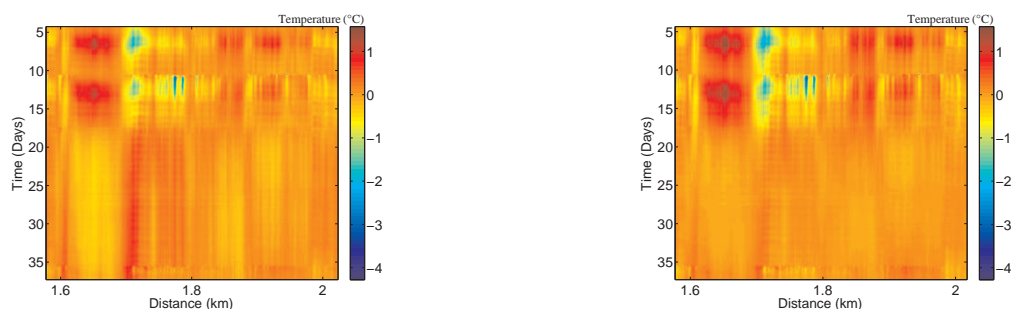


Fig. B.4: Sous-espace résidue  $\mathbf{Y}_{residue}^{LS}$  obtenu par une approche de moindres carrés à l'aide de HOS.

Le résidu issu de la méthode proposée est plus net que celui issu de la SVD. La singularité du terrain présente dans la région, 1.63 – 1.7 (entourée dans la figure) est plus énergique dans le résidu issu de la SVD que dans le résidu HOS-LS. On en conclue que cette singularité est alors mieux estimée dans l'espace des sources par la méthode proposée puisqu'elle est moins présente dans le résidu.

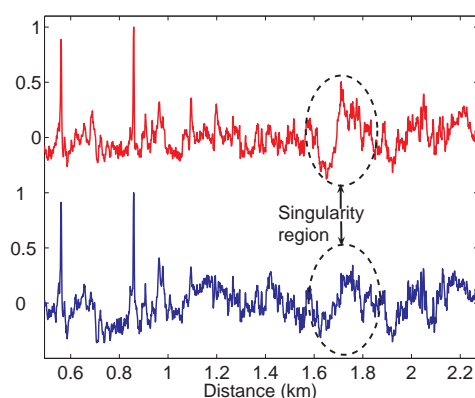
De plus, nous constatons que l'approche proposée (HOS-LS) permet de mieux mettre en évidence les phénomènes intéressants. Par exemple, un zoom des deux résidus du jour 28, proche de la zone de  $L1$  (figure B.7) montre que la fuite  $L1$  a un RSB plus élevé en  $\mathbf{Y}_{residue}^{LS}$  (calculé par rapport au background dans la proximité de  $L1$ ) qu'en  $\mathbf{Y}_{residue}^{SVD}$ . C'est un résultat important du point de vue de la détection des fuites.



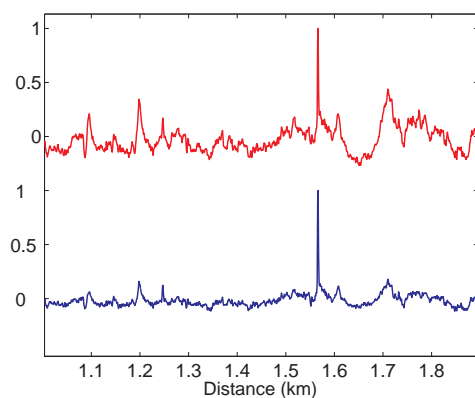
(a) Zoom sur le résidu issu de la SVD.

(b) Zoom sur le résidu issu des HOS.

**Fig. B.5:** Zoom sur le résidu lissée obtenu par la méthode de moindres carrés (droite) et des effets de compensation dus à la SVD (gauche).



**Fig. B.6:** Résidu au jour 21 (en haut  $\mathbf{Y}_{residue}^{SVD}$  et en bas  $\mathbf{Y}_{residue}^{LS}$ ) montrant l'efficacité de la méthode proposée.



**Fig. B.7:** Zoom sur le résidu au jour 28 (en haut  $\mathbf{Y}_{residue}^{SVD}$  et en bas  $\mathbf{Y}_{residue}^{LS}$ ) montrant que la détection de la fuite  $L1$  par la méthode proposée est plus discriminante que celle donnée par la méthode SVD.

# Annexe C

Annexes relatives au chapitre 2

## C.1 Définition des Cumulants

Les cumulants d'une variable aléatoire  $x$ , notés  $\kappa_{x(1)}$ ,  $\kappa_{x(2)}$ , ...,  $\kappa_{x(r)}$ , sont définis par l'identité en  $t$  de l'équation [Kendall 63, pp. 67] :

$$\begin{aligned} \exp\left\{\kappa_{x(1)}(jt) + \kappa_{x(2)}\frac{(jt)^2}{2!} + \dots + \kappa_{x(r)}\frac{(jt)^r}{r!} + \dots\right\} = \\ = 1 + \mu'_{x(1)}\frac{(jt)}{1!} + \dots + \mu'_{x(r)}\frac{(jt)^r}{r!} + \dots = \Phi_x(t) \end{aligned} \quad (\text{C.1})$$

où  $\Phi_x(t)$  est la première fonction caractéristique de  $x$  définie comme :

$$\Phi_x(t) = \int_{-\infty}^{\infty} e^{jtu} f_x(u) du$$

avec  $f_x(u)$  la densité de probabilité de  $x$ . Souvent une deuxième définition est utilisée pour les cumulants d'une variable aléatoire à partir de la seconde fonction caractéristique. Le cumulants d'ordre  $r$  de la variable aléatoire  $x$  est la dérivée d'ordre  $r$  de la seconde fonction caractéristique calculée en  $t = 0$  [Lacoume 90, chap. 1] :

$$\kappa_{x(r)} = (-j)^r \frac{d^r}{dt^r} [\Psi_x(t)]_{t=0} \quad (\text{C.2})$$

où  $\Psi_x(t)$  est la seconde fonction caractéristique de  $x$  connue également sous le nom fonction génératrice des cumulants et définie comme :

$$\Psi_x(t) = \ln(\Phi_x(t)) \quad (\text{C.3})$$

Les cumulants d'ordre  $r$  d'une variable aléatoire sont aussi notés dans la littérature [Lacoume 90] :

$$\kappa_{x(r)} \stackrel{\text{not}}{=} \underbrace{\text{Cum}[x, x, \dots, x]}_{r \text{ fois}} \quad (\text{C.4})$$

où “ $\stackrel{\text{not}}{=}$ ” dénote “par notation”.

## C.2 Position des points de mesure pour les données thermométriques simulées.

Les coordonnées des points de mesure par rapport à la figure 2.1 (page 46) sont affichées dans le tableau C.1. La valeur  $z$  dénote l'élévation par rapport à la base de la digue et la valeur  $y$  dénote la largeur de la digue mesurée par rapport au point extrême situé en bas à gauche. Chaque point pourra correspondre à un emplacement de la fibre optique. Le point de mesure 25 a été choisi pour simuler les zones sans fuites alors que le point 2 pour simuler les fuites.

**Tab. C.1:** Cordonnées des points de mesures.

| Point d'observation | $y$ (m) | $z$ (m) |
|---------------------|---------|---------|
| 1                   | 30,5    | 0       |
| 2                   | 29      | 0,5     |
| 3                   | 29      | 0       |
| 4                   | 27,5    | 1       |
| 5                   | 27,5    | 0,5     |
| 6                   | 27,5    | 0       |
| 7                   | 26      | 1,5     |
| 8                   | 26      | 1       |
| 9                   | 26      | 0,5     |
| 10                  | 26      | 0       |
| 11                  | 24,5    | 2       |
| 12                  | 24,5    | 1,5     |
| 13                  | 24,5    | 1       |
| 14                  | 24,5    | 0,5     |
| 15                  | 24,5    | 0       |
| 16                  | 23      | 2,5     |
| 17                  | 23      | 2       |
| 18                  | 23      | 1,5     |
| 19                  | 23      | 1       |
| 20                  | 21,5    | 3       |
| 21                  | 21,5    | 2,5     |
| 22                  | 21,5    | 2       |
| 23                  | 20      | 3,5     |
| 24                  | 20      | 3       |
| 25                  | 18,5    | 4       |
| 26                  | 17      | 4,5     |
| 27                  | 17      | 4       |
| 28                  | 17      | 3,5     |
| 29                  | 17      | 3       |
| 30                  | 17      | 2,5     |
| 31                  | 17      | 2       |

suite page suivante...

TAB. C.1 – suite

| <b>Point d'observation</b> | <b><math>y</math> (m)</b> | <b><math>z</math> (m)</b> |
|----------------------------|---------------------------|---------------------------|
| 32                         | 17                        | 1,5                       |
| 33                         | 17                        | 1                         |
| 34                         | 17                        | 0,5                       |
| 35                         | 17                        | 0                         |
| 36                         | 14                        | 4,5                       |
| 37                         | 9,5                       | 3                         |
| 38                         | 3,5                       | 1                         |





