



Mémoire de fin d'étude

Pour l'obtention du diplôme d'Ingénieur

Filière : Automatique  
Spécialité : Automatique

Présenté par :

Asmaa BENDAHMANE et Sarah HAMDY CHERIF

Thème

**Introduction de l'intelligence artificielle  
pour le diagnostic et la détection de  
défauts dans un système automatique.**

Soutenu à huis clos, le 11 / 07 / 2021, devant le jury composé de :

Dr Fouad BOUKLI HACENE	MCA	ESSA. Tlemcen	Président
Dr Sidi Mohammed ABDI	MCB	ESSA. Tlemcen	Directeur de mémoire
Dr Mustapha A. BRAHAMI	MCB	ESSA. Tlemcen	Co-Directeur de mémoire
Pr Brahim CHERKI	PR	ESSA. Tlemcen	Examineur 1
Dr Ghouti ABDELLAOUI	MCB	ESSA. Tlemcen	Examineur 2
Dr Fouad MALIKI	MCB	ESSA. Tlemcen	Invité 1
Dr Sidi Mohammed KHEFIF	MCB	ESSA. Tlemcen	Invité 2



# Dédicaces

*À nos parents,  
À Wassim et Chahrazed,  
À Naila et Abderahmane,  
À nos amis et à notre promotion,  
À Meriem et Meriem (elles se reconnaîtront),  
Sans oublier Chouquette, Pichu et Kirby.*

# Remerciements

Nous tenons à exprimer notre gratitude envers nos encadrants, M. Sidi Mohammed ABDI et M. Mustapha Anwar BRAHAMI pour leur patience et leur indulgence. Nous les remercions aussi pour l'aide qu'ils nous ont fourni, les connaissances qu'ils nous ont transmises et surtout pour leur constante disponibilité.

Nous remercions également M. Sidi Mohammed KHEFIF pour son aide et ses précieux conseils.

Nous remercions les membres du jury : Pr B. CHERKI, Dr F. BOUKLI-HACENE, et Dr G. ABDELLAOUI d'avoir accepté d'évaluer notre travail.

Pour finir, nous ne saurons oublier de trop remercier nos familles qui ont su nous encourager par leur enthousiasme pour ce projet même si ils n'y comprenaient pas grand chose.

# Table des matières

## Liste des abréviations

<b>Introduction</b>	<b>1</b>
<b>1 Généralités sur le diagnostic</b>	<b>3</b>
1 Introduction . . . . .	3
2 Aperçu sur le diagnostic . . . . .	3
2.1 Terminologie . . . . .	3
2.2 Types de défauts . . . . .	4
3 Diagnostic à base de modèle . . . . .	5
4 Méthodes de détection de fuites . . . . .	6
4.1 Méthodes de détection passive (vision-based) . . . . .	6
4.2 Méthodes sans modèle (data-driven) . . . . .	6
4.3 Méthodes basées sur un modèle hydraulique . . . . .	6
5 Les méthodes basées sur l'analyse des transitoires (TBDM) . . . . .	6
5.1 L'Analyse transitoire inverse (ITA) . . . . .	7
5.2 Analyse dans le domaine fréquentiel . . . . .	7
5.3 Analyse transitoire directe . . . . .	8
5.3.1 Analyse de la réflexion de l'onde transitoire (TRM) . . . . .	8
5.3.2 Analyse de l'amortissement de l'onde transitoire (TDM) . . . . .	8
6 Conclusion . . . . .	9
<b>2 Présentation du cas étudié</b>	<b>10</b>
1 Introduction . . . . .	10
2 Analyse transitoire inverse (ITA) . . . . .	10
3 Présentation du système . . . . .	11
4 Présentation du modèle . . . . .	11
5 Matched field processing (MFP) . . . . .	12
6 L'Estimateur du maximum de vraisemblance (EMV) . . . . .	15
6.1 Le principe de l'EMV . . . . .	16
6.2 La borne inférieure de Cramer-Rao . . . . .	16
6.3 Équivalence entre MFP et EMV . . . . .	18
7 Conclusion . . . . .	18
<b>3 Les Algorithmes Génétiques</b>	<b>19</b>
1 Introduction . . . . .	19
2 Optimisation . . . . .	19
2.1 Définition : . . . . .	19
2.2 Terminologie : . . . . .	19
2.3 Types de problèmes d'optimisation : . . . . .	19
2.4 Classes de difficulté : . . . . .	20

2.5	Types de méthodes d'optimisation . . . . .	20
2.6	Les méthodes exactes : . . . . .	21
2.7	Les méthodes approchées : . . . . .	21
3	Les Métaheuristiques . . . . .	22
3.1	Caractéristiques . . . . .	22
3.2	Classification . . . . .	22
4	L'Algorithme Génétique . . . . .	23
4.1	Origine . . . . .	23
4.2	Analogie avec la biologie . . . . .	23
4.3	Etapas de l'Algorithme . . . . .	23
4.4	Les opérateurs de l'algorithme génétique . . . . .	24
4.5	Paramètres de l'algorithme génétique . . . . .	25
4.6	Avantages de l'algorithme génétique . . . . .	25
4.7	Inconvénients de l'algorithme génétique . . . . .	25
5	Conclusion . . . . .	25
<b>4</b>	<b>Application de l'Algorithme Génétique sur le modèle</b>	<b>26</b>
1	Introduction . . . . .	26
2	Description de l'Algorithme Génétique . . . . .	26
2.1	Codage des solutions . . . . .	26
2.2	Fonction Objectif . . . . .	26
2.3	Opérateurs génétiques adoptés . . . . .	27
2.4	Procédure de vérification de contraintes et de réparation des solutions infaisables . . . . .	28
3	Expériences numériques et interprétations . . . . .	29
3.1	Génération d'instances . . . . .	29
3.2	Résultats obtenus . . . . .	29
3.3	Interprétation des résultats : . . . . .	30
3.4	Conclusion de l'analyse . . . . .	32
4	Conclusion . . . . .	35
	<b>Conclusion</b>	<b>36</b>
	<b>Conclusion générale</b>	<b>38</b>

# Liste des tableaux

- 4.1 Résultats obtenus par l'AG pour les petites distances . . . . . 30
- 4.2 Résultats obtenus par l'AG pour les grandes distances . . . . . 30

# Table des figures

1.1	Différents types de défauts d'un système physique [32] . . . . .	5
1.2	Processus de génération de résidus[15] . . . . .	5
1.3	Onde transitoire dans le domaine temporel (a et b); et (c et d) dans le domaine fréquentiel, pour (a et c) une canalisation intacte est considérée et (b et d) une canalisation défaillante (avec des fuites) est considérée.[48] . . . . .	9
2.1	Localisation de deux fuites éloignées en utilisant MFP . . . . .	14
2.2	Localisation de deux fuites proches en utilisant MFP . . . . .	14
2.3	Localisation de deux fuites proches en utilisant EMV . . . . .	15
2.4	Le cas ou EMV satisfait la CRLB . . . . .	17
2.5	Le cas ou EMV ne satisfait pas la CRLB . . . . .	18
3.1	Classification des méthodes d'optimisation [25] . . . . .	21
4.1	Évolution du temps d'exécution par rapport à la longueur des canalisations . . . . .	31
4.2	Pour une canalisation de longueur $L = 1km$ le temps de convergence est $t_{conv} = 6min$ . . . . .	32
4.3	Pour une canalisation de longueur $L = 1,5km$ le temps de convergence est $t_{conv} = 6min$ . . . . .	33
4.4	Pour une canalisation de longueur $L = 4km$ le temps de convergence est $t_{conv} = 13min$ . . . . .	33
4.5	Pour une canalisation de longueur $L = 8km$ le temps de convergence est $t_{conv} = 10min$ . . . . .	34
4.6	Pour une canalisation de longueur $L = 12km$ le temps de convergence est $t_{conv} = 23min$ . . . . .	34

# Liste des abréviations

<b>AG</b>	:	Algorithme génétique
<b>CRLB</b>	:	Cramer-Rao lower bound
<b>EMV</b>	:	Estimation du maximum de vraisemblance
<b>ITA</b>	:	Inverse transient analysis
<b>MFP</b>	:	Matched field processing
<b>MH</b>	:	Métaheuristiques
<b>RE</b>	:	Recherche exhaustive
<b>SNR</b>	:	Signal-to-noise ratio
<b>TBDM</b>	:	Transient based detection methods

# Introduction

A correct diagnosis is three-fourths the  
remedy.

---

Mahatma Ghandi

Pour s'assurer de la qualité et de la longévité d'un processus, il est nécessaire d'avoir un système de supervision capable de nous prévenir en cas de dysfonctionnement. Il est encore plus important d'être capable de localiser la source du problème et de pouvoir émettre un diagnostic fiable et rapide. Les méthodes de détection de fautes sont multiples et ne cessent d'accroître en nombre et en efficacité avec les années. Le développement des équipements industriels a permis l'implémentation de techniques toujours plus développées et complexes, notamment celles qui découlent de l'Intelligence Artificielle et qui sont réputées plus robustes et moins chronophages.

Nous avons choisi de nous intéresser à un problème qui sans être récent reste toujours d'actualité : la localisation de fuites dans une canalisation d'eau. Allant sur plusieurs kilomètres, les canalisations sont sujettes à plusieurs types de perturbations - comme le trafic routier, des travaux dans les alentours ou encore les tremblements de terre dans les zones qui y sont sensibles - qui peuvent provoquer une mauvaise pression sur les matériaux, en addition à des défauts de fabrication, des faiblesses causées par la corrosion ou par le phénomène de coup de bélier. Ce sont des constructions qui sont destinées à fonctionner dans des conditions stressantes, ce qui peut provoquer bon nombres de défauts, dont les fuites.

Les pertes qui résultent de ces fuites peuvent devenir rapidement très conséquentes si on ne dispose pas d'un diagnostic rapide et précis. En effet, en plus du coût financier supporté par les services publics, il y'a aussi des risques sanitaires via la contamination du liquide - qui n'est plus isolé -, ainsi que la charge environnementale associée au gaspillage de l'énergie.

Vu l'importance du problème, de nombreuses méthodes ont été mises au point [5] : séparés en deux grandes classes. Les méthodes hardware, qui reposent essentiellement sur les capteurs ; on peut citer les systèmes de détection des fuites basé sur les ondes de pression négative [44], à l'aide de capteurs à fibre optique [23] ou encore par capteurs d'émission acoustique piézoélectriques. [37]

La deuxième classe concerne les méthodes software : à base de traitement de signal numérique. On peut citer la simulation numérique de la dynamique des fluides [4], l'analyse harmonique en ondelettes [22], ou encore la détection par apprentissage à base de réseaux de neurones flous. [13]

Une hybridation de ces méthodes peut aussi se révéler plus efficace dans certains cas, comme l'analyse inverse des transitoires, combiné avec des techniques d'optimisation tel que l'algorithme de Levenberg-Marquardt [30], ou encore les algorithmes génétiques [46].

Les méthodes acoustiques sont réputées être les plus efficaces, et elles sont nombreuses. La plupart exploitent le régime transitoire de l'onde acoustique qui est plus fourni en informations. Une de ces méthodes est l'ITA : l'analyse inverse des transitoires [30], elle consiste à créer un modèle numérique des pressions manométriques dans une canalisation. Elle procède à la localisation des fuites en minimisant une fonction objectif qui représente la différence entre les mesures réelles et celles modélisées.

Cette méthode n'est pas adaptée aux environnements bruités et utilise une technique d'optimisation exacte qui peut fausser les résultats une fois que l'espace de recherche s'agrandit en bloquant sur des minimums locaux.

Les travaux de Wang et Ghidaoui [48], [47] ont présenté deux méthodes basées sur l'ITA, mais qui proposent des solutions aux deux problèmes soulevées. Celles-ci utilisent la notion de recherche exhaustive - afin d'optimiser la fonction objectif - qui résulte en un temps d'exécution important et qui augmente avec l'espace de recherche. Le temps de diagnostic est proportionnel à la longueur de la canalisation.

Nous proposons l'introduction d'une technique de l'intelligence artificielle, à savoir les algorithmes génétiques, afin de remplacer la recherche exhaustive. Le but est d'arriver à un diagnostic rapide et précis.

Les algorithmes génétiques sont des méthodes d'optimisation pratiques pour exploiter de grands espaces de recherches, même si elles sacrifient l'exactitude de la solution au profit d'un temps de résolution acceptable.

Dans le premier chapitre, nous exposerons des notions de diagnostic et principalement les méthodes qui ont été pertinentes pour ce travail.

Puis, dans le deuxième chapitre, nous présenterons le système étudié, son modèle analytique ainsi que les notions importantes des travaux de Wang et Ghidaoui sur lesquels nous nous sommes basées.

Le troisième chapitre est dédié aux algorithmes génétiques, leurs structures et les notions d'optimisation qui y sont liés.

Finalement, le quatrième chapitre contient la présentation de l'algorithme génétique que nous avons implémenté ainsi que l'interprétation des résultats que nous avons obtenus. Nous terminons avec une conclusion générale qui fait le point sur l'ensemble du travail.

# Chapitre 1

## Généralités sur le diagnostic

### 1 Introduction

Un système est forcément amené à présenter des défauts durant son cycle de vie, qu'ils soient innés ou provoqués. C'est pour cela que le diagnostic prend de plus en plus d'importance aux cotés de la maintenance pour assurer et le bon fonctionnement et la longévité des processus.

Ce chapitre introduit d'importantes notions sur le diagnostic et principalement celles dont nous avons eu besoin dans notre travail. En effet, les approches de détection de fautes sont nombreuses et diverses. Nous avons donc préféré nous restreindre à la présentation du diagnostic à base de modèle, suivi d'un bref état de l'art sur les méthodes - anciennes et modernes - de la détection de fuites dans une canalisation afin de contextualiser notre travail.

### 2 Aperçu sur le diagnostic

Les définitions du diagnostic et de la détection des fautes sont nombreuses, nous en avons choisi une qui nous semble complète :

« Comme dans son acception médicale, le diagnostic des systèmes automatiques consiste à remonter des symptômes perçus vers les causes. Il est également défini comme l'ensemble des actions mises en œuvre en vue de détecter, localiser et identifier tout phénomène anormal sur un système à partir de la connaissance disponible sur ce dernier. La connaissance disponible inclut, non seulement, les sorties du contrôleur (qui a pour but d'améliorer les performances du système), les mesures fournies par les capteurs mais aussi, le plus souvent, des informations sur la structure et/ou un modèle du système. » [35]

Comme cité plus haut, les principales étapes suivies pour un bon diagnostic sont la détection, la localisation puis l'identification de toute anomalie. On entend par ce dernier terme les différents dysfonctionnements qui peuvent apparaître durant le cycle de vie d'un processus : des défauts de fabrications en passant par les défaillances abruptes jusqu'aux usures causées par la vieillesse des matériaux ou le mauvais entretien.

