

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département d'Electrotechnique
Laboratoire de Recherche en Électrotechnique

Thèse de doctorat présenté par :
Ahmed Khaled CHAOU

Master en Électrotechnique

Pour l'obtention du titre de
Docteur Troisième Cycle (D/LMD) en Électrotechnique

Thème

*Techniques de diagnostic et de classification de
l'état de surface d'un isolateur pollué sous
tension alternative 50 Hz*

Soutenue publiquement le Jeudi 29 Octobre 2015 devant le jury composé de :

Président	A. BOUBAKEUR	Professeur, ENP
Rapporteurs	A. MEKHALDI M. TEGUAR	Professeur, ENP Professeur, ENP
Examineurs	L. MOKHNACHE O. TOUHAMI H. MOULAI	Professeur, UHL Batna Professeur, ENP Professeur, USTHB
Invité	A. BAYADI	Professeur, UFA Sétif 1

Octobre 2015

يتناول هذا العمل تشخيص وتوصيف حالة نموذج مسطح لعازل HT ملوث تحت توتر متزايد 50 هرتز. يشمل بحثنا منهجين رئيسيين يرتكزان على أساس معالجة الصورة والإشارات. نطرح في أول عمل خوارزمية الجمع بين أساليب مختلفة وأدوات معالجة الصور للكشف عن الأفراس الكهربائية. بداية تبدأ هذه الخوارزمية بتجزئة الصور عن طريق أسلوب أوتسو. تسمح هذه الخطوة باستخراج التفريغ الكهربائي من خلفية الصورة. يتم بعد ذلك إجراء الترشيح المورفولوجي من خلال الجمع بين عمليات التعرية و التمدد. يتم حساب شعاع خصائص التوسيم للمكونات المجاورة بعد ترشح الصورة و استخراج التفريغ الكهربائي. يحتوي هذا المتجه عدد التفريغات، مطال أكبر تفريغ، طوله و عرضه بالبكسال. يكون هذا المتجه مدخل اساليب التصنيف (سند آلة المتجهات، النظرية الافتراضية الساذجة وك أقرب الجيران) للصورة إلى فئتين؛ قوس أو لا قوس. تناولنا في العمل الثاني معالجة إشارة تيار التسرب قصد تشخيص حالة سطح عازل ملوث. و منه، نقوم عندئذ بتحليل إشارة التيار المتسرب بإستعمال تحويل الموجات المنفصلة (DWT) قبل حساب مخطط التكرار (RP). على أساس هذا المخطط، تم تطبيق التحليل الكمي للتكرار (RQA) للحصول على ثمانية مؤشرات أو مقاييس. تتغير هذه المؤشرات إلى حد كبير مع قيمة ناقليّة طبقة التلوث. لذلك، يتم استخدام متوسط قيمها كمدخل أساليب التصنيف المذكورة أعلاه. تسمح هذه الطرق بتوفير قيمة ناقليّة التلوث. توضح النتائج المتحصل عليها في هذا العمل أن خوارزميات معالجة الصور والإشارات التي قمنا بتطويرها تمكننا من تحقيق نتائج تصنيف ممتازة.

الكلمات المفتاحية: تشخيص، تصنيف، شدة التلوث، تيار التسرب، قوس الكهربائي، طريقة أوتسو، ترشح مورفولوجي، تعرية، تمدد، توسيم المكونات المجاورة، ك أقرب الجيران، سند آلة المتجهات، النظرية الافتراضية الساذجة، تحويل الموجات المنفصلة، مخطط التكرار، التحليل الكمي للتكرار.

Résumé

Ce présent travail porte sur le diagnostic et la caractérisation de l'état de surface d'un modèle plan d'isolateur HT pollué sous tension alternative 50 Hz. Notre investigation comporte deux principales approches basées sur les traitements d'image et de signal. En effet, dans un premier travail, un algorithme combinant différentes méthodes et outils de traitement d'image est proposé pour la détection des arcs électriques. Cet algorithme procède d'abord en segmentant les images par la méthode d'Otsu. Cette étape permet d'extraire la décharge électrique de son arrière-plan. Par la suite, le filtrage morphologique est entrepris en combinant les opérations d'érosion et de dilatation, Une fois l'image filtrée et la décharge électrique extraite, le calcul du vecteur des caractéristiques par l'étiquetage des composantes connexes est effectué. Ce vecteur comporte le nombre de décharges, l'amplitude de la plus grande décharge, sa longueur et sa largeur en pixels. Ce vecteur sera l'entrée des méthodes de classification (Support Vecteur Machine, Naïve Bayésienne et k-plus proches voisins) de l'image en deux classes ; arc ou non arc. Dans un second travail, nous utilisons le traitement du signal du courant de fuite pour le diagnostic de l'état de surface de l'isolateur sous pollution. En effet, une fois que le signal du courant de fuite est décomposé à travers la transformée en ondelettes discrètes (DWT), le diagramme de récurrence (RP) est calculé. Sur la base de ce diagramme, l'analyse quantitative de récurrence (RQA) est appliquée pour obtenir huit indicateurs, ou mesures. Ces indicateurs varient sensiblement avec la valeur de la conductivité de la couche de pollution. De ce fait, leurs valeurs moyennes sont utilisées comme entrées aux méthodes de classification susmentionnées. Ces dernières méthodes permettent d'obtenir la valeur de la conductivité de la pollution. Les résultats obtenus dans ce travail montrent que les algorithmes de traitement d'image et de signal que nous avons élaborés nous permettent d'aboutir à d'excellents résultats de classification.

Mots-Clés : diagnostic, classification, sévérité de pollution, courant de fuite, arc électrique, méthode d'Otsu, filtrage morphologique, érosion, dilatation, étiquetage des composantes connexes, k-plus proches voisins, support vecteur machine, naïve bayésienne, transformée en ondelette discrète, diagramme de récurrence, analyse quantitative de récurrence.

Abstract

This present work deals with the diagnosis and characterization of the surface condition of HV polluted insulator plane model under 50 Hz alternative voltage. Our investigation has two main approaches based on image processing and signal processing. Indeed, in a first work, an algorithm combining different methods and image processing tools is provided for detecting electric arcs. First, this algorithm first performs segmenting images by Otsu method. This step allows to extract the electrical discharge of its background. Thereafter, the morphological filtering is undertaken by combining the erosion and dilation operations. Once the image is filtered and the electric discharge is extracted, calculation of the vector of characteristics by labeling connected components is performed. This vector includes the number of discharges, the magnitude of the largest discharge, its length and its width in pixels. This vector is the entry of classification methods (Support Vector Machine, Naive Bayesian and k-nearest neighbors) of the image into two classes; arc or not arc. In a second work we use the leakage current signal processing for the diagnosis of the surface condition of the polluted insulator. Indeed, once the signal of leakage current is decomposed through the discrete wavelet transform (DWT), the recurrence plot (RP) is calculated. On the basis of this diagram, the quantitative analysis of recurrence (RQA) is applied to obtain eight indicators or measurements. These indicators vary substantially with the value of the conductivity of the pollution layer. Therefore, their mean values are used as inputs to the above classification methods. These methods provide the value of the conductivity of the pollution. The results obtained in this work show that the image and signal processing algorithms that we have developed enables us to achieve excellent classification results.

Keywords: diagnosis, classification, pollution severity, leakage current, electric arc, Otsu method, morphological filtering, erosion, dilation, labeling of connected components, k-nearest neighbors, Support Vector Machine, Naive Bayesian, discrete wavelet transform, recurrence plot, quantitative analysis of recurrence.

REMERCIEMENTS

Avant tout, je tiens tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui m'a donné la force et la patience d'accomplir ce modeste travail.

En premier lieu, je tiens à remercier Messieurs **Abdelouahab Mekhaldi** et **Madjid Tegar**, Professeurs à l'Ecole Nationale Polytechnique, pour avoir accepté de diriger et de suivre ce travail avec disponibilité, patience et bienveillance.

En second lieu, je suis très reconnaissant envers Monsieur **Ahmed Boubakeur**, Professeur à l'Ecole Nationale Polytechnique, d'avoir accepté de présider le jury de cette thèse.

Aussi, je tiens à remercier Madame **Leila Mokhnache**, Professeur à l'Université de Batna, Monsieur **Hocine Moulai**, à l'Université des Sciences et de la Technologie Houari Boumediene, Monsieur **Omar Touhami**, Professeur à l'Ecole Nationale Polytechnique et **Abdelhafid Bayadi** Professeur à l'Université de Ferhat Abbas Sétif 1, d'avoir accepté d'examiner ce travail et de participer à ce jury de thèse.

Je tiens aussi à remercier tous les enseignants ayant contribué à ma formation et ce depuis le tronc commun jusqu'à la dernière année de graduation, sans oublier l'ensemble de mes camarades de la promotion 2012.

Enfin, je ne saurais clore ce volet sans remercier mes parents, qui sans eux, mes études et ce travail n'aurait jamais abouti. Je ne les remercierai jamais assez pour leur patience, courage, conseils, exemples et dévouement.

A toute ma famille, mes parents, ma sœur et Rania, ainsi que tous mes amis.

Sommaire

Introduction générale	1
-----------------------------	---

Chapitre I

TECHNIQUES DE DIAGNOSTIC DE L'ÉTAT DE SURFACE D'UN ISOLATEUR POLLUÉ

I.1. Introduction	3
I.2. La transformée en ondelettes continue	3
I.2.1. La transformée de Fourier à Court Terme	3
I.2.1.1. Définition	3
I.2.1.2. Résolution dans le plan temps-fréquence [10]	4
I.2.2. La transformée en ondelette continue	5
I.2.2.1. Définition	5
I.2.2.2. Résolution dans le plan temps-échelle [8-11]	6
I.2.2.3. Discrétisation de la CWT	7
I.3. L'analyse multi-résolution [28]	7
I.3.1. Définition	7
I.3.2. Algorithme pyramidal	8
I.3.2.1. Présentation de l'algorithme	8
I.3.2.2. Filtrage par bande	9
I.4. Décomposition en paquets d'ondelettes, bases de paquets d'ondelettes [18]	10
I.4.1. Décomposition en paquets d'ondelettes	10
I.4.2. Arbre admissible	13
I.5. La technique du diagramme de récurrence ou Recurrent Plot (RP)	13
I.5.1. La théorie du chaos	14
I.5.1.1. Origine et définition	14
I.5.1.2. L'attracteur de Lorenz	14
I.5.2. Systèmes dynamiques	15
I.5.2.1. Modèle déterministe, modèle stochastique et modèle chaotique	15
I.5.2.2. Espace des phases	15
□ Construction de Ruelle-Takens (méthode des délais)	15
I.5.3. Le diagramme de récurrence (RP)	16

I.5.3.1. Définition	16
I.5.3.2. Les structures dans les diagrammes de récurrence	17
I.5.4. Analyse quantitative du diagramme de récurrence (RQA).....	19
I.5.4.1. Les mesures de la RQA [40].....	19
I.6. Conclusion.....	21

Chapitre II

METHODOLOGIE DE TRAITEMENT D'IMAGE

II.1. Introduction	22
II.2. Différentes étapes de propagation des décharges jusqu'au processus de contournement	22
II.3. Notions basiques sur le traitement d'image	25
II.3.1. Segmentation d'image	25
II.3.2. Méthode de segmentation d'Otsu [46-51].....	26
II.3.3. Filtrage morphologique [37-42]	29
II.3.3.1. Nécessité du filtrage morphologique.....	29
II.3.3.2. Principe du filtrage morphologique et de l'élément structurant [57]	30
II.3.3.3. Outils de base du filtrage morphologique [52-57]	31
A. Érosion	31
B. Dilatation.....	32
C. Ouverture.....	33
D. Fermeture	34
II.3.4. Etiquetage des composantes connexes [57]	35
II.4. Conclusion.....	35

Chapitre III

ELABORATION D'UN ALGORITHME DE TRAITEMENT D'IMAGE POUR L'ETUDE DES DECHARGES ELECTRIQUES

III.1. Introduction.....	37
III.2. Elaboration d'un algorithme de traitement d'image pour l'étude des décharges électriques	37
III.2.1. Segmentation et filtrage	38

III.2.2. Extraction des caractéristiques	39
III.2.2.1. Détection de la plus grande décharge (calcul de N_l et N_p)	39
III.2.2.2. Calcul des deux indicateurs quantifiant la plus grande décharge (L et W)	40
III.2.2.3. Exemple pratique de calcul des quatre indicateurs	40
III.2.3. Classification des images	41
III.2.3.1. Principe des méthodes de classification	41
III.2.3.2. Machines à vecteurs de support (SVM) [58-79]	43
III.2.3.2.1. Principe de la technique SVM [58-79]	43
III.2.3.2.2. SVM linéaires [58]	44
III.2.3.2.3. SVM pour le cas multi-classes [61-79]	49
III.2.3.2.4. Avantages et désavantages des SVM [53-56]	50
III.2.3.3. Méthodologie de l'apprentissage bayésien [80-85]	50
III.2.3.3.1. Description du modèle bayésien [81]	50
III.2.3.3.2. Estimation des paramètres [82]	51
III.2.3.4. Les k plus proches voisins [86-89]	52
III.2.3.4.1. Principe général	52
III.2.3.4.2. Fondements théoriques	53
III.2.3.4.3. Algorithme de la k-ppv	53
III.2.3.4.4. Amélioration des capacités de généralisation [88]	54
III.2.3.4.5. Avantages et inconvénients [89]	54
III.3. Résultats de l'algorithme	55
III.3.1. Résultats de la segmentation et filtrage	55
III.3.2. Résultats du vecteur des caractéristiques	57
III.4. Conclusion	61

Chapitre IV

L'ANALYSE QUANTITATIVE DE RECURRENCE (RQA) POUR LA CLASSIFICATION DE L'ETAT DE SURFACE D'UN ISOLATEUR POLLUE

IV.1. Introduction	64
IV.2. La technique RP pour l'étude du phénomène de contournement des isolateurs	64

IV.2.1. La technique RP appliquée directement au signal du courant de fuite	64
IV.2.2. La technique RP appliquée au détail du signal du courant de fuite	65
IV.2.3. Choix des paramètres m , τ et ε	67
IV.3. RP des composantes fréquentielles du signal de courant de fuite.....	67
IV.3.1. Cas d'une conductivité de 0,01 mS/cm.....	67
IV.3.2. Cas d'une conductivité de 1,2 mS/cm.....	68
IV.3.3. Cas d'une conductivité de 10,1 mS/cm.....	70
IV.4. RP de la composante haute fréquence du signal de courant de fuite	74
IV.4.1. Cas d'une conductivité de 0,01 mS/cm.....	74
IV.4.2. Cas d'une conductivité de 1,2 mS/cm.....	76
IV.4.3. Cas d'une conductivité de 10,1 mS/cm.....	78
IV.2.4. Discussion des résultats	80
IV.5. Application de l'analyse quantitative de récurrence (RQA) au signal de courant de fuite	80
IV.5.1. Cas d'une conductivité de 0,01 mS/cm.....	80
IV.5.2. Cas d'une conductivité de 1,2 mS/cm.....	82
IV.5.3. Cas d'une conductivité de 10,1 mS/cm.....	84
IV.6. La relation entre RP et RQA pour l'étude de la performance d'un isolateur pollué.....	86
IV.7. Extraction des caractéristiques à partir de la RQA	88
IV.8. Résultats de la classification	89
IV.9. Conclusion.....	91
Conclusion générale	92
Références bibliographiques	94

INTRODUCTION GENERALE



Compte tenu de leur importance primordiale dans les réseaux électriques, les isolateurs ont été l'objet de diverses investigations [1-27]. De tels isolateurs doivent remplir efficacement leur tâche d'isolation dans les environnements et les conditions les plus hostiles ; c'est pourquoi, le diagnostic de l'état de leur surface est sujet à de nombreux travaux de recherche [5-27].

Un travail de diagnostic et de classification se doit d'offrir comme résultat final une décision fiable concernant l'action à adopter vis-à-vis de l'isolateur (nettoyage, remplacement, réparation, etc...). Pour parvenir à une telle décision, différentes méthodologies ont été mises en œuvre en se basant sur les techniques d'intelligence artificielle parmi lesquelles celles d'apprentissage et de classification [13-27]. Cependant, ces dernières méthodes nécessitent un traitement préalable des données (signaux, images, vidéos, son, etc).

Les essais réalisés au Laboratoire de Haute Tension de l'Ecole Nationale Polytechnique d'Alger (ENP) sur des modèles d'isolateurs plans uniformément pollués sous tension alternative 50 Hz, ont permis d'obtenir une importante base de données de résultats expérimentaux. Ces données sont soit de types signaux échantillonnés (signaux de courant de fuite et de la tension appliquée), soit de type vidéo (décharges électriques). Une telle base de données a été utilisée dans notre présent travail en vue de diagnostiquer et caractériser l'état de surface d'un modèle plan d'isolateur pollué sous tension alternative 50 Hz. En effet, deux principales approches basées sur les traitements d'image et de signal ont été développées. La première basée sur le traitement d'images combine différentes méthodes et outils à savoir la segmentation des images, le filtrage morphologique, l'étiquetage et la classification. La deuxième approche est basée sur le traitement du signal du courant de fuite pour le diagnostic de l'état de surface de l'isolateur sous pollution. Elle utilise la transformée en ondelettes discrètes (DWT), le diagramme de récurrence (RP), l'analyse quantitative de récurrence (RQA) et la classification. Ainsi, le manuscrit est structuré en quatre chapitres distincts.

Dans le premier chapitre, nous présentons les nouvelles méthodes qui ont été employées dernièrement pour l'étude de la performance des isolateurs sous pollution [13-27]. Ces méthodes, basées principalement sur l'étude des signaux du courant de fuite et de la tension appliquée, visent à extraire les informations les plus pertinentes pouvant caractériser de manière fiable et robuste l'état réel de l'isolateur pollué.

Dans le second chapitre, nous introduisons le traitement de l'image en définissant son principe, ainsi que ses outils les plus fondamentaux. Etant pratiquement un domaine quasi-vierge en Electrotechnique, et plus spécialement dans le diagnostic et l'étude de la performance des isolateurs pollués, le traitement de l'image est très répondu dans le domaine médical [28-36], où le but est de diagnostiquer la bénignité ou malignité des tumeurs sur des tests d'imageries médicales. Dès lors, nous avons jugé qu'il serait très intéressant d'user de ces mêmes techniques pour le diagnostic de l'état des isolateurs pollués.

En se basant sur la banque de données de vidéos de contournement d'isolateurs plan préalablement enregistrées, nous avons mis en œuvre, dans le troisième chapitre, un algorithme. Ce dernier vise à détecter les décharges de type « arc », permettant ainsi de suivre l'évolution des décharges précédant le contournement, et de prédire si l'apparition de ce dernier phénomène est

proche ou lointaine. L'algorithme que nous avons développé, combine divers outils de traitement d'images basés sur la segmentation, le filtrage morphologique, l'étiquetage et la classification [37-42]. La segmentation permet d'extraire la décharge électrique de son arrière-plan. Le filtrage morphologique permet d'éliminer les pixels qui ne représentent pas la décharge électrique en question. L'étiquetage des composantes connexes permet de calculer le vecteur des caractéristiques de la décharge. Ce vecteur, composé du nombre de décharges, de l'amplitude de la plus grande décharge, de sa longueur et de sa largeur en pixels, est utilisé comme entrée aux méthodes de classification à savoir la Support Vecteur Machine, la décision bayésienne et la classification par le plus proche voisin [43-76]. La sortie de ces méthodes de classification indique s'il existe une décharge de type arc sur la surface isolante ou non.

Enfin, dans le quatrième et dernier chapitre, nous utilisons la méthode du traitement de signal. Traditionnellement, l'analyse temporelle et fréquentielle sont les méthodes de base du traitement de signal [4-13]. Cependant, ces méthodes ont fait l'objet de très nombreux travaux, et ne représentent pas de méthodologie ou d'approches innovantes, si elles sont appliquées au courant de fuite ou à la tension appliquée (voire même le signal de la charge) [8-12]. C'est pourquoi, durant nos travaux, nous nous sommes intéressés à une nouvelle méthode de traitement de signal, communément appelé « Recurrent Plot » ou « RP ». Cette méthode, basée sur la théorie du Chaos [14-16], a été sujette à de nombreuses recherches dans le domaine médical [31-36], compte tenu des signaux étudiés dans ce domaine (signaux cardiaques, neurologiques ou cérébraux) qui présentent souvent des dynamiques de types transitoire, non-périodique et intermittent. Dès lors, ces différentes dynamiques sont mises en avant lors de l'application de la RP sur le courant de fuite en présence de pollution à la surface de l'isolateur. En effet, la RP permet d'étudier, tout au long du processus de contournement, la dynamique du courant de fuite. Cependant, cette méthode est considérée comme une analyse qualitative, car elle n'offre pas de paramètres permettant de quantifier la dynamique du signal étudié. C'est pourquoi, une méthode complémentaire à la RP, appelée « Recurrence Quantification Analysis », ou « RQA », permettant de quantifier cette dynamique à travers de nombreux indicateurs, est utilisée. Ces derniers sont utilisés comme entrées à diverses méthodes de classification susmentionnées, en vue de classer la sévérité de la pollution à laquelle sont soumis les isolateurs testés.

Le diagnostic de la performance des isolateurs a été traité, durant nos travaux, suivant des approches inédites et innovantes, en se basant sur un traitement de signal et d'image. Les outils d'intelligence artificielle utilisés apportent un complément valable et permettent de prémunir les isolateurs du phénomène de contournement. Ce travail rentre dans la démarche visant à assurer un bon transport de l'énergie électrique, ainsi que la sécurité des exploitants et éléments de ce réseau.

Chapitre I

LES TECHNIQUES DE DIAGNOSTIC DE L'ÉTAT DE SURFACE D'UN ISOLATEUR POLLUÉ

I.1. Introduction

Dans ce premier chapitre, nous présenterons les principaux outils habituellement utilisés pour le diagnostic de l'état de surface d'un isolateur pollué. Etant certainement l'outil le plus utilisé dans le traitement de signal, nous présenterons en premier lieu la transformée de Fourier. Puis, nous exposerons la transformée en ondelettes continue. Cette dernière méthode comble le manque de la transformée classique de Fourier, à l'aide de sa bonne représentation temps-fréquence. Par la suite, nous parlerons de la transformée en ondelettes discrète et de l'analyse multi-résolution. Sur la base de cette dernière méthode, nous présenterons l'algorithme de décomposition en paquets d'ondelettes. Enfin, nous finirons par présenter les nouvelles méthodes des diagrammes de récurrences (Recurrent Plot, RP) ainsi que l'Analyse Quantitative de Recurrence (Recurrence Quantitative Analysis, RQA) pour l'étude du courant de fuite dans le cas d'un isolateur pollué.

I.2. La transformée en ondelettes continue

La transformée en ondelettes d'un signal permet de représenter le signal sur un espace bidimensionnel appelé plan temps-échelle, fournissant sur le signal des informations conjointes en temps et en fréquence. Le pavage du plan temps-fréquence induit par cette transformée a pour particularité de permettre une résolution temporelle fine aux hautes fréquences et une résolution fréquentielle fine aux basses fréquences. Cette propriété permet souvent une analyse intéressante du signal mais reste rigide. La décomposition en paquets d'ondelettes est une extension de la transformée en ondelettes discrète permettant de choisir le pavage du plan temps-fréquence. Ce choix est réalisé à travers la sélection d'une base de paquets d'ondelettes. En général, la base est sélectionnée selon le signal traité et selon un critère répondant aux contraintes de l'application. Cette base sera appelée meilleure base [1-12].

Afin de clarifier le propos, nous introduisons des notions utiles par l'intermédiaire de définitions sur la transformée de Fourier à court terme puis nous définissons la transformée en ondelettes continue.

I.2.1. La transformée de Fourier à Court Terme

I.2.1.1. Définition

La transformée de Fourier est un outil permettant de connaître le comportement fréquentiel d'un signal. En utilisant cette transformation, nous perdons toute information relative au temps [4]. Pour remédier à cela, et dans le cadre des signaux à énergie finie ($x(t) \in L^2(\mathbb{R})$), nous utilisons un outil « temps-fréquence » : on restreint l'existence du signal autour d'une date t , grâce à une fenêtre d'analyse $g(u - t)$ centrée sur cette date, puis on applique sa transformée de Fourier :

$$\int_{\mathbb{R}} x(u)g(u - t)e^{-i2\pi v u} du. \quad (\text{I.1})$$

On fait alors glisser cette fenêtre le long du signal, ce qui permet de mesurer le contenu spectral au cours du temps. Cette transformation est appelée la transformée de Fourier à Court Terme, on la note STFT (Short Term Fourier Transform) [4].

$$T_x(v, t) = \int_{\mathbb{R}} x(u)g(u - t)e^{-i2\pi v u} du = \int_{\mathbb{R}} x(u)g_{v,t}^*(u) du = \langle x(u), g_{v,t}(u) \rangle. \quad (\text{I.2})$$

Cette transformation peut être vue comme la projection du signal sur des atomes temps fréquences, les $g_{v,t}$. Ces vecteurs sont obtenus par applications successives de deux opérateurs élémentaires à une fonction-mère de référence g . L'opérateur de translation temporelle déplace celle-ci le long de l'axe des temps, tandis que l'opérateur de translation (ou de modulation) fréquentielle la fait glisser le long de l'axe des fréquences. On obtient alors la définition suivante des atomes :

$$g_{v,\tau}(t) = g(t - \tau)e^{i2\pi vt}. \quad (I.3)$$

I.1.1.2. Résolution dans le plan temps-fréquence [10]

La STFT ne contient pas plus d'informations que la transformée de Fourier, elle fournit simplement une représentation du signal sur un espace bidimensionnel. $T_x(v, t)$ ne peut cependant pas décrire le contenu du signal strictement à l'instant t et à la fréquence v car l'atome d'analyse $g_{v,\tau}$ est caractérisé par ses extensions conjointes temporelle Δt et fréquentielle Δv .

Une localisation temps-fréquence idéale, infiniment précise, ($\Delta t = 0$ et $\Delta v = 0$) est interdite par le principe de Gabor-Eisenberg qui stipule que la résolution conjointe temps-fréquence est minorée :

$$\Delta t \Delta v \geq \frac{1}{4\pi}, \quad (I.4)$$

où

$$\Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt} \text{ et } \Delta v^2 = \frac{\int v^2 |G(v)|^2 dv}{\int |G(v)|^2 dv}, \quad (I.5)$$

G étant la transformée de Fourier de la fenêtre d'analyse g .

Une représentation dans le plan temps-fréquence conduit à un pavage en cellules élémentaires, dont la forme ne varie ni avec le temps, ni avec la fréquence, comme l'indique la figure I.1.

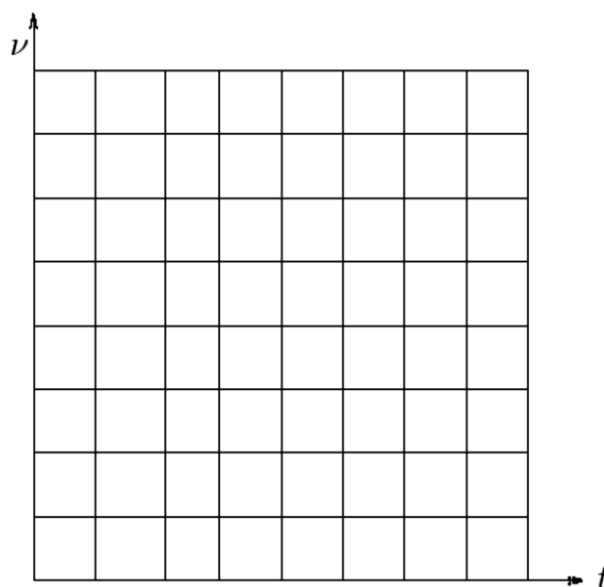


Figure I.1. Pavage du plan temps-fréquence

Sur la figure I.1, le choix d'une fenêtre longue conduit à une bonne résolution en fréquence, mais à une résolution temporelle assez mauvaise. De même, une fenêtre courte privilégie la résolution temporelle au détriment de la résolution fréquentielle.

En pratique, nous avons souvent des signaux composés de bouffées d'activité de courte durée, contenant des hautes fréquences, superposées à des composantes basses fréquences de longue durée. Il s'avère alors nécessaire de disposer d'une grande résolution temporelle dans les hautes fréquences afin de déterminer les instants d'occurrence de ces bouffées, tandis que dans les basses fréquences, une bonne résolution fréquentielle aura l'avantage de mieux caractériser les composantes de longues durées.

Ce problème peut être résolu en utilisant des fenêtres telles que $\frac{\Delta v}{v} = Q = cste$. C'est ce que réalise la transformée en ondelettes.

I.2.2. La transformée en ondelette continue

II.2.2.1. Définition

La transformée en ondelettes continue réalise une projection sur un ensemble de fonctions appelées classiquement ondelettes et dont la construction diffère de celle de la STFT : on remplace la variable de fréquence v par celle d'échelle a . Cette modification est induite par l'utilisation d'un nouvel opérateur élémentaire pour la construction des vecteurs de base. Partant d'une fonction ψ de $L^2(\mathbb{R})$ de moyenne nulle, l'ondelette mère, ceux-ci sont obtenus par action conjointe des opérateurs de dilatation en échelle :

$$\psi(t) \mapsto \psi_a(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t}{a}\right), \quad (\text{I.6})$$

et de translation en temps :

$$\psi(t) \mapsto \psi_\tau(t) = \psi(t - \tau). \quad (\text{I.7})$$

L'opérateur de dilatation fait subir à une fonction ψ un changement d'échelle de facteur a : cela revient à pratiquer sur le graphe $(t, \psi(t))$ une homothétie de paramètre a sur l'axe des temps et $\frac{1}{\sqrt{a}}$ sur celui des amplitudes. Le choix du facteur multiplicatif $\frac{1}{\sqrt{a}}$ est guidé par la volonté de préserver l'énergie du motif analysant. Les atomes s'écrivent donc :

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right), \text{ où } a \in \mathbb{R}_+^* \text{ et } \tau \in \mathbb{R} \quad (\text{I.8})$$

et définissent l'espace bidimensionnel : le plan temps-échelle, dans lequel l'information du signal $x(t)$ va être représentée. La transformée en ondelette continue CWT (Continuous Wavelet Transform) peut alors être définie par le produit scalaire suivant (équation I.9) :

$$L^2(\mathbb{R}) \mapsto L^2(\mathbb{R}^2)$$

$$x(t) \mapsto CWT_x(a, t) = \int_{\mathbb{R}} x(u) \psi_{a,\tau}^*(u) du = \langle x(u), \psi_{a,\tau}(u) \rangle. \quad (\text{I.9})$$

Cette transformation est inversible, à condition que l'ondelette mère vérifie la condition, dite d'admissibilité :

$$\int_{\mathbb{R}} |\Psi(v)|^2 \frac{dv}{v} = C_{\Psi} < +\infty, \quad (\text{I.10})$$

où $\Psi(v)$ est la transformée de Fourier de $\psi(t)$. Cette condition signifie que l'ondelette oscille, c'est-à-dire :

$$\Psi(0) = \int_{\mathbb{R}} \psi(t) dt = 0. \quad (\text{I.11})$$

II.2.2.2. Résolution dans le plan temps-échelle [8-11]

La condition d'admissibilité donne à la fonction ondelette un caractère de type passe bande qui permet de lui associer une fréquence caractéristique $v_{\psi} = \int_0^{-\infty} v |\Psi(v)|^2 dv$ où $\Psi(v)$ est la transformée de Fourier de l'ondelette mère $\psi(t)$.

La transformée de Fourier de l'ondelette ψ_a , dilatée de ψ à l'échelle a , s'écrit $\Psi_a(v) = \sqrt{a} \Psi(av)$. L'opérateur de dilatation translate donc la fréquence caractéristique selon :

$$v_{\psi a} = \frac{v_{\psi}}{a}. \quad (\text{I.12})$$

Nous verrons alors l'analyse temps-échelle comme une exploration particulière du plan temps-fréquence. L'aire des cellules élémentaires, appelés *logons*, qui couvrent le plan temps-échelle est ainsi préservée mais ceux-ci se déforment et s'allongent le long de l'axe temporel à mesure que l'échelle a augmente, comme le montre le pavage présenté par la figure I.2.

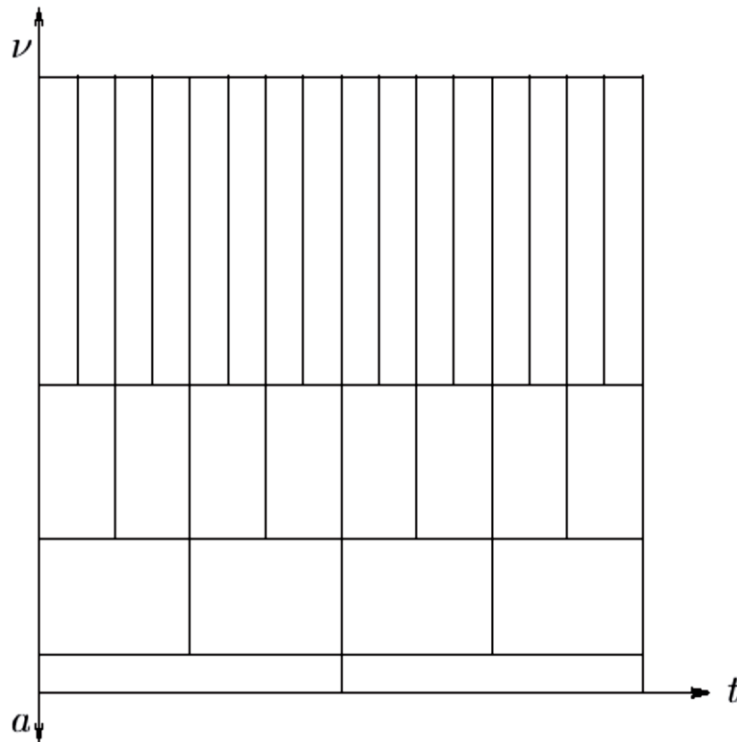


Figure I.2. Pavage du plan temps-échelle

Ce mécanisme de déformation des cellules contient l'essentiel de la richesse de l'analyse temps-échelle. Elle permet, lorsqu'on sélectionne une échelle a petite, d'effectuer une analyse du signal très localisée en temps (vue de détail) et pour une échelle a grande, de réaliser une analyse sur un horizon beaucoup plus grand (vue d'ensemble).

La CWT travaille donc à résolution relative constante. L'analyse des hautes fréquences est réalisée à résolution fréquentielle pauvre, mais permet une localisation temporelle fine (et inversement).

II.2.2.3. Discrétisation de la CWT

Le facteur d'échelle a et le pas de translation τ sont des réels et la transformée en ondelettes est continue et donc redondante. Le plan temps-fréquence est sur-analysé. Il est donc évident qu'une discrétisation de la transformée doit être envisagée si on souhaite obtenir une transformation non redondante. Le pavage temps-fréquence obtenu par la transformation en ondelettes suggère une méthode de discrétisation exponentielles pour les échelles et pour le temps. Dans l'expression $\psi\left(\frac{t-\tau}{a}\right)$ le pas de translation à l'échelle a est $\frac{\tau}{a}$. On posera donc :

$$a = a_0^m \text{ et } \tau = n\tau_0 a_0^m \text{ avec } a_0, \tau_0 \in \mathbb{Z}, \quad (\text{I.13})$$

d'où l'expression de la transformée en ondelettes discrète donnée ci-après ($n \in N$).

$$DWT_x(m, n) = CWT_x(a = a_0^m, \tau = n\tau_0 a_0^m) = \int_{\mathbb{R}} x(u) \psi_{j,k}^*(u) du = \langle x(u), \psi_{j,k}(u) \rangle. \quad (\text{I.14})$$

Si on choisit $a_0 = 2$ et $\tau_0 = 1$, on parle alors de *transformée dyadique*.

Certaines ondelettes peuvent être générées par analyse multi-résolution (AMR). Nous allons voir dans le paragraphe suivant comment les coefficients de la transformée en ondelettes discrète d'un signal peuvent être obtenus par l'AMR.

I.3. L'analyse multi-résolution [28]

I.3.1. Définition

L'analyse en multi-résolution permet de décomposer un signal en différentes bandes de fréquences, ce qui donne une vue de la plus fine à la plus grossière.

Soit φ la fonction échelle. Elle doit être L^2 et avoir une moyenne non nulle. On forme une base de fonctions d'échelle pour tout $j \in \mathbb{Z}$ comme suit :

$$\varphi_{j,k}(t) = 2^{-j/2} \varphi(2^{-j}t - k), \text{ où } k \in N \quad (\text{I.15})$$

et de la même manière la base d'ondelette ψ :

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (\text{I.16})$$

En notant que $h(k)$ et $g(k)$ sont les filtres passe bas et passe haut respectivement lors d'une décomposition par ondelettes, le facteur d'échelle dyadique mène à :

$$\varphi(t) = \sum_k 2^{-k} h(k) \varphi(2t - k) \quad (\text{I.17})$$

$$\psi(t) = \sum_k 2^{-k} g(k) \varphi(2t - k) \quad (\text{I.18})$$

Les expressions I.17 et I.18 représentent la décomposition de la fonction échelle et de la fonction ondelette en combinaisons linéaires de la fonction échelle.

I.3.2. Algorithme pyramidal

I.3.2.1. Présentation de l'algorithme

L'analyse multi-résolution d'un signal revient à le décomposer en différentes échelles, en approximation et en détail. S. Mallat [28] propose un algorithme rapide permettant de calculer les coefficients de détails et d'approximations en utilisant des filtrages et divisions successifs.

La figure I.3 présente cet algorithme : les coefficients de détail (les d_i) sont obtenus par filtrage passe haut (filtre h_1) puis décimation par 2 ; les approximations (les a_i) sont obtenues de la même manière par filtrage passe bas (g_1). Pour obtenir les coefficients de détails aux résolutions supérieures, il suffit de réitérer ces étapes sur les coefficients d'approximation.

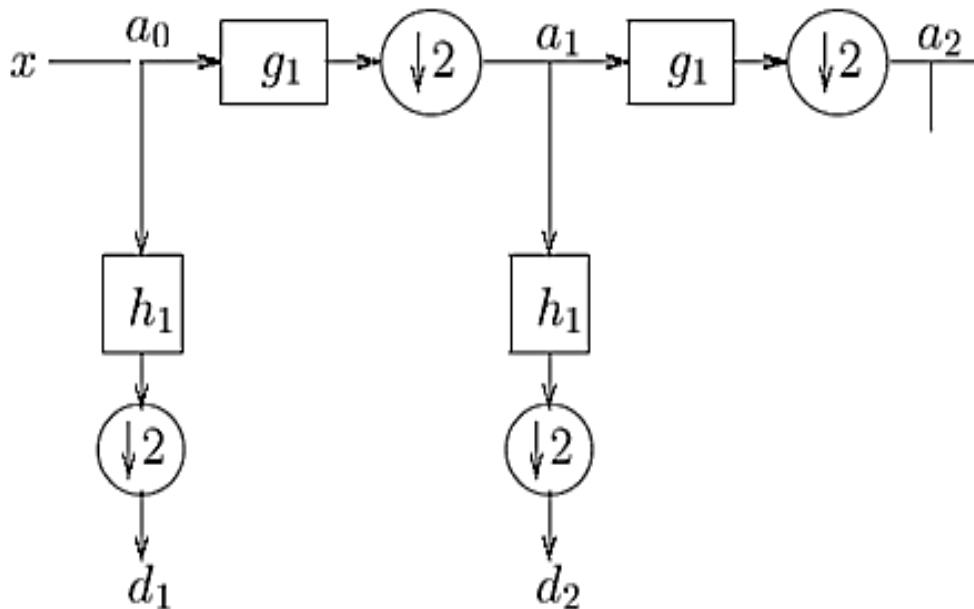


Figure I.3. Algorithme pyramidal de Mallat

On peut reconstruire le signal grâce à des filtres h_2 et g_2 selon l'algorithme présenté à la figure I.4. L'approximation a_n à un niveau donné n est la somme des coefficients de détail d_{n+1} et d'approximation a_{n+1} du niveau supérieur préalablement filtrés et ré-échantillonnés.

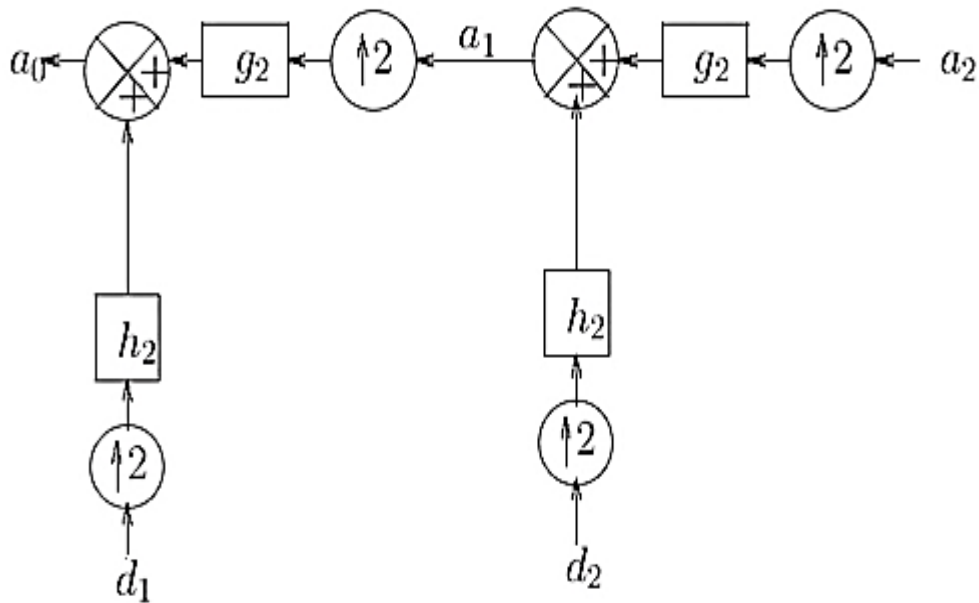


Figure I.4. Algorithme de reconstruction du signal

I.3.2.2. Filtrage par bande

D'un point de vue fréquentiel, le signal apparaît comme décomposé suivant différentes bandes. La figure I.5 présente ce point de vue.

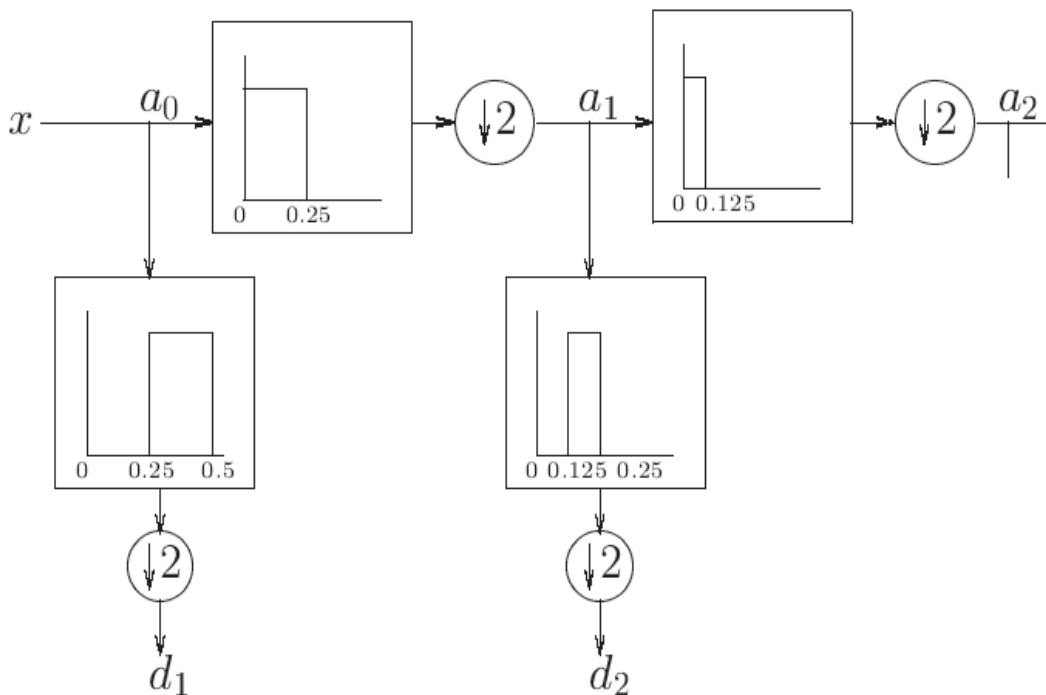


Figure I.5. Algorithme pyramidal de Mallat : point de vue fréquentiel

Les filtres h_l et g_l sont liés aux filtres exhibés par les relations I.17 et I.18. Ils sont symétriques par rapport à $\pi/2$, de puissances complémentaires. Ce sont des filtres miroirs en quadrature (QMF : Quadrature Mirror Filter) [28].