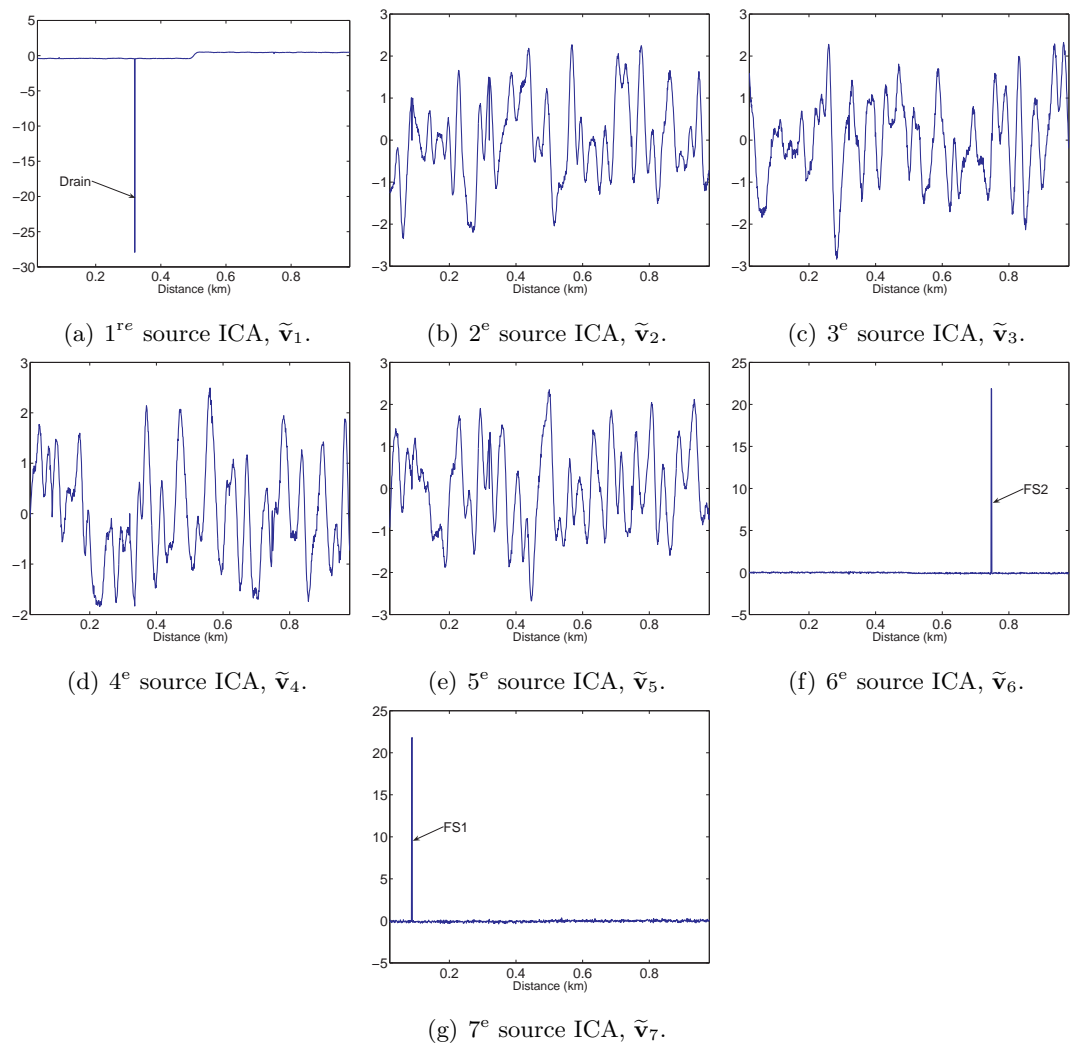
# Annexe D

Annexes relatives au chapitre 3

## D.1 Analyse des données simulées

### D.1.1 Analyse sans fenêtrage glissant en temps des données simulées avec des fuites d'un jour.

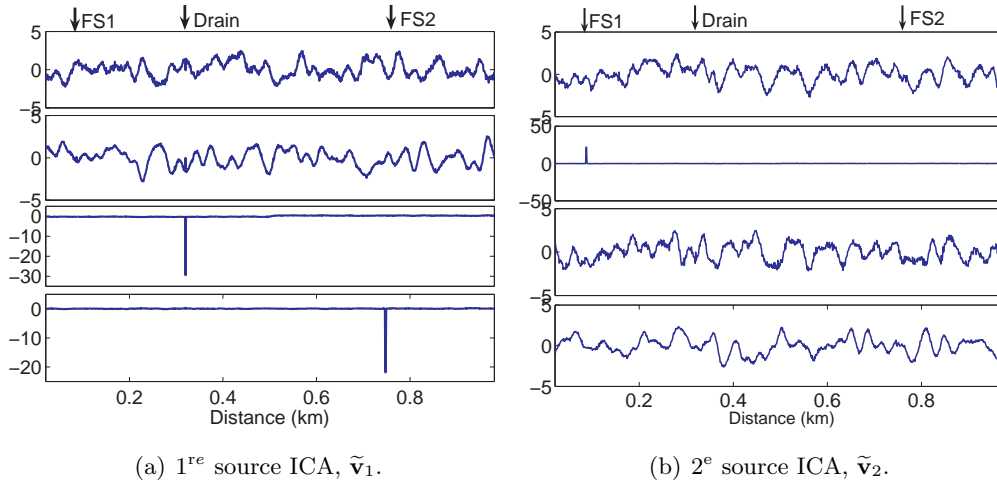
Nous avons vu (figure 3.6 à la page 104) que sur les 6 sources estimées par l'ICA sur le résidu de SVD  $\mathbf{Y}_{résidu}$ , la fuite *FS1* n'est pas identifiable. La figure D.1 montre les sources ICA pour  $m = 1$  et  $p = 7$  dans l'algorithme SVD-ICA. Augmentant le nombre de sources à estimer par l'ICA, la fuite *FS1* est identifiée sur les sources si on pose au moins  $p = 7$  sources à estimer par ICA. Les sources sont ordonnées par rapport à la décroissance des valeurs singulières modifiées  $\tilde{\sigma}_j$ . L'énergie de la fuite *FS1* reste alors faible par rapport les données entières.



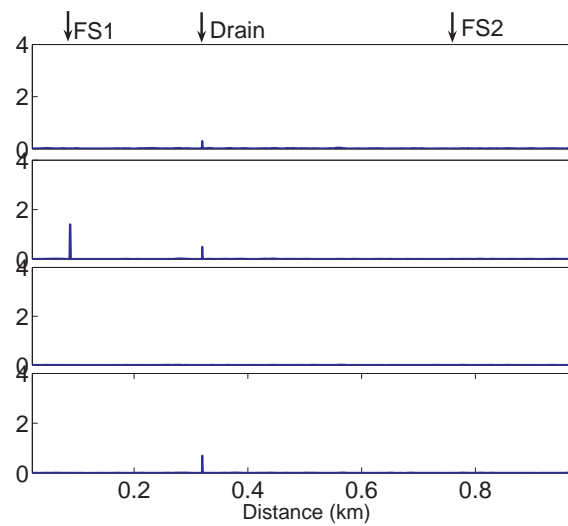
**Fig. D.1:** Sources obtenues dans la deuxième étape de l'algorithme en appliquant l'ICA pour estimer 7 sources sur le résidu  $\mathbf{Y}_{résidu}$  sans fenêtrage glissant en temps.

### D.1.2 Analyse des fuites simulées d'une durée d'un jour par fenêtrage glissant de 15 jours

Des tests avec différentes tailles de fenêtre d'analyse par SVD-ICA ont été faits. Nous montrons ici un exemple d'analyse des fuites simulées d'une durée d'un jour par une fenêtre de taille 15 jours. La première source SVD estime la réponse du sol. ICA est appliquée sur le résidu  $\mathbf{Y}_{résidu}$  pour estimer 4 sources, les sources étant ordonnées en termes de décroissance des valeurs singulières modifiées. Les deux premières sources sont présentées sur la figure D.2. La fuite  $FS2$  étant relativement importante dans cette fenêtre d'analyse sort sur la 1<sup>re</sup> source ICA. Le résidu  $\mathbf{Z}_{résidu}$  avec  $i_2 = 1$  ainsi obtenu ne détecte pas la fuite  $FS2$  (figure D.3). Il faut rappeler ici que la fuite  $FS2$  a un débit assez important, et ce dernier est également un facteur important dans l'énergie des valeurs singulières modifiées.



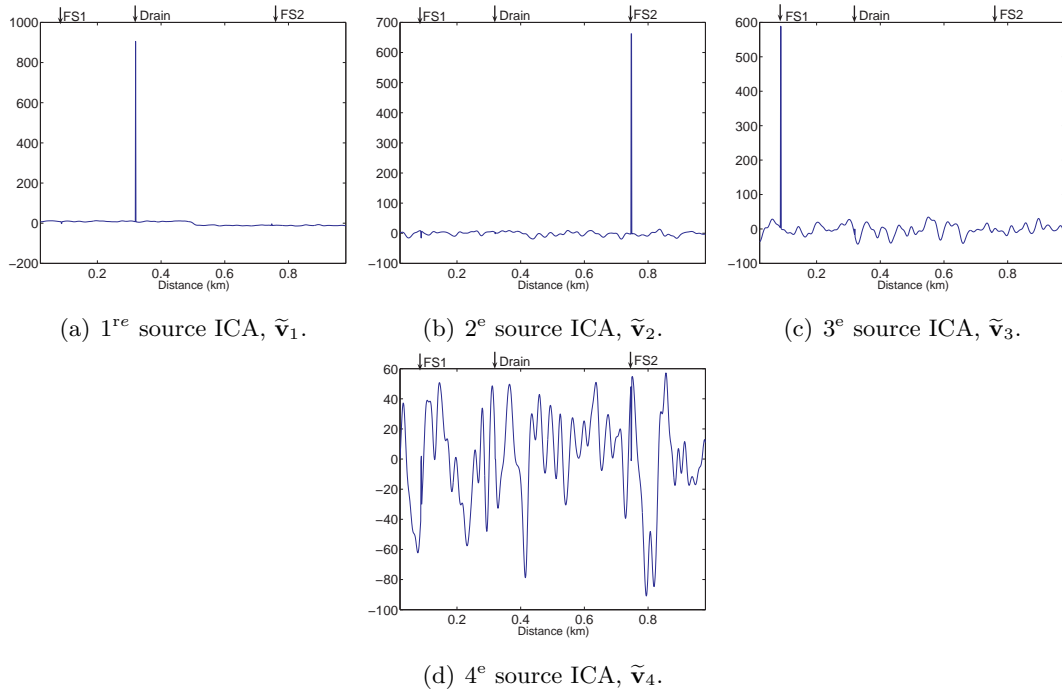
**Fig. D.2:** Les 2 premières sources ICA pour 4 différentes zones temporelles. De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . Le fuite  $FS2$  apparaît sur la 1<sup>re</sup> source ICA, la fuite  $FS2$  continue d'apparaître sur la deuxième.



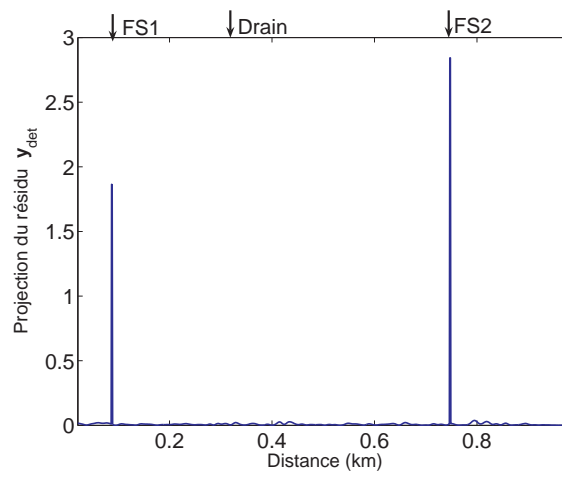
**Fig. D.3:** Les projections,  $y_{det}(x)$ , du résidu  $\mathbf{Z}_{résidu}$  sur l'axe de distances pour les 4 différentes zones temporelles.  $N_p = 15 * 12$ . De haut en bas : zone au début de données sans fuites ; zone contenant la fuite  $FS1$  ; zone sans fuites ; zone contenant la fuite  $FS2$ . La fuite  $FS1$  est très bien détectée mais  $FS2$  n'est pas détectée à cause de sa présence dans le 2<sup>e</sup> sous espace résidu.