#### 2.1 Terminologie

*Défaut* : est un écart entre la caractéristique observée et la caractéristique de référence. Un défaut peut conduire à une défaillance du système tout comme il peut exister sans altérer son

fonctionnement général.

*Défaillance* : est une conséquence possible d'une altération importante des caractéristiques. Elle correspond à l'incapacité du système à réaliser la fonction souhaitée. Elle peut être passagère ou permanente.

*Panne* : est l'interruption permanente du système à réaliser la fonction souhaitée.

*Résidus* : Indicateurs d'anomalie. Expriment la différence entre les valeurs théoriques et celles mesurées. Présentent une valeur nulle lors du fonctionnement normal.

*Monitoring* : c'est la surveillance en temps réel du système, il permet la détection précoce des défauts.

*Supervision* : c'est la surveillance du système et la prise de décisions rapide dans le cas de l'existence d'un défaut.

*Détection de défaut* : il s'agit de détecter rapidement les anomalies et de prendre une décision rapide quant à l'existence ou non d'un défaut dans le système.

*Isolation des défauts* : il s'agit de déterminer le composant qui cause problème, par exemple, quel capteur ou actionneur est devenu défectueux.

*Identification du défaut* : il s'agit d'estimer la taille, le type ou la nature du défaut.

*Diagnostic* : La détermination du type, de la taille, de la localisation et de l'instant d'occurrence d'un défaut ; il suit la détection de défauts et inclut l'isolation et l'identification. [?]

## 2.2 Types de défauts

Dans ce qui suit, nous décrirons les différentes sources de défauts dans un système [42].

### **Défauts sur les capteurs :**

Les capteurs représentent l'interface de sortie d'un système, et transmettent des informations sur le comportement de ce dernier. Par conséquent, les défauts du capteur peuvent entraîner une dégradation des performances du processus et compromettre l'intégrité des données ce qui revient à fausser la prise de décisions.

Les capteurs sont utilisés soit pour mesurer directement les états du système, soit pour générer des estimations d'état pour la loi de contrôle. Ainsi, la présence de défauts dans les capteurs peut détériorer les estimations d'état et par conséquent entraîner un diagnostic inexact.

### **Défaut sur les actionneurs :**

Dans de nombreux systèmes les signaux de commande du contrôleur ne peuvent pas être directement appliqués au système. Des actionneurs sont nécessaires pour les transformer en signaux d'actionnement appropriés. Cependant, l'apparition d'anomalies dans cet organe peut amener à une perte de contrôle totale.

### Défauts sur le processus :

Le défaut sur le processus survient généralement lorsque certaines conditions changent dans le système, rendant invalide l'équation dynamique nominale du système.

La figure ci-dessous représente un récapitulatifs des types de défauts présentés.

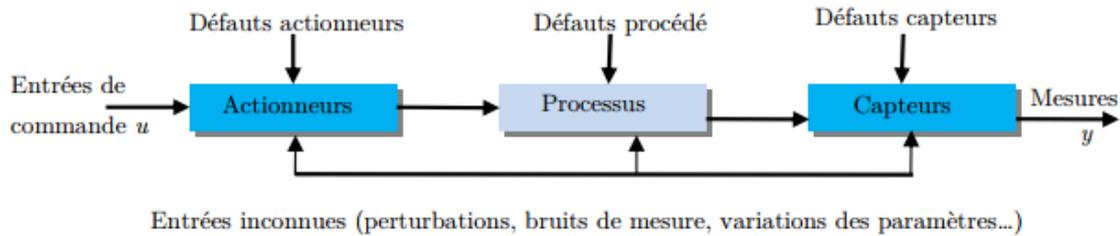


FIGURE 1.1 – Différents types de défauts d'un système physique [32]

## 3 Diagnostic à base de modèle

Les méthodes basées sur des modèles effectuent un diagnostic de panne basé sur une redondance analytique dans laquelle la cohérence entre les mesures et le comportement attendu du procédé sont vérifiés par des modèles analytiques, ces derniers se présentent généralement sous forme mathématique.

Sur cette base des méthodes de génération résiduelle telles que le filtre de Kalman, l'identification du système et les relations de parité sont utilisées pour obtenir ce qu'on appelle des résidus qui indiquent la présence de défaut dans un processus. Enfin, les résidus sont évalués pour réaliser la détection, l'isolation et l'identification des défauts[24].

Ce processus est illustré par la figure qui suit :



FIGURE 1.2 – Processus de génération de résidus[15]

En règle générale, les approches basées sur un modèle peuvent être plus efficaces si un modèle correct et précis est construit.

Cependant, un modèle mathématique parfaitement précis et complet d'un système physique n'est jamais disponible ou difficile à mettre en œuvre. Habituellement, les paramètres du système peuvent varier avec le temps de manière incertaine, et les caractéristiques des perturbations et du bruit sont inconnues car ces phénomènes ne peuvent pas être modélisés avec précision [3]. Par conséquent, il y a toujours un décalage entre le processus réel et son modèle mathématique même s'il y a aucun défaut de processus.

## 4 Méthodes de détection de fuites

### 4.1 Méthodes de détection passive (vision-based)

Le plus ancien système de détection passive des fuites consiste à signaler toute indication d'accumulation à la surface du sol, par exemple une croissance anormale de la végétation peut être considérée comme une fuite possible [12].

### 4.2 Méthodes sans modèle (data-driven)

Cette approche repose sur la collecte de données afin d'effectuer le traitement du signal et l'analyse statistique pour la détection des fuites. L'avantage de cette approche est qu'elle ne nécessite aucune connaissance approfondie du système. Elle n'a besoin d'apprendre que des données historiques collectées couplées à des outils statistiques ou de reconnaissance de formes (M. Romano *et al.* 2014).

Le principal désavantage de cette méthode est qu'une grande quantité de données est nécessaire pour développer une classification [50] et qu'elles peuvent être difficiles à trouver. La classification risque donc d'être mal réalisée si certaines données manquent, ou que les mesures des capteurs sont bruitées [14].

Ces inconvénients démontrent la nécessité d'avoir une certaine connaissance du système hydraulique afin de faire la différence entre les données fiables et celles erronées.

### 4.3 Méthodes basées sur un modèle hydraulique

L'approche basée sur un modèle implique généralement l'utilisation de fonctions ou de formules mathématiques pour représenter ou reproduire le fonctionnement d'un réseau de canalisations. On peut déterminer l'emplacement approximatif des fuites (Y. Wu *et al.*, 2018) en comparant les mesures de pression avec leurs estimations obtenues à l'aide du modèle. La condition est que le modèle soit basé sur des fonctions mathématiques et des lois de la physique ainsi que de l'hydraulique[12].

## 5 Les méthodes basées sur l'analyse des transitoires (TBDM)

Aussi appelée méthodes par émission acoustique, elles sont efficaces et moins coûteuses en capteurs puisque souvent couplés à du traitement de signal numérique.

L'idée est d'exploiter les propriétés hydrauliques du fluide afin de détecter les anomalies, cela nécessite d'étudier la canalisation en état de fonctionnement. S'il est possible de créer un modèle basé sur le régime permanent, il risque d'être peu fiable et assez éloigné de la réalité. Les caractéristiques, à savoir : le débit, la vitesse d'écoulement et la pression, restent constantes. Une telle étude permet de réaliser un dimensionnement du conduit étudié, mais ne permet pas la détection précise d'anomalies.

C'est en 1920 que *Babbitt* proposa l'étude par onde transitoire pour exploiter les phénomènes de pressions créées par les fuites. Le régime transitoire est plus riche en informations, les caractéristiques du fluide varient par rapport au temps, ce qui permet une meilleure analyse du conduit. L'idée principale des TBDM est d'exciter le liquide par l'intermédiaire d'un événement tel que la fermeture d'une vanne, qui aura l'effet d'un coup de bélier. Cela créera une onde de pression hydraulique qui fera des allers-retours entre la vanne et le réservoir jusqu'à se dissiper

complètement. Si cet évènement est déclenché plusieurs fois de suite, suivant une période, alors l'onde hydraulique deviendra périodique [10].

Au cours de son parcours, un signal transitoire acquiert des propriétés relatives à la configuration et à l'intégrité du système. L'analyse de ce signal permet d'exposer et de localiser les fuites. Ces transitoires injectées sont conçues pour avoir de faibles amplitudes et peuvent être utilisés régulièrement pour la surveillance du système sans devoir craindre des répercussions sur les matériaux.

Une fuite est un phénomène hydraulique, une singularité qui est repérable lorsqu'on étudie la réponse de l'onde de pression transitoire. Les effets les plus importants que causent une fuite sont :

- L'augmentation du taux d'amortissement de l'onde transitoire.
- La réflexion de l'onde hydraulique au niveau de la trace de pression résultante.

A partir de là, certaines méthodes exploitent la nature amortissante de la fuite (TDM), d'autres sa nature réfléchissante (TRM) et d'autres exploitent les deux. Dans un article dédié à la revue de littérature des TBDM ; Colombo, Lee et Karney [11] ont classifié les méthodes basées sur l'analyse des transitoires comme suit en trois groupes :

## 5.1 L'Analyse transitoire inverse (ITA)

Cette approche consiste à faire l'étalonnage inverse d'un modèle numérique du système par rapport aux données transitoires observées. Cette analyse est effectuée dans le domaine temporel.

C'est en 1992 que Liggett et Pudar [38] publient un article pionnier ; ce sont les premiers travaux à suggérer d'utiliser les mesures de pressions pour détecter les fuites et aussi les premiers à considérer que c'est un problème inverse. Cette étude a été faite en régime stationnaire et n'a donc pas donné de résultats satisfaisants. C'est en 1994 que Liggett et Chen pousse l'étude plus loin en exploitant le régime transitoire, améliorant ainsi la précision de la détection grâce au surplus d'informations qui y est disponible [30].

L'idée générale est de minimiser une fonction objectif au sens des moindres carrés. Elle représente la différence de pression manométrique entre les points du modèle numérique et des mesures observées. Plusieurs méthodes ont été développées sur ce principe, utilisant des algorithmes d'optimisation différents afin d'avoir de meilleurs résultats [46].

## 5.2 Analyse dans le domaine fréquentiel

Ces méthodes considèrent que la présence de la fuite modifie également la réponse en fréquence du système, et donc la variation associée de la réponse du système peut éventuellement être détectée et décodée.

Au lieu de surveiller la réponse transitoire à partir d'une variété d'emplacements dans un système, comme dans l'ITA standard, les techniques du domaine fréquentiel exigent généralement la mesure de l'historique de la pression à une seule section de mesure. En outre, elles

impliquent un dispositif actionné périodiquement, généralement une vanne de contrôle.

En pratique, la méthode implique la procédure suivante : Une vanne située à l'extrémité aval d'une canalisation est périodiquement actionnée selon un schéma défini, produisant une excitation oscillatoire régulière dans le système. À l'emplacement de la vanne, les amplitudes de la pression manométrique et du débit sont enregistrées. Ces actions sont ensuite répétées pour une gamme de fréquences en ajustant la période des oscillations de la vanne, ce qui donne lieu à un diagramme de réponse en fréquence obtenu expérimentalement pour le système. Ce dernier peut alors être comparé à celui d'un système "sans fuite", lui-même obtenu par modélisation numérique du système [10],[48].

### 5.3 Analyse transitoire directe

Ces techniques tentent d'isoler les effets induits par la fuite dans la trace de pression, en cherchant à exposer directement les particularités du signal qui proviennent de la fuite.

#### 5.3.1 Analyse de la réflexion de l'onde transitoire (TRM)

En général, une fuite crée un jet turbulent qui interagit avec la paroi du conduit. Ce jet entraîne d'importantes fluctuations de pression, générant ainsi un son qui peut traverser la paroi de la canalisation ou atteindre la surface. De plus, lorsque l'eau passe par le trou de la fuite, sa vitesse augmente. Si cette dernière est suffisamment élevée, la pression au point de fuite peut tomber en dessous de la pression de vapeur du liquide et former des bulles. L'implosion de ces bulles provoquent des minuscules ondes de choc, leur contact avec la paroi dure de la canalisation provoque un son qui est détecté par les capteurs (microphones, accéléromètres, hydrophones) puis traité avec des techniques de traitement de signal telles que la méthode des ondelettes [27].

#### 5.3.2 Analyse de l'amortissement de l'onde transitoire (TDM)

La taille et l'emplacement de la fuite sont déterminés analytiquement à l'aide d'une analyse de Fourier appliquée directement aux versions modifiées des équations directrices afin de déterminer la relation entre les taux d'amortissement des différentes harmoniques. Il a été découvert que le taux d'amortissement de chaque harmonique dans un signal transitoire est différent pour une canalisation qui présente des fuites [12].

Les figures 1.3 ci-dessous montrent un exemple de signal d'onde de pression dans le domaine temporel et dans le domaine fréquentiel. Il est clair qu'une fuite modifie les deux signaux. Dans le premier, les singularités observées peuvent être expliquées par la réflexion de l'onde au niveau de la fuite. Dans le domaine fréquentiel, la chute de pression manométrique amortit la résonance de manière périodique.

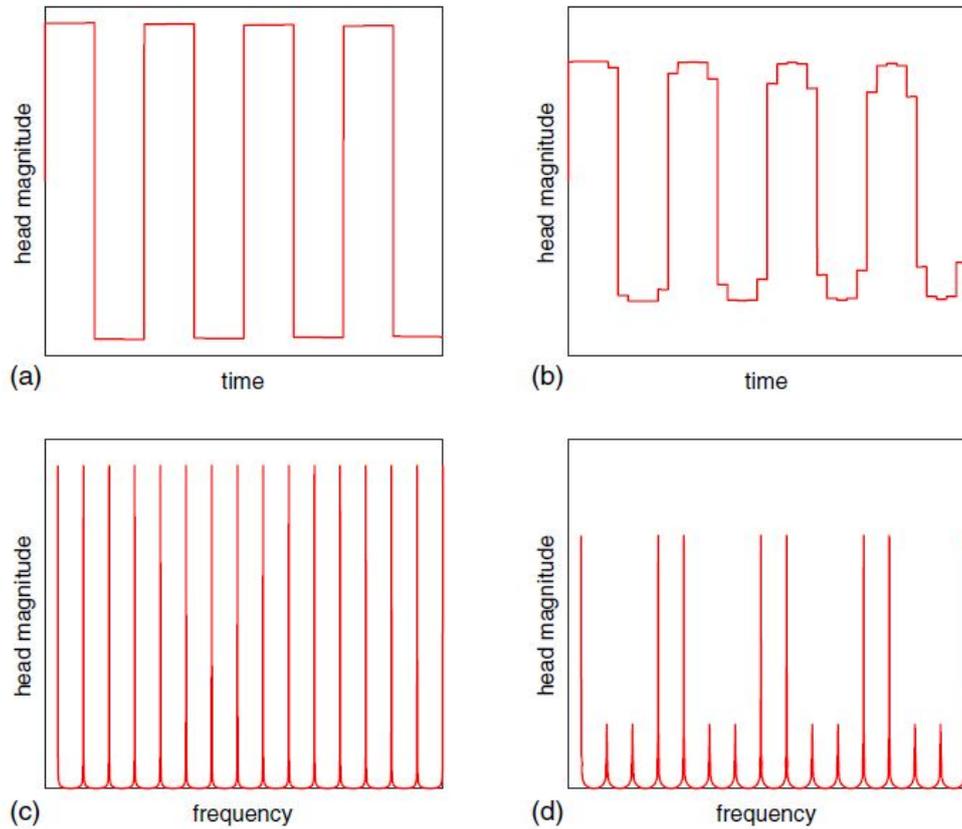


FIGURE 1.3 – Onde transitoire dans le domaine temporel (a et b); et (c et d) dans le domaine fréquentiel, pour (a et c) une canalisation intacte est considérée et (b et d) une canalisation défailante (avec des fuites) est considérée.[48]

## 6 Conclusion

Un large éventail de méthodes de diagnostic est disponible, le choix de l'approche adéquate et déterminé par la structure et la complexité du problème.