Les choix sur les filtres déterminent les ondelettes associées. Ce sont aussi ces filtres qui permettent de représenter au mieux la décomposition en sous-bandes du signal.

La figure I.6 présente la décomposition en base d'ondelettes du signal x composé de N points. L'arbre de décomposition a $\log_2(N)$ niveaux. A chaque niveau, la résolution temporelle est divisée par 2, au dernier niveau les coefficients de détails sont représentés par un unique point, la résolution temporelle est nulle et la résolution fréquentielle est maximum. Le signal est décomposé en $N-1$ coefficients de détail et 1 coefficient d'approximation (la composante de plus basse fréquence du signal).

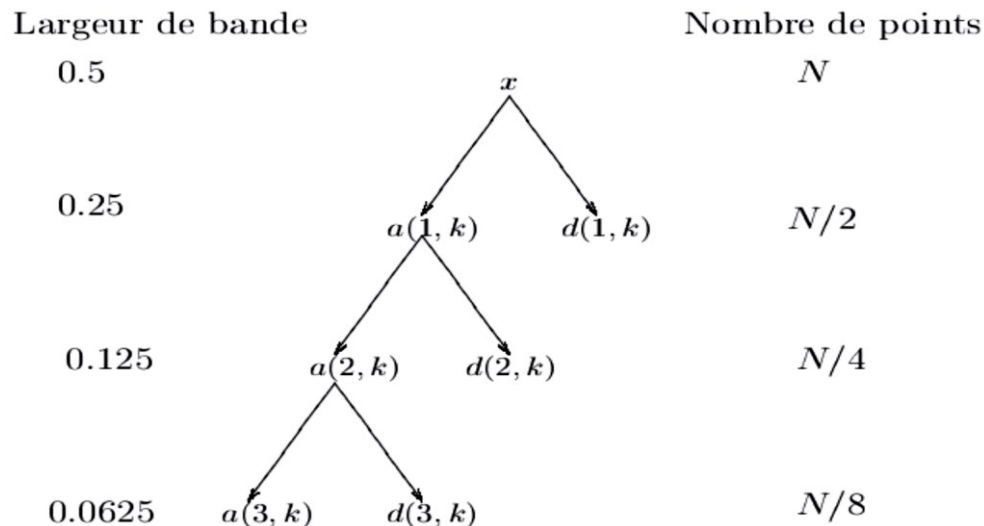


Figure I.6. Arbre de décomposition d'un signal sur une base d'ondelettes

I.4. Décomposition en paquets d'ondelettes, bases de paquets d'ondelettes [18]

La caractéristique principale de la transformée en ondelettes est de permettre une analyse à des résolutions variables selon les échelles d'analyse. Un phénomène lent de longue durée ne nécessite pas une bonne localisation temporelle, tandis que les événements rapides sont traités avec une résolution temporelle précise, au détriment de la précision fréquentielle. L'approche par paquets d'ondelettes est différente : elle offre une librairie d'ondelettes organisées en fonction de leurs propriétés d'analyse et de localisation temps-fréquence, et donc de filtrage passe-bande, selon une architecture d'arbre binaire. Cette architecture permet alors de mettre en œuvre des algorithmes de recherche de bases adaptés à la fois aux propriétés temps-fréquence souhaitées et au signal analysé, que l'on dénomme classiquement "meilleures bases".

De plus, cette décomposition présente l'avantage d'être définie et mise en œuvre à partir d'outils théoriques et d'algorithmes développés pour la transformée discrète en ondelettes (DWT).

I.4.1. Décomposition en paquets d'ondelettes

Les paquets d'ondelettes ont été définis par en généralisant la relation entre les approximations multi-résolution et les ondelettes. Un espace V_j d'une approximation multi-résolution se décompose en un espace de résolution plus basse V_{j+1} et un espace de détails W_{j+1} . Alors que la décomposition en ondelettes se contente de décomposer les espaces d'approximations

V_j et garde intacts les espaces de détails W_j , la décomposition en paquets d'ondelettes opère encore sur les espaces de détails pour en déduire de nouvelles bases.

Les figures I.7 et I.8 décrivent d'une manière simple la différence entre la décomposition en ondelettes et celle en paquets d'ondelettes. La première consiste à décomposer V_j en V_{j+1} et W_{j+1} puis à réitérer cette procédure uniquement pour V_{j+1} ensuite V_{j+1} jusqu'à l'obtention de V_j et W_j alors que la deuxième applique cette décomposition à V_{j+1} mais aussi à W_{j+1} et redécompose chacun des sous espaces obtenus en utilisant le même banc de filtres associé à l'analyse multi-résolution [18].

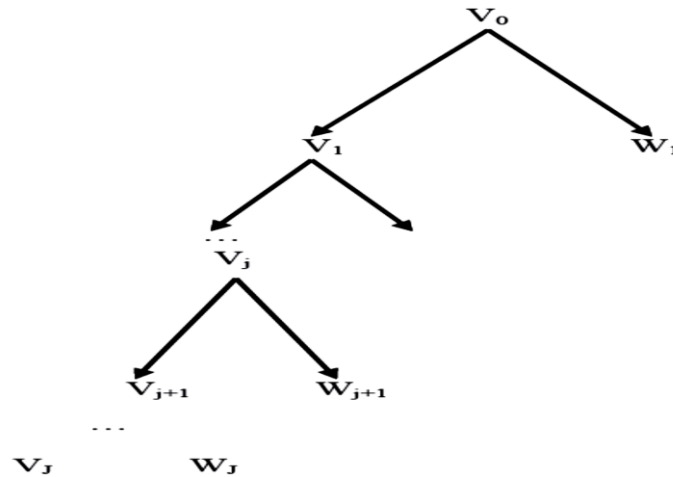


Figure I.7. Décomposition en ondelettes à l'ordre J

Pour mieux illustrer la décomposition en paquets d'ondelettes associée à la construction des bases, on donne le schéma suivant : soit $j \in \mathbb{Z}$, pour chaque échelle j . Les valeurs possibles du paramètre n qui numérote les paquets sont : $0, 1, \dots, 2^j - 1$. Dans la suite, l'espace V sera remplacé par W .

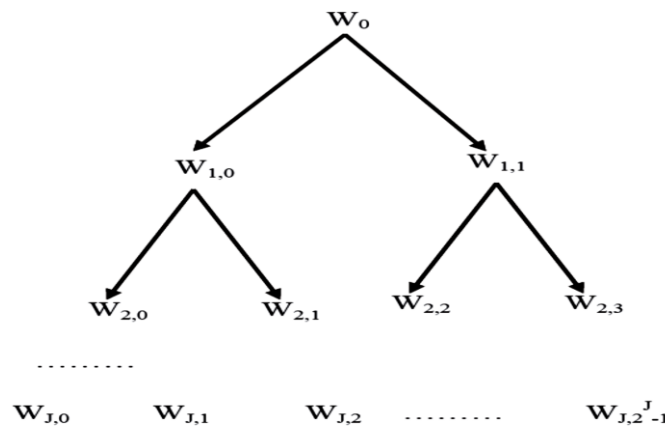


Figure I.8. Décomposition en paquets d'ondelettes à l'ordre J

Les paquets d'ondelettes sont indexés par trois paramètres : l'échelle, l'index et la position. j peut être interprété comme un paramètre d'échelle, n comme un index de chaque nœud de l'échelle j et k correspond à la bande spectrale [28]. Pour les différentes valeurs de j et n , la décomposition en paquets d'ondelettes est organisée en arbres. L'arbre tracé en figure I.9 est créé

pour un niveau de décomposition égal à quatre, où g_1 et h_1 sont les filtres QMF associés aux fonctions échelle φ et ondelette ψ .

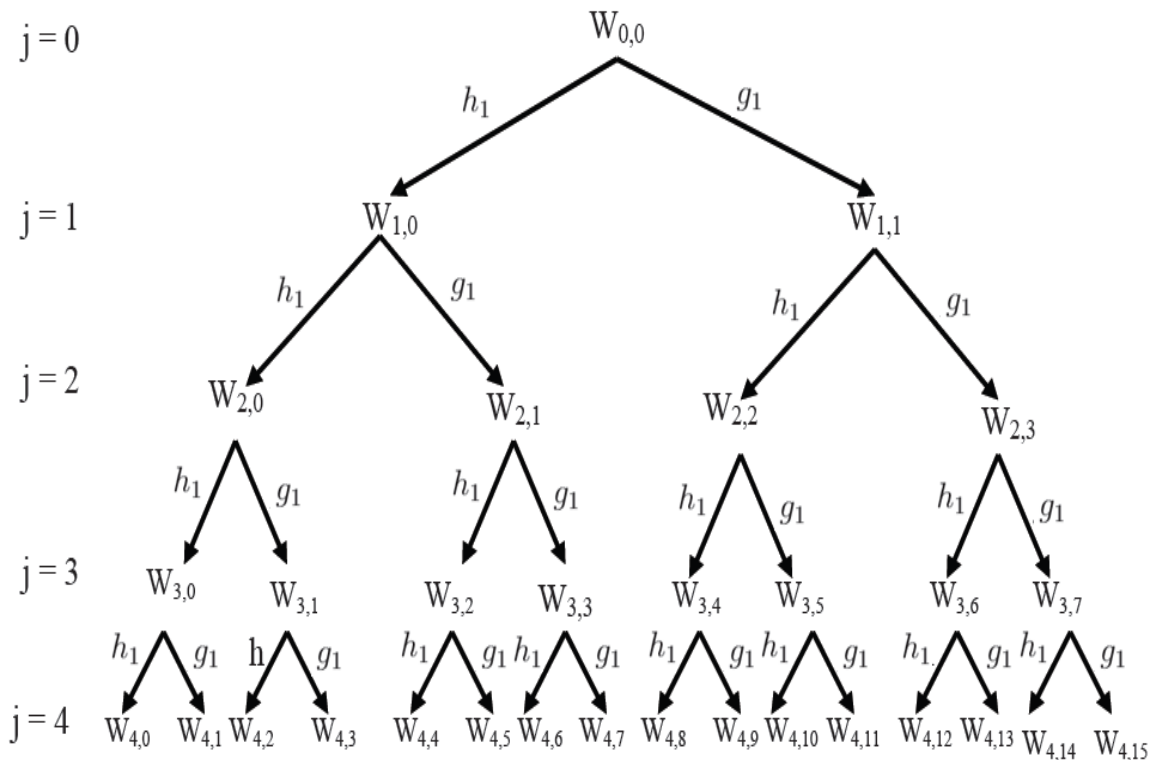


Figure I.9. Arbre obtenu après décomposition en paquets d'ondelettes

L'arbre obtenu après décomposition en paquets d'ondelettes peut être présenté en introduisant le niveau d'échelle et le numéro de paquet pour chaque niveau de décomposition (figure I.9). Une autre façon de représenter l'arbre est de numéroter les paquets par ordre croissant (figure I.10) [18].

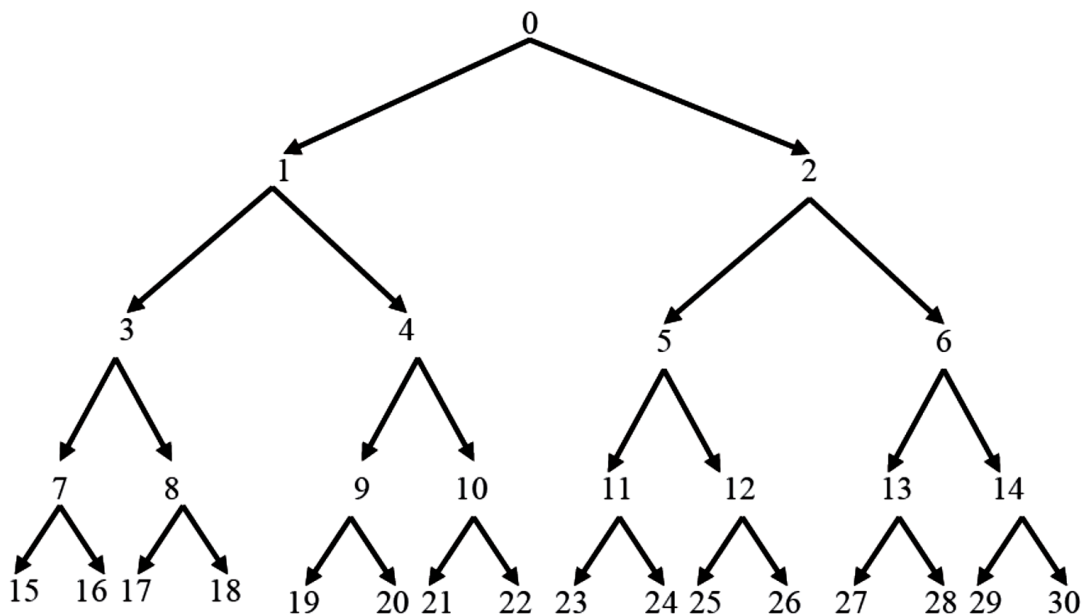


Figure I.10. Arbre obtenu après décomposition en paquets d'ondelettes, paquets numérotés

I.4.2. Arbre admissible

L'arbre binaire de décomposition en paquets d'ondelettes donne une représentation fortement redondante du signal. Si l'on souhaite travailler avec une représentation non redondante, il faut choisir une base de paquets, c'est-à-dire un ensemble de nœuds dont la projection dans l'espace temps-fréquence forme une partition [44]. Celle-ci est obtenue à partir de la notion d'arbre admissible. On appelle arbre admissible tout arbre qui permet une reconstruction exacte du signal de départ. Il est représenté par un arbre binaire dont chaque nœud est 0 ou 2 fils ; la base sera constituée de tous les nœuds n'ayant pas de fils.

La figure I.11 représente deux exemples de bases de paquets d'ondelettes. Les nœuds choisis dans la constitution de la base sont entourés par des carrés. On voit que si on élague les arbres au niveau des nœuds choisis, on obtient des arbres admissibles.

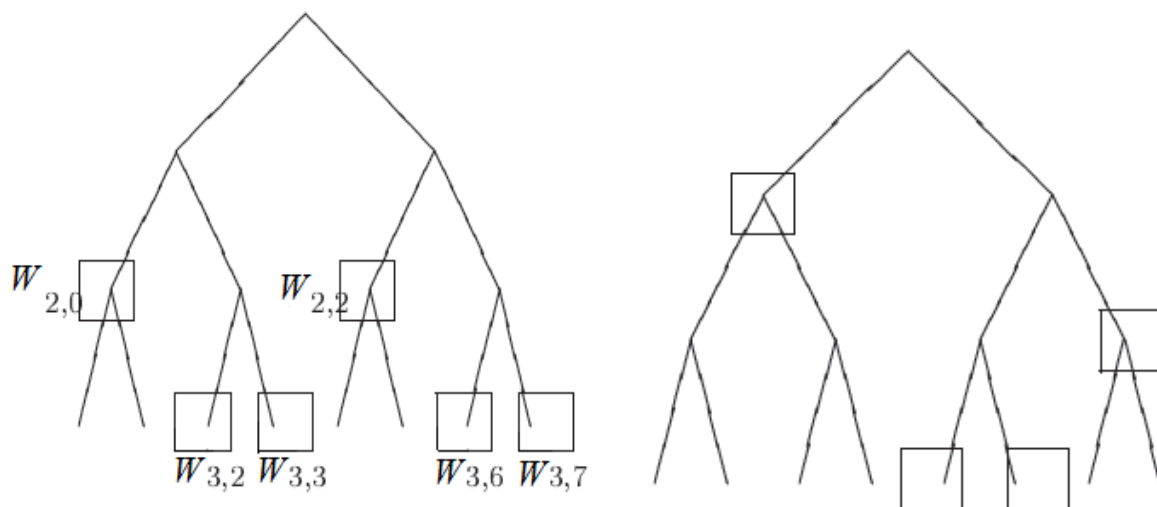


Figure I.11. Exemples de bases de paquets d'ondelettes

Obtenir la meilleure base de décomposition, revient à rechercher l'ensemble des coefficients d'approximation et de détail les mieux adaptés pour représenter le signal et l'information qui nous intéresse. Nous présenterons des exemples d'algorithmes de sélection de meilleure base dans ce qui suit.

I.5. La technique du diagramme de récurrence ou Recurrent Plot (RP)

Le diagramme de récurrence, ou Recurrent Plot (RP) concerne l'aspect qualitatif de la théorie du chaos. En se basant sur l'étude de la récurrence d'états d'un système dynamique, cette méthode permet une reconstruction sans pertes des données, et donne un aperçu immédiat de l'état du processus étudié. Par ailleurs, l'analyse quantitative de récurrence (RQA), permet de quantifier le comportement d'un système dynamique dans un espace de phase bien défini, en décrivant ces récurrences à l'aide de différentes mesures plus ou moins complexes. Mais en réduisant le signal sur un certain nombre de points, il perd un nombre plus ou moins important d'informations.

Ainsi, avant de décrire la structure des RP, il va de soi qu'un bref aperçu soit donné sur la théorie du chaos, quelques définitions et généralités concernant les systèmes dynamiques, pour décrire les différentes structures du RP, ainsi que l'analyse quantitative qui lui est associée. La structure RP sera appliquée aux composantes fréquentielles du signal du courant de fuite issues de sa décomposition en DWT. Cette étude sera établie pour voir l'influence de la classe de

pollution, de la conductivité de la solution contaminante ainsi que la largeur de la couche polluante. Finalement une analyse quantitative de récurrence sera appliquée au signal du courant de fuite pour étudier les mêmes caractéristiques [29-33].

I.5.1. La théorie du chaos

I.5.1.1. Origine et définition

La théorie du chaos est l'étude des systèmes qui ont une évolution extrêmement complexe et qui paraissent désordonnés. Cette théorie fut développée à partir des années soixante, lorsque Edward Lorenz, professeur de mathématique au MIT (Massachusetts Institut of Technologie) et père de la théorie du chaos [29-33], montra le caractère chaotique de l'atmosphère terrestre. Ce météorologue avait construit un modèle mathématique simple de douze variables pour prédire le déplacement des grandes masses d'air dans l'atmosphère. En faisant des simulations sur son ordinateur, il s'aperçut qu'une modification infime des données initiales pouvait changer radicalement l'évolution météorologique. Ainsi, en 1961, Lorenz venait de mettre en évidence l'une des principales caractéristiques du chaos : l'extrême sensibilité aux conditions initiales [29].

La théorie du chaos peut être définie comme étant la théorie qui traite des systèmes dynamiques rigoureusement déterministes, mais qui présentent un phénomène fondamental d'instabilité appelé « sensibilité aux conditions initiales » qui, modulant une propriété supplémentaire de récurrence, les rend non prédictibles en pratique à « long » terme [29-31]. La théorie du chaos est donc la théorie qui cherche de l'ordre dans le désordre.

I.5.1.2. L'attracteur de Lorenz

En 1963, à la suite de sa découverte, Lorenz entreprit alors de réaliser un nouveau modèle météorologique en réduisant le nombre de variable à trois au lieu de douze. De ce fait, Lorenz décida de représenter graphiquement la solution de son système à trois dimensions. Il obtint alors des figures géométriques dans l'espace (y_1, y_2, y_3), qu'il appellera : les attracteurs [29-30] (figure I.12). Ainsi, en apportant d'infimes changements aux conditions initiales de son système, il obtient des résultats différents. Ces derniers, décrivent une courbe ressemblant fortement à des ailes de papillon. C'est pourquoi le chaos a souvent été explicité par ce qu'on appelle « l'effet papillon », d'où la célèbre citation de Lorenz : « le battement d'aile d'un papillon aujourd'hui à Pékin engendre dans l'air suffisamment de remous pour influencer sur l'ordre des choses et provoquer une tempête le mois suivant à New York ».

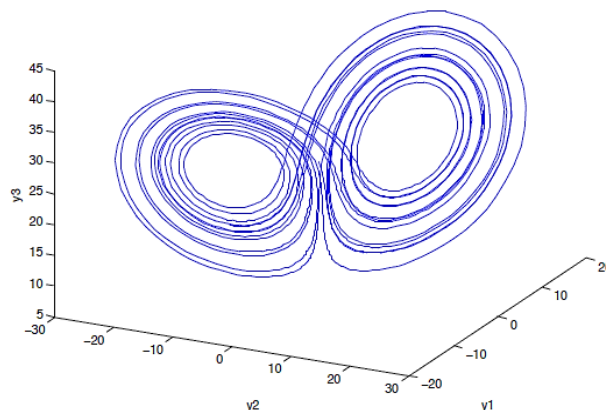


Figure I.12. Attracteur étrange de Lorenz

I.5.2. Systèmes dynamiques

Un système dynamique est un système physique qui évolue soit dans le temps soit par rapport à une autre variable suivant l'espace de phases considéré. La trajectoire d'un objet en mouvement dans le temps est donc un système dynamique, ainsi que le nombre d'individus d'une population quelconque dans le temps, ou encore les valeurs d'une fonction (par exemple : $y = 2x$) par rapport à la valeur de x .

I.5.2.1. Modèle déterministe, modèle stochastique et modèle chaotique

On peut citer trois sortes de systèmes dynamiques : les systèmes aléatoires (aussi appelés systèmes stochastiques), les systèmes déterministes et les systèmes chaotiques.

Les systèmes aléatoires évoluent comme leur nom l'indique au hasard dans tout l'espace sans qu'aucune équation ne les régit, sans qu'aucune prévision exacte ne soit possible dans le temps.

Les systèmes déterministes sont des systèmes régis par des lois mathématiques bien connues. On peut donc prévoir exactement leur évolution dans le temps.

Les systèmes chaotiques, quant à eux, ont un comportement infiniment complexe. La trajectoire qu'ils décrivent se rapproche d'une figure géométrique de structure également infiniment complexe sur laquelle ils semblent errer au hasard, mais sans jamais la quitter, ni repasser deux fois par le même point. Les attracteurs qui caractérisent ces systèmes semblent inclure à la fois des lois déterministes et des lois aléatoires, ce qui rend impossible toute prévision à long terme.

I.5.2.2. Espace des phases

Pour suivre l'évolution de l'état d'un système dynamique, préalablement décrit avec des lois physiques, on définira à un instant donné un point dans un repère. Ce point caractérise l'état du système dans l'espace à cet instant. Cet espace est appelé « l'espace des phases » de dimension N . Lorsque la variable d'évolution change de valeur (quand le temps s'écoule, par exemple), le point figurant à l'état du système décrit en général une courbe dans cet espace.

- **Construction de Ruelle-Takens (méthode des délais)**

En pratique, il n'est pas possible de mesurer toutes les composantes d'un vecteur \vec{x} décrit dans un espace de phases à N dimensions. Ainsi, en 1981, Ruelle et Takens [31] prouvent qu'il est possible de reconstruire l'attracteur d'une série temporelle. Dès lors, l'espace de phase est reconstruit en un « espace d'immersion » (ou espace de plongement) de dimensions m inférieur à N en utilisant les « vecteurs d'immersion » [32-36] décrit comme suit :

$$\vec{x}(t_i) = [x(t_i), x(t_i - \tau), \dots, x(t_i - (m - 1) \cdot \tau)], \quad (\text{I.19})$$

où m représente la dimension d'immersion, τ est le retard (ou le pas) et $x(t_i)$ est la valeur de la série temporelle à l'instant t_i .

Ce théorème permet donc de se renseigner sur le comportement d'un système non linéaire, en l'étudiant sur un espace d'immersion reconstruit à partir d'une seule série temporelle, pourvu que les paramètres m et τ de cette dernière soit connus [31].

Pour l'estimation de ces deux paramètres, différentes méthodes sont adoptées [32-36]. En ce qui concerne la dimension m , nous citons:

- False Nearest Neighbours ;
- Bad Prediction Method ;
- Fractal and Correlation Dimensions.

Pour le calcul de τ , nous pouvons utiliser les méthodes suivantes :

- Autocorrelation and Power Spectrum Functions ;
- Average Mutual Information (AMI) Function ;
- Degree of Separation Function ;
- Exposants de Lyapunov.

Ces méthodes sont des méthodes analytiques. Cependant, il existe des méthodes empiriques, nous citons : « les réseaux de neurones » et les algorithmes génétiques. Tout de même, il existe une nouvelle méthode reconnue pour son efficacité, elle permet de déterminer à la fois ces deux paramètres, elle est appelée la « C-C Method » [36].

I.5.3. Le diagramme de récurrence (RP)

I.5.3.1. Définition

Le premier concept de récurrence fut introduit par Henry Poincaré en 1890, qui non seulement, a introduit le premier concept de point récurrent [37], mais a aussi introduit un phénomène fondamental d'instabilité appelé « sensibilité aux conditions initiales » qui, modulant une propriété supplémentaire de récurrence, rend les systèmes dynamiques non prédictibles en pratique à « long terme ». Cette théorie, n'a pu être exploitée qu'après 70 ans, et ce en attendant le développement des ordinateurs et machines de calcul.

En 1987, Eckmann et al [38] ont introduit la méthode des RP pour visualiser les récurrences des systèmes dynamiques. Supposons qu'on ait une trajectoire \vec{x} d'un système dans son espace de phase [42]. Ainsi, le RP correspondant est construit comme suit : chaque point de l'espace de phase \vec{x} est testé s'il est proche ou non d'un autre point x_j . Ainsi, si la distance entre ces deux points est inférieure à un seuil ε prédéfini, les points sont dits récurrents et sont associés à un point noir. Dans le cas contraire, ils sont non récurrents et sont associés à un point blanc. Cela se traduit par une matrice $N \times N$, telle que :

$$R_{i,j} = \begin{cases} 1 & \text{si } \vec{x}_i \approx \vec{x}_j, \\ 0 & \text{si } \vec{x}_i \not\approx \vec{x}_j, \end{cases} i, j = 1, \dots, N \quad (\text{I.20})$$

où N est le nombre d'états considérés et $\vec{x}_i \approx \vec{x}_j$ indique l'égalité avec une certaine erreur ε .

Ainsi, l'équation I.2 peut être réécrite sous sa forme matricielle :

$$\mathbf{R}_{i,j}(\varepsilon) = \theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \quad i, j = 1, \dots, N \quad (\text{I.21})$$

où N est le nombre de points \vec{x}_i mesurés, ε est le seuil, $\theta(\cdot)$ est la fonction échelon ($\theta(x) = 0$ si $x < 0$, et $\theta(x) = 1$ autrement) et $\|\cdot\|$ est une norme.

Pour un ε fixé (pour les états voisins de ε), on introduit la notion suivante :

$$\vec{x}_i \approx \vec{x}_j \Leftrightarrow \mathbf{R}_{i,j} = 1. \quad (\text{I.22})$$

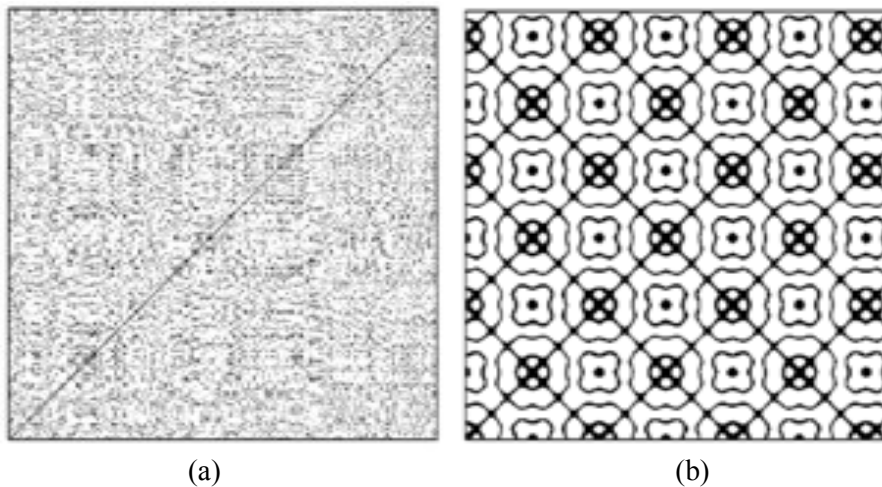
Puisque $\mathbf{R}_{i,j} = 1$ par définition, le diagramme de récurrence se caractérise par une ligne noire diagonale, appelé « la ligne d'identité » (Line Of Identity, LOI), avec un angle de $\pi/4$.

Le paramètre ε est crucial lors d'un diagramme de récurrence. Ainsi, une attention spéciale doit lui être accordée. Si la valeur de ε est trop grande, la plupart des points seront voisins d'autres points, ce qui produira un diagramme condensé, et donc des points non pertinents. D'autre part, si la valeur de ε choisie est trop petite, on peut perdre la plupart des points récurrents, et donc aucune information utile ne sera tirée. Par ailleurs, il n'existe pas de méthode standard pour choisir la valeur de ε . La plupart des chercheurs la choisissent comme étant égale à 0.25σ , σ représente la déviation standard du signal étudié [22-24].

I.5.3.2. Les structures dans les diagrammes de récurrence

Il existe 4 types de diagramme de récurrence (figure I.13), et cela dépendamment de leurs structures :

- diagramme homogène ;
- diagramme périodique et quasi-périodique ;
- diagramme dévié : une déviation est causée par des systèmes à faible variations des paramètres (systèmes non stationnaires) ;
- diagramme interrompu : les changements abrupts causent des surfaces blanches sur le diagramme [40-41].



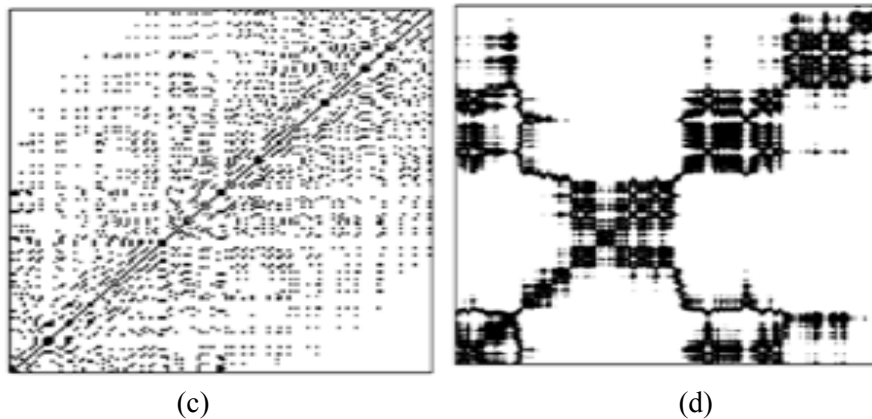


Figure I.13. Topologie caractéristique des diagrammes de récurrence
(a) Homogène, (b) périodique, (c) dévié et (d) interrompu

La technique RP consiste donc en points isolés, points joints dans des segments de lignes et espaces blancs. Les points denses et les segments blancs reflètent dans notre cas la caractéristique non linéaire du signal du courant de fuite. Un « point récurrent » signifie que la distance entre deux vecteurs voisins dans l'espace de phase est inférieure ou égale à la valeur du seuil ε . La diagonale principale existe dans toutes les structures, ce qui correspond au fait que la distance entre le vecteur et la diagonale elle-même est nulle. Le point à la plus haute densité indique que la distance entre les vecteurs reconstruits dans l'espace de phase change légèrement. Il indique que les vecteurs incrustés sont très similaires aux vecteurs voisins dans m -dimensions d'espace de phase [40].

Les points à faible densité représentent un dispersément évident dans le processus. Alors que les espaces blancs indiquent un changement rapide d'amplitude. Nous résumons les significations à propos de la topologie des structures du diagramme de récurrence dans le tableau suivant.

Tableau I.1. Exemple typique de diagramme de récurrence et leurs significations.

Exemples	Signification
(1) Homogénéité	Le processus est stationnaire.
(2) Eclaircissement des coins haut gauche et bas droit	Données non stationnaires ; le processus contient un mouvement ou un décalage.
(3) Disruptions (bandes blanches)	Données non stationnaires ; certains état sont rares ; présence de transitions.
(4) Modèle périodique/ quasi-périodique	Processus cyclique ; la période correspond à la distance entre les lignes ; différentes distances entre les lignes diagonales révèlent un processus quasi-périodique.
(5) Points isolés	Forte fluctuation du processus ; si seuls les points isolés sont présents, le processus étudié est faiblement ou non corrélé.

(6) Points denses	La densité des points indique l'intensité du processus étudié. Si les points gagnent en densité, l'activité du processus étudié s'intensifie.
(7) Lignes diagonales (parallèles à la LOI)	L'évolution est similaire à d'autres périodes ; le processus peut être déterministe. Si ces lignes diagonales sont présentes à côté des points isolés, le processus peut être chaotique.
(8) Lignes diagonales (orthogonale à la LOI)	L'évolution est similaire à d'autres périodes mais en temps inverse ; parfois, cela indique une dimension d'immersion m insuffisante.
(9) Lignes verticales et horizontales	Certains états présentent des changements peu nombreux ou lents ; indiquent des états laminaires.
(10) Lignes longues arquées	L'évolution est similaire à d'autres périodes mais à une vitesse différente ; la dynamique du système peut être variable.

I.5.4. Analyse quantitative du diagramme de récurrence (RQA)

Dans le but d'aller au-delà de l'analyse visuelle des diagrammes de récurrence, et sachant la difficulté relativement importante à les interpréter, différentes mesures quantifiant les structures précédemment décrites par les diagrammes ont été proposées. J.P. Zbilut et C.L. Jr. Webber [42] ont développé un outil puissant qui quantifie et analyse la structure des RP, appelé « analyse quantitative de récurrence » (Recurrence Quantification Analysis, RQA). Cet outil définit la mesure de complexité en utilisant la densité des points récurrents et les structures des lignes diagonales et verticales dans le diagramme de récurrence. Il permet de déceler les transitions entre les états périodiques, laminaires et chaotiques lors d'un processus donné.

Quelques études basées sur ces mesures de RQA montrent que ces mesures sont aptes à trouver des points de bifurcation, plus spécialement les transitions chaos-ordre. La RQA impose un seuil ε , ainsi le diagramme de récurrence est symétrique.

I.5.4.1. Les mesures de la RQA [40]

On peut distinguer deux types de mesure RQA, les mesures basées sur les lignes diagonales, ainsi que celles basées sur les lignes verticales (similaires aux lignes horizontales vu la symétrie).

Ainsi, nous introduisons l'histogramme $P(\varepsilon, l)$, représentant le nombre total de lignes diagonales de longueur l dans un RP avec un seuil ε fixé :

$$P(\varepsilon, l) = \sum_{i,j=1}^N \left(1 - R_{i-1,j-1}(\varepsilon)\right) \left(1 - R_{i+1,j+1}(\varepsilon)\right) \prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon) \quad (\text{I.23})$$

Par analogie, le nombre de lignes verticales est calculé en remplaçant l par v , où v représente la longueur de la ligne verticale. Dans les prochaines équations, nous omettrons de citer le symbole ε des mesures de RQA, pour une raison de simplicité ($P(l) = P(\varepsilon, l)$).

Les principales mesures RQA sont les suivantes.

- *Taux de récurrence (RR)*. Le RR compte le nombre de points noirs dans le RP. Il correspond à la probabilité qu'un état se répète et peut être utilisé pour détecter les changements dans la dynamique du système. Il est défini comme suit :

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\varepsilon) \quad (I.24)$$

- *Déterminisme (DET)*. Il renseigne sur la prédictibilité du processus étudié et représente le rapport entre le nombre de points récurrents formant les diagonales (au moins de longueur l_{min}) et le nombre totale de tous les points récurrents. Les signaux non corrélés, ou faiblement corrélés, apparaissent sous formes de courtes diagonales sur les RP. Les signaux déterministes impliquent des diagonales plus longues, ainsi que des points récurrents isolées.

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (I.25)$$

- *Laminarité (LAM)*. Par analogie, le LAM représente la quantité de points récurrents obtenus à partir des lignes verticales, il est donné par l'expression suivante :

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (I.26)$$

où $P(v)$ est le nombre de lignes verticales de longueurs v ayant au moins une longueur de v_{min} .

- *Longueur maximale des lignes diagonales (L_{max})*. L définit la moyenne temporelle pendant laquelle deux segments de trajectoire sont proches l'une de l'autre, et peut être interprétée comme la moyenne temporelle de prédiction.

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (I.27)$$

Au lieu de cette moyenne, la RQA utilise la « longueur maximale » des lignes diagonales, ou son inverse, la « divergence », définie comme suit :

$$L_{max} = \max(\{l_i ; i = 1 \dots N_1\}) \quad (I.28)$$

$$DIV = \frac{1}{L_{max}}. \quad (I.29)$$

Pour des systèmes déterministes, de longues diagonales sont produites, alors que pour des données aléatoires ou fortement fluctuantes, de courtes diagonales en sont le résultat.

- *Entropie (ENTR)*. L'entropie fait référence à l'entropie de Shannon de la distribution des lignes diagonales du RP : elle reflète la complexité de la structure du RP par rapport aux lignes diagonales. Pour des signaux faiblement corrélés, la valeur ENTR est faible, indiquant ainsi une faible complexité.

$$ENTR = - \sum_{l=l_{min}}^N p(l) \cdot \ln p(l) \quad (I.30)$$

où $p(l)$ représente la probabilité que la ligne diagonale ait exactement la longueur l , et peut être estimée à partir de la distribution de fréquence $P(l)$, avec :

$$p(l) = \frac{P(l)}{\sum_{l=l_{min}}^N P(l)} \quad (I.31)$$

- *Trapping Time (TT)*. Il représente la valeur moyenne temporelle qu'un système se répète sur un seul et même état. Il est défini par l'expression (I.32) :

$$TT = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N P(v)} \quad (I.32)$$

où $P(v)$ est le nombre de lignes diagonales de longueur v . TT estime la moyenne temporelle où le système demeure dans un même état.

- *Recurrence Time Entropy (RTE)*. Le RTE sert à déterminer la périodicité ou la répétitivité d'un signal. Il peut être appliqué aux systèmes dynamiques, ainsi qu'aux processus stochastiques. Il se calcule via l'expression (I.33):

$$RTE = -(\ln T_{max})^{-1} \sum_{t=1}^{T_{max}} p(t) \cdot \ln p(t) \quad (I.33)$$

où T_{max} est la plus grande période du signal. Le RTE est très utilisé pour détecter les changements subtils dans les dynamiques des signaux stochastiques.

Les indicateurs précédemment cités informent sur les différents états d'un processus étudié, et sont en réalité fortement inter-corrélés. Les mesures basées sur les lignes verticales sont capables de trouver les transitions chaos-chaos. Ces indicateurs permettent alors, d'étudier l'intermittence, même pour les séries de données courtes et non stationnaires. De plus, les transitions chaos-ordre peuvent être identifiées pour des systèmes dynamiques périodiques dont les mesures sont nulles.

I.6. Conclusion

Dans ce chapitre, nous avons présenté les plus célèbres méthodes utilisées actuellement pour le traitement de signal dans le domaine des isolateurs pollués. Compte tenu de leur nouveauté, les méthodes RP et RQA ont spécialement retenu notre attention. De ce fait, nous avons développé une méthodologie inédite de traitement du signal du courant de fuite qui sera exposée dans le chapitre IV de ce travail.

Chapitre II

METHODOLOGIE DE TRAITEMENT D'IMAGE

II.1. Introduction

L'information, dans sa définition la plus basique, correspond à un ensemble de données pouvant être exploitées et analysées en vue d'y extraire des indicateurs fiables. Ces derniers seront, par la suite, étudiés pour aboutir à des conclusions plus ou moins pertinentes. Or, l'information peut être véhiculée principalement sous deux formes, celle d'un signal ou d'une image.

Pour ce qui est de l'information sous forme d'un signal, bons nombres d'études et de techniques ont été élaborées et mises en place afin de garantir une analyse fiable. Dans le domaine de l'étude de la pollution sur les isolateurs électriques, nous pouvons citer la célèbre analyse de Fourier. Cette méthode, visant à décomposer un signal donné en une somme de fonctions trigonométriques, a fait preuve d'un franc succès quand il s'agit de diagnostiquer l'état de surface de l'isolateur pollué [1-7]. Aussi, la technique des ondelettes, de par son originalité, a réussi à combler les lacunes de l'analyse de Fourier et à apporter de nouveaux résultats relatifs à l'étude de la performance des isolateurs pollués [8-12].

Cependant, dans le cas où l'information est véhiculée sous forme d'une image, rares sont les études qui ont été effectuées. Ceci est dû au fait que le traitement d'image demeure un terrain vierge et non exploité dans le domaine de la pollution des isolateurs.

Dans ce chapitre, nous proposons une nouvelle méthodologie visant à utiliser le traitement d'image pour étudier, caractériser et diagnostiquer l'état de surface d'un isolateur pollué. Pour cela, le travail expérimental entrepris sera décrit. Puis, les différents outils de base du traitement d'image seront explicités.

II.2. Différentes étapes de propagation des décharges jusqu'au processus de contournement

En vue d'élaborer un algorithme de traitement d'image pour étudier les décharges électriques sur une surface isolante plane, les différentes étapes d'apparition et de propagation des décharges doivent être minutieusement étudiées et analysées. Pour cela, nous présentons dans cette section, un exemple d'images enregistrées (figure II.1), sous diverses valeurs de la tension appliquée, décrivant la dynamique de la décharge électrique jusqu'au contournement sur une surface plane soumise à une pollution continue (mélange d'eau distillée et NaCl) de 1,2 mS/cm. Les essais correspondants ont été réalisés par M. A. Douar et al [7].

Pour obtenir les images montrées sur la figure II.1, nous devons préalablement analyser les vidéos acquises. Puis, à partir de ces vidéos, nous procédons à l'extraction des images pour chaque niveau de tension. Ces images seront analysées et étudiées pour tirer les phénomènes typiques des décharges électriques conduisant au contournement. Dans notre travail, sept étapes, ou décharges typiques, précédents le contournement, sont relevées, comme montré par la figure II.1.

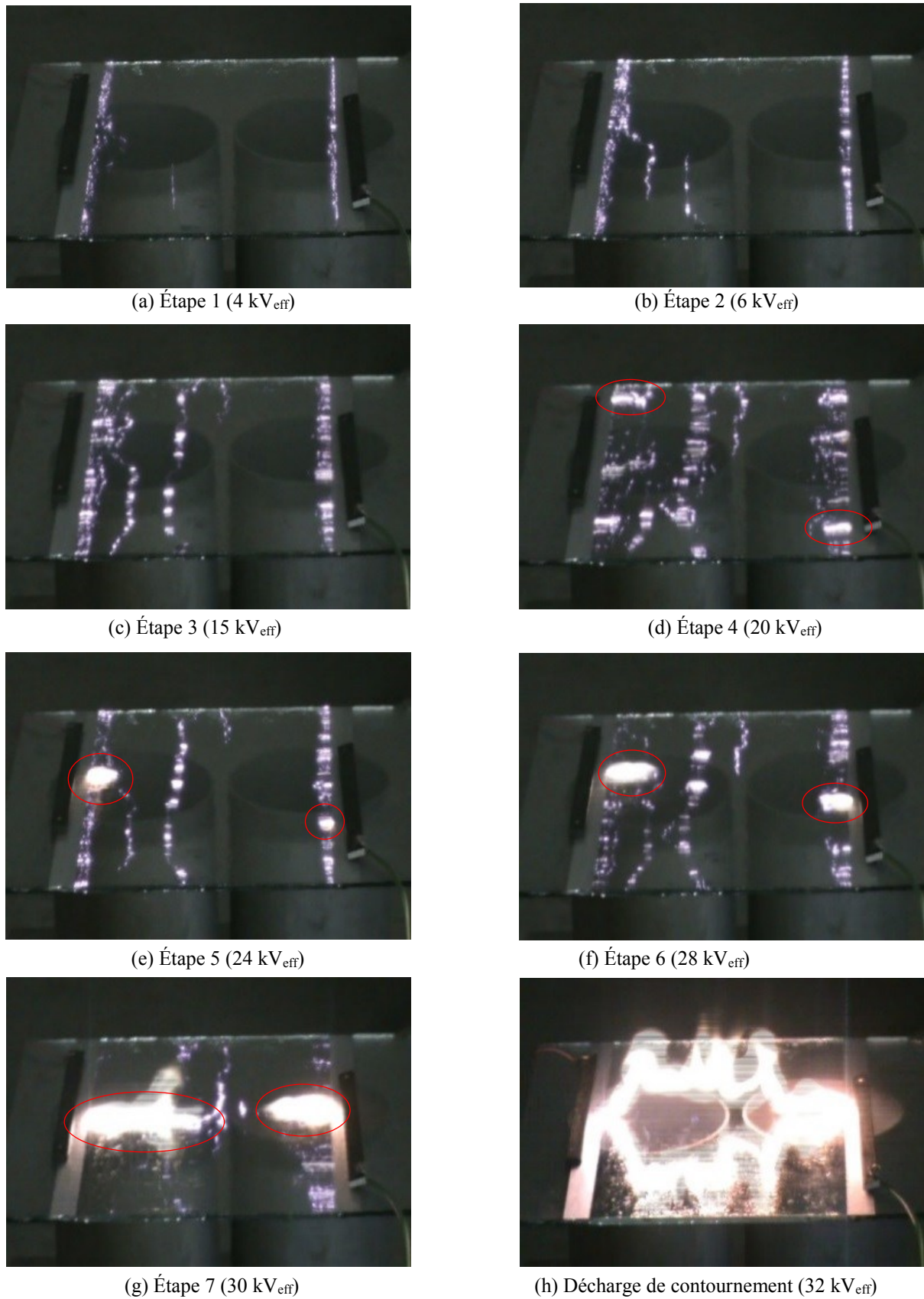


Figure II.1. Étapes du processus de contournement [7]

Suite à l'augmentation du niveau de tension à partir de 0 kV_{eff}, aucune activité sur la surface isolante n'est observé, et ce jusqu'à une tension appliquée de 3 kV_{eff} approximativement

Dès que le niveau de tension de $4 \text{ kV}_{\text{eff}}$ est atteint, nous observons l'apparition d'intensités lumineuses semblables à des points épars sur les zones proches des électrodes. Ces points lumineux sont l'étape initiale de formation d'une décharge électrique, comme montré par la figure II.1.a.