### D.1.3 Étude par rapport à la durée des fuites : Exemple de fuites simulées d'une durée de 5 jours

Pour étudier l'influence de la durée de fuite, nous présentons ici un exemple des deux fuites simulées avec une durée de 5 jours. Les données sont analysées par un fenêtrage glissant de 100 jours pour détecter ces fuites dans le résidu  $\mathbf{Z}_{résidu}$ . Les 4 sources estimées en appliquant l'ICA sur le sous-espace résidu SVD sont présentées dans la figure D.4. La fuite  $FS2$  est estimée par la 2<sup>e</sup> source ICA et la fuite  $FS1$  par la 3<sup>e</sup> source. Le résidu obtenu avec  $i_2 = 1$  permet alors une bonne séparation des deux fuites D.5.



**Fig. D.4:** Les 4 sources ICA estimées pour les fuites  $FS1$  et  $FS2$  avec une durée de 5 jours, analysées par un fenêtrage glissant de 100 jours. Les fuites  $FS2$  et  $FS1$  apparaissent respectivement sur les 2<sup>e</sup> et 3<sup>e</sup> sources.

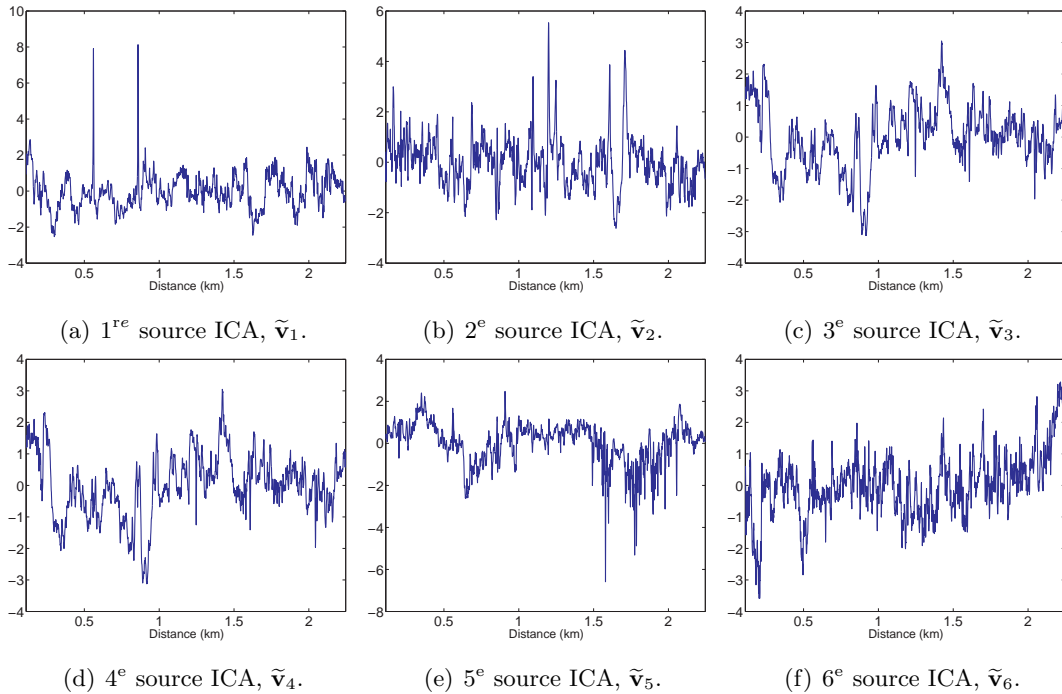


**Fig. D.5:** La projection,  $y_{det}(x)$ , du résidu  $\mathbf{Z}_{résidu}$  sur l'axe de distances pour une zone temporelle.  $N_p = 100 * 12$ . Les fuites  $FS1$  et  $FS2$  sont très bien séparées.

## D.2 Analyse des données d'Oraison

### D.2.1 Sources ICA pour $m = 1$ , $p = 6$ , pour l'analyse globale des données enregistrées sur le site d'Oraison

Nous avons vu qu'une analyse des données entières enregistrées sur le site d'Oraison ne permet la localisation des fuites artificielles,  $L1$ ,  $L2$  et  $L3$  sur les sources (figure 3.16 à la page 115). Même en posant plusieurs sources à estimer par l'ICA, nous n'arrivons pas à localiser les fuites artificielles,  $L1$ ,  $L2$  et  $L3$  sur les sources à cause de leur faible énergie par rapport aux données entières. La figure D.6 montre un exemple de l'estimation de 6 sources par ICA sur le résidu  $\mathbf{Y}_{résidu}$  mais les fuites artificielles ne sont toujours pas identifiées sur les sources.



**Fig. D.6:** Sources obtenues dans la deuxième étape de l'algorithme en appliquant l'ICA pour estimer 6 sources sur le résidu  $\mathbf{Y}_{résidu}$  sans fenêtrage glissant en temps. Les fuites artificielles à cause de leur faible énergie ne sont toujours pas identifiées sur ces sources.



### D.2.2 Résultats de l’algorithme de séparation de sources pour les données d’Oraison avec différents prétraitements.

Le tableau de la figure D.7 présente un résumé des résultats obtenus en l’appliquant l’algorithme SVD-ICA avec différents prétraitements et dans différentes orientations des données d’Oraison. Cette comparaison ici est faite sur le résidu obtenue en termes des fuites artificielles. Après avoir fait la normalisation, nous pouvons appliquer l’opérateur de gradient spatial. Le gradient spatial consiste à prendre le gradient pour chaque profil sur toute la longueur de la fibre. Le gradient spatial atténue les phénomènes variant lentement en distance comme la réponse du sol ainsi que les précipitations. En appliquant ce prétraitement, la première étape de SVD dans l’algorithme SVD-ICA pourra être évitée. Dans un deuxième temps, nous avons essayé également d’appliquer l’algorithme SVD-ICA sur les données transposées. Dans cette orientation les sources estimées seront en fonction du temps. Les résultats ont été analysés en termes de localisation spatiale et temporelle des fuites artificielles ainsi qu’en termes de puissance relative des fuites artificielles et des drains dans le résidu  $\mathbf{Z}_{résidu}$ . Il a été retrouvée que les données traitées dans leur orientation originale (les sources SVD et ICA en fonction de la distance) et le prétraitement de gradient spatial donne des meilleurs résultats.

|    | Without Gradient     |           |       |            |           |     | Spatial Gradient     |           |           |            |           |           |
|----|----------------------|-----------|-------|------------|-----------|-----|----------------------|-----------|-----------|------------|-----------|-----------|
|    | Original Orientation |           |       | Transposed |           |     | Original Orientation |           |           | Transposed |           |           |
|    | L.T                  | L.D       | R.S   | L.T        | L.D       | R.S | L.T                  | L.D       | R.S       | L.T        | L.D       | R.S       |
| L1 | Very Good            | Very Good | Worst | Worst      | Bad       | Bad | Very Good            | Very Good | Very Good | Bad        | Good      | Good      |
| L2 | Very Good            | Very Good | Worst | Worst      | Bad       | Bad | Very Good            | Very Good | Very Good | Bad        | Good      | Good      |
| L3 | Very Good            | Very Good | Worst | Worst      | Bad       | Bad | Very Good            | Very Good | Very Good | Bad        | Good      | Good      |
| D1 | N/A                  | Very Good | Good  | N/A        | Very Good | Bad | N/A                  | Very Good | Bad       | N/A        | Very Good | Very Good |
| D2 | N/A                  | Very Good | Good  | N/A        | Very Good | Bad | N/A                  | Very Good | Bad       | N/A        | Very Good | Very Good |

Very Good    Good  
Bad    Worst

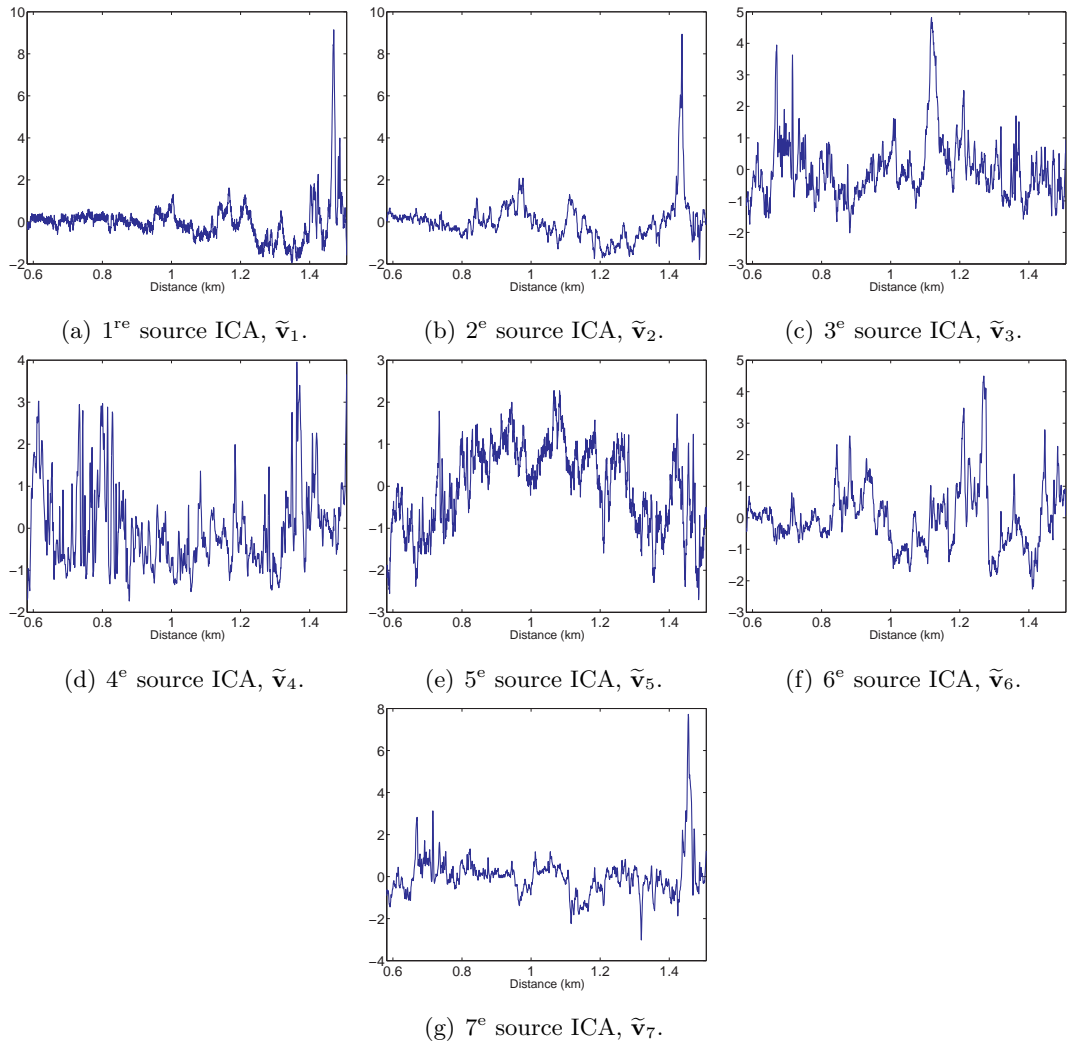
L.T = localization in time;  
L.D = localization in distance; R.S = Relative strength