Nous avons surtout détaillé dans ce chapitre les méthodes basées sur l'analyse de transitoires afin de pouvoir passer à la présentation du problème.

Une des exigences fonctionnelles d'un diagnostiqueur est **la rapidité du diagnostic**, d'où la nécessité d'utiliser un algorithme d'optimisation de type méta-heuristique.

# Chapitre 2

## Présentation du cas étudié

### 1 Introduction

Comme cas d'étude, nous avons considéré un problème d'actualité : la localisation de fuites dans une canalisation d'eau. Les pertes qui résultent de ces fuites peuvent devenir rapidement très conséquentes si on ne dispose pas d'un diagnostic rapide et précis. Sans parler des risques de contaminations puisque le liquide n'est plus isolé.

Ce chapitre a pour but de décrire le processus choisi comme application, ainsi que diverses notions figurant dans les travaux de Xun WANG et Mohamed S. GHIDAOUI [48],[47] qui ont servi de base à notre travail.

### 2 Analyse transitoire inverse (ITA)

Comme il a déjà été vu dans le premier chapitre, cette méthode a été introduite par Liggett et Chen en 1994 [30] et fait partie des méthodes basées sur l'analyse des transitoires (TBDM).

Après avoir introduit une onde de pression hydraulique en guise de perturbation, des mesures de la pression manométrique sont effectuées, grâce à des hydrophones positionnés stratégiquement dans la canalisation. En parallèle, nous avons un modèle numérique de la canalisation qui considère chaque point comme fuite potentielle. A partir de là, dans les travaux de Liggett et Chen, une fonction objectif est proposée [30] :

$$E = \sum_{i=1}^J (h_i^m - h_i)^2 \quad (2.1)$$

avec :

$h^m$  : les mesures de pression manométrique par les capteurs (hydrophones).

$h_i$  : les pressions calculées par le modèle.

$J$  : le nombre de fréquences.

Il s'agit de minimiser au sens des moindres carrés cette différence afin de trouver l'emplacement des fuites. Le problème est que cette approche ne prend pas en compte le bruit et devient donc peu fiable dans des conditions réelles.

Ce dernier inconvénient est l'objet d'étude des travaux de Wang et Ghidaoui, leur objectif étant d'aboutir à une méthode robuste face aux perturbations.

### 3 Présentation du système

L'étude portera sur une canalisation horizontale connectée à deux réservoirs dont les coordonnées sont  $x = x^U = 0$  et  $x = x^D$  et leurs pressions manométriques respectives  $H1 = 25m$  et  $H2 = 20m$ .

Le facteur de friction Darcy-Weisbach est  $f = 0.02$ , le débit en régime permanent est  $Q_0^L = 0.0153m^3/s$ , et la vitesse de l'onde est  $a = 1200m/s$ .  $H_0^L$  représente la pression manométrique en régime permanent. Il y'a trois capteurs (des hydrophones) situés stratégiquement au début et à la fin de la canalisation. Leurs coordonnées sont représentées par  $x_M$ .

On considère  $x^{L_n^*}$  les positions réelles des fuites et  $x^{L_n}$  les positions potentielles (avec  $n = 1, \dots, N$ ). La taille de la fuite est donnée par  $s^L = C^d A^L$  avec  $C^d$  le coefficient de débit de la fuite et  $A^L$  la surface de l'orifice [47].

### 4 Présentation du modèle

Afin d'obtenir le modèle de l'état transitoire du système, les auteurs ont procédé comme suit :

Après avoir perturbé le système avec une onde de pression causée par le mouvement d'ouverture de la vanne, les variations du débit et de la pression manométrique dues à la dynamique du fluide sont représentées par  $q$  et  $h$ . Les équations de continuité et de quantité de mouvement linéarisée selon Chaudhry [10] sont :

$$\begin{aligned} \frac{a^2}{gA} \frac{\partial q}{\partial x} + \frac{\partial h}{\partial t} &= 0 \\ \frac{1}{gA} \frac{\partial q}{\partial t} + \frac{\partial h}{\partial x} + Rq &= 0 \end{aligned} \quad (2.2)$$

avec :

$a$  : la vitesse d'onde.

$A$  : la surface de la canalisation.

$R$  : le terme de la résistance en régime permanent.

$q$  : le débit.

$g$  : l'accélération gravitationnelle.

$h$  : la pression manométrique.

En considérant  $q(x^U)$  et  $h(x^U)$ , les quantités en  $x^M$  peuvent être représentées tel que :

$$\begin{pmatrix} q(x^M) \\ h(x^M) \end{pmatrix} = \left( M^{NL}(x^M) + s^L M^{SL}(x^L) \right) \begin{pmatrix} q(x^U) \\ h(x^U) \end{pmatrix}, \quad (2.3)$$

avec  $M^{NL}$  la matrice de transfert qui concerne le cas d'une canalisation sans fuites (no leaks) et  $M^{SL}$  est en fonction d'une seule position  $x^L$  (single leak) mais est indépendante de la taille de la fuite.

Le cas pour plusieurs fuites est présenté comme suit [48] :

$$\begin{pmatrix} q(x^M) \\ h(x^M) \end{pmatrix} = \left( M^{NL}(x^M) + \sum_{n=1}^N s^L M^{SL}(L_n, x^{L_n}, x^M) \right) \begin{pmatrix} q(x^U) \\ h(x^U) \end{pmatrix} \quad (2.4)$$

Il est noté que nous nous sommes intéressées au cas où  $N=2$ , donc la détection d'un couple de fuites.

La formule de la pression manométrique en présence de bruit blanc gaussien est la suivante :

$$h(\omega_j, x_M) = h^{NL}(\omega_j, x_M) + \sum_{n=1}^N s^{L_n} G(\omega_j, x^{L_n}, x_M) + n_{jM} \quad (2.5)$$

avec :

$G$  : la fonction qui contient les fuites potentielles qu'on cherche à faire coïncider avec les véritables fuites.

$n$  : un bruit blanc gaussien centré en 0 et de variance  $\sigma^2$ .

Cette étude est faite spécifiquement pour bruit blanc et tout autre bruit devra y être assimilé.

Si on considère :

$$\Delta h_{jm} \equiv h(\omega_j, x_M) - h^{NL}(\omega_j, x_M) \quad (2.6)$$

alors

$$\Delta h = G(x^L) s^L + n \quad (2.7)$$

$G_n$  une matrice de dimension  $JM \times N$  dont la  $n^{ime}$  colonne correspond à :

$$G_n(x^{L_n}) = (G(\omega_1, x^{L_n}, x_1), \dots, G(\omega_J, x^{L_n}, x_1), \dots, G(\omega_1, x^{L_n}, x_M), \dots, G(\omega_J, x^{L_n}, x_M))^T \quad (2.8)$$

avec :

$J$  : le nombre de fréquences,  $M$  : le nombre de capteurs et  $N$  : le nombre de fuites.

## 5 Matched field processing (MFP)

Dans un premier article [48], le problème de la localisation des fuites est résolu en utilisant la méthode du matched field processing (MFP), c'est une technique de traitement de signal qui permet d'estimer les paramètres du modèle en corrélant les réponses modélisées aux données observées [45].

Comme cité plus haut, la fonction objectif (2.1) utilisée par Ligett et Chen (1994) n'est pas robuste face aux bruits. Les travaux de X.Wang et M.Ghidaoui tentent de remédier à ce problème.

L'idée est de maximiser le rapport signal-bruit (SNR), ce qui revient à augmenter le taux d'information du signal de l'onde hydraulique tout en minimisant le bruit. Puisque utiliser des ondes aux amplitudes élevées comme excitation n'est pas conseillé, notamment parce que ça pourrait compromettre l'intégrité de la canalisation ; ils ont pensé à traiter le signal avec une méthode numérique : MFP, qui dans notre cas [47] revient à se servir d'un filtre optimal pour maximiser SNR. De plus, cette méthode exploite toutes les fréquences et pas seulement les

fréquences résonantes, ce qui signifie qu'elle a plus d'informations pour établir une localisation précise des fuites.

Le principe de MFP revient à ajuster un vecteur unitaire  $\omega = (\omega_1, \dots, \omega_J)^T$  avec  $|\omega| = 1$  afin qu'il ait la même direction que la mesure. Le résultat  $B$  est défini par le produit entre le vecteur de pondération  $\omega$  et la différence de pression manométrique  $\Delta h$ .

$$B \equiv \langle \omega, \Delta h \rangle = \omega^H \Delta h \quad (2.9)$$

H désigne la transconjuguée. Le poids optimal est obtenu en maximisant :

$$|B|^2 = |\omega^H \Delta h|^2 = \omega^H \Delta h \Delta h^H \omega \quad (2.10)$$

Après des calculs laborieux, le résultat obtenu est :

$$\hat{\omega} = \operatorname{argmax} E(|B|^2) = \operatorname{argmax}((s^L)^2 \omega^H G G^H \omega + \sigma^2) = \operatorname{argmax}(\omega^H G G^H \omega) = \pm \frac{G}{\sqrt{G^H G}} \quad (2.11)$$

Ce même résultat peut être obtenu en cherchant à appliquer un filtre à la différence de pression manométrique.

$$\omega^H \Delta(h) = s^L \omega^H G + \omega^H n \quad (2.12)$$

Et on obtient le filtre  $\omega^H$  en maximisant SNR.

La méthode de MFP est limitée par la condition d'échantillonnage de Shannon-Nyquist, ainsi différencier deux fuites proches séparées par une distance inférieure à  $0.5\lambda_{\min}$  (avec  $\lambda_{\min}$  la longueur d'onde minimale) devient impossible.

Pour illustrer la limitation de MFP, on considère une canalisation de 2 km avec un couple de fuites éloignées  $x^{L_1} = 300m$  et  $x^{L_2} = 700m$  (figure 2.1).

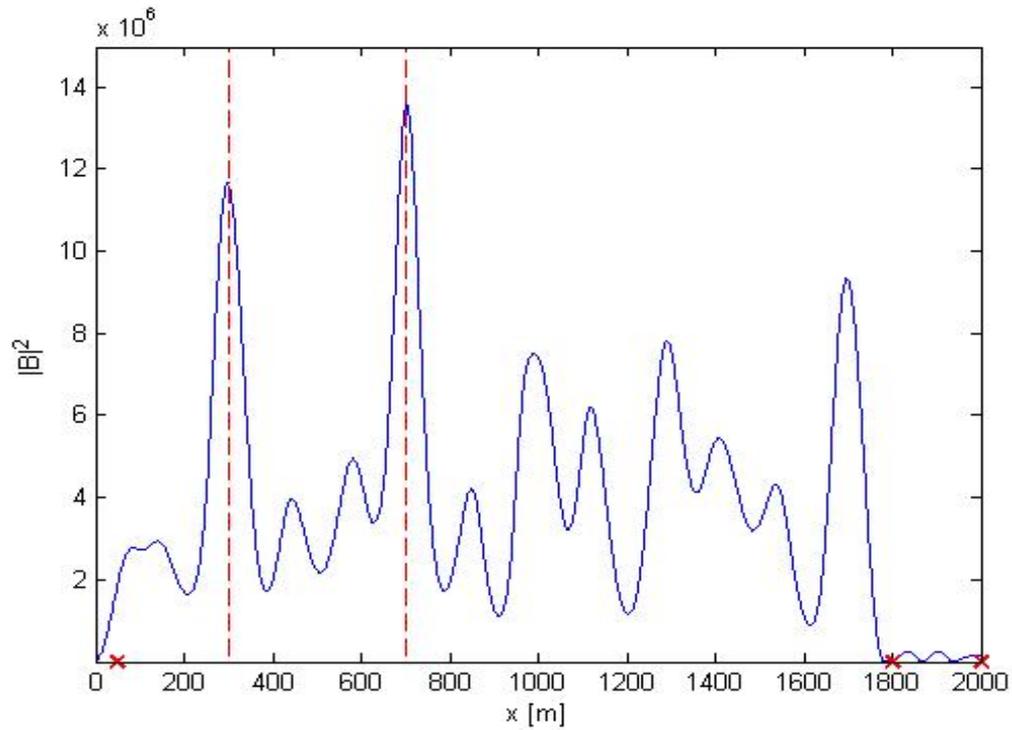


FIGURE 2.1 – Localisation de deux fuites éloignées en utilisant MFP

On répète la même opération mais pour deux fuites proches avec  $x^{L1} = 400m$  et  $x^{L2} = 460m$  ( $|x^{L1} - x^{L2}| = 60m = 0.23\lambda_{\min}$ ) (figure 2.2).

