

### 3.7 Conclusions

Nous avons défini un modèle de normalité à partir de la transformée en curvelets pour les différents patchs de la subdivision. Ce modèle est issu de l'apprentissage du support normal des atomes des curvelets sur les transformées en curvelet de patchs sans signature inusuelle. Le support normal correspond aux atomes activés redondants dans ces représentations. La reconstruction à partir de ce dictionnaire donne une estimation normale du patch, des artéfacts liés aux signatures inusuelles et aux basses fréquences de la transformée restent présents sur cette reconstruction.

Malgré la représentation dans le dictionnaire des curvelets, cette représentation ne permet pas de détecter les signatures inusuelles en fixant un score de détection sur l'intégralité du patch. Les signatures atypiques restent noyées du fait de leur faible dimension et intensité.

Une étude ponctuelle intrinsèque des résidus positifs et négatifs donne de bons résultats de détection des différents points provenant de signatures atypiques. Les approches permettent de fixer automatiquement des seuils de détection à partir de tests statistiques sur la queue de distribution des résidus ou à partir de valeurs extrêmes. Ces deux méthodes donnent des résultats assez proches. L'algorithme détecte bien les signatures inusuelles mais également les points issus de vibrations liées au  $N_1$  qui sont décalées sur les différents spectrogrammes. La valeur du seuil fixée par les résidus intrinsèques (sans tenir compte des autres données) entraîne nécessairement des points détectés indépendamment du fait que le patch contienne des signatures inusuelles ou non. Cette approche engendre donc de fausses détections, particulièrement pour les patchs normaux.

La définition du modèle de normalité à partir du dictionnaire des curvelets avec les résidus de reconstruction pour détecter les signatures atypiques permettent la mise en évidence automatique des signatures inusuelles sur les patchs sans a priori sur ces dernières.



## Chapitre 4

# Représentation par dictionnaire data-driven - Non-Negative Matrix Factorization (NMF)

## 4.1 Les dictionnaires adaptatifs/data-driven

### 4.1.1 Définition mathématique du problème

Dans cette partie, nous présentons deux types de dictionnaires.

- Les dictionnaires prédéfinis à partir de fonctions génériques dont les atomes ne sont pas modifiés suivant les données étudiées. On parle alors de dictionnaire non-adaptatif comme ses atomes restent les mêmes. Le dictionnaire est choisi de manière à ce que ses atomes soient par nature adaptés à la structure des données. La complémentarité entre le dictionnaire et les données provient de l'avis d'un expert. Ils forment en général des bases.
- Les dictionnaires data-driven définis par apprentissage sur les données. Les atomes de ces dictionnaires correspondent à des fonctions calculées et apprises à partir d'un échantillon des données. On parle de dictionnaire adaptatif car les atomes sont modifiés en fonction des données. Ils sont donc spécifiques aux données. L'adaptation du dictionnaire aux données est établie à partir de la procédure d'apprentissage. Ces dictionnaires font partie du domaine de dictionary learning [87, 79].

Dans ce chapitre, nous nous intéressons au second type de dictionnaire. Les atomes de ce dernier sont appris de manière à caractériser le mieux possible les différents éléments présents dans les données. Ces dictionnaires sont appris sur un jeu de données  $X \in \mathbb{R}^{p \times n}$  contenant  $n$  données de dimension  $p$  en résolvant une fonctionnelle de la forme :

$$[D^*, C^*] = \underset{D, C}{\operatorname{argmin}} \{g(X, DC)\} \text{ avec } X \approx DC \quad (4.1)$$

$D \in \mathbb{R}^{p \times r}$  et  $C \in \mathbb{R}^{r \times n}$  représentent respectivement le dictionnaire et les coefficients dans ce dictionnaire et  $g$  la fonction de coût mesurant la distance entre les données  $X \in \mathbb{R}^{n \times p}$  et leurs estimations selon leurs représentations à partir du dictionnaire  $D$ . La distance  $l_2$  est la plus courante pour apprendre les dictionnaires, il est également possible d'utiliser la divergence de Kullback-Leibler. Il s'agit d'une forme de décomposition matricielle, la dimension  $r$  du dictionnaire (son nombre d'atomes) doit être définie également.

- Si  $r < p$  on parle de réduction de dimension, les données sont projetées dans un espace de dimension réduite pour être étudiées ou représentées. Cette approche permet de caractériser les éléments redondants (c'est-à-dire observés plusieurs fois) dans les données.
- Si  $r > p$  on parle alors de dictionnaire surcomplet, les données sont projetées dans un espace de plus grande dimension permettant une décomposition plus fine des données. Ce type de dictionnaire peut permettre une représentation parcimonieuse des données en seignant les coefficients non significatifs. Le dictionnaire ainsi défini est alors considéré comme redondant (plusieurs atomes peuvent caractériser les mêmes éléments).
- Si  $r = p$  les données sont projetées dans un espace de même dimension définie par une matrice de passage.

La fonctionnelle définie par l'équation 4.1 ne possède pas de solution unique. Afin de trouver une solution robuste et intéressante, des contraintes  $\phi$  doivent être ajoutées (4.2). Ces contraintes peuvent être sur les coefficients  $C$  et/ou le dictionnaire  $D$ . Il peut s'agir de contraintes de parcimonie, de régulation  $l_2$ , de positivité, d'invariance par translation ou par rotation,...

$$[D^*, C^*] = \underset{D, C}{\operatorname{argmin}} \{g(X, DC) + \lambda_C \phi_C(C) + \lambda_D \phi_D(D)\} \quad (4.2)$$

Dans la littérature récente [98, 93], les problèmes étudiés attribuent des contraintes de parcimonie sur les coefficients  $C$  (4.3). Cette parcimonie provient d'une pénalisation  $l_0$  des coefficients ajoutée à la fonction de coût. Cette formulation n'est cependant pas convexe, elle est approximée en remplaçant la pseudo-norme  $l_0$  par la norme  $l_1$  (4.4) qui peut être vue comme la norme "la plus proche" de  $l_0$  [87, 79, 4]. Ces formulations sont donc composées d'un terme d'attache aux données et d'un terme de parcimonie sur les coefficients.

$$[D^*, C^*] = \underset{D, C}{\operatorname{argmin}} \left\{ \frac{1}{2} \|X - DC\|_2^2 + \lambda \|C\|_0 \right\} \quad (4.3)$$

$$[D^*, C^*] = \underset{D, C}{\operatorname{argmin}} \left\{ \frac{1}{2} \|X - DC\|_2^2 + \lambda \|C\|_1 \right\} \quad (4.4)$$

avec  $\|C\|_0 = \operatorname{card}(C \neq 0)$

#### 4.1.2 Les méthodes de résolution

Les problématiques (4.3) et (4.4) se résolvent généralement par des optimisations alternées itératives des paramètres  $C$  et  $D$  en supposant l'un ou l'autre fixé. L'optimisation du paramètre

$C$  avec les contraintes de parcimonie (en supposant  $D$  fixé) est effectuée à partir de méthodes de sparse coding [87, 41]. Ce problème est similaire aux problématiques de type LASSO [110].

$$C^k = \underset{C}{\operatorname{argmin}} \left\{ \frac{1}{2} \|X - D^k C\|_2^2 + \lambda \|C\|_1 \right\} \quad (4.5)$$

avec  $D^k$  l'estimation du dictionnaire à l'itération  $k$ . Les algorithmes de matching pursuit (MP) [82] permettent de donner une solution approximative de (4.5) en considérant la norme  $l_0$ . Il s'agit d'une méthode itérative consistant à déterminer l'atome du dictionnaire ayant la plus forte corrélation avec les signaux et de retrancher la contribution de cet atome des signaux. Le processus est répété à l'itération suivante avec le signaux résiduels. L'algorithme s'arrête à un nombre d'atomes pré-déterminé par l'utilisateur ou à un niveau d'erreur de reconstruction. L'orthogonal matching pursuit (OMP) [43] est une extension du MP où une nouvelle représentation est calculée à chaque itération à partir des atomes sélectionnés et retranchés des signaux. Les algorithmes d'optimisation de basis pursuit (BP) [28] prennent en compte la norme  $l_1$  et utilisent la programmation linéaire pour résoudre (4.5). Dans le même cadre, l'algorithme FOCUSS (FOCal Underdetermined System Solver) [51] résout le problème (4.5) en remplaçant la norme  $l_1$  par une pseudo-norme  $l_p$  avec  $0 < p \leq 1$ . Ces pseudo-normes sont plus proches de la pseudo-norme  $l_0$  que la norme  $l_1$ . Les algorithmes proximaux [88] permettent également de résoudre cette problématique.

La mise à jour du dictionnaire  $D$  peut s'effectuer par une méthode similaire aux moindres carrés, la méthode des directions optimales (MOD) [46], si aucune contrainte n'est donnée à  $D$ . Généralement, pour éviter de trop grandes valeurs du dictionnaire, des contraintes limitant la norme 2 des atomes du dictionnaire sont ajoutées.

#### 4.1.3 Quelques exemples de dictionnaires data-driven (adaptatifs)

L'ACP peut être considérée comme le dictionnaire appris le plus répandu. Les atomes du dictionnaire correspondent aux composantes principales maximisant la variance au sein des données.

Les k-means [49] sont une méthode permettant de construire un dictionnaire data-driven. Les atomes du dictionnaire sont définis par les centroïdes des clusters. Les données sont alors représentées par un coefficient dont la dimension est égale au nombre de clusters et prenant la valeur 1 au niveau du cluster attribué et 0 ailleurs. Il s'agit de la plus parcimonieuse des représentations. Les K-SVD [4] sont présentés comme une généralisation des k-means avec une représentation des données par plus d'un atome et dont les coefficients peuvent prendre des valeurs différentes de 0 ou 1. La particularité de ce dictionnaire l'apparentant au k-means provient de la mise à jour itérative du dictionnaire atome par atome (en considérant tous les autres fixés) à partir d'une décomposition en valeur singulière (SVD).

Les méthodes précédentes permettent de trouver une bonne représentation des données  $X$

dans les atomes appris de  $D$  en imposant une parcimonie sur les coefficients  $C$ . Les dictionnaires ne sont pas définis uniquement à partir des contraintes de parcimonie. La Non-Negative Matrix Factorization (NMF) [70] permet de proposer un dictionnaire en imposant la positivité au dictionnaire et aux coefficients. Cette représentation impose donc uniquement des combinaisons linéaires positives. Une description plus détaillée de la méthode est donnée par la suite. Il est également possible de combiner les contraintes. Dans [5], l'algorithme K-SVD est appliqué avec une contrainte supplémentaire de positivité et d'invariance par translation dans [78]. Des contraintes de parcimonie sont également ajoutées à la NMF dans [93]. Les réseaux de neurones [66, 68] tels que les autoencodeurs [109] peuvent également correspondre à des dictionnaires appris à partir des données. Les éléments des différentes couches donnent une représentation de la donnée, les poids au niveau de ces couches correspondent au dictionnaire. Ce type de décomposition possède la particularité de ne pas être linéaire.

Les dictionnaires data-driven vus ici sont donc définis comme une décomposition matricielle sous contraintes où les atomes du dictionnaire et les coefficients peuvent être mis à jour alternativement. Les différentes méthodes d'optimisation définissent différents dictionnaires.

#### 4.1.4 Les dictionnaires appris pour caractériser les spectrogrammes

Dans le chapitre précédent, nous avons sélectionné le dictionnaire non-adaptatif des curvelets pour caractériser le modèle de normalité car la structure de ces atomes permet de représenter les formes courbes que sont les signatures vibratoires de manière parcimonieuse (avec un seuillage des coefficients). Cependant, les atomes basses fréquences de ce dictionnaire restent activés et entraînent une reconstruction grossière des signatures atypiques. L'utilisation de ce dictionnaire permet de réduire les signatures inusuelles sur la reconstruction mais pas de les supprimer. Les dictionnaires data-driven s'adaptent à la donnée. Ils peuvent apprendre les éléments normaux spécifiques et redondants sur les différents patchs normaux des spectrogrammes afin de définir le modèle de normalité et permettre la reconstruction de ces éléments uniquement.

La dimension de nos patchs ( $128 \times 128$ ) est nettement plus grande que le nombre de données dont nous disposons (de l'ordre de 500). Il nous est donc impossible de définir un dictionnaire surcomplet (comme pour les curvelets) car les atomes correspondraient alors aux données utilisées pour l'apprentissage du dictionnaire. Nous nous plaçons donc dans un cadre de réduction de dimension avec une nouvelle dimension  $r$  très inférieure à la dimension des patchs et au nombre de données utilisées pour l'apprentissage. Cette réduction de dimension permet de considérer principalement les éléments redondants au sein des données qui correspondent au caractères normaux de nos patchs. Nous ne cherchons donc pas à trouver une représentation parcimonieuse à l'instar de la transformée en curvelets mais cherchons une représentation des vibrations normales en dimension réduite. Cela est possible car le nombre de raies présentes sur les différents patchs est limité. Cependant, un des désavantages des dictionnaires data-driven par rapport aux dictionnaires génériques est la non-unicité de la représentation. En fonction des données d'apprentissage, les dictionnaires définis ne sont pas les mêmes.

Dans ce chapitre nous utilisons la Non-Negative Matrix Factorization (NMF) pour caractériser le modèle de normalité des spectrogrammes. Nous considérons ainsi les spectrogrammes comme une combinaison positive des différents atomes/sources du dictionnaire. La NMF a principalement été utilisée pour effectuer de la séparation de sources aveugle [30] et de la décomposition d'images [69]. Nous utilisons une nouvelle fois un dictionnaire linéaire car les spectrogrammes vibratoires peuvent être considérés comme une combinaison de plusieurs sources, chaque source étant caractérisée par une raie vibratoire. Cette approche est pertinente car les spectrogrammes définissent des données positives ou chaque nouvelle source s'ajoute aux précédentes (ou du moins contribue positivement à l'ensemble). La NMF semble donc être en adéquation avec la structure des spectrogrammes. Nous définissons donc un modèle de normalité basé sur le dictionnaire appris par la NMF et donnons des résultats de détection d'anomalies sur le patch et ponctuellement.

## 4.2 Non-Negative Matrix Factorization (NMF)

### 4.2.1 Formulation mathématique

La NMF, définie dans [70], consiste à décomposer une matrice positive  $X \in \mathbb{R}_+^{p \times n}$  en un produit de deux matrices positives  $D \in \mathbb{R}_+^{p \times r}$  et  $C \in \mathbb{R}_+^{r \times n}$  (4.6).  $r$  définit la nouvelle dimension de représentation dans laquelle les données  $X$  sont décomposées. La valeur de  $r$  doit être renseignée au préalable.

$$[D^*, C^*] = \underset{D > 0, C > 0}{\operatorname{argmin}} \|X - DC\|_2^2 \quad (4.6)$$

La matrice  $D$  correspond au dictionnaire appris sur les données  $X$  et la matrice  $C$  aux coefficients de ces mêmes données dans le dictionnaire. Il est également possible de définir comme fonction de perte une distance proche de la divergence de Kullback-Leibler.

### 4.2.2 Résolution de la problématique

Ce problème est bi-convexe, c'est-à-dire qu'il est convexe selon chacune des variables mais n'est pas convexe en fonction des 2 variables simultanément. Comme pour les apprentissages de dictionnaire, ce problème est résolu par une optimisation alternée en  $D$  et  $C$ . A partir des mises à jour définies dans [70] avec  $i$  et  $j$  les indices des matrices :

$$\begin{aligned} C_{ij}^{k+1} &= C_{ij}^k \frac{(D^{k^T} X)_{ij}}{(D^{k^T} D^k C^k)_{ij}} \\ D_{ij}^{k+1} &= D_{ij}^k \frac{(X C^{k+1^T})_{ij}}{(D^k C^{k+1} C^{k+1^T})_{ij}} \end{aligned} \quad (4.7)$$

la fonction  $k \mapsto \|X - D^k C^k\|_2^2$  est décroissante.  $C^k$  et  $D^k$  sont les matrices de coefficients et de dictionnaire à l'itération  $k$ .

Ces mises à jour permettent donc la décroissance de la fonction de coût minorée par 0, l'algorithme converge donc bien vers une solution de (4.6). De plus, tant que l'initialisation (définie aléatoirement en général), est positive, la série  $(C^k, D^k)_{k \in \mathbb{N}}$  reste positive.

## 4.3 Le modèle de normalité défini à partir de la NMF

### 4.3.1 L'apprentissage du dictionnaire

Le modèle de normalité est comme précédemment défini à partir du dictionnaire dans lequel les données sont projetées. Le dictionnaire défini par la NMF permet de caractériser les éléments redondants présents au sein des données. Dans le cas de nos spectrogrammes, il s'agit des raies présentes sur la majorité des patchs. De plus, la positivité impose des combinaisons linéaires positives empêchant les atomes du dictionnaire de reconstruire des raies inusuelles. Tous les atomes du dictionnaire sont utilisés pour représenter la donnée en dimension réduite, ainsi une signature atypique présente sur un des atomes sera retrouvée sur l'ensemble des reconstructions par la combinaison additive. De même, une signature présente sur un patch ne peut pas être reconstruite si elle n'est pas caractérisée par les atomes. Représenter cette signature reviendrait à modifier les coefficients de représentation pour différents atomes, ce qui entraînerait une modification d'une plus grande ampleur sur la reconstruction due aux aspects linéaire et positif. Ce dictionnaire possède donc la particularité de ne pas pouvoir supprimer ce qui est ajouté par les atomes. Cette contrainte de positivité liée à la relation linéaire est essentielle pour définir le modèle de normalité. Seules les signatures redondantes dans les données utilisées dans l'apprentissage sont représentées dans les atomes et seules ces signatures sont reconstruites. Ce dictionnaire est donc établi pour apprendre et reconstruire uniquement les éléments normaux de nos spectrogrammes.