Fig. D.7: Sommaire général des résultats d’application d’algorithme SVD-ICA pour les données Oraison en utilisant différents prétraitements et différentes orientations des données.

## D.3 Analyse des données de Kembs

### D.3.1 Sources ICA pour l'analyse globale avec $m = 0$ , $p = 7$ .

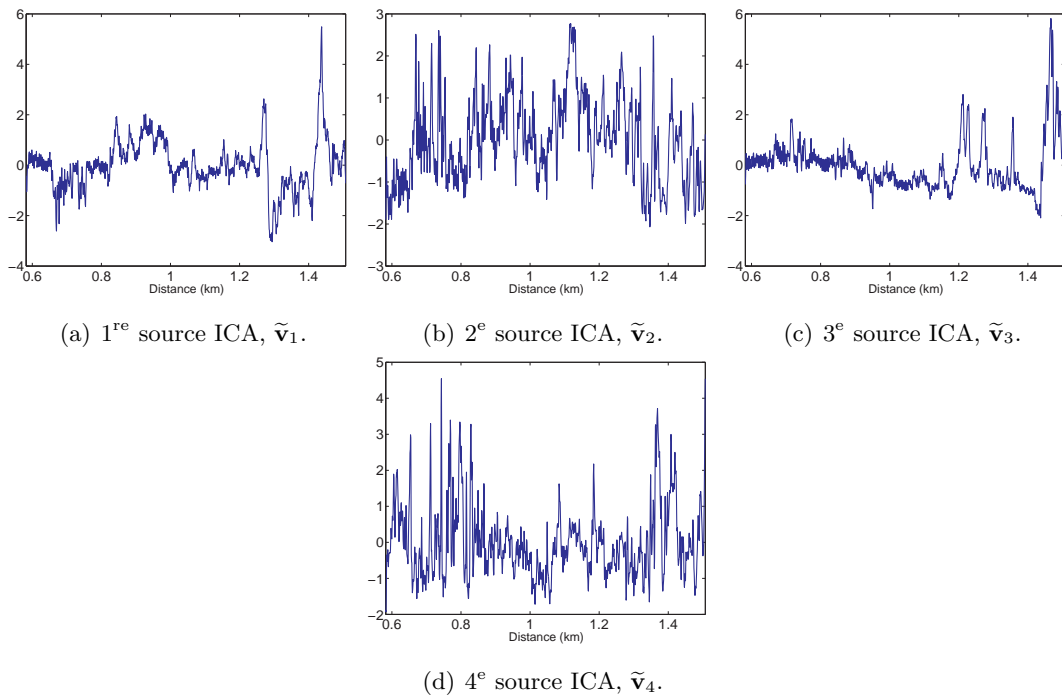
Les 7 sources ICA obtenues en appliquant l'algorithme SVD-ICA avec  $m = 0$  et  $p = 7$  sans fenêtrage glissant en temps sont montrées dans la figure D.8. Comme remarqué lors de l'analyse avec 4 sources ICA (figure 3.26 à la page 125), les différents phénomènes ne sont pas très bien séparables même si on estime plus de sources avec l'ICA.



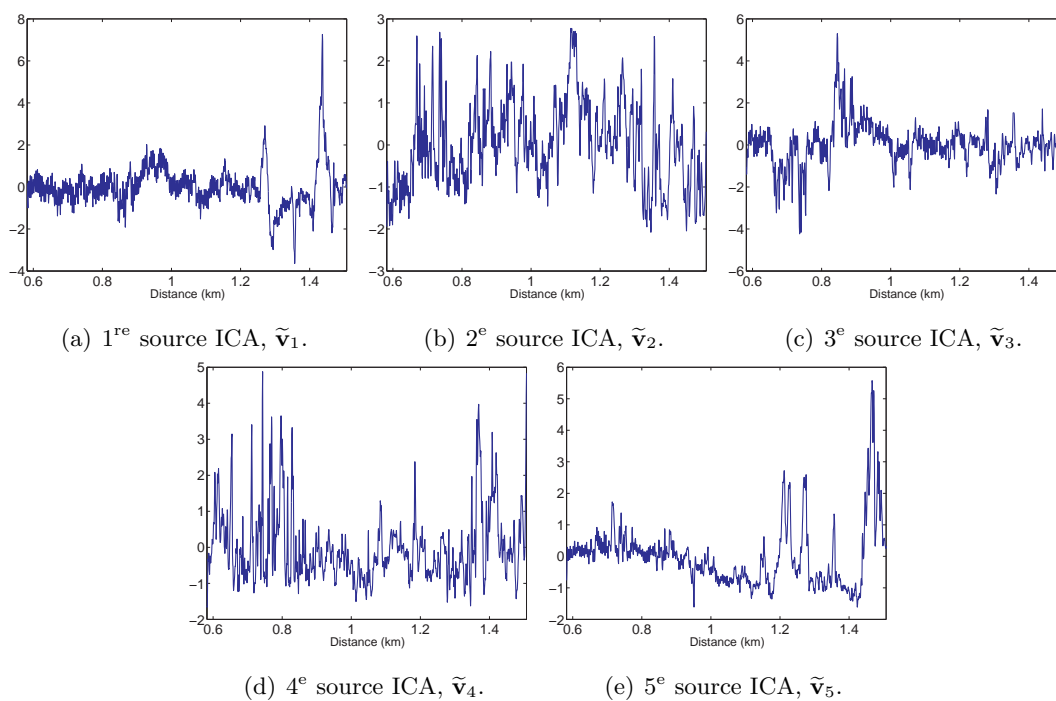
**Fig. D.8:** Sources obtenues dans la deuxième étape de l'algorithme en appliquant l'ICA pour estimer 7 sources sur le résidu  $\mathbf{Y}_{résidu}$  sans fenêtrage glissant en temps.

### D.3.2 Sources ICA pour l'analyse par fenêtrage glissant de 14 jours pour la période entre 29 avril et 13 mai 2007.

Nous présentons ici les sources ICA obtenues en posant  $p = 4$  et  $p = 5$  pour les données de Kembs analysées par fenêtrage glissant de 14 jours pour la période entre 29 avril et 13 mai 2007. Les 4 sources ICA avec  $m = 0$  et  $p = 4$  sont montrées dans la figure D.9 ainsi que les 5 sources obtenues avec  $m = 0$  et  $p = 5$  dans la figure D.10. Par rapport à l'information qui a été récupérée en posant  $p = 3$  (cf. figure 3.27 à la page 126), aucune information complémentaire n'est pas obtenue en augmentant le nombre de sources à estimer.



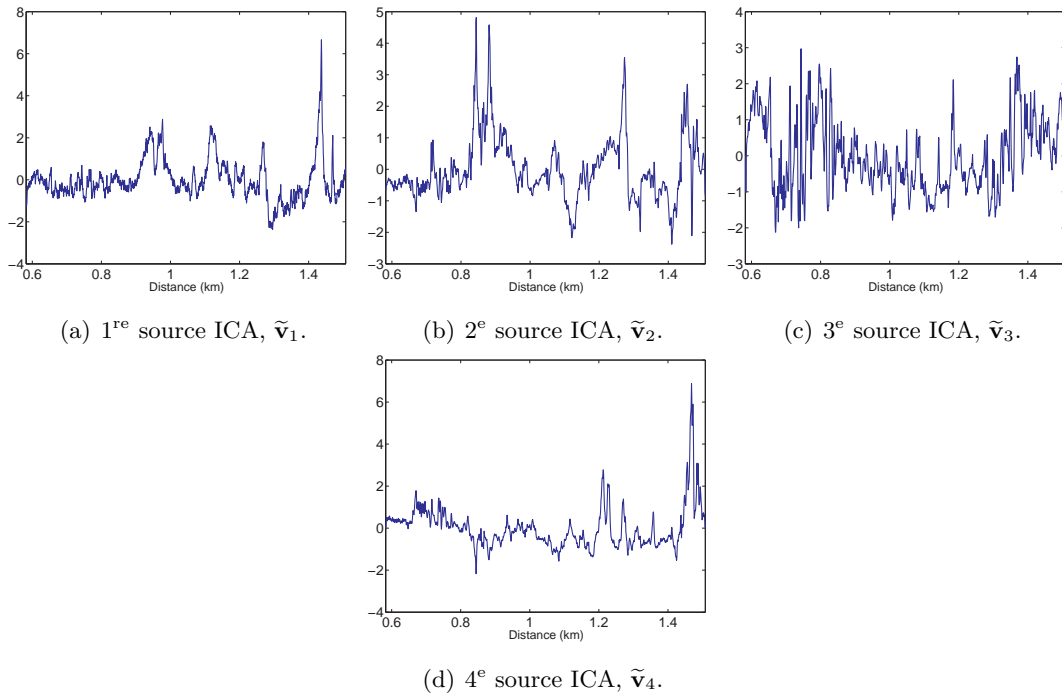
**Fig. D.9:** Sources obtenues par l'application de l'ICA (pour estimer  $p = 4$  sources) pour les données de Kembs avec un fenêtrage glissant en temps de 14 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 2<sup>e</sup> fenêtre temporelle (du 29 avril au 13 mai 2007).