Par la suite, la tension appliquée au niveau de l'isolateur est augmentée pour atteindre les $6 \text{ kV}_{\text{eff}}$. Pour cette valeur de la tension, nous remarquons une intensification de la densité des points lumineux (figure II.1.b). Ces derniers ne sont plus aussi épars et dispersés, mais commencent à se regrouper pour former de faibles étincelles localisées principalement dans les zones proches des électrodes. Pour un niveau de tension de $15 \text{ kV}_{\text{eff}}$, ces faibles étincelles, appelées décharge brosse (brush discharge), constituent ainsi la deuxième étape du processus de contournement (figure II.1.c).

Cependant, dès que le niveau de tension de $20 \text{ kV}_{\text{eff}}$ est atteint, nous remarquons l'apparition d'une structure différente des figures précédentes ; les décharges électriques qui étaient, auparavant dispersées et éparses se relient entre elles pour former de petits arcs partiels. Cette étape, décrite par la figure II.1.d, marque la transition de la décharge électrique vers un arc. Ce dernier demeure un paramètre très important pouvant indiquer l'état de surface de l'isolateur et l'apparition du phénomène de contournement.

Sur la figure II.1.e, nous remarquons que les deux petits arcs partiels (entourés en rouge sur la figure II.1.d) ont disparu, laissant place à la formation de deux autres arcs à des endroits proches de ceux des précédents. De tels arcs sont appelés « arcs non-localisés » ; puisqu'ils changent de place très rapidement, avec ou sans augmentation du niveau de la tension appliquée. Cependant, pour une tension de $24 \text{ kV}_{\text{eff}}$, nous noterons que les arcs ont gagné en épaisseur et en intensité, devenant ainsi des arcs électriques denses.

Si le niveau de tension de $28 \text{ kV}_{\text{eff}}$ est atteint, les arcs précédemment relevés sur la figure II.1.e sont toujours présents aux mêmes endroits sur la figure II.1.f. Ces « arcs localisés » augmentent en épaisseur et en densité mais aussi en longueur. Leur couleur est devenue très claire et très vive. Nous remarquons également le début de formation d'arc au niveau du milieu de l'isolateur.

Pour une tension appliquée de $30 \text{ kV}_{\text{eff}}$, les arcs partiels, qui étaient localisés pour une tension inférieure, gagnent très considérablement en épaisseur et en densité. De plus, comme indiqué sur la figure II.1.g, l'arc présent sur la gauche de la figure II.1.f se connecte au petit arc au milieu de l'isolateur pour former un arc d'une longueur bien plus importante. Ce dernier tend à progresser pour se relier à l'arc progressant à partir de l'électrode de droite. Les deux arcs, l'un à gauche et l'autre à droite, sont appelés « arcs principaux », et ont une couleur qui tend vers le rouge, alors qu'elle était claire et lumineuse pour les niveaux de tensions inférieures.

La décharge de contournement apparaît alors pour une tension de $32 \text{ kV}_{\text{eff}}$. Elle naît à partir des deux arcs principaux qui se relient entre eux. Cette décharge représente l'étape finale du contournement.

Les étapes de contournement que nous avons relevées dans ce présent travail ont été confirmées par d'autres auteurs dans le cas d'un isolateur réel [25]. Ces étapes sont résumées dans le tableau II.1. A noter que sur ce tableau, nous proposons la discrimination entre deux principaux types de décharges : les décharges de types arc et celles de type non-arc.

Tableau II.1. Décharges électriques typiques précédant le contournement

Étape	Décharge électrique typique	Classe
1	Point lumineux	Décharge non arc
2	Faible étincelle	
3	Décharge de couleur violette (brushes)	
4	Petits arcs partiels	Décharge arc
5	Arcs denses de longueur moyenne	
6	Arcs localisés de couleur claire	
7	principaux arcs de couleur rouge	

A la base de la discrimination présentée dans le tableau ci-dessus, nous développerons dans les prochaines sections II. 3 et II.4 un algorithme qui vise à contrôler l'état de surface et l'activité des décharges à la surface de l'isolateur. Cet algorithme a pour vocation de déceler la présence des décharges électriques de type arc à partir des images enregistrées. A cet effet, les outils de traitement d'image nécessaire seront soigneusement développés dans la section II.3.

II.3. Notions basiques sur le traitement d'image

Etant une discipline de l'informatique et des mathématiques appliquées, le traitement d'image a principalement deux buts : améliorer la qualité des images en question et extraire les informations utiles. Ce dernier point est l'objet de notre travail. Les méthodes de traitement d'image ont pour rôle de manipuler des pixels. Ces derniers sont l'unité de base constituant une image numérique. Cette image est alors représentée sous forme de matrice. Chaque élément de cette matrice est un chiffre représentant un pixel de l'image.

Une large panoplie des méthodes permettant de manipuler les images peut être citée [43-57]. Cependant, dans notre travail, nous nous limiterons aux outils nécessaires au bon fonctionnement de notre algorithme de détection des décharges de type arc. La présente section sert donc à décrire ces différents outils [52-57].

II.3.1. Segmentation d'image

L'image, étant la représentation d'une perception visuelle donnée, véhicule généralement des informations sur de nombreux objets, parfois superposés ou mélangés. En vue d'extraire les informations sur un objet bien précis, ou une zone d'intérêt, une opération appelée « segmentation » est utilisée. En effet, si l'être humain est naturellement doté de l'aptitude de séparer les objets perçus, il demeure indispensable, pour une machine informatique, de développer des algorithmes de segmentation en vue d'extraire les zones d'intérêt à étudier.

Il existe trois grandes catégories de segmentations :

- Segmentation par région : ce type de segmentation exploite les caractéristiques propres de chaque région (surface, intensité lumineuse, colorimétrie, texture, etc.), pour fusionner les régions qui partagent le plus de caractéristiques [43].
- Segmentation par contour : basée sur l'hypothèse que l'image est composée de différentes régions connexes qui peuvent être séparées en détectant la frontière, ou transition, qui les sépare [44].
- Segmentation par seuillage (thresholding) : cette méthode se base sur l'intensité des pixels qui composent l'image, pour la classifier dans une certaine catégorie. Ainsi, pour chaque pixel, une opération de comparaison est effectuée avec un seuil (threshold) bien déterminé, puis une classe lui est attribuée [46-51].

Cette dernière méthode fera l'objet de notre travail et sera explicitée dans la prochaine section II.3.2.

II.3.2.Méthode de segmentation d'Otsu [46-51]

Comme il a déjà été dit dans la section précédente, le but de notre étude est de développer un algorithme fiable pour la détection des décharges de type arc. De ce fait, la première étape de traitement de ces images consiste à extraire la décharge en question. Pour cela, une segmentation par seuillage est opérée.

Lors de l'étape de segmentation d'un ensemble d'images, deux approches pour le choix du seuil sont utilisées : le choix d'un seuil constant tout au long de cette étape, ou bien le choix d'user d'un seuil dynamique, calculé pour chaque image. Cette dernière approche est le principe même de la méthode de segmentation d'Otsu.

Techniquement, le principe de la segmentation par seuillage consiste à assigner un seuil T à chaque pixel d'une image donnée. Ce pixel sera alors classé comme étant un composant de l'objet d'intérêt, ou bien une partie de l'arrière-plan de cet objet. Dès lors, en assumant f comme étant l'image initiale à segmenter, l'image segmentée g est binaire (composée de 0 et 1) et est donnée par :

$$g(x, y) = \begin{cases} 0, & f(x) < T \\ 1, & f(x) \geq T \end{cases} \quad (\text{II.1})$$

Un exemple préliminaire de cette segmentation est donné par la figure II.2, où l'image numérique originale (figure II.2.a), représentant des maisons ou habitations, comporte plusieurs niveaux de gris. Après le choix d'un seuil T arbitraire, les pixels de couleur claire ou blanche, étant de valeur inférieure à T , sont représentés par un 0 dans la matrice de g et donc un point blanc sur l'image segmentée (figure II.2.b). Alors que les pixels de la figure II.2.a, de couleur grise claire ou foncée, de valeur supérieure à T , sont représentés par un 1 dans la matrice g et donc un point noir (figure II.2.b).



Figure II.2. Exemple de segmentation par la méthode du seuil constant, (a) Image originale et (b) Image segmentée

Comme présenté sur la figure ci-dessus, le choix du seuil est primordial. Ainsi, si l'objet d'intérêt est les habitations représentées sur la figure II.2, il est évident que la segmentation appliquée est faillible. Le seuil que nous avons choisi a pour effet de considérer les pixels à haute intensité comme zone d'intérêt, alors que les pixels à faible intensité sont éliminés. En d'autres termes, le choix d'un tel seuil peut ignorer des détails très importants sur l'image. D'autant plus si l'étude vise à comparer ou étudier une variété d'images, le choix d'un seul seuil pour le traitement de toutes les images devient encore plus compliqué.

De ce fait, l'automatisation du choix du seuil optimal doit être entreprise. Dès lors, nous proposons l'utilisation de la méthode de segmentation d'Otsu. Cette dernière assume que l'image à segmenter contient deux classes de pixels, ceux appartenant au premier plan (foreground) et d'autres à l'arrière-plan (background). Puis, la méthode d'Otsu vise à calculer le seuil optimal séparant ces deux plans. Un tel seuil optimal est choisi pour minimiser l'intra-variance entre ces deux classes, donc maximiser l'inter-variance. Pour rappel, l'intra-variance est définie comme étant la somme pondérée des variances des deux classes, comme décrit par l'équation (II.2) :

$$\sigma_{intra}^2(T) = w_B(T)\sigma_B^2(T) + w_F(T)\sigma_F^2(T) \quad (II.2)$$

où σ_B^2 et σ_F^2 sont les variances des pixels de l'arrière-plan (en dessous du seuil) et le premier plan (au-dessus du seuil) respectivement. w_B et w_F sont les poids (probabilités) des classes de pixels calculés par les équations (II.3) and (II.4) respectivement :

$$w_B(T) = \sum_{i=0}^{T-1} P(i) \quad (II.3)$$

$$w_F(T) = \sum_{i=T}^{N-1} P(i) \quad (II.4)$$

où $[0, N - 1]$ est l'intervalle des niveaux d'intensité de l'image.

Pour calculer l'inter-variance, nous soustrayons alors l'intra-variance de la variance totale, comme montré par l'équation (II.5) :

$$\sigma_{inter}^2(T) = \sigma^2(T) - \sigma_{intra}^2(T) \quad (II.5)$$

En remplaçant l'équation (II.2) dans (II.5), nous obtenons :

$$\sigma_{inter}^2(T) = w_B(T)[\mu_B(T) - \mu]^2 + w_F(T)[\mu_F(T) - \mu]^2 \quad (II.6)$$

où σ^2 est la variance totale de toute l'image et μ est sa moyenne (calculée à partir de l'équation (II.7)).

$$\mu = w_B(T) \mu_B(T) + w_F(T) \mu_F(T) \quad (\text{II.7})$$

En substituant l'équation (II.7) dans (II.6), nous obtenons une formulation bien plus simplifiée de l'inter-variance, comme montré par l'équation (II.8).

$$\sigma_{inter}^2(T) = w_B(T)w_F(T)[\mu_B(T) - \mu_F(T)]^2 \quad (\text{II.8})$$

Le seuil optimal T maximise l'inter-variance. Dès lors, T peut être obtenu avec de plus simples relations de récurrences, basées sur l'incrémentement de $w_B(T)$, $w_F(T)$ et leurs moyennes respectives $\mu_B(T)$ et $\mu_F(T)$, comme présentés par les équations (II.9) à (II.12).

$$w_B(T + 1) = w_B(T) + w_T \quad (\text{II.9})$$

$$w_F(T + 1) = w_F(T) - w_T \quad (\text{II.10})$$

$$\mu_B(T + 1) = \frac{\mu_B(T)w_B(T) + w_T T}{w_B(T+1)} \quad (\text{II.11})$$

$$\mu_F(T + 1) = \frac{\mu_F(T)w_F(T) - w_T T}{w_F(T+1)} \quad (\text{II.12})$$

La figure II.3 montre un exemple pratique, de décharge pendant sa propagation Uniformément polluée (la conductivité étant égale à 1,2 mS/cm) [7]. Pour prouver l'efficacité de la méthode d'Otsu comparée à la méthode du seuil fixe, nous proposons de comparer les résultats de cette première méthode (figure II.3.c), par rapport aux résultats donnés par la segmentation à seuil constant (figure II.3.b et II.3.d).

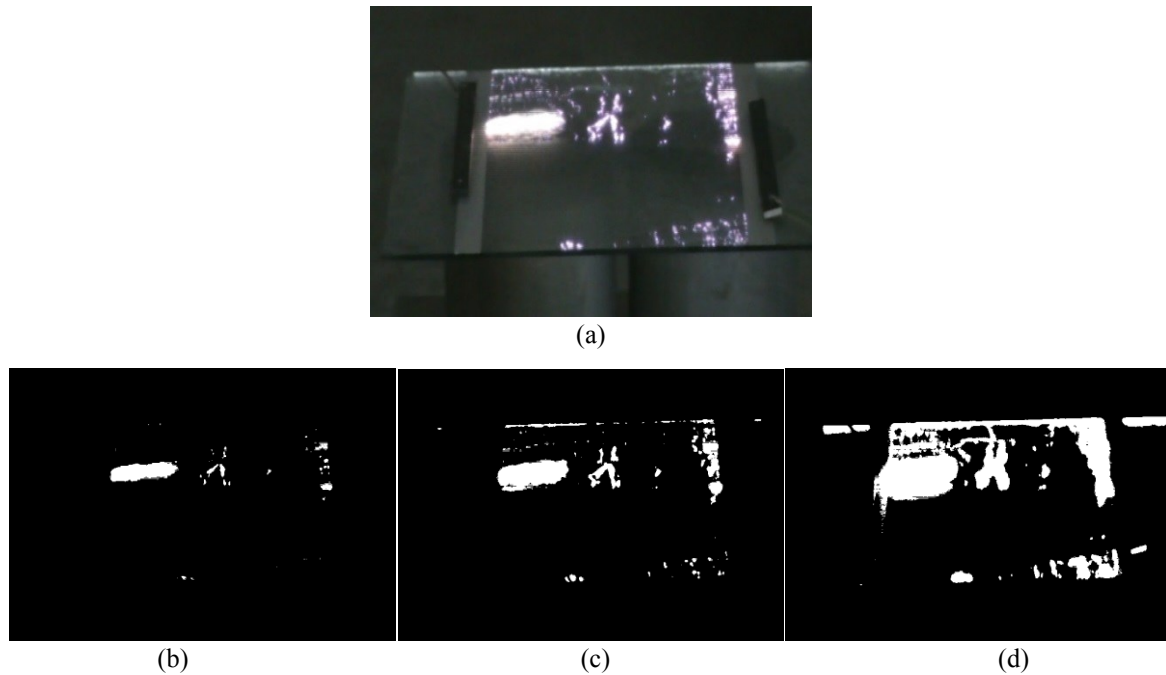


Figure II.3. Exemple de segmentation avec l'image (a) originale, (b) segmentée pour un seuil de 0,3, (c) pour un seuil optimal de 0,65 (méthode d'Otsu) et (d) pour un seuil de 0,9 [7]

Nous choisissons, tout d'abord, de fixer le seuil à une faible valeur de 0,3 (ce dernier varie entre 0 et 1). Nous remarquons alors sur l'image segmentée (figure II.3.b) que de nombreuses informations concernant les décharges qui apparaissent sur la surface de l'isolateur ont été supprimées. Les luminiscences et petits arcs proches de l'électrode de droite ont été presque entièrement effacés sur la figure II.3.b. De même pour les décharges se trouvant au milieu de l'isolateur, ainsi que l'arc se propageant à partir de l'électrode de gauche qui a été sensiblement rétréci.

Puis, nous présentons le résultat de la segmentation par la méthode d'Otsu (seuil optimal). Dans ce cas, l'image segmentée (figure II.3.c) expose d'une manière bien plus appropriée l'activité des décharges sur la surface isolante. En effet, les luminosités montrées sur les rebords du modèle de l'isolateur (figure II.3.a) sont désormais visibles sur l'image segmentée (figure II.3.c). De même pour les faibles et éparses décharges en bas à droite de l'isolateur. Comme observation générale, nous noterons que la segmentation d'Otsu, permet grâce à son seuil optimal de ne pas ignorer d'importants détails de l'activité des décharges à la surface isolante, à l'inverse du choix d'un seuil en dessous de l'optimal (figure II.3.b).

Enfin, nous proposons d'analyser le cas où nous fixons le seuil à une valeur supérieure à l'optimale (0,9). Présenté par la figure II.3.d, le résultat de la segmentation expose clairement une représentation exagérée et non adéquate des décharges électriques. Les petites luminosités sont très grossièrement traduites sur l'image segmentée. Etant supérieur à sa valeur optimal, le choix d'un tel seuil traduit d'une façon excessive chaque petite luminosité en pixel blanc sur l'image segmentée, offrant ainsi une représentation incorrecte de l'activité des décharges électriques à la surface isolante.

Suivant l'analyse comparative effectuée dans ce paragraphe, la méthode de segmentation d'Otsu démontre une supériorité incontestée face à la segmentation par seuil constant. Dès lors, afin d'élaborer notre algorithme de détection des arcs sur la surface isolante, notre choix pour la méthode d'Otsu devient évident.

II.3.3. Filtrage morphologique [37-42]

II.3.3.1. Nécessité du filtrage morphologique

L'image traitée dans ce chapitre, même étant capturée dans un laboratoire, est loin d'être sans bruit. En effet, l'image, peu importe l'environnement duquel elle est prélevée, véhicule toujours une certaine quantité, plus ou moins importante, de bruit.

Dans notre cas, notre travail s'est porté sur l'étude des décharges sur un modèle d'isolateur plan (voir annexe). Cependant, ce choix ne demeure pas sans conséquence. En effet, le choix d'un tel profil de l'isolateur permet, certes, d'avoir une meilleure vision sur la propagation des décharges, mais en contrepartie, cause la réflexion de lumière sur le bord supérieur de ce modèle. Considérée comme bruit, cette réflexion de lumière est, par la suite, confondue avec l'activité des décharges électriques, rendant ainsi notre diagnostic perturbé. Ce constat est mis en avant par la figure II.4 [7].

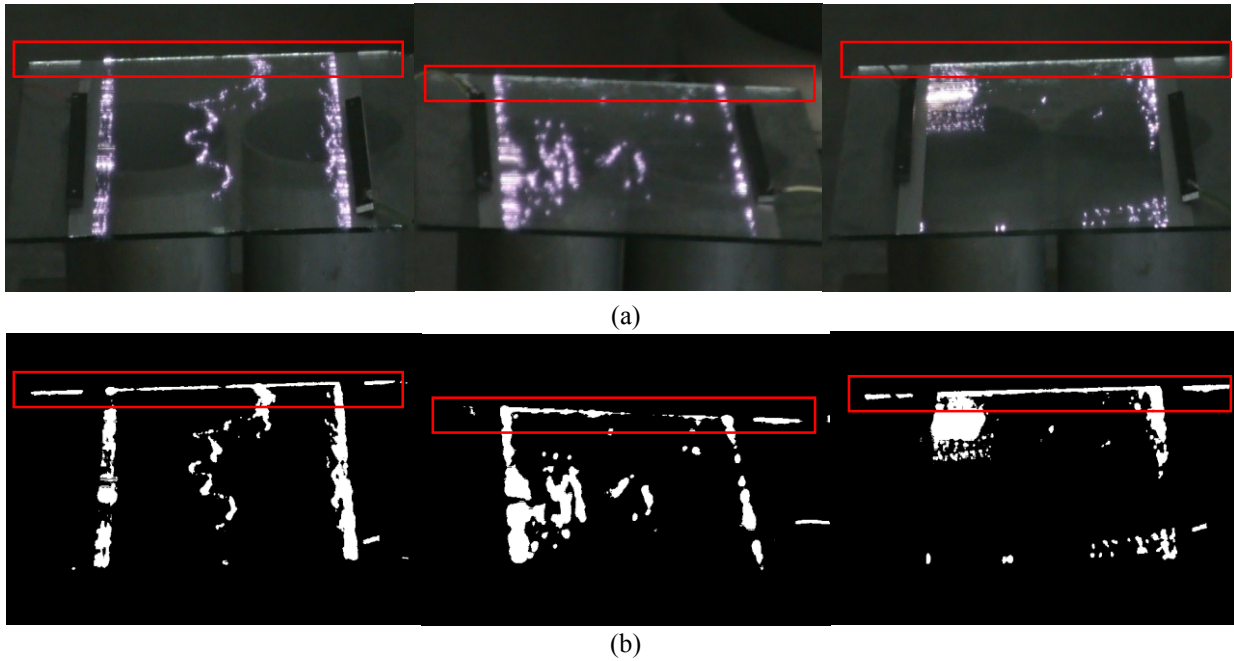


Figure II.4. Bruit présent sur les images, causé par la réflexion de la lumière sur le bord supérieur de l'isolateur plan (a) image originale [7] et (b) image segmentée par la méthode d'Otsu

Entourée en rouge sur la figure II.4, la lumière du jour et celle de l'éclairage du laboratoire d'essai sont réfléchies sur le bord supérieur de l'isolateur. A noter que ce bruit est présent sur toutes les images prélevées au laboratoire et demeure inévitable étant donné que nous ne pouvons exempter nos tests de la lumière du laboratoire d'essai. De ce fait, une méthodologie bien adaptée doit être mise en application pour supprimer ce bruit, tout en minimisant son impact sur le reste de l'image. Dès lors, la présente section a pour but de mettre en exergue l'utilité du filtrage morphologique des images pour permettre un diagnostic correct de l'état des décharges sur la surface isolante. A noter que le filtrage morphologique sera désormais opéré sur l'image segmentée par la méthode d'Otsu, et non sur l'image originale.

II.3.3.2. Principe du filtrage morphologique et de l'élément structurant [57]

Le principe du filtrage morphologique consiste à utiliser un élément structurant B pour apporter des modifications sur une image donnée A . L'élément structurant B est un ensemble de pixels, possédant deux caractéristiques principales : il a une forme géométrique bien connue et il est repéré par son origine O . Dès lors, cet élément structurant B sera utilisé comme sonde, il sera placé sur chaque pixel de l'image à traiter, pour vérifier s'il est inclus dans l'ensemble initial de cette image. Autrement dit, l'origine O de l'élément B sera posée sur chaque pixel de l'image A . Puis, un certain nombre de pixels sera rajouté ou supprimé, en fonction de si B est inclus dans A ou pas.

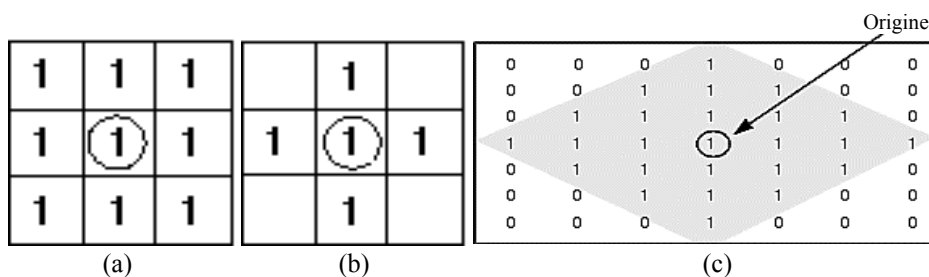


Figure II.5. Exemple d'élément structurant : (a) carré de taille 3, (b) losange de taille 2 et (c) losange de taille 4

II.3.3.3. Outils de base du filtrage morphologique [52-57]

En vue d'utiliser l'élément structurant pour effectuer un filtrage sur les images, différents outils doivent être utilisés [52-57]. Dans ce paragraphe, nous explicitons les quatre opérations de base du filtrage morphologique.

A. Érosion

Dans le cas de l'érosion, quand l'élément structurant est confondu avec le pixel de l'image à traiter, s'il n'est pas contenu par cette image, alors les pixels de cette dernière image seront tous remplacés par 0. En d'autres termes, l'image érodée contient un 1 uniquement au cas où l'élément structurant est contenu dans l'image. Dès lors, l'érosion supprime les bords de l'image à traiter et tend à la rétrécir.

Soit E l'espace euclidien, et A est l'image binaire appartenant à E . L'érosion de l'image binaire A par l'élément structurant B est donnée par :

$$A \ominus B = \{z \in E | B_z \subseteq A\} \quad (\text{II.13})$$

où B_z est la translation de B par le vecteur z , i.e:

$$B_z = \{b + z | b \in B\}, \forall z \in E \quad (\text{II.14})$$

Autrement dit, en assumant que l'origine de B se trouve à son centre. Nous superposons l'origine de B en chaque pixel de A . Si B est complètement contenu dans A , alors le pixel en question est retenu, sinon il est supprimé.

Etant donné que dans notre travail, le filtrage morphologique est effectué sur une image segmentée par la méthode d'Otsu (une image binaire), nous présentons un exemple d'érosion sur une matrice binaire, comme suit :

Soit A une image binaire et B un élément structurant, tels que :

$$A = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \quad \text{et} \quad B = \begin{matrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{matrix}$$

Alors l'érosion de A par B est donnée par :

```

0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0

```

B. Dilatation

L'érosion, étant un outil de suppression de pixels, apporte bien évidemment une perte de l'information. L'érosion doit donc être utilisée avec précaution. Dès lors, nous préconisons l'emploi d'un autre outil, analogue à l'érosion, appelé « la dilatation ». En effet, la dilatation a pour but de rajouter des pixels dans certaines zones de l'image, en remplissant les vides et en connectant les objets dispersés. L'image dilatée contient alors des 1 dans tous les pixels où l'élément structurant touche l'image à traiter. La dilatation de A par B est donnée par:

$$A \oplus B = \{z \in E \mid (B^s)_z \cap A \neq \emptyset\} \tag{II.15}$$

où B^s indique la matrice symétrique de B et est définie par:

$$B^s = \{x \in E \mid -x \in B\} \tag{II.16}$$

C'est-à-dire, si B a son origine à son centre, la dilatation de A par B peut être représentée par tous les points touchés par B quand son centre se déplace à l'intérieure de A .

Comme exemple, soit A une image binaire et B un élément structurant, tels que :

<pre> 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 A = 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 </pre>	et	<pre> 1 1 1 </pre>
--	----	--------------------

Alors la dilatation de A par B est donnée par:

```

0 0 0 1 1 1 1 1 0 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
0 0 0 1 1 1 1 1 1 1 0 0 0
    
```

Les figures II.6 et II.7 suivantes représentent un exemple d'application des opérations d'érosion et de dilatation sur une matrice binaire et une figure géométrique respectivement.

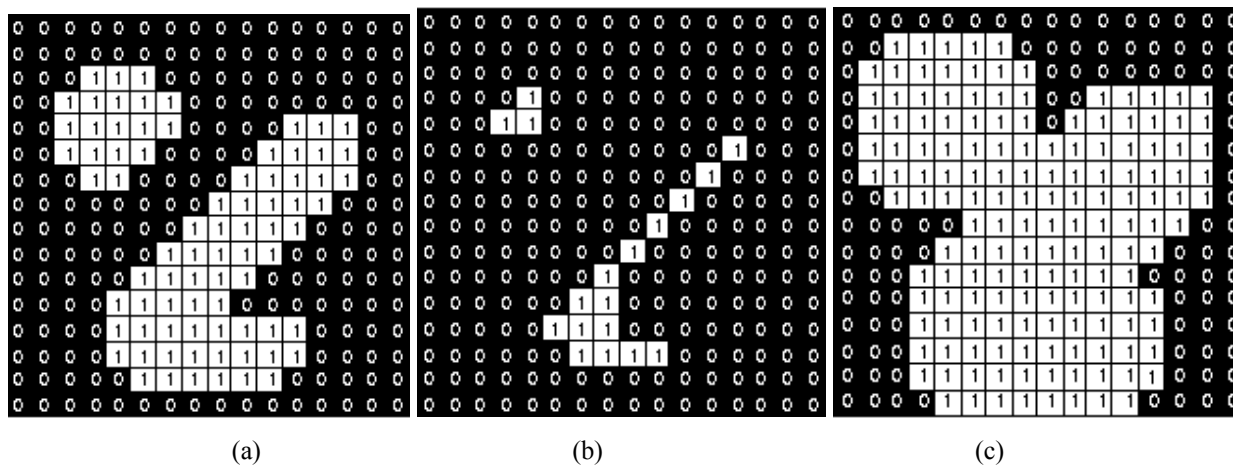


Figure II.6. Filtrage morphologique sur une image binaire :
 (a) image original, (b) image érodée et (c) image dilatée par un élément structurant de type carré 3x3

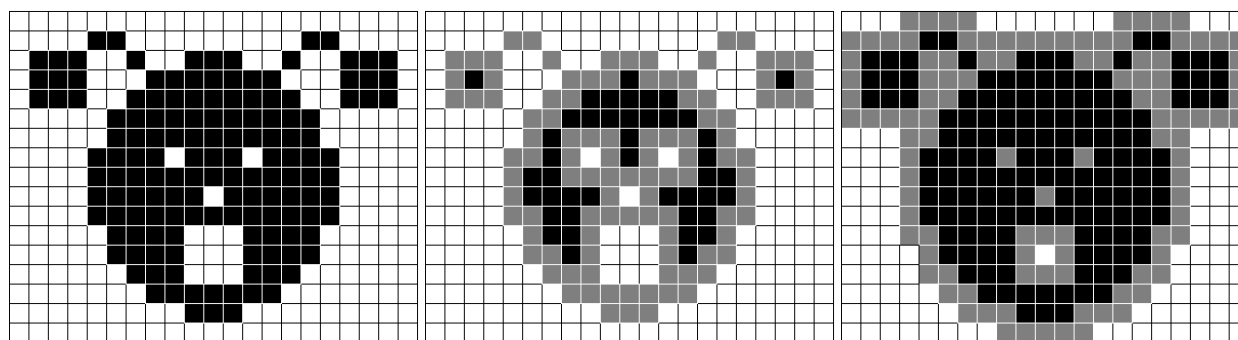


Image originale (en noir : l'objet ; en blanc : le fond) Érosion par un carré 3x3 : seuls les pixels noirs font partie de l'ensemble résultant Dilatation par un carré 3x3 : les pixels noirs et gris font partie de l'ensemble résultant

Figure II.7. Exemple d'opération d'érosion et de dilatation sur une figure géométrique

C. Ouverture

L'érosion et la dilatation, étant des opérations de base de traitement morphologique d'image, peuvent être combinées. En effet, si une image donnée est tout d'abord érodée, puis dilatée par le même élément structurant, nous dirons que cette image a subi une opération d'ouverture. Dès lors, l'ouverture de A par B est donnée par :

$$A \circ B = (A \ominus B) \oplus B \tag{II.17}$$

L'opération d'ouverture a pour effet de préserver les régions de l'avant-plan, qui ont une forme similaire à l'élément structurant, alors que les régions de l'arrière-plan sont éliminées. Par conséquent, l'opération d'ouverture semble bien adaptée à notre cas, où notre but est de supprimer les pixels causés par la réflexion de lumière sur le bord supérieur du modèle d'isolateur plan, alors que nous voulons garder intact le reste des pixels de la décharge électrique.

D. Fermeture

Analogue à l'ouverture, la fermeture de A par B est donnée par :

$$A \bullet B = (A \oplus B) \ominus B \tag{II.18}$$

L'opération de fermeture sert à élargir les régions et frontières de l'avant plan, alors qu'elle permet de supprimer celles de l'arrière-plan. De ce fait, la fermeture préserve uniquement les régions de l'arrière-plan pouvant contenir l'élément structurant, et élimine les autres régions, comme illustré par la figure II.8.

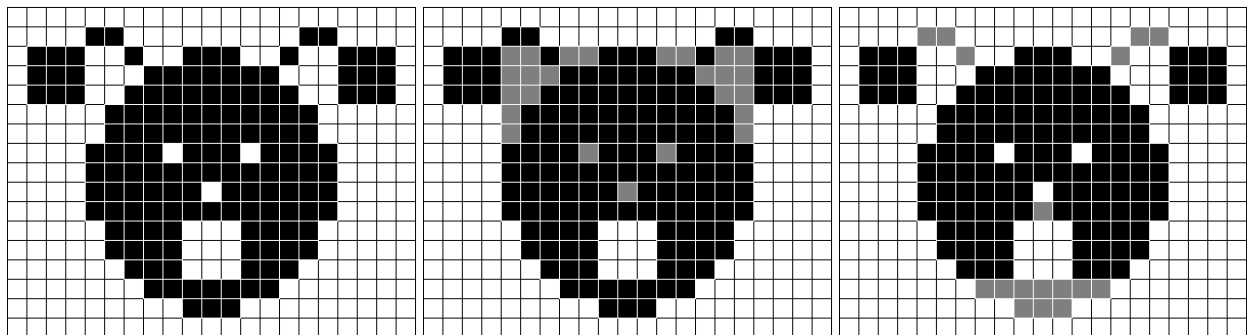


Image originale

Fermeture par un carré 3x3 : les pixels noirs et gris font partie de l'ensemble résultant

Ouverture par un carré 3x3 : seuls les pixels noirs font partie de l'ensemble résultant

Figure II.8. Exemple d'opération d'ouverture et de fermeture on sur une figure géométrique

Afin de bien comprendre le fonctionnement de ces quatre opérations de base du filtrage morphologique, nous proposons le tableau II.3 qui résume leurs principales caractéristiques.

Tableau II.3. Résumé des opérations de base de filtrage morphologique

Opération morphologique	Opération ensembliste	Opération logique	Principe
Erosion	Intersection	ET	Minimum (de l'élément structurant dans l'image) ; intersection de l'élément structurant avec l'image
Dilatation	Union	OU	Maximum (de l'élément structurant dans l'image) ; union de l'élément structurant avec l'image
Ouverture	Intersection puis union	ET puis OU	Erosion suivie d'une dilatation
Fermeture	Union puis intersection	OU puis ET	Dilatation suivie d'une érosion

II.3.4. Etiquetage des composantes connexes [57]

Suite au filtrage de l'image préalablement segmentée avec la méthode d'Otsu, nous obtenons une matrice binaire, composée de pixels ayant donc les valeurs 0 et 1. Les pixels ayant les valeurs 1 appartiennent à l'avant plan et représentent la décharge électrique. Ils sont décrits par un point blanc sur l'image segmentée et filtrée. Alors que les pixels de valeur 0 appartiennent à l'arrière-plan. Sachant que les pixels de valeur 1 sont bien évidemment l'objet de notre attention, nous devons donc utiliser un outil pour pouvoir, d'abord, les extraire ou les isoler du reste de l'arrière-plan pour pouvoir mieux les étudier et en tirer le maximum d'informations.

Pour cela, nous proposons de mettre en œuvre la méthode d'étiquetage des composantes connexes. Pour chaque région ou composante de l'image, cet outil a pour but d'assigner une étiquette à chaque composante connexe d'une image donnée. A noter qu'une composante connexe est définie comme étant un ensemble de pixels de même valeur, rassemblé ou réuni, sous forme d'un groupe ou d'une région. Dans ce présent travail, l'étiquetage des composantes connexes sert à donner une étiquette à chaque région de pixels de valeur 1. De cette manière, chaque ensemble de pixels, définissant une décharge électrique plus ou moins importante, peut être étudié et mesuré. Un exemple de cette méthode d'étiquetage est explicité par la figure II.9.

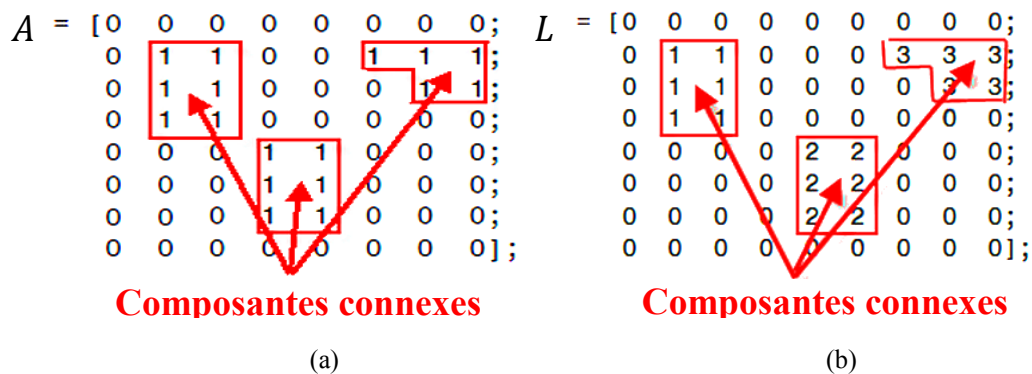


Figure II.9. Exemple d'étiquetage des composantes connexes

Soit A , la matrice représentant l'image segmentée par la méthode d'Otsu. Cette image étant de type binaire, elle ne contient que des pixels de valeur 0 et 1. La première étape d'étiquetage des composantes connexes consiste à détecter ces composantes. Pour cela, chaque région, ou ensemble, de pixels de valeur 1, est détectée et entourée par une couleur rouge, comme sur la figure II.9.a. Puis, la matrice L est créée pour contenir les étiquettes de chaque composante connexe de la matrice A . De gauche à droite, des étiquettes, portant des numéros croissant à partir de 1, sont assignées pour chaque région comportant des composantes de valeur 1, comme montré sur la figure II.9.b.

II.4. Conclusion

Dans ce chapitre, les notions basiques sur le traitement d'image ont été exposées. De plus, ces différentes notions ont été appliquées sur les images montrant l'activité des décharges électriques sur un modèle d'isolateur plan.

Etant donné que le chapitre III vise à développer un algorithme de traitement d'image, ce présent chapitre est donc d'une importance primordiale. En effet, les différents outils décrits

seront par la suite combinés et utilisés afin de constituer un algorithme efficace pour le diagnostic de l'état de surface d'un isolateur pollué.

Chapitre III

ELABORATION D'UN ALGORITHME DE TRAITEMENT D'IMAGE POUR L'ETUDE DES DECHARGES ELECTRIQUES

III.1. Introduction

Etant donné que dans le chapitre précédent, nous avons introduit différentes notions relatives au traitement d'images de décharges électriques se produisant sur un isolateur plan, nous proposons dans ce chapitre de mettre en application ces différentes notions.

En effet, nous nous sommes intéressés à élaborer un algorithme pour la caractérisation des décharges électriques sur une surface isolante plane soumise à divers degrés de pollution. L'algorithme sera testé et validé en utilisant des méthodes de classification nous permettant de déceler l'apparition des décharges électriques de type arc sur la surface isolante.

III.2. Elaboration d'un algorithme de traitement d'image pour l'étude des décharges électriques

Comme il a été déjà explicité dans la section II.2 du précédent chapitre, le processus de contournement peut être subdivisé en 8 étapes. Chaque étape représente un état bien défini de l'évolution des décharges électriques à la surface de l'isolateur. Parmi ces 8 étapes, nous avons décelé une transition majeure de l'état des décharges, passant de petites décharges localisées, jusqu'à la formation d'arcs partiels court-circuitant une partie, plus ou moins importante, de la surface isolante polluée. De ce fait, nous suggérons l'élaboration d'un algorithme, à base de traitement d'image, pour détecter l'apparition de décharges électriques de type arc.

Pour cela, notre algorithme se subdivisera en quatre étapes, comme indiqué par le schéma de la figure III.1 :

1. Segmentation des images par la méthode du seuil optimal d'Otsu.
2. Filtrage morphologique des images par ouverture (érosion suivie d'une dilatation).
3. Extraction du vecteur des caractéristiques à partir de l'image segmentée.
4. Classification de l'image étudiée pour déterminer la présence d'arcs électriques.

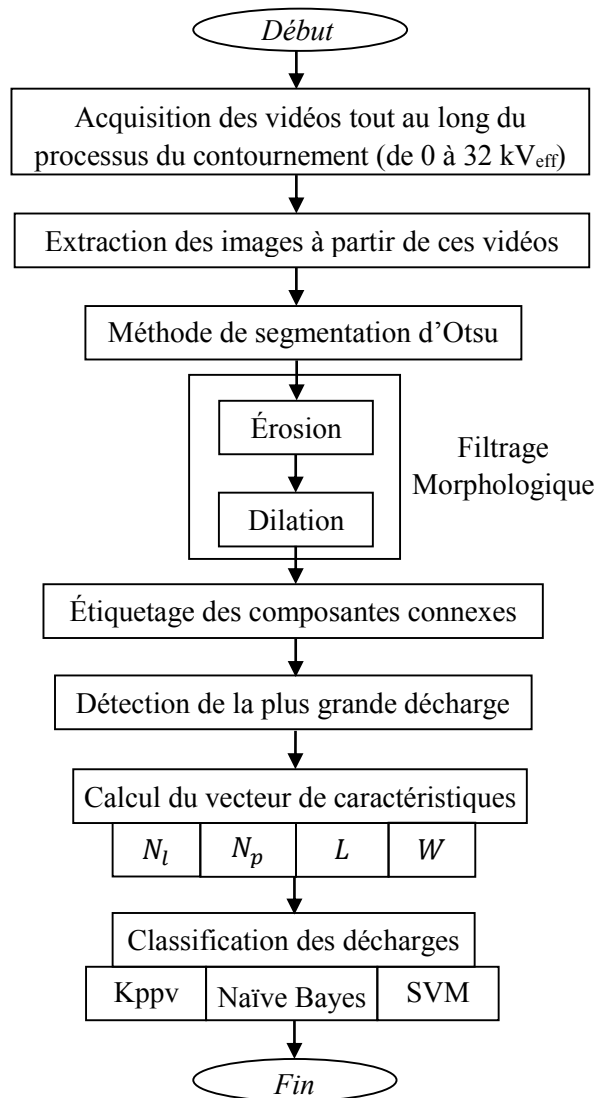


Figure III.1. Algorithme de traitement d'image proposé pour la détection des arcs électriques

III.2.1. Segmentation et filtrage

Une fois que l'acquisition des vidéos de l'évolution des décharges électriques jusqu'au contournement est accomplie, la première étape de l'algorithme consiste en la segmentation de ces images par la méthode d'Otsu. Cette étape est la base de notre algorithme. Puis, à partir de l'image segmentée, la seconde étape de filtrage morphologique prend place.

La figure III.2 montre l'utilité du filtrage morphologique qui succède à la segmentation. En effet, le reflet de la lumière du jour et celle de l'éclairage du laboratoire d'essai sont bien visibles sur le bord supérieur du modèle plan de l'isolateur (figure III.2.a). Cette lumière est représentée par des pixels blancs sur l'image segmentée (figure III.2.b). Si nous ne procédons pas au filtrage morphologique, alors la détection de la plus grande décharge à la surface isolante sera erronée. Ce résultat est montré par la figure III.2.c, où nous voyons le contour, de forme carré et de couleur bleu, encercler la totalité de la longueur de la surface isolante. Ainsi, la plus grande décharge a été surdimensionnée et donc faussement détectée par notre algorithme. Un tel résultat est dû au bruit présent sur l'image. De ce fait, nous appliquons le filtrage morphologique sur l'image 2.b, par la succession d'érosion et de dilatation (ouverture).