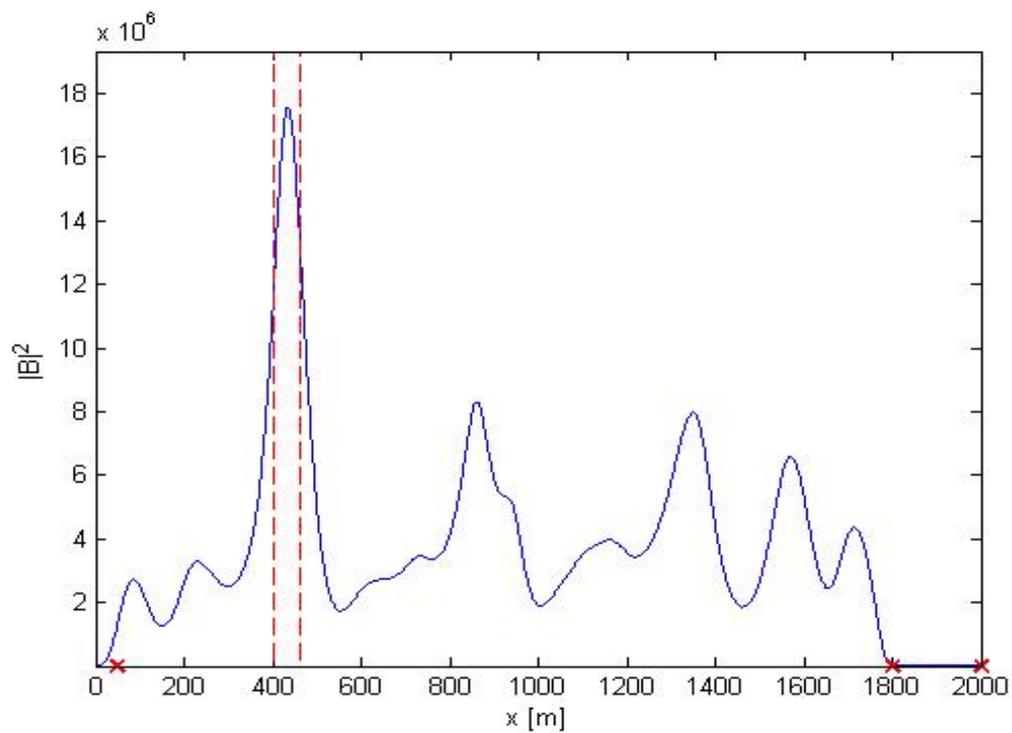


FIGURE 2.2 – Localisation de deux fuites proches en utilisant MFP

La distance entre les deux fuites étant inférieure à la moitié de la longueur d'onde minimale, MFP ne peut les localiser séparément.

## 6 L'Estimateur du maximum de vraisemblance (EMV)

Pour contourner ce problème, les auteurs ont présenté un deuxième article [47] où la détection de fuite est faite grâce à la méthode de l'estimation du maximum de vraisemblance (EMV).

Considérée comme une technique de super-résolution, cette méthode devrait être capable d'identifier plusieurs fuites proches même lorsque la distance les séparant est inférieure à la condition de Shannon-Nyquist, et cela sans sacrifier l'efficacité, la robustesse (vis-à-vis du SNR/bruit) ou la précision de la méthode. En effet, elle élargit le champ de localisation (la résolution) de 35%.

Afin de montrer que la méthode EMV renvoie une estimation de super-résolution, nous allons reprendre l'exemple cité en haut où les fuites sont proches, c'est à dire  $x^{L_1} = 400m$  et  $x^{L_2} = 460m$  ( $|x^{L_1} - x^{L_2}| = 60m = 0.23\lambda_{\min}$ ).

On remarque que la condition de Shannon-Nyquist n'est pas respectée, mais que la localisation est précise, ce qui revient à dire que l'estimateur de vraisemblance propose une meilleure résolution que celle imposée par la condition de Shannon-Nyquist.

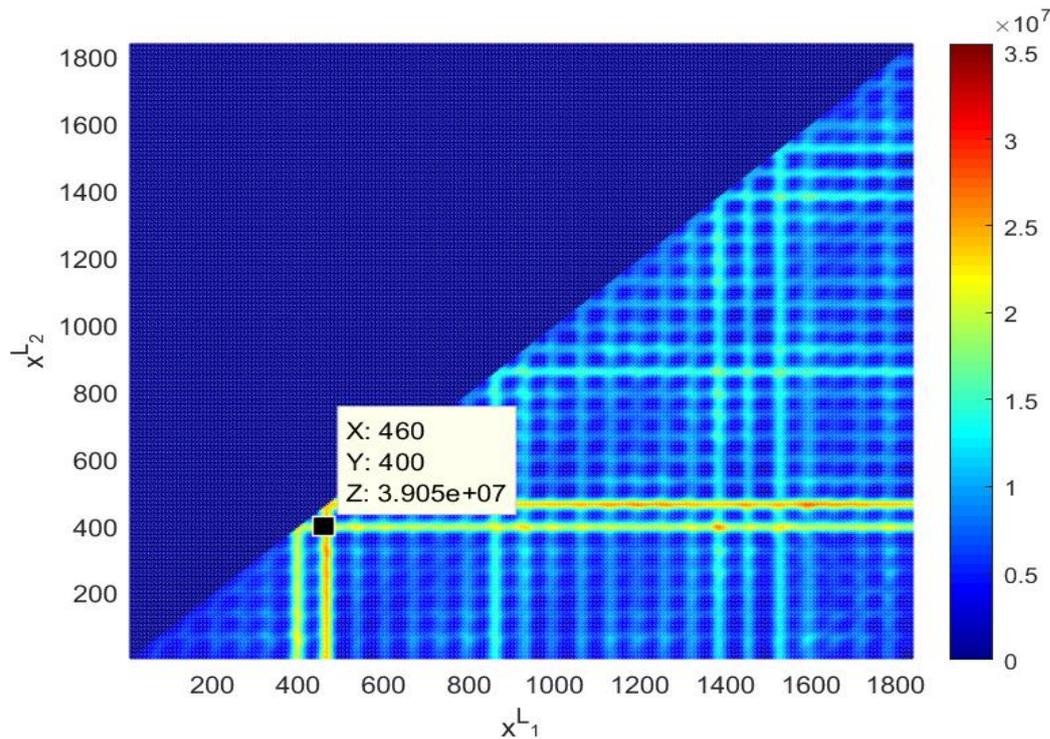


FIGURE 2.3 – Localisation de deux fuites proches en utilisant EMV

## 6.1 Le principe de l'EMV

L'estimateur du maximum de vraisemblance peut être appliqué pour la plupart des problèmes, il a un fort attrait intuitif et donne souvent une estimation raisonnable des paramètres  $\theta$  (dans notre cas  $\hat{x}^L$ ). De plus, si l'échantillon est grand, la méthode donnera une excellente estimation de  $\theta$ . Pour ces raisons, EMV est la méthode d'estimation la plus utilisée en statistiques[26].

Pour mettre en œuvre cet estimateur, l'expression mathématique de la fonction densité de probabilité  $p(\Delta(h); x^L, s^L)$  des données échantillonnées (qu'on nommera fonction de vraisemblance) est présentée comme suit [47] :

$$p(\Delta h; x^L, s^L) = (\pi\sigma^2)^{-JM} \exp\left(-\frac{1}{\sigma^2} \|\Delta h - G(x^L)s^L\|^2\right) \quad (2.13)$$

Afin de simplifier les calculs, il est plus commode d'utiliser le logarithme de la fonction de vraisemblance. (Puisque la fonction logarithme est croissante, le maximum de la densité de probabilité et de la vraisemblance logarithmique coïncident).

$$\log L(x^L, s^L; \Delta h) = -JM \log(\pi\sigma^2) - \frac{1}{\sigma^2} \|\Delta(h) - G(x^L)s^L\|^2 \quad (2.14)$$

Cette fonction sera maximisée par un algorithme de recherche exhaustive, (dans cette étude il s'agit de trouver le paramètre  $x_L$  qui représente la position des fuites).

$$\begin{aligned} \hat{x}^L &= \underset{x^L}{\operatorname{argmin}} \|\Delta h - G(x^L)(G^H(x^L)G(x^L))^{-1}G^H(x^L)\Delta h\|^2 \\ &= \underset{x^L}{\operatorname{argmax}} (\Delta h^H G(x^L)(G^H(x^L)G(x^L))^{-1}G^H(x^L)\Delta h) \end{aligned} \quad (2.15)$$

## 6.2 La borne inférieure de Cramer-Rao

Afin de mesurer l'efficacité de l'estimateur, la notion de la borne inférieure de Cramer-Rao (CRLB) est introduite. Elle correspond dans notre cas à la distance minimale acceptable entre deux fuites pour que l'estimation soit correcte.

La borne de Cramer-Rao est inversement proportionnelle à la matrice d'information de Fisher, cette dernière représente le taux d'informations qui parvient à l'estimateur. Donc forcément si ce taux est élevé on a une meilleure précision [39].

Elle s'écrit comme suit [47] :

$$I(\Theta) = -E\left(\frac{\partial^2 \log p(x|\Theta)}{\partial \Theta^2}\right) \quad (2.16)$$

Intuitivement la précision de l'estimation de  $\hat{\theta}$  (ici notée :  $\hat{x}^L$ ) dépend de la courbure de la fonction de vraisemblance logarithmique[34].

Si la fonction de vraisemblance logarithmique est très courbée ou raide autour de  $\hat{\theta}$ , alors  $\theta$  sera estimé avec une grande précision, ce qui correspond a un taux d'information élevé. En revanche, une fonction de vraisemblance logarithmique plate ou qui révèle une petite courbure près de  $\hat{\theta}$ , donnera une estimation peu précise de  $\theta$ .

Dans le cas étudié, la borne de Cramer-Rao s'écrit comme suit :

$$Cov(\hat{\Theta}) \geq (TI\Theta)^{-1} \quad (2.17)$$

avec :  $T$  le temps d'échantillonnage et  $\Theta = (\theta_1, \dots, \theta_n)$ .

Les auteurs [47] sont parvenus à ces résultats :

- L'estimation de  $x_L$  est jugée fiable si la distance entre deux fuites est supérieure à 20 m dans le cas d'une longueur de 2 km ; c'est-à-dire  $0.15\lambda_{min}$ .
- L'utilisation de plus de fréquences, ainsi que de plusieurs capteurs, donnera une meilleure estimation car avec l'augmentation du taux d'information disponible, la valeur de la CRLB diminue.

La figure 2.4 présente le cas où la CRLB est respectée. Le pic représentant la fuite est facilement distinguable. La figure 2.5 illustre le cas où la CRLB n'est pas respectée et où il est difficile de distinguer la fuite.

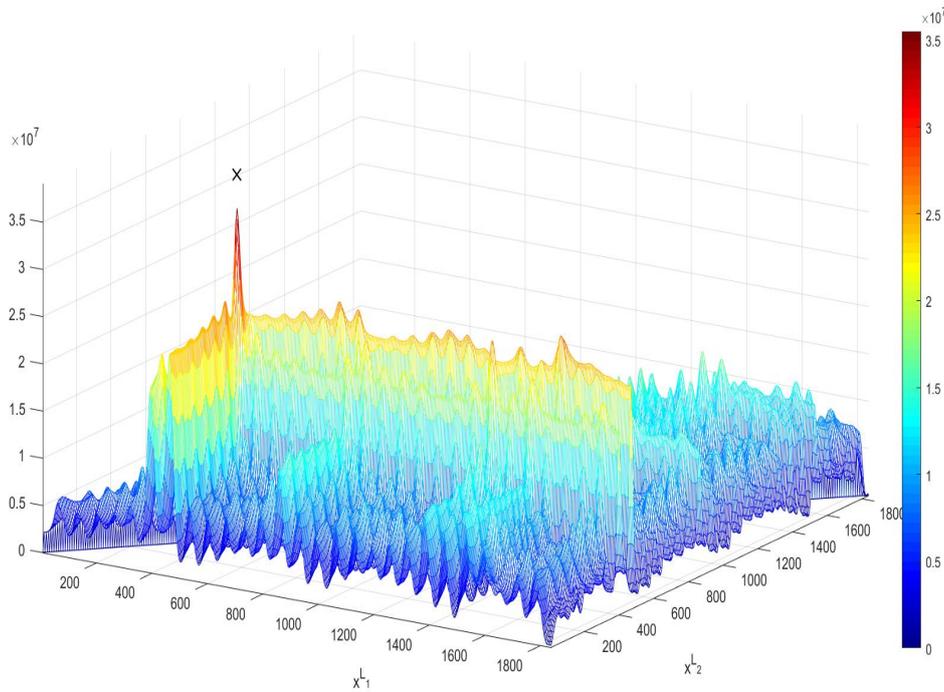


FIGURE 2.4 – Le cas où EMV satisfait la CRLB

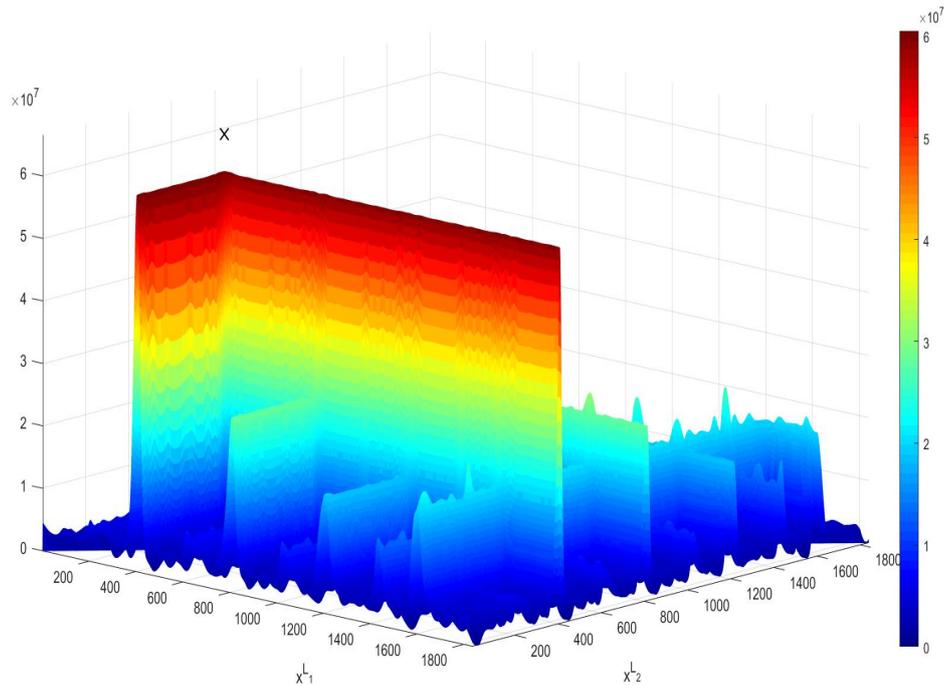


FIGURE 2.5 – Le cas ou EMV ne satisfait pas la CRLB

### 6.3 Équivalence entre MFP et EMV

MFP est une méthode de traitement de signal utilisée majoritairement à des fins de localisation de sources dans des environnements bruyants [1] [16], elle apparaît dans plusieurs domaines tels que l’océanographie et la sismologie. Dans notre cas elle emploie un algorithme de recherche exhaustive de **la meilleure correspondance** entre les réponses modélisées et celles mesurées par les hydrophones, ce qui équivaut à une estimation du maximum de **vraisemblance** [29].