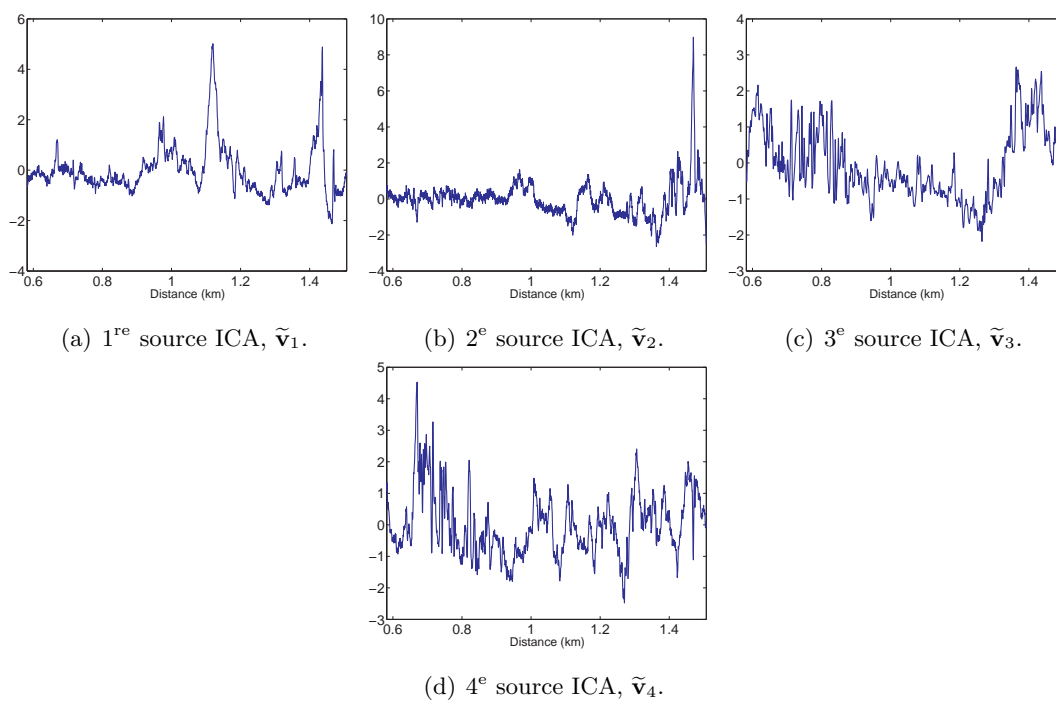
**Fig. D.10:** Sources obtenues par l'application de l'ICA (pour estimer  $p = 5$  sources) pour les données de Kembs avec un fenêtrage glissant en temps de 14 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 2<sup>e</sup> fenêtre temporelle (du 29 avril au 13 mai 2007).

### D.3.3 Analyse avec un fenêtrage glissant de 40 jours.

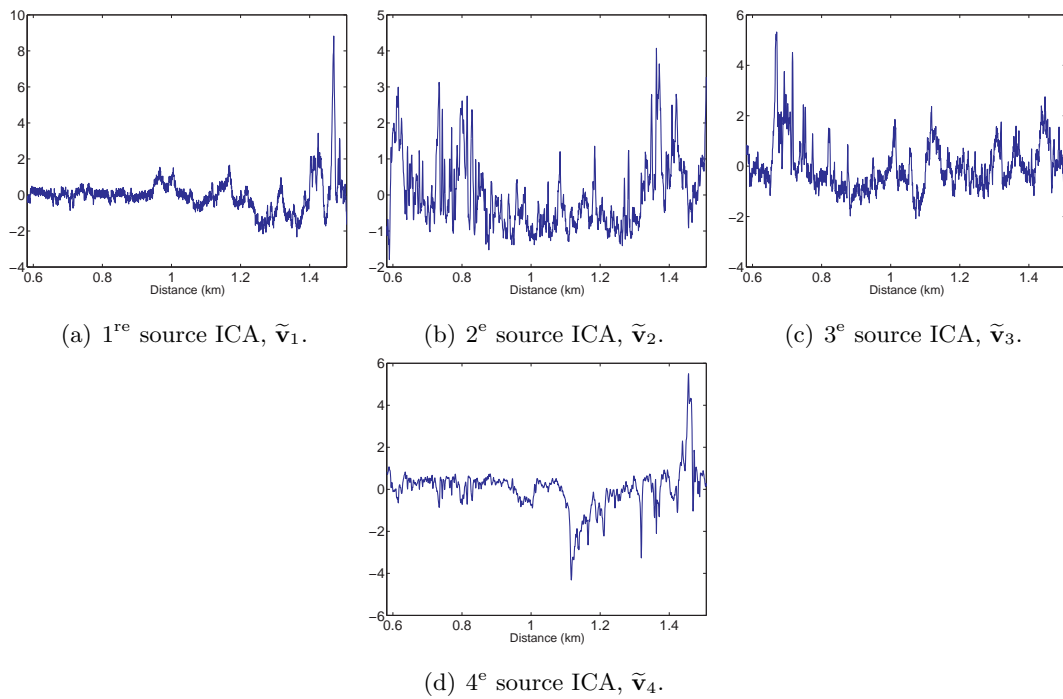
A ce stade, nous ne sommes pas en mesure à donner une valeur empirique pour la taille de la fenêtre d'analyse pour les données de Kembs. Un exemple d'analyse par fenêtrage glissant de 40 jours est présenté ici. Les figures D.11 à D.13 montrent les 4 sources ICA obtenues en appliquant l'algorithme SVD-ICA avec  $m = 0$  et  $p = 4$  pour différentes périodes d'analyse de 40 jours entre 15 avril et 13 août 2007. Les différentes anomalies énumérées dans le tableau 3.2 (page 128) sont identifiées à nouveau pendant les différentes périodes d'analyse. En plus, il existe des anomalies qui sont identifiées sur une fenêtre d'analyse mais pas pour les autres fenêtres d'analyse. Par exemple, les anomalies autour de 0.844 km et 0.882 km existent sur la première période d'analyse entre 15 avril au 25 mai (cf. figure D.11(b)) mais pas sur les autres périodes d'analyse (cf. figures D.12 et D.13).



**Fig. D.11:** Sources obtenues par l'application de l'ICA (pour estimer  $p = 4$  sources) pour les données de Kembs avec un fenêtrage glissant en temps de 40 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 1<sup>e</sup> fenêtre temporelle (du 15 avril au 25 mai 2007).



**Fig. D.12:** Sources obtenues par l'application de l'ICA (pour estimer  $p = 4$  sources) pour les données de Kembs avec un fenêtrage glissant en temps de 40 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 2<sup>e</sup> fenêtre temporelle (du 25 mai au 04 juillet 2007).



**Fig. D.13:** Sources obtenues par l'application de l'ICA (pour estimer  $p = 4$  sources) pour les données de Kembs avec un fenêtrage glissant en temps de 40 jours. Les sources montrées ici sont celles qui ont été obtenues dans la 3<sup>e</sup> fenêtre temporelle (du 04 juillet au 13 août 2007).

# Annexe E

## Annexe relative au chapitre 4

### E.1 Analyse des données du contre-canal de Kembs

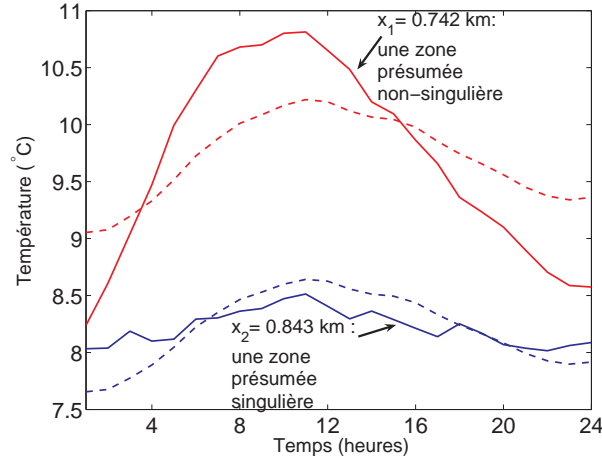
Les hypothèses faites pour la méthode discutée dans le chapitre 4 (section 4.2) ne sont pas valables pour les données enregistrées sur le site de Kembs. Premièrement, comme nous avons vu, Kembs est un cas extrême avec plusieurs anomalies. L'hypothèse d'avoir beaucoup plus de zones singulières que des zones singulières n'est pas nécessairement justifiable ici. Deuxièmement, le fait que la fibre de contre canal est trop proche de l'eau, parfois immergée, ne permettra pas une bonne distinction entre les singularités et les non-singularités avec la définition de dissimilarité proposée dans la section 4.2.

Nous présentons ici quelques résultats d'analyse des données de Kembs avec la méthodologie du chapitre 4 sans pourtant d'être capables à conclure sur la validité des résultats.