Nous utilisons des patchs sans signature inusuelle de  $\Omega_{App}^j$  pour définir la matrice des données d'apprentissage. Nous notons  $n_{App}^j = \text{card}(\Omega_{App}^j)$  le nombre de données utilisées pour l'apprentissage. Nous notons  $\vec{Z}_{\mathcal{K}_j^{128}}^i \in \mathbb{R}_+^{128^2}$  le patch vectorisé permettant de créer la matrice d'apprentissage  $\vec{Z}_{\Omega_{App}^j} \in \mathbb{R}_+^{128^2 \times n_{App}^j}$  par concaténation des patchs vectorisés d'apprentissage. A l'inverse des matrices en analyse de données, les lignes de la matrice  $\vec{Z}_{\Omega_{App}^j}$  caractérisent les différentes variables correspondant aux points du patch tandis que les colonnes caractérisent les différentes observations de ce même patch. Le dictionnaire issu de la NMF est alors obtenu en décomposant la matrice  $\vec{Z}_{\Omega_{App}^j}$  en produit des matrices  $\mathcal{D}_j^{NMF} \in \mathbb{R}_+^{128^2 \times r_j}$  correspondant au dictionnaire appris et  $C_{\Omega_{App}^j} \in \mathbb{R}_+^{r_j \times n_{App}^j}$  correspondant aux coefficients dans  $\mathcal{D}_j^{NMF}$  des patchs de  $\Omega_{App}^j$  (4.8).  $r_j$  représente le rang de la représentation. Un dictionnaire est donc calculé pour chaque patch  $j$ . Il

n'y a aucune interaction entre les dictionnaires des différents patchs.

$$\underbrace{[\vec{Z}_{\mathcal{K}_j^{128}}^{i_1} \dots \vec{Z}_{\mathcal{K}_j^{128}}^{i_k} \dots \vec{Z}_{\mathcal{K}_j^{128}}^{i_{n_{App}^j}}]}_{\vec{Z}_{\Omega_{App}^j}} = \underbrace{[d_1^j \dots d_{r_j}^j]}_{\mathcal{D}_j^{NMF}} \times \underbrace{[c_{i_1} \dots c_{i_k} \dots c_{i_{n_{App}^j}}]}_{C_{\Omega_{App}^j}}, \{i_1, \dots, i_k, \dots, i_{n_{App}^j}\} \in \Omega_{App}^j$$

Les colonnes  $d_k^j$  et  $c_{i_k}$  correspondent respectivement aux atomes du dictionnaire et aux coefficients dans le dictionnaire des patchs de  $\Omega_{App}^j$ .

Nous imposons une contrainte de norme unitaire sur les atomes du dictionnaire. Pour cela nous nous inspirons de la méthode des K-SVD [4], à chaque mise à jour de  $\mathcal{D}_j^{NMF}$  et  $C_{\Omega_{App}^j}$ , nous divisons chaque colonne du dictionnaire par leurs normes et impactons cette norme sur les coefficients correspondants. Ainsi le produit  $\mathcal{D}_j^{NMF} \times C_{\Omega_{App}^j}$  reste inchangé. Nous cherchons donc à résoudre le problème suivant :

$$\forall j, [\mathcal{D}_j^{NMF}, C_{\Omega_{App}^j}] = \underset{D>0, C>0}{\operatorname{argmin}} \| \vec{Z}_{\Omega_{App}^j} - DC \|_2^2 \text{ avec } \forall k, \|d_k^j\|_2^2 = 1. \quad (4.8)$$

Nous considérons que la méthode a convergé lorsque l'erreur de reconstruction devient inférieure à un seuil sur plusieurs itérations consécutives.

Les dictionnaires  $\{\mathcal{D}_j^{NMF}\}_{j \in \mathcal{K}^{128}}$  définissent donc le modèle de normalité sur lequel nous cherchons à représenter les données de spectrogrammes patch par patch. Le rang  $r_j$  du dictionnaire est une variable indispensable pour définir le modèle de normalité comme il s'agit de la dimension dans laquelle les patchs sont décomposés. Il est déterminé indépendamment sur chaque patch.

### 4.3.2 Définition du rang du dictionnaire

Dans les méthodes de dictionary learning, les rangs des matrices utilisées pour la décomposition doivent être définis au préalable. Dans le cadre de la séparation de sources, ce rang correspond au nombre de sources décomposant le signal qui est généralement connu. Pour nos données de spectrogrammes, cette valeur est inconnue. Les raies présentes dans les patchs correspondent aux différentes sources vibratoires des moteurs d'avions. Cependant les moteurs sont sujets à de nombreuses sources vibratoires dont le nombre est variable. La transformée de Hough [10] pourrait permettre de déterminer le nombre de raies présentes sur les patchs normaux de  $\Omega_{App}^j$ . Les raies vibratoires normales correspondent à des équations linéaires du  $N_1$ ,  $N_2$  ou RDS. Une application du seuillage sur la transformée détermine les raies présentes. Le nombre de raies différentes dans la base d'apprentissage pourrait alors correspondre au nombre d'atomes dans le dictionnaire. Cependant, cette approche entraînerait de considérer un atome pour chaque raie normale décalée sur le patch. De plus les raies présentes sur les patchs ne sont pas nécessairement complètes mais seulement partielles. Définir le nombre d'atomes comme le nombre de raies différentes présentes sur les patchs d'apprentissage ne semble pas adapté à notre problématique.

Nous considérons qu'il existe un nombre d'atomes optimal pour la représentation par la NMF. Nous faisons donc l'hypothèse, pour un patch quelconque  $j$ , qu'au bout d'un certain rang  $r_j^*$ , le gain de représentation obtenu en ajoutant un atome supplémentaire au dictionnaire  $\mathcal{D}_j^{NMF}$  n'est plus suffisant pour justifier l'augmentation de la complexité de représentation. De plus, si le rang est trop important et est proche du nombre de données d'apprentissage, le dictionnaire sera surappris et conviendra parfaitement pour la base d'apprentissage mais ne conviendra pas pour de nouvelles données avec une erreur de reconstruction augmentant en fonction du rang. La détermination du rang optimal est effectuée sur une base de validation  $\Omega_{Val}^j$  disjointe de la base d'apprentissage  $\Omega_{App}^j$  afin d'éviter ces problèmes de surapprentissage. Le rang optimal correspond à un faible taux d'accroissement entre l'erreur de reconstruction au rang  $r + 1$  et cette même erreur au rang  $r$  (4.9).

$$r_j^* = \underset{r}{\operatorname{argmin}} \left\{ \frac{\|\vec{Z}_{\Omega_{Val}^j} - \mathcal{D}_{j,r+1}^{NMF} C_{\Omega_{Val}^j}^{r+1}\|_2^2 - \|\vec{Z}_{\Omega_{Val}^j} - \mathcal{D}_{j,r}^{NMF} C_{\Omega_{Val}^j}^r\|_2^2}{\|\vec{Z}_{\Omega_{Val}^j} - \mathcal{D}_{j,r}^{NMF} C_{\Omega_{Val}^j}^r\|_2^2} \leq \epsilon \right\} \quad (4.9)$$

$\mathcal{D}_{j,r}^{NMF}$  et  $C_{\Omega_{Val}^j}^r$  définissent respectivement le dictionnaire et les coefficients au rang  $r$ , et  $\epsilon$  le niveau d'erreur à partir duquel on considère qu'il n'est plus nécessaire d'augmenter le rang.

La figure 4.1 illustre l'erreur de reconstruction moyenne sur la base de validation  $\Omega_{Val}$  établie à partir du dictionnaire appris de la NMF en fonction du rang pour 6 différents patchs sur 10 ensembles d'apprentissage différents. Nous pouvons remarquer que pour les premiers rangs l'erreur de reconstruction diminue pour la base de validation  $\Omega_{Val}$ . Cependant pour certains patchs, à partir d'un certain rang l'erreur croît en fonction du rang, cela signifie que le modèle a surappris sur les données d'apprentissage et ne généralise pas les éléments présents sur les patchs. Pour d'autres patchs, le modèle ne surapprend pas, cela est dû au fait que la variabilité entre les différentes données pour ces patchs en particulier n'est pas importante et donc le surapprentissage du modèle n'impacte pas la reconstruction des patchs de  $\Omega_{Val}$ . Indépendamment du patch considéré, les comportements des erreurs de reconstruction sont similaires pour les différentes bases d'apprentissage normales utilisées. Le rang est donc approximativement le même pour les différents apprentissages de la NMF. Cela signifie que le rang n'est pas dépendant des données d'apprentissage ou de validation mais est intrinsèque au patch et est indépendant des données utilisées pour la calibration du dictionnaire.

### 4.3.3 Représentation de la normalité à partir du dictionnaire de la NMF

Le dictionnaire est calculé à partir de l'équation (4.8) avec une initialisation aléatoire. Pour chaque patch  $j$  le rang  $r_j^*$  est défini au préalable (4.9). Nous présentons en figure 4.2 les différents atomes du dictionnaires obtenus pour deux patchs possédant des structures différentes. Les atomes du dictionnaire sont proches des patchs malgré l'initialisation aléatoire. Les différentes raies normales sont bien représentées dans le dictionnaire. A la différence du dictionnaire défini par les curvelets qui permet la reconstruction des raies bout à bout, ce dictionnaire caractérise l'intégralité des raies. Les atomes ont des structures très proches des patchs considérés. Les dif-

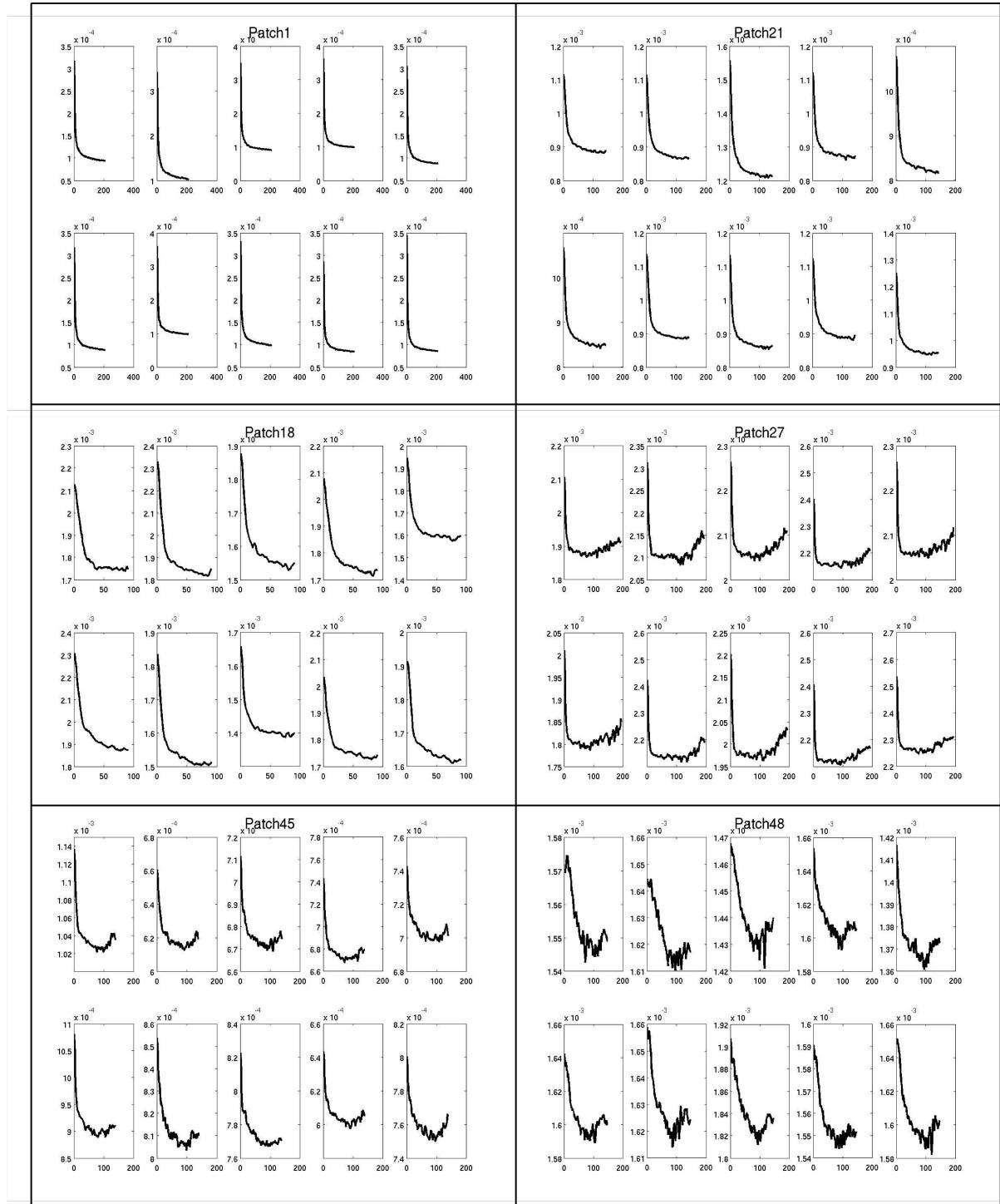


FIGURE 4.1 – Erreur de reconstruction en validation en fonction du rang de la NMF pour 6 différents patchs. Chaque bloc de l'image correspond à un des patchs et chaque courbe dans ces blocs correspond aux résultats de reconstruction pour différentes données d'apprentissage. Les comportements pour les différentes données d'apprentissage sur chaque bloc sont les mêmes. Les rangs optimaux  $r^*$  ne sont pas dépendants des données d'apprentissage mais dépendent de la structure intrinsèque du patch considéré.

férentes informations vibratoires normales sont présentes et interprétables sur ces atomes, cela provient de la positivité imposée par la NMF qui entraîne une combinaison additive des atomes. A partir d'une méthode de définition de dictionnaire telle que l'ACP, les atomes possèderaient des structures non similaires aux patchs. De plus, les atomes du dictionnaire  $\mathcal{D}^{NMF}$  tiennent compte des différents comportements des raies  $N_1$  (décalage sur les différents spectrogrammes) en représentant plusieurs comportements sur chaque atome. Il s'agit d'une zone d'incertitude pour les dictionnaires. Ils tiennent également compte de la structure non complète des raies normales. Les dictionnaires  $\{\mathcal{D}_j^{NMF}\}_{j \in \mathcal{K}^{128}}$  permettent donc de caractériser les raies normales présentes sur les différents patchs des spectrogrammes et définissent donc un modèle de normalité pertinent.

Pour que ce dictionnaire définisse le modèle de normalité, il faut lui associer une méthode pour caractériser les données non issues de l'apprentissage, c'est-à-dire déterminer les coefficients de représentation dans ce dictionnaire  $\mathcal{D}_j^{NMF}$ . Comme le dictionnaire est considéré appris, la fonctionnelle définit dans l'équation (4.8) devient une optimisation en une unique variable  $C$ . Nous voulons garder la positivité des coefficients dans la représentation pour combiner positivement les atomes du dictionnaire et garder la structure de ce dernier. Pour résoudre cette problématique, nous utilisons la méthode de mise à jour (4.7) correspondant uniquement à la variable  $C$ . Nous savons déjà que cette mise à jour permet de réduire l'erreur à chaque itération et que les coefficients sont positifs tant que l'initialisation est positive.

$$C_j^i = \underset{C > 0}{\operatorname{argmin}} \|\vec{Z}_{\mathcal{K}_j^{128}} - \mathcal{D}_{j,r_j^*}^{NMF} C\|_2^2$$

avec  $i$  le spectrogramme et  $j$  le patch.

Nous pouvons donc obtenir les représentations des différents patchs des spectrogrammes dans les dictionnaires de normalité  $\{\mathcal{D}_j^{NMF}\}_{j \in \mathcal{K}^{128}}$  à partir desquelles nous donnons une estimation normale de ces patchs par reconstruction. Nous notons  $\hat{Z}_{\mathcal{K}_j^{128}}^{i,NMF}$  l'estimation dans le domaine normal de  $Z_{\mathcal{K}_j^{128}}^{i,NMF}$ .

$$\hat{Z}_{\mathcal{K}_j^{128}}^{i,NMF} = \mathcal{D}_{j,r_j^*}^{NMF} C_j^i$$

La figure 4.3 présente les reconstructions des patchs comportant des signatures atypiques à partir du dictionnaire  $\mathcal{D}_j^{NMF}$ . Nous représentons le patch étudié, sa reconstruction à partir du dictionnaire ainsi que les résidus de reconstruction. Les signatures inusuelles ne sont pas du tout reconstruites à partir de ce dictionnaire, elles se retrouvent dans les résidus positifs de la reconstruction. Des artefacts sont créés par le dictionnaire mais consistent uniquement en la complétion ou l'ajout de certaines signatures vibratoires normales. Ces artefacts correspondent à des résidus négatifs. De plus, la reconstruction à partir du dictionnaire  $\mathcal{D}^{NMF}$  a permis un débruitage du patch, le bruit semble visuellement atténué dans la reconstruction.

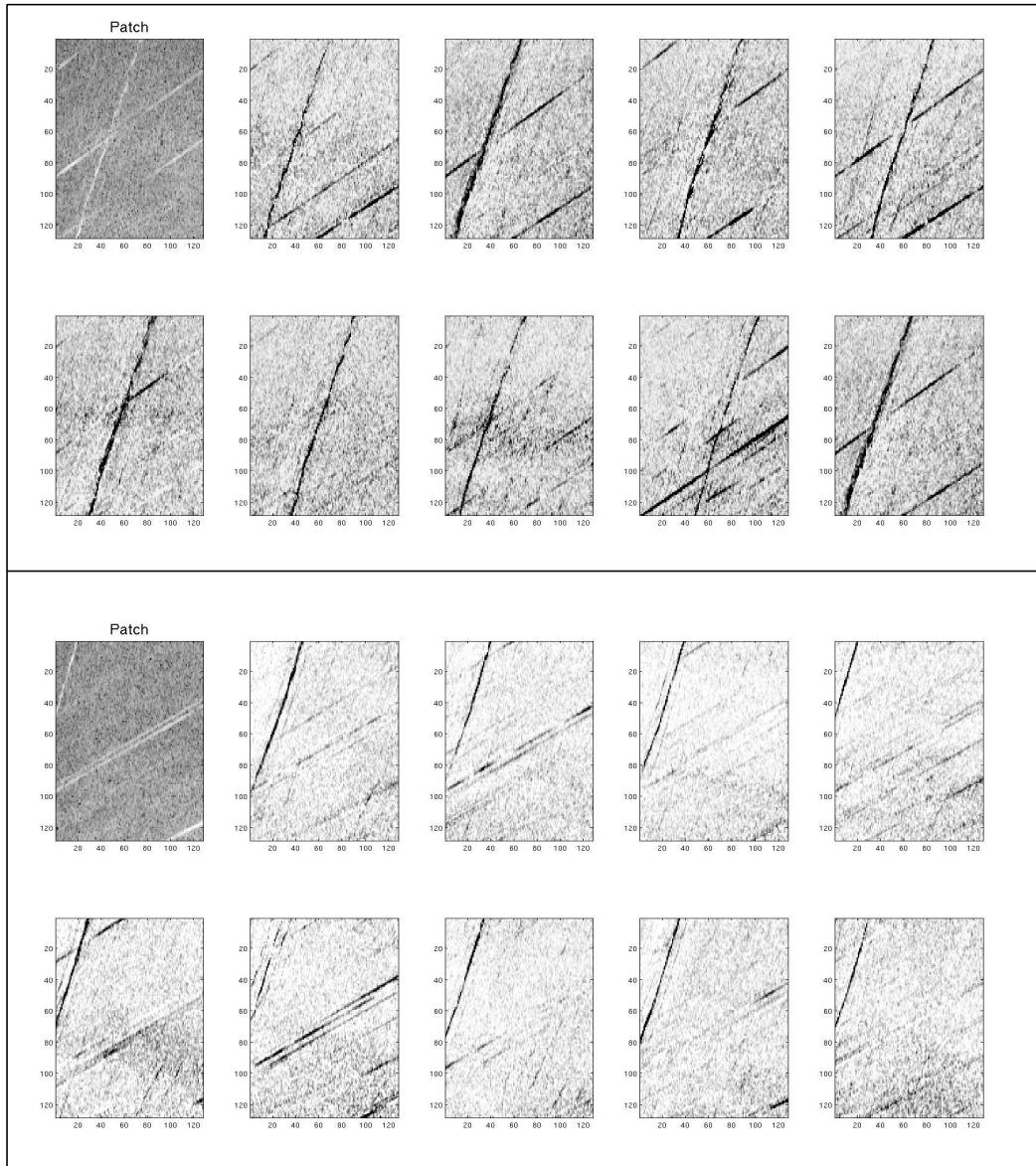


FIGURE 4.2 – Atomes du dictionnaire de la NMF appris sur les données normales pour 2 différents patchs. Les atomes des dictionnaires appris possèdent des structures proches des patchs à partir desquels ils ont été appris. Ils permettent de caractériser les différents comportements des raies normales avec la prise en compte des raies décalés dans les atomes ainsi que la non-complétude des raies normales.