## 7 Conclusion

La maximisation de la fonction de vraisemblance est réalisée par une recherche exhaustive, c’est-à-dire que l’algorithme va parcourir tout le champ des solutions possibles et attribuer un score à chaque possibilité. La valeur maximale correspond aux bonnes estimations des positions des fuites. Cette méthode donne des résultats corrects mais en contrepartie le temps d’exécution devient très important à mesure que la complexité du problème augmente.

Afin de palier à cet inconvénient, l’utilisation d’une métaheuristique à la place de l’algorithme de recherche pourrait être une solution adéquate. En effet, elles sont réputées pour obtenir des résultats acceptables en un temps réduit grâce à une meilleure stratégie de recherche de l’espace des solutions.

# Chapitre 3

## Les Algorithmes Génétiques

### 1 Introduction

Dans ce chapitre, nous allons nous intéresser à la méthode d'intelligence artificielle que nous avons adapté à notre problème : les algorithmes génétiques. Une métaheuristique qui ne cesse d'augmenter en popularité et qui est utilisée très fréquemment dans les problèmes d'ingénierie requérant l'optimisation d'une fonction dont l'espace de recherche est important, comme dans notre cas.

### 2 Optimisation

L'optimisation est une branche des mathématiques. Il s'agit de modéliser un problème concret, l'analyser puis le résoudre de façon analytique ou numérique.

#### 2.1 Définition :

On appelle problème d'optimisation en dimension fini, tout problème de recherche des points  $x'$  dans  $R^n$  qui vérifient :

$$f(x') = \min_{x \in C} f(x)$$

ou

$$f(x') = \max_{x \in C} f(x)$$

C'est donc la recherche des points optimaux, qu'ils soient minimums ou maximums, locaux ou globaux selon le problème considéré et les contraintes qui lui sont propres.

#### 2.2 Terminologie :

$f(x)$  : fonction objectif à optimiser.

$x'$  : solution du problème.

$C$  : ensemble des contraintes. S'il n'existe pas de contraintes alors  $C = R^n$ .

#### 2.3 Types de problèmes d'optimisation :

Il existe deux grandes familles : l'optimisation continue et l'optimisation discrète. Cette classification dépend de la nature des variables impliquées dans la fonction objectif.

- Optimisation continue : si l'ensemble des valeurs acceptables des variables est continu et donc infini.
- Optimisation discrète : Si l'ensemble des valeurs acceptables des variables est fini ou dénombrable.

A noter qu'il est possible d'avoir des problèmes mixtes puisque la limitation des variables est souvent associée aux contraintes imposées par le problème. [41]

## 2.4 Classes de difficulté :

Les méthodes d'optimisation sont nombreuses et diverses ; chacune s'adaptant à un niveau de difficulté. Il existe différentes façons de classifier la complexité des problèmes d'optimisation, nous nous contenterons de présenter celles concernant les problèmes de décision.[43]

Un problème d'optimisation peut être assimilé à un problème de décision qui tranche si oui ou non il existe une meilleure solution que l'actuelle. Ils partagent donc la même complexité, qui est déterminée par les classes suivantes :

**Classe P** : regroupe les problèmes solvables par un algorithme déterministe de résolution polynomiale.

**Classe NP** : regroupe les problèmes solvables par un algorithme non-déterministe de résolution polynomiale. S'il est possible de résoudre lesdits problèmes dans un temps polynomial on parle de NP-complets, dans le cas inverse, on parle de problèmes NP-difficiles.

A savoir que les problèmes NP-difficiles sont NP-complets mais la réciproque est fausse.

## 2.5 Types de méthodes d'optimisation

Il existe plusieurs classifications, celle de Dréo [25] suit cette hiérarchie :

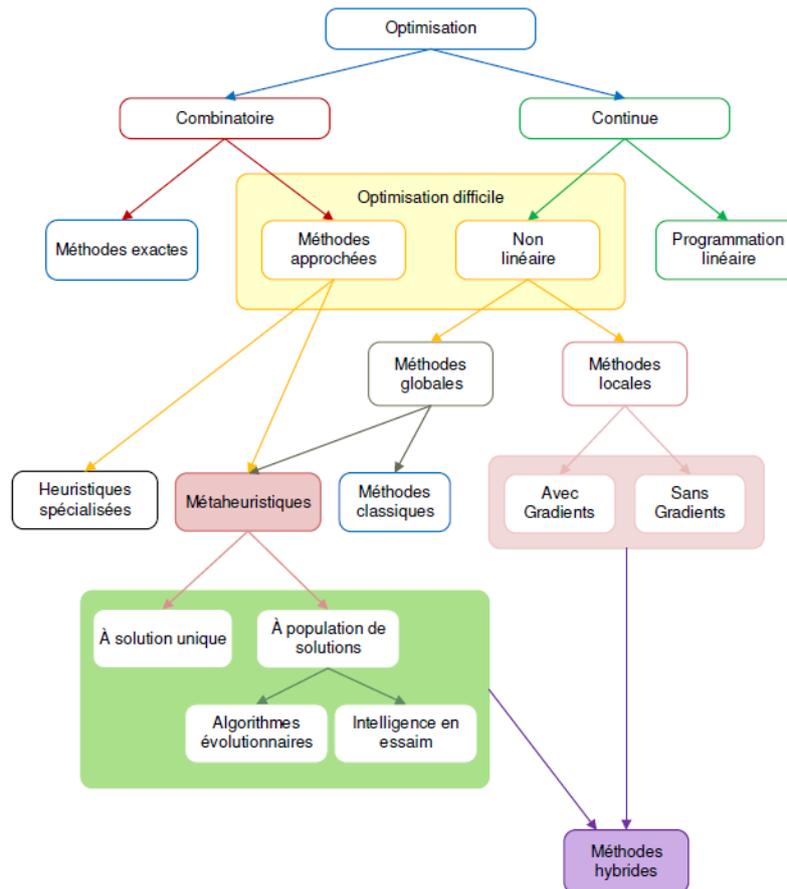


FIGURE 3.1 – Classification des méthodes d’optimisation [25]

Les méthodes de l’optimisation continue sont composées de la programmation linéaire et non linéaire; tandis que les méthodes de l’optimisation discrète (combinatoire) sont divisées entre les méthodes exactes et les méthodes approchées. On trouve en commun entre les deux branches, les métaheuristiques.

Nous détaillons ci-dessous les techniques de l’optimisation combinatoire.

## 2.6 Les méthodes exactes :

Ces méthodes procèdent à l’exploration complète de l’espace des solutions. Elles finissent toujours par trouver l’optimum ou bien par réfuter son existence. Ces méthodes risquent de se heurter au problème de l’explosion combinatoire dans le cas où la taille de l’espace de recherche est trop importante; autrement dit, le temps de résolution devient excessivement long.

## 2.7 Les méthodes approchées :

Afin de contrer l’explosion combinatoire, ces méthodes explorent l’espace de recherche de façon intuitive et pas forcément en son entièreté. Il est possible qu’en négligeant une partie des solutions, elles ratent l’optimum; cependant, elles arrivent souvent à un résultat qui est suffisamment proche et cela en un temps raisonnable. Elles sont divisées en deux familles : les heuristiques et les métaheuristiques.

*Les Heuristiques* : d'origine grec, le mot heuristique signifie « je trouve », connu plus largement grâce au fameux « Eureka » d'Archimède. Une heuristique est un algorithme qui repère une solution acceptable en un temps raisonnable – donc, polynomial – mais qui ne garantit pas l'optimalité de ladite solution. Sa façon d'explorer l'espace de recherche est construite selon les contraintes du problème étudié, elle lui est donc propre et ne peut pas être généralisée pour d'autres cas [6].

*Les Métaheuristiques* : c'est là qu'interviennent les métaheuristiques. Méta signifie « au-delà » en grec, et effectivement ces algorithmes ont une structure plus globale que les heuristiques. Au lieu de suivre des contraintes spécifiques, la diversification des solutions est assurée par des « opérateurs » inspirés de phénomènes naturels tels que la théorie de l'évolution (algorithmes évolutionnaires), la métallurgie (recuit-simulé) ou encore l'éthologie (PSO, algorithme des chauves-souris) [6].

### 3 Les Métaheuristiques

Les métaheuristiques (MH) sont des algorithmes d'optimisation stochastiques et itératifs capables de faire une exploration intelligente de l'espace des solutions en utilisant des approches statistiques ne nécessitant pas le calcul des dérivés de la fonction coût [40].

Un des points intéressants des MH est qu'elles sont capables de s'adapter aux problèmes discrets et aussi aux problèmes continus. La caractéristique intelligente des métaheuristiques est représenté par leur capacité à diversifier et intensifier leur manière d'exploiter l'espace de solutions.

#### 3.1 Caractéristiques

- Elles sont, au moins en partie, stochastiques : ce qui permet de faire face à l'explosion combinatoire des solutions possibles.
- Généralement d'origine discrète, elles ont l'avantage, décisif dans le cas continu d'être directes (ne nécessitent pas le calcul des dérivés.)
- Inspirées par des analogies, avec la physique (recuit simulé), la biologie (algorithmes évolutionnaires, recherche tabou) ou avec l'éthologie (colonie de fourmis, colonie d'abeilles.)
- Leur inconvénient majeur est la difficulté de réglage des paramètres [41].

#### 3.2 Classification

Il existe plusieurs façons de classer les métaheuristiques, l'une d'entre elle se base sur la stratégie d'évolution des solutions.

- *Métaheuristiques à solution unique* : aussi appelées méthodes de trajectoire ; elles manipulent une seule solution initiale et la font évoluer à travers les itérations, créant ainsi une trajectoire vers la meilleure solution. Des exemples de ces métaheuristiques sont la méthode de descente, le recuit simulé [28] et la recherche par tabous [17].

- *Métaheuristique à population* : manipulent un groupe de solutions initiales et les font évoluer à travers les itérations. La taille de la population est fixée par l'utilisateur et ne subit pas de modifications. Il faut noter qu'à la fin, nous retrouvons une seule solution : la meilleure. Ces méthodes sont divisées en deux grandes familles :
- Les algorithmes évolutionnaires (les algorithmes génétiques [18], les stratégies d'évolution, la programmation évolutive, et la programmation génétique).
- Les algorithmes d'intelligence par essaim (l'optimisation par colonie d'abeilles , les systèmes immunitaires artificiels, l'optimisation par essais particuliers. . .)

## 4 L'Algorithme Génétique

### 4.1 Origine

proposé par J. Holland en 1972 puis popularisé par Goldberg en 1988 [18], les algorithmes génétiques découlent des algorithmes évolutionnaires. Leur fonctionnement est inspiré de la théorie de l'évolution de Charles Darwin, elles reprennent le principe de la sélection naturelle et du besoin d'adaptation pour la survie des espèces [21].

### 4.2 Analogie avec la biologie

Les algorithmes génétiques reprennent plusieurs termes issus de la biologie mais adaptés à l'optimisation mathématique.

Un individu – ou solution – est représenté par un **chromosome**. Ce dernier est constitué de plusieurs **gènes** qui sont les valeurs de chaque solution. Chaque gène a une position précise dans le chromosome qui est appelée **locus**.

Au lieu d'itérations on parle de **générations**, et chaque génération de solution laisse place à une autre plus adaptée à son environnement. En effet, la théorie de Darwin stipule que seuls les individus les plus forts survivent et peuvent se reproduire, créant ainsi une progéniture encore plus robuste, ce qui revient à dire des solutions toujours plus proches de l'optimum.

La **force** d'un individu dans l'algorithme génétique est représentée par sa "fitness". Selon que l'on ait à faire à un problème de minimisation ou de maximisation, on définit ce que l'on entend par une meilleure fitness.

### 4.3 Etapes de l'Algorithme

Les fondements théoriques des AG ont été détaillés dans l'œuvre de David Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning (1989)*.

*Type de codage* : Avant de procéder aux étapes de l'AG, il faut tout d'abord choisir une représentation pour les chromosomes. Selon les problèmes, ça peut être une représentation binaire, réelle ou en base n.

*La première étape* est la génération de la population initiale, aussi appelée la genèse. Elle peut être faite de façon aléatoire, ou si l'on a une bonne connaissance de l'espace de recherche, on peut rajouter des contraintes pour mieux orienter la naissance d'individus proches de la solution. Une fois la génération créée, la fitness de chaque solution est évaluée.

C'est à partir de *la deuxième étape* que l'on commence l'itération. Des individus sont sélectionnés pour servir de parents à la prochaine génération, pour cela, il y'a plusieurs méthodes qui le plus souvent retiennent les meilleurs éléments.

Lors de *la troisième étape*, une population d'enfants est créée à partir de la recombinaison des gènes des parents suivant une probabilité de croisement ; sinon on reprend les parents comme individus dans la nouvelle génération.

*La quatrième étape* est celle de la mutation. Une mutation génétique est une modification rare de l'information génétique, et peut toucher une ou plusieurs parties du génome. Dans l'algorithme génétique, sa rareté est représentée par une probabilité de mutation assez basse – bien que cela dépende des applications -. Il y'a ici aussi plusieurs méthodes pour procéder à la modification d'un gène.

A partir de là, soit une condition d'arrêt est satisfaite et on récupère la meilleure solution de la dernière génération, ou alors on réitère depuis la deuxième étape [7].

## 4.4 Les opérateurs de l'algorithme génétique

La fonction d'adaptation (fitness) : C'est la fonction qui permet d'évaluer les individus afin de leur attribuer un score selon qu'ils soient proches ou éloignés de la solution optimale. Elle est souvent assimilée à la fonction objectif.

Sélection : c'est un algorithme qui va sélectionner des individus - selon leur fitness - pour se reproduire et ainsi sauvegarder leur patrimoine génétique, c'est-à-dire des parents. Il existe plusieurs méthodes pour ce faire ; sélection par tournoi, sélection par probabilités et élitisme.

Croisement : Cet opérateur est inspiré du processus de reproduction biologique. Après que deux individus soient sélectionnés dans la population sélectionnée. Deux autres chromosomes seront créés à partir du mélange des gènes des parents. Le but du croisement est de diversifier la population, en produisant de nouvelles solutions à partir d'individus considérés forts.

Il existe plusieurs méthodes pour ce faire : croisement à un point, le croisement multi-points et le croisement arithmétique (ou barycentre).

Mutation : La mutation consiste à modifier aléatoirement un ou plusieurs gènes d'un chromosome selon une probabilité de mutation définie à l'avance et qui est souvent choisie très basse. En effet, tout comme dans la génétique, la mutation dans l'algorithme génétique est censée être un phénomène rare qui permet la création d'individus inédits. Il existe plusieurs méthodes : dans le cas de la représentation binaire : la méthode bit-flip, et pour la représentation réelle : la mutation uniforme et la mutation non-uniforme.

La mutation permet d'atteindre la propriété d'ergodicité, elle garantit un balayage de tout l'espace recherche ; donc plus de chances de trouver l'optimum globale [19].