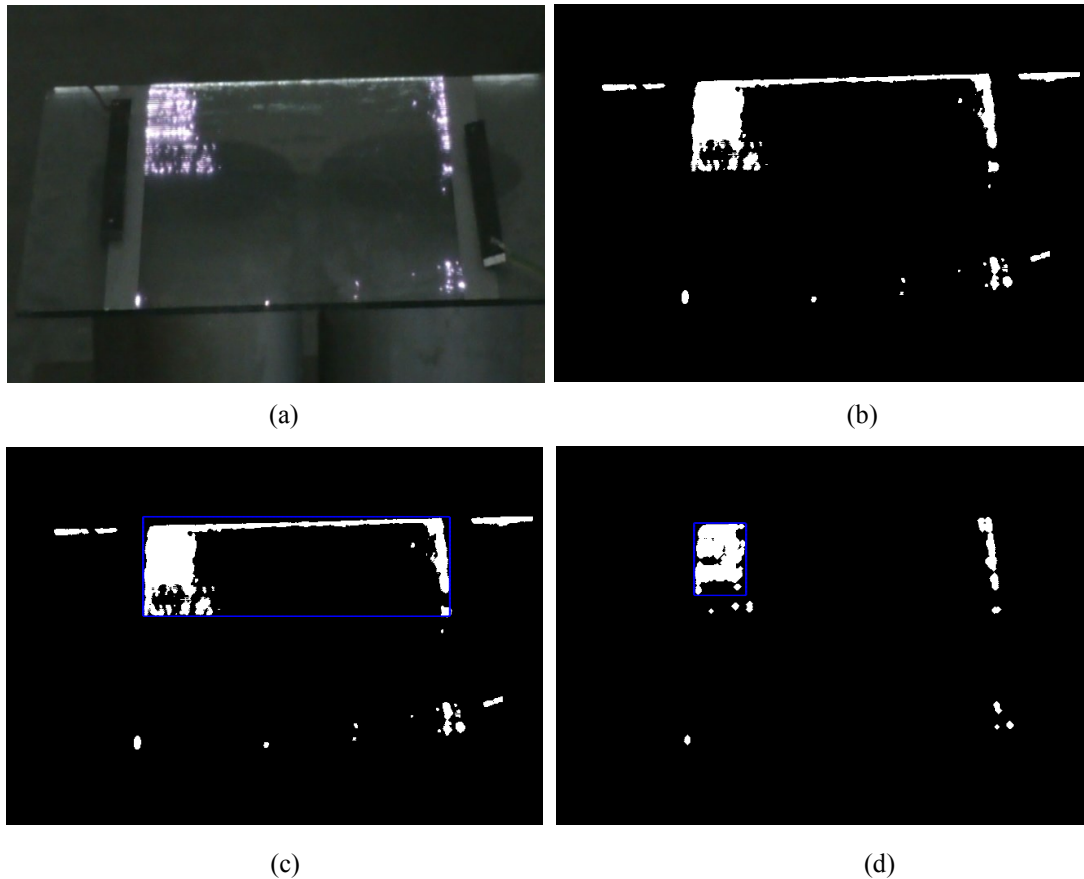


Figure III.2. Nécessité du filtrage morphologique après segmentation pour détecter la plus grande décharge
 (a) Image originale [7], (b) image segmentée, (c) image étiquetée sans filtrage et (d) image étiquetée avec filtrage

La nécessité du filtrage morphologique sur l'image segmentée a été mise en avant par la figure III.2. Cette dernière démontre un exemple de détection de la plus grande décharge dépendamment de la présence ou non du filtrage morphologique. Le point suivant visera à expliciter en détails comment la détection de la plus grande décharge se déroule.

III.2.2. Extraction des caractéristiques

L'extraction des caractéristiques à partir de l'image filtrée est la troisième étape de l'algorithme proposé. Cette étape est cruciale, car elle donne des indicateurs quantitatifs (N_l , N_p , L et W) aptes à représenter l'état des décharges électriques à la surface isolante. Cette présente section servira à décrire et justifier en détails comment ces indicateurs, au nombre de quatre, sont calculés, ainsi que leur utilité pour l'étude des décharges électriques.

Le calcul du vecteur des caractéristiques se fait en deux étapes. La première consiste en la détection de la plus grande décharge, permettant le calcul de deux indicateurs, à savoir N_l et N_p . Puis, sur la base de cette décharge, deux autres indicateurs seront calculés, à savoir L et W .

III.2.2.1. Détection de la plus grande décharge (calcul de N_l et N_p)

Comme il a été préalablement décrit dans la section II.2 du précédent chapitre, le contournement d'un isolateur survient à partir de la plus grande décharge. Cette dernière, suite à une légère augmentation de la tension appliquée, se développe et arrive à court-circuiter la totalité

de la surface isolante. De ce fait, il va de soi que l'algorithme proposé dans ce chapitre se base sur les caractéristiques de cette plus grande décharge pour évaluer l'état de la surface sous conditions de pollution.

Pour détecter la plus grande décharge sur la surface isolante, nous procédons d'abord par étiquetage des composantes connexes, comme préalablement décrit dans le paragraphe II.3.4 du précédent chapitre. Puis, pour chaque zone étiquetée, nous calculons le nombre de pixels de valeur 1 qu'elle contient. Dès lors, deux indicateurs peuvent être définis :

- N_l : représente le nombre de composantes connexes étiquetées sur l'image après segmentation et filtrage morphologique. En d'autres termes, N_l correspond au nombre de décharges électriques, de taille plus ou moins importante, sur la surface isolante. Un faible nombre de N_l implique une présence de peu de décharges, alors qu'un grand nombre indique de nombreuses décharges. A noter que N_l ne prend pas en compte la taille des décharges, mais uniquement leur nombre.
- N_p : indique le nombre de pixels que contient la plus grande composante connexe étiquetée (entourée par un carré bleu sur la figure III.2.c). En pratique, N_p représente le nombre de pixels que couvre la plus grande décharge électrique. Contrairement à N_l , N_p est relié à la taille de la décharge. Si N_p est important, alors la plus grande décharge à la surface isolante occupe une grande partie de cette dernière. Sinon, si N_p est faible, il indique que l'activité des décharges est concentrée sur une petite zone.

III.2.2.2. Calcul des deux indicateurs quantifiant la plus grande décharge (L et W)

Dès que la plus grande décharge est bien cernée et le nombre de pixels qu'elle contient est bien connu, nous pouvons alors calculer deux indicateurs géométriques qui la quantifient. Pour cela, nous proposons le calcul de sa longueur (L) et de sa largeur (W) :

- L : correspond à la longueur de la plus grande composante connexe étiquetée. Dès lors, il représente la longueur de la plus grande surface comportant les décharges électriques.
- W : définit la largeur de la plus grande composante connexe étiquetée. Autrement dit, il symbolise la longueur de la plus grande décharge électrique.

III.2.2.3. Exemple pratique de calcul des quatre indicateurs

Afin de clarifier et d'explicitier le calcul des quatre indicateurs formant le vecteur de caractéristique, nous proposons un exemple pratique sur une matrice binaire.

Soit A , la matrice binaire d'une image segmentée et filtrée, décrivant la surface isolante pour un certain état d'activité des décharges électriques.

$$A = \begin{matrix} & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{matrix}$$

Tout d'abord, nous commençons par entourer les composantes connexes avec un carré rouge, comme suit :

$$B = \begin{matrix} 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{matrix}$$

Puis, l'étiquetage de ces composantes connexes est effectué. De gauche à droite, nous donnons un label (numéro croissant) à partir de 1 pour chaque composante. La matrice étiquetée L est donnée par :

$$L = \begin{matrix} 0 & 0 & 2 & 2 & 0 & 4 & 4 & 4 & 4 \\ 1 & 0 & 2 & 2 & 0 & 0 & 4 & 4 & 4 \\ 1 & 0 & 0 & 0 & 0 & 0 & 4 & 4 & 4 \\ 1 & 0 & 0 & 3 & 3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 3 & 3 & 0 & 5 & 5 & 5 \\ 1 & 0 & 0 & 0 & 3 & 0 & 5 & 5 & 5 \end{matrix}$$

Par conséquent, nous déterminons le nombre de composantes connexes. Ce nombre est égal au plus grand label que nous avons assigné lors de l'étiquetage. Par conséquent, $N_l = 5$.

Par la suite, pour chaque composante connexe, nous calculons le nombre de pixels qu'elle contient N_i . Puis, N_p correspond à la valeur maximale de N_i .

$$N_p = \max_{i=1 \dots N_l} N_i \quad (\text{III.1})$$

Par exemple, dans notre cas, pour la première composante connexe ($i = 1$), N_i (ou N_1) est égal à 5. De même, pour les deuxième, troisième, quatrième et cinquième composantes connexes, N_i est respectivement égal à 4, 5, 10 et 6. Par conséquent, $N_p = 10$.

Sachant que N_p est obtenu pour la 4^{ème} composante connexe, dès lors cette même composante est la plus grande décharge sur la surface isolante. De ce fait, L et W représente respectivement sa longueur et sa largeur en pixels. Finalement, L et W sont égales à 4 et 3 respectivement.

III.2.3. Classification des images

Afin de décider si une décharge représentée par une image donnée, fait partie des décharges de types arc, nous proposons l'utilisation des méthodes de classification basées sur différents processus d'apprentissage supervisé. Pour cela, nous décrivons, dans ce qui suit, les notions de bases des méthodes de classification et explicitons leurs principes.

III.2.3.1. Principe des méthodes de classification

La classification consiste à attribuer à un objet une classe le caractérisant : dans le monde industriel, les classes peuvent être "pièce avec défaut" versus "pièce sans défaut" par exemple.

Dans le cas des isolateurs, les classes sont généralement "échantillon pollué" versus "échantillon propre" ; mais nous pouvons aussi avoir des cas de multi-classes (supérieurs à 2 classes). L'objet à classer est représenté par des attributs qui ont chacun une valeur : l'ensemble de ces valeurs forme un vecteur dans l'espace de description. Ainsi, nous pouvons parler soit de l'objet (monde réel), soit de sa représentation mathématique x ($x \in \mathbb{R}^n$, n est le nombre d'attributs de description de l'objet, c'est la dimension de l'espace de description). Dans notre travail, les deux classes sont : décharge de type arc et non arc.

Il existe ainsi deux grandes familles de méthodes de classification [43].

- Classification supervisée : Dans ce cas, un expert doit avoir l'aptitude de définir la classe adéquate pour tout objet qui lui est présenté. Le classifieur doit alors être capable de prédire la classe d'appartenance ω_j d'un objet de vecteur de description x_i avec un minimum d'erreur. Pour cela, il estimera la probabilité d'appartenance à la classe ω_i du prototype x_i notée $P(\omega_i/x_i)$. Nous considérons une machine devant classer L objets d'un ensemble D . Chaque échantillon est donc représenté par un doublet $\{x_i, y_i\}$ où x_i est le vecteur représentant le $i^{\text{ème}}$ objet dans l'espace de description (typiquement \mathbb{R}^n) et y_i est l'étiquette attribuée à la classe d'appartenance de l'objet. Nous cherchons donc à définir la loi de probabilité $P(x, y)$ qu'un x donné appartienne à la classe d'étiquette y .
- Classification non supervisée : Dans ce cas, il n'existe pas d'expert capable de définir la classe de chaque objet. Les méthodes de classification ne peuvent pas se baser sur l'estimation de probabilités puisqu'elles ne sont ni définissables ni estimables dans la mesure où nous ne pouvons connaître l'appartenance d'un objet à une classe. Ces méthodes essayent donc de découper la représentation ponctuelle dans l'espace de description en parties homogènes dites clusters par le biais d'agrégations. La principale difficulté de ces méthodes réside dans la détermination du critère d'homogénéité permettant de former les clusters.

Que la classification soit supervisée ou non, les méthodes nécessitent un grand ensemble de données pour être statistiquement valables. Cet ensemble de données doit être divisé en deux parties [58].

- Base d'apprentissage : Cet ensemble d'objets x_i permettra de déterminer soit le modèle dans le cas d'une classification supervisée, soit les clusters pour une classification non supervisée. Il doit être le plus représentatif possible de l'ensemble de la population finale qui sera observée.
- Base de test : C'est un ensemble d'éléments à partir duquel sera menée la phase de validation. Sa construction doit être aussi rigoureuse que celle de l'ensemble d'apprentissage, car c'est à partir de l'ensemble de test que les performances de l'algorithme seront mesurées.

Dans notre cas, la base de donnée est composée d'images représentant les décharges à la surface de l'isolateur jusqu'à l'apparition de l'arc du contournement. Le nombre total des images est de 747 pour tous les niveaux de tension. 70 % de cette base est dédiée à l'apprentissage des modèles de classification. Alors que les 30 % restant représente la base de test, dédiée à la comparaison de 3 méthodes de classification (décrites dans les sections III.2.3.2 à III.2.3.4), comme présenté dans le tableau III.1, ci-dessous. A noter que, pour garantir un apprentissage

optimal du modèle, le nombre d'échantillons d'images contenant une décharge de type arc est égal à celui des images sans arcs.

Tableau III.1. Base de données des images collectées

Base	Nombre d'images
Apprentissage	521
Test	226
Totale	747

III.2.3.2. Machines à vecteurs de support (SVM) [58-79]

Les SVM (Support Vector Machines) sont de nouvelles techniques d'apprentissage statistique initiées par V. Vapnik en 1995 [59]. Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent. Elles permettent d'aborder des problèmes très divers, dont la classification. SVM est une méthode particulièrement bien adaptée pour traiter des données de très haute dimension telles que les textes et les images. Depuis leur introduction dans le domaine de la reconnaissance de formes, plusieurs travaux ont pu montrer l'efficacité de ces techniques, principalement dans les domaines du traitement d'image et de la parole [64-79].

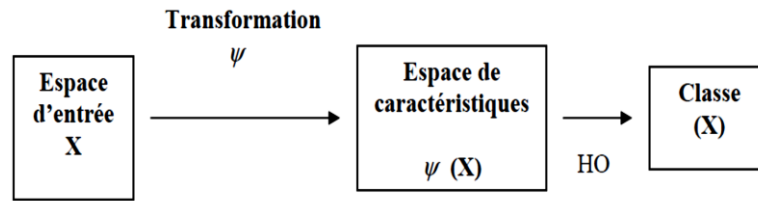
Cette technique est initialement une méthode de classification à deux classes qui tente de séparer linéairement les exemples positifs de ceux négatifs dans l'ensemble des exemples. Chaque exemple, ou échantillon, doit être représenté par un vecteur de dimension n . La méthode cherche alors l'hyperplan qui sépare les exemples positifs de ceux négatifs, en garantissant que la marge entre le plus proche des positifs et celui des négatifs soit maximale. Intuitivement, cela garantit un bon niveau de généralisation car de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan mais être tout de même situés franchement d'un côté ou l'autre de la frontière. L'efficacité des SVM est très bonne pour la reconnaissance de formes et l'intelligence artificielle. Un autre intérêt est la sélection de vecteurs supports qui représentent les vecteurs discriminant grâce auxquels est déterminé l'hyperplan. Les exemples utilisés lors de la recherche de l'hyperplan ne sont alors plus utiles et seuls ces vecteurs supports sont utilisés pour classer un nouveau cas. Cela en fait une méthode très rapide.

III.2.3.2.1. Principe de la technique SVM [58-79]

Le principe des SVM consiste à projeter les données de l'espace d'entrée (appartenant à deux classes différentes) non linéairement séparables dans un espace de plus grande dimension appelé espace de caractéristiques de façon à ce que les données deviennent linéairement séparables. Dans cet espace, nous construisons un hyperplan optimal séparant les classes tel que:

- les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan ;
- la plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale.

Le principe de la technique SVM est représenté sur la figure III.3 ci-dessous.



HO : Hyperplan Optimal.

Figure III.3. Principe de la technique SVM

III.2.3.2.2. SVM linéaires [58]

- **Cas des données linéairement séparables**

On considère une base d'apprentissage B de n objets à deux classes linéairement séparables. Chaque objet est alors décrit par le couple $\{x_i, y_i\}$, $x_i \in \mathbb{R}, y_i \in \{-1, 1\}$ où x_i est le vecteur décrivant chaque objet dans l'espace de description de dimension d et y_i l'étiquette de la classe de cet objet (figure III.4).

Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en x [58]. On peut exprimer une telle fonction par $w \in \mathbb{R}$ et $b \in \mathbb{R}$ tel que :

$$h(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i + b \quad (\text{III.2})$$

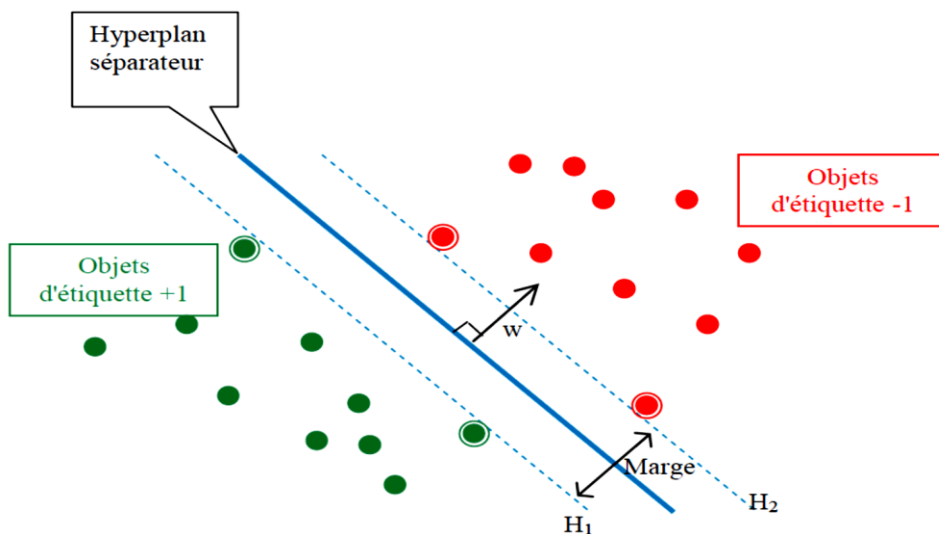


Figure III.4. Principe des SVM : trouver l'hyperplan séparant les groupes d'éléments en maximisant la marge

On suppose qu'il existe un hyperplan séparant les points de chaque classe, l'équation d'un tel hyperplan s'écrit alors :

$$\langle w, x \rangle + b = 0 \quad (\text{III.3})$$

où w est une normale à l'hyperplan et b une constante.

Il existe une infinité d'hyperplans respectant cette équation, il faut trouver le meilleur. La notion de meilleur hyperplan est introduite en considérant deux hyperplans parallèles (H_1 et H_2) passant chacun par les points les plus proches de la frontière des classes qu'ils limitent et n'ont aucun point de la base d'apprentissage B entre eux (figure III.5). Leurs équations respectives sont :

$$H_1: \langle w, x \rangle + b = 1 \text{ et } H_2: \langle w, x \rangle + b = -1 \quad (\text{III.4})$$

Pour garantir qu'il n'y ait aucun point de B entre eux, tout point de B doit satisfaire l'inégalité suivante :

$$\forall i \in [1..n], y_i \cdot (\langle w, x_i \rangle + b) \geq 1 \quad (\text{III.5})$$

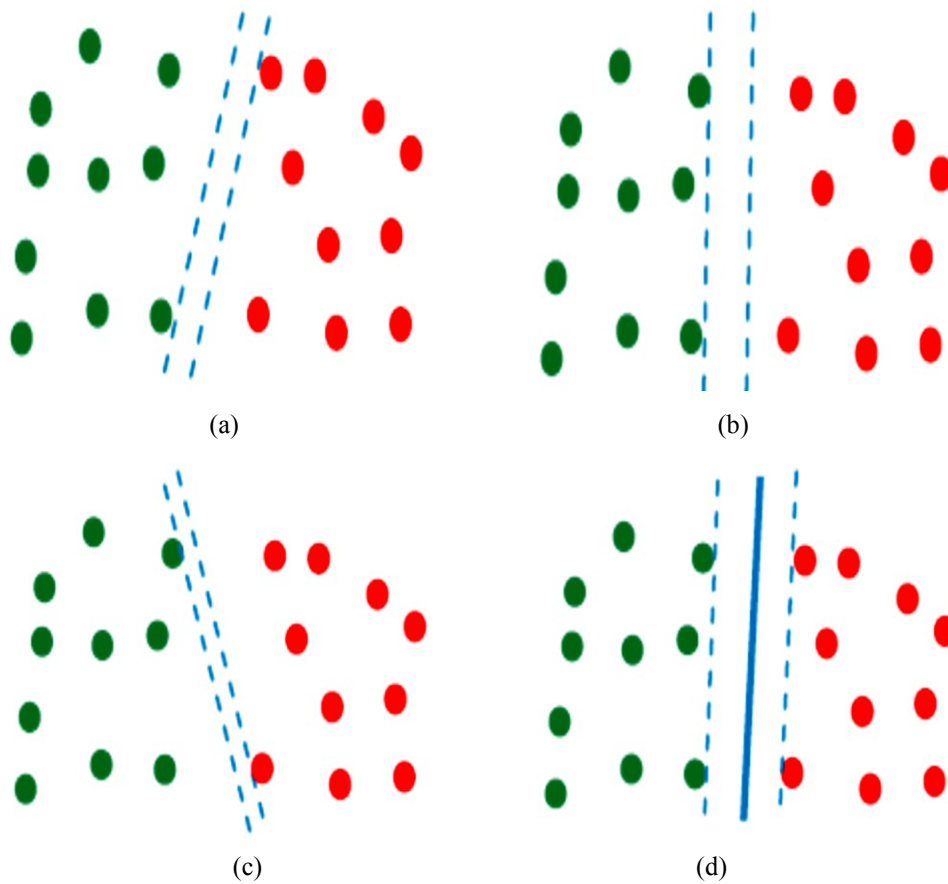


Figure III.5. Exemple d'hyperplans séparateurs
(a), (b), (c) sont des hyperplans quelconques et (d) est un hyperplan à marge maximale

La distance entre deux hyperplans (la marge) est $\frac{2}{\|w\|}$, donc pour maximiser sa valeur, il faut minimiser $\|w\|$, où $\| \cdot \|$ est défini par la distance euclidienne. Ceci est un problème d'optimisation quadratique. Le meilleur hyperplan séparateur est alors celui qui est à égale distance de H_1 et H_2 . Pour résoudre ce problème, nous utilisons la théorie de Lagrange pour reformuler le problème de minimisation sans la contrainte d'inégalité. On obtient alors la fonction de Lagrange définie par :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (\langle w, x_i \rangle + b) + \sum_{i=1}^n \alpha_i, \alpha_i \geq 0 \quad (\text{III.6})$$

où les α_i sont les coefficients de Lagrange. Cette fonction est alors minimisée par rapport à w et b .

En prenant la dérivée partielle de $L(w, b, \alpha)$ par rapport à w et b , nous obtenant :

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i \quad (\text{III.7})$$

et

$$\frac{\partial L(w, b, \alpha)}{\partial b} = - \sum_{i=1}^n \alpha_i y_i \quad (\text{III.8})$$

En annulant les deux dérivées partielles, on obtient le système suivant :

$$\sum_{i=1}^n \alpha_i y_i x_i = w \quad (\text{III.9})$$

et

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (\text{III.10})$$

On peut obtenir un problème dual en prenant en considération les éléments de la base d'apprentissage B par couple x_i et y_i . Ce problème s'écrit alors :

$$L(w, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (\text{III.11})$$

$L(w, b, \alpha)$ est alors maximisé selon les α_i en respectant les contraintes $\sum_{i=1}^n \alpha_i y_i = 0$ et $\alpha_i \geq 0$.

La décision de classification est alors obtenue pour tout x n'appartenant pas à l'ensemble d'apprentissage par :

$$y = \text{sign}(\langle w, x \rangle + b) \quad (\text{III.12})$$

En remplaçant par w sa valeur de l'équation III.47, nous obtenons la fonction de décision suivante :

$$f(x) = \sum_{i=1}^n \alpha_i y_i \langle x_i, x \rangle + b \quad (\text{III.13})$$

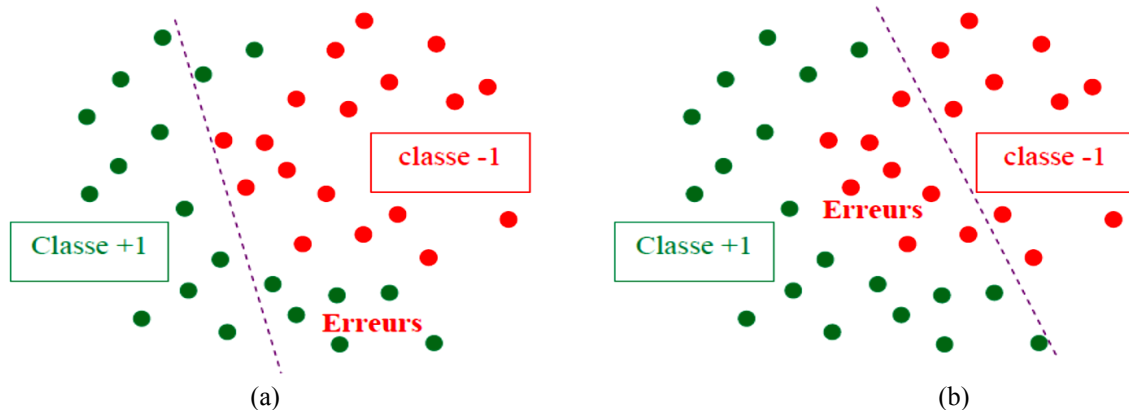
Chaque coefficient multiplicateur de Lagrange α_i correspond à un élément x_i de la base d'apprentissage B. Les x_i qui ont contribué à la maximisation de la marge sont associés à des α_i non nuls et sont donc appelés vecteurs supports. Les autres vecteurs de la base d'apprentissage ne participent donc pas à la décision finale et l'équation III.13 devient :

$$f(x) = \sum_{i \in \text{SV}} \alpha_i y_i \langle x_i, x \rangle + b \quad (\text{III.14})$$

où SV est l'ensemble des indices des vecteurs de support (vecteurs pour lesquels les coefficients de Lagrange sont non nuls).

- **Cas des données non linéairement séparables**

Le cas non linéairement séparable se présente lorsqu'il n'existe pas d'hyperplan d'équation linéaire séparant les deux classes sans commettre d'erreur (figure III.6).



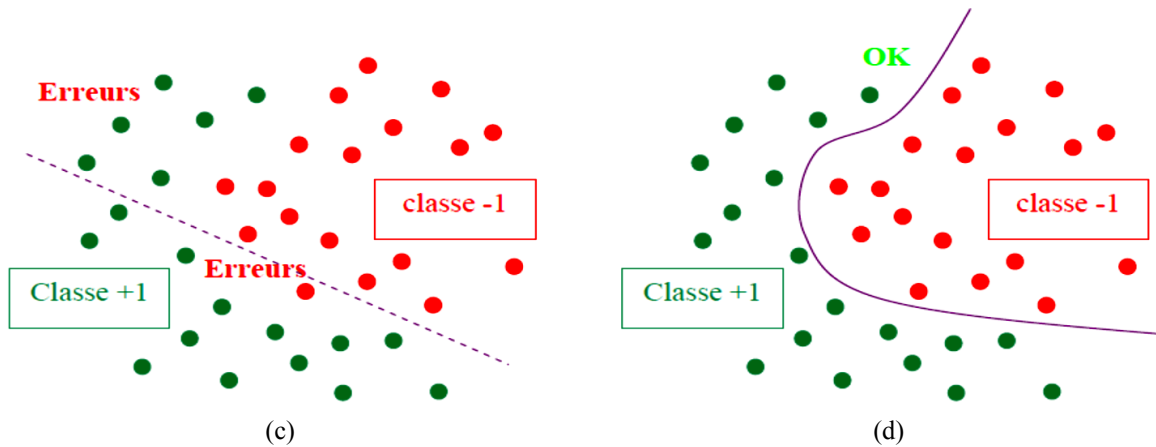


Figure III.6. Deux classes non linéairement séparables ;
 il n'existe pas d'hyperplan séparant les deux classes sans erreur (a, b, c), seul un hyperplan d'équation non linéaire le permet (d)

Pour apporter une solution à ce problème, nous introduisons la notion de marge souple (soft margin), qui autorise un certain nombre d'éléments d'une classe à se trouver du mauvais côté de la frontière lors de l'apprentissage (figure III.7). Ces éléments sont appelés des « outiliers ».

On applique donc une marge d'erreur ξ_i à l'équation (III.2) pour obtenir :

$$\forall i \in [1..n], y_i \cdot (\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad (\text{III.15})$$

Pour maximiser la marge, il ne suffit plus de minimiser $\|w\|$ mais la fonction $\|w\| + C \sum_{i=1}^n \xi_i$, où C est le coût attribué à un échantillon mal classé.

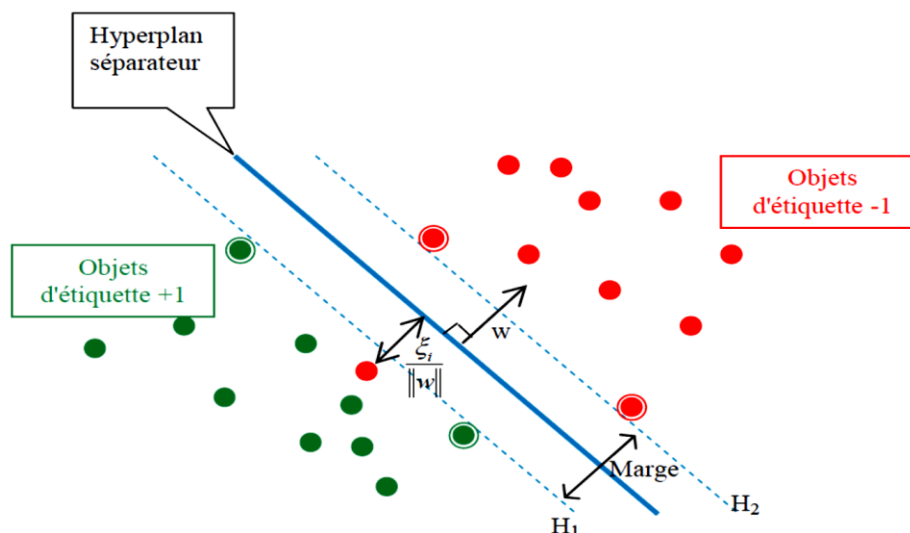


Figure III.7. Principe de "l'outlier" ;
 certains éléments sont "autorisés" à se retrouver du mauvais côté de la frontière lors de la phase d'apprentissage

La frontière définie est là aussi linéaire mais décalée en fonction du coût fixé. L'erreur due à ce type de séparation est souvent plus importante que l'erreur générée par une frontière non linéaire, c'est pourquoi une modélisation non linéaire a été intégrée, toujours en vue de minimiser le risque empirique. Cette méthode dite « astuce du noyau » (kernel trick) a été introduite par Aizerman en 1964 [60].

Les équations de base font intervenir les données d'apprentissage (les échantillons et leur classe) sous la forme d'un produit scalaire. Il a été démontré qu'il est possible que les données initiales (de l'espace \mathbb{R}^d) soient linéairement séparables dans un espace euclidien $H = \mathbb{R}^e$ (avec $e > d$ fini ou non), dans lequel nous projetons l'espace de description \mathbb{R}^d à l'aide d'une fonction ϕ telle que :

$$\phi: \mathbb{R}^d \rightarrow \mathbb{R}^e, \phi(x) = \tilde{x} \quad (\text{III.16})$$

où \tilde{x} est un exemple estimé.

En effet, plus la dimension de l'espace de description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les classes est élevée. En transformant l'espace d'entrée en un espace de redescription H de très grande dimension, éventuellement infinie, il devient donc possible d'envisager d'utiliser la méthode de résolution quadratique précédente dans le cas de classes non linéairement séparables.

Si ces espaces sont munis d'un produit scalaire, nous pouvons définir sur \mathbb{R}^d une fonction noyau telle que :

$$K: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^e \times \mathbb{R}^e, K(x, y) = \langle \phi(x), \phi(y) \rangle = \langle \tilde{x}, \tilde{y} \rangle \quad (\text{III.17})$$

L'algorithme des SVM est alors le même que dans le cas linéaire : il suffit de remplacer tous les produits scalaires par une fonction noyau non linéaire. Avec cette astuce, les SVM peuvent définir le meilleur hyperplan séparateur linéaire dans le nouvel espace \mathbb{R}^e en utilisant la fonction noyau appropriée.

En appliquant la technique à noyau à la fonction de Lagrange, on obtient la fonction suivante :

$$L(w, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (\text{III.18})$$

La fonction de décision devient donc :

$$y = \text{sign}(\sum_{i \in \text{SV}} \alpha_i y_i K(x_i, x_j) + b) \quad (\text{III.19})$$

où SV est l'ensemble des indices des vecteurs support issus de la phase d'apprentissage.

Toute fonction satisfaisant au critère de Mercer peut être utilisée comme noyau. Ce critère est le suivant : soit k une fonction symétrique, on définit alors la matrice $(k(x_i, x_j))_{i,j}$. Si cette matrice est définie positive, c'est-à-dire que pour chaque vecteur colonne χ de cette matrice $\chi^T M \chi > 0$, alors il existe un espace F et une fonction ϕ tels que $k(x, y) = \langle \phi(x), \phi(y) \rangle$.

On peut ainsi définir notre propre noyau, cependant il existe plusieurs noyaux prédéfinis dont les plus employés sont :

- le noyau polynomial (d'ordre p) : $K(x, y) = (a \langle x, y \rangle)^p$;
- le noyau RBF : $K(x, y) = \exp(-\gamma \|x - y\|^2)$, avec $\gamma > 0$;
- le noyau RBF gaussien $K(x, y) = \exp\left(-\frac{\|x - y\|}{2\sigma^2}\right)$;

- le noyau tangente hyperbolique : $K(x, y) = \tanh(k(x, y) - \theta)$.

Malgré l'apparente complexité de la méthode, sa mise en œuvre est relativement simple. La prédiction d'appartenance à une classe d'un élément s'effectue à partir d'un calcul d'une somme de produits dans laquelle intervient la fonction noyau. La complexité de la fonction de décision est donc essentiellement liée au nombre de vecteurs supports et la complexité de K .

III.2.3.2.3. SVM pour le cas multi-classes [61-79]

La plupart des problèmes ne se contentent pas de deux classes de données. Nous avons choisi de considérer les problèmes multi-classes comme un cumul de problèmes à deux classes (voir section IV.8). Il existe plusieurs stratégies pour faire la classification multi-classes.

- La stratégie « un contre tous »

La solution la plus simple pour résoudre un problème multi-classes à l'aide de SVM consiste à le décomposer en un ensemble de sous-problèmes binaires et à construire indépendamment un SVM pour chacun d'entre eux. Ainsi, Cortes et Vapnik [62] proposent d'utiliser une stratégie de décomposition très intuitive et facile à mettre en place. Cette stratégie, communément nommée « un contre tous » ou en anglais « One Against All, OAA », consiste à construire autant de SVM qu'il y a de classes. Chaque SVM est alors entraîné à séparer les données d'une classe qui seront étiquetées +1, de celles de toutes les autres classes qui seront étiquetées -1. La règle de décision finale est l'application du principe « winner take all ». Pour chaque classifieur un score va être établi et l'étiquette attribuée à l'entrée x_i est celle associée au classifieur qui obtient le meilleur score. Cependant, étant donné que les sorties des SVM ne sont pas calibrées (pour un classifieur ω_i nous pouvons avoir un très petit nombre d'exemples de la classe i et un grand nombre de contre exemples), la décision peut parfois être faussée. À partir de cette constatation, Mayoraz et Alpaydin [63] proposent soit de calibrer indépendamment les sorties de chaque SVM, soit d'utiliser un perceptron qui recevra les sorties brutes des différents SVM et retournera des valeurs calibrées. Dans le même esprit, il peut être intéressant d'utiliser les sorties des différents SVM pour estimer les probabilités a posteriori d'appartenance aux différentes classes et d'utiliser ensuite ces valeurs pour prendre la décision.

- La stratégie « un contre un »

Une autre stratégie classique consiste à construire un SVM pour chaque paire de classes, soit $c(c-1)/2$ SVM pour un problème à c classes. Chaque classifieur est donc entraîné à séparer les données d'une classe de celles d'une autre classe. Nous parlerons alors de stratégie « un contre un » ou en anglais « One Against One, (OAO) ». Notons par ailleurs, que cette stratégie peut aussi être rencontrée dans la littérature sous les noms de « pairwise coupling », « all pairs » ou encore « round robin » [64-65]. La règle de décision généralement utilisée est le vote majoritaire. Chaque SVM vote alors pour une classe et la donnée inconnue est finalement associée à la classe ayant reçu le plus grand nombre de votes. Cependant, il est possible que plusieurs classes reçoivent le même nombre de votes. Alors, partant du principe que les cas d'égalité sont très peu fréquents, Hsu et Lin [66] proposent de prendre la décision de manière aléatoire. Par ailleurs, une approche originale est proposée par Platt et al [67] qui proposent d'utiliser un arbre de décision dont chaque nœud est un SVM et chaque feuille une des classes du problème.

III.2.3.2.4. Avantages et désavantages des SVM [53-56]

L'avantage de la création d'une fonction de décision avec l'algorithme des SVM est que la solution produite correspond à l'optimum d'une fonction convexe. Elle ne possède donc pas plusieurs optima locaux comme pour les réseaux de neurones (dans leur formulation classique), mais un optimum global. Cet optimum correspond à la minimisation du risque structurel et donc à la recherche d'une hypothèse possédant de bonnes capacités de généralisation à partir d'un espace d'hypothèses donné. L'espace d'hypothèses dépend du choix de la fonction noyau, ce qui est à la fois un avantage et un inconvénient. Un avantage car l'algorithme d'apprentissage reste identique. L'inconvénient est que la fonction noyau a souvent des paramètres libres, comme la largeur de bande avec un noyau gaussien, et que la recherche de la valeur optimale de ces paramètres libres ne correspond plus à la recherche du minimum d'une fonction convexe.

La recherche des valeurs idéales des paramètres optimaux de la fonction noyau, appelée communément sélection de modèles, est critique [69-79]. La sélection d'une mauvaise fonction noyau (ou de ses paramètres) peut conduire à produire un effet de sur-apprentissage et c'est le principal désavantage des SVM.

Un autre désavantage des SVM est la durée importante de la phase d'entraînement, en particulier si une sélection de modèles doit être réalisée, bien que dans les dix dernières années plusieurs optimisations de l'algorithme initial aient été proposées.

Un dernier désavantage est la complexité de la fonction de décision produite lorsque la base d'apprentissage est de taille importante.

III.2.3.3. Méthodologie de l'apprentissage bayésien [80-85]

La méthode bayésienne est un ensemble de techniques statistiques utilisées pour modéliser des problèmes, extraire de l'information de données brutes et prendre des décisions de façon cohérente et rationnelle. Son cadre d'application est général, mais ses avantages sont déterminants lorsque l'information disponible est incertaine ou incomplète.

Bien que les premiers travaux d'inspiration bayésienne datent du XVII^{ème} siècle, cette méthode connaît un regain de popularité depuis quelques décennies. Ce renouveau est sensible dans des domaines très variés, en partie grâce à la disponibilité de calculateurs puissants, mais aussi à cause d'une évolution de la pensée statistique et des problèmes abordés.

La méthode bayésienne se distingue des autres méthodes statistiques par la sémantique qu'elle donne à la notion de probabilité. Une méthode statistique est d'inspiration bayésienne si les probabilités qu'elle manipule sont subjectives. A l'opposé, les outils des statistiques classiques, comme les estimateurs sans biais et les tests d'hypothèses, considèrent une interprétation objective des probabilités [65].

III.2.3.3.1. Description du modèle bayésien [81]

De manière abstraite, le modèle probabiliste pour un classificateur bayésien est un modèle conditionnel. Il se base sur la règle de Bayes qui s'énonce de la manière suivante :

$$P(A|B_1, B_2, \dots, B_n) = \frac{P(B_1, B_2, \dots, B_n|A) \times P(A)}{P(B_1, B_2, \dots, B_n)} \quad (\text{III.20})$$

La probabilité d'avoir l'évènement A étant donné B_1, \dots, B_n est donnée par le rapport entre la probabilité d'avoir les évènements B_1, \dots, B_n étant donné A et la probabilité que B_1, \dots, B_n se soient produits. Tant que le dénominateur ne dépend pas de l'évènement A, on peut considérer la probabilité $P(B_1, \dots, B_n)$ comme étant constante. Le numérateur peut être écrit encore de la manière suivante :

$$\begin{aligned} P(B_1, B_2, \dots, B_n|A) \times P(A) &= P(A, B_1, B_2, \dots, B_n) \\ &= P(A) \times P(B_1|A) \times P(B_2, \dots, B_n|A, B_1) \\ &= P(A) \times P(B_1|A) \times P(B_2|A, B_1) \times P(B_3, \dots, B_n|A, B_1, B_2). \end{aligned} \quad (\text{III.21})$$

La décomposition de $P(A, B_1, B_2, \dots, B_n)$ se termine lorsqu'on a parcouru l'ensemble des classes B_1, \dots, B_n . Le caractère « naïf » de ce théorème vient du fait qu'on suppose l'indépendance des différentes classes B_i, \dots, B_j , ce qui en d'autres termes se traduit par :

$$P(B_i|A, B_j) = P(B_i|A) \quad (\text{III.22})$$

Cette hypothèse permet également d'écrire :

$$P(A|B_1, \dots, B_n) = P(A) \times P(B_1|A) \times \dots \times P(B_n|A) = P(A) \times \prod_{i=1}^n P(B_i|A) \quad (\text{III.23})$$

III.2.3.3.2. Estimation des paramètres [82]

A l'aide d'un ensemble d'apprentissage, nous allons estimer les paramètres du modèle présenté en (III.20). On utilise classiquement la méthode de maximisation de la log-vraisemblance sur un corpus d'entraînement afin de calculer ces paramètres.

Soit (x_1, \dots, x_n) un ensemble de n exemples de composantes $x^j = (x_1^j, \dots, x_{|x^j|}^j)$ où $|x^j|$ représente la longueur de la séquence. La vraisemblance de θ s'écrit :

$$l_\theta = \prod_{j=1}^n P(x^j/\theta). \quad (\text{III.24})$$

Nous allons donc calculer l'ensemble des paramètres qui maximisent cette vraisemblance:

$$\theta_{\text{finaux}} = \text{argmax } l_\theta. \quad (\text{III.25})$$

Pour cela, nous allons plutôt maximiser la log-vraisemblance $L_\theta = \log(l_\theta)$ qui est une fonction convexe. Cette maximisation s'obtient donc en annulant la dérivée :

$$\forall v \in V, \frac{\partial L_\theta}{\partial \theta_v} = 0 \text{ sous la contrainte } \sum_{v \in V} \theta_v = 1. \quad (\text{III.26})$$

V est un ensemble discret quelconque $V = (v_1, \dots, v_n)$ appelé vocabulaire.

En utilisant la méthode des multiplicateurs de Lagrange, on obtient le système suivant :

$$\forall v \in V, \frac{\partial (L_\theta + \lambda(1 - \sum_{v' \in V} \theta_{v'}))}{\partial \theta_v} = 0 \Leftrightarrow \forall v \in V, \frac{\partial (\sum_{j=1}^n \sum_{i=1}^{|x^j|} \log P(x_i^j/\theta) + \lambda(1 - \sum_{v' \in V} \theta_{v'}))}{\partial \theta_v} = 0$$

$$\Leftrightarrow \forall v \in V, \frac{\partial \left(\sum_{j=1}^n \sum_{i=1}^n \log P \left(\theta_{x_i^j} \right) + \lambda (1 - \sum_{v' \in V} \theta_{v'}) \right)}{\partial \theta_v} = 0, \quad (\text{III.27})$$

où $\theta_{x_i^j}$ est le paramètre qui correspond à l'estimation de $P(x_i^j/\theta)$.