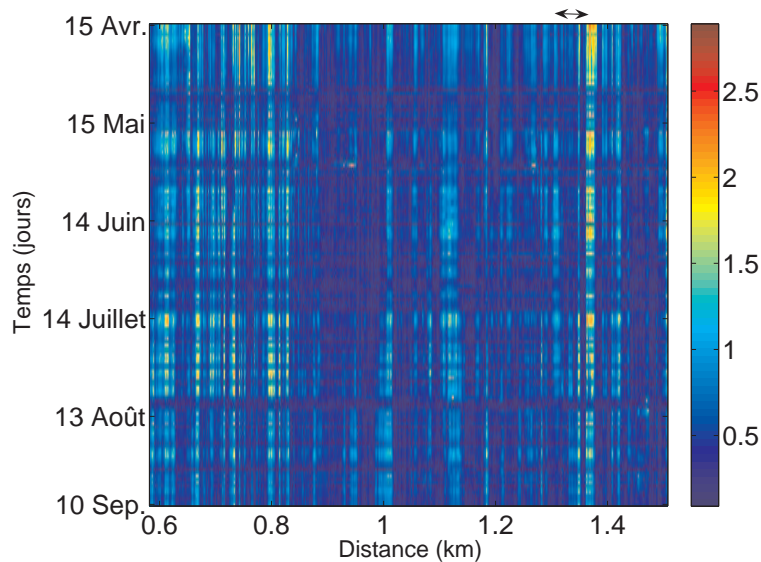
Lors de l'analyse des données de Kembs dans le chapitre 3, nous avons identifié une zone d'anomalies autour de  $x_2 = 0.843$  km. Nous considérons cette zone comme un exemple de singularité et, par le même raisonnement, la zone  $x_1 = 0.742$  km comme un exemple de zone non-singulière. Les profils journaliers de température pour ces deux zones sont montrés dans la figure E.1 avec des lignes continues. Les sous-espaces signaux correspondant à ces zones sont également affichés en lignes pointillées. Nous constatons que pour aucune de ces deux zones, le vecteur de référence n'est pas identique au profil journalier correspondant. Au contraire de ce qui était prévu, la zone présumée singulière présente moins de déviations par rapport à la référence que la zone non-singulière. Ce qui nous amène à conclure que le vecteur de référence étant plus proche des signatures de zones singulières, le site présente plus de zones singulières que des zones non-singulières. La mesure de dissimilarité alors ne fera pas ressortir les zones singulières. En revanche, selon cette interprétation, une mesure de similarité caractérisera mieux les singularités. La mesure de dissimilarité obtenue pour toute la période d'analyse est présentée dans la figure E.2 en fonction de la distance (m) et du temps (jours). Nous pouvons constater que plusieurs zones sont mise en évidence, plus particulièrement une zone marquée par la double flèche apparaît quasiment toujours. Ceci est mis en évidence aussi si nous affichons la dissimilarité  $d^i(x)$  pour 4 jours non-consécutives choisis arbitrairement (figure E.3). Nous ne sommes pas encore capables d'associer une notion physique à cette zone comme pour les autres anomalies existantes sur le site de Kembs. Il est intéressant de noter que sur la deuxième fenêtre de cette figure, nous récupérons l'anomalie autour de 0.843 km (cf. figure E.3 marquée par flèche), alors que dans les autres fenêtres, nous la voyons pas.

Les résultats obtenus sont alors inexploitable et l'étude pour le site de Kembs reste incomplète.

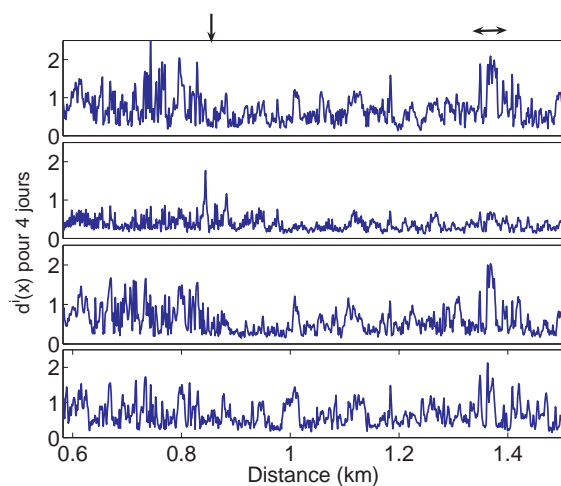




**Fig. E.1:** *Lignes continues* : Profils de température enregistrés sur 24 heures à 2 distances différentes pour le contre-canal sur le site de Kembs. La zone présumée non-singulière est  $x_1$  (0.742 km) et la zone présumée singulière est  $x_2$  (0.843 km). *Lignes pointillées* :  $\mathbf{Y}_{sig}^i$  le même jour et aux mêmes distances. Il est difficile d'établir une distinction très précise (en termes des écarts par rapport au vecteur de référence) entre la zone singulière et la zones non-singulière.



**Fig. E.2:** Mesure de dissimilarité  $d^i(x)$  pour les données du contre-canal de Kembs acquises entre le 15 avril et le 10 septembre 2007. Plusieurs zones sont mises en évidence, les résultats étant inexploitable faute d'un manque d'expertise effectué le même temps sur le site.



**Fig. E.3:** La mesure de dissimilarité  $d^i(x)$  pour 4 jours. De haut en bas : 23 avril, 14 mai, 13 juin et 12 août. Plusieurs zones sont mises en évidence. Notamment la zone marquée par la double flèche apparaît toujours. La flèche simple marque la distance 0.843 km, la zone présumée de singularité.



# Bibliographie

- [Akaike 73] H. Akaike. *Information theory and an extension of the maximum likelihood principle*. In B. N. Petrov & F. Csaki, éditeurs, Second International Symposium on Information Theory, pages 267–281, 1973.
- [Armbruster 89] H. Armbruster, J. Brauns, W. Mazur & G. P. Merkler. *Effect of leaks in dams and trials to detect leakages by geophysical means*. In Merkler GP et al., éditeur, Lecture Notes in Earth Sciences : Detection of subsurface flow phenomena. Springer-Verlag, Berlin Heidelberg, 1989.
- [Astola 90] J. Astola, P. Haavisto & Y. Neuvo. *Vector median filters*. Proc. IEEE, vol. 78, pages 678 – 689, April 1990.
- [Bao 94] X. Bao, D. J. Webb & D. A. Jackson. *Combined distributed temperature and strain sensor based on Brillouin loss in an optical fiber*. Optics Letters, vol. 19, no. 2, pages 141–143, 1994.
- [Barnoski 78] M. K. Barnoski & S. M. Jansen. *Fiber waveguides : A novel Technique for Investigating of Attenuation Characteristics*. Applied Optics, vol. 16, pages 2112–2115, 1978.
- [Bell 95] A. J. Bell & T. J. Sejnowski. *An information-maximisation approach to blind separation and blind deconvolution*. Neural Computation, vol. 7, pages 1129–1159, 1995.
- [Bihan 06] N. Le Bihan, V. Vrabie & J. I. Mars. *Multi-dimensional seismic data decomposition by High Order SVD and unimodal ICA*. In C. H. Chen, éditeur, Signal and Image Processing for Remote sensing. Taylor and Francis, 2006.
- [Bièvre 05] G. Bièvre & C. Norgeot. *Utilisation des méthodes géophysiques pour l’auscultation des digues en eau : Etude de cas sur le canal du Centre (France)*. Bulletin des laboratoires des ponts et chaussées ISSN 1269-1496, vol. 254, pages 85–107, 2005.
- [Bolève 09] A. Bolève. *Etude des zones de fuite par la méthode du potentiel spontané dans les digues et les barrages. Application sur ouvrages hydrauliques réels*. PhD thesis, Université de Savoie, France, 2009.
- [Cardoso 89] J. F. Cardoso. *Source separation using higher order moments*. In ICASSP, pages 2109–2112, 1989.
- [Cardoso 93] J. F. Cardoso & A. Souloumiac. *Blind beamforming for non gaussian signals*. IEE Proceedings-F, vol. 140, no. 6, pages 362–370, 1993.

- [Cardoso 98] J. F. Cardoso. *Blind signal separation : statistical principles*. Proc. of IEEE, vol. 86, no. 10, pages 2009–2025, 1998.
- [Cardoso 99] J. F. Cardoso. *High-order contrasts for independent component analysis*. Signal Processing, vol. 11, pages 157–192, 1999.
- [Cemagref 05] Cemagref. Sécurité des digues fluviales et de navigation. Cemagref, 2005.
- [Charlot 96] F. Charlot, B. Falourd, N. Vivier & T. Xouillot. *La Loire. Histoire des protections contre les crues*. Rapport technique, Direction Régionale de l'Équipement du Centre, 1996.
- [Common 07] P. Common & C. Jutten. Séparation de sources 1 : concepts de base et analyse en composantes indépendantes. Lavoisier, 2007.
- [Comon 94] P. Comon. *Independent component analysis, A new concept ?* Signal Processing, vol. 36, no. 3, pages 287–314, 1994.
- [Corwin 90] R. F. Corwin. *Geotechnical and environmental geophysics*. In H. S. Ward, editeur, The self-potential method for environmental and engineering applications. Soc of Exploration Geophysicists, 1990.
- [CSB 05] CSB. *Dispositif d'auscultation des barrages - Appareils et méthodes de mesures*. Rapport technique, Swiss Committee on Dams, 2005.
- [Cullagh 87] P.Mc. Cullagh. Tensor methods in statistics. Chapman and Hall Ltd., 1987.
- [Dahlin 93] T. Dahlin. *On the automation of 2D resistivity surveying for engineering and environmental applications*. PhD thesis, Lund University, 1993.
- [Dahlin 01] T. Dahlin. *The development of DC resistivity imaging techniques*. Computers and Geosciences, vol. 27, pages 1019–1021, 2001.
- [Dakin 85] J. P. Dakin, D. Pratt & G. W. Bibby. *Distributed optical fiber Raman temperature sensor using a semiconductor light source and detector*. Electron. Lett., vol. 21, pages 569–570, 1985.
- [Dakin 90] J. P. Dakin. The Distributed Fiber Optic Sensing Handbook. Wiley, 1990.
- [DeCusatis 06] Casimer DeCusatis & Carolyn DeCusatis. Fiber optic essentials. Elsevier, 2006.
- [Delfosse 95] N. Delfosse & P. Loubaton. *Adaptive blind separation of independent sources : a deflation approach*. Signal Processing, vol. 45, pages 59–83, 1995.
- [Dempster 77] A. P. Dempster, N. M. Laird & D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. J. Roy. Stat. Soc. B, vol. 39, pages 1–38, 1977.
- [Deville 07] Y. Deville. *Applications*. In P. Common & C. Jutten, editeurs, Séparation de sources 2 : au-delà de l'aveugle et applications, pages 519–560. Lavoisier, 2007.
- [Diamantaras 96] K. I. Diamantaras & S. Y. Kung. Principal component neural networks : Theory and applications. Wiley, 1996.