## 4.5 Paramètres de l’algorithme génétique

Nombre d’individus : En règle générale, ce nombre dépend de la taille de l’espace de recherche. Plus l’espace est grand plus on a besoin d’une population fournie afin d’en faire le tour.

Nombre de générations : Ou nombre d’itérations ; plus on augmente ce nombre, plus la solution a de chances pour converger. Le temps d’exécution en dépend fortement, il faut donc préférer un nombre qui réalise une solution approchée dans un temps raisonnable.

Probabilité de croisement : Souvent fixé à 0.7, on considère que plus ce taux est élevé, plus les solutions sont diversifiées et la convergence de la solution est assurée.

Probabilité de mutation : On considère que plus ce taux est petit, mieux c’est. La mutation doit rester un phénomène rare.

Dans certains cas où l’espace de recherche est très grand, on peut avoir recours à une probabilité de mutation élevée afin de pouvoir prétendre à *l’ergodicité*.

Critère d’arrêt : Si on a déjà une idée de l’optimum, alors l’algorithme ne s’arrête que lorsque ses solutions s’en approchent (ou l’égalent). Sinon, on peut soit définir un temps d’exécution maximal ou un certain nombre de générations à atteindre.

## 4.6 Avantages de l’algorithme génétique

- Peut s’adapter à différent problèmes.
- Trouve des solutions approchées en un temps raisonnable.
- Est moins susceptible de bloquer dans des optimums locaux.

## 4.7 Inconvénients de l’algorithme génétique

- Impossibilité de s’assurer que la solution trouvée est bien l’optimum (sauf si connaissance préalable de ladite solution).
- Le réglage des paramètres de l’algorithme afin d’avoir de bons résultats peut prendre un certain temps.

# 5 Conclusion

L’algorithme génétique est une métaheuristique qui offre de nombreux avantages, le plus important étant le temps d’exécution réduit et sa souplesse pour éviter les optimums locaux.

Nous avons trouvé que cette méthode est adaptée à notre problème, et présenterons dans le prochain chapitre le squelette de l’algorithme que nous avons implémenté ainsi que l’interprétation des résultats obtenus.

# Chapitre 4

## Application de l'Algorithme Génétique sur le modèle

### 1 Introduction

Dans ce chapitre, nous allons décrire l'algorithme génétique développé en expliquant le codage et les opérateurs génétiques adoptés. Ensuite, nous présentons et interprétons les résultats obtenus pour différentes instances du problème étudié..

### 2 Description de l'Algorithme Génétique

#### 2.1 Codage des solutions

Parmi les trois types de codage existants pour l'Algorithme Génétique, le codage réel s'est imposé puisque nous avons seulement deux paramètres exprimés en mètre et représentés par des réels. Le chromosome est composé de deux gènes : la position de la première fuite (x1) et celle de la deuxième (x2).

x1	x2
----	----

La valeur du premier gène est choisie suivant un vecteur de valeurs équidistantes distribuées sur le long de l'intervalle imposé par la longueur de la canalisation.

Celle du deuxième gène est choisie aléatoirement entre deux valeurs : cette position doit être supérieure à la première avec une distance suivant la limite de l'estimateur par maximum de vraisemblance ( $0.15\lambda_{min}$ ), et inférieure à la position du premier hydrophone en aval de la canalisation.

*Remarque* : La valeur de  $\lambda_{min}$  est proportionnelle à la longueur de la canalisation.

#### 2.2 Fonction Objectif

Comme ce fut expliqué dans le chapitre précédent, nous avons pris comme fonction coût l'estimation d'un couple de fuites par maximum de vraisemblance comme détaillé dans l'article de Wang et Ghidaoui. [47]

Elle prend comme entrées deux fuites potentielles et retourne un coût. Notre objectif est trouver le couple de positions pour lequel la fonction est maximale.

$$\begin{aligned}\hat{x}^L &= \underset{x^L}{\operatorname{argmin}} \|\Delta h - G(x^L)(G^H(x^L)G(x^L))^{-1}G^H(x^L)\Delta h\|^2 \\ &= \underset{x^L}{\operatorname{argmax}} (\Delta h^H G(x^L)(G^H(x^L)G(x^L))^{-1}G^H(x^L)\Delta h)\end{aligned}\tag{4.1}$$

avec

$$x^L = (x_1, x_2)$$

## 2.3 Opérateurs génétiques adoptés

Comme expliqué dans le chapitre précédent, l'AG comporte trois types d'opérateurs qui permettent aux générations d'évoluer et de se diversifier à travers les itérations.

**L'opérateur de sélection** : permet de sélectionner les individus qui vont faire l'objet de l'opération de croisement pour donner naissance à la prochaine génération. Dans notre AG, nous avons utilisé la sélection binaire par tournoi.

*la sélection binaire par tournoi* : Soit  $N$  le nombre d'individus de la génération, il faut en sélectionner  $\frac{N}{2}$  puisque chaque paire de parents va engendrer deux enfants.

Dans le cas de la sélection binaire, deux individus sont choisis au hasard et seul celui qui présente la meilleure fitness peut accéder à la génération intermédiaire.

Cette étape est répétée jusqu'à ce que la génération intermédiaire soit remplie ( $N$  composants). Il est tout à fait possible que les mêmes individus participent à plusieurs tournois : s'ils gagnent plusieurs fois, ils auront donc le droit d'être autant de fois dans la génération intermédiaire, ce qui favorisera la pérennité de leurs gènes.

**L'opérateur de croisement** : permet de créer de nouveaux individus à partir de la combinaison des gènes des parents sélectionnés en amont (population intermédiaire). Nous avons utilisé le croisement arithmétique (barycentre).

*Croisement arithmétique* : Si les croisements en un point/multi-points peuvent être implémentés pour n'importe quel choix de codage. Il est plus intéressant de procéder au croisement arithmétique lorsqu'il s'agit d'une représentation réelle.[31]

On définit un paramètre de pondération aléatoire alpha, souvent prit entre [0 1]. Puis l'on sélectionne des gènes  $p1(i)$  et  $p2(i)$  dans chaque parent au même locus  $i$  et on crée deux nouveaux gènes par combinaison linéaire.

**L'opérateur de mutation** : permet de modifier les valeurs d'un ou de plusieurs gènes, juste suffisamment pour avoir une nouvelle solution. Nous avons utilisé une mutation non uniforme, avec une fonction définissant l'écart de mutation présentée dans un article de Cordon et Herrera en 2001. [36].

*La mutation non uniforme* : Après avoir tiré au sort quels gène(s) vont être modifiés, un taux de mutation est calculé ce taux sera à son maximum durant les premières itérations et

selon le moment où l'on veut stopper la mutation, il décroîtra jusqu'à devenir nul. [8]

La fonction définissant le taux de mutation non uniforme est présentée comme suit :

$$\Delta(t, y) = y * (1 - r^{(1-t/T)^\beta}) \quad (4.2)$$

$$x'_k = \begin{cases} x_k + \Delta(t, x_k^{max} - x_k) & si \\ rand = 0 \\ x_k - \Delta(t, x_k - x_k^{min}) & si \\ rand = 1 \end{cases} \quad (4.3)$$

$T$  : l'indice de la génération pour laquelle on cesse la mutation.

$t$  : l'indice de la génération en cours de traitement.

$r$  : un nombre aléatoire dans l'intervalle  $[0, 1]$ .

$\beta$  : paramètre de l'opérateur de mutation.

$y$  : l'écart entre la position à muter et la limite inférieure/supérieure de l'intervalle de recherche selon la décision binaire prise.

$rand$  : un nombre aléatoire dans l'intervalle  $[0, 1]$

## 2.4 Procédure de vérification de contraintes et de réparation des solutions infaisables

Les contraintes qui sont imposées par le système et par la méthode de EMV sont :

1) La distance entre deux fuites doit être au minimum  $0.15\lambda_{min}$  pour que l'estimateur donne un résultat fiable.

2) Les solutions proposées doivent rester dans l'intervalle imposé , ou bien si les positions se trouvent à la fin de la canalisation : par la position du premier hydrophone.

La première contrainte est prise en compte lors de la création de la population initiale. Avec une condition qui permet de fixer une distance minimum entre deux fuites.

Elle est également vérifiée après l'opération de mutation.

La deuxième contrainte est vérifiée après la mutation de la population enfant, et donc juste avant son évaluation. La stratégie de réparation consiste à désigner le minimum entre la valeur proposée et la limite inférieure de l'intervalle, puis le maximum entre ladite valeur et la limite supérieure.

### 3 Expériences numériques et interprétations

Nous présentons dans cette section les résultats numériques obtenus par notre algorithme génétique.

Le code représentant le modèle de la canalisation et la fonction objectif appartient à *X. Wang et M. Ghidaoui*.

Le code original de l'algorithme génétique appartient à *Mostapha Kalami Heris* [20], nous l'avons modifié afin de l'adapter au problème étudié.

#### 3.1 Générations d'instances

La génération d'instances est liée à l'adaptation des paramètres de l'algorithme génétique au problème décrit dans (Wang et Ghidaoui, 2018). Il s'agit de déterminer le nombre d'individus, le nombre d'itération et les probabilités de mutation et de croisement afin d'avoir un algorithme fonctionnel et des résultats satisfaisants.

Nous avons choisi une canalisation de 1 Km pour nous servir de référence, avec une limite imposée par EMV de 30 m de distance entre deux fuites. L'instance de 1,5 Km nous a servi de comparaison avec les cas présentés dans l'article de Wang et Ghidaoui.

Nous avons ensuite testé l'AG sur de grandes distances – à savoir 4Km, 8Km et 12Km - afin de vérifier son aptitude à fonctionner dans un grand espace de recherche.

- Les fuites ont été choisies avec une distance de  $0.5\lambda_{min}$  entre elles.
- Les paramètres de l'algorithme ont dû être réadaptés au fur et à mesure de l'augmentation de l'espace de recherche.
- Le temps d'exécution de la recherche exhaustive a été limité à 100 min.

L'inconvénient majeur de cette partie de la simulation est que l'on peut difficilement vérifier les erreurs commises par l'AG sans les résultats de la recherche exhaustive.

#### 3.2 Résultats obtenus

**Paramètres de l'algorithme génétique adopté :** Nous avons testé les différents paramètres afin d'optimiser la recherche pour une canalisation d'une longueur de 1 km. Nous avons procédé par tâtonnement, en commençant par des valeurs faibles et en augmentant selon les résultats. Nous sommes parvenus à cette structure :

Codage	Réel
Nombre de gènes	2
Croisement	Arithmétique avec $P_c = 0.6$
Mutation	Mutation non uniforme avec $P_m = 0.4$
Taille de la population	150
Nombre d'itérations	350

Le tableau suivant présente les résultats pour l'implémentation de l'AG sur toute la longueur des canalisations (respectivement 1 km et 1,5 km), avec comme instance des fuites proches et

éloignés pour chaque cas.

Les fuites prises pour 1.5 km sont des exemples pris de l'article de Wang et Ghidaoui (2018) et qui permettent de confirmer l'efficacité de l'algorithme génétique.

$N$	$L$ (Km)	Intervalle (m)	Fuites réelles (m)	Résultats (m)		Temps d'exécution (min)	
				RE	AG	RE	AG
1	1 Km	1000 m	(45 80)	(45 80)	(40 80)	25min	6min
			(180 650)	(180 650)	(180 650)	25min	6min
2	1,5 Km	1500 m	(260 350)	(260 350)	(258 351)	61min	6min
			(140 470)	(140 470)	(141 368)	61min	6min

TABLE 4.1 – Résultats obtenus par l'AG pour les petites distances

### Résultats pour de grandes distances :

$N$	$L$ (Km)	Intervalle (m)	Fuites réelles (m)	Résultats (m)		Temps d'exécution (min)	
				RE	AG	RE	AG
3	4 Km	4000 m	(820 1000)	(- -)	(821 1000)	>100min	15min
4	8 Km	8000 m	(4500 4820)	(- -)	(4497 4827)	>100min	17min
5	12 Km	12000 m	(6125 6550)	(- -)	(6129 6557)	>100min	20min

TABLE 4.2 – Résultats obtenus par l'AG pour les grandes distances

### 3.3 Interprétation des résultats :

Avant de commencer à analyser la performance de l'algorithme génétique, il faudrait éclaircir les erreurs commises par l'algorithme original (EMV + recherche exhaustive (RE)) et qui ont été reproduites par l'AG.

EMV est une méthode dont l'efficacité dépend du taux d'information qui lui est transmis par l'onde acoustique se propageant dans la canalisation, à travers les mesures prises par les hydrophones. Les auteurs ont précisé dans leurs articles que plus l'espace de recherche augmenterait, plus il serait judicieux d'augmenter le nombre de fréquences exploitables.

Puisque la distance minimale entre deux fuites est définie par  $\lambda_{\min}$  et que  $a = \frac{\lambda_{\min}}{f}$ . Augmenter les fréquences permettrait d'atténuer la croissance de  $\lambda_{\min}$  (avec  $a$  la vitesse d'onde et

$f$  les fréquences exploitées).

Nous avons préféré garder le nombre de fréquences utilisé par défaut dans les articles parce que ça n'influence pas la performance de l'algorithme génétique. Le résultat est une perte de précision du côté de EMV et qui est retrouvée par la recherche exhaustive et donc par l'AG.

L'AG a donné des résultats équivalents à ceux de la recherche exhaustive tant que l'intervalle sur lequel il était exécuté restait inférieur à 1500 m. A partir de là, des erreurs commencent à apparaître et bien qu'elles soient négligeables c'est le signe qu'il faut ajuster à nouveau les paramètres de l'algorithme génétique afin qu'il puisse s'adapter à de plus grands espaces de recherche.

A partir de 4 Km, le temps d'exécution de la recherche exhaustive ne nous permet plus d'avoir des résultats en moins de 100 min, il devient donc difficile d'évaluer avec précision la performance de l'AG.

Nous déduisons de l'écart dans l'estimation que la précision est inversement proportionnelle à l'augmentation de la taille de l'espace de recherche. Malgré la réadaptation des paramètres, il est difficile d'obtenir un résultat exact.

D'un autre côté, nous avons réussi à avoir une estimation suffisamment proche des fuites réelles en un temps raisonnable en comparaison avec la méthode de recherche exhaustive.