Ce système admet une solution analytique unique. Soit N_v^x le nombre de fois que l'élément v apparaît dans l'exemple x, on obtient :

$$\forall v \in V, \theta_v = \frac{\sum_{j=1}^n N_v^{x_j}}{\sum_{v' \in V} \sum_{j=1}^n N_{v'}^{x_j}}. \quad (\text{III.28})$$

XI.3.3.3. Avantages et inconvénients [85]

Le point fort de la classification bayésienne est sa simplicité d'emploi car elle fournit une méthode élégante pour combiner nos connaissances a priori avec les informations provenant de ces données.

Cependant, la force de la méthode de classification bayésienne est aussi sa faiblesse en même temps. En effet, le classifieur bayésien étant naïf, il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. Ceci rend la méthode bayésienne parfois très faible si elle est utilisée dans le cas de données vastes ou complexes.

De plus, l'hypothèse d'indépendance des variables permet de se contenter de la variance de chacune d'entre elle pour chaque classe, sans avoir à calculer de matrice de covariance. Or, en pratique, la classification bayésienne donne parfois de mauvais résultats si les données utilisées sont dépendantes entre elles, ou fortement corrélées.

III.2.3.4. Les k plus proches voisins [86-89]

III.2.3.4.1. Principe général

C'est une des méthodes de classification les plus simples, qui donne généralement de bons résultats. Après plus de 35 ans, les classificateurs utilisant la règle des « Plus Proches Voisins (PPV) » ou en anglais « Nearest Neighbors (NN) » sont encore largement étudiés et utilisés pour résoudre des problèmes de reconnaissance des formes. L'idée de base est d'inférer la classe d'un nouvel exemple en choisissant celle de l'exemple qui lui est le plus proche dans la base d'apprentissage [86]. Cela consiste à choisir pour un exemple x la classe $y = \omega_j$ de son plus proche voisin j :

$$D_{\text{PPV}}(x) = \omega_j | j = \operatorname{argmin}_{1 \leq i \leq m} d(x, x_i) \quad (\text{III.29})$$

avec $d(x, x_i)$ une métrique permettant de mesurer la distance entre deux exemples.

Pour éviter d'être trop sensible à des données bruitées, on réalise souvent la recherche des k plus proches voisins. Notons $N_{\text{voisin}}(x, k, \omega_i, Z_m)$ la fonction retournant le nombre d'exemples de la classe ω_i parmi les k plus proches voisins de x. La classe sélectionnée est celle qui est majoritaire parmi ces k voisins, ce qui correspond à la fonction de décision suivante :

$$D_{k\text{-ppv}}(x) = \operatorname{argmax}_{\omega_i \in Y} N_{\text{voisin}}(x, k, \omega_i, Z_m) \quad (\text{III.30})$$

Par exemple, dans un problème de classification, on retiendra la classe la plus représentée parmi les k sorties associées aux k entrées les plus proches de la nouvelle entrée x .

La figure III.8 ci-dessous montre bien le principe ainsi que la problématique majeure de la k -ppv. Ainsi, en bleu et rouge sont tracées les classes 1 et 2. L'objet circulaire en couleur verte, symbolise un échantillon de test à classifier. Il peut être classé soit dans la classe 1, soit la classe 2. Si $k = 3$, il sera assigné à la classe 2, car il existe 2 triangles rouges, et seulement un carré bleu à l'intérieur du cercle. Par contre, si $k = 5$, l'objet de test sera assigné à la classe 1 (3 carrés bleus et 2 triangles rouges à l'intérieur du cercle). Le bon choix de k s'avère donc être capital lors de la classification.

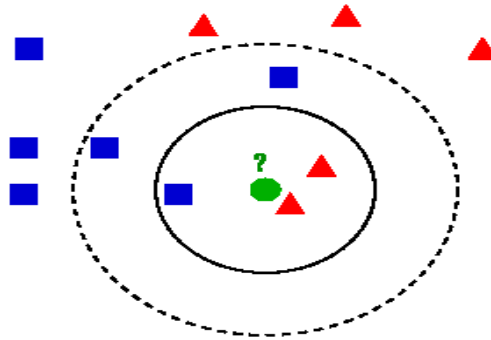


Figure III.8. Exemple de la k -ppv

III.2.3.4.2. Fondements théoriques

Bien qu'élégante et simple, cette méthode a des fondements théoriques importants liés à une approche probabiliste. Supposons k fixé. Soit un ensemble d'exemples Z_m issu d'une distribution D_Z correspondant à un ensemble de points dans un espace \mathbb{R}^n qui ont chacun une classe associée. Pour un exemple x , la densité de probabilité $P(x)$ dont il est issu (c'est-à-dire D_x) peut être approximée localement comme étant égale à :

$$P(x) \approx \frac{k}{m \cdot V_R}, \quad (\text{III.31})$$

avec V_R le volume de la sphère centrée sur x contenant les k plus proches exemples de x . Notons k_i le nombre d'exemples de classe ω_i parmi les k plus proches voisins et m_i le nombre d'exemples de classe ω_i dans Z_m . La densité de probabilité $P(x|\omega_i)$ conditionnellement à la classe ω_i peut être approximée suivant le même principe :

$$P(x|\omega_i) \approx \frac{k_i}{m_i \cdot V_R}. \quad (\text{III.32})$$

III.2.3.4.3. Algorithme de la k -ppv

Soit A un ensemble d'apprentissage défini comme suit :

$$A = \{(x, c) \text{ tel que } x \in \mathbb{R}^n, c \in C\}, \quad (\text{III.33})$$

avec x l'ensemble des échantillons d'apprentissage, c la classe qui lui est assignée, et C l'ensemble contenant toutes les classes de A .

Soit, x' l'exemple dont on souhaite déterminer la classe. L'algorithme de la k-ppv peut s'établir comme suit :

```

début
|
|   pour chaque (exemple(x, c) ∈ A) faire
|   |   calculer la distance d(x', x)
|   fin
|   pour chaque { x ∈ kppv(x') } faire
|   |   compter le nombre d'occurrences de chaque classe
|   fin
|   Attribuer à x' la classe la plus fréquente
fin

```

Habituellement, la distance euclidienne est choisie lors de la k-ppv, mais cela reste applicable seulement dans le cas de variables continues. Autrement, la distance dite de « Hamming » sera à privilégier et est égale à $d(x', x) = \sqrt{(x' - x)^2}$.

III.2.3.4.4. Amélioration des capacités de généralisation [88]

Pour mieux appréhender les problèmes de la règle du plus proche voisin, abordons-le comme un problème de recherche d'une hypothèse à partir d'un principe d'inférence de minimisation de l'erreur empirique. La sélection d'une hypothèse est implicite pour la construction du classificateur PPV. Il n'y a pas de phase d'apprentissage à proprement parler lorsque k est fixé. On peut considérer que l'espace des hypothèses est un ensemble de m points dans un espace \mathbb{R}^n . Lorsque l'ensemble Z_m est présenté, l'hypothèse sélectionnée est tout simplement Z_m . Si on suppose en plus que k est fixé à 1, l'erreur empirique correspondant à la règle du plus proche voisin est forcément nulle pour cette hypothèse (sauf si des exemples peuvent avoir la même description mais avec des classes différentes). Cela correspond donc à utiliser un principe inductif de minimisation de l'erreur empirique pour la sélection de l'hypothèse optimale. La possibilité de sélectionner systématiquement une hypothèse d'erreur empirique nulle correspond à utiliser un espace implicite d'hypothèse vaste. La variance est donc grande et la fonction de décision produite est fortement sensible aux variations de Z_m . Il n'y a aucune généralisation à proprement parler avec l'algorithme du plus proche voisin. La fonction de décision produite est donc coûteuse en place mémoire et en temps de décision. Plusieurs méthodes ont alors été proposées pour réduire la variance de la règle du PPV, ainsi que le coût de la fonction de décision.

La méthode la plus simple consiste à utiliser une valeur de k supérieure à un avec la règle k-ppv pour atténuer une variance élevée. K agit comme un mécanisme de régularisation en augmentant le biais. Le choix d'une valeur appropriée de k est essentiel et cette valeur est généralement déterminée par validation croisée.

III.2.3.4.5. Avantages et inconvénients [89]

Voici les principaux avantages de la méthode k-ppv :

- Sa qualité s'améliore en introduisant de nouvelles données sans nécessiter la reconstruction d'un modèle ;
- La classe attribuée à un objet peut être expliquée en exhibant les plus proches voisins qui ont amené à ce choix, ce qui apporte une bonne clarté des résultats ;
- La méthode peut s'appliquer à tout type de données, même les données complexes telles que des informations géographiques, des textes, des images, du son, elle est ainsi robuste au bruit ;
- Elle est facile à mettre en œuvre.

Pour les inconvénients, notons les points suivants :

- La méthode ne nécessite pas d'apprentissage, ce qui implique que tous les calculs sont effectués lors de la classification. Contrairement aux autres méthodes (telle que la SVM) qui nécessitent un apprentissage (éventuellement long) mais qui sont plus rapides en classification.
- Elle donnera de mauvais résultats si le nombre d'attributs pertinents est faible relativement au nombre total d'attributs, car la proximité sur les attributs pertinents sera noyée par les distances sur les attributs non pertinents.
- Les performances de la méthode dépendent du choix de la distance, du nombre de voisins et du mode de combinaison des réponses des voisins.

III.3. Résultats de l'algorithme

Comme il a été précédemment décrit dans la section 5, l'algorithme développé dans ce chapitre peut se subdiviser en quatre étapes distinctes. Ainsi, pour présenter les résultats de cet algorithme, la présente section décrira d'abord les résultats obtenus dans les étapes de segmentation et de filtrage. Puis, l'étape d'extraction des caractéristiques sera explicitée, en présentant les valeurs du vecteur de caractéristique obtenues pour la totalité des images enregistrées. Enfin, les résultats des méthodes de classifications seront exposés et comparés.

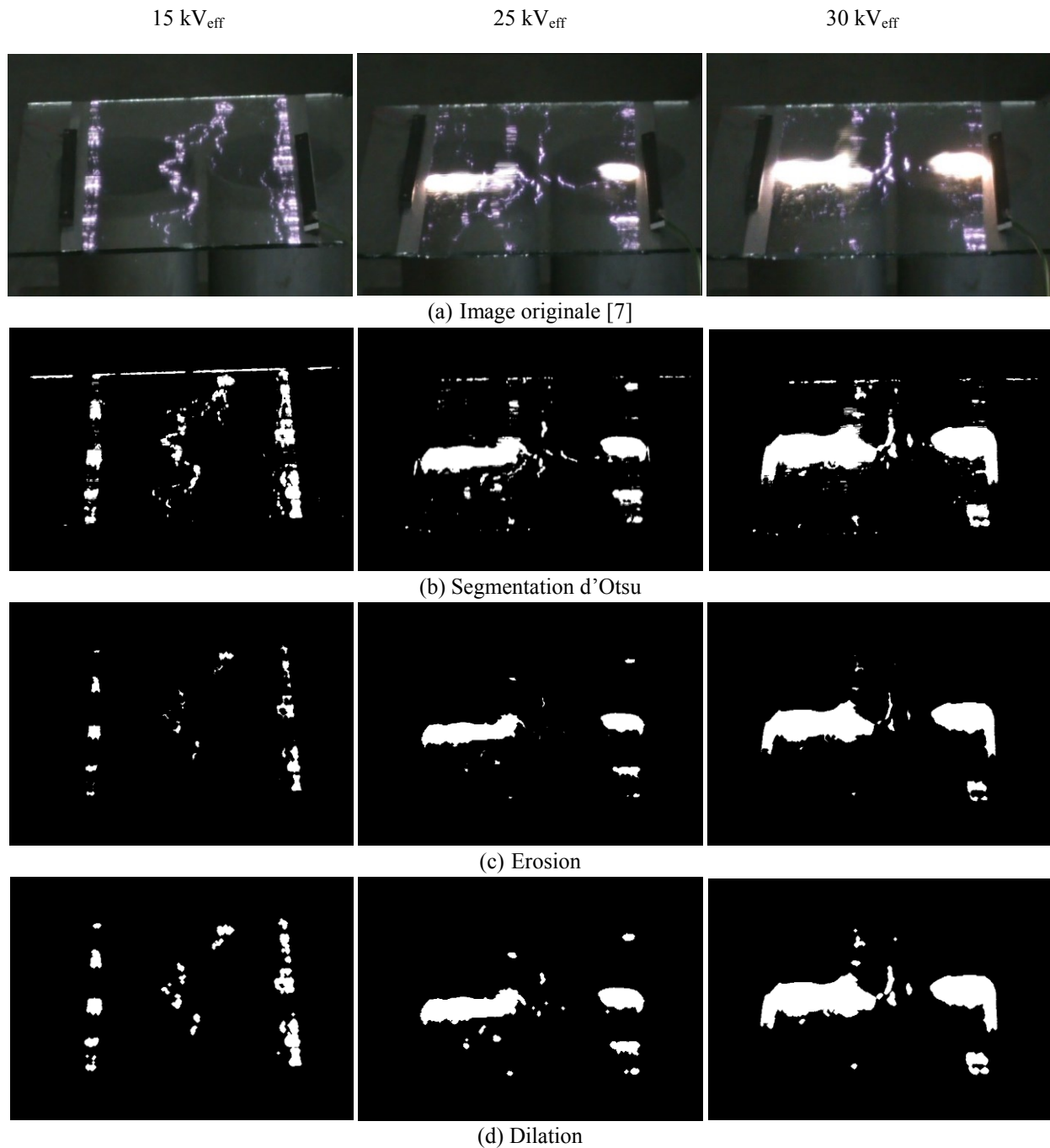
III.3.1. Résultats de la segmentation et filtrage

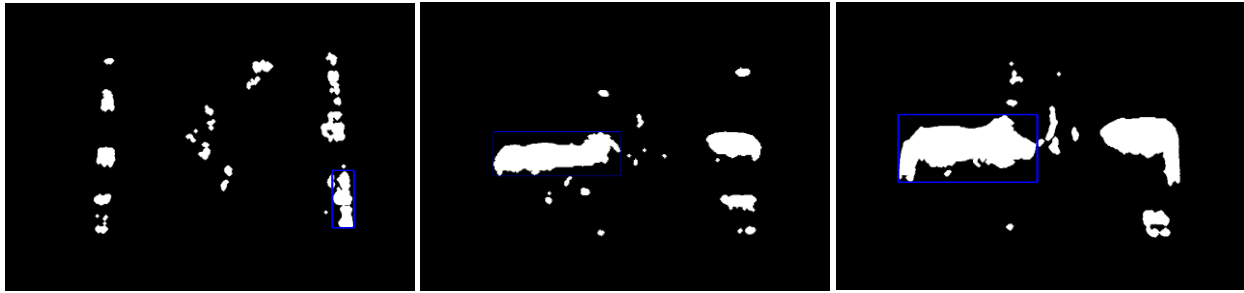
En faisant varier la tension de 0 kV_{eff} jusqu'à la tension de contournement, suivant un pas de 1 à 2 kV_{eff} par seconde, les images sont enregistrées (figure III.9). Nous proposons d'analyser les résultats de l'algorithme proposé pour trois niveaux de tension.

Comme montré sur la figure III.9, la surface de l'isolateur pour un niveau de tension de 15 kV_{eff} ne contient pas d'arc, à l'inverse des niveaux de tension 25 et 30 kV_{eff}, où des arcs partiels bien visibles y sont montrés sur la figure III.9.a [7]. L'acquisition des images étant faite, l'algorithme développé procède alors à la première étape. Par conséquent, le seuil optimal d'Otsu est calculé. L'image est alors segmentée sur la figure III.9.b. Par la suite, l'étape de filtrage morphologique prend place. Les pixels blancs reflétés par la lumière du jour et celle de l'éclairage du laboratoire sur le modèle de l'isolateur étant bien visibles, nous procédons à l'opération d'érosion qui a pour but de les éliminer, comme montré sur la figure III.9.c. Cependant, cette opération d'érosion n'est pas sans perte d'informations. En effet, l'érosion ronge les bords des décharges électriques causant leur rétrécissement. Par la suite, nous utilisons la dilatation, qui a

pour but de compenser cette perte d'information, en rétablissant les pixels supprimés par l'érosion, comme montré sur la figure III.9.d.

Sur la base de l'image dilatée, présentée sur la figure III.9.d, l'extraction des caractéristiques est accomplie. Pour ce faire, nous commençons tout d'abord par l'étiquetage des composantes connexes, comme décrit dans la section 5.2, permettant ainsi la détection de la plus grande décharge. A travers cette dernière décharge, le calcul des quatre indicateurs du vecteur de caractéristiques est procédé. Dès que la plus grande décharge est détectée, elle est entourée par un carré bleu, comme illustré sur la figure III.9.e.





(e) Détection de la décharge la plus importante

Figure III.9. Algorithme de traitement d'image proposé pour la détection des décharges de type arc

De prime à bord et du point de vue géométrique, il est clair que dans le cas de présence d'arc sur la surface isolante polluée, la plus grande décharge tend à prendre de la longueur et perdre de la largeur, à l'inverse du cas où les arcs sont absents, où elle est plus large que longue. Aussi, il est clair que la surface au contour bleu s'agrandit et prend plus de place avec l'augmentation de la tension appliquée et à l'approche du contournement. De ce fait, les mesures géométriques du carré bleu peuvent servir à diagnostiquer la présence d'arcs à la surface de l'isolateur. Par conséquent, une telle corrélation ne peut être que confirmée par l'analyse profonde et minutieuse des indicateurs géométriques du vecteur des caractéristiques. Dès lors, la section suivante servira à décrire en détail la variation de ce dernier.

III.3.2. Résultats du vecteur des caractéristiques

Contenant les quatre indicateurs explicités dans la section 5.2, le vecteur des caractéristiques est calculé pour divers niveaux de tension, commençant par $1 \text{ kV}_{\text{eff}}$, jusqu'à la tension de contournement ($32 \text{ kV}_{\text{eff}}$), comme montré sur la figure III.10. A l'aide de la caméra utilisée, l'analyse visuelle et minutieuse permet de déceler le niveau de tension de $20 \text{ kV}_{\text{eff}}$ comme étant tension minimale où les décharges de type arcs commencent à apparaître. Ainsi, la présente section sert à confirmer la corrélation préalablement supposée dans la section précédente (III.3.1) entre les indicateurs du vecteur de caractéristiques et l'apparition des arcs à la surface de l'isolateur.

Tout d'abord, N_l , qui représente le nombre de composantes connexes étiquetées, exhibe des valeurs bien plus grandes dans le cas où les arcs sont présents. Un tel résultat demeure prévisible car il prouve que le nombre de décharges augmente à la surface de l'isolateur avec l'augmentation de la tension et l'approche du contournement. Cependant, N_l montre de nombreuses fluctuations durant son augmentation. Une telle fluctuation est causée par l'aspect intermittent des décharges qui apparaissent avant le contournement. En effet, l'analyse visuelle des vidéos enregistrées confirme ce phénomène d'intermittence. Nous voyons clairement sur les vidéos que les décharges apparaissent et disparaissent très rapidement dans le deux cas où les arcs sont présents ou bien absents.

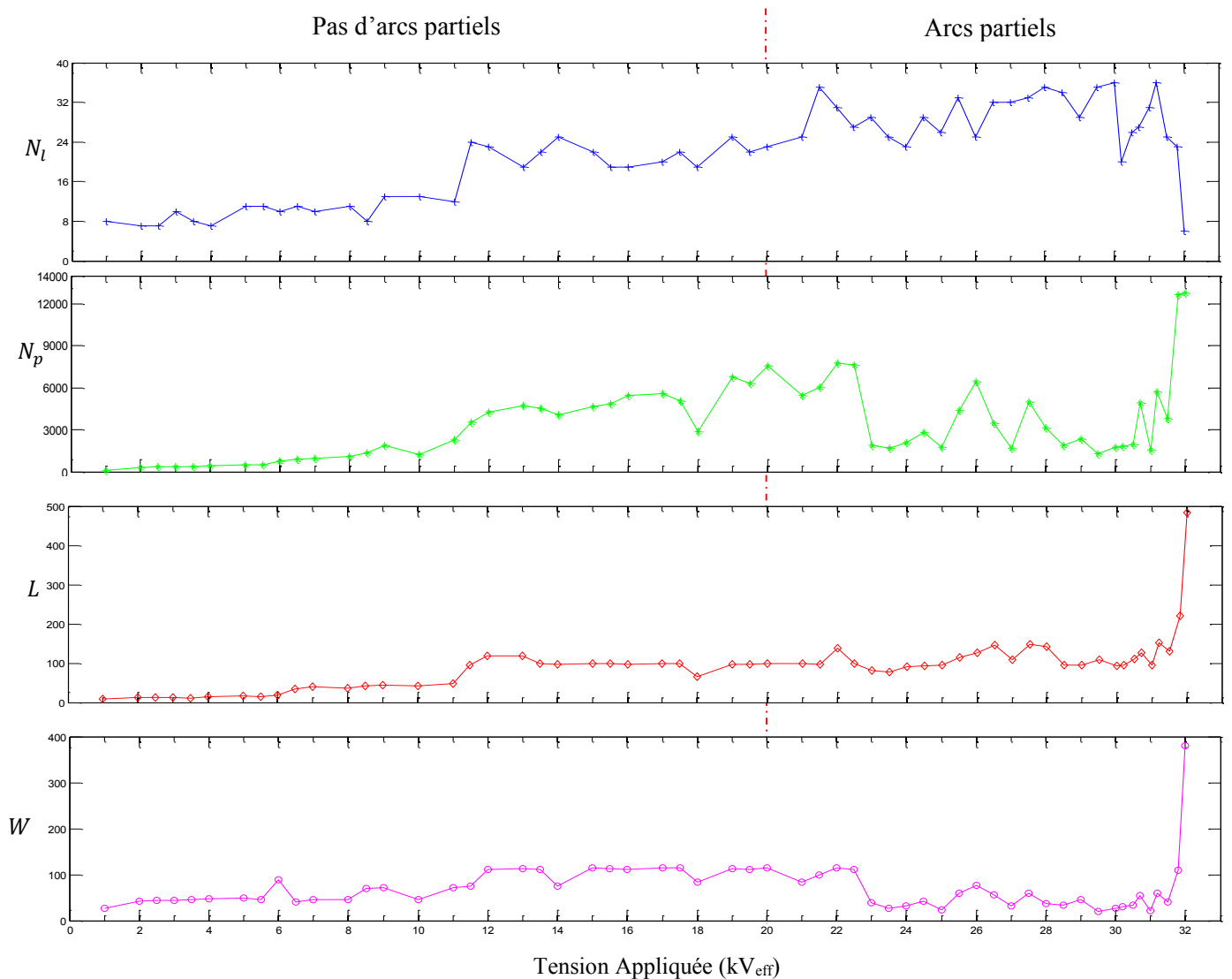


Figure III.10. Indicateurs du vecteur de caractéristiques pour une conductivité de 1,2 mS/cm

Il est utile aussi de signaler que N_l augmente jusqu'à la valeur de 36 pixels pour une tension de 31 kV_{eff} et diminue juste avant le contournement jusqu'à une valeur de 5 pixels. La diminution de N_l représente un résultat très important indiquant que le nombre de décharges diminue drastiquement. Une telle diminution représente alors un facteur révélateur d'un contournement imminent. Ce dernier se compose initialement de petites décharges qui vont

former plus tard l'arc de contournement, causant ainsi l'extinction des décharges avoisinantes et donc la diminution de N_l .

N_p représente le nombre de pixels de la plus grande décharge. Il montre de très faibles valeurs (moins que 1200 pixels) pour de faibles niveaux de la tension appliquée (inférieurs à 11 kV_{eff}). De telles valeurs de N_p montrent l'absence d'importantes décharges électriques. Cependant, N_p augmente rapidement pour atteindre la valeur de 7600 pixels pour 20 kV_{eff}, indiquant ainsi que les décharges se regroupent et se joignent pour occuper une décharge bien plus importante, sans pour autant former un arc sur la surface isolante.

Cependant, en présence d'arcs, N_p présente de faibles valeurs (au tour de 1700 pixels) entre 23 jusqu'à 25 kV_{eff} et 29 jusqu'à 30.5 kV_{eff}) accompagnées par de nombreux pics atteignant 6413, 4965, 4943 et 5696 pixels pour 26, 27.5, 30.7 et 31.2 kV_{eff} respectivement, comme montré par la figure III.10. Les faibles valeurs indiquent que les décharges sont disjointes et séparées. En fait, les petites décharges tendent à disparaître dans les voisinages des arcs électriques. A noter que les fortes fluctuations de N_p confirment l'apparition et disparition rapide de ces arcs.

L indique la longueur de la plus grande décharge et montre une lente augmentation avec la tension appliquée. Pour les niveaux de tension en dessous de 11 kV_{eff}, L reste inférieur à 45 pixels. De telles valeurs de L prouvent que les décharges occupent une petite surface sur l'isolateur. Cependant, pour des niveaux de tension entre 11 et 30 kV_{eff} (avant le contournement), L est plus ou moins égal à 100 pixels malgré la présence de petites fluctuations. Par contre, durant les derniers niveaux de tension précèdent le contournement (entre 30 et 32 kV_{eff}), nous notons l'augmentation rapide de L jusqu'à 484 pixels indiquant que la décharge finale commence à prendre forme et s'allonge rapidement pour aboutir à la décharge de contournement. Par ailleurs, la présence d'arcs sur certaines zones de la surface isolante résulte en la création de bandes sèches sur ces mêmes zones. Alors, L peut être considéré comme un outil pour mesurer la longueur des bandes sèches.

W désigne la largeur de la plus grande décharge. Il augmente très lentement pour atteindre son maximum (115 pixels) pour 20 kV_{eff}. Cette augmentation indique clairement que les décharges s'élargissent en absence d'arcs. Cependant, en présence d'arcs, W diminue drastiquement de 110 pixels (obtenus entre 20 et 22.5 kV_{eff}) à moins que 40 pixels (entre 23 et 31 kV_{eff}). Cette diminution indique que les arcs, qui se forment à partir des décharges électriques, tendent à perdre leur largeur durant leur élongation pour former l'arc de contournement final. De manière analogue à L , W peut être considéré comme la largeur de la bande sèche. De ce fait, de faibles valeurs de W sont justifiées par le fait que les arcs court-circuitent la bande sèche sur sa longueur totale, sans pour autant couvrir toute sa largeur. Par conséquent, la largeur des arcs est toujours inférieure à la largeur de la bande sèche qu'il court-circuite. Juste avant le contournement (31 kV_{eff}), W augmente rapidement, indiquant que l'arc final devient très large pour former la décharge de contournement.

De par les résultats exposés dans cette section, nous pouvons affirmer que les indicateurs du vecteur de caractéristique décrivent avec précision le comportement des décharges électriques à la surface de l'isolateur, plus spécialement des arcs. Dès lors, ces indicateurs peuvent être utilisés

comme entrées pour des méthodes d'apprentissage supervisé en vue de détecter la présence de décharges électriques de type arc.

XII.3. Résultats de la classification

Avant de passer à la validation de l'algorithme proposé en le testant sur le nombre d'échantillons décrit par le tableau 1, nous devons d'abord configurer nos méthodes de classification, plus spécialement les méthodes des Support Vecteur Machines (SVM) et la méthode des k plus proches voisins. Dans ce travail, le paramétrage le plus simple de ces dernières méthodes est adopté. Pour la SVM, le noyau linéaire est sélectionné, ainsi qu'une stratégie d'apprentissage de « One vs. All », alors que pour la k-ppv, la distance euclidienne est choisie, la valeur de k est fixée à 1.

Pour chaque échantillon, nous jugeons d'abord de la présence ou non d'arcs électriques pour chaque image. De ce fait, nous assignons une classe à chaque échantillon durant la phase d'apprentissage. Puis, durant la phase de test, la classe de sortie (présence ou bien absence d'arc électrique) de chaque échantillon de test est comparée à celle que nous lui avons assignée durant l'apprentissage. Pour évaluer la performance de chaque méthode, nous nous basons sur la mesure du taux d'erreur ou bien le taux de réussite. Etant un pourcentage, le taux de réussite correspond au nombre d'échantillons classés avec succès sur le nombre d'échantillons total formant l'ensemble de test. Les résultats sont représentés dans le tableau III.2.

Tableau III.2. Résultats de la classification des décharges

	Kppv	Bayes	SVM
Taux de réussite (%)	92,47	88,05	95,57

Les résultats comparatifs des trois méthodes de classification, présentés dans le tableau III.2, annonce la méthode des SVM comme le meilleur choix pour la détection des décharges de type arcs. La supériorité des SVM peut être expliquée par leur habilité à traiter tant les problèmes de classification linéaires que non linéaires. D'autant plus que d'après la figure III.10, les données d'entrée sont fortement non linéaires et présentent une corrélation très complexe avec la classe de sortie. Par ailleurs, la méthode du Kppv présente de très bons résultats malgré la simplicité de son principe de classification. Ces résultats peuvent être améliorés encore, en optimisant la valeur de k. Enfin, la classification bayésienne donne le moins bon taux de réussite. Ceci est dû à son principe basé sur la probabilité et son modèle à caractéristiques statistiquement indépendantes [90-91].

Néanmoins, il est très intéressant de noter que les résultats de classification sont, de manière générale, très bons. Ceci dénote l'efficacité de l'algorithme proposé pour la détection d'arcs électriques. Cette efficacité est non seulement due aux méthodes de classification, mais aussi à l'extraction des caractéristiques et des étapes qui la précèdent. Dès lors, nous pouvons affirmer que les différents outils utilisés, tels que la méthode de segmentation, le filtrage morphologique et l'extraction des caractéristiques sont en bonne adéquation et offrent d'excellents résultats une fois combinés.

III.4. Conclusion

Le travail décrit dans ce chapitre expose une nouvelle méthodologie pour l'analyse de l'activité des décharges électriques et le monitoring du contournement. Les tests ont été effectués sur une surface plane polluée pour assurer une meilleure visibilité de l'apparition et propagation des décharges. Après une analyse minutieuse de l'activité de ces décharges, nous avons décelé 8 états des décharges sur la surface de l'isolateur. Parmi ces 8 états, 4 incluent la formation d'arcs électriques. De ce fait, nous avons élaboré un algorithme en vue de déceler l'apparition de ces arcs. Pour cela, nous proposons de calculer quatre indicateurs (N_l , N_p , L et W) décrivant et quantifiant l'activité des décharges électriques sur la surface isolante. Puis, ces indicateurs sont utilisés comme entrées pour différentes méthodes de classification (Kppv, Naïve Bayes and SVM) en vue de discerner le cas de présence de décharge de type arc.

Dans notre travail, la méthode de segmentation d'Otsu permet d'obtenir de très bons résultats et représente assez bien la texture des images extraites à partir des vidéos d'essais montrant la propagation des décharges électriques sur le modèle plan de l'isolateur. Cependant, les images segmentées par la méthode d'Otsu contiennent aussi du bruit. Ce dernier est dû aussi bien à l'éclairage du laboratoire d'essai qu'à la lumière du jour qui sont réfléchis sur le bord supérieur de l'isolateur. De ce fait, nous proposons l'utilisation du filtrage morphologique. Ce filtrage, combine les opérations d'érosion et de dilatation pour supprimer les pixels non représentatifs des décharges électriques. Ainsi, nous constatons que le filtrage morphologique est indispensable pour permettre une vision correcte des décharges.

Les quatre indicateurs ainsi obtenus permettent de quantifier et de caractériser des phénomènes de décharges très intéressants :

- Les fluctuations de N_l et N_p précédant le contournement permettent de mettre en avant le phénomène d'intermittence des décharges électriques.
- Avec l'augmentation de la tension appliquée, le nombre de décharges électriques sur la surface isolante, quantifié par N_l , augmente. Cependant, ce nombre décroît juste avant le contournement ($1 \text{ kV}_{\text{eff}}$ avant la tension de contournement) indiquant l'apparition de l'arc de contournement final.
- N_p étant le nombre de pixels de la plus grande décharge, permet de savoir si les arcs sont présents. En effet, une fluctuation de N_p est synonyme d'activité d'arcs électriques.
- L et W permettent de quantifier l'espace pris par les arcs électriques. En effet, lors de faibles niveaux de la tension appliquée, les décharges sont dispersées sur des surfaces plus ou moins larges.
- La diminution de W est synonyme de formation d'arcs électriques.
- Les quatre indicateurs permettent de prévenir le phénomène de contournement des isolateurs. En effet, N_p , L et W décrivent la décharge la plus proéminente, ils augmentent

donc à l'approche du contournement. Cependant, N_l décompte le nombre de décharges à la surface de l'isolateur, il diminue donc avant le contournement.

Le choix de la méthode de classification est loin d'être anodin. Il dépend grandement du type de donnée d'entrée. Dans notre travail, les SVM donnent le meilleur taux de classification permettant de prévenir l'apparition du contournement.

Le travail effectué tout au long de ce chapitre prouve l'efficacité du traitement d'image pour le monitoring et la surveillance de l'état de surface d'un isolateur plan sous pollution. A travers un algorithme combinant la segmentation, l'extraction de caractéristique et la classification, l'apparition des arcs électriques a été détectée avec succès, permettant ainsi de prévenir le phénomène de contournement électrique.

Chapitre IV

L'ANALYSE QUANTITATIVE DE RECURRENCE (RQA) POUR LA CLASSIFICATION DE L'ETAT DE SURFACE D'UN ISOLATEUR POLLUE

IV.1. Introduction

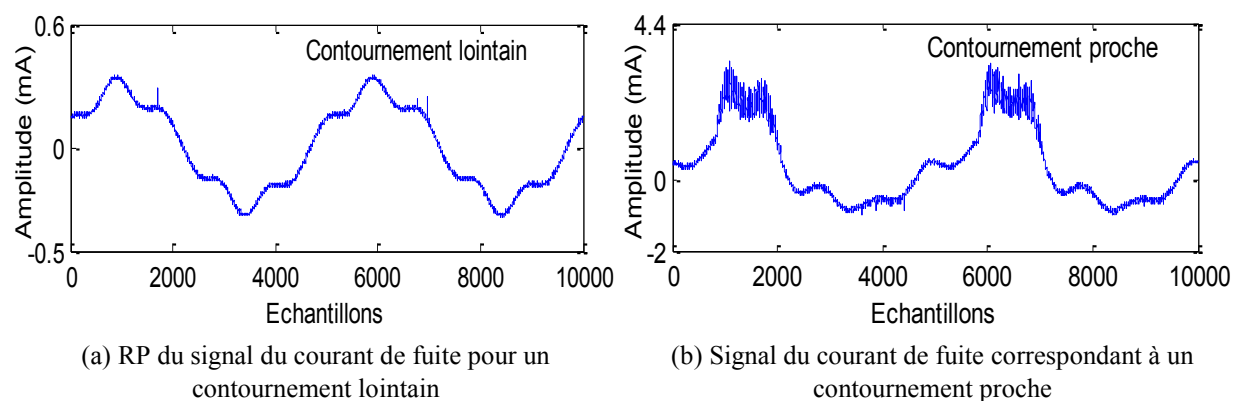
Récemment, la technique du diagramme de récurrence (Recurrent Plot, ou RP) a été introduite pour l'étude du signal du courant de fuite dans le cas d'isolateurs pollués [21-23]. En effet, étant basée sur une représentation graphique plus ou moins complexe, le diagramme de récurrence assure uniquement une analyse qualitative de l'état de surface de l'isolateur en question. En vue de surmonter ce manque, nous exposons dans ce présent chapitre une toute nouvelle technique appelée l'Analyse Quantitative de Récurrence (Recurrence Quantification Analysis, ou RQA). Cette dernière permet de quantifier la dynamique du signal du courant de fuite tout au long du processus de contournement. RQA est alors introduite pour analyser les structures qui apparaissent sur le RP et pour extraire différentes caractéristiques du signal du courant de fuite en vue d'étudier l'état de surface de l'isolateur pollué et de diagnostiquer sa performance. Pour ce faire, l'enregistrement du signal du courant de fuite est accompli sur un modèle d'isolateur plan uniformément pollué avec une solution saline. Puis, sur la base de ce dernier signal, huit indicateurs de la RQA sont calculés sous divers niveaux de pollutions. Par la suite, les valeurs moyennes de ces huit indicateurs sont proposées pour caractériser et étudier la performance de l'isolateur. En effet, ces indicateurs sont utilisés comme entrées à trois méthodes de classification (K-plus proches voisins, Naïve Bayes and Support Vector Machines) pour classer la sévérité de la pollution en 5 classes. Les résultats obtenus démontrent une excellente corrélation entre les indicateurs de la RQA et le niveau de la sévérité de pollution.

IV.2. La technique RP pour l'étude du phénomène de contournement des isolateurs

Connaissant la caractéristique non-linéaire du courant de fuite durant le processus du contournement, deux choix s'offrent à nous quand il s'agit d'utiliser les RP pour l'étude de ce phénomène. On applique la technique des RP soit directement aux signaux du courant de fuite, soit à leurs décompositions fréquentielles.

IV.2.1. La technique RP appliquée directement au signal du courant de fuite

La figure IV.1 montre un exemple de la technique RP appliquée directement à un signal du courant de fuite typique lors du processus de contournement. Les figures IV.1.a et IV.1.b montrent les signaux originaux des courants de fuite correspondant respectivement à un contournement lointain et proche.



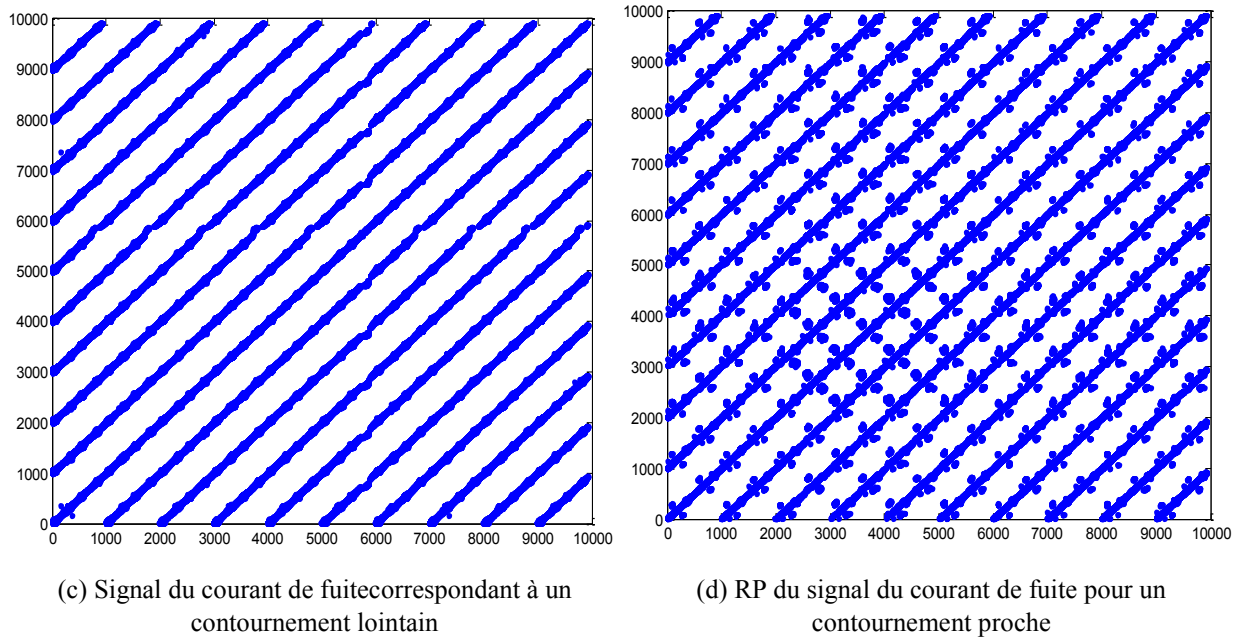


Figure IV.1. Exemple de la technique RP appliquée directement sur le signal du courant de fuite

Pour un contournement lointain (faible tension appliquée), nous remarquons la présence de lignes parallèles à la diagonale principale (LOI) sur le RP de la figure IV.1.c, ce qui indique qu'on est en présence d'un phénomène périodique. Ces lignes sont d'une grande longueur, signifiant ainsi un phénomène déterministe. De plus, on observe de courtes lignes orthogonales à la LOI, ce qui indique une évolution du courant de fuite similaire à d'autres périodes (voir tableau I.1).

Pour un niveau de tension proche de celui du contournement, nous remarquons l'absence des lignes verticales à la LOI sur le RP (figure IV.1.d), laissant place uniquement aux lignes parallèles à cette dernière. Une telle structure signifie que le processus étudié est déterministe.

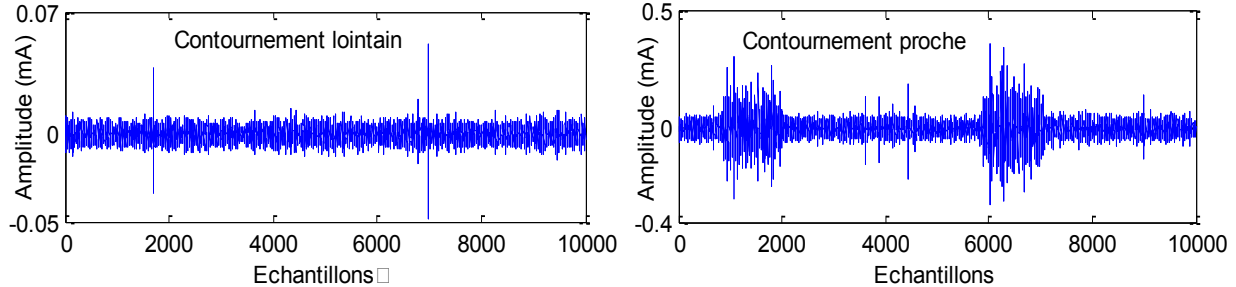
Ces conclusions ne reflètent pas du tout la réalité du processus de contournement. C'est pourquoi, durant notre étude, nous appliquerons le diagramme de récurrence aux composantes fréquentielles du signal du courant de fuite.

IV.2.2. La technique RP appliquée au détail du signal du courant de fuite

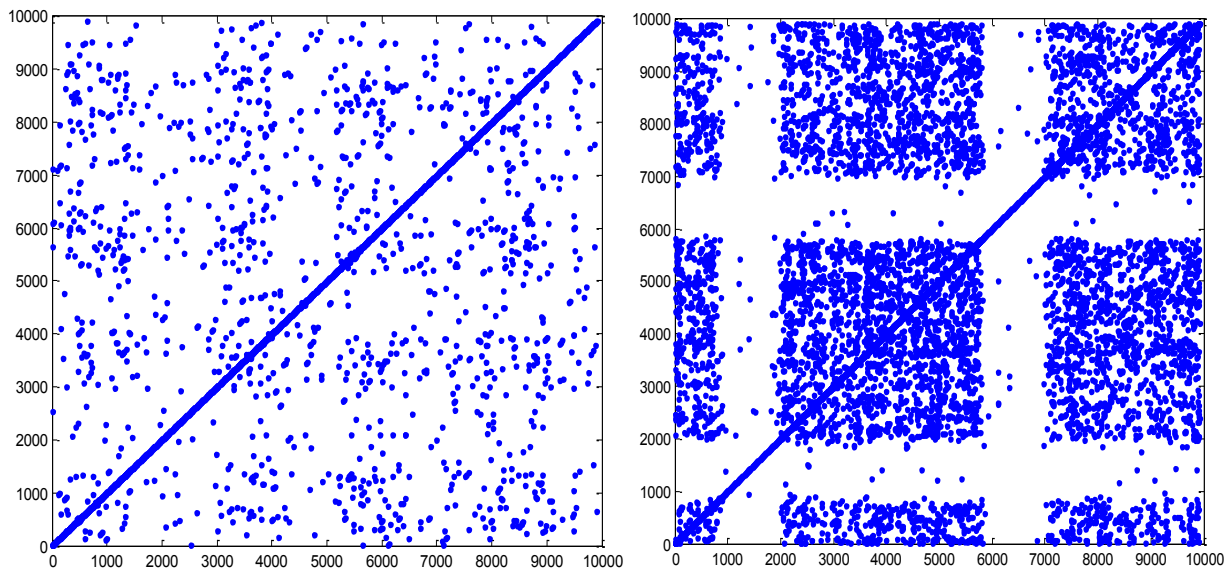
Dans notre travail, nous allons appliquer la DWT en utilisant l'ondelette de Daubechies 4 pour décomposer le signal du courant de fuite afin d'identifier les transitions pour la pollution non-uniforme, qui sont causées par des perturbations de hautes fréquences [21-23]. La décomposition par DWT est effectuée comme expliqué dans le paragraphe 3, du chapitre I.