- 
- [Diersch 05] H. J. G. Diersch & O. Kolditz. *WASY Software FEFLOW : Finite element subsurface flow and transportation simulation system*. In White Papers - vol. II WASY Institute for Water Resources Planning and Systems Research, 2005.
- [Dornstädter 97] J. Dornstädter. *Detection of Internal Erosion in Embankment Dams*. In ICOLD 19th Congress, Florence, 1997.
- [Fauchard 04] C. Fauchard & P. Mériaux. *Méthodes géophysiques et géotechniques pour le diagnostic des digues de protection contre les crues, guide pour la mise en oeuvre et l'interprétation*. Cemagref éditions, 2004.
- [Fell 07] R. Fell & J. J. Fry. *Internal erosion of dams and their foundations*. Taylor and Francis, 2007.
- [Ferdinand 90] P. Ferdinand, Y. Denayrolles, C. Mersier, J. Plantey, N. Recrosio, M. Pays & D. Vielpeau. *The potential for distributed sensors and optical fibre sensor networks in the electric power industry*. Measurement Science and Technology, vol. 1, pages 908–916, 1990.
- [Fernandez 05] A. F. Fernandez, P. Rodeghiero & B. Brichard. *Radiation-tolerant Raman distributed temperature monitoring system for large nuclear infrastructures*. IEEE Trans. Nucl. Sci., vol. 52, no. 6, pages 2689–2694, December 2005.
- [Fry 97] J. J. Fry. *Internal erosion and surveillance*. In ICOLD 19th Congress, Florence, 1997.
- [Fry 04] J. J. Fry. *Détection de fuite sur les digues par acquisition de profils de température le long d'une fibre optique*. In Sécurité des digues fluviales et de navigation, Actes de colloque du CFGB, Orléans, France, 2004.
- [Gaeta 90] M. Gaeta & J.-L. Lacoume. *Source separation without prior knowledge : the maximum likelihood solution*. In EUSIPCO, pages 621–624, 1990.
- [Golub 89] G. H. Golub & C. F. Van Loan. *Matrix computation*. John Hopkins, 1989.
- [Grosswig 01] St. Grosswig, A. Graupner & 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, pages 9–17, 19–21, June 2001.
- [Grosswig 05] S. Grosswig, E. Hurtig, S. Luebbecke & B. Vogel. *Pipeline leakage detection using distributed fibre optical temperature sensing*. In 17th International Conference on Optical Fibre Sensors, pages 9–17, Mexico, 2005.
- [Guidoux 08] C. Guidoux. *Développement et validation d'un système de détection et de localisation par fibre optiques de zones de fuites dans les digues en terre*. PhD thesis, Université Joseph Fourier, France, 2008.
- [Hartog 83] A. H. Hartog. *A distributed temperature sensor based on liquid-core optical fibers*. IEEE Journal of Lightwave Tech., vol. 1, no. 3, pages 498–509, Sept. 1983.

- [Hartog 00] A. H. Hartog. *Distributed fiber-optic temperature sensors : principles and applications*. In K. T. Grattan & B. T. Meggitt, editeurs, *Optical Fiber Sensor Technology*, pages 241–301. Kluwer, 2000.
- [Hartog 02] A. H. Hartog. *Progress in distributed fiber-optic temperature sensing*. In M. A. Marcus & B. Culshaw, editeurs, *Proc. SPIE, Fiber Optic Sensor Technology and Applications*, 2001, volume 4578 of *0277-786X/02*, pages 43–52, 2002.
- [Hastie 01] T. Hastie, R. Tibshirani & J. H. Friedman. *The elements of statistical learning : data mining, inference, and prediction*. Springer, 2001.
- [Heinz 01] D. C. Heinz & C.-I Chang. *Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery*. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 39, no. 3, pages 529–545, March 2001.
- [Henault 06] J. M. Henault. *Projet PAREOT-Évaluation métrologique de l'appareil de mesure de température répartie le long d'une fibre optique (DTS) Sensornet Sentinel*. Edf technical report, EDF, 2006.
- [Henderson 97] P. J. Henderson, N. E. Fischer & D. A. Jackson. *Current metering using fiber grating based interrogation of a conventional current transformer*. In *Proc. 12th Intl. Conf. Optical Fiber Sensors*, pages 186–189, Williamsburg, VA, 1997.
- [Hyvärinen 99a] A. Hyvärinen. *Fast and robust fixed-point algorithms for independent component analysis*. *IEEE Transactions on Neural Networks*, vol. 10, pages 626–634, 1999.
- [Hyvärinen 99b] A. Hyvärinen. *Survey on independent component analysis*. *Neural Computing Surveys*, vol. 2, pages 94–128, 1999.
- [Hyvärinen 99c] A. Hyvärinen, J. Särelä & R. Vigario. *Spikes and bumps : Artefacts generated by independent component analysis with insufficient sample size*. In *Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 425–420, Aussois, France, 1999.
- [Hyvärinen 00] A. Hyvärinen & E. Oja. *Independent component analysis : algorithms and applications*. *Neural Networks*, vol. 13, pages 411–430, 2000.
- [Hyvärinen 01] A. Hyvärinen, J. Karhunen & E. Oja. *Independent component analysis*. Wiley, New York, 2001.
- [Johansson 91] S. Johansson. *Localization and quantification of water leakage in ageing embankment dams by regular temperature measurements*. In *Proc. ICOLD, 17th Congress*, Vienna, 1991.
- [Johansson 96] S. Johansson & T. Dahlin. *Seepage monitoring in an earth embankment dam by repeated resistivity measurements*. *European Journal of Engineering and Environmental Geophysics*, vol. 1, pages 229–247, 1996.
- [Johansson 97] S. Johansson. *Seepage monitoring in embankment dams*. PhD thesis, Royal Inst. of Tech., Stockholm, 1997.

- [Johansson 98] S. Johansson & T. Dahlin. *Seepage monitoring in Hällby embankment dam by continuous resistivity measurements*. In Proc. 8th Congress of the International Ass. of Eng. Geology and the Environment, Vancouver, 1998.
- [Johansson 04] S. Johansson & P. Sjö Dahl. *Downstream seepage detection using temperature measurements and visual Inspection - monitoring experiences from Rösuvatn field test dam and large embankment dams in Sweden*. In Proc. Intl. Seminar on Stability and Breaching of Embankment Dams, Oslo, Norway, Oct. 2004.
- [Kappelmeyer 57] O. Kappelmeyer. *The use of near surface temperature measurements for discovering anomalies due to causes at depths*. Geophysical Prospecting, vol. 3, pages 239–258, 1957.
- [Kay 91] S. M. Kay. *Fundamentals of statistical signal processing : Estimation theory*. Prentice Hall, 1991.
- [Kendall 63] M. G. Kendall & A. Stuart. *The advanced theory of statistics, volume1 - distribution theory, second edition*. Addison-Wesley, London, 1963.
- [Kersey 99] A. D. Kersey. *Optical fiber sensors for downwell monitoring applications in the oil and gas industry*. In Proc. 13th Intl. Conf. Optical Fiber Sensors, pages 326–331, 1999.
- [Khan 08a] A. A. Khan, V. Vrabie, G. d’Urso & J. I. Mars. *Blind source separation techniques for percolation-type leakage detection in dikes using fiber optic DTS signals*. In Proc. SPIE OFS19, Perth, Australia, 2008.
- [Khan 08b] A. A. Khan, V. Vrabie, G. d’Urso & J. I. Mars. *Multicomponent filtering for leakage detection using bi-dimensional thermometric data*. In Proc. IEEE Conference on Communication, Bucharest, Romania, 2008.
- [Khan 08c] A. A. Khan, V. Vrabie, G. d’Urso & J. I. Mars. *SVD based automated dike monitoring system using DTS data*. In Proc. IEEE IECON’08, Orlando, USA, 2008.
- [Khan 08d] A. A. Khan, V. Vrabie, J. I. Mars & A. Girard. *A least square approach for bidimensional source separation using higher order statistics criteria*. In EUSIPCO, Lausanne, Switzerland, Aug 2008.
- [Khan 08e] A. A. Khan, V. Vrabie, J.I. Mars, A. Girard & 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, pages 1118–1129, 2008.
- [Khan 09a] A. A. Khan, V. Vrabie, J. I. Mars & G. d’Urso. *Séparation de sources thermométriques pour la détection des fuites dans des digues en terre*. In XIIe Colloque GRETSI, Dijon, France, Sep 2009.
- [Khan 09b] A. A. Khan, V. Vrabie, J.I. Mars, A. Girard & G. d’Urso. *Automatic monitoring system for singularities detection in dikes by DTS data measurement*. IEEE Transactions on Instrumentation and Measurement, 2009.
- [Lacoume 90] J. L. Lacoume, P. O. Amblard & P. Comon. *Statistiques d’ordre supérieur pour le traitement du signal*. Wiley, 1990.



- [Le Bihan 01] N. Le Bihan. Traitement algébrique des signaux vectoriels : application en séparation d'ondes sismiques. Thèse de doctorat, Institut National Polytechnique de Grenoble, 2001.
- [Lee 98] T.-W. Lee. Independent component analysis : Theory and applications. Kluwer Academic Publishers, 1998.
- [Lewis 07] E. Lewis, C. Sheridan, M. O'Farrell, D. King, C. Flanagan, W. B. Lyons & C. Fitzpatrick. *Principal component analysis and artificial neural networks based approach to analysing optical fiber sensors signals*. Sensors and Actuators A, vol. 136, pages 28–38, 2007.
- [Loke 00] M. H. Loke. *Electrical imaging surveys for environmental and engineering studies : A practical guide to 2-D and 3-D surveys*. Rapport technique, M. H. Loke, 2000.
- [Lu 98] K. Lu & J. Macnae. *The international campaign on intercomparison between electrodes for geoelectrical measurements*. Exploration Geophysics, vol. 29, no. 4, pages 484–488, 1998.
- [Lukac 05] R. Lukac, B. Smolka, K. Martin, K.N. Plataniotis & A.N. Venetsanopoulos. *Vector Filtering for Color Imaging*. IEEE Signal Processing Magazine, vol. 22, pages 74 – 86, Jan. 2005.
- [Mars 06] J. I. Mars, N. L. Bihan, V. Vrabie & C. Paulus. *Multicomponent seismic data filtering by multilinear methods*. In Proc. ISCCSP, Morocco, 2006.
- [Mazet 05] V. Mazet, C. Carteret, D. Brie, J. Idier & B. Humbert. *Background removal from spectrum by designing and minimising a non-quadratic cost function*. Chemometrics and Intelligent Laboratory Systems, vol. 1, pages 121–133, 2005.
- [McLachlan 08] G. J. McLachlan & T. Krishnan. The em algorithm and extensions. Wiley, 2008.
- [Mendel 91] J. M. Mendel. *Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory : Theoretical Results and Some Applications*. Proceedings of IEEE, vol. 79, pages 278–309, 1991.
- [Morawetz 07] R. Morawetz, J. Schön, C. Wolfahrt & M. Röck. *Dike investigations using geophysical methods - techniques for the future ?* In S. Begum et al., editeur, Flood Risk Management in Europe. Springer, 2007.
- [Niederleithinger 07] E. Niederleithinger, A. Weller, R. Lewis, U. Stötzner, Th. Fechner, B. Lorenz & J. Nieben. *Evaluation of geophysical methods for river embankment investigation*. Rapport technique, 2007. Available at.
- [Niklès 07a] Marc Niklès. *Fibre Optic Distributed Scattering Sensing System : Perspectives and Challenges for High Performance Applications*. In Proc. Third European Workshop on Optical Fibre Sensors, Napoli, Italy, 2007.
- [Niklès 07b] Marc Niklès. *Leakage detection using fiber optics distributed temperature monitoring*. In Proc. Smart Structures and Materials 2004 : Smart Sensor Technology and Measurement Systems, San Diego, CA, USA, 2007.