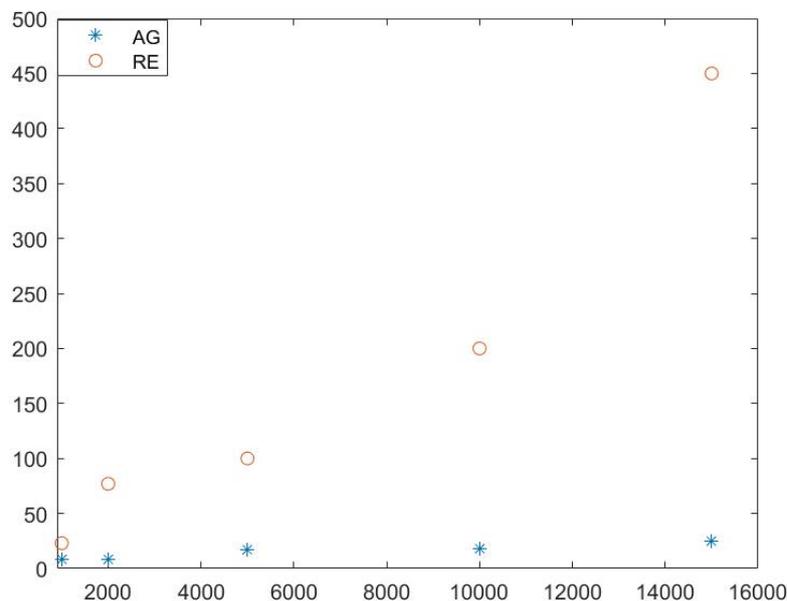


FIGURE 4.1 – Évolution du temps d'exécution par rapport à la longueur des canalisations

Puisque nous sommes en train de traiter un problème NP-difficile, nous avons limité le temps d'exécution de la recherche exhaustive à 100 min et n'avons donc pas conduit les exécutions jusqu'au bout. En comparaison, l'augmentation du temps de convergence de l'AG est plus raisonnable et nous permet de conclure que l'utilisation d'une métaheuristique est une nécessité pour les grandes instances de ce problème (grandes distances pour les canalisations).

### 3.4 Conclusion de l'analyse

Les métaheuristiques ont pour but de résoudre les problèmes NP-difficiles en un temps raisonnable, même si il n'y a aucune garantie sur la qualité d'optimalité de la solution trouvée.

C'est ce que nous avons observé dans cette étude. Malgré l'explosion combinatoire qui rend le temps d'exécution de la recherche exhaustive très excessif, l'AG a su exploiter l'espace de recherche de façon adéquate jusqu'à un certain point. La réadaptation de ses paramètres devient plus fréquente à chaque fois que l'on augmente la taille de la canalisation.

On considère en règle générale que le nombre de générations nécessaires pour atteindre la convergence augmente avec la taille de l'espace de recherche ; mais puisque l'AG est un algorithme stochastique, son temps de convergence change à chaque exécution.

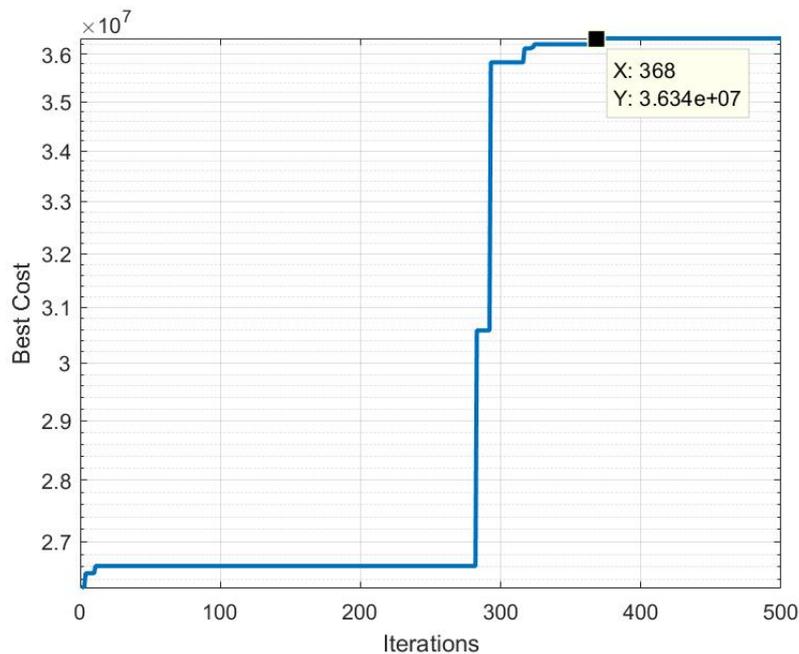


FIGURE 4.2 – Pour une canalisation de longueur  $L = 1km$  le temps de convergence est  $t_{conv} = 6min$

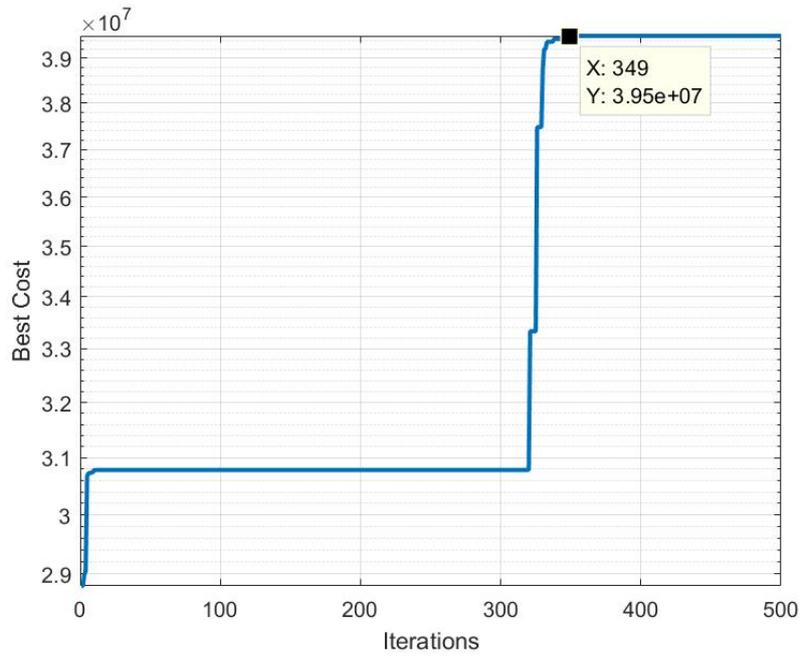


FIGURE 4.3 – Pour une canalisation de longueur  $L = 1,5 \text{ km}$  le temps de convergence est  $t_{conv} = 6 \text{ min}$

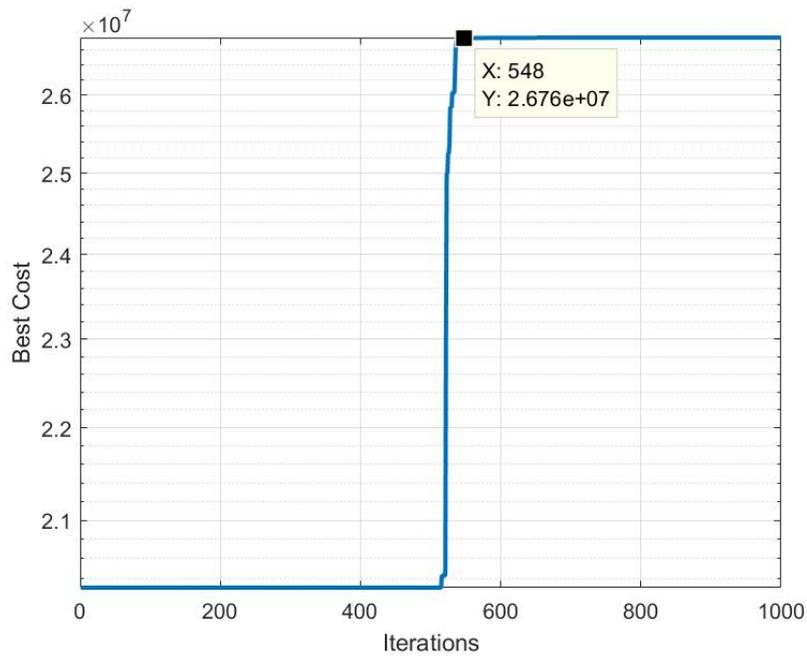


FIGURE 4.4 – Pour une canalisation de longueur  $L = 4 \text{ km}$  le temps de convergence est  $t_{conv} = 13 \text{ min}$

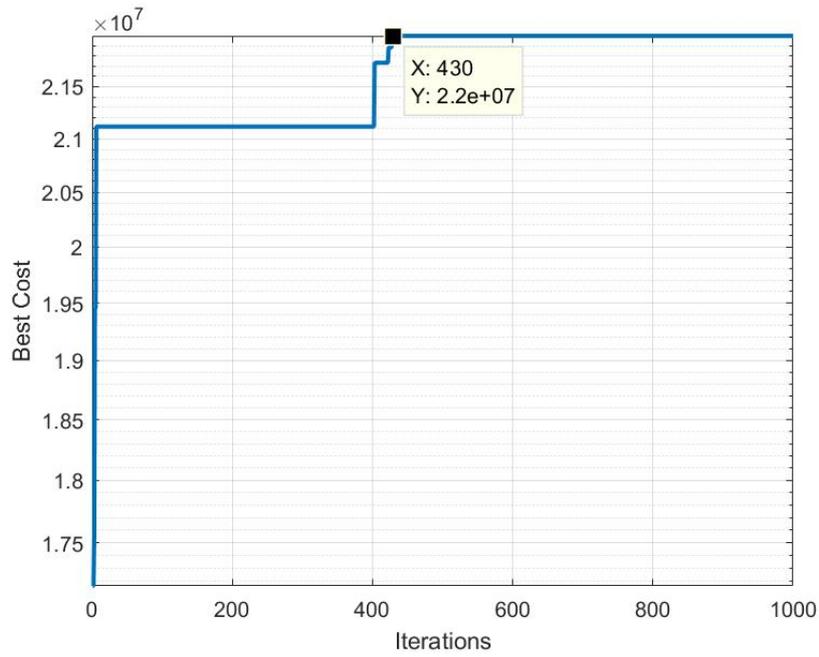


FIGURE 4.5 – Pour une canalisation de longueur  $L = 8km$  le temps de convergence est  $t_{conv} = 10min$

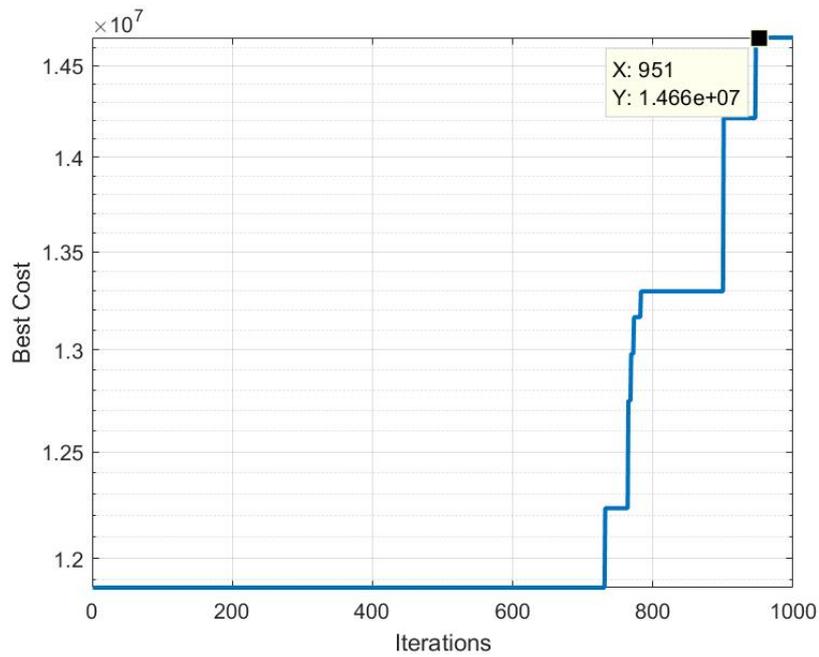


FIGURE 4.6 – Pour une canalisation de longueur  $L = 12km$  le temps de convergence est  $t_{conv} = 23min$

A travers les figures (4.2 4.3 4.4 4.5 4.6), on peut clairement voir la nature stochastique de l'algorithme génétique, les temps de convergences n'évoluent pas forcément proportionnellement à la taille de l'espace de recherche. Cela dépend de la qualité des solutions initialisées et de sa rapidité à se rapprocher de la solution exacte.

## 4 Conclusion

Nous avons présenté dans ce chapitre les résultats obtenus par l'algorithme génétique. L'AG a réalisé les objectifs fixés, à savoir un diagnostic rapide et fiable.

# Conclusion générale

Ce travail a été fait dans l'optique d'optimiser une technique de détection de fuites grâce à une méthode de l'intelligence artificielle. Pour cela, nous avons tout d'abord présenté des généralités sur l'art du diagnostic dans le premier chapitre, avec un résumé des techniques de détection de fuites existantes. Nous avons particulièrement insisté sur les techniques d'émission acoustiques qui sont au cœur de notre travail. Le deuxième chapitre a été dédié à la présentation du problème comme il a été considéré dans les articles de Wang et Ghidaoui [48], [47], nous avons détaillé les techniques de diagnostic utilisées par les auteurs, une hybridation de méthodes mêlant des méthodes hydrauliques et statistiques ; le problème principal résidait dans la recherche exhaustive qui est chronophage et ne permet pas un diagnostic rapide.

C'est pour cette raison que nous avons introduit les algorithmes génétiques (chapitre 3) qui sont des métaheuristiques, donc des techniques d'optimisation dédiées aux problèmes NP-difficile. Basé sur la théorie de Darwin, ils sont caractérisés par leur intuitivité pour exploiter l'espace de recherche ainsi que leur temps de résolution réduit par rapport aux méthodes d'optimisation classiques. Dans le quatrième chapitre, nous avons présenté notre version de l'AG, adapté à notre problème. Nous avons insisté sur les contraintes - à la fois en lien avec les méthodes de détection de défauts implémentées et le modèle de la canalisation - afin d'orienter au maximum la recherche de l'algorithme et réduire un peu plus le temps de résolution.

Nos résultats sont satisfaisants. En comparaison avec les résultats de Wang et Ghidaoui, l'AG était tout aussi précis que la recherche exhaustive et beaucoup moins chronophage. Nos instances ont été choisies pour démontrer l'évolution du temps d'exécution de chaque méthode par rapport à la longueur de la canalisation étudiée. A partir d'une certaine longueur de la canalisation, l'espace de solution devient trop grand et les paramètres de l'AG doivent être réadaptés.

Nous pensons qu'il serait intéressant d'implémenter des métaheuristiques différentes (PSO, algorithme des chauves-souris, colonie de fourmis) et de comparer les résultats obtenus. Une autre piste serait l'utilisation d'autres méthodes de l'IA ; par exemple les algorithmes d'apprentissage comme les réseaux de neurones. La difficulté sera l'extraction des paramètres mais aussi la création d'une base de données suffisamment fournie pour pouvoir entraîner le réseau.