La figure IV.2 représente le RP des composantes hautes fréquences du signal du courant de fuite, plus précisément du détail D1. Le choix de ce détail sera expliqué par la suite. Pour un faible niveau de tension appliquée, le RP correspondant au détail D1 (figure IV.2.c) est composé de points de faible densité dispersés tout au long du diagramme. Aucune forme particulière n'est visible. Une telle distribution des points démontre parfaitement l'état des décharges sur la surface de l'isolateur. Ainsi, pour un contournement loin, il n'existe que de faibles décharges partielles, ce qui est en accord avec l'observation visuelle.

Pour un niveau de tension proche de celui du contournement, le RP correspondant au détail D1 (figure IV.2.d) est composé de deux éléments : les bandes blanches, et les points de forte densité. La présence de bandes blanches indique que le processus étudié est de nature non stationnaire et qu'il présente des états de transition (tableau I.1). Par contre, les points de forte densité, parfois isolés, indique une activité intense sur la surface de l'isolateur.



(a) Détail D1 de la décomposition en ondelettes du signal du courant de fuite pour un contournement lointain (b) Détail D1 de la décomposition en ondelettes du signal du courant de fuite pour un contournement proche



(c) RP du détail D1 de la décomposition en ondelettes pour un contournement lointain

(d) RP du détail D1 de la décomposition en ondelettes pour un contournement proche

Figure IV.2. Exemple de la technique RP appliquée au détail D₁

En pratique, à l'approche du contournement, nous remarquerons une intensification des décharges, représentée par les points à forte densité. On observe aussi une intermittence de ces décharges sur la surface de l'isolateur, ceci est décrit sur la figure IV.2.d par les bandes blanches. Le rôle de ces dernières s'avère donc capital pour déceler l'état de transition menant au contournement.

Les résultats obtenus par l'application de la technique des RP aux décompositions fréquentielles sont parfaitement en accord avec l'observation visuelle du phénomène de contournement.

IV.2.3. Choix des paramètres m , τ et ε

D'après les études menées, les valeurs de la dimension de plongement m et du retard τ n'ont pratiquement pas d'effet sur la structure topologique du RP [21-23]. Néanmoins, elles sont choisies, via la «C-C method», égales à 5 et 20 respectivement [36].

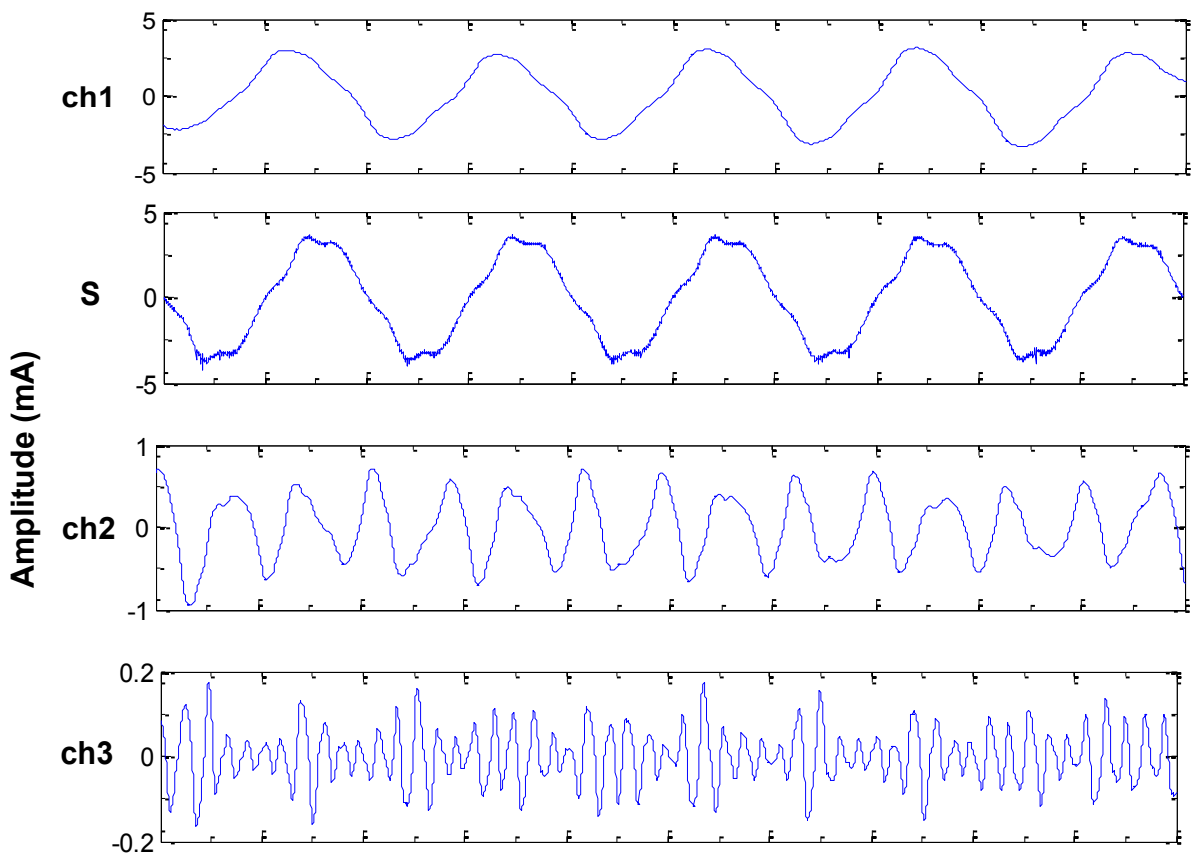
La valeur de ε est critique. Si elle est trop grande elle produira des points incohérents et non pertinents. Par contre, si elle est trop petite, l'information utile sera perdue. Cependant, il n'existe pas de méthode standard pour déterminer la valeur optimale de ε . La plupart des chercheurs la choisissent égale à $0,25\sigma$, où σ est la standard déviation du signal étudié [21-23].

IV.3. RP des composantes fréquentielles du signal de courant de fuite

D'après ce qui a été montré précédemment, la décomposition fréquentielle du signal de courant de fuite s'avère être nécessaire avant d'utiliser la techniques RP. Différents détails correspondant aux différents niveaux de décomposition sont donc extraits.

Dans le présent paragraphe, nous décomposons le signal du courant de fuite précédant un contournement par la DWT à l'aide de l'ondelette Daubechies 4 en 10 niveaux. On extrait alors les différents détails correspondant aux différentes bandes de fréquence pour les trois classes de pollution; cette étape est illustrée par les figures IV.3 à IV.5. Enfin, nous appliquerons la technique RP aux différents niveaux de décomposition obtenus pour la dernière classe de pollution, comme montré sur la figure IV.6.

IV.3.1. Cas d'une conductivité de 0,01 mS/cm



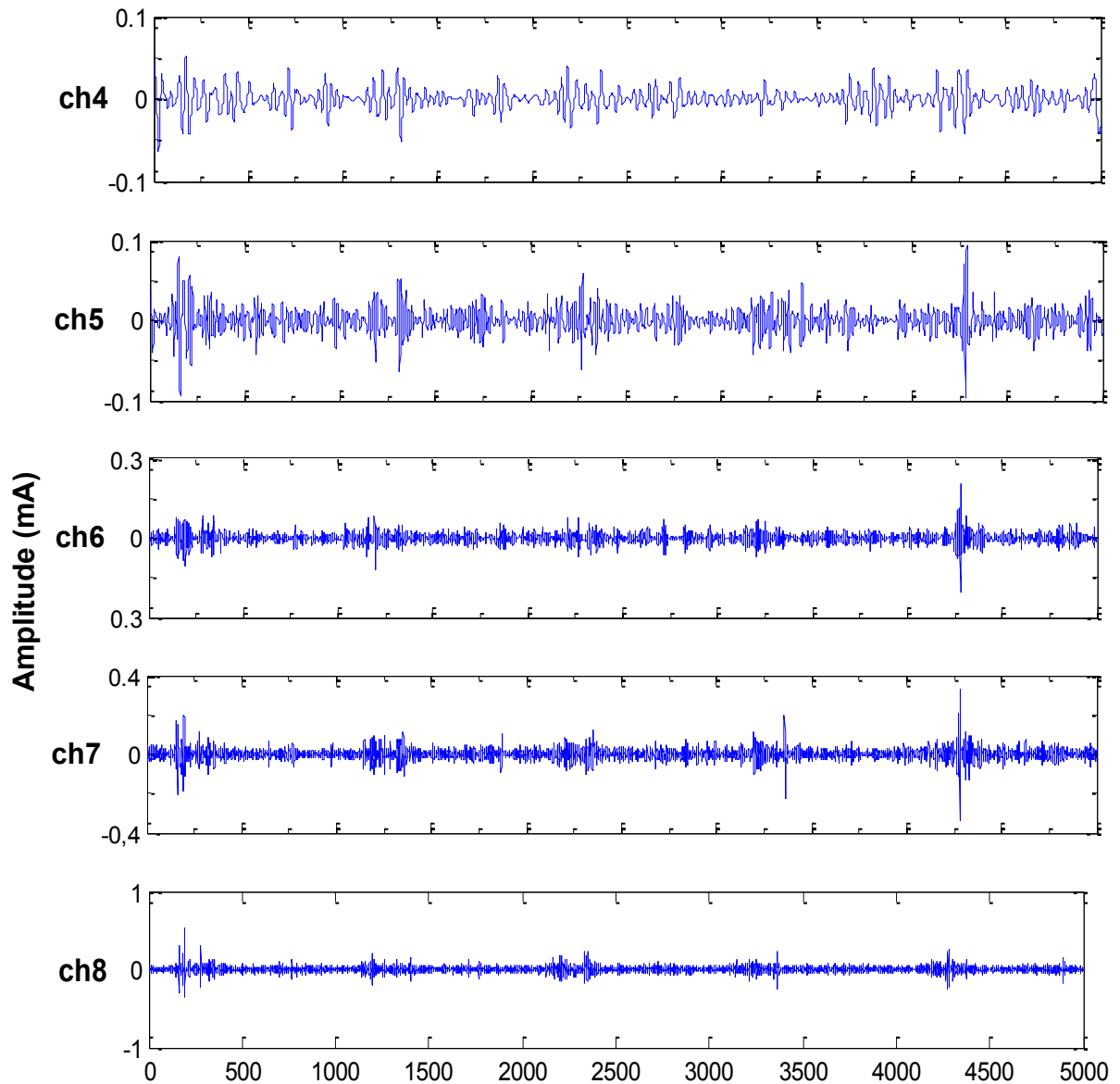
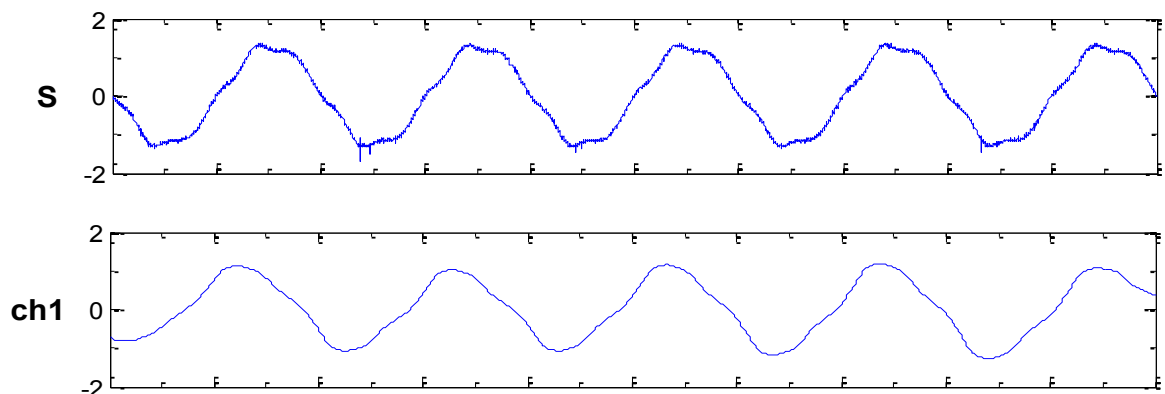


Figure IV.3. Extraction des composantes fréquentielles du courant de fuite précédant le contournement pour une conductivité de 0,01 mS/cm:
 s indique le signal original, ch1 définit le coefficient D_{10} (50 Hz), ch2 définit D_8 (390- 1953 Hz), ch3 définit D_6 (781 – 1562 Hz), ch4 définit D_5 (1,56 – 3,12 kHz), ch5 définit D_4 (3,12 – 6,25 kHz), ch6 définit D_3 (6,25 – 12,5 kHz), ch7 définit D_2 (12,5 - 25 kHz), ch8 définit D_1 (25 - 50 kHz)

IV.3.2. Cas d'une conductivité de 1,2 mS/cm



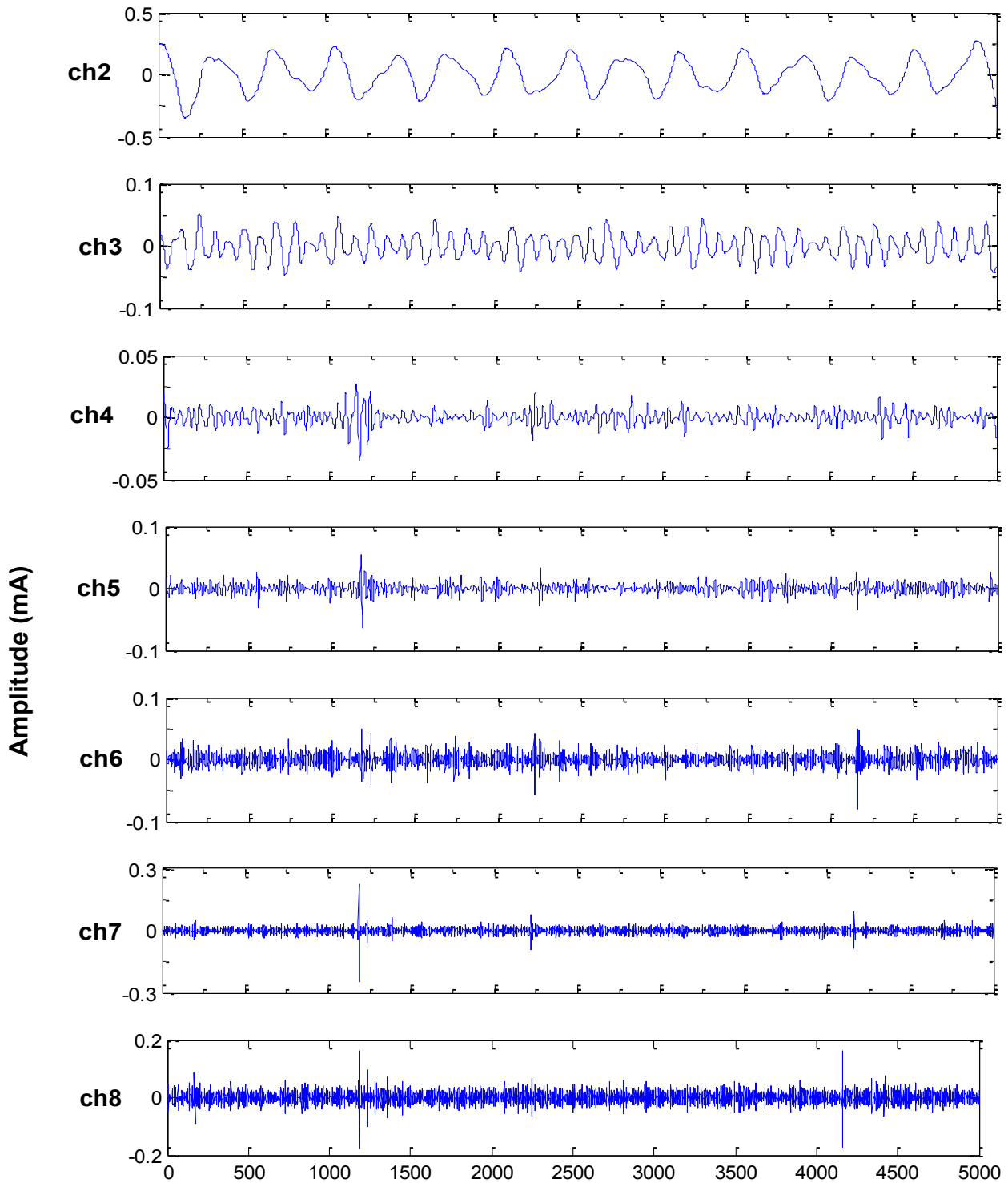
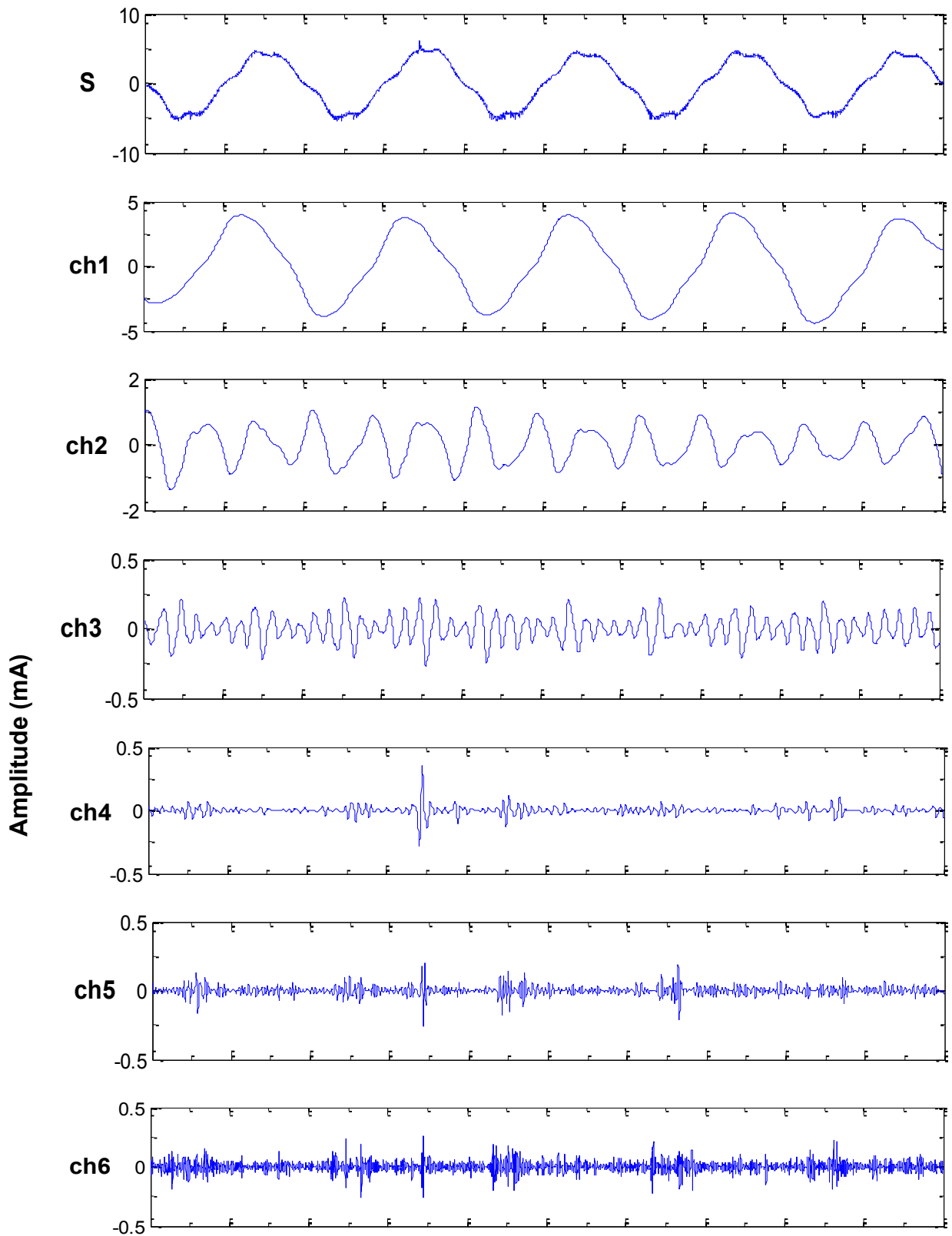


Figure IV.4. Extraction des composantes fréquentielles du courant de fuite précédant le contournement pour une conductivité de 1,2 mS/cm:
 s indique le signal original, ch1 définit le coefficient D_{10} (50 Hz), ch2 définit D_8 (390- 1953 Hz), ch3 définit D_6 (781 – 1562 Hz), ch4 définit D_5 (1,56 – 3,12 kHz), ch5 définit D_4 (3,12 – 6,25 kHz), ch6 définit D_3 (6,25 – 12,5 kHz), ch7 définit D_2 (12,5 - 25 kHz), ch8 définit D_1 (25 - 50 kHz)

IV.3.3. Cas d'une conductivité de 10,1 mS/cm



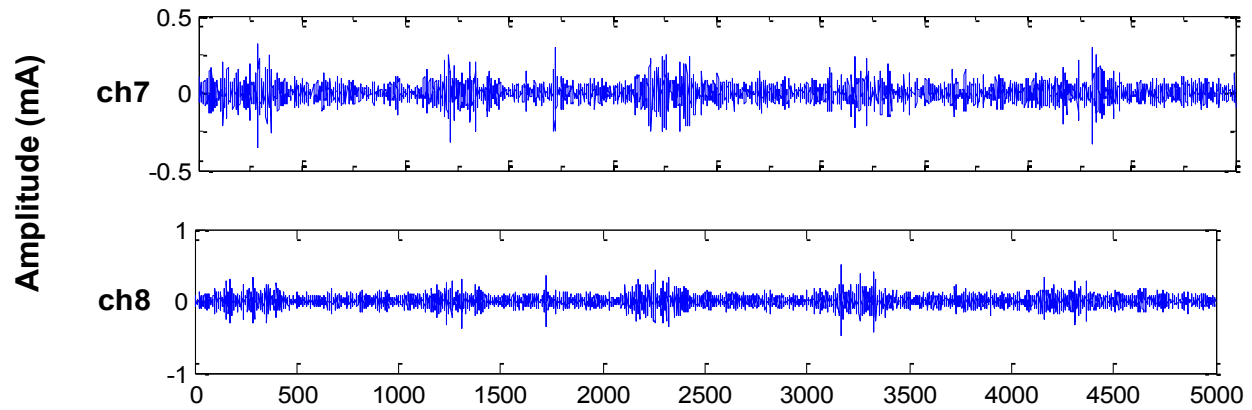


Figure IV.5. Extraction des composantes fréquentielles du courant de fuite précédant le contournement pour une conductivité de 10,1 mS/cm:

s indique le signal original, ch1 définit le coefficient D_{10} (50 Hz), ch2 définit D_8 (390- 1953 Hz), ch3 définit D_6 (781 – 1562 Hz), ch4 définit D_5 (1,56 – 3,12 kHz), ch5 définit D_4 (3,12 – 6,25 kHz), ch6 définit D_3 (6,25 – 12,5 kHz), ch7 définit D_2 (12,5 - 25 kHz), ch8 définit D_1 (25 - 50 kHz)

D'après les études menées [21-23], les composantes fréquentielles peuvent être séparées en trois intervalles : basses (ch1 à ch3), moyennes (ch4 à ch7) et hautes fréquences (ch8).

En ce qui concerne les composantes faibles fréquences (50 – 1562 Hz), la figure IV.5.a comporte très peu de structures ; elle n'est composée que de quelques fragments de lignes parallèles à la LOI. Une structure si pauvre indique très peu de décharges électriques sur la surface de l'isolateur. Alors que les courtes lignes diagonales indiquent que le processus étudié comporte des composantes déterministes. Une telle structure est bien évidemment totalement contraire à la théorie, ainsi qu'à l'observation visuelle de l'état de décharge.

Cependant, nous remarquons sur la figure IV.5.b décrivant ch2 que les lignes diagonales ont tendance à se rassembler, et à gagner en densité. Les bandes blanches sont bien visibles, indiquant un état très instable du processus étudié.

La figure IV.5.c, décrivant ch3, est elle aussi très pauvre en structures. Elle ne contient pratiquement plus de lignes diagonales à la LOI, mais hérite quand même des bandes blanches. On notera aussi une diminution des points, comparativement à la figure IV.5.b décrivant ch2.

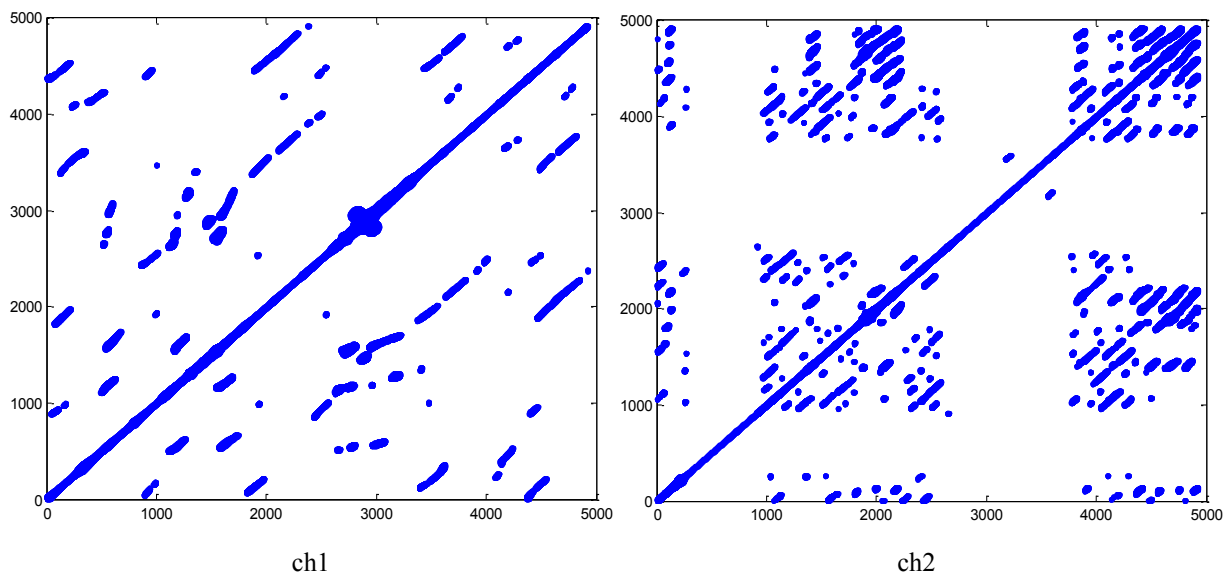
Bien que les figures IV.5.b et IV.5.c comportent des bandes blanches (élément essentiel indiquant la transition vers un contournement imminent), elles restent inadéquates pour décrire l'état des décharges sur la surface de l'isolateur, car beaucoup de lignes diagonales parallèles à la LOI y sont représentées, ce qui est totalement en désaccord avec la théorie.

En ce qui concerne les composantes moyennes fréquences (1,56 – 25 kHz), elles sont représentées par les figures IV.5.d à IV.5.g. En dépit de la constante présence des bandes blanches, nous remarquons une structure radicalement différente comparée aux composantes basses fréquences. Ainsi, on observe une forte augmentation de la densité des points, comparé à la figure la précédant. Une telle intensification des points indique une forte activité des décharges à la surface de l'isolateur. Cette conclusion peut être tirée également pour toutes les figures correspondant aux composantes moyennes fréquences. Ainsi, la figure IV.5.d se rapproche sensiblement de la figure IV.5.e : nous remarquons la présence de petites zones vides à l'intérieur de grandes zones à points très intenses. On notera une structure du RP de plus en plus opaque.

Quant à la figure IV.5.f, représentant ch6, elle se caractérise par une forte densité des points, rendant la structure quasi-opaque, ce qui indique une activité extrême des décharges sur la surface de l'isolateur. La densité de points augmente encore pour la figure IV.5.g représentant ch7, créant ainsi une structure totalement opaque.

Les composantes moyennes fréquences décrivent sans doute l'activité des décharges à la surface de l'isolateur mieux que les composantes basses fréquences. Cependant, l'activité des décharges décrites par les figures IV.5.f et IV.5.g nous laisse quelque peu perplexes. Une si forte activité reste contraire à la théorie. Une structure si opaque décrit maladroitement l'activité de décharges au niveau de l'isolateur.

Concernant la figure IV.5.h, représentant le RP de la composante fréquentielle D_1 (25 – 50 kHz), elle paraît être très proche de celles qui la précèdent, mais est en réalité bien plus intéressante. Elle garde tout naturellement les bandes blanches, préconisant l'apparition d'un état de transition imminent. Elle se différencie surtout par une représentation équilibrée et juste des décharges à la surface de l'isolateur. A l'opposé des figures qui la précèdent, elle ne comporte pas de structure opaque décrivant à tort l'activité des décharges, mais bien une structure composée de points denses et de petits vides, ce qui indique une activité intense de décharges mais qui reste modérée et acceptable, conformément à la théorie et à ce que nous pouvons trouver dans la littérature [21-23].



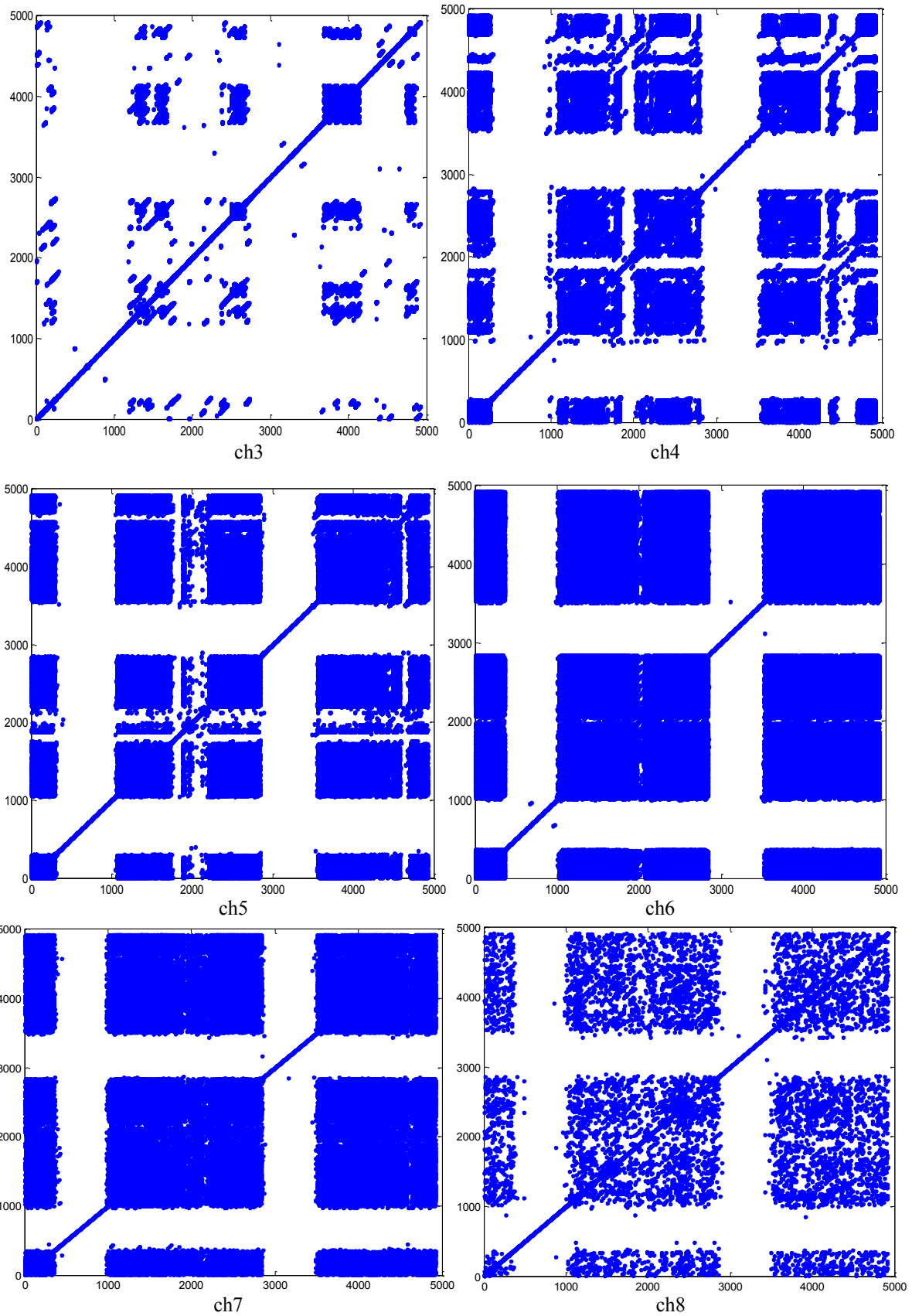


Figure IV.6. RP de la décomposition en ondelettes du signal du courant de fuite avant le contournement. ch1 définit le coefficient D10 (50 Hz), ch2 définit D8 (390- 1953 Hz), ch3 définit D6 (781 –1562 Hz), ch4 définit D5 (1,56 – 3,12 kHz), ch5 définit D4 (3,12 – 6,25 kHz), ch6 définit D3 (6,25 – 12,5 kHz), ch7 définit D2 (12,5 - 25 kHz), ch8 définit D1 (25 - 50 kHz)

De la figure IV.6, nous pouvons donc tirer les conclusions suivantes.

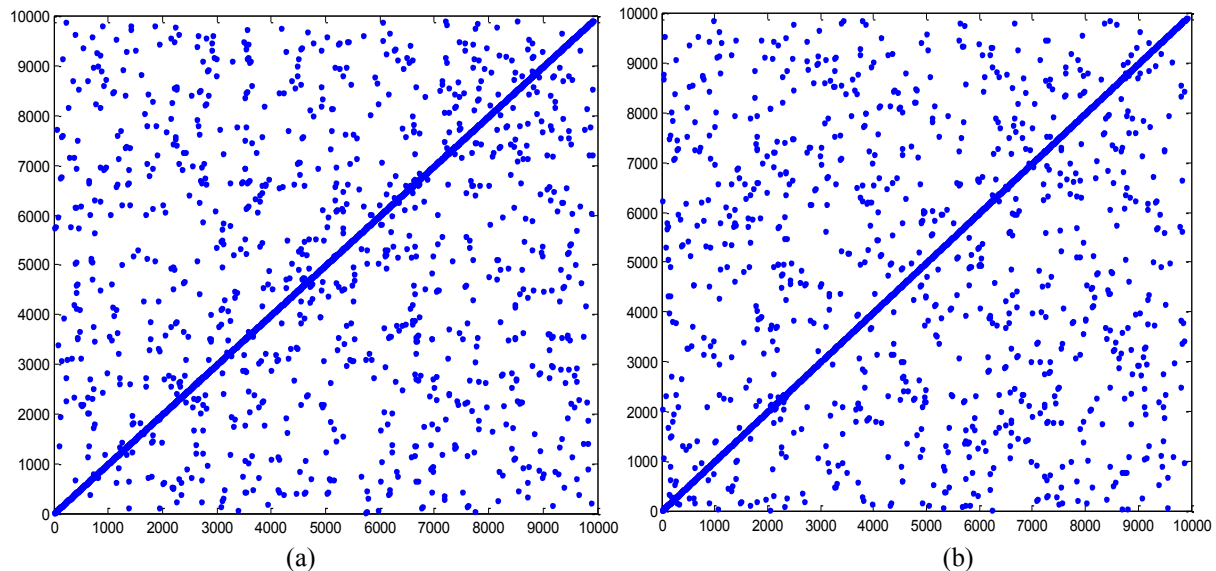
- Les composantes basses fréquences (50 – 1562 Hz) sont totalement inadéquates quand il s'agit d'étudier le phénomène de contournement des isolateurs. Elles présentent une analyse totalement erronée de la situation réelle. Elles sont à proscrire indubitablement.
- Les composantes moyennes fréquences (1,56 – 25 kHz) décrivent maladroitement le comportement des décharges. Le RP qu'elles fournissent manque grossièrement de précision. Sa représentation opaque est démesurée. Leur utilisation est donc fortement déconseillée.
- La composante haute fréquence (25 – 50 kHz) est la meilleure pour la description de l'activité des décharges sur la surface de l'isolateur, ainsi que pour l'étude du phénomène de contournement car elle est composée de bandes blanches, indiquant une forte intermittence des décharges, et de points denses. Leur représentation justifie la présence de décharges intermittentes et irrégulières précédant le contournement. Ces décharges vont en s'affaiblissant, laissant place à un seul arc de décharge. Ce dernier, avec une très légère augmentation de la tension, donnera suite à un arc de contournement imminent.

IV.4. RP de la composante haute fréquence du signal de courant de fuite

On notera que désormais, dans ce paragraphe, la technique RP sera exclusivement appliquée au détail haute fréquence D_1 résultant de la décomposition préalable à l'aide de la DWT su signal de coutant de fuite en utilisant l'ondelette de *Daubechies 4*.

IV.4.1. Cas d'une conductivité de 0,01 mS/cm

Pour cette classe de pollution, nous nous intéressons à un signal du courant de fuite sous une pollution non uniforme d'une conductivité de 0,01 mS/cm.



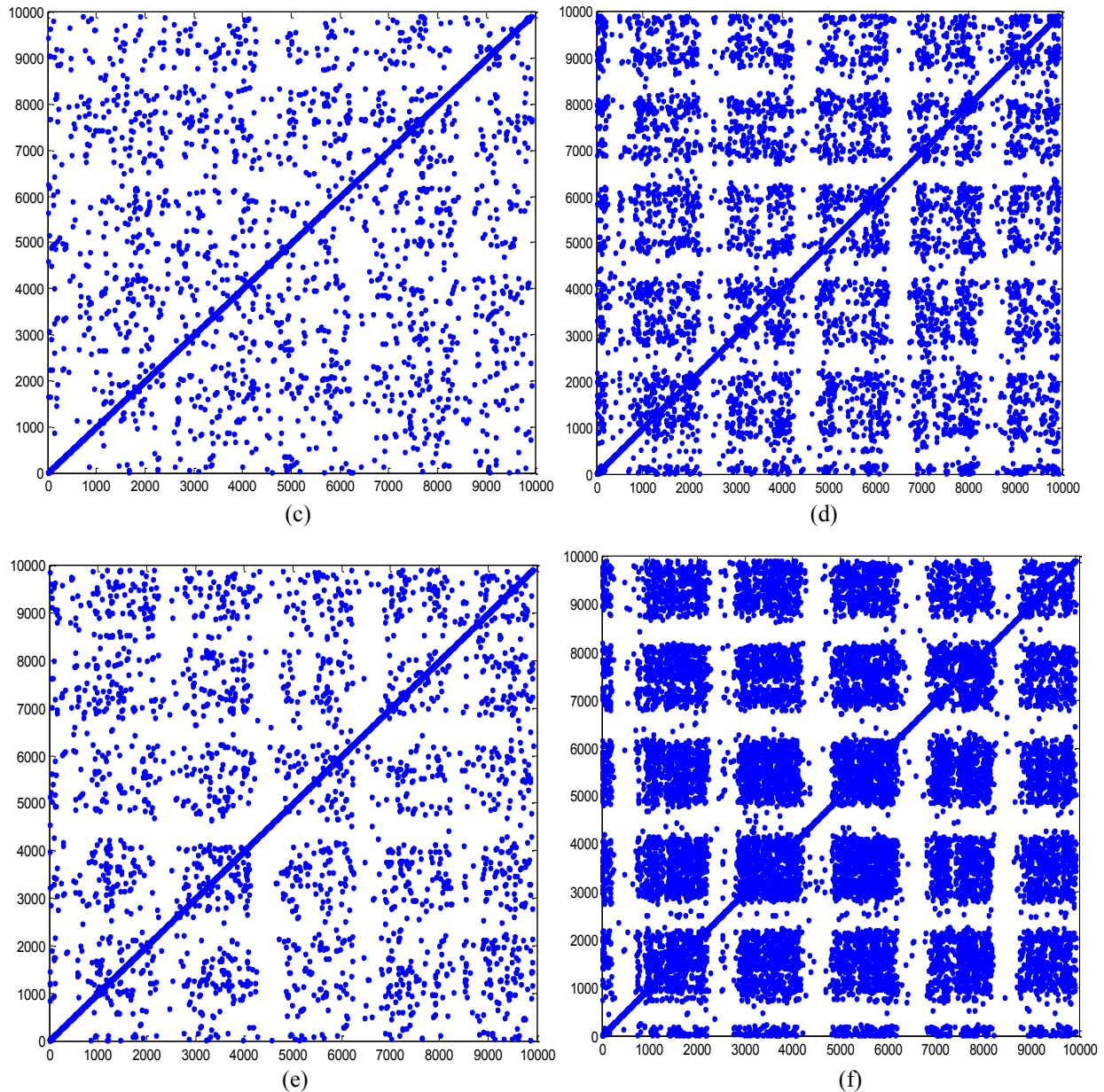


Figure IV.7. RP de la composante haute fréquence (D_1) du signal du courant de fuite en fonction de la tension appliquée pour une pollution de conductivité $0,01 \text{ mS/cm}$:

(a) $5 \text{ kV}_{\text{eff}}$, (b) $10 \text{ kV}_{\text{eff}}$, (c) $15 \text{ kV}_{\text{eff}}$, (d) $20 \text{ kV}_{\text{eff}}$, (e) $25 \text{ kV}_{\text{eff}}$, (f) $30 \text{ kV}_{\text{eff}}$

La figure IV.7 montre les différents RP de la composante haute fréquence, le détail D_1 (25-50 kHz), pour une pollution de $1,2 \text{ mS/cm}$ en fonction de la tension appliquée.

D'après la figure IV.7.a, pour une faible tension de $5 \text{ kV}_{\text{eff}}$ appliquée à l'isolateur, nous remarquons la présence de quelques points épars. Il en est de même pour la figure IV.7.b représentant le RP pour une tension appliquée de $10 \text{ kV}_{\text{eff}}$. Ainsi, une telle structure indique la présence de rares décharges à la surface de l'isolateur. Ces dernières sont d'une intensité faible. L'analyse visuelle lors des essais permet de vérifier et de confirmer cela, car nous remarquons que pour de faibles tensions appliquées, des décharges apparaissent dans les zones propres, mais restent toujours faibles et localisées. Par conséquent, nous soulignerons l'absence d'arc durant ces niveaux de tensions.

Pour une tension appliquée supérieure de 15 kV_{eff}, la figure IV.7.c se compose de points denses. Leur présence est plus consistante comparativement aux deux figures précédentes. Leur intensité est elle aussi plus importante. Une telle transition indique une augmentation de l'activité des décharges à la surface de l'isolateur. Ces dernières deviennent donc plus denses et plus importantes, ce qui est en conformité avec notre analyse visuelle (figure II.1). Mais elles restent toujours faibles pour pouvoir créer des arcs électriques.

Par contre, pour une tension appliquée de 20 kV_{eff}, la figure IV.7.d se différencie totalement des figures précédentes. On remarque d'une part, des points bien plus denses, plus concentrés, mais aussi une apparition de petites bandes blanches. L'augmentation de la densité des points indique une activité des décharges bien plus importante comparée à la figure précédente. Alors que l'apparition de petites bandes blanches indique l'approche d'un état menant vers une transition certaine. Les décharges se font donc de plus en plus intermittentes. Notre analyse visuelle nous permet de confirmer cela. Ainsi, pour ce niveau de tension, nous remarquons une saturation complète des bandes propres. La présence des décharges partielles y est très importante. De cet état, nous pouvons prévoir une future apparition d'arcs électriques (figure II.1).

Ainsi, pour une tension appliquée de 25 kV_{eff}, la figure IV.7.e garde relativement la même structure que la figure qui la précède, mais avec une diminution de densité des points qui la composent. Une telle transition peut sembler contradictoire, pourtant, elle explique parfaitement la réalité des décharges à la surface de l'isolateur. En effet, une diminution de densité des points indique une diminution de l'activité des décharges. Cela indique l'apparition d'arcs électriques. Ces conclusions sont en total accord avec l'observation visuelle (figure II.1). Ainsi, nous remarquons la formation de petits arcs électriques aptes à court-circuiter les zones propres d'une part, mais aussi à commencer une propagation dans les zones polluées. Cela prédit la diminution de l'effet capacitif du courant de fuite, et l'augmentation de son effet résistif.

Enfin, pour une tension appliquée de 30 kV_{eff}, la figure IV.7.f se démarque par une structure unique : une forte présence des bandes blanches, ainsi qu'une augmentation évidente de l'intensité des points qui la composent comparativement à la figure qui la précède. Une telle structure indique à la fois une intensification des décharges au niveau de l'isolateur, et la présence d'un état fort intermittent qui approche. La combinaison de ces deux éléments prédit l'apparition imminente d'un arc de contournement. Visuellement, nous remarquons la propagation des arcs électriques sur les zones polluées, ces derniers court-circuitent totalement les zones propres, et tendent se connecter entre eux (figure II.1). Si cela se réalise, nous observerons l'arc de contournement liant directement la borne HT à la borne terre, indiquant dès lors un comportement quasi-résistif du courant de fuite.