## 4.4 Détection d'anomalies sur les patchs

### 4.4.1 Les scores de détection

Nous cherchons à détecter les patchs contenant des signatures inusuelles à partir de cette représentation qui donne une caractérisation en dimension réduite des patchs. Dans un premier temps, nous avons étudié des approches supervisées afin de classifier les patchs à partir des méthodes classiques du domaine comme les random forests et les SVM [49] appliquées aux repré-

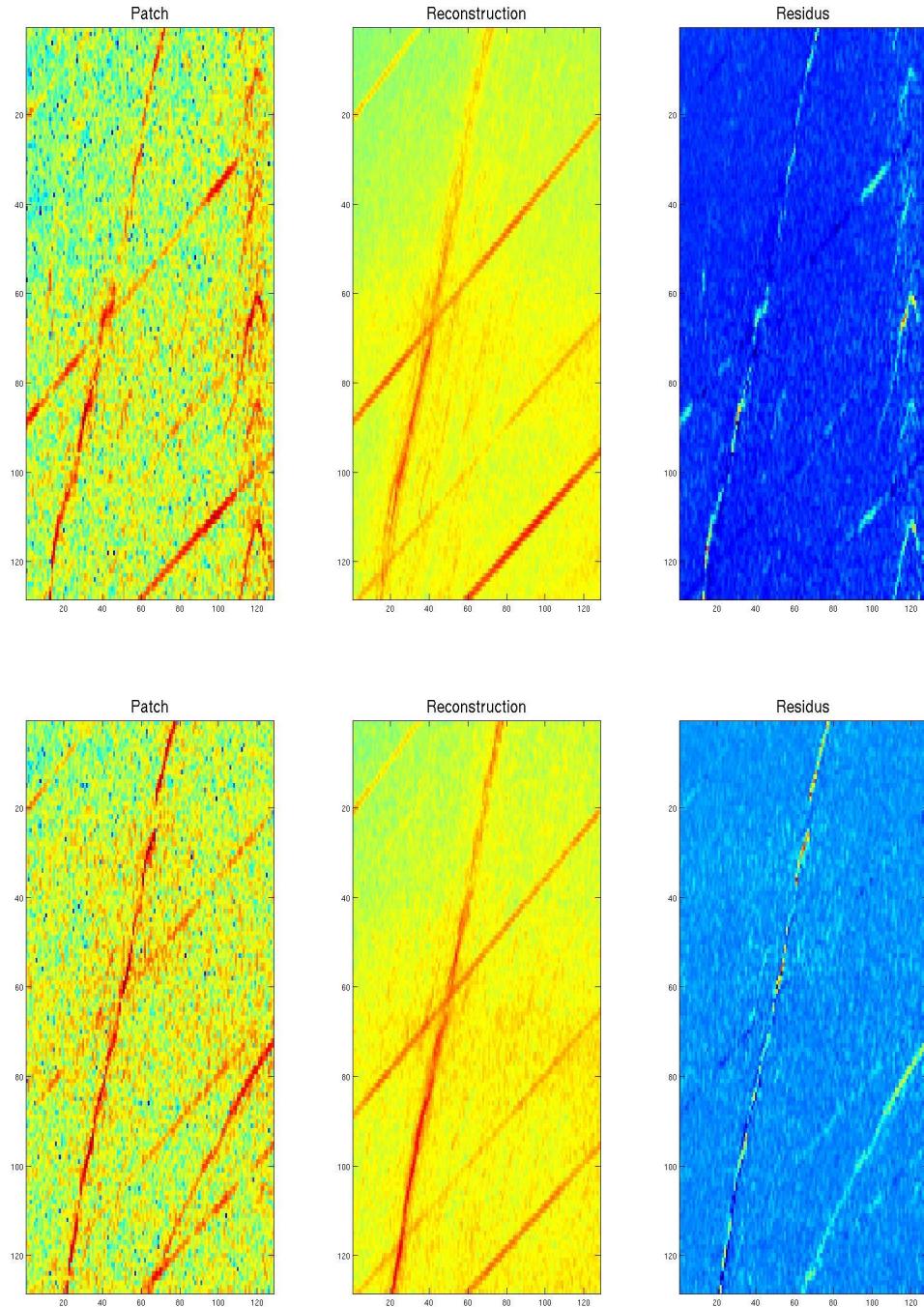


FIGURE 4.3 – Représentation de 2 patches contenant des signatures inusuelles à partir de  $\mathcal{D}^{NMF}$ . Chaque ligne correspond au patch étudié avec sa reconstruction à partir du dictionnaire et les résidus de cette reconstruction. Les signatures atypiques ne sont pas reconstruites à partir du dictionnaire  $\mathcal{D}_j^{NMF}$ . Ces signatures inusuelles se trouvent dans les résidus de la reconstruction.

sentations dans le dictionnaire de la NMF de faible dimension. Cependant les résultats n'étaient pas satisfaisants avec une courbe ROC proche de la première bissectrice. Dans un second temps, nous avons étudié les distances entre les représentants dans les dictionnaires des patchs. Les scores basés sur ces distances sont calculés à partir de l'ensemble des données normales de validation  $\Omega_{Val}^j$  non utilisées pour l'apprentissage du dictionnaire. Nous avons analysé les résultats sur deux distances classiques :

- la distance  $l_2$  minimale entre la représentation du patch étudié et les représentations des patchs dans la base de validation  $\Omega_{Val}^j$  :

$$d_{l_2}(Z_{\mathcal{K}_j^{128}}^i) = \min_{k \in \Omega_{Val}^j} \|C_j^i - C_j^k\|_2^2$$

les patchs atypiques sont censés avoir des représentations éloignées des patchs normaux de la base de validation,

- la distance de Mahalanobis entre les représentations pour tenir compte des corrélations entre les coefficients comme les atomes des dictionnaires  $\{\mathcal{D}_j^{NMF}\}_{j \in \mathcal{K}^{128}}$  ne sont pas orthogonaux,

$$d_M(Z_{\mathcal{K}_j^{128}}^i) = (C_j^i - \mu_j)^T \Sigma_j^{-1} (C_j^i - \mu_j),$$

avec  $\mu_j$  et  $\Sigma_j$  la moyenne et covariance des représentations des patchs de  $\Omega_{Val}^j$ .

Nous représentons la matrice des distances  $l_2$  entre tous les patchs sur la figure 4.4. Les premières lignes et colonnes (flèches vertes) de la matrice correspondent aux patchs normaux, les suivantes (flèches rouges) contiennent l'ensemble des patchs inusuels. Nous n'observons pas de structure diagonale par bloc dans cette matrice attestant d'une différence des distances entre les patchs normaux et les patchs inusuels. Les résultats pour la distance de Mahalanobis sont similaires et ne permettent donc pas de discriminer les patchs atypiques des patchs normaux. Les approches exploitant uniquement la représentation dans le dictionnaire  $\mathcal{D}_j^{NMF}$  ne sont pas efficaces pour discriminer les différentes catégories de patch. Les signatures inusuelles ne consistent qu'en une infime partie du patch, ils engendrent donc des modifications moindres sur les coefficients de représentation ne permettant pas la détection des patchs atypiques.

La projection dans le domaine normal engendré par  $\mathcal{D}_j^{NMF}$  permet une reconstruction sans aucune signature inusuelle, les informations atypiques se trouvent donc entièrement dans les résidus de la reconstruction (figure 4.3). L'erreur de reconstruction (4.10) est donc un score de détection d'anomalies pertinent à partir de ce dictionnaire.

$$d_{ER}^{NMF}(Z_{\mathcal{K}_j^{128}}^i) = \|Z_{\mathcal{K}_j^{128}}^i - \mathcal{D}_j^{NMF} C_j^i\|_2^2 = \|Z_{\mathcal{K}_j^{128}}^i - \hat{Z}_{\mathcal{K}_j^{128}}^{i,NMF}\|_2^2 \quad (4.10)$$

Comme pour les curvelets, le score de normalité est défini à partir de la p-valeur d'un test statistique d'adéquation à la distribution des erreurs de reconstruction des patchs normaux modélisée par une loi gamma. Nous notons  $pval_j^i$  la p-valeur associée à  $Z_{\mathcal{K}_j}^i$ , le patch  $j$  du spectrogramme  $i$ . Elle est calculée de manière analogue à celle des curvelets (3.12). Le classifieur de patch cor-

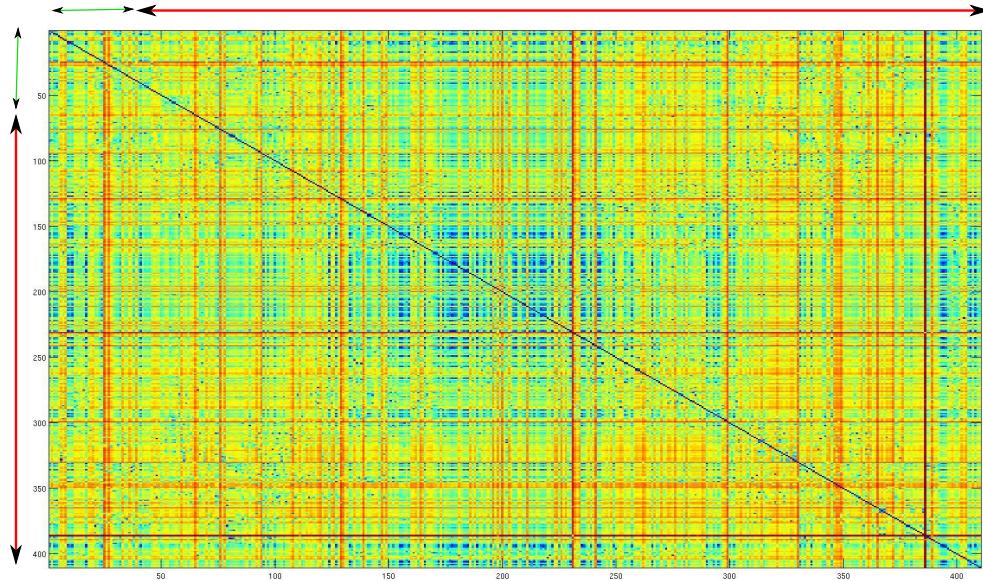


FIGURE 4.4 – Matrices des distances euclidiennes des représentations des patchs dans le dictionnaire  $\mathcal{D}_j^{NMF}$ . Les premiers patchs de la matrice correspondent aux patchs normaux (flèche verte) les suivants aux patchs inusuels (flèche rouge). La matrice ne possède pas de structure diagonale par bloc. La distance entre les coefficients de la NMF des différents patchs ne permet pas de détecter les patchs atypiques.

respond à un seuil  $s = 0.1$  sur les p-valeurs de ces tests statistiques.

$$\hat{Y}_{Z_{\mathcal{K}_j}}^i = \begin{cases} 1 & \text{si } pval_j^i \leq s \\ 0 & \text{sinon} \end{cases}$$

#### 4.4.2 Résultats sur la base de test

##### Étude du moteur avec endommagement

A l'instar de l'approche développée à partir du dictionnaire des curvelets, nous comparons dans un premier temps les résultats sur différents patchs du moteur endommagé et des moteurs normaux. Nous présentons ensuite des résultats de détection sur les différents patchs de la subdivision  $\mathcal{K}^{128}$  de la base de test  $\Omega_{Test}$ .

Le tableau 4.1 donne les p-valeurs pour le moteur endommagé et des moteurs considérés comme normaux sur le patch contenant les signatures anormales du moteur endommagé et sur un patch sans signature inusuelle pour l'ensemble des spectrogrammes de la base de test. Le moteur endommagé se distingue très bien des autres avec une p-valeur très faible pour le patch contenant les signatures anormales caractérisant son endommagement. Le moteur endommagé n'est pas détecté sur un patch normal et possède une p-valeur proche de celle des moteurs normaux. L'approche est donc pertinente pour détecter les anomalies de fortes intensités représentées par

de nombreuses raies caractéristiques d'un endommagement uniquement sur le patch contenant les signatures correspondantes.

TABLE 4.1 – P-valeurs des tests statistiques issus de l'erreur de reconstruction du moteur endommagé et des moteurs normaux sur des patchs avec et sans signatures anormales.

	Patch avec signature inusuelle	Patch sans signature inusuelle
Moteur avec endommagement	$< 10^{-16}$	$0.34 \pm 0.05$
Moteur sans endommagement	$0.52 \pm 0.03$	$0.49 \pm 0.04$

### Étude sur tous les patchs des moteurs contenant des signatures inusuelles

La figure 4.5 contient la matrice correspondant à la vérité terrain  $Y_{Z_{K_j}}^i$  patch par patch (colonne) pour les différents moteurs de la base de test  $\Omega_{Test}$  (ligne) sur la partie supérieure, les résultats de l'algorithme de détection  $\hat{Y}_{Z_{K_j}}^i$  avec un seuil sur les p-valeurs de 0.1 se trouvent sur la partie inférieure. De nombreux points au niveau des matrices ne concordent pas, c'est-à-dire qu'ils sont détectés par l'algorithme à tort ou ne sont pas détectés alors qu'ils le devraient.

Le calcul de l'erreur de reconstruction sur le patch entier noye les signatures inusuelles à travers la considération des erreurs de tous les points du patch malgré la définition d'un dictionnaire adapté au patch. Le nombre de points caractérisant les signatures atypiques n'est pas suffisant pour impacter l'erreur de reconstruction globale lorsqu'il ne s'agit pas d'une signature liée à une endommagement réel. Pour distinguer les signatures inusuelles, il peut être important de considérer les erreurs au niveau de chaque point du patch et non dans leur globalité.

## 4.5 Les erreurs ponctuelles issues du dictionnaire de la NMF

### 4.5.1 Les résidus de la NMF

Les résidus  $R_{NMF}$  issus des dictionnaires de la NMF  $\{\mathcal{D}_j^{NMF}\}_{j \in \mathcal{K}^{128}}$  correspondent à la différence point par point du patch et de sa reconstruction à partir du domaine normal défini par le dictionnaire.

$$R_{NMF}(Z_{\mathcal{K}_j^{128}}^i) = Z_{\mathcal{K}_j^{128}}^i - \hat{Z}_{\mathcal{K}_j^{128}}^{i,NMF}$$

Les résidus sont analysés pour chaque moteur indépendamment des autres, nous considérons donc les résidus intra-patch pouvant être modélisés par une gaussienne (Figure 4.6). Les histogrammes des résidus des patchs normaux sont similaires à ceux des patchs inusuels, les résidus sont centrés en 0 avec des valeurs de résidus positifs assez importantes. Les résidus ne sont pas comparés aux résidus de représentation des autres données. Nous considérons que les éléments atypiques présents sur les patchs correspondent aux éléments qui ne sont pas correctement représentés par le dictionnaire de la NMF et se trouvant dans les extrêmes des résidus. Les différents résidus de la représentation n'ont pas la même signification.

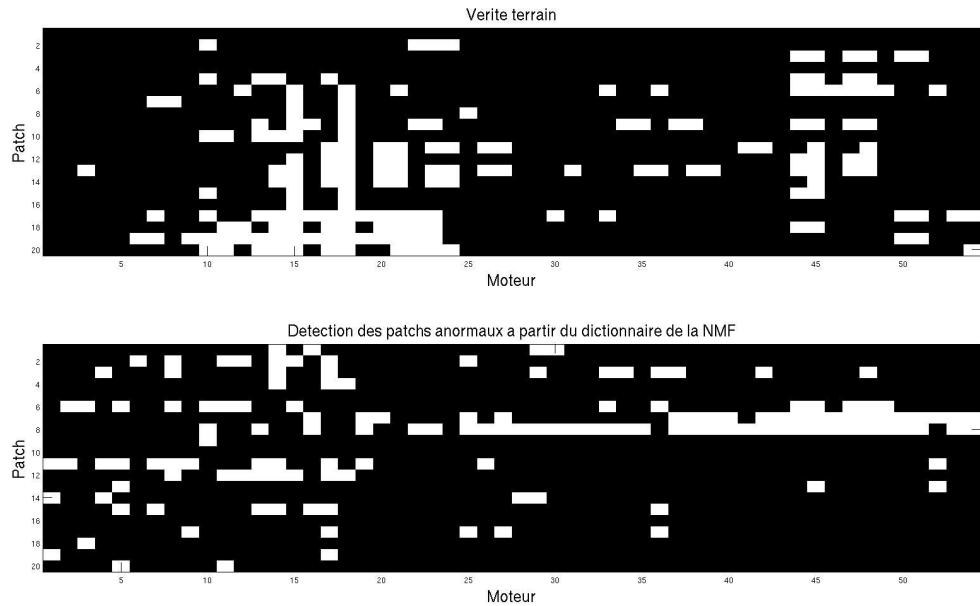


FIGURE 4.5 – Vérité terrain de la présence de signatures atypiques sur les patchs (partie supérieure) et classification à partir de l'erreur de reconstruction obtenue à partir de la représentation dans le dictionnaire de la NMF (partie inférieure). Chaque ligne des matrices correspond à un spectrogramme de la base de test et chaque colonne à un patch. Les deux matrices possèdent plusieurs différences avec des patchs détectés à tort et non détectés. L'erreur de reconstruction considérée globalement sur le patch ne permet pas la détection des signatures atypiques faibles.

- Les résidus autour de 0 correspondent aux points correctement représentés et donc reconstruits par le dictionnaire (figure 4.6 cadre noir).
- Les résidus "fortement" négatifs correspondent à des artefacts créés par le dictionnaire de la NMF, (figure 4.6 cadre bleu).
- Les résidus "fortement" positifs correspondent aux points qui sont mal caractérisés par le dictionnaire et qui sont donc mal reconstruits, nous les considérons comme les points inusuels présents sur les patchs (figure 4.6 cadre rouge).