## Résumé

**Résumé** Dans ce travail, nous analysons les travaux de Wang et Ghidaoui (2018) concernant des méthodes de détection de fuites dans un environnement bruité. Ces méthodes mélangeant des techniques de l'hydraulique (ITA) et des statistiques (EMV) sont d'une grande précision mais sont particulièrement chronophages. Le problème est résolu par la maximisation d'une fonction de coût basée sur la recherche exhaustive. Nous avons tenté de réduire le temps de diagnostic en implémentant un algorithme génétique à la place de l'algorithme de recherche exhaustive original dans l'optique d'avoir un gain de temps sans perdre en précision.

**Mots-Clés :** Diagnostic, Intelligence Artificielle, Détection de fuites, Algorithme Génétique, EMV, Traitement de Signal, TBDM, Super-résolution

---

## Abstract

**Abstract** In this work, we analyze the work of Wang and Ghidaoui (2018) concerning leak detection methods in a noisy environment. These methods combining hydraulic techniques (ITA) and statistics (MLE) are very precise but are particularly time-consuming. The problem is solved by maximizing a cost function based on the exhaustive search. We have attempted to reduce diagnostic time by implementing a genetic algorithm instead of the original exhaustive search to save time without losing precision.

**Key words :** Diagnosis, Artificial Intelligence, Leak detection, Genetic Algorithm, MLE, TBDM, ITA, Signal Processing, Super-resolution

---

## ملخص

في هذا العمل، نقوم بتحليل عمل وانغ و غداوي (2018) فيما يتعلق بطرق اكتشاف التسرب في بيئة صاخبة. تعتبر هذه الطرق التي تجمع بين التقنيات الهيدروليكية (ITA) والإحصاءات (EMV) دقيقة للغاية ولكنها تستغرق وقتاً طويلاً. يتم حل المشكلة عن طريق زيادة دالة التكلفة إلى الحد الأقصى بناءً على البحث الشامل. لقد حاولنا تقليل وقت التشخيص من خلال تنفيذ خوارزمية جينية بدلاً من خوارزمية البحث الشاملة الأصلية من أجل تقليل الوقت دون فقدان الدقة في التشخيص.

الكلمات الرئيسية: التشخيص ، الذكاء الاصطناعي ، كشف التسرب ، MLE ، TBDM ، الخوارزمية الجينية ، معالجة الإشارات

# Conclusion générale

Nous avons présenté dans ce travail des méthodes de l'apprentissage et de l'optimisation fréquemment utilisées dans le diagnostic dans le but d'étudier les stratégies utilisées par ces techniques qui les rendent plus performantes que d'autres plus conventionnelles.

Les méthodes se basant sur l'apprentissage telles que les réseaux de neurones artificiels demandent beaucoup d'efforts et de données afin d'être mises en place mais le résultat est un classificateur autonome avec une capacité d'adaptation aux nouvelles situations. L'apprentissage non supervisé, bien que compliqué à atteindre, est l'une des méthodes les plus prometteuses du diagnostic automatique.

Les méthodes se basant sur l'optimisation, et principalement les métaheuristiques, lorsqu'elles sont utilisées à des fins de diagnostic de défauts, dans la grande majorité des cas, elles représentent une méthodologie de soutien pour d'autres techniques (traitement de signal, statistique), en particulier pour les tâches de réglage des paramètres. L'objectif récurrent lors de l'utilisation des métaheuristiques est d'obtenir un résultat approché en un temps raisonnable. Il existe également des approches qui les utilisent comme une technique autonome pour effectuer le diagnostic.

Nous avons illustré l'intérêt de ces méthodes pour le diagnostic avec un exemple concret qui est la détection des fuites, en utilisant un algorithme génétique pour optimiser une méthodologie déjà existante.[47],[48] Les résultats très prometteurs obtenus ont été présentés et interprétés.

Un prochain travail pourrait se pencher sur la mise au point d'une technique hybride mêlant apprentissage et optimisation.

## Résumé

**Résumé** Dans ce travail, nous analysons les travaux de Wang et Ghidaoui (2018) concernant des méthodes de détection de fuites dans un environnement bruité. Ces méthodes mélangeant des techniques de l'hydraulique (ITA) et des statistiques (EMV) sont d'une grande précision mais sont particulièrement chronophages. Le problème est résolu par la maximisation d'une fonction de coût basée sur la recherche exhaustive. Nous avons tenté de réduire le temps de diagnostic en implémentant un algorithme génétique à la place de l'algorithme de recherche exhaustive original dans l'optique d'avoir un gain de temps sans perdre en précision.

**Mots-Clés :** Diagnostic, Intelligence Artificielle, Détection de fuites, Algorithme Génétique, EMV, Traitement de Signal, TBDM, Super-résolution

---

## Abstract

**Abstract** In this work, we analyze the work of Wang and Ghidaoui (2018) concerning leak detection methods in a noisy environment. These methods combining hydraulic techniques (ITA) and statistics (MLE) are very precise but are particularly time-consuming. The problem is solved by maximizing a cost function based on the exhaustive search. We have attempted to reduce diagnostic time by implementing a genetic algorithm instead of the original exhaustive search to save time without losing precision.

**Key words :** Diagnosis, Artificial Intelligence, Leak detection, Genetic Algorithm, MLE, TBDM, ITA, Signal Processing, Super-resolution

---

## ملخص

في هذا العمل، نقوم بتحليل عمل وانغ و غداوي (2018) فيما يتعلق بطرق اكتشاف التسرب في بيئة صاخبة. تعتبر هذه الطرق التي تجمع بين التقنيات الهيدروليكية (ITA) والإحصاءات (EMV) دقيقة للغاية ولكنها تستغرق وقتاً طويلاً. يتم حل المشكلة عن طريق زيادة دالة التكلفة إلى الحد الأقصى بناءً على البحث الشامل. لقد حاولنا تقليل وقت التشخيص من خلال تنفيذ خوارزمية جينية بدلاً من خوارزمية البحث الشاملة الأصلية من أجل تقليل الوقت دون فقدان الدقة في التشخيص.

الكلمات الرئيسية: التشخيص ، الذكاء الاصطناعي ، كشف التسرب ، MLE ، TBDM ، الخوارزمية الجينية ، معالجة الإشارات

# Bibliographie

- [1] A.B. Baggeroer, W.A. Kuperman, and P.N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, 18(4) :401–424, 1993.
- [2] Michel Batteux. *Diagnosticabilité et diagnostic de systèmes technologiques pilotés*. Theses, Université Paris Sud - Paris XI, December 2011.
- [3] Nassim Belkharci. *Contribution à l'étude du diagnostic et de la commande tolérante aux fautes par l'approche structurelle*. PhD thesis, Université de Toulouse, 2011.
- [4] R. Ben-Mansour, M.A. Habib, A. Khalifa, K. Youcef-Toumi, and D. Chatzigeorgiou. Computational fluid dynamic simulation of small leaks in water pipelines for direct leak pressure transduction. *Computers & Fluids*, 57 :110–123, mar 2012.
- [5] Pascal BOULANGER. *Nouvelles méthode de détection et de localisation de fuites par émission acoustique*. PhD thesis, UNIVERSITÉ DE PARIS-SUD, CENTRE D'ORSAY, 1993.
- [6] Ilhem Boussaid. *Perfectionnement de métaheuristiques pour l'optimisation continue*. PhD thesis, Université Paris-Est, 2013.
- [7] M.A Brahami. *Contributions à la conception des réseaux logistiques : Résolution des problèmes intégrés de localisation et de conception des réseaux de transport sous contraintes environnementales et de capacités*. PhD thesis, Université de Tlemcen, 2020.
- [8] Mustapha Anwar Brahami, Mohammed Dahane, Mehdi Souier, and M'hammed Sahnoun. Sustainable capacitated facility location/network design problem : a non-dominated sorting genetic algorithm based multiobjective approach. *Annals of Operations Research*, may 2020.
- [9] T. K. Chan, Cheng Siong Chin, and Xionghu Zhong. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*, 6 :78846–78867, 2018.
- [10] M. Hanif Chaudhry. *Applied Hydraulic Transients*. Springer New York, 2014.
- [11] Andrew F. Colombo, Pedro Lee, and Bryan W. Karney. A selective literature review of transient-based leak detection methods. *Journal of Hydro-environment Research*, 2(4) :212–227, apr 2009.
- [12] D. Covas, H. Ramos, N. Graham, and C. Maksimovic. Application of hydraulic transients for leak detection in water supply systems. *Water Supply*, 4(5-6) :365–374, dec 2004.
- [13] Henrique V. da Silva, Celso K. Morooka, Ivan R. Guilherme, Tiago C. da Fonseca, and José R.P. Mendes. Leak detection in petroleum pipelines using a fuzzy system. *Journal of Petroleum Science and Engineering*, 49(3-4) :223–238, dec 2005.
- [14] B. Farley, S. R. Mounce, and J. B. Boxall. Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal*, 7(6) :345–356, dec 2010.
- [15] Erik Frisk. Residual generation for fault diagnosis. *Linköping University*, 2001.

- [16] M. Gal, A. M. Reading, N. Rawlinson, and V. Schulte-Pelkum. Matched field processing of three-component seismic array data applied to rayleigh and love microseisms. *Journal of Geophysical Research : Solid Earth*, 123(8) :6871–6889, aug 2018.
- [17] Fred Glover. Tabu search : A tutorial. *Interfaces*, 20(4) :74–94, aug 1990.
- [18] David E. Goldberg and John H. Holland. *Machine Learning*, 3(2/3) :95–99, 1988.
- [19] Ouahib GUENOUNOU. *Méthodologie de conception de contrôleurs intelligents par l'approche génétique. Application à un bioprocédé*. PhD thesis, Université de Toulouse, 2009.
- [20] Mostapha Kalami Heris. Mostapha kalami heris, practical genetic algorithms in python and matlab – video tutoria. <https://yarpiz.com/632/ypga191215-practical-genetic-algorithms-in-python-and-matlab>), 2020.
- [21] F. Herrera, M. Lozano, and J.L. Verdegay. *Artificial Intelligence Review*, 12(4) :265–319, 1998.
- [22] Jinqiu Hu, Laibin Zhang, and Wei Liang. Detection of small leakage from long transportation pipeline with complex noise. *Journal of Loss Prevention in the Process Industries*, 24(4) :449–457, jul 2011.
- [23] Shih-Chu Huang, Wu-Wen Lin, Meng-Tsan Tsai, and Mao-Hsiung Chen. Fiber optic in-line distributed sensor for detection and localization of the pipeline leaks. *Sensors and Actuators A : Physical*, 135(2) :570–579, apr 2007.
- [24] Rolf Isermann. *Fault-Diagnosis Systems*. Springer Berlin Heidelberg, 2006.
- [25] Éli D Taillard Patrick Siarry J Dréo, A Petrowski. *Métaheuristiques pour l'optimisation difficile*. 2013.
- [26] Steven M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*. PRENTICE HALL SIGNAL PROCESSING SERIES, 1993.
- [27] Y. A. Khulief, A. Khalifa, R. Ben Mansour, and M. A. Habib. Acoustic detection of leaks in water pipelines using measurements inside pipe. *Journal of Pipeline Systems Engineering and Practice*, 3(2) :47–54, may 2012.
- [28] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598) :671–680, may 1983.
- [29] H. Krim and M. Viberg. Two decades of array signal processing research : the parametric approach. *IEEE Signal Process. Mag.*, 13 :67–94, 1996.
- [30] James A. Liggett and Li-Chung Chen. Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering*, 120(8) :934–955, aug 1994.
- [31] G. Y. Lin and C. J. Wu. Design of fuzzy logic controllers using genetic algorithms. *IEEE Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, 6 :104–109, 1999.
- [32] Salowa METHNANI. *Diagnostic, reconstruction et identification des défauts capteurs et actionneurs : application aux stations d'épurations des eaux usées*. PhD thesis, École Doctorale Mathématiques et Informatique [Équipe Estimation-Commande-Diagnostic (ES-CODI)], 2012.
- [33] Zbigniew Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1992.
- [34] Mario Alberto Coutino Minguez. Identification of room boundaries for sound field estimation. Master's thesis, TU Delft, The Netherlands, 2016.
- [35] Abdouramane Moussa Ali, Cédric Join, and Frédéric Hamelin. Diagnostic des systèmes dynamiques linéaires sans modèle explicite. *Journal Européen des Systèmes Automatisés (JESA)*, 45(Spécial : Nouvelles tendances en automatique) :339–362, January 2011.

- [36] F. Herrera O. Cordon and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems*, pages 667—674, 2001.
- [37] Didem Ozevin and James Harding. Novel leak localization in pressurized pipeline networks using acoustic emission and geometric connectivity. *International Journal of Pressure Vessels and Piping*, 92 :63–69, apr 2012.
- [38] Ranko S. Pudar and James A. Liggett. Leaks in pipe networks. *Journal of Hydraulic Engineering*, 118(7) :1031–1046, jul 1992.
- [39] Philippe Rigollet. Lecture 10 : Maximum likelihood estimation, 2016, Massachusetts Institute of Technology (MIT).
- [40] Slami Saadi. *Introduction à l'optimisation métaheuristiques*.
- [41] Patrick Siarry. *Métaheuristiques*.
- [42] Ehsan Sobhani-Tehrani and Khashayar Khorasani. *Fault Diagnosis of Nonlinear Systems Using a Hybrid Approach*. Springer US, 2009.
- [43] Christine Solnon. *Optimisation par colonies de fourmis*. 2021.
- [44] Liang Sun and Ning Chang. Integrated-signal-based leak location method for liquid pipelines. *Journal of Loss Prevention in the Process Industries*, 32 :311–318, nov 2014.
- [45] A. Tolstoy, K.V. Horoshenkov, and M.T. Bin Ali. Detecting pipe changes via acoustic matched field processing. *Applied Acoustics*, 70(5) :695–702, may 2009.
- [46] John P. Vítkovský, Angus R. Simpson, and Martin F. Lambert. Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management*, 126(4) :262–265, jul 2000.
- [47] Xun Wang and Mohamed S. Ghidaoui. Identification of multiple leaks in pipeline : Linearized model, maximum likelihood, and super-resolution localization. *Mechanical Systems and Signal Processing*, 107 :529–548, jul 2018.
- [48] Xun Wang and Mohamed S. Ghidaoui. Pipeline leak detection using the matched-field processing method. *Journal of Hydraulic Engineering*, 144(6) :04018030, jun 2018.
- [49] Xun Wang and Mohamed S. Ghidaoui. Identification of multiple leaks in pipeline II : Iterative beamforming and leak number estimation. *Mechanical Systems and Signal Processing*, 119 :346–362, mar 2019.
- [50] Yipeng Wu and Shuming Liu. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal*, 14(9) :972–983, feb 2017.