IV.4.2. Cas d'une conductivité de 1,2 mS/cm

Pour cette classe de pollution, nous nous intéressons au RP du signal du courant de fuite sous une pollution non uniforme d'une conductivité de 1,2 mS/cm.

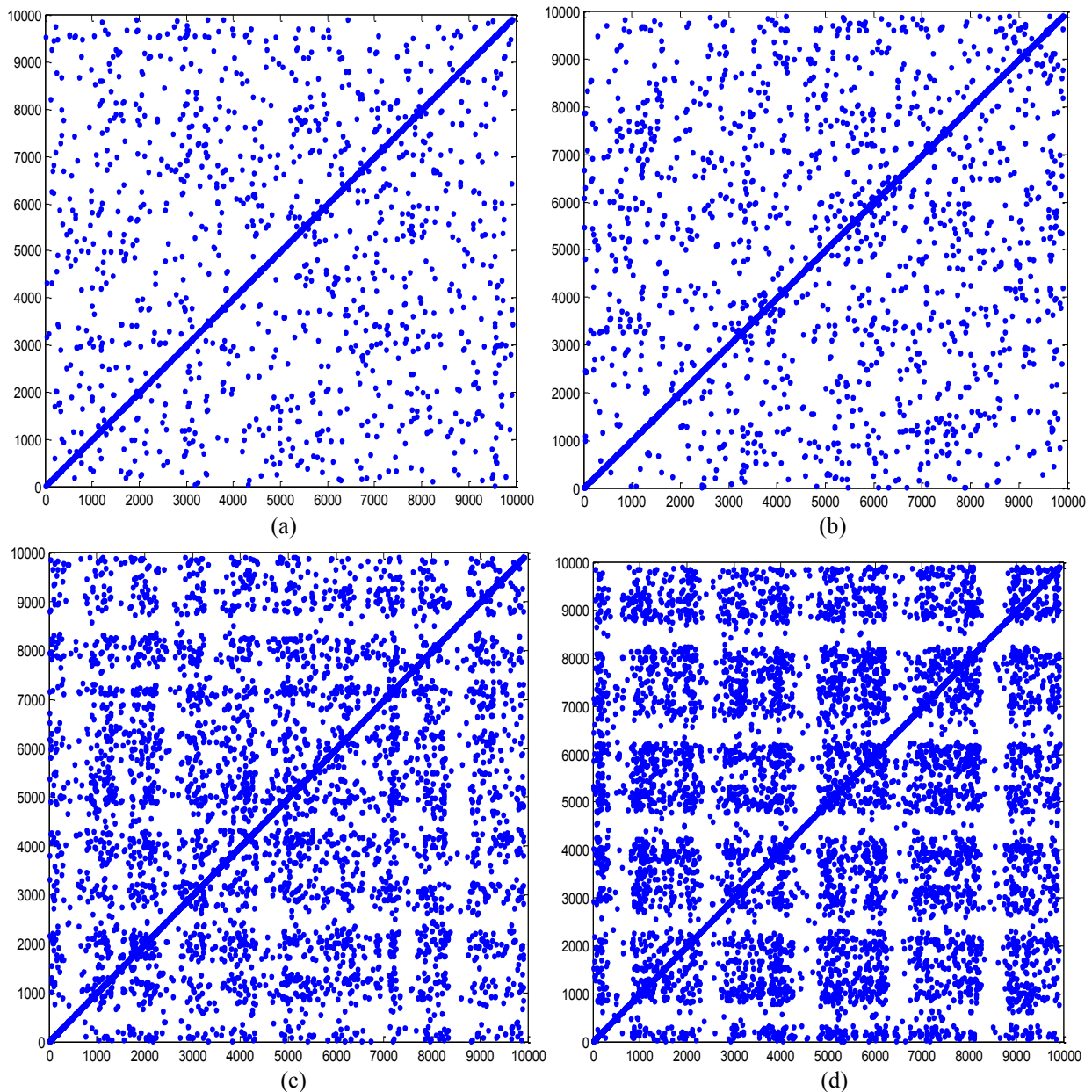


Figure IV.8. RP des composantes hautes fréquences (D_1) du signal du courant de fuite en fonction de la tension appliquée pour une conductivité de pollution de $1,2 \text{ mS/cm}$: (a) $15 \text{ kV}_{\text{eff}}$, (b) $20 \text{ kV}_{\text{eff}}$, (c) $25 \text{ kV}_{\text{eff}}$ et (d) $30 \text{ kV}_{\text{eff}}$

La figure IV.8 montre les différents RP des composantes hautes fréquences du détail D_1 (25 à 50 kHz) pour une pollution de $1,2 \text{ mS/cm}$ en fonction de la tension appliquée.

Pour les niveaux de tensions 5 et $10 \text{ kV}_{\text{eff}}$, leurs RP correspondant est totalement similaire à celui de $15 \text{ kV}_{\text{eff}}$ (c'est pour cela qu'ils ne sont pas présentés sur la figure IV.8). Ce dernier n'est composé que de quelques points, ce qui indique une faible activité des décharges à la surface de l'isolateur. En ce qui concerne la figure IV.8.b représentant le RP pour une tension appliquée de $20 \text{ kV}_{\text{eff}}$, une très légère augmentation du nombre de points est à signaler, indiquant une légère intensification des décharges au niveau de l'isolateur. L'analyse visuelle lors des essais permet de vérifier cela, car nous remarquons que pour de faibles tensions appliquées, les décharges apparaissent dans les zones propres, mais restent faibles. Par conséquent, nous soulignerons

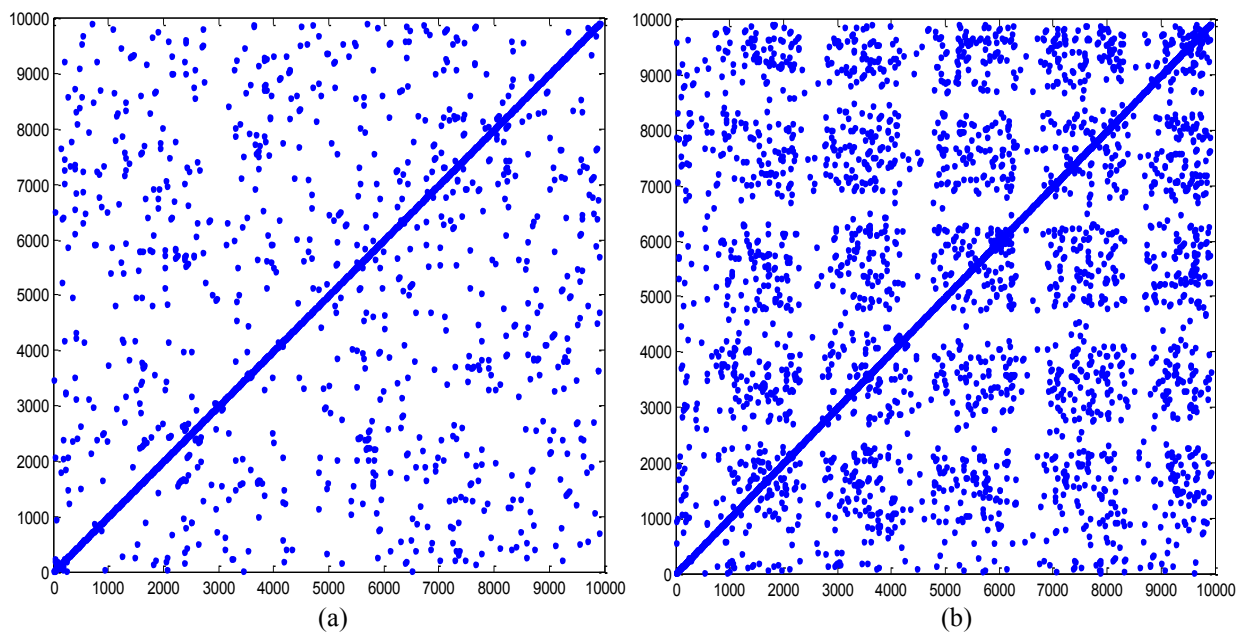
l'absence d'arcs pour ces niveaux de tension. Cela indique le fort effet capacitif du courant de fuite, alors que l'effet résistif est quasi-inexistant.

Par contre, pour une tension appliquée de 25 kV_{eff}, la figure IV.8.c se différencie des figures précédentes. On remarque d'un part, des points bien plus denses, mais aussi une apparition de petites bandes blanches. L'augmentation de la densité des points indiquent une activité des décharges plus importantes comparées à la figure précédente. Alors que l'apparition de petites bandes blanches indique l'approche d'un état menant vers une transition certaine. Les décharges se font donc plus intenses, et quelques peu instables.

Dès lors, pour une tension appliquée de 30 kV_{eff}, la figure IV.8.d garde relativement la même structure que la figure qui la précède, mais avec une augmentation de la densité des points qui la composent et un élargissement des bandes blanches. Ainsi, nous remarquons la formation de petits arcs électriques et leurs propagations dans les zones polluées. Cela prédit la diminution de l'effet capacitif du courant de fuite, et l'augmentation de son effet résistif. Une si forte présence des bandes blanches prédit l'apparition imminente d'un arc de contournement.

IV.4.3. Cas d'une conductivité de 10,1 mS/cm

Dans ce cas, nous appliquerons la technique RP à un signal du courant de fuite correspondant à une conductivité de 10,1 mS/cm.



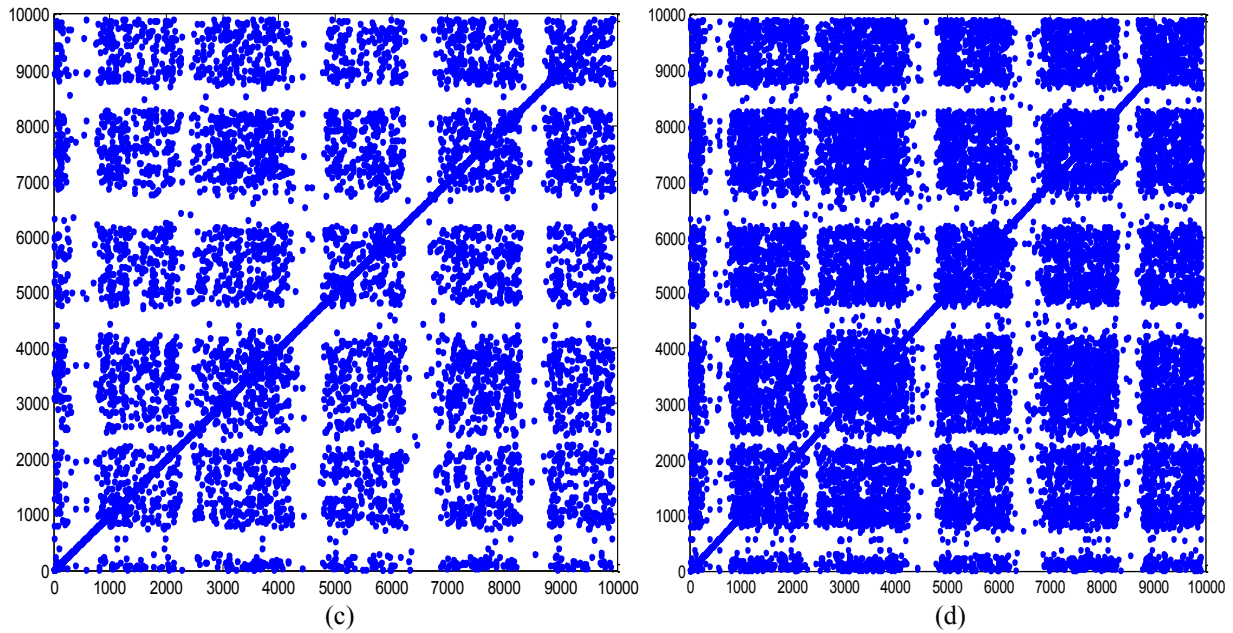


Figure IV.9. RP des composantes hautes fréquences (D_1) du signal du courant de fuite en fonction de la tension appliquée pour une conductivité de pollution de 10,1 mS/cm : (a) 5 kV_{eff}, (b) 10 kV_{eff}, (c) 15 kV_{eff} et (d) 20 kV_{eff}

La figure IV.9 montre les différents RP des composantes hautes fréquences du détail D1 (25 à 50 kHz) pour une pollution de 10,1 mS/cm.

D'après la figure IV.9.a, pour une faible tension de 5 kV_{eff} appliquée au bord de l'isolateur, nous remarquons la présence de quelques points épars, indiquant une faible activité des décharges à la surface de l'isolateur. Alors que pour la figure IV.9.b représentant le RP pour une tension appliquée de 10 kV_{eff}, nous remarquons non seulement une augmentation du nombre de points composant le RP, mais aussi un début de formation des bandes blanches. Ce dernier élément est inédit, c'est la première fois que nous remarquons une apparition si prématurée des bandes blanches pour une si faible tension appliquée. Nous expliquons ce fait par la forte pollution étudiée.

Ainsi, une telle structure indique le début de présence d'un état non stationnaire ; les transitions sont donc proches. En pratique, nous visualisons quelques décharges à la surface de l'isolateur, elles sont plus prononcées que dans les cas de classes de pollution précédentes.

Pour une tension supérieure appliquée (15 kV_{eff}), la figure IV.9.c se compose de points denses et de bandes blanches clairement établies. L'intensité des points est là aussi plus importante, et les bandes blanches sont plus imposantes. Une telle transition indique une augmentation de l'activité des décharges à la surface de l'isolateur. Ces dernières deviennent donc plus denses et plus importantes, ce qui est en conformité avec notre analyse visuelle.

Par contre, pour une tension appliquée de 20 kV_{eff}, la figure IV.9.d se situe dans la continuité de la figure qui la précède. On remarque d'une part, des points bien plus denses, plus concentrés, avec des bandes blanches demeurant inchangées. L'augmentation de la densité des points indique une activité des décharges très intense, alors que les de bandes blanches indiquent l'approche d'un état transitoire et instable. Les décharges se font donc de plus en plus intermittentes. Ainsi, pour ce niveau de tension, nous remarquons une saturation complète des

décharges sur les bandes propres. La présence des décharges partielles y est très importante. De cet état, nous pouvons prévoir une future apparition d'arcs électriques.

Ce qui différencie cette classe de pollution de celles qui la précède, est l'absence de la diminution de l'intensité des points lors de l'analyse minutieuse des RP. Ce point est très important, car il nous permet d'affirmer qu'à ce niveau de pollution, le phénomène de contournement se produit sans formation d'arcs électriques reliant les différentes décharges partielles. On notera aussi une intensité de points inédite, elle est la plus importante pour une pollution de 10,1 mS/cm. Ces conclusions sont en total accord avec l'observation visuelle. Ainsi, nous remarquons dès l'augmentation de la tension, un accroissement de l'activité des décharges à la surface de l'isolateur. Cet accroissement est suivi directement de l'arc de contournement, sans formation préalable d'arcs électriques de faible intensité reliant les décharges électriques.

IV.2.4. Discussion des résultats

L'étude de la structure topologique du RP en fonction de la classe de pollution nous a permis de tirer les conclusions suivantes.

En ce qui concerne les pollutions de 0,01 et 1,2 mS/cm, nous remarquons un RP composé principalement de points épars pour les faibles et moyennes tensions, avec une apparition des bandes blanches à partir de 20 kV_{eff}. On notera une nette diminution de l'intensité des points durant la pollution de 0,01 mS/cm (voir figure IV.7), indiquant la forte formation d'arcs électriques reliant les décharges partielles.

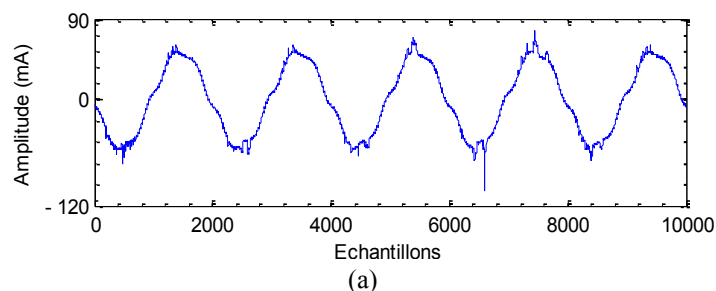
Par contre, pour la pollution de 10,1 mS/cm, nous noterons une rapide apparition des bandes blanches dès 15 kV_{eff}, et une grande intensité de points composant le RP comparé aux classes de pollution inférieures. Pour cette même classe de forte pollution, nous noterons que le phénomène de contournement s'établit sans formation préalable d'arcs électriques locaux. Un tel résultat n'aurait pas été conclu sans l'étude comparative des RP correspondant aux différentes classes de pollution.

IV.5. Application de l'analyse quantitative de récurrence (RQA) au signal de courant de fuite

On s'intéressera à appliquer l'analyse quantitative de récurrence directement au signal du courant de fuite correspondant aux différentes classes de pollution, et cela, sous une tension appliquée qui demeurera constante et égale à 20 kV_{eff}.

IV.5.1. Cas d'une conductivité de 0,01 mS/cm

Dans ce cas, nous nous intéresserons à appliquer la RQA à un signal de courant de fuite de pollution de 0,01 mS/cm.



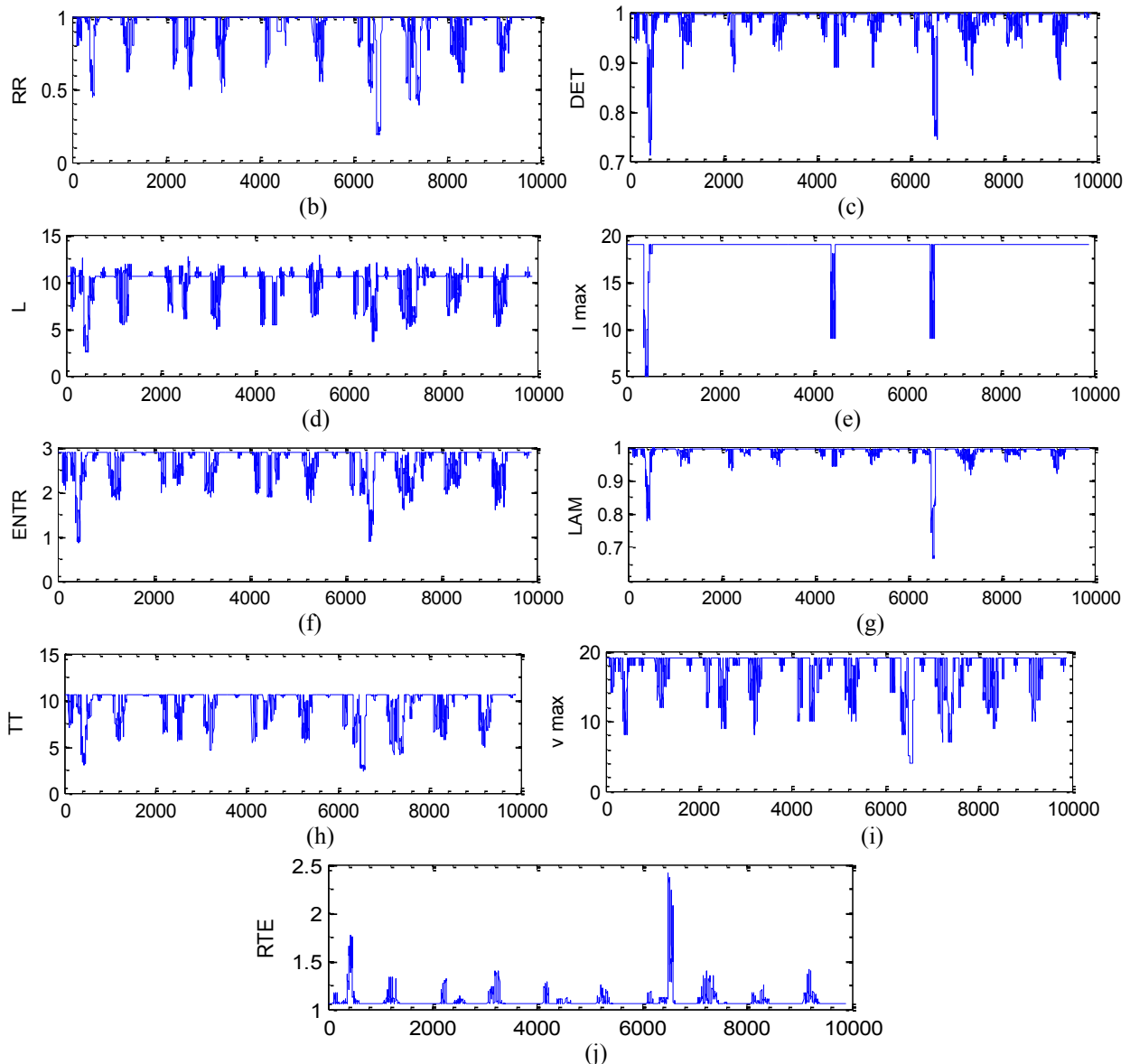


Figure IV.10. Analyse RQA pour une pollution de 0,01 mS/cm à 20 kV_{eff}
 (a) indique le signal d'origine, (b) indique RR, (c) indique DET, (d) indique L, (e) indique l_{max} , (f) indique ENTR, (g) indique LAM, (h) indique TT, (i) indique v_{max} , (j) indique RTE

Les différentes mesures calculées par la RQA sont peu variables. Elles ne montrent pas de zones à présence de larges pics. Bien au contraire, les pics sont très étroits, indiquant l'absence de zones à comportement chaotique.

Le RR, comme décrit sur la figure IV.10.b, révèle la nature du diagramme de récurrence. Il s'avère très peu variable. Il est composé de 11 pics, de largeur très étroite, ne dépassant pas une vingtaine d'échantillons. La valeur moyenne de RR est égale à 0,9374. Une telle structure indique l'absence de zones laminaires. Ainsi, nous pouvons déduire un comportement quasi-périodique du courant de fuite.

Les mêmes remarques peuvent être observées pour le DET. Il ne présente pas de zones creuses ou vides. Cela confirme l'absence du comportement chaotique du courant de fuite. Pour 99 % des données, le DET est au-dessus de 0,9, ce qui privilégie la très forte présence de structures déterministes.

L , représenté par la figure IV.10.d, exprime la durée moyenne de la similarité dans la dynamique du système. Pour ce signal, la valeur moyenne du L est de 9,93, indiquant la présence de longues lignes diagonales. La valeur maximale de ces lignes est donnée par la figure IV.10.e indiquant que l_{max} est égale à 18,78.

La valeur de ENTR est supérieure à zéro, et varie de 0,85 à 2,89. De grandes valeurs de ENTR indiquent une forte complexité du courant de fuite.

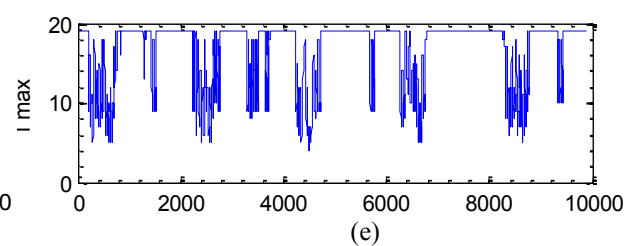
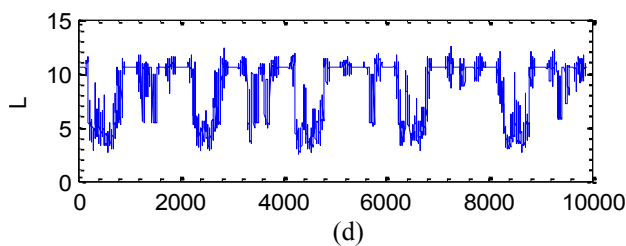
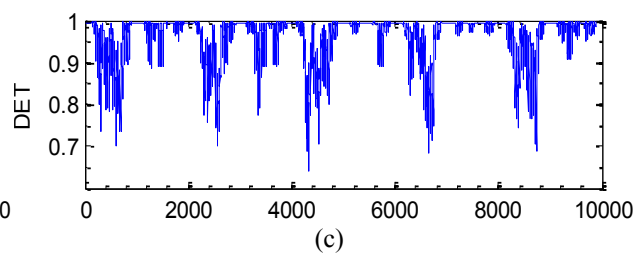
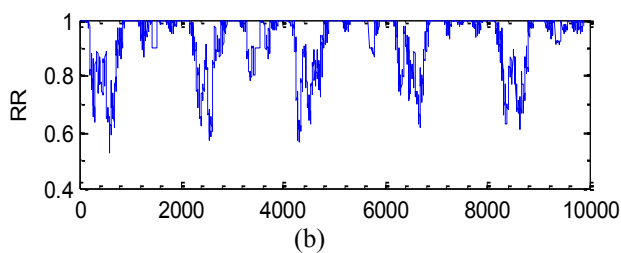
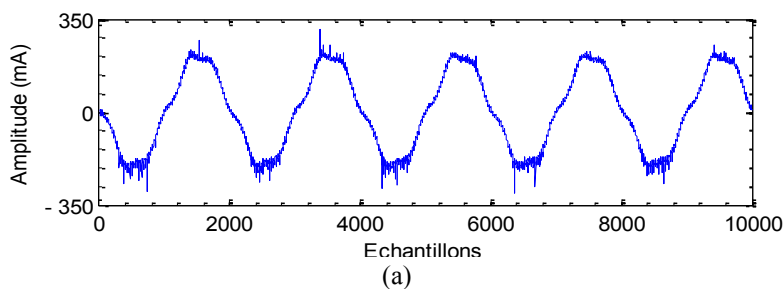
Le LAM ne présente pas lui non plus de pics distingués. 97 % des données ont un LAM supérieur à 0,9, ce qui traduit un état stationnaire du courant de fuite ainsi que la présence de nombreux états stables.

Les valeurs de TT sont peu oscillantes, ce qui indique un temps de récurrence très peu variable des points, spécifiant l'absence d'intermittences et donc de bandes blanches sur le RP. On est donc bien loin du contournement.

Le RTE, représenté par la figure IV.10.j, sert à déterminer la période d'un signal. Il varie de peu, entre 1,05 à 2,4. Seulement 14 % des données présentent un RTE supérieur à 1,1. Le signal de courant de fuite est donc bien périodique dans les cas des faibles conductivités de la pollution.

IV.5.2. Cas d'une conductivité de 1,2 mS/cm

Pour ce cas de figure, nous prendrons un signal pour une conductivité de pollution de 1,2 mS/cm.



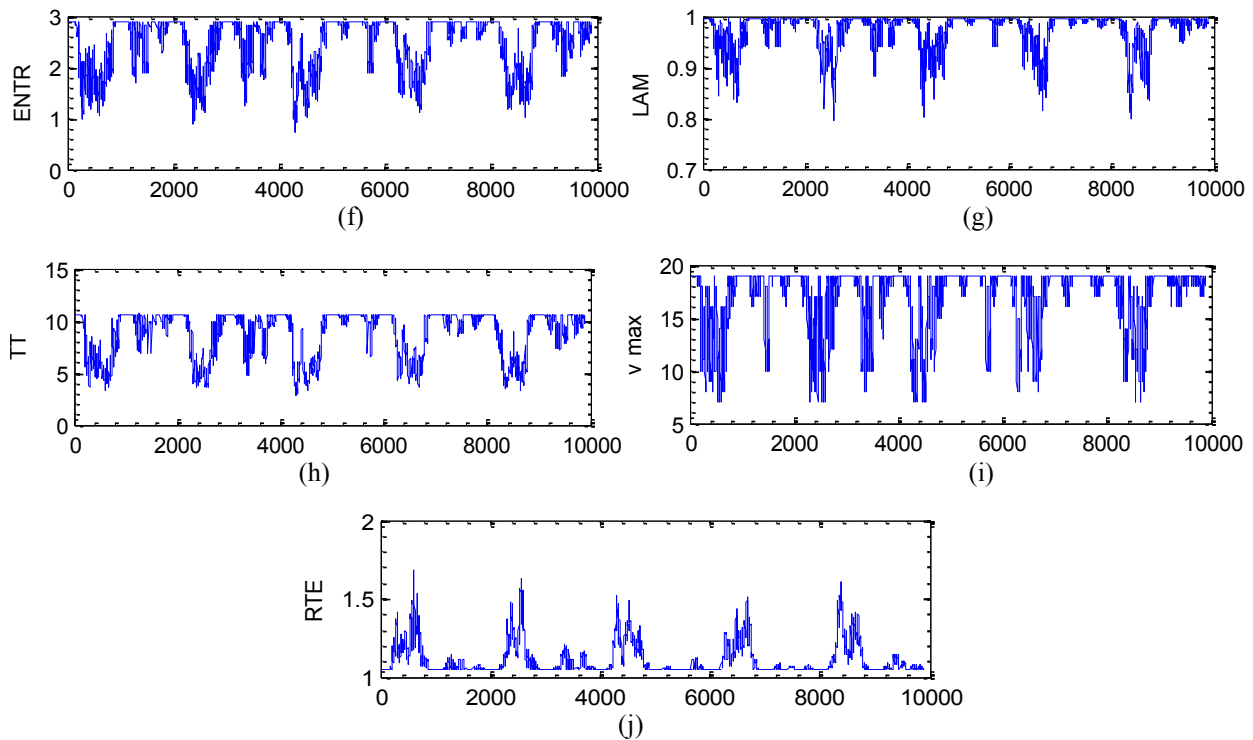


Figure IV.11. Analyse RQA pour une pollution de 1,2 mS/cm à 20 kV_{eff}
 (a) indique le signal d'origine, (b) indique RR, (c) indique DET, (d) indique L, (e) indique l_{max} ,
 (f) indique ENTR, (g) indique LAM, (h) indique TT, (i) indique v_{max} , (j) indique RTE

Dans ce cas, les différentes mesures calculées par la RQA, présentées sur la figure IV.11, sont plus variables comparativement à la figure IV.10. Elles se caractérisent par des pics plus larges, ce qui indique la présence de quelques zones à comportement chaotique du courant de fuite.

Le RR, comme décrit sur la figure IV.11.b, révèle la nature du diagramme de récurrence. Il s'avère être variable. Il est composé de 5 pics, larges de plus de 800 échantillons. La valeur moyenne de RR est égale à 0,92. Une telle structure indique la présence de zones laminaires. Ainsi, nous pouvons déduire un comportement non périodique du courant de fuite durant certains intervalles.

De même, le DET ne présente pas de pics larges, ni de grandes amplitudes. Cela confirme le comportement très irrégulier du courant de fuite. Plus que 85% des données ont un DET supérieur à 0,9, indiquant une baisse de ce dernier comparé à la classe de pollution précédente, ce qui privilégie un début d'absence de structures déterministes.

L représenté par la figure IV.11.d, possède une valeur moyenne égale à 8,55. Cette valeur est en baisse comparée à la pollution de 0,01 mS/cm. Les lignes diagonales sont donc de moins en moins longues, ce qui signale une diminution des structures déterministes. La valeur maximale de ces lignes est donnée par la figure IV.11.e indiquant que l_{max} n'est plus égal qu'à 16,7. Alors que la valeur maximale des lignes verticales, indiqué par la figure IV.11.i, est de 17,3.

La valeur de ENTR est supérieure à zéro, et varie de 0,73 à 2,89. Elle présente plus de variation comparée à la classe de pollution inférieure. Une augmentation de la variation indique une structure de moins en moins complexe durant certains intervalles du signal de courant de fuite.

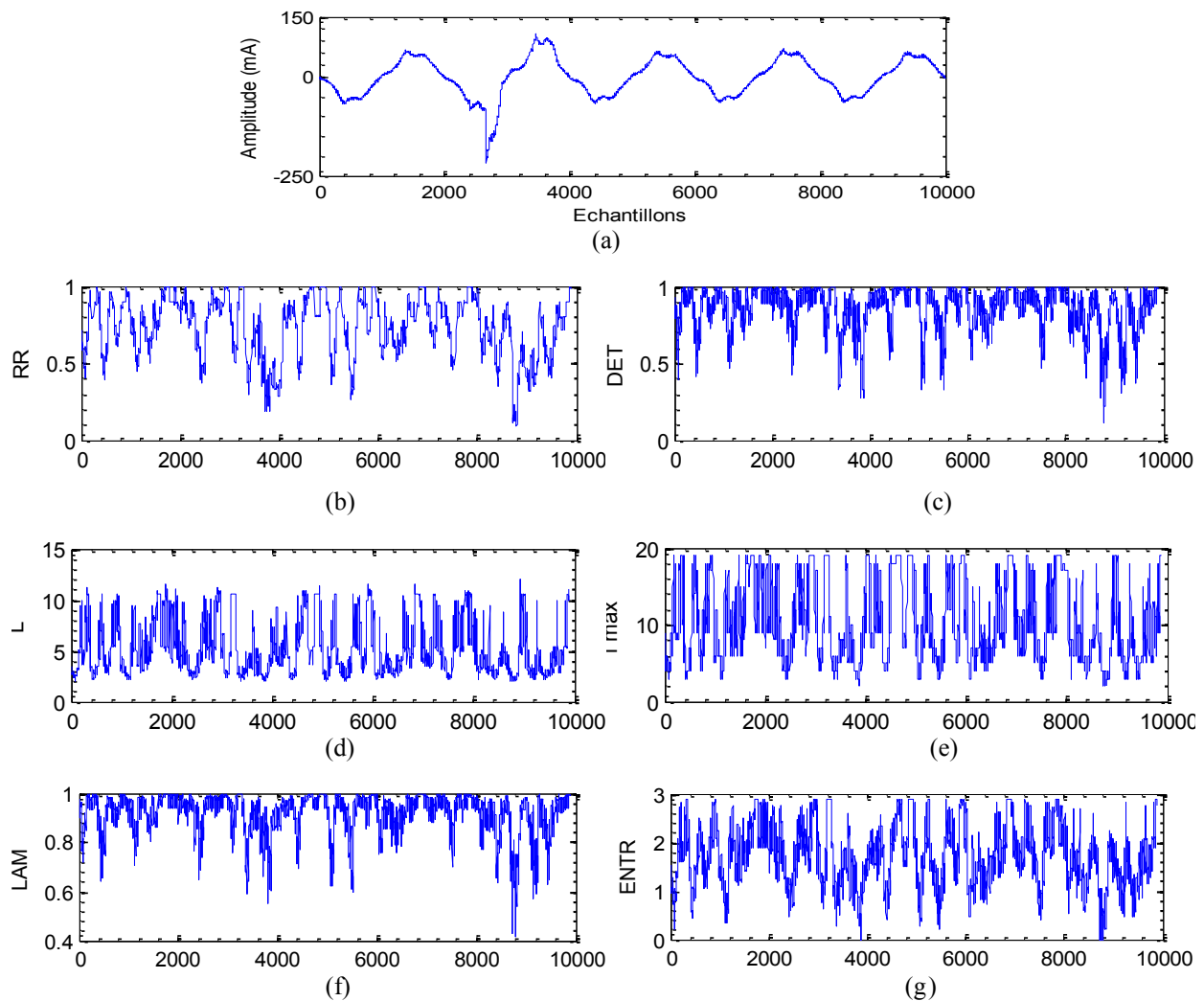
Le LAM ne présente pas lui non plus de pics distingués. 90 % des données ont un LAM supérieur à 0,9, ce qui traduit une dynamique peu stationnaire du signal de courant de fuite ainsi que la présence de nombreux états transitoires.

Les valeurs de TT sont oscillantes, ce qui indique un temps de récurrence variable des points. Sa valeur moyenne est de 8,7, il varie de 2,7 à 10,7. Une si forte variation spécifie la présence d'intermittences et donc de bandes blanches sur le RP. On est donc assez éloigné du contournement.

Le RTE, représenté par la figure IV.11.j, sert à déterminer la période d'un signal. Il varie très peu, entre 1,05 à 1,6. Plus de 40% des données présentent un RTE supérieur à 1,1. Le signal de courant de fuite est donc loin d'être périodique durant les moyennes pollutions. Il présente un comportement assez chaotique.

IV.5.3. Cas d'une conductivité de 10,1 mS/cm

Pour ce cas final, nous étudierons un signal du courant de fuite sous une conductivité de 10,1 mS/cm. Les résultats obtenus sont présentés sur la figure IV.12.



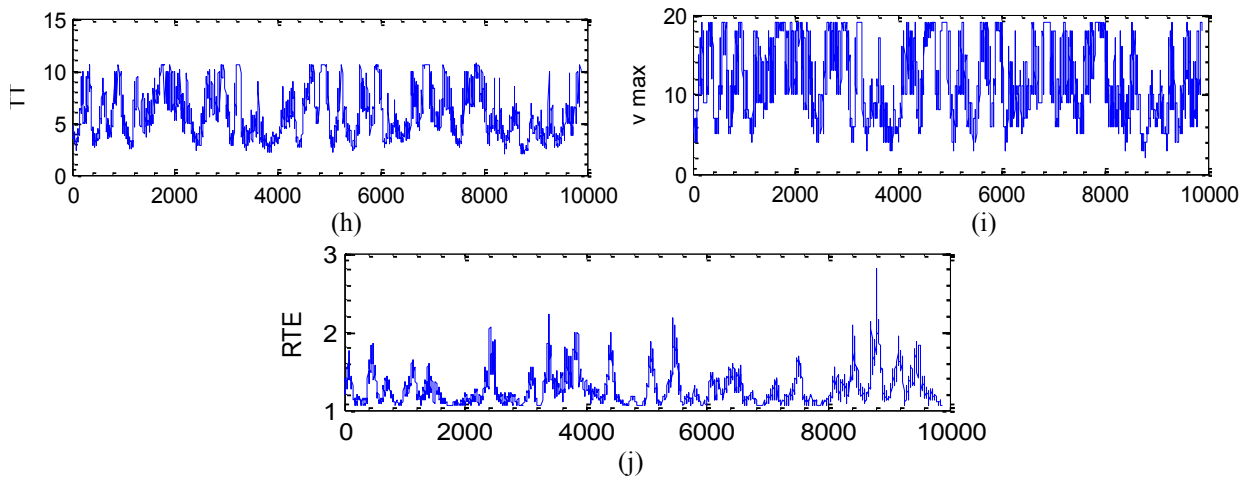


Figure IV.12. Analyse RQA pour une pollution de 10,1 mS/cm à 20 kV_{eff}

(a) indique le signal d'origine, (b) indique RR, (c) indique DET, (d) indique L, (e) indique I_{\max} , (f) indique ENTR, (g) indique LAM et (h) indique TT, (i) indique v_{\max} , (j) indique RTE

L'analyse visuelle rapide de la figure IV.12 permet de remarquer une forte variation de tous les paramètres en fonction des échantillons étudiés, ce qui prédit un comportement fort chaotique du courant de fuite.

La figure IV.12.b représentant le RR montre deux larges pics. Ils s'étendent durant les intervalles [2800-4400] et [8000-10000]. Ces intervalles correspondent au comportement chaotique du courant de fuite, car une forte variation ou une forte baisse du RR indique la présence de bandes blanches sur le RP correspondant. La valeur moyenne du RR chute jusqu'à 0,73.

De même pour le DET, des valeurs instables ou décroissantes sont trouvées dans les zones correspondant aux zones chaotiques. Le DET est au-dessus de 0,9, pour seulement 50 % des données, ce qui ne privilégie la présence de rares structures déterministes.

Le L, représenté par la figure IV.12.d oscille énormément. Sa valeur moyenne se voit chuter à 5,04, ce qui indique la présence de lignes diagonales très courtes. Cela s'interprète par l'approche d'un état instable et d'une grande probabilité de contournement.

L'ENTR s'annule sur quelques échantillons durant les zones chaotiques, ce qui n'a pas été observé durant les classes de pollutions précédentes. Une faible valeur de ENTR indique une faible complexité de la structure du signal de courant de fuite.

Le LAM se caractérise par de nombreux pics et de grandes amplitudes. Lui aussi s'annule durant certains échantillons, mais est en moyenne égal à 0,92. Plus que 70% des données ont un LAM supérieur à 0,9, ce qui indique l'existence d'un changement dans la dynamique du courant de fuite, préconisant donc un contournement proche, ainsi que l'existence de rares états stationnaires.

Les valeurs de TT sont quant à elles très oscillantes, elles varient de 2 à 10,6, avec une valeur moyenne égale à 5,7, qui demeure en baisse comparée aux classes de pollution précédentes, ce qui nous renseigne sur la période de récurrences des points. De telles valeurs signalent de forts points intermittents, ce qui indique la présence de fortes décharges sur le RP.

En ce qui concerne le RTE, ses valeurs sont très faibles. Sa valeur moyenne est de 1,2, sa valeur maximale ne dépasse pas 2,8. Une telle variation indique l'intense présence des points récurrents et donc des décharges durant cette classe de pollution.

IV.6. La relation entre RP et RQA pour l'étude de la performance d'un isolateur pollué

Dans les paragraphes 4 et 5, les analyses approfondies du diagramme de récurrence (RP) et de la RQA ont été effectuées. Cependant, ces deux méthodes, comme nous l'avons cité dans le chapitre I, ne sont pas séparées ou distinctes l'une de l'autre. Bien au contraire, la RQA décrit quantitativement les structures qui apparaissent sur le RP. De ce fait, ce paragraphe est dédié à l'étude de l'analogie entre la RP et la RQA.

La figure IV.13 montre les RP et RQA calculés pour des ondes du courant de fuite obtenues pour trois valeurs de la conductivité (0,01 ; 1,2 et 10,1 mS/cm). De prime abord, il est évident que plus nous augmentons la conductivité, plus la densité des points qui composent le RP augmentent, plus les valeurs des indicateurs de la RQA diminuent et s'approche de 0.

En augmentant la valeur de la conductivité de 0,01 à 10,1 mS/cm, les valeurs du courant de fuite augmentent. Aussi des distorsions, pics et déformations apparaissent sur le signal du courant de fuite, plus spécialement à proximité des maxima et minima de ce dernier. Ces distorsions sont traduites par des bandes blanches sur le diagramme du RP. Alors que l'augmentation des valeurs du courant de fuite implique l'augmentation de la densité des points composant le RP. Concernant les graphiques obtenus de la RQA. De manière globale, l'augmentation de la conductivité induit l'augmentation de l'intervalle de variations des indicateurs de la RQA.

Ainsi, nous observons des valeurs de plus en plus faibles des huit indicateurs étudiés. Aussi, ces derniers indicateurs montrent des pics étroits, sur de petits intervalles, pour de faibles conductivités. Ces pics et intervalles augmentent avec l'augmentation de la pollution, diminuant ainsi les valeurs des indicateurs de la RQA. Aussi, il est important de souligner que les structures apparaissant sur le RP sont retrouvées sur les graphiques des indicateurs de la RQA. En effet, les bandes blanches visibles sur le diagramme RP pour une conductivité de 10,1 mS/cm sont traduites par des zones à faibles valeurs des indicateurs de la RQA.

En conclusion, il est possible de dire que l'augmentation de la conductivité est suivie par la diminution des valeurs des indicateurs de la RQA. En effet, ces derniers s'approchent de 0, suggérant ainsi une dynamique non périodique du courant de fuite. Aussi, il est important de noter, que de telles valeurs et formes des indicateurs de la RQA résultent du fait que le signal du courant de fuite contient des pics et distorsions. Les zones auxquelles appartiennent ces derniers sont interprétées par des bandes blanches sur le diagramme RP. Par conséquent, ces bandes blanches sont dépourvues de points récurrents, de ce fait, ils sont traduits à travers la RQA, par exemple, par des valeurs très basses du RR. De même pour les restants des autres indicateurs.

Dans ce paragraphe, l'efficacité de la RQA pour l'étude de la performance d'un isolateur pollué a été démontrée. En effet, la RQA, à travers ses différents indicateurs, permet d'avoir une vision précise et quantifiée sur l'état de surface de l'isolateur en question.