- 
- [Ogilvy 69] A. A. Ogilvy, M. A. Ayed & V. A. Bogoslovsky. *Geophysical studies of water leakage from reservoirs*. Geophysical Prospecting, vol. 22, pages 32–62, 1969.
- [Orrell 98] P. R. Orrell & A. P. Leach. *Use of fibre optic distributed temperature sensing as a heat detector in a fire detection system*. In Proc. Safety in road and rail tunnels, Nice, France, 1998.
- [Panthulu 01] T. V. Panthulu, C. Krishnaiah & J. M. Shirke. *Detection of seepage paths in earth dams using self-potential and electrical resistivity methods*. Engineering Geology, vol. 59, pages 281–295, 2001.
- [Papoulis 91] A. Papoulis. Probability, random variables, and stochastic processes. McGraw Hill Series in Electrical Engineering, 1991.
- [Paulus 05] C. Paulus, J. I. Mars & P. Gounon. *Wideband spectral matrix filtering for multicomponent sensors array*. Signal Processing, vol. 85, pages 1723 – 1743, Sept. 2005.
- [Perriera 99] F. Perriera, M. Triquea, J. Aupiais, U. Gautamb & P. Shrestha. *Electric potential variations associated with periodic spring discharge in western Nepal*. Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science, vol. 328, pages 73–79, 1999.
- [Pham 92] D.-T. Pham, P. Garrat & C. Jutten. *Separation of a mixture of independent sources through a maximum likelihood approach*. In EUSIPCO, pages 771–774, 1992.
- [Pham 96] D. T. Pham. *Blind separation of instantaneous mixture of sources via an independent component analysis*. IEEE Trans. Signal Processing, vol. 44, pages 2768–2779, 1996.
- [Revil 03] A. Revil, V. Naudet, J. Nouzaret & M. Pessel. *Principles of electrography applied to self-potential electrokinetic sources and hydrogeological applications*. Water Resources Research, vol. 39, no. 5, 2003.
- [Revil 06] A. Revil, K. Titov, C. Doussan & V. Lapenna. *Applications of the self-potential method to hydrological problems*. In H. Vereecken et al., editeur, Applied Hydrogeophysics. Springer, 2006.
- [Rogers 02] A. Rogers. *Distributed Optical Fiber Sensing*. In J. M. Lopez-Higuera, editeur, Handbook of Optical Fiber Sensing Technology, pages 271–308. Wiley, 2002.
- [Rozycki 05] A. Rozycki, J. M. Ruiz & A. Caudra. *Detection and evaluation of horizontal fractures in earth dams using the self potential method*. Engineering Geology, vol. 82, no. 3, pages 145–153, 2005.
- [Scharf 91] L. L. Scharf. Statistical signal processing : Detection, estimation, and time series analysis. Addison-Wesley, New York, 1991.
- [Schervish 01] M. Schervish & M. H. DeGroot. Probability and statistics. Addison Wesley, 2001.
- [Schlumberger 20] C. Schlumberger. Étude sur la prospection électrique du sous-sol. Gaultier-Villars et Cie, 1920.

- [Schwarz 78] G. Schwarz. *Estimating the dimension of a model*. Annals of Statistics, vol. 6, pages 461–464, 1978.
- [Sensornet 09] Sensornet. [www.sensornet.co.uk/news/case-studies](http://www.sensornet.co.uk/news/case-studies), 2009.
- [Sheffer 04] M. Sheffer, E. Haber, D. Oldenburg, R. Shekhtman & J. Chen. *SP3D : a tool for the analysis of self-potential data in seepage investigations*. In Proceedings of the International Seminar on the Stability and Breaching of Embankment Dams, Oslo, Norway, 2004.
- [Sjödahl 06] P. Sjödahl. *Resistivity investigation and monitoring for detection of internal erosion and anomalous seepage in embankment dams*. PhD thesis, Lund University, Sweeden, 2006.
- [Sjödahl 08] P. Sjödahl, T. Dahlin, S. Johansson & M. H. Lokec. *Resistivity monitoring for leakage and internal erosion detection at Hällby embankment dam*. Journal of Applied Geophysics, vol. 65, pages 155–164, 2008.
- [Sorin 92] W. V. Sorin. *High-resolution optical fiber reflectometry techniques*. In Proc. SPIE Distributed and Multiplexed Fiber Optic Sensors II, Boston, USA, 1992.
- [Soutter 07] M. Soutter, A. Mermoud & A. Musy. *Ingénierie des eaux et du sol*. PPUR presses polytechniques, 2007.
- [Tan 09] C. O. Tan, J. A. Taylor, A. S. H. Ler & M. A. Cohen. *Detection of Multifiber Neuronal Firings : A Mixture Separation Model Applied to Sympathetic Recordings*. IEEE Trans. Biomed. Eng., vol. 56, pages 147–158, 2009.
- [Titov 00] K. Titov, V. Lokhmanov & A. Potapov. *Monitoring of water seepage from a reservoir using resistivity and self polarization methods : case history of the Petergoph fountain water supply system*. First Break, vol. 18, pages 431–435, 2000.
- [Udd 91] E. Udd. *The emergence of fiber optic sensor technology*. In E. Udd, editeur, Fiber Optic Sensors : An Introduction for Engineers and Scientists, pages 1–8. Marcel Dekker Inc., 1991.
- [Udd 06] E. Udd. *Overview of Fiber Optic Sensors*. In S. Yin, P. B. Ruffin & F. T. Yu, editeurs, Fiber Optic Sensors, pages 1–34. Taylor and Francis Group, 2006.
- [Vercoutere 03] G. Vercoutere. *Installation de la fibre optique à l’usine d’Oraison*. Edf technical report, EDF, 2003.
- [Vercoutere 05] G. Vercoutere. *Essais de fuites artificielles sur le site d’Oraison*. Edf technical report, EDF, 2005.
- [Vogel 01] B. Vogel, C. Cassens, A. Graupner & A. Trostel. *Leakage detection systems by using distributed fiber optical temperature measurements*. In D. Inaudi E. Udd, editeur, Proc. SPIE Smart Structures and Materials, 2001, volume 4328, pages 23–34, 2001.
- [Vrabie 04] V. Vrabie, J. I. Mars & J-L . Lacoume. *Singular value decomposition by means of independent component analysis*. Signal Processing, vol. 84, no. 3, pages 645–652, 2004.

- [Wade 04] S. A. Wade, K. T. Grattan & B. McKinley. *Incorporation of fiber-optic sensors in concrete specimens : Testing and evaluation*. IEEE Sensors Journal, vol. 4, no. 1, pages 127–134, Feb. 2004.
- [Wegmüller 01] M. Wegmüller, P. Oberson, J. P. von derWeid, O. Guinnard, L. Guinnard, C. Vinegoni, M. Legré & N. Gisin. *Overview of coherent reflectometry techniques : characterization of components and small systems*. In P. A. Williams & G. W. Day, éditeurs, Technical Digest Symposium on Optical Fiber Measurements. NIST, 2001.
- [Yamate 08] T. Yamate. *Fiber Optic Sensors for the Exploration of Oil and Gas*. In Proc. SPIE OFS19, Perth, Australia, 2008.
- [Yin 00] S. Yin. *Distributed Fiber Optic Sensors*. In F. T. Yu & S. Yin, éditeurs, Fiber Optic Sensors, pages 201–229. Marcel Dekker Inc., 2000.
- [Yu 94] F. T. S. Yu, J. Zhang, S. Yin & P. B. Ruffin. *A novel distributed fiber sensor based on the Fourier spectrometer technique*. In Proc. OSA Annual Meeting, Texas, 1994.





---

## RÉSUMÉ

Ce travail de thèse aborde un problème de la détection des fuites d'eau dans des digues en terre. L'écoulement préférentiel, un indicateur de destruction interne, résulte dans une anomalie thermique. Les capteurs DTS basés sur la fibre optique offrent une solution efficace pour les mesures spatio-temporelles de température. Les données thermométriques sont influencées par différents facteurs comme la réponse du sol, les variations saisonnières, l'environnement géomécanique, etc. Afin d'écarter des phénomènes comme les précipitations, nous proposons un critère basé sur le skewness et le kurtosis. La détection des fuites étant formulée comme un problème de séparation de sources, nous présentons un système basé sur la SVD et l'ICA pour séparer les informations utiles liées aux fuites des autres facteurs. Dans le cas où le nombre d'acquisitions en temps est limité, nous proposons un détecteur de singularités exploitant les dissimilarités entre les mesures journalières de température aux différentes distances.

---

## MOTS-CLÉS

détection des fuites, mesures thermométriques, capteurs à fibre optique, analyse en composantes indépendantes, décomposition en valeurs singulières, statistiques d'ordre supérieure, surveillance

---

## ABSTRACT

The aim of this research is to address the problem of leakage detection in embankment dikes. Significant flow of water through the dike, an indicator of internal erosion, results in a thermal anomaly. The measurement of temperature is therefore capable of revealing information linked to leakage. Fiber optic based distributed temperature sensors present an economically viable solution for recording spatio-temporal temperature data. The data are influenced by several factors amongst them the leakages, the response of the near surface, the seasonal variations, geomechanical environment, etc. In order to remove phenomena like precipitations, we propose a criterion based on skewness and kurtosis. Formulating leakage detection as a source separation problem, firstly, we present a system based on SVD and ICA for the separation of useful information relevant to the leakages from other "non relevant" factors. Secondly, for the case of limited acquisitions, we propose a singularity detector exploiting the dissimilarity of daily temperature variations at different distances.

---

## KEY WORDS

leakage detection, temperature measurements, fiber optic sensors, independent component analysis, singular value decomposition, higher order statistics, surveillance

---