#### 4.5.2 La détection des points inusuels

A la différence des curvelets, nous considérons uniquement les résidus positifs (4.11) contenant les signatures inusuelles au niveau de leurs valeurs extrêmes. Le dictionnaire de la NMF ne produit pas d'artefacts lors de la reconstruction identifiant les signatures inusuelles. Ces artefacts correspondent à la construction de raies normales non présentes sur les données (figure 4.3). Les valeurs extrêmes négatives n'ont donc pas d'utilité pour la détection des points atypiques sur les patchs.

$$R_{NMF}^+(Z_{\mathcal{K}_j^{128}}^i) = \max(R_{NMF}(Z_{\mathcal{K}_j^{128}}^i), 0) \quad (4.11)$$

La détection des points inusuels des patchs s'effectue uniquement sur valeurs extrêmes des

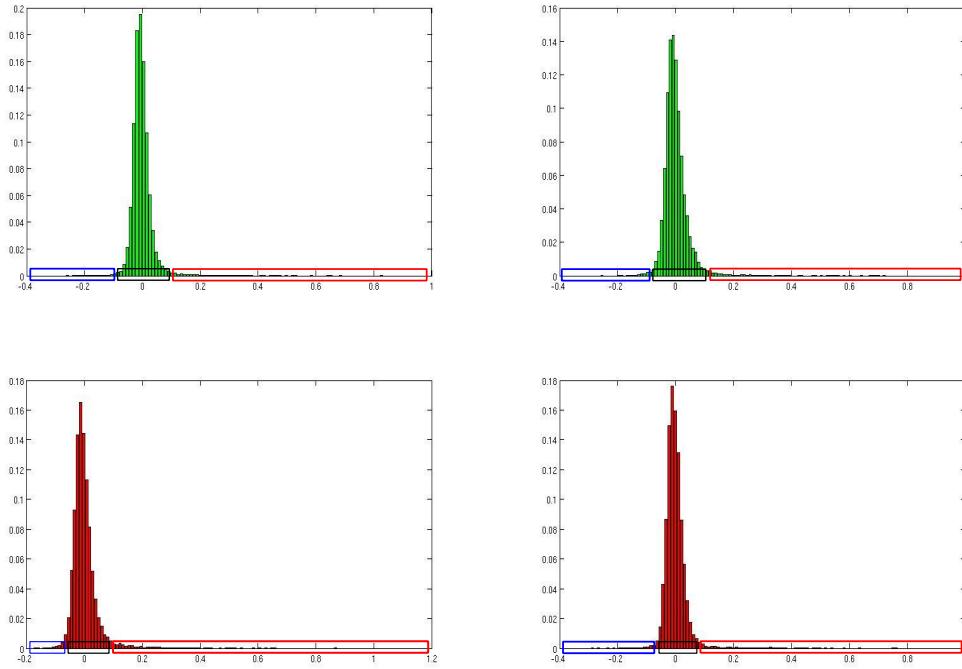


FIGURE 4.6 – Histogrammes des résidus de reconstruction par la NMF pour 2 patchs normaux (partie supérieure) et 2 patchs inusuels (partie inférieure). Les cadres noirs correspondent aux points qui ont été correctement représentés par le dictionnaire, les cadres bleus aux artéfacts créés par le dictionnaire et les points rouges aux points mal représentés par le dictionnaire. Les résidus peuvent être modélisés par des gaussiennes.

résidus positifs, il s'agit de fixer un seuil pour les déterminer. Ce seuil est défini pour tous les patchs et ne doit donc pas être sensible aux différences d'ordres de grandeur des résidus entre les différents patchs. Nous définissons les seuils de détection à partir des deux mêmes approches définies pour les curvelets.

- Le seuil est défini sur les p-valeurs des tests statistique d'adéquation à la distribution des résidus modélisée par une gausienne.

$$pval^+(Z_{\mathcal{K}_j(f, N_2)}^i) \approx \mathbb{P}_{\mathcal{N}(\hat{\mu}_j^i, \hat{\sigma}_j^i)} \left( X > R_{NMF}^+(Z_{\mathcal{K}_j(f, N_2)}^i) \right) = 1 - F_{\mathcal{N}(\hat{\mu}_j^i, \hat{\sigma}_j^i)} \left( R_{NMF}^+(Z_{\mathcal{K}_j(f, N_2)}^i) \right)$$

$\hat{\mu}_j^i$  et  $\hat{\sigma}_j^i$  représentent les paramètres des distributions gaussiennes des résidus  $R_{NMF}$ . La détection s'effectue alors par comparaison de la p-valeur de chaque point au seuil de détection  $s_+$ .

$$\tilde{Y}_{f, N_2}^i = \mathbf{1} \left\{ pval^+(Z_{\mathcal{K}_j(f, N_2)}^i) \leq s_+ \right\}$$

- Le seuil  $s_{max}$  est défini par le mode de la distribution des valeurs extrêmes estimée à partir des  $n_{max}$  résidus positifs maximaux.

$$s_{max} = \operatorname{argmax} \left( f_{R_{NMF, n_{max}}^+}^e \right)$$

La détection s'effectue par comparaison des résidus positifs au  $s_{max}$ .

$$\tilde{Y}_{f,N_2}^i = \mathbb{1} \left\{ R_{NMF}^+(Z_{\mathcal{K}_j(f,N_2)}^i) \geq s_{max} \right\}$$

Ces approches permettent de définir automatiquement des seuils de détection sur les données. Ces seuils sont calibrés sur la base de validation  $\Omega_{Val}$  de la base de données  $\mathcal{B}_1$ , les données dont nous disposons d'une annotation de quelques points. Les résultats sont présentés sur la base de test  $\Omega_{Test}$  de la même base de données. Nous ajoutons un filtrage de voisinage pour peaufiner la détection en considérant qu'un point est inusuel si et seulement si 2 de ses voisins directs  $\mathcal{V}_{f,N_2}$  sont inusuels également comme les signatures atypiques correspondent à des raies.

$$\hat{Y}_{f,N_2}^{i,NMF} = \begin{cases} 0 & \text{si } \tilde{Y}_{f,N_2}^i = 0 \\ 0 & \text{si } \tilde{Y}_{f,N_2}^i = 1 \text{ \& } \sum_{k \in \mathcal{V}_{f,N_2}} \tilde{Y}_k^i < 2 \\ 1 & \text{si } \tilde{Y}_{f,N_2}^i = 1 \text{ \& } \sum_{k \in \mathcal{V}_{f,N_2}} \tilde{Y}_k^i \geq 2 \end{cases} \quad (4.12)$$

#### 4.5.3 Calibration des paramètres

Les points de la base de données  $\mathcal{B}_1$  sont répartis en 4 classes (cf 1.5.3), les points normaux, les points inusuels, le bruit et les points normaux décalés sur les spectrogrammes. Seuls les points inusuels sont à détecter.

##### Calibration des paramètres des tests d'hypothèses

Le seul paramètre devant être calibré pour cette approche est le seuil de détection. Pour le définir nous représentons le taux de détection des différentes classes de points de la base de validation  $\Omega_{Val}$  en fonction de différents seuils sur les p-valeurs (Figure 4.7). Pour le dictionnaire basé sur les curvelets, nous disposons de deux niveaux de détection le premier au niveau des résidus positifs et le second au niveau des résidus négatifs. Cela permettait de laisser une certaine marge au niveau du choix du seuil. Cependant, cette double validation n'est pas présente pour ce dictionnaire adapté aux données, le seuil doit donc être sélectionné de manière à détecter l'ensemble des points inusuels tout en évitant la détection à tort des points normaux. Le seuil sélectionné est donc moins conservateur que celui défini pour les curvelets.

Nous observons un fort taux de bonnes détections des points inusuels pour des niveaux de seuil faibles, ainsi qu'un fort taux de fausses détections des points décalés (Figure 4.7). Cela signifie que les raies décalées (donc les raies liées au  $N_1$ ) sont mal représentées par le dictionnaire entraînant des résidus de fortes intensités au niveau de ces raies. Ces raies possèdent des intensités vibratoires plus importantes que les raies inusuelles entraînant donc une détection équivalente entre les points atypiques et les points normaux décalés. De plus, le taux de détection des autres points normaux augmente rapidement avec le seuil (beaucoup plus que pour le dictionnaire

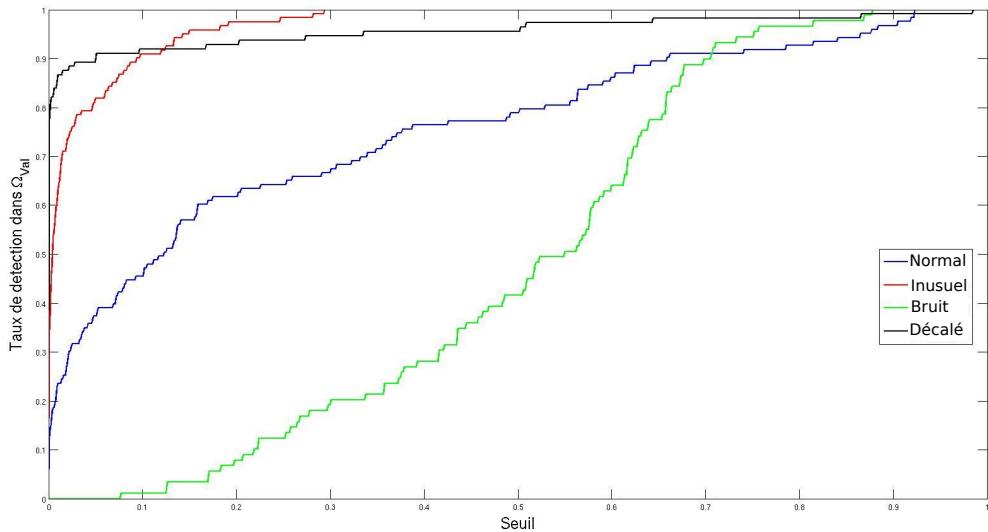


FIGURE 4.7 – Taux de détection pour les différentes classes de points de la base de validation  $\Omega_{Val}$  en fonction du seuil sur les p-valeurs. Les points atypiques (rouge) et les points se trouvant sur des raies normales décalées (noir) sont fortement détectés pour des petites valeurs de seuils. Les points normaux (bleu) et du bruit (vert) sont faiblement détectés mais augmentent rapidement en fonction du seuil de détection. Nous fixons donc un niveau de seuils bas.

des curvelets), pour un seuil de 0.1, il atteint les 40%. Nous avons donc choisi comme seuil de détection sur les p-valeurs :

$$s = 0.03.$$

Ce seuil permet un taux de bonne détection d'environ 80% et un taux de détection des points normaux d'environ 25% et de 90% des points décalés dans la base de validation. La figure 4.8 présente des images de détection de différents patchs contenant des signatures inusuelles pour différentes valeurs de seuil. Lorsque le seuil de détection est trop faible, le modèle est trop strict et les signatures atypiques ne sont pas détectées. En cas de seuil trop important de nombreux points normaux sont détectés à tort. Le seuil choisi permet de faire un compromis entre ces 2 effets, il est reporté sur les autres patchs.

### Calibration des paramètres pour les valeurs extrêmes

Pour ce système de détection, le paramètre impactant le seuil de détection est le nombre de points utilisés pour calibrer la distribution des valeurs extrêmes. Nous représentons le taux de détection des différentes classes de points de la base de validation en fonction de la taille des résidus positifs dont est issue la distribution des extrêmes (Figure 4.9). Le taux de bonnes détections augmente avec le nombre de résidus pris en compte car le seuil devient de plus en plus faible. Cependant le taux de fausses détections augmente également en fonction du nombre de données. Nous sélectionnons comme nombre de données pour estimer la distribution des

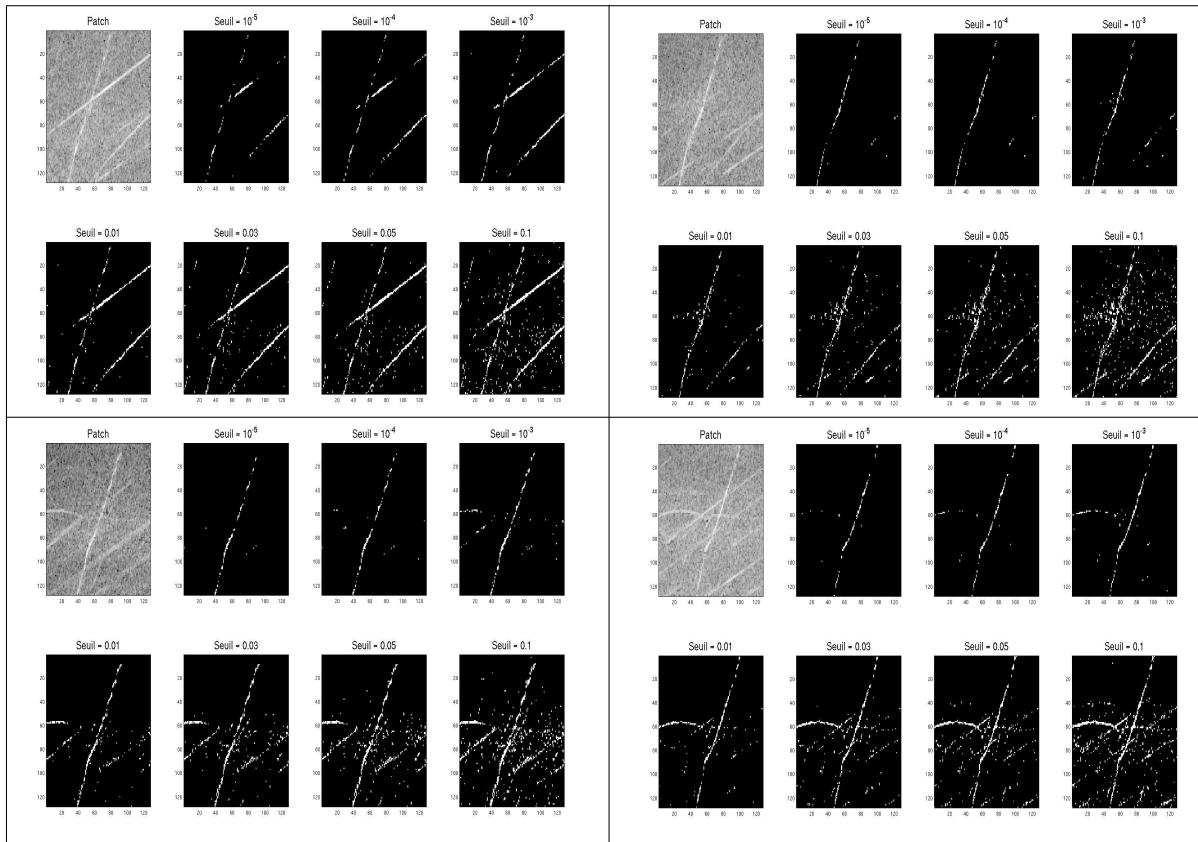


FIGURE 4.8 – Points détectés pour différents patchs de la base de validation à partir d'un seuil sur les p-valeurs des résidus issus du dictionnaire de la NMF pour différentes valeurs de seuil. Pour des seuils trop faibles, les signatures atypiques ne sont pas détectables, tandis que des seuils trop importants entraînent la détection des signatures normales.

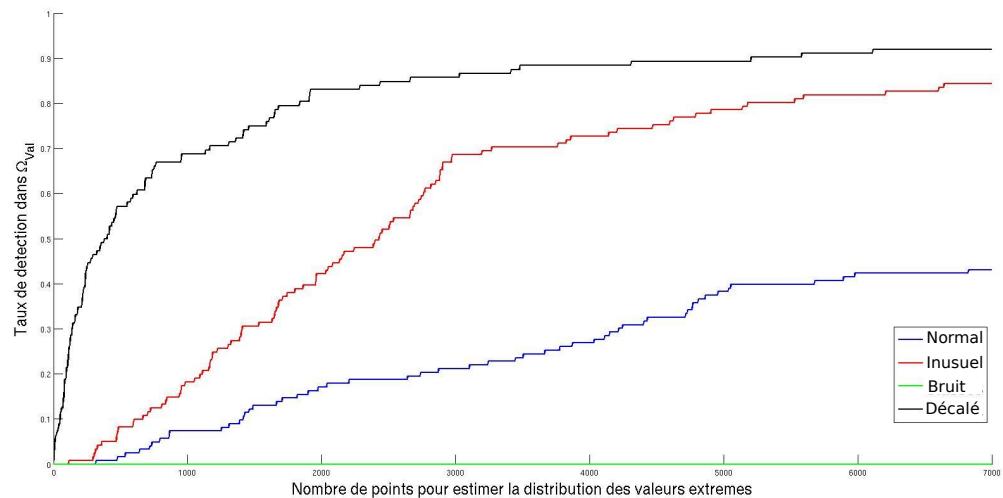


FIGURE 4.9 – Taux de détection pour les différentes classes de points de la base de validation  $\Omega_{Val}$  en fonction du nombre de résidus utilisés pour calibrer la distribution des valeurs extrêmes. Les détections des points atypiques (rouge) et des points se situant sur des raies normales décalées (noir) augmentent fortement en fonction du nombre de points pris en compte pour l'estimation de la distribution des valeurs extrêmes. Le taux de détection des points normaux (bleu) augmente beaucoup plus faiblement. Le nombre de points pris en compte pour l'estimation de la distribution doit donc être important.

extrêmes :

$$n_{max} = 5178.$$

Cette valeur permet un taux de bonnes détections de 80%, tout en limitant le taux de fausses détections des points normaux de la base de validation de  $\mathcal{B}_1$ . Le taux de détection des points décalés reste supérieur. Par manque de points annotés dans les autres patchs, cette valeur est reportée sur les autres.

#### 4.5.4 Résultats sur la base de test $\Omega_{Test}$

##### Résultats sur la base de test annotée ponctuellement

Le tableau 4.2 donne les taux de détection des différentes classes de points annotés de la base de test  $\Omega_{Test}$  de  $\mathcal{B}_1$  obtenus à partir de différentes méthodes de définition de seuil sur les résidus de reconstruction issus du dictionnaire de normalité de la NMF appris sur différentes données d'apprentissage.

TABLE 4.2 – Taux de détection des différentes classes de points de la base de test  $\Omega_{Test}$  à partir du modèle défini par la NMF avec les scores définis par les p-valeurs et les valeurs extrêmes des résidus pour différentes données d'apprentissage

Classe de points	P-valeurs sur les résidus	Valeurs extrêmes sur les résidus
Classe "normal"	$30.8\% \pm 3.6\%$	$35.8\% \pm 3.2\%$
Classe "inusuel"	$79.7\% \pm 1.2\%$	$83.1\% \pm 1.2\%$
Classe "bruit"	0%	0%
Classe "décalé"	$87.8\% \pm 4.4\%$	$88.5\% \pm 3.3\%$

Cette approche donne de bons résultats de détection pour les points atypiques mais détectent aussi les points décalés, les raies  $N_1$  sont donc fortement détectées par cette méthode. Les résultats des deux méthodes de définition de seuil sont très proches. La détection des autres points normaux provient principalement de points isolés ou des points situés sur des raies normales de fortes intensités. Afin de limiter cet effet, nous pouvons appliquer un filtre médian sur les détections pour supprimer les points dont aucun voisin ne fait partie des points détectés.

##### Résultats sur l'ensemble des patchs

Le tableau 4.4 présente des résultats sur l'ensemble des patchs de la base de test en considérant un critère d'erreur identique à celui défini pour les curvelets (3.15). Nous calculons la proportion de points détectés faisant partie d'une zone atypique pour les patchs contenant des signatures inusuelles et le taux de points détectés pour les patchs ne contenant pas de signatures inusuelles. Un point classifié comme inusuel par notre modèle est considéré comme une bonne détection s'il se trouve dans une des zones atypiques extraites de notre base de données. Nous mettons NA pour les patchs dans lesquels aucun moteur de la base de test ne possède de signature inusuelle.

Le tableau 4.4 est divisé en plusieurs sous-tableaux, chacun d'entre eux correspondant à un des patchs. Le contenu de ces sous-tableaux est donné dans le tableau 4.3.

TABLE 4.3 – Structure des sous-tableaux de résultats pour le modèle défini par le dictionnaire de la NMF

Taux de bonnes détection à partir de l'approche des p-valeurs pour les patchs contenant des signatures inusuelles	Taux de points détectés à partir de l'approche des p-valeurs pour les patchs sans signatures inusuelles
Taux de bonnes détection à partir de l'approche des valeurs extrêmes pour les patchs contenant des signatures inusuelles	Taux de points détectés à partir de l'approche des valeurs extrêmes pour les patchs sans signatures inusuelles

Ce tableau permet de voir un bon pourcentage de points détectés pour plusieurs patchs contenant des signatures inusuelles. Certains patchs possèdent cependant un faible taux de détection, il peut s'agir de patchs dont les points considérés comme inusuels constituent les points normaux ajoutés lors de l'extraction ou des points inusuels. De plus, il est à remarquer que ces approches rejettent un certain pourcentage de points normaux dû au seuil fixé intra-patch que ce soit pour les p-valeurs ou les valeurs extrêmes et imposent donc nécessairement des points à classifier comme inusuels sur tous les patchs. Ce pourcentage de détection reste néanmoins faible comparativement à la taille du patch. Ces résultats sont à relativiser car les points détectés dans les zones atypiques peuvent correspondre à des points normaux situés dans ces cadres lors de l'extraction.

#### 4.5.5 Complémentarité des approches adaptatives et non-adaptatives

Les modèles définis à partir des dictionnaires issus de la NMF et des curvelets permettent tous deux la détection des signatures atypiques présentes sur les patchs. Le dictionnaire des curvelets fonctionnent de manière locale sur les patchs tandis que celui de NMF gère les patchs dans leur globalité. Cela est visible à travers les atomes des dictionnaires qui correspondent à des parties de raies pour les curvelets et à des raies entières pour la NMF. Les deux modèles entraînent également la détection de plusieurs points normaux sur les patchs. Le dictionnaire  $\mathcal{D}^{NMF}$  permet la reconstruction des raies au sein d'un patch mais ne donne pas une estimation correcte des intensités vibratoires les plus fortes. Les atomes du dictionnaire caractérisent l'intégralité du patch et non pas une partie seulement. Etant donné que seules les opérations additives sont appliquées, caractériser correctement une raie vibratoire de forte intensité impliquerait une augmentation des coefficients associés aux atomes. Cette augmentation n'impactera pas uniquement la raie vibratoire mais tout l'atome et donc a fortiori l'intégralité du patch augmentant ainsi l'erreur de reconstruction globale. Ainsi, le dictionnaire de la NMF permet la reconstruction de la structure du patch mais pas des intensités, entraînant des détections sur les raies normales de fortes intensités. A l'inverse, les curvelets permettent une bonne reconstruction des intensités vibratoires et de la structure du patch, malgré des raies vibratoires reconstruites de manière plus grossière. Chaque atome du dictionnaire correspond à une sous-zone du patch et donc l'augmentation des coefficients n'impacte pas l'intégralité du patch. Cependant le dictionnaire des curvelets entraîne

TABLE 4.4 – Taux de détection sur l'ensemble des patchs de la base de test à partir du modèle de normalité défini à partir de la NMF

<i>NA</i>	3.55% ± 0.03%	<i>NA</i>	3.93% ± 0.03%	1.8% ± 0.3%	3.49% ± 0.04%
<i>NA</i>	5.10% ± 0.13%	<i>NA</i>	5.32% ± 0.05%	1.4% ± 0.3%	4.85% ± 0.05%
<i>NA</i>	4.02% ± 0.03%	<i>NA</i>	4.25% ± 0.04%	<b>70.8% ± 0.6%</b>	4.16% ± 0.02%
<i>NA</i>	5.62% ± 0.06%	<i>NA</i>	5.73% ± 0.04%	<b>69.5% ± 0.7%</b>	5.67% ± 0.02%
25.5% ± 0.4%	4.27% ± 0.01%	3.5% ± 0.6%	4.1% ± 0.01%	<b>64.5% ± 2.1%</b>	3.46% ± 0.02%
25.7% ± 0.5%	6.14% ± 0.06%	4.3% ± 0.5%	5.43% ± 0.01%	<b>63.1% ± 1.6%</b>	4.5% ± 0.01%
<b>68.5% ± 0.6%</b>	4.1% ± 0.01%	41.6% ± 0.3%	3.44% ± 0.01%	<b>52.1% ± 0.5%</b>	4.28% ± 0.01%
<b>65.8% ± 0.4%</b>	5.62% ± 0.01%	41.7% ± 0.3%	4.3% ± 0.03%	<b>51.2% ± 0.1%</b>	5.97% ± 0.05%
31.9% ± 0.3%	4.40% ± 0.03%	29.8% ± 0.3%	3.19% ± 0.03%	34.1% ± 0.4%	3.66% ± 0.03%
31.1% ± 0.2%	6.12% ± 0.03%	30.3% ± 0.3%	4.06 ± 0.06%	34.3% ± 0.3%	4.49% ± 0.05%
<b>55.7% ± 0.5%</b>	4.68% ± 0.02%	<b>54.4% ± 0.2%</b>	4.38% ± 0.01%	<b>82.2% ± 0.2%</b>	2.17% ± 0.02%
<b>56.0% ± 0.4%</b>	6.63% ± 0.02%	<b>53.3% ± 0.2%</b>	6.31% ± 0.03%	<b>82.1% ± 0.1%</b>	2.40% ± 0.02%
<b>51.1% ± 0.3%</b>	4.61% ± 0.02%	<b>50.5% ± 0.4%</b>	4.41% ± 0.01%	<b>55.8% ± 0.3%</b>	2.10% ± 0.01%
<b>52.6% ± 0.4%</b>	6.74% ± 0.06%	49.6% ± 0.4%	6.38% ± 0.01%	<b>55.5% ± 0.5%</b>	2.49% ± 0.01%
18.5% ± 0.6%	4.41% ± 0.01%	35.0% ± 0.3%	4.02% ± 0.01%	47.9% ± 0.3%	4.36% ± 0.02%
19.0% ± 0.5%	6.27% ± 0.03%	35.1% ± 0.2%	5.27% ± 0.03%	49.4% ± 0.2%	6.29% ± 0.02%
7.9% ± 0.9%	4.61% ± 0.02%	23.6% ± 0.9%	4.12% ± 0.03%	37.3% ± 0.5%	4.35% ± 0.02%
7.13% ± 0.5%	6.65% ± 0.03%	23.5% ± 0.7%	5.77% ± 0.03%	38.0% ± 0.6%	6.37% ± 0.01%
<i>NA</i>	4.73% ± 0.02%	<i>NA</i>	4.24% ± 0.01%	17.9% ± 0.7%	4.20% ± 0.03%
<i>NA</i>	6.64% ± 0.01%	<i>NA</i>	6.10% ± 0.05%	19.8% ± 0.4%	5.73% ± 0.04%
25.8% ± 0.8%	4.53% ± 0.02%	<i>NA</i>	4.35% ± 0.03%	21.9% ± 0.6%	3.87% ± 0.01%
21.7% ± 0.8%	6.70% ± 0.05%	<i>NA</i>	6.14% ± 0.02%	21.0% ± 0.8%	4.89% ± 0.03%
7.0% ± 0.3%	4.26% ± 0.01%	14.1% ± 1.4%	3.26% ± 0.02%	5.5% ± 0.2%	2.99% ± 0.01%
7.8% ± 0.4%	6.31% ± 0.05%	13.8% ± 1.2%	4.03% ± 0.06%	5.5% ± 0%	3.52% ± 0.03%
7.2% ± 0.2%	4.51% ± 0.02%	18.7% ± 0.3%	2.75% ± 0.01%	12.3% ± 0.4%	1.89% ± 0.01%
8.1% ± 0.6%	6.97% ± 0.05%	18.9% ± 0.2%	3.37% ± 0.03%	13.3% ± 0.2%	2.22% ± 0.01%
<i>NA</i>	4.54% ± 0.01%	3.6% ± 0.3%	3.01% ± 0.01%	48.8% ± 0.9%	3.31% ± 0.03%
<i>NA</i>	7.20% ± 0.01%	3.8% ± 0.1%	3.57% ± 0.05%	48.7% ± 0.7%	4.16% ± 0.05%
<i>NA</i>	4.18% ± 0.02%	<b>60.5% ± 0.3%</b>	2.07% ± 0.02%	<b>51.9% ± 0.1%</b>	4.14% ± 0.03%
<i>NA</i>	6.40% ± 0.03%	<b>59.7% ± 0.3%</b>	2.28% ± 0.04%	<b>50.9% ± 0.2%</b>	5.61% ± 0.01%
14.8% ± 4%	4.47% ± 0.01%	45.0% ± 0.7%	2.16% ± 0.02%	42.0% ± 0.3%	3.66% ± 0.01%
16.0% ± 0.2%	6.80% ± 0.02%	44.0% ± 0.5%	2.56% ± 0.05%	41.0% ± 0.4%	4.50% ± 0.01%
<b>92.7% ± 1.3%</b>	3.89% ± 0.03%	45.9% ± 0.8%	2.36% ± 0.01%	16.7% ± 0.1%	3.35% ± 0.02%
<b>91.5% ± 1.4%</b>	6.08% ± 0.12%	45.3% ± 1.1%	3.01% ± 0.04%	16.3% ± 0%	4.05% ± 0.02%
30.0% ± 3.2%	3.21% ± 0.04%	6.3% ± 1.1%	3.05% ± 0.02%	22.4% ± 0.1%	3.28% ± 0.01%
31.6% ± 1.8%	4.83% ± 0.11%	6.1% ± 0.8%	4.29% ± 0.02%	21.6% ± 0.1%	3.90% ± 0.02%

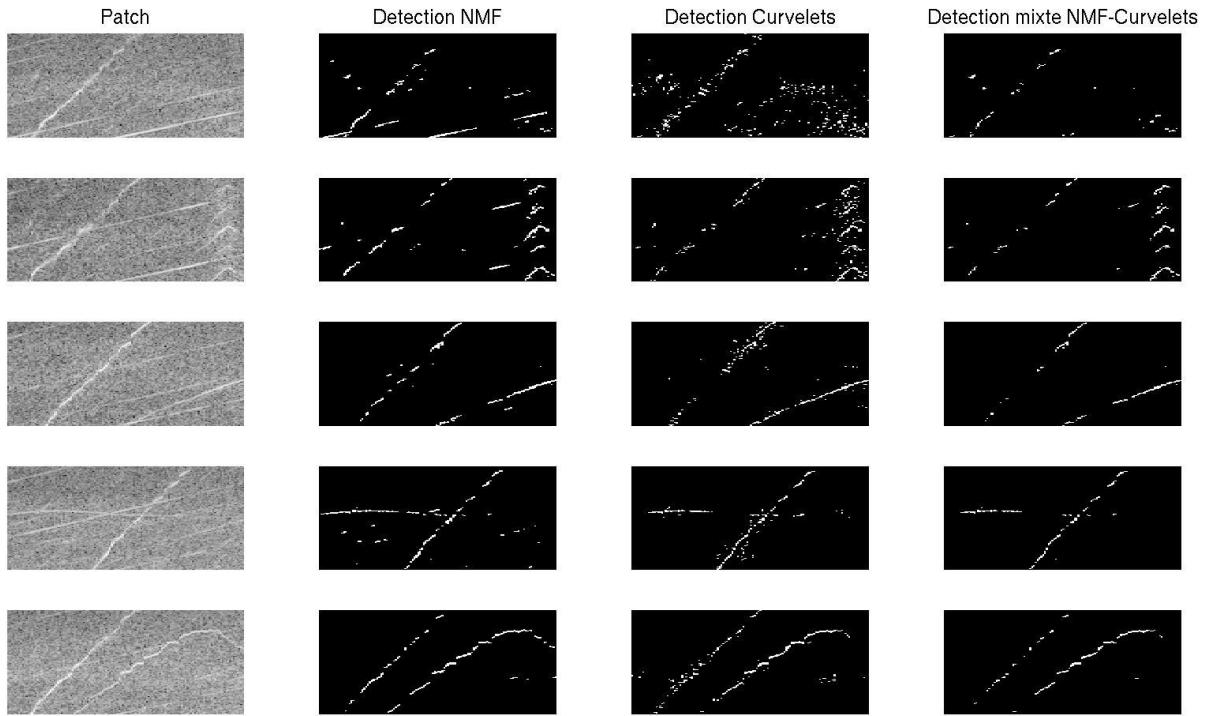


FIGURE 4.10 – Détection des points inusuels à partir des modèles de normalité définis par la NMF, les curvelets et l’approche mixte combinant les 2 dictionnaires.

la détection de plusieurs points autour des signatures atypiques et des raies liées au  $N_1$ . Ainsi les deux dictionnaires permettent la détection des signatures inusuelles mais n’entraînent pas les mêmes fausses détections. Il est donc intéressant de les combiner sous la forme d’un vote pour les limiter. Un point est considéré comme inusuel dans cette approche mixte lorsque les 2 modèles le considèrent comme inusuel.

$$\hat{Y}_{f,N_2}^i = \min(\hat{Y}_{f,N_2}^{i,\mathcal{C}}, \hat{Y}_{f,N_2}^{i,NMF}) \quad \forall(f, N_2)$$

La figure 4.10 présente la détection des points atypiques sur 5 patchs (définis dans la figure 2.3, un patch normal (1ère ligne) et 4 patchs contenant différentes signatures inusuelles) à partir des modèles de normalité définis à partir de la NMF combiné à un filtrage de voisinage, des curvelets et du classifieur combinant ces deux dictionnaires par un système de vote. Les deux dictionnaires permettent la détection des points atypiques, nous pouvons remarquer qu’en mélangeant les deux approches nous réduisons les points détectés à tort et conservons une bonne détection des signatures inusuelles. La combinaison des dictionnaires est donc efficace pour la détection d’anomalies sur les spectrogrammes.

## 4.6 Conclusions

### Le dictionnaire adaptatif

Le dictionnaire adaptatif issu de la NMF permet une bonne caractérisation du modèle de normalité des patchs. Un dictionnaire est appris pour chaque patch séparément et uniquement sur ceux ne présentant aucune signature inusuelle. A partir de ce dictionnaire, les patchs reconstruits constituent une estimation normale de ces derniers. Ils ne présentent plus usuellement de signatures atypiques étayant l'hypothèse d'une bonne caractérisation de la normalité à partir du dictionnaire adaptatif en dimension réduite.

L'étude du patch dans sa globalité ne permet pas la détection des patchs inusuels. Les coefficients de représentation sont en dimension réduite par rapport à la dimension des patchs et n'ont pas la possibilité d'être affecté par les anomalies. Ils permettent uniquement la caractérisation des signatures redondantes au sein du patch. L'erreur de reconstruction globale dilue l'impact issu des signatures atypiques par rapport à l'ensemble des résidus issus des éléments normaux sur le patch. Cette mesure permet la détection des patchs anormaux lorsque les signatures sont de fortes intensités et en grand nombre, c'est-à-dire lors d'un réel endommagement sur le moteur.

La considération des résidus ponctuels de reconstruction permet de considérer les erreurs au niveau de chaque point et donc de détecter ceux dont le modèle de normalité ne donne pas une bonne caractérisation. Cela permet de détecter finement les signatures inusuelles au sein du patch. Cependant, la détection est effectuée à partir des seuils fixés sur les résidus intra-patch, le seuil est fixé sur la distribution des résidus ou à partir de la distribution des valeurs extrêmes. Cela entraîne une détection systémataque des points extrêmes des résidus y compris sur les patchs ne contenant pas de signatures inusuelles.

### Les approches de représentation de la normalité par des dictionnaires

Les approches par dictionnaire ont permis de caractériser la normalité permettant une détection fine des signatures inusuelles à partir de la reconstruction normale des patchs. Les dictionnaires non-adaptatifs permettent de considérer des dictionnaires surcomplets malgré un faible nombre de données par rapport à la taille des patchs. Les atomes du dictionnaire sont prédéfinis et consistent en de petites parties de raies permettant de caractériser les patchs localement. Chaque atome du dictionnaire caractérise donc une zone du patch. Les dictionnaires data-driven sont en dimension réduite pour éviter le sur-apprentissage, les différents atomes caractérisent différentes structures de l'intégralité du patch et permettent une reconstruction des patchs sans aucune signature atypique. Ce type de dictionnaire est donc idéal pour définir le modèle de normalité. La détection des signatures inusuelles s'effectue point par point sur chaque patch en fixant un seuil spécifique aux données sur les résidus de reconstruction des patchs à partir du modèle de normalité. Dans le cas du dictionnaire non-adaptatif, le dictionnaire engendre des

artéfacts autour des signatures inusuelles permettant de les utiliser comme un filtre sur les résidus positifs caractérisant les points inusuels. Dans le cas des dictionnaires adaptatifs, seuls les résidus positifs permettent de détecter les points inusuels. La combinaison des deux dictionnaires apporte une détection plus fine des points atypiques et permet d'effectuer une double vérification sur les dictionnaires afin de limiter les mauvaises détections.

La détection fine des signatures inusuelles s'effectue à partir des résidus intra-patchs et entraîne la détection de points sur tous les patchs contenant ou ne contenant pas de signatures atypiques. Cette détection est effectuée sans tenir compte des autres données associées au même patch. Le seuil de détection est fixé à partir des résidus du patch et est identique pour tous les points présents sur le patch. Cependant, un point de forte intensité peut être mal caractérisé par le dictionnaire du patch et entraîné de forts résidus de représentation. Cependant bien que l'intensité de ce point peut être considérée comme importante par rapport au patch étudié, elle peut ne pas être inusuelle par rapport aux points identiques sur les données normales. Ce point n'est donc une anomalie mais est considéré comme tel par l'algorithme de détection basé sur les dictionnaires comme ce point est vu comme un extrême de la distribution des résidus. Il semble donc pertinent de considérer un modèle de normalité au niveau de chaque point des différents patchs. Cela permettrait de tenir compte de la variabilité des mêmes points sur les différentes données afin de détecter uniquement les points dont l'intensité vibratoire ne correspond pas à la distribution des points normaux aux mêmes coordonnées.

Les approches par dictionnaire permettent donc une bonne détection des signatures atypiques sur les différents patchs du spectrogramme sans a priori sur la forme ou position de ces signatures inusuelles.

## Troisième partie

# Analyse ponctuelle des spectrogrammes



## Introduction

Dans la partie précédente, nous avons montré que l'étude des résidus ponctuels permet une détection plus fine des signatures atypiques sur les patchs par rapport à une étude globale de ces derniers. Chaque point de ces données à régime et fréquence fixés correspondant à une intensité vibratoire. Considérer les points du spectrogramme ponctuellement permet donc de se rapprocher de la physique en étudiant chaque vibration à une fréquence et un régime précis. Dans cette partie, nous considérons le spectrogramme ponctuellement. Chaque point du spectrogramme est donc assimilé à une donnée à part entière, à cette échelle la base de données est plus fortement déséquilibrée qu'à l'échelle du patch. Chaque point possède un nombre limité de données labélisées comme inusuelles. Nous restons donc dans un contexte de détection de nouveautés.

Dans cette partie, nous proposons d'étudier un modèle de normalité pour chaque point du spectrogramme. Ce modèle de normalité est basé sur la distribution des données normales du point considéré. Cette distribution est estimée de manière non paramétrique à partir de l'estimation de la densité par noyau [49] sur les points normaux. Nous définissons différents modèles dans cette partie correspondant à la relation entre chaque point et leurs voisins :

- en considérant les points des spectrogrammes comme indépendants,
- en considérant la dépendance des points par rapport à tous leurs voisins d'ordre 1 (voisins directs à distance 1 du point étudié),
- en considérant la dépendance des points par rapport à leurs voisins dans différentes directions et de distance supérieure à 1.

La complexité de ce type d'approche est plus importante que celles des approches par dictionnaire étant donné qu'un modèle de normalité doit être défini pour chaque point. Les temps de calculs restent convenables grâce à la parallélisation de l'étude de chaque point et aux choix du noyau permettant une simplification des calculs.

Dans cette partie, nous montrons la pertinence des approches ponctuelles pour la détection d'anomalies sur les spectrogrammes et soulignons la complémentarité de ces approches avec les méthodes par dictionnaire. Nous caractérisons également les signatures atypiques à partir des approches ponctuelles.



# Chapitre 5

## Analyse ponctuelle indépendante des spectrogrammes

### 5.1 Introduction

#### 5.1.1 Considération ponctuelle des points des spectrogrammes

Dans cette partie, nous considérons une modélisation au niveau des points du spectrogramme afin de détecter les points inusuels à partir des comportements normaux des points aux mêmes coordonnées. En effet les spectrogrammes correspondent à des mesures physiques où chaque point est une intensité vibratoire à une fréquence  $f$  donnée et un régime  $N_2$  donné. Nous étudions donc chaque point des spectrogrammes séparément et définissons un modèle de normalité en 1 dimension pour chacun d'entre eux (Figure 5.1). Il s'agit là d'une autre subdivision du spectrogramme en patch de taille  $1 \times 1$  paramétré par le couple de coordonnées  $(f, N_2)$  associé au point étudié. Cette interprétation du spectrogramme est en adéquation avec la physique des vibrations. Nous considérons chaque point comme une mesure vibratoire du moteurs dont l'intensité aux mêmes fréquence et régime ne doit pas s'écarte d'une normalité à définir. Nous disposons donc d'un jeu de données unidimensionnel pour chaque point du spectrogramme.

$$\{S_{f,N_2}^i\}_{i=1,\dots,n} \quad \forall(f, N_2)$$

A partir des dictionnaires, les méthodes de détection ponctuelle étaient basées sur les résidus de reconstruction intrinsèques au patch. La détection d'un point était donc fortement liée aux résidus de l'ensemble des points du patch, y compris des points indépendants du point considéré. Dans cette partie, la détection est effectuée en fonction des points aux mêmes coordonnées dans les données normales. En se figurant le jeu de données des patchs comme une grande matrice en 3 dimensions symbolisant la largeur et la hauteur du patch et la profondeur correspondant au nombre de données, il est possible de visualiser les algorithmes de détection par dictionnaire

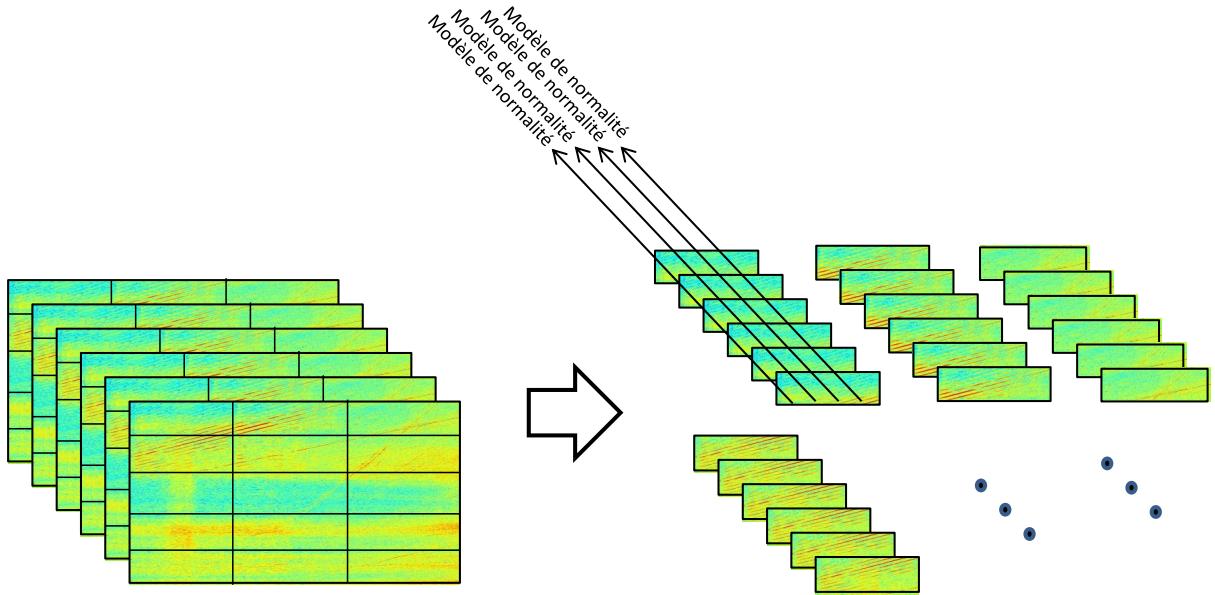


FIGURE 5.1 – Analyse ponctuelle des points du spectrogramme par patch.

comme agissant sur toutes les dimensions pour calibrer le dictionnaire et uniquement sur les deux premières dimensions pour la détection sans tenir compte de la 3ème dimension de profondeur. Les approches développées dans ce chapitre agissent uniquement sur la dimension de profondeur pour la définition du modèle de normalité et la détection (figure 5.1) et dans le chapitre suivant en fonction des trois dimensions.

### 5.1.2 La base de données

Nous ne disposons pas de l'annotation de l'ensemble des points des spectrogrammes. Nous utilisons la base de données construite contenant les zones atypiques (cf section 1.4) afin de définir pour un point donné l'ensemble des spectrogrammes dont le point constitue une vibration normale et l'ensemble des spectrogrammes dont l'une des zones atypiques sur le spectrogramme contient le point étudié. Nous notons  $\Omega^{f,N_2}$  l'ensemble des points aux coordonnées  $(f, N_2)$  sur les spectrogrammes de la base de données. Nous notons  $Y_{f,N_2}^i$  la vérité terrain dans la base de données sur l'appartenance du point aux coordonnées  $(f, N_2)$  du spectrogramme  $i$  à une zone atypique.

$$\forall (f, N_2) \in \mathcal{K}_j, Y_{f,N_2}^i = \begin{cases} 1 & \text{si } \exists k : (f, N_2) \in \text{zone}_{ano}^i(k) \\ 0 & \text{sinon} \end{cases}$$

$\text{zone}_{ano}$  correspond à l'ensemble des zones atypiques extraites dans la base de données de spectrogrammes et  $\mathcal{K}_j$  à la subdivision à laquelle appartient le couple  $(f, N_2)$ . Par simplification, nous considérons que tous les points dans les zones atypiques extraites sont eux-mêmes inusuels malgré la présence de plusieurs points normaux. Nous définissons l'erreur  $E_j^i$  (avec  $i$  le spectrogramme et  $j$  l'élément de la subdivision en patchs) de manière identique à (3.15) comme le pourcentage de bonne détection se trouvant dans ces zones atypiques pour les patchs contenant des signatures inusuelles et le pourcentage de points détectés (fausse détection) pour les patchs

sans signature inusuelle.

Nous disposons également de la base de données  $\mathcal{B}_1$  contenant la labélisation par classe (cf section 1.5.3) de points sur un patch spécifique à partir de laquelle nous calibrons les modèles mis en place et présentons les résultats. Nous fournissons également des résultats visuels de détection.

### 5.1.3 Les modèles de normalité

Dans ce chapitre nous modélisons la normalité par la distribution estimée sur des points normaux. Les différents points des spectrogrammes sont considérés indépendants. Différentes approches ont été étudiées :

- la modélisation paramétrique des points par une loi gamma
- la modélisation non-paramétrique de la distribution des points par l'estimation de densité par noyau [49]. Nous avons sélectionné pour cela deux noyaux différents qui offrent quelques propriétés intéressantes :
  - le noyau gaussien,
  - le noyau gamma [29].

La nouveauté est décidée par un test statistique d'adéquation du point à la distribution estimée. Cela signifie donc un nombre de tests égal au nombre de points sur le patch étudié entraînant alors des problématiques de tests multiples [96].

## 5.2 Modélisation paramétrique de la distribution de normalité

### 5.2.1 Le modèle de normalité

La modélisation des points du spectrogramme par des distributions permet de fixer des seuils de détection sur chaque point indépendamment du point considéré. Dans [63], les points des spectrogrammes sont normalisés et comparés à la moyenne de points normaux pour détecter les points inusuels. La figure 5.2 présente les histogrammes de différents points normaux sélectionnés aléatoirement. L'allure générale de la distribution des points est proche des lois gamma. Dans un premier temps, nous modélisons de manière paramétrique ces distributions par des lois gamma afin d'expliquer les procédures de détection des points inusuels et d'étudier les tests multiples [96] pour la détection d'anomalies sur les spectrogrammes. Un test de Mann-Whitney [85] à 5% confirme l'hypothèse de loi gamma pour la plupart des points du spectrogramme.

Les lois gamma  $\hat{\Gamma}_{f,N_2}$ , dont les paramètres sont estimés sur la base d'apprentissage de points normaux  $\Omega_{App}^{f,N_2}$ , constituent le modèle de normalité sur lequel nous nous basons dans cette section pour détecter les points inusuels. L'étude d'un patch du spectrogramme passe donc par la

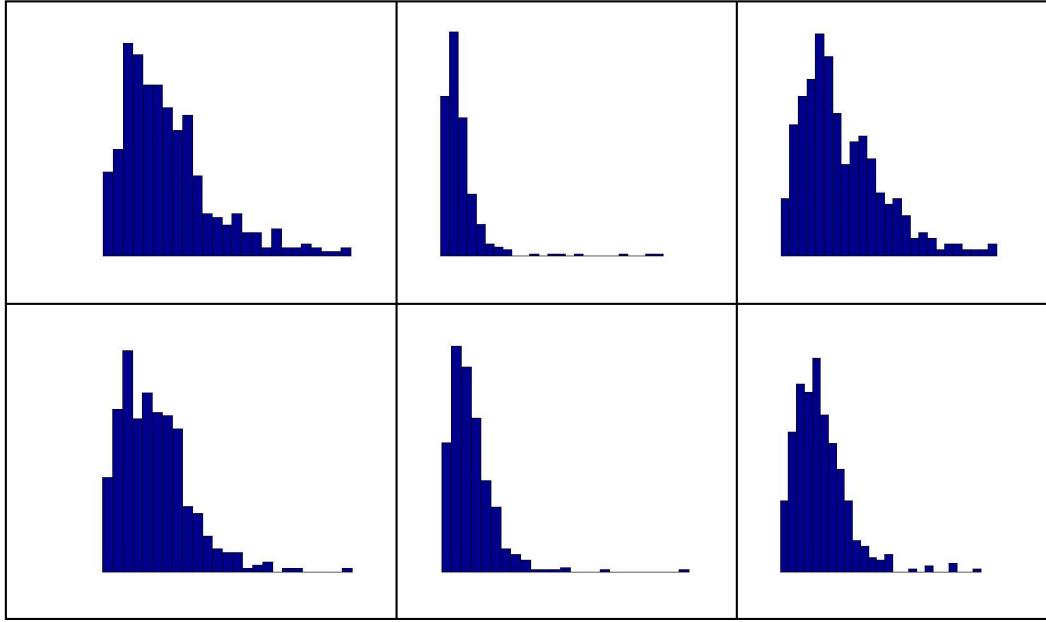


FIGURE 5.2 – Histogrammes des intensités des différents points du spectrogramme. Ces histogrammes ont des allures proches des distributions gamma.

caractérisation de chaque point du patch par une loi gamma.

$$\{\hat{\Gamma}_{f,N_2}\}_{(f,N_2) \in \mathcal{K}_j}$$

### 5.2.2 Le score de détection

Le score de détection d'un point  $S_{f,N_2}^i$  du spectrogramme correspond au test statistique opposant l'hypothèse  $\mathcal{H}_0^{f,N_2}$  selon laquelle le point  $S_{f,N_2}^i$  est normal à l'hypothèse  $\mathcal{H}_1^{f,N_2}$  d'anomalie du point. Ce test correspond à un test d'adéquation de l'intensité vibratoire du point aux coordonnées  $(f, N_2)$  à la loi sous  $\mathcal{H}_0^{f,N_2}$ . Cette loi correspond à  $\hat{\Gamma}_{f,N_2}$ , la distribution estimée des points normaux aux mêmes coordonnées. La p-valeur (5.1) associée à ce test correspond au score de détection sur lequel un seuil est fixé pour classifier un point comme normal ou atypique.

$$pval_{f,N_2}^i = \mathbb{P}_{\mathcal{H}_0^{f,N_2}}(X > S_{f,N_2}^i) \approx \mathbb{P}_{\hat{\Gamma}_{f,N_2}}(X > S_{f,N_2}^i) = 1 - F_{\hat{\Gamma}_{f,N_2}}(S_{f,N_2}^i) \quad (5.1)$$

avec  $X$  une variable aléatoire suivant la loi sous  $\mathcal{H}_0^{f,N_2}$  correspondant à la loi gamma estimée  $\hat{\Gamma}_{f,N_2}$  sur les points normaux et  $F_{\hat{\Gamma}_{f,N_2}}$  la fonction de répartition associée.

La classification des points correspond alors à la comparaison des p-valeurs au seuil  $s$  de détection qui reste à calibrer.

$$\tilde{Y}_{f,N_2}^i = \mathbf{1}\{pval_{f,N_2}^{i,\hat{\Gamma}_{f,N_2}} \leq s\} \quad (5.2)$$

Le processus de décision à partir de (5.2) considère les tests comme indépendants et ne

tient pas compte de la multiplicité de ces derniers. D'après la théorie des tests multiples [96], la probabilité de commettre au moins une fausse détection sur un patch est équivalente à 1 étant donné la dimension  $128 \times 128$  du patch correspondant au nombre de tests. Les procédures de tests multiples permettent limiter les fausses détections sur l'ensemble des tests réalisés simultanément en contrôlant à des niveaux souhaités des grandeurs telles que la probabilité de commettre au moins une fausse détection (FWER) ou la proportion moyenne de fausses détections sur l'ensemble des tests (FDR). Une introduction aux tests multiples est donnée en Annexe A. Nous associons les procédures de tests multiples à notre méthode de détection. Les seuils de décision sont paramétrés par le niveau  $\alpha$  du test considéré. Ce niveau est le même pour les différentes approches. Il est estimé sur la base de validation  $\Omega_{Val}$  sans tenir compte de la multiplicité des tests. Ainsi chaque test est contrôlé unitairement au niveau  $\alpha$ . Les approches de tests multiples permettent dans un second temps de contrôler l'ensemble des tests en tenant compte de leur multiplicité à ce même niveau  $\alpha$ . Nous avons choisi de comparer différentes approches de tests multiples pour la détection d'anomalies :

- une approche unitaire où le seuil de détection sur les p-valeurs correspondant au niveau des tests est déterminé sans tenir compte de la multiplicité,
- une approche contrôlant le Family-wise error rate (FWER) à partir de la procédure de Bonferroni [59], il s'agit de diminuer le niveau des tests d'un facteur correspondant au nombre de tests, il s'agit d'une approche très conservatrice et non adaptative car le seuil de détection est le même pour toutes les données,
- une approche contrôlant le FWER à partir de la procédure de Romano-Wolf [95], la méthode peut être considérée comme adaptative comme le seuil de détection varie selon les données étudiées,
- un approche contrôlant le False discovery rate (FDR) à partir de la procédure Benjamini-Hochberg (BH) [14] qui est par nature adaptative aux données avec un seuil de détection dépendant de ces dernières. Le seuil de décision est défini comme la dernière intersection entre la droite de pente  $\alpha/m$  et la courbe des p-valeurs ordonnées.

### 5.2.3 Calibration des seuils de détection sur la base de validation $\Omega_{Val}$

Le niveau des tests est déterminé par rapport au taux de détection sur les tests unitaires des différentes classes de points sur la base de validation  $\Omega_{Val}$  de  $\mathcal{B}_1$ , nous utilisons ensuite les tests multiples à ce même niveau. Les résultats de détection sont fournis sur la base de test  $\Omega_{Test}$ .

Le seuil de détection  $s$  (correspondant au niveau  $\alpha$  des tests unitaires) est déterminé en comparant les taux de détections des différentes classes de points à partir des p-valeurs des tests pour différentes niveaux  $\alpha = s$  (Figure 5.3). La caractérisation de chaque point du spectrogramme par la loi gamma permet un fort taux de détection des points inusuels pour de petites valeurs de  $s$ . Cependant, comme pour les méthodes de dictionnaire, cette approche entraîne aussi un fort taux de détection des points situés sur des raies décalées sur les différents spectrogrammes. Les niveaux de détection des autres classes sont assez faibles et montrent une bonne caractérisation des points normaux. Nous choisissons comme seuil de détection  $\alpha = s = 0.1$  sur les p-valeurs

indépendantes permettant une bonne détection tout en limitant les fausses détections sur la base de validation  $\Omega_{Val}$ .

Ce niveau est également appliqué pour les procédures de tests multiples. Le seuil sur les p-valeurs de l'approche FWER non-adaptative est alors égal à  $s/128^2$ . Ce seuil est donc très faible et entraîne une approche très conservatrice. Les seuils de décision des approches FWER adaptative et du FDR s'adaptent automatiquement suivant les données.

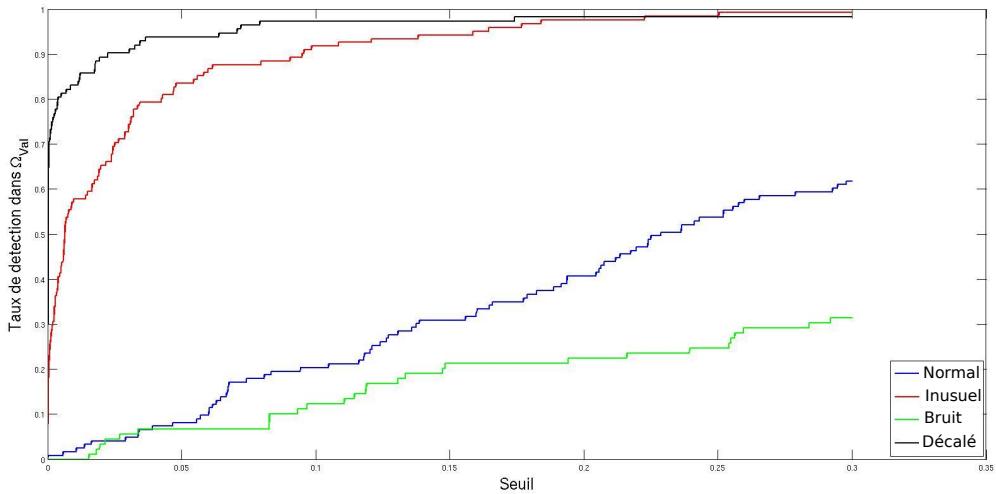


FIGURE 5.3 – Détection des différentes classes de points à partir de la modélisation de chaque point du spectrogramme par une distribution Gamma en fonction du niveau des tests unitaires. Les points inusuels et les points normaux situés sur des signatures décalées sont fortement détectés pour de petites valeurs de seuil de décision. Les points normaux et le bruit sont faiblement détectés pour ces mêmes seuils. Nous définissons donc un seuil de détection au niveau de ces valeurs.

#### 5.2.4 Résultats sur la base de test $\Omega_{Test}$

##### Résultats des tests multiples

La figure 5.4 présente pour plusieurs patchs contenant des points atypiques la détection de ces derniers à partir des 4 approches proposées. A partir du niveau de test choisi, les signatures inusuelles sont totalement ou partiellement visibles dans les points détectés des tests unitaires (colonne 2). Cependant plusieurs points normaux sont également détectés consistant en de fausses détections. Les approches de tests multiples sont bien plus conservatrices. Les approches de FWER et FWER adaptative (colonne 3 et 4) sont très conservatrices avec une mauvaise détection des signatures inusuelles. L'approche FDR (colonne 5) est moins conservatrice et permet la détection d'une partie des signatures inusuelles pour certains patchs, ceux dont les points inusuels sont d'intensités conséquentes par rapport aux mêmes points normaux. Cependant une bonne partie des signatures inusuelles sont elles aussi non détectées. Du fait de leurs décalages et de leurs intensités vibratoires très importantes, les raies  $N_1$  sont fortement détectées. Les points

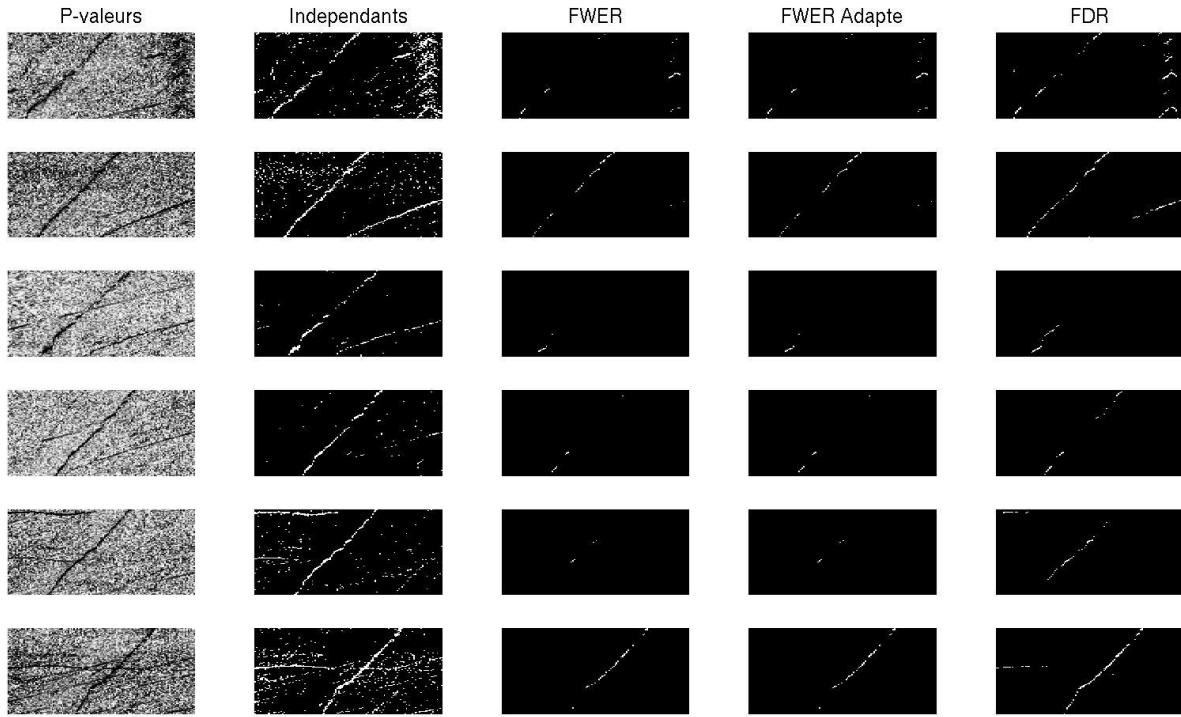


FIGURE 5.4 – Détection des points atypiques sur des patchs contenant des signatures inusuelles par un seuillage sur les p-valeurs des tests statistiques au niveau de chaque point. Le seuil de détection est défini par les approches indépendantes, FWER, FWER avec une valeur du seuil adaptée aux données et FDR. Les signatures sont visuellement détectables à partir de l'approche indépendante. Les approches basées sur le FWER sont trop conservatrices et ne permettent pas la détection des signatures atypiques. Le FDR permet la détection de certaines signatures sur les patchs mais restent également trop conservateur.

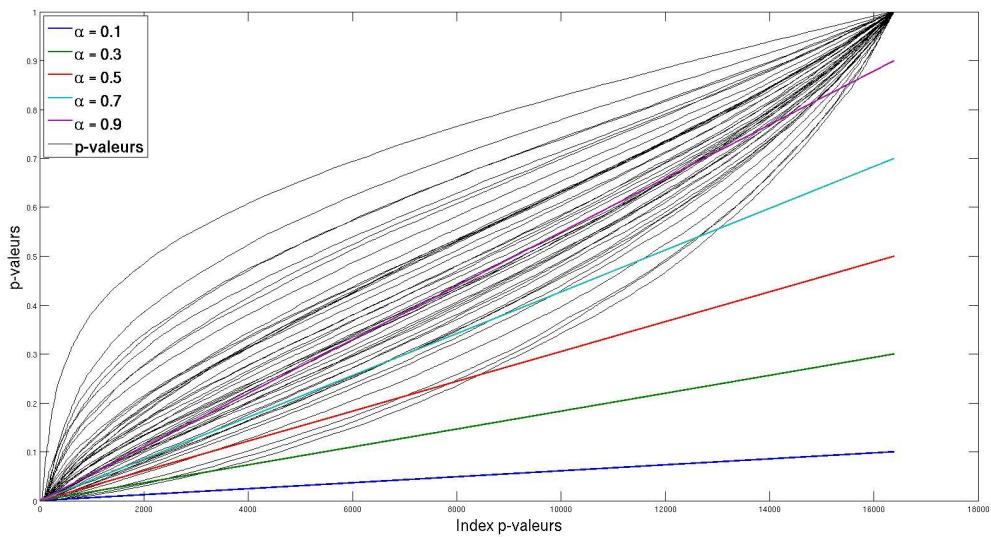


FIGURE 5.5 – Courbes des p-valeurs ordonnées des mêmes patchs de différents spectrogrammes. Comparaison aux droites d'équation  $\alpha/m$  pour  $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ . Le seuil de détection de l'approche basée sur le FDR correspond à la dernière intersection entre ces la courbe des p-valeurs ordonnées et la droite d'équation  $\alpha/m$ . Les courbes des p-valeurs ordonnées possèdent différents comportements (convexe ou concave) entraînant une grande variabilité dans les détections.

situés sur ces raies ont des valeurs de densités très faibles car ces densités sont calibrées sur des points ne faisant pas partie en général de signatures vibratoires liées au  $N_1$ . Les tests statistiques ont donc plus fortement tendance à rejeter l'adéquation de ces points aux distributions estimées.

Les approches de tests multiples sont donc trop conservatrices pour détecter les signatures inusuelles sur les patchs. Ces approches ont tendance à accepter le plus possible les hypothèses  $\mathcal{H}_0^i$  plutôt que les rejeter à tort. Ce comportement n'est pas compatible avec la détection d'anomalies sur les moteurs d'avions où nous cherchons à déterminer l'ensemble des signatures inusuelles présentes. Les méthodes basées sur le FWER sont bien trop conservatrices pour être mises en place car le nombre de tests diminue fortement le seuil de détection. Les approches basées sur le FDR permettent la définition d'un seuil adapté aux données suivant le niveau voulu. L'approche est basée sur le comportement de la courbe des p-valeurs ordonnées. Cependant ces courbes n'ont pas les mêmes caractéristiques suivant les différents spectrogrammes dont les patchs sont extraits. Certaines des courbes des p-valeurs sont fortement concaves entraînant des seuils de détection très faible car l'intersection avec la droite de pente  $\alpha/m$  s'effectue très tôt. Très peu de tests rejettent alors l'hypothèse  $\mathcal{H}_0^i$  et l'approche peut-être considérée comme conservatrice. D'autres patchs possèdent des courbes de p-valeurs convexes entraînant une valeur importante du seuil de détection, l'approche a alors tendance à rejeter la majorité des tests sur le patch étudié ce qui n'est pas réaliste pour nos données de spectrogrammes. La figure 5.5 présente les courbes de p-valeurs ordonnées du même patch de différents spectrogrammes avec les droites de pente  $\alpha/m$  pour différentes valeurs de  $\alpha$ . Nous pouvons voir les différents comportements des courbes des p-valeurs ordonnées aussi bien convexes que concaves entraînant une grande variabilité du seuil de détection sur les patchs.

## Résultats des tests unitaires

Les approches de tests multiples étudiées ne sont pas pertinentes pour la détection de signatures inusuelles sur les données vibratoires des moteurs d'avions. La considération des tests unitaires (avec définition des seuils sur la base de validation  $\Omega_{Val}$ ) est donc plus pertinente et permet la mise en évidence des signatures inusuelles malgré plusieurs fausses détections (Figure 5.4). Afin de limiter les fausses détections, un filtrage à partir du voisinage est mis en place de manière similaire à (4.12). Un point est considéré comme atypique si et seulement sa p-valeur associée est inférieure au seuil de détection et si deux points parmi son voisinage direct  $\mathcal{V}$  (points à distance de 1 du point considéré) sont également détectés (5.3). Les signatures correspondant à des raies sur les spectrogrammes, cela permet d'éliminer les détections des points isolés.

$$\hat{Y}_{f,N_2}^i = \begin{cases} 1 & \text{si } \tilde{Y}_{f,N_2}^i \& \sum_{k \in \mathcal{V}_{f,N_2}} \tilde{Y}_k^i \geq 2 \\ 0 & \text{sinon} \end{cases} \quad (5.3)$$

$\tilde{Y}_{f,N_2}^i$  correspond à la décision par test unitaire sans filtrage défini dans (5.2).

Le tableau 5.1 présente le taux de détections pour les différentes classes de points de la base

de test  $\Omega_{Test}$  annotée de  $\mathcal{B}_1$  à partir des tests unitaires avec et sans filtrage pour différentes données d'estimation de la distribution. La méthode permet une très bonne détection des points atypiques mais aussi des points des signatures vibratoires liées au  $N_1$ . Les points liés au bruit ne sont pas détectés et les points normaux liés au  $N_2$  sont partiellement détectés (il s'agit de fausses détections). Ce taux de détection reste inférieur à celui défini à partir du dictionnaire de la NMF. Le filtrage permet une légère diminution des taux de détection sur les différentes classes, son effet est plus visible sur une représentation visuelle des résultats.

TABLE 5.1 – Taux de détection à partir de la modélisation gamma des points avec et sans filtrage des différentes classes de points de la base de test  $\Omega_{Test}$  pour différentes données d'apprentissage

Classe de points	P-valeurs Gamma sans filtrage	P-valeurs Gamma avec filtrage
Classe "normal"	$20.0\% \pm 2.15\%$	$17.44\% \pm 2.15\%$
Classe "inusuel"	$87.0\% \pm 1.27\%$	$81.40\% \pm 2.85\%$
Classe "bruit"	$3.12\% \pm 0\%$	$3.12\% \pm 0\%$
Classe "décalé"	100%	$94.36\% \pm 2.15\%$

La figure 5.6 présente les résultats visuels de détection pour différents patchs de la base de test (patchs utilisés pour illustrer les résultats dans la partie précédente) sans filtrage et avec filtrage. Le premier patch (1ère ligne) consiste en un patch sans signatures inusuelles, les suivants contiennent tous différentes signatures inusuelles afin de vérifier que la détection n'est pas dépendante du type de signature. Pour les différents patchs, les signatures inusuelles sont mises en évidence, il ne s'agit pas nécessairement de la signature inusuelle dans sa globalité. L'effet du filtrage est également plus visible sur le patch complet avec une forte diminution des points détectés à tort, tandis que les signatures inusuelles restent détectées. Seule la raie liée au  $N_1$  est détectée pour le patch normal (1ère ligne). Visuellement, la définition d'un modèle de normalité sur chaque point du spectrogramme (sans tenir compte de la multiplicité) à partir d'une distribution permet une bonne caractérisation de la normalité et la détection les signatures inusuelles.

La modélisation unitaire par loi gamma est aisée à mettre en place et donne de bons résultats. Cependant cette loi n'est pas valide pour plusieurs points du spectrogramme. La figure 5.7 donne l'ensemble des points rejetant l'hypothèse gamma du point à partir d'un test de Mann-Whitney au niveau 5%. Le choix de la loi gamma a été effectué à but exploratoire en observant les histogrammes de différents points. Cette distribution était pratique pour étudier le comportement des tests multiples dans notre problématique. Le choix de loi gamma suppose une distribution unimodale des points, cette affirmation est correcte dans notre base de données du fait du faible nombre de données. Dans les cas d'une base de données beaucoup plus volumineuse, cette affirmation est fausse. En effet, les signatures normales ne sont pas toujours présentes ou complètes sur les spectrogrammes, elles peuvent également être décalées. De ce fait, il est fort possible que la distribution des points faisant partie des signatures soit bimodale, un mode correspondant au passage d'une raie vibratoire par ce point sur le spectrogramme et un autre à son absence. Dans la suite du manuscrit, la distribution des points est estimée sur des données normales de manière non paramétrique sans a priori afin de définir au mieux cette distribution.

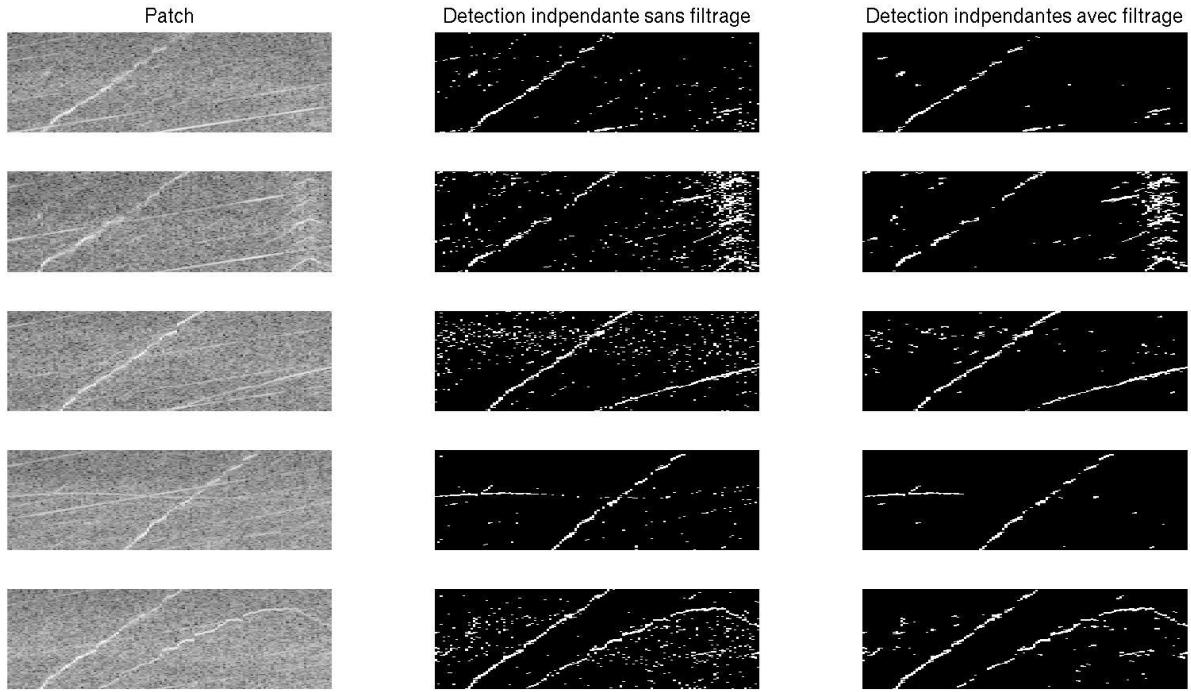


FIGURE 5.6 – Détection à partir de la modélisation gamma des points de différents patchs avec (3ème colonne) et sans filtrage (2ème colonne). Le premier patch ne contient pas de signatures inusuelles, les suivants contiennent différentes signatures inusuelles. La méthode permet une bonne mise en évidence des signatures atypiques sur les patchs. L’application du filtrage de voisinage réduit les fausses détections et ne détériore pas la détection des signatures atypiques.



FIGURE 5.7 – Résultats des tests statistiques de Mann-Whitney vérifiant l’hypothèse gamma pour chaque point du spectrogramme. Les points blancs sont les points rejetant l’hypothèse gamma. Plusieurs points rejettent l’hypothèse de caractérisation de la normalité par une loi gamma.

## 5.3 Estimation non paramétrique de la densité par noyau

### 5.3.1 Formulation

L'estimation de densité par noyau (ou fenêtre de Parzen) [89, 49] permet de caractériser la distribution d'un jeu de données de manière non paramétrique. En considérant un ensemble de variables aléatoires indépendantes et identiquement distribuées  $X = X_1, \dots, X_n$ , l'estimation  $\hat{f}_X$  de la densité  $f_X$  par noyau de ce jeu de données est donnée par :

$$\hat{f}_{X,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (5.4)$$

$K$  correspond au noyau utilisé pour caractériser la distribution et  $h$  à l'échelle appliquée au noyau. Toutes les observations contribuent à l'estimation de densité en tout point  $x$ . La valeur de la densité en un point  $x$  correspond à la moyenne des différentes contributions. Les zones ayant peu de points dans  $(X_1, \dots, X_n)$  sont de densités faibles, à l'inverse les zones contenant de nombreux points dans  $X$  ont des valeurs de densités importantes (Figure 5.8). Les noyaux utilisés pour l'estimation sont le plus souvent symétriques.

Cette méthode est une version continue de l'estimation de densité par histogramme. A l'inverse de l'estimation par histogramme limitée par la valeur maximale des données, la méthode à noyau permet de donner une valeur de la densité en tout point de l'espace. Sans perte de généralité, une autre façon de voir l'estimation de densité à partir de noyau symétrique consiste à considérer une application du noyau centré sur la valeur  $x$  sur les différentes observations de  $X$ . La valeur de la densité en  $x$  correspond alors à la moyenne des applications et l'estimateur

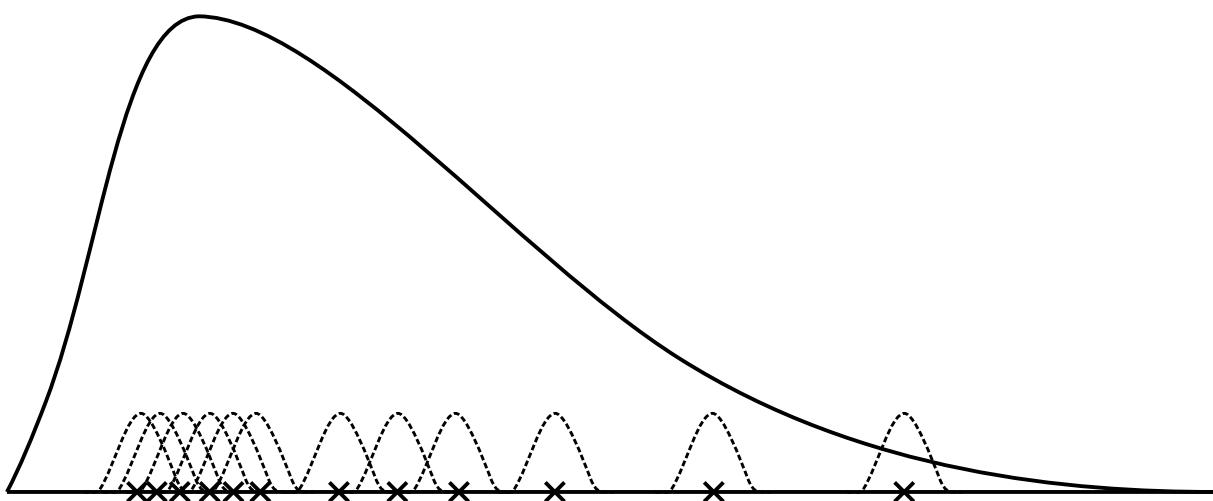


FIGURE 5.8 – Illustration de l'estimation de densité par noyau. Les croix correspondent aux points  $X = (X_1, \dots, X_n)$ , les distributions associées à chacun de ces points (en pointillées) aux noyaux et la distribution globale (trait plein) à la l'estimation de la distribution des données  $X$ .

peut se réécrire :

$$\hat{f}_{X,h}(x) = \frac{1}{nh} \sum_{i=1}^n K_{x,h}(X_i), \quad (5.5)$$

avec  $K_{x,h}$  le noyau centré  $x$  et d'échelle  $h$ .

Le noyau le plus couramment utilisé est le noyau gaussien (5.6) symétrique et permettant des facilités de calcul.

$$K_N(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \quad (5.6)$$

Avec ce type de noyau, l'approche est similaire à un mélange de gaussiennes [15] où le nombre de mélanges correspond au nombre de points d'apprentissage, les moyennes des différentes gaussiennes du mélange sont les points d'apprentissage et les variances correspondent aux échelles des noyaux.

Le biais et la variance de l'estimateur (5.4) s'expriment lorsque la distribution à estimer est suffisamment lisse et que le moment d'ordre 4 du noyau est fini [102] :

$$\begin{aligned} biais(\hat{f}_{X,h}(x)) &= \frac{h^2}{2} \mu_2(K) f''(x) + o(h^2), \\ var(\hat{f}_{X,h}(x)) &= \frac{1}{nh} R(K) f(x) + o\left(\frac{1}{nh}\right), \end{aligned}$$

avec  $\mu_2(K) = \int y^2 K(y) dy$  et  $R(K) = \int K^2(y) dy$ .

Le choix du noyau n'a pas une influence forte sur l'estimation de la densité. Les noyaux symétriques permettent l'estimation des distributions sur  $[0, +\infty[$  mais entraînent un biais à l'origine car la densité est non-nulle pour les valeurs négatives. Certains noyaux, comme le noyau gamma, permettent de caractériser les distributions sur les valeurs positives uniquement. L'échelle  $h$  attribuée au noyau joue un rôle fondamental dans l'estimation. Plus cette échelle est petite, plus les noyaux sont concentrés autour des données de calibration de la densité, les différentes contributions sont alors généralement faibles. On observe alors un surapprentissage de la densité. Si cette échelle est trop grande, les noyaux s'aplatissent donnant alors des contributions significatives pour les points éloignés des points d'apprentissage, la densité est alors totalement lissée avec des modes de la distribution mal caractérisées.

### 5.3.2 L'estimation de la densité par noyau gamma

Le noyau gamma [29] permet une bonne caractérisation des distributions sur  $[0, +\infty[$  et de ne pas avoir de biais à l'origine, il atteint le niveau de convergence optimal parmi les estimateurs de densité positive [29]. Une première formulation de l'estimation de densité par noyau gamma  $\hat{f}^{K_\Gamma}$  pour un jeu de données  $X = (X_1, \dots, X_n)$  est donnée par :

$$\hat{f}_{X,h}^{K_\Gamma}(x) = \frac{1}{n} \sum_{i=1}^n K_{\frac{x}{h}+1,h}(X_i), \quad x \in \mathbb{R}_+$$

avec  $h$  l'échelle du noyau et  $K_{\frac{x}{h}+1,h}$  le noyau gamma.

$$K_{u,v}(t) = \frac{t^{u-1} e^{-\frac{t}{v}}}{v^u \Gamma(u)} = \Gamma(u, v), \quad (u, v) \in \mathbb{R}_+ \times \mathbb{R}_+$$

Cette formulation est similaire à (5.5) du noyau gaussien et peut donc être vue comme l'application du noyau gamma paramétré par  $x$  à tous les points de  $X = (X_1, \dots, X_n)$  utilisés pour définir la distribution.

A la différence du noyau gaussien ayant la même forme pour les points sur lesquels il est appliqué, la forme du noyau gamma change pour chaque point. Une seconde différence par rapport au noyau gaussien provient de sa non-symétrie autour de la valeur  $x$ . Pour le noyau gaussien, cette valeur  $x$  pouvait être vue comme la moyenne du noyau. Cependant dans le cas du noyau gamma, la valeur  $x$  ne correspond pas à la moyenne mais au mode du noyau. Il est possible de modifier les paramètres du noyau en  $\Gamma(x/h, h)$  de telle sorte que  $x$  en soit la moyenne, le noyau ne serait plus alors borné en 0. Dans [29], le noyau de l'estimateur est défini de manière différente aux voisinage de 0 et en dehors du voisinage de 0 (5.7).

$$\hat{f}_{X,h}^{K_\Gamma}(x) = \frac{1}{n} \sum_{i=1}^n K_{\rho_h(x),h}(X_i), \quad x \in \mathbb{R}_+ \quad (5.7)$$

avec

$$\rho_h(x) = \begin{cases} \frac{x}{h} & \text{si } x \geq 2h \\ \frac{1}{4} \left(\frac{x}{h}\right)^2 + 1 & \text{si } x \in [0, 2h[ \end{cases}$$

Le biais de cet estimateur augmente avec la valeur de  $x$  tandis que sa variance diminue lorsque  $x$  croît.

### 5.3.3 Estimation de l'échelle du noyau

L'estimation de l'échelle appliquée au noyau passe par la minimisation de grandeurs estimant l'écart entre la densité réelle et la densité estimée :

- asymptotic mean squared error (AMSE) mesure la distance entre la densité et son estimation

$$AMSE \left( \hat{f}_{X,h}(x) \right) = var(\hat{f}_{X,h}(x)) + bias^2(\hat{f}_{X,h}(x)) \approx \frac{1}{nh} R(K) f(x) + \frac{h^4}{4} \mu_2^2(K) [f''(x)]^2$$

- mean integrated square error (MISE) permet de juger de l'écart moyen

$$MISE \left( \hat{f}_{X,h}(x) \right) = \mathbb{E} \left\{ \int \left( \hat{f}_{X,h}(y) - f(y) \right)^2 dy \right\} = \int bias^2(\hat{f}_{X,h}(y)) dy + \int var(\hat{f}_{X,h}(y)) dy$$

- Asymptotic mean integrated squared error (AMISE)

$$AMISE \left( \hat{f}_{X,h}(x) \right) = \frac{1}{nh} R(K) + \frac{h^4}{4} \mu_2^2(K) R(f'')$$

— Integrated squared error (ISE)

$$ISE\left(\hat{f}_{X,h}(x)\right) = \int \left(\hat{f}_{X,h}(y) - f(y)\right)^2 dy$$

Dans le cas de la grandeur AMISE, il existe une valeur théorique de  $h$  la minimisant :

$$\hat{h}_{AMISE} = \left[ \frac{R(K)}{\mu_2^2(K)R(f'')} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (5.8)$$

Cependant il est nécessaire de connaître la densité réelle pour trouver cette valeur optimale ce qui n'est pas le cas en pratique. Différentes approches permettent alors d'approximer cette échelle optimale. Des états de l'art des méthodes d'estimation de l'échelle du noyau pour l'estimation de densité ont été réalisés dans [102, 57].

### La règle du pouce

Cette première approche consiste à remplacer  $R(f'')$  par sa valeur dans une famille paramétrique (comme les gaussiennes) afin d'estimer la valeur optimale de AMISE. Dans [104], l'estimation de  $h$  dans un cadre gaussien est donnée par  $\hat{h}_{AMISE} = 1.06\sigma n^{-1/5}$  et  $\hat{h}_{AMISE} = 0.79IQRn^{-1/5}$ , avec  $\sigma$  et  $IQR$  estimée sur les données. De plus, afin de détecter de manière plus efficace les différents modes de la distribution, le facteur 1.06 est réduit à 0.9 avec une sélection du minimum entre l'écart-type et l'écart inter-quantile.

$$\hat{h} = 0.9 \min(\hat{\sigma}, IQR/1.34)n^{-\frac{1}{5}}$$

Ces estimations dépendent intégralement de jeu de données avec la variance et l'IQR de ce dernier et du nombre d'échantillons. Il n'y a donc pas d'optimisation à réaliser ce qui rend cette estimation la plus rapide dans un cadre industriel.

### Les méthodes basées sur le maximum de vraisemblance

Ces approches consistent à calculer le maximum de vraisemblance sur les données à partir de la distribution estimée par rapport au paramètre d'échelle  $h$  du noyau. Dans [74], ce maximum est calculé dans un cadre gaussien multidimensionnel à partir d'une formulation de point fixe dans un cadre sphérique où la matrice de covariance est diagonale et dans un cadre sans apriori avec une matrice de covariance pleine.

Il existe d'autres méthodes (autres que la règle du pouce et le maximum de vraisemblance) permettant d'estimer l'échelle optimale basées sur la validation croisée [102, 100] et des méthodes plug-in [103]. Il s'agit de méthodes basées sur des optimisations. Dans le cadre des spectrogrammes, cela revient à effectuer une optimisation pour chacun des points.

## 5.4 Modèle de normalité défini à partir de l'estimation de densité par noyau gaussien

### 5.4.1 Le modèle de normalité

De même que dans la section 5.2 où un modèle de normalité était défini pour chaque point par une distribution gamma, chaque point du spectrogramme (paramétré par un régime  $N_2$  et une fréquence  $f$ ) est caractérisé par une distribution estimée à partir des données normales de ce point et caractérisant le modèle de normalité de ce point uniquement. Aucun a priori n'est apporté à la distribution obtenue à partir de l'estimation de cette densité par noyau gaussien (5.9). La distribution de normalité est donc caractérisée par les données normales, le type de noyau et l'échelle du noyau.

$$\begin{aligned} \forall(f, N_2), \quad \hat{f}_{f, N_2}(x | \Omega_{App}^{f, N_2}, h) &= \frac{1}{\text{card}(\Omega_{App}^{f, N_2})h} \sum_{i \in \Omega_{App}^{f, N_2}} K_N\left(\frac{x - S_{f, N_2}^i}{h}\right) \\ &= \frac{1}{\text{card}(\Omega_{App}^{f, N_2})h} \sum_{i \in \Omega_{App}^{f, N_2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - S_{f, N_2}^i)^2}{h^2}\right) \end{aligned} \quad (5.9)$$

avec  $\Omega_{App}^{f, N_2}$  les données d'apprentissage normales des spectrogrammes aux coordonnées  $(f, N_2)$  et  $S_{f, N_2}^i$  le point du spectrogramme  $i$  sur ces mêmes coordonnées.

Le modèle de normalité est donc entièrement caractérisé à partir des données normales au niveau de chaque point, de la nature du noyau et de l'échelle associée permettant de donner une estimation de la distribution de normalité.

### Estimation de l'échelle du noyau

La modélisation de la distribution permet une détection des points atypiques en analysant la queue de distribution. Le choix de  $h$  est donc important afin d'éviter, dans le cas où  $h$  est trop grand, de définir des points inusuels comme normaux car la distribution aura été fortement lissée et, dans le cas où  $h$  est trop petit, de détecter des points normaux également dû à un surapprentissage de cette distribution. Cette valeur doit être calibrée pour chaque point du spectrogramme. Nous utilisons deux méthodes afin de calibrer ce paramètre :

- issue de la règle du pouce [104] qui permet de donner une approximation directe de la valeur de  $h$  à partir des données d'apprentissage normale sans aucune optimisation

$$\hat{h}_{RP} = 1.06\hat{\sigma}n^{-\frac{1}{5}},$$

avec  $\hat{\sigma}$  la variance estimée sur les données d'apprentissage de la distribution,

- issue du maximum de vraisemblance par rapport à  $h$  calibrée sur des données de validation

normales (non utilisées pour apprendre la distribution), il aurait été également possible de faire du leave-one-out sur la base d'apprentissage.

$$\hat{h}_{ML} = \operatorname{argmax}_h \left\{ \prod_{j \in \Omega_{Val}^{f,N_2}} \hat{f}_{f,N_2}(S_{f,N_2}^j | \Omega_{App}^{f,N_2}, h) \right\}.$$

La figure 5.9 présente la vraisemblance des points de  $\Omega_{Val}^{f,N_2}$  en fonction de différentes valeurs de  $h$ , nous observons un pic de la vraisemblance pour une unique valeur de  $h$  définissant  $\hat{h}_{ML}$ . Cette estimation entraîne cependant plus de variabilité que l'estimation de  $h$  donnée par la règle du pouce  $\hat{h}_{RP}$  (Tableau 5.2). Bien que les deux estimations restent dans les mêmes ordres de grandeur, l'estimation  $\hat{h}_{RP}$  de  $h$  à partir de la règle du pouce est toujours inférieure à celle à partir du maximum de vraisemblance  $\hat{h}_{ML}$  (Tableau 5.2). Cela entraîne une distribution estimée à partir du maximum de vraisemblance plus lisse que celle à partir de la règle du pouce. La figure 5.10 représente les estimations des densités pour différents points du spectrogramme à partir de la règle du pouce (trait plein) et du maximum de vraisemblance (pointillé). Les distributions estimées sur plusieurs points des spectrogrammes possèdent des modes sur différentes intensités vibratoires et des queues de distribution différentes illustrant la variabilité au niveau des mêmes points sur les spectrogrammes. La définition d'un modèle de normalité sur chaque point des spectrogrammes est donc pertinente.

TABLE 5.2 – Estimation de l'échelle à partir de la règle du pouce et du maximum de vraisemblance pour différents points du spectrogramme et avec différentes données d'apprentissage pour estimer la distribution.

$\hat{h}_{RP}$	$\hat{h}_{ML}$
$(1.41 \pm 0.07) \times 10^{-2}$	$(4.58 \pm 1.71) \times 10^{-2}$
$(1.87 \pm 0.52) \times 10^{-2}$	$(4.17 \pm 2.81) \times 10^{-2}$
$(1.35 \pm 0.06) \times 10^{-2}$	$(3.67 \pm 1.07) \times 10^{-2}$
$(1.2 \pm 0.05) \times 10^{-2}$	$(3.67 \pm 1.07) \times 10^{-2}$
$(1.45 \pm 0.06) \times 10^{-2}$	$(4.18 \pm 1.88) \times 10^{-2}$

#### 5.4.2 Le score de détection

La décision est caractérisée, pour chaque point du spectrogramme, par un seuil sur les p-valeurs du test statistique confrontant l