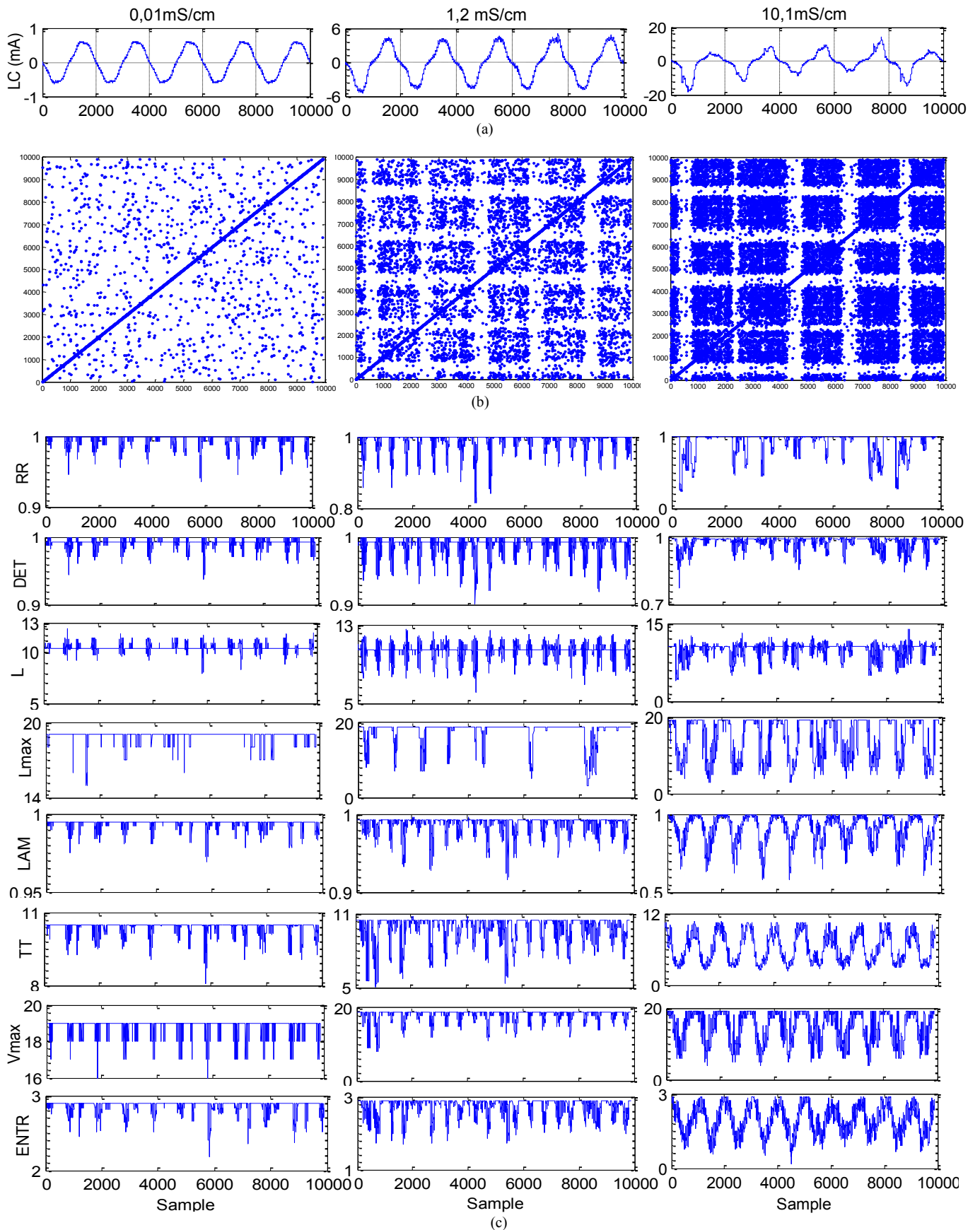


Figure IV.13. Comparatif de la RQA, (a) signal du courant de fuite pour différentes conductivités, (b) Diagramme de récurrence, (c) Indicateurs de la RQA

IV.7. Extraction des caractéristiques à partir de la RQA

Comme il a été précédemment prouvé dans la section IV.5, et d'après les figures IV.10, IV.11 et IV.12, les indicateurs de la RQA démontrent une utilité incontestable pour diagnostiquer l'état de surface d'un isolateur pollué. Dans les sections IV.5 et IV.6 précédentes, la corrélation entre les indicateurs de la RQA et le niveau de pollution a été prouvée. Cependant, étant représentés sur 10000 points, les indicateurs de la RQA sont quelque peu difficiles à interpréter. De ce fait, le présent paragraphe vise à introduire des mesures secondaires qui auront pour utilité de faciliter l'interprétation des indicateurs de la RQA.

Etant l'une des mesures statistiques les plus simples et les plus communes, la moyenne, ou communément appelée « la moyenne arithmétique », permet de quantifier un ensemble donné composé de valeurs en faisant la somme des valeurs numériques (de cet ensemble) divisée par le nombre de ces valeurs numériques. Par exemple, la moyenne de RR est calculée par l'équation suivante :

$$\text{Moyenne (RR)} = \frac{1}{N} \sum_{i=1}^N RR_i \quad (\text{IV.1})$$

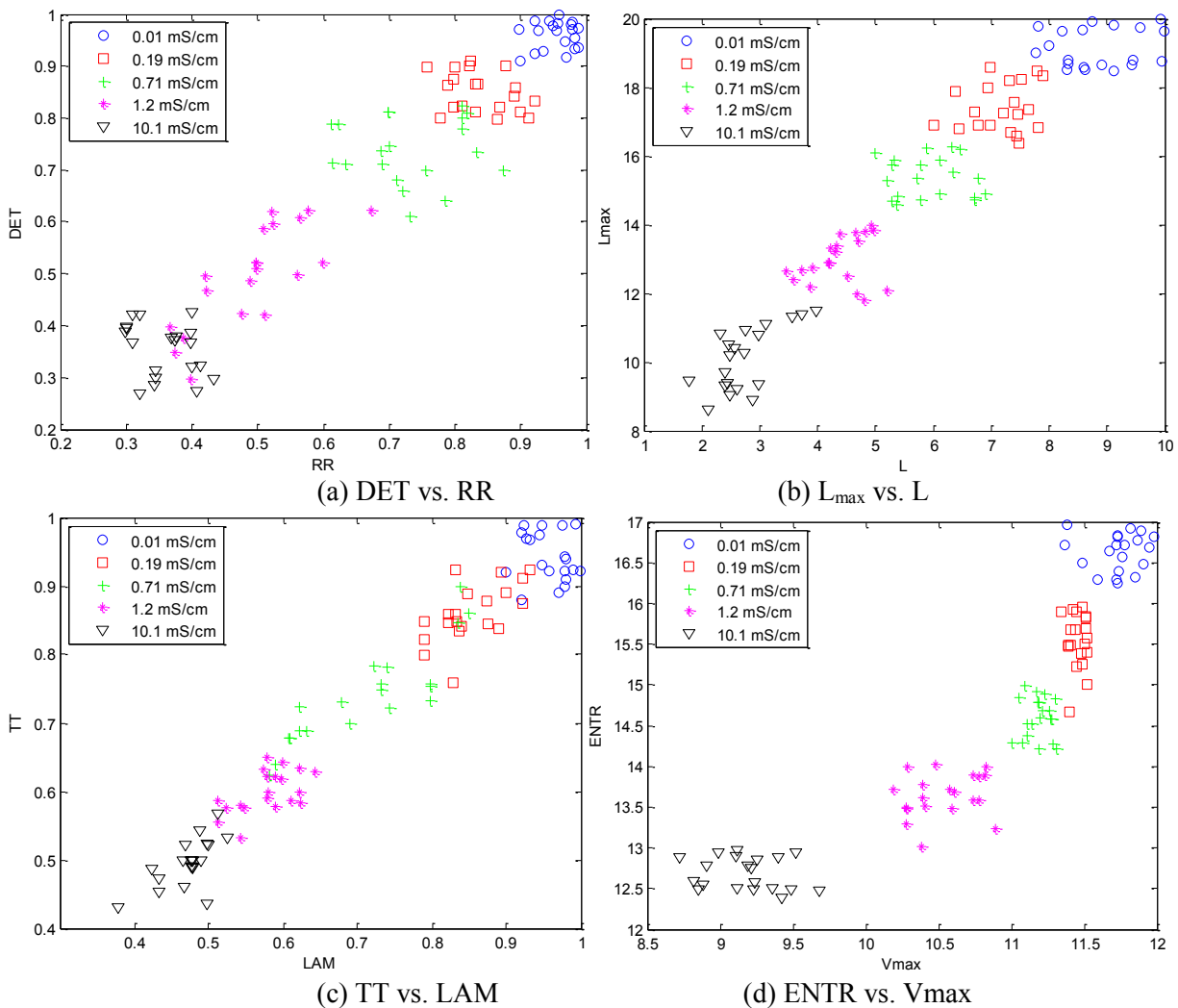


Figure IV.14. Valeurs moyennes des indicateurs de la RQA pour différents niveaux de pollution

Dès lors, nous procédons aux calculs des valeurs moyennes des indicateurs de la RQA pour la totalité des signaux de la base de données pour différents niveaux de pollution. Afin de mettre en exergue le lien entre les valeurs moyennes et la valeur de la conductivité, nous choisissons de tracer les valeurs moyennes deux à deux. Certes, pour tracer ces huit valeurs deux à deux, 28 combinaisons sont possibles. Parmi elles, nous présentons le choix montré par la figure IV.14.

Sur la figure IV.14, il apparaît que les valeurs moyennes des indicateurs de la RQA diminuent avec l'augmentation de la pollution. Ce constat confirme les observations précédemment citées dans les paragraphes 5 et 6. En effet, tant que l'augmentation de la valeur de la pollution implique la diminution des valeurs des indicateurs de la RQA, il est évident que la valeur moyenne de ces derniers diminue aussi. De plus, l'observation de la figure IV.14 permet aussi de voir une distribution assez particulière des données tracées. En effet, les valeurs moyennes tracées sont sous la forme de cinq groupes. Ces groupes correspondent aux cinq valeurs de la conductivité de pollution.

Par conséquent, le présent paragraphe, à travers la figure IV.14, confirme l'aptitude des valeurs moyennes des indicateurs de la RQA à discerner entre les différents niveaux de pollution. De ce fait, la RQA est choisie comme méthode d'extraction des caractéristiques à partir des signaux du courant de fuite.

IV.8. Résultats de la classification

Afin de tester objectivement la véracité de la méthode RQA en tant que technique d'extraction des caractéristiques à partir du courant de fuite, nous décidons d'utiliser les méthodes de classification afin de discerner entre les différentes classes de pollution. Comme nous l'avons déjà abordé dans le chapitre III, les trois méthodes de classification précédemment utilisées sont sollicitées encore une fois dans ce présent chapitre, à savoir la méthode des k plus proches voisins, la classification bayésienne et la méthode des supports vecteurs machine (SVM) [90-91].

Cependant, avant d'utiliser les méthodes de classification, bons nombres d'étapes doivent être respectées. Premièrement, l'acquisition des signaux du courant de fuite doit être effectuée. Durant les expérimentations, les 5 valeurs de la conductivité précédemment utilisées dans le chapitre III, à savoir 0,01 ; 0,19 ; 0,71 ; 1,2 et 10,1 mS/cm, sont gardées les mêmes.

En second lieu, la reconstruction de l'espace de phase est accomplie en tenant compte des paramètres m , du retard τ et du seuil ϵ , c'est à dire: 5, 20 et $0,25\sigma$, où σ est la standard déviation du signal étudié, comme citée dans le paragraphe 2.3 de ce présent chapitre. Troisièmement, la RQA et ses différents indicateurs sont calculés. Puis, l'extraction des caractéristiques à partir des indicateurs de la RQA est procédée par le calcul de la valeur moyenne de ces derniers. Enfin, ces dernières valeurs moyennes sont utilisées comme entrées aux trois méthodes de classification afin de discerner entre les 5 classes de conductivité de pollution utilisées.

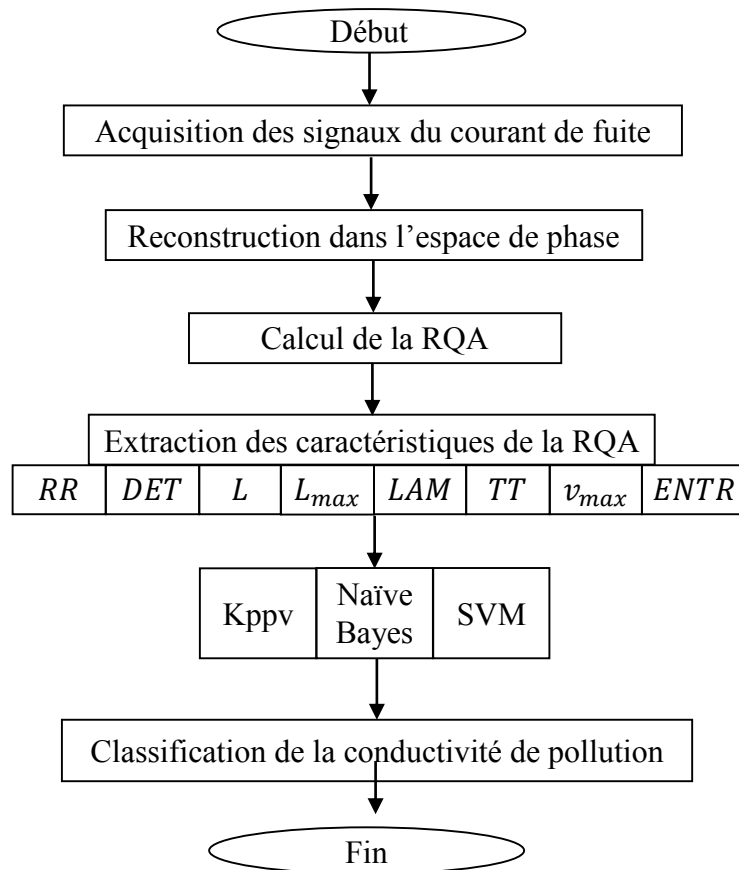


Figure IV.15. Diagramme de l'algorithme de classification de la sévérité de pollution du courant de fuite base sur la RQA

En pratique et afin de mettre en œuvre la classification de la pollution, cinq labels sont assignés aux cinq valeurs de la conductivité de pollution de la plus légère à la plus grande, de tel sorte que la class 1 correspond à 0,01 mS/cm, alors que la class 5 correspond à 10,1 mS/cm, comme décrit sur le tableau IV.1. Par ailleurs, il est utile de rappeler que la base de données composée de 100 signaux du courant de fuite, précédemment schématisée par la figure IV.14, est utilisée cette fois-ci pour la classification. 90% de ces signaux sont dédiés à l'entraînement des méthodes de classification, alors que le reste des signaux sont pour le test.

Tableau IV.1. Résultats de la classification

Echantillon	RR	DET	L	L _{max}	LAM	TT	V _{max}	ENTR	Conductivité de pollution (mS/cm)	Classe assignée	Classe obtenue		
											Kppv	Naïve Bayes	SVM
1	0,877	0,899	9,410	17,55	1,740	0,743	9,407	16,428	10,1	5	5	5	5
2	0,749	0,942	3,799	11,28	0,631	0,665	9,281	17,128	10,1	5	5	5	5
3	0,89	0,989	9,974	18,40	2,221	0,897	10,73	18,634	1,2	4	4	4	4
4	0,875	0,995	9,937	17,38	3,933	0,898	9,676	16,412	1,2	4	4	3*	4
5	0,954	0,992	4,724	12,48	1,766	0,900	10,046	17,033	0,710	3	3	3	3
6	0,718	0,853	2,344	3,852	0,636	0,540	10,546	14,389	0,710	3	3	4*	3
7	0,954	0,992	5,219	12,98	2,061	0,900	10,341	17,228	0,190	2	2	2	2
8	0,718	0,853	2,839	4,347	0,931	0,541	11,249	14,578	0,190	2	2	3*	2
9	0,728	0,863	3,544	4,852	1,436	0,552	11,146	14,783	0,01	1	1	1	1
10	0,820	0,914	11,59	19,79	3,580	0,945	10,757	17,368	0,01	1	1	2*	1

Comme montré par le tableau IV.1 ci-dessous, les résultats de la classification sont réussis pour la totalité des échantillons de l'ensemble de test (10 échantillons) dans le cas des deux algorithmes SVM et kppv. Cependant, pour la méthode de classification bayésienne, cette dernière offre le plus mauvais des résultats, avec 4 échantillons sur 10 faussement classés. Sur ce point, il est possible de noter que les résultats de classification obtenus étaient quelque peu prédictibles. En effet, la SVM est connue pour avoir la capacité de traiter les problèmes linéaire et non linéaire. Aussi, la kppv, malgré sa simplicité d'emploi, permet aussi d'avoir d'aussi bons résultats. Cependant, la classification bayésienne démontre une faible aptitude à traiter avec succès une telle base de données.

Les résultats de classification obtenus à travers ce chapitre démontrent l'importance primordiale de choisir tant la méthode d'extraction des caractéristiques, que la méthode de classification. En effet, afin de concevoir un modèle de classification robuste, il est nécessaire que les méthodes choisies soit performantes mais aussi en bonne adéquations entre elles.

IV.9. Conclusion

Dans le but de diagnostiquer l'état de surface et la performance d'un isolateur sous pollution, l'analyse approfondie du signal du courant de fuite, à travers la méthode de la RQA a été établie. Démontrant une efficacité indubitable pour l'étude de la dynamique du courant de fuite, huit indicateurs de la RQA ont été examinés à travers différents niveaux de pollution. Ces indicateurs caractérisent très bien la dynamique de variation du courant de fuite. En effet, plus la valeur de la conductivité de la solution polluante augmente, plus les valeurs des indicateurs de la RQA diminuent. Pour quantifier cette diminution, nous optons pour le calcul des valeurs moyennes de ces indicateurs. Cette diminution est mise en évidence par les figures montrant les valeurs moyennes de ces derniers tracées deux à deux. Sur ces mêmes figures, il est aussi possible de voir l'apparition de cinq groupes, correspondant aux cinq valeurs de la conductivité. De ce fait, les valeurs moyennes des indicateurs de la RQA sont utilisées comme entrées à trois différentes méthodes de classification. Les résultats obtenus annoncent la combinaison de la RQA, comme méthode d'extraction des caractéristiques, ainsi que de la SVM ou kppv, comme méthodes de classification, comme choix optimal assurant de bonnes performances de classification.

CONCLUSION GENERALE



Tout au long de ce travail, de nouvelles méthodes, permettant de diagnostiquer avec fiabilité l'état de surface des isolateurs pollués, ont été mises en avant. Ces méthodes peuvent être développées sous diverses méthodologies. En effet, durant cette thèse, nous avons mis en application, non seulement, le traitement d'image, mais aussi le traitement de signal, pour la caractérisation de la performance d'un isolateur sous conditions de pollution.

Dans un premier lieu, à travers cette thèse, nous avons exposé les fondements du traitement d'image. Puis, sur la base de ces derniers, nous avons exposé divers outils qui peuvent être combinés afin de constituer un algorithme efficace pour le diagnostic de l'état de surface d'un isolateur pollué. Sur une surface plane polluée, l'activité des décharges électriques est enregistrée à l'aide d'une caméra numérique. Suivant une disposition uniforme de la pollution, l'acquisition des vidéos d'essais est faite pour diverses valeurs de la conductivité de la solution polluante (0,01 ; 0,19 ; 0,71 ; 1,2 et 10,1 mS/cm). De prime abord, l'observation de ces vidéos d'essais permet de déceler 8 états des décharges électriques sur la surface de l'isolateur. Parmi ces 8 états, 4 incluent la formation d'arcs électriques. De ce fait, nous avons élaboré un algorithme en vue de déceler l'apparition de ces arcs.

Afin d'effectuer la reconnaissance des arcs électriques sur la surface isolante, nous avons proposé la méthodologie de traitement suivante :

1. Premièrement, la méthode de segmentation d'Otsu est appliquée. Elle permet de segmenter l'image traitée, en effectuant l'extraction de la décharge électrique de son background (fond d'image) ;
2. Deuxièmement, compte tenu du bruit que contiennent les images segmentées par la méthode d'Otsu, nous proposons l'utilisation du filtrage morphologique. Ce filtrage, combine les opérations d'érosion et de dilatation pour éliminer les pixels non représentatifs des décharges électriques.
 - a. En effet, l'opération d'érosion permet de supprimer les pixels dus aussi bien à l'éclairage du laboratoire d'essai qu'à la lumière du jour qui sont réfléchis sur le bord supérieur de l'isolateur.
 - b. Cependant, l'opération d'érosion réduit les informations présentes sur l'image. Dès lors, nous utilisons l'opération de dilatation pour recouvrir et retrouver les pixels appartenant à la décharge électrique.
3. En troisième lieu, une fois que le filtrage morphologique, qui permet d'obtenir une vision correcte des décharges, est effectué, nous proposons le calcul de quatre indicateurs qui décrivent et quantifient l'activité des décharges électriques sur la surface isolante polluée. Ces quatre indicateurs (N_l , N_p , L et W) nous permettent de conclure que :
 - a. Les fluctuations de N_l et N_p précédant le contournement permettent de mettre en avant le phénomène d'intermittence des décharges électriques. Ce dernier correspondant à une rapide apparition suivie d'une disparition des décharges sur la surface isolante, est d'autant plus accentué lors des dernières étapes menant au contournement.
 - b. N_l , quantifiant le nombre de décharges électriques sur la surface isolante polluée, augmente avec l'augmentation de la tension appliquée et diminue juste avant le

- contournement ($1 \text{ kV}_{\text{eff}}$ avant la tension de contournement). Une telle diminution de N_l indique l'extinction des petites décharges, qui donnent naissance à l'arc de contournement final.
- c. N_p étant le nombre de pixels de la plus grande décharge, permet de savoir si les arcs sont présents. En effet, une fluctuation de N_p est synonyme d'activité d'arcs électriques, alors qu'une monotonie dénote l'absence des arcs.
 - d. L et W permettent de quantifier l'espace pris par les arcs électriques. En effet, lors de faibles niveaux de tension appliqués, les décharges sont dispersées sur des surfaces plus ou moins larges. Cependant, la diminution de W est synonyme de formation d'arcs électriques, car ces derniers occupent des surfaces moins larges que celles des décharges éparses.
 - e. Les quatre indicateurs permettent de prévenir le phénomène de contournement des isolateurs. En effet, N_p , L et W décrivent la décharge la plus proéminente, ils augmentent donc à l'approche du contournement. Cependant, N_l décompte le nombre de décharges à la surface de l'isolateur, il diminue donc avant le contournement.
4. Finalement, compte tenu de la corrélation de ces quatre indicateurs avec le niveau de pollution, nous proposons de les utiliser comme entrées pour différentes méthodes de classification (Kppv, Naïve Bayes and SVM) en vue de discerner le cas de présence de décharge de type arc. Différents taux de réussite sont obtenus, la SVM devance la Kppv et la classification bayésienne, avec un taux de réussite égale à 95,57 %.

Dans un second lieu, nous mettons en œuvre le diagnostic de l'état de surface et la performance d'un isolateur sous pollution à travers la méthode de la RQA appliquée sur le signal du courant de fuite. Huit indicateurs de la RQA ont été examinés à travers différents niveaux de pollution. L'analyse de ces indicateurs permet de conclure que :

1. L'augmentation de la valeur de la conductivité de la couche polluante est suivie par la diminution des valeurs des indicateurs de la RQA ;
2. Pour quantifier cette diminution, nous optons pour le calcul des valeurs moyennes de ces indicateurs. Ces derniers sont tracés deux à deux et permettent de faire apparaître cinq groupes, correspondant aux cinq valeurs de la conductivité ;
3. Dès l'utilisation des valeurs moyennes des indicateurs de la RQA comme entrées aux trois différentes méthodes de classification (Kppv, Naïve Bayes and SVM), les résultats obtenus annoncent la combinaison de la RQA comme excellente méthode d'extraction des caractéristiques, ainsi que de la SVM ou kppv, comme meilleures méthodes de classification.

Comme perspective, il serait très intéressant dans le futur d'utiliser les méthodologies, développées tout le long de ce travail sur un modèle d'isolateur plan, pour l'étude et le diagnostic des isolateurs réels en exploitation.

REFERENCES
BIBLIOGRAPHIQUES



V. Références bibliographiques

- [1] Kumagai and N. Yoshimura, "Leakage Current Characterization for Estimating the Conditions of Ceramic and Polymeric Insulating Surfaces", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 11, pp. 681-690, 2004.
- [2] G. K. Amaldo and F. B. Geraldo, "Leakage Current Monitoring of Insulators Exposed to Marine and Industrial Pollution", IEEE Internal Symposium on Electrical Insulation (ISEI), Montreal, Canada, pp. 271-274, 1996.
- [3] M. Tegar, A. Mekhaldi and A. Boubakeur, "Algorithm for HV Insulators Flashover under Discontinuous pollution", Archives of Electr. Eng., Vol. LI, No.2, pp 137-146, 2002.
- [4] T. Suda, "Frequency Characteristics of Leakage Current Waveforms of an Artificially Polluted Suspension Insulator", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 8, pp. 705-710, 2001.
- [5] Y.C. Song and D.H. Choi, "High-frequency Components of Leakage Current as Diagnostic Tool to Study Ageing of Polymer Insulators under Salt Fog", Electronics Letters, No.41, Vol. 12, pp. 684-685, 2005.
- [6] J. Li, C. Sun, W. Sima, Q. Yang and J. Hu, "Contamination Level Prediction of Insulators Based on the Characteristics of Leakage Current", IEEE Transactions On Power Delivery, Vol. 25, No. 1, pp. 417-424, January 2010.
- [7] M. A. Douar, M. C. Bouzidi et A. Mekhaldi, *Caractérisation de l'état de surface d'un isolateur pollué à l'aide des signaux du courant de fuite et de la tension appliquée*, Mémoire de projet de fin d'études, Ecole Nationale Polytechnique, Alger, Juin 2009.
- [8] M. A. Douar, A. Mekhaldi and M. C. Bouzidi, "Flashover Process and Frequency Analysis of the Leakage Current on Insulator Model under non-Uniform Pollution Conditions", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 17, No. 4, pp. 1284-1297, 2010.
- [9] M. A. Douar, A. Mekhaldi and M. C. Bouzidi, "Frequency analysis of the leakage current under non uniform polluted conditions on one insulator plane model", IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), pp. 1-4, West Lafayette, IN, 17-20 October, 2010.
- [10] M. A. Douar, A. Mekhaldi and M. C. Bouzidi, "Time-scale analysis of the applied voltage waveform in uniform pollution conditions as a new monitoring technique of the insulator flashover", IEEE International Conference on Solid Dielectrics (ICSD), pp. 1-4, Potsdam, 4-9 July, 2010.
- [11] M. A. Douar, A. Mekhaldi and M. C. Bouzidi, "Welch periodogram analysis of the leakage current on insulator model under wetted contaminated conditions for flashover prediction", IEEE Mediterranean Electrotechnical Conf. (MELECON), pp. 1636-1641, Valletta, 26-28 April 2010.
- [12] B. Moula, A. Mekhaldi, M. Tegar and A. Haddad, "Characterization of Discharges on non-Uniformly Polluted Glass Surfaces Using a Wavelet Transform Approach", Transactions on Dielectrics and Electrical Insulation. Vol. 20, No. 4, pp. 1457-1466, 2013.
- [13] **A.K. Chaou, A. Mekhaldi and M. Tegar, "Classification of Pollution Severity on Insulator Model using Recurrence Quantification Analysis", T&D Conference and**

Exposition, 2014 IEEE Power and Energy Society (PES) Transmission & Distribution Conference & Exposition 2014, Chicago, DOI: 10.1109/TDC.2014.6863188, pp. 1-5, April 14-17, 2014.

[14] A. K. Chaou, A. Mekhaldi and M. Teguar, "Recurrence Quantification Analysis as a novel feature extraction technique for the classification of pollution severity on High Voltage Insulators", International Conference on Electrical Engineering and Automatic Control, Setif, November 24-26, 2013.

[15] A. K. Chaou, A. Mekhaldi and M. Teguar, "Recurrent Plot for Leakage Current Investigation and Performance Monitoring of HV Insulator Model under Discontinuous Pollution", 2014 International Conference on High Voltage Engineering and Application, 8-11 September 2014, Poznan, Poland, DOI: 10.1109/ICHVE.2014.7035444, pp. 1-4, September 8-11, 2014.

[16] A. K. Chaou, A. Mekhaldi and M. Teguar, "Elaboration of novel image processing algorithm for arcing discharges recognition on HV polluted insulator model", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 22, No. 2, DOI: 10.1109/TDEI.2015.7076800, pp. 990-999, April 2015.

[17] A.K. Chaou, A. Mekhaldi, B. Moula, M. Teguar, 'The use of wavelets for the monitoring and diagnostic of surface state of HV polluted insulators', International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM), Tunis, DOI: 10.1109/CISTEM.2014.7076974, pp. 1-8, November 3-6, 2014.

[18] A. K. Chaou, A. Mekhaldi, M. Teguar and I. Fofana and F. Meghnefi, "Wavelet Packet Transform based Multi Resolution Analysis Technique for Classification of LC Waveforms on Polluted Insulating Surfaces", IEEE Conference on Electrical Insulation and Dielectric Phenomena, Des Moines, Iowa, USA, DOI:10.1109/CEIDP.2014.6995845, pp. 291-294, October 19-22, 2014.

[19] A. K. Chaou, A. Mekhaldi and M. Teguar, "Classification of Leakage Current Waveforms using Wavelet Packet Transform on High Voltage Insulator", 2014 International Conference on High Voltage Engineering and Application, Poznan, Poland, DOI: 10.1109/ICHVE.2014.7035489, pp. 1-4, September 8-11, 2014.

[20] I. Ramirez-Vazquez and J. L. Fierro-Chavez, "Criteria for the Diagnostic of Polluted Ceramic Insulators Based on the Leakage Current Monitoring Technique", IEEE Conf. Electr. Insul. Dielectr. Phenomena (CEIDP), Austin, USA, pp. 715-718, 1999.

[21] B. X. Du, Y. Liu and H. J. Liu, "Recurrent plot analysis of leakage current for monitoring outdoor insulator performance", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 16, No. 1, pp.139-146, February 2009.

[22] Y. Liu and B. X. Du, "Recurrent plot analysis of leakage current on flashover performance of rime-iced composite insulator", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 17, No. 2, pp. 465-472, April 2010.

[23] Y. Liu and B. X. Du, "Recurrent Plot Analysis of Leakage Current in Dynamic Drop Test for Hydrophobicity Evaluation of Silicone Rubber Insulator", IEEE Transactions on Power Delivery, Vol. 28, No. 4, pp. 1996-2003, October 2013.

[24] J. Li, W. Sima, C. Sun and S.A. Sebo, "Use of Leakage Currents of Insulators to Determine the Stage Characteristics of the Flashover Process and Contamination Level

- Prediction”, IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 17, No. 2, pp. 490-501, 2010.
- [25] X. Jiang X., Y. Shi, C. Sun and Z. Zhang, “Evaluating the safety condition of porcelain insulators by the time and frequency characteristics of LC based on artificial pollution tests”, IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 17, No. 2, pp. 481-489, 2010.
- [26] D. Pylarinos, K. Theofilatos, K. Siderakis, E. Thalassinakis, I. Vitellas, A.T. Alexandridis and E. Pyrgioti, “Investigation and Classification of Field Leakage Current Waveforms”, IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 19, No. 6, pp. 2111- 2118, 2012.
- [27] D. Pylarinos, S. Lazarou, G. Marmidis and E. Pyrgioti, “Classification of Surface Condition of Polymer Coated Insulators using Wavelet Transform and Neural Networks”, International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '07, pp. 658-663, Beijing, China, 2-4 November, 2007.
- [28] S. Mallat, *A wavelet tour of signal processing*, Elsevier, Academic Press, San Diego, CA, ISBN 13. 978-0-12-374370-1, 1999.
- [29] E.N. Lorenz, “Deterministic non periodic flow”, Journal of Atmospheric Sciences, Vol. 20, pp.130-141, 1968.
- [30] E.N. Lorenz, “The statistical prediction of solutions of dynamic equations”, Proc. Internat. Sympos. Numerical Weather Prediction, pp. 629-635, Tokyo 7-13 November, 1960.
- [31] F. Takens, “Detecting strange attractors in turbulence”, In Lecture notes in mathematics, Dynamical systems and turbulence, Vol. 898, page 366, Berlin, 1981.
- [32] N. Packard , J. Crutchfield, J. Farmer, and R. Shaw, “Geometry from a time series”. Phys. Rev. Lett., pp. 45-712, 1980.
- [33] J.P. Huke, “Embedding non-linear dynamical systems : a guide to Takens theorem”, Manchester Institute of Mathematical Sciences, ISSN 1749-9097, 2006.
- [34] H. S. Kim, R. Eykholt, and J.D. Salas, “Nonlinear dynamics, delay times, and embedding windows”, Physica D, Vol. 127, pp. 48–60, 1999.
- [35] H.G. Ma, X. H. Li and G.H. Wang, “Selection of embedding dimension and delay time in phase space reconstruction”, J. Xi’an Jiao tong University, Vol. 38, pp. 335–338, 2004.
- [36] L. Tang and J. Liang, “C-C Method to Phase Space Reconstruction Based on Multivariate Time Series”, 2nd International Conference on Intelligent Control and Information Processing (ICICIP), pp. 438 – 441, Harbin, 25-28 July, 2011.
- [37] H. Poincaré, "Sur le problème des trois corps et les équations de la dynamique", Acta Mathematica, Vol. 13, pp. 1–271, 1890.
- [38] J. P. Eckmann, S.O. Kamphorst and D. Ruelle, “Recurrence plots of dynamical systems”, Europhysics Le [29tters 4, Vol. 9, pp. 973-977, 1987.
- [39] J. Belaire-Franch, D. Contreras, L. Tordera-Lledó, “Assessing nonlinear structures in real exchange rates using recurrence plot strategies”, Physica D: Nonlinear Phenomena, No. 171, Vol. 4, pp. 249-264. 2002
- [40] N. Marwan, M. C. Romano and M. Thiel, “Recurrence Plots for the Analysis of Complex Systems”, Physics Reports, Vol. 438, pp.237-329, 2007.

- [41] A. Facchini, C. Mocenni, N. Marwan, A. Vicino and E. Tiezzi, “Nonlinear time series analysis of dissolved oxygen in the Orbetello Lagoon”, *Ecological Modelling*, Vol. 3-4, pp. 339-348. 2007.
- [42] Jr. C. L. Webber and J. P. Zbilut, “Recurrence quantification analysis of nonlinear dynamical systems”, *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pp. 26-94. 2005.
- [43] R. Nock and F. Nielsen, “Statistical Region Merging”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 26, No 11, pp 1452-1458, 2004
- [44] T. Lindeberg and M. Li, M.-X, “Segmentation and classification of edges using minimum description length approximation and complementary junction cues”, *Computer Vision and Image Understanding*, Vol. 1, No.67, pp. 88–98. 1997.
- [45] S. Oz and Y.S. Dogrusoz, “Fast 3D reconstruction from medical image series based on thresholding method”, 15th National Biomedical Engineering Meeting (BIYOMUT), pp.1-4, Antalya, 21-24 April 2010.
- [46] G.M.N.R. Gajanayake, R.D. Yapa and B. Hewawithana, “Comparison of standard image segmentation methods for segmentation of brain tumors from 2D MR images”, *International Conference on Industrial and Information Systems (ICIIS)*, pp.301-305, Sri Lanka, 28-31 December, 2009.
- [47] R. Moussa, M. Beurton-Aimar and P.Desbarats, “Multi-agent segmentation for 3D medical images”, 9th International Conference on Information Technology and Applications in Biomedicine (ITAB), pp. 1-5, 2009.
- [48] M.K. Osman, M.Y. Mashor and H. Jaafar, “Performance comparison of clustering and thresholding algorithms for tuberculosis bacilli segmentation”, *International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1-5, Amman, 14-16 May, 2012.
- [49] H. Ting-lei and X. Bai, “An Improved Algorithm for Medical Image Segmentation”, 2nd International Conference on Genetic and Evolutionary Computing, WGEC '08, pp.289-292, 2008.
- [50] R. Helen, N. Kamaraj, K. SelviandV. Raja Raman, “Included in Your Digital Subscription Segmentation of pulmonary parenchyma in CT lung images based on 2D Otsu optimized by PSO”, *International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT)*, pp.536–541, Tamil Nadu, 23-24 March, 2011.
- [51] D. Hang and L. Xi, “The Color Components' Exchanging on Different Color Spaces for Image Segmentation of Hematocyte”, *International Conference on Multimedia Information Networking and Security (MINES)*, pp. 10-13, Nanjing, Jiangsu, 4-6 Nov, 2010.
- [52] Jean Serra, “Introduction to mathematical morphology”, *Computer Vision, Graphics, and Image Processing*, Science Direct, Vol. 35, No. 3, pp. 283–305, September 1986.
- [53] Jean Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Orlando, FL, USA, ISBN. 0126372403, 1983.
- [54] H. Cao and Y. Shen, “Application of MATLAB Image Processing Technology in Sewage Monitoring System”, *The Ninth International Conference on Electronic Measurement & Instruments ICEMI'2009*, pp. 3-993 - 3-995, Beijing, 16-19 August, 2009.

- [55] S. Pierre, P. Martino and O. Georgios, “Mathematical morphology and its applications to image processing”, 10th International Symposium, ISMM 2011, Verbania-Intra, Italy, pp 132-142, July 6-8, 2011.
- [56] C. Hendriks, G. Borgefors and R. Strand, “Mathematical Morphology and Its Applications to Signal and Image Processing”, 11th International Symposium, ISMM 2013, Uppsala, Sweden, May 27-29, 2013.
- [57] Image Processing Toolbox™ User's Guide, Understanding Connected-Component Labeling, The MathWorks, Inc, pp. (10-35)-(10-41), 2015.
- [58] Sébastien Parfait, *Classification de spectres et recherche de biomarqueurs en spectroscopie par résonance magnétique nucléaire du proton dans les tumeurs prostatiques*, thèse de doctorat, Instrumentation et Informatique de l'Image, Université de Bourgogne, pp. 117-127, 2010.
- [59] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, ISBN. 0-387-94559-8, 2000.
- [60] Aizerman A., Braverman E.M. and Rozner L.I., “Theoretical foundations of the potential function method in pattern recognition learning”, *Automation and Remote Control*, Vol. 25, p. 821-837, 1964.
- [61] Jonathan Milgram, *Contribution à l'intégration des machines à vecteurs de support au sein des systèmes de reconnaissance de formes: application à la lecture automatique de l'écriture manuscrite*, thèse de doctorat, Ecole de Technologie Supérieure, Université du Québec, pp. 28-56, 2007.
- [62] C. Cortes and V. N. Vapnik. Support-vector networks. *Machine Learning*, No. 20, Vol. 3, pp. 273-297, 1995.
- [63] E. Mayoraz and E. Alpaydin. “Support vector machines for multi-class classification”. *International Work-Conference on Artificial Neural Network, IWANN*, pp. 833-842, 1999.
- [64] J. H. Hong, J. K. Min, U. K. Cho and S. B. Cho, “Fingerprint classification using one-vs-all support vector machines dynamically ordered with naïve Bayes classifiers”, *Pattern Recognition*, Vol. 41, pp. 662 – 671, 2008.
- [65] M. Moreira and E. Mayoraz. Improved pairwise coupling classification with correcting classifiers. *European Conference on Machine Learning*, Chemnitz, Germany, pp. 160-171, April 21-24, 1998.
- [66] C. W. Hsu and C. J. Lin. A comparison of methods for multi-class support vector machines. *IEEE transactions on Neural Networks*, Vol. 13, No. 2, pp. 415-425, 2002.
- [67] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large Margin DAGs for Multiclass Classification. *Advances in Neural Information Processing Systems*, pp. 547-553, 2000.
- [68] Gilles Lebrun et Hubert Cardot, *Sélection de modèles pour la classification supervisée avec des SVM (Séparateurs à Vaste Marge) : application en traitement et analyse d'images*, thèse de doctorat, Université de Caen/Basse-Normandie, 2006.
- [69] J. Platt, “Fast Training of Support Vector Machines using Sequential Minimal Optimization, *Advances in Kernel Methods-Support Vector Learning*”, MIT Press, pp. 185–208, 1999.

- [70] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya and K. R. K. Murthy, “Improvements to Platt’s SMO Algorithm for SVM Classifier Design”, *Neural computation*, Vol. 13, pp. 637–649, 2001.
- [71] R. E. Fan, P. H. Chen and C. J. Lin, “Working Set Selection Using the Second Order Information for Training SVM”, *JMLR*, Vol. 6, pp. 1889-1918, 2005.
- [72] G. Anthony, H. Gregg and M. Tshilidzi, “Image classification using SVM’s: One-against-One Vs One-against-All”, 28th Asian Conference on Remote Sensing, Computer Vision and Pattern Recognition, 2007.
- [73] R. McCue, “A comparison of the accuracy of support vector machine and naive Bayes algorithms in spam classification”, University of California at Santa Cruz, pp. 1-17, 2009.
- [74] H. Bouirouga, S. El Fkihi, A. Jilbab and D. Aboutajdin, “Comparison of performance between different SVM kernels for the identification of adult video”, *World Academy of Science, Engineering and Technology*, Vol. 5-05-21, pp. 129-132, 2011.
- [75] L. Yi and Y. Zheng, “One-against-all multi-class SVM classification using reliability measures”, *IEEE International Joint Conference on Neural Networks, IJCNN '05*, Vol.2, pp. 849 – 854, 31 July - 4 August, 2005.
- [76] T.C. Mota and A. Thome, “One-Against-All-based multiclass SVM strategies applied to vehicle plate character recognition”, *International Joint Conference on Neural Networks, IJCNN '09*, pp. 2153 – 2159, Atlanta, GA, 14-19 June, 2009.
- [77] K.B. Duan¹, S. Keerthi, “Which is the best multiclass SVM method: an empirical study”, *Springer-Verlag, Berlin Heidelberg*, pp. 278–285, 2005.
- [78] Vincent Vandewalle, *Estimation et sélection en classification semi-supervisée, thèse de doctorat*, Mathématiques Appliquées, Université Lille 1, pp. 23-30, 2009.
- [79] M. Samb, F. Camara, S. Ndiaye, Y. Slimani et M. Esseghir, « Approche de sélection d’attributs pour la classification basée sur l’algorithme RFE-SVM », *Special issue CARI’12, ARIMA Journal*, Vol. 17, pp. 197-219, 2014.
- [80] Christophe Salperwyck, « Apprentissage incrémental en ligne sur flux de données », thèse de doctorat, spécialité informatique, Université Lille 3, pp. 70-80, 2012.
- [81] Eric Ngouana et Serge Mayaya, « Classification Bayésienne Naïve des Textes », *Faculté Polytechnique de Mons, 5ème Electricité, Certificat Applicatif Multimédia*, pp. 1-14, 13 Décembre 2005.
- [82] Marine Campedel, Pierre Hoogstoël, « Sémantique et multimodalité en analyse de l’information », *Recherche d’information et Web*, Edition Lavoisier, pp. 73-79, 2015.
- [83] I. Rish, “An empirical study of the naive Bayes classifier”, *Workshop on Empirical Methods in Artificial Intelligence, Computer Science, RC 22230*, Vol. W0111-014, pp. 1-6, 2 November, 2001.
- [84] F. Liwei, *Independent component analysis for naïve bayes classification*, Thesis of Doctor of Philosophy, Dalian University Of Technology, Department of Industrial & Systems Engineering, National University of Singapore, 2010.

- [85] Loïc Barrault, *Diagnostic pour la combinaison de systèmes de reconnaissance automatique de la parole*, thèse de doctorat, Laboratoire d'Informatique d'Avignon, Académie d'Aix-Marseille, Université d'Avignon et des Pays de Vaucluse, pp. 9-46, 2007.
- [86] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification", *IEEE Trans. Inf. Theory*, Vol. 13, No. 1, pp. 21–27, 1967.
- [87] Fatiha Barigou, *Contribution à la catégorisation de textes et à l'extraction d'information*, thèse de doctorat, département d'informatique, Université d'Oran, pp. 46-49, 2013.
- [88] Koudri Mohammed, *Modèle de mélange Gaussien, Application sur image cytologique, Mémoire de fin d'études*, Département d'Informatique, Faculté de Technologie, Université Abou Bakr Belkaid de Tlemcen, pp. 22-26, 2011.
- [89] Philippe Besse et Béatrice Laurent, "Apprentissage Statistique : modélisation, prévision et data mining", 5ème année GMM – MMS, Institut de Mathématiques de Toulouse — UMR CNRS C5219 Département Génie Mathématique et Modélisation Institut National des Sciences Appliquées de Toulouse.
- [90] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang and H. Motoda, "Top 10 algorithms in data mining", *Knowledge and Information Systems*, Vol. 14, No. 1, pp. 1–37, 2008.
- [91] C. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, ISBN. 978-0-387-31073-2, 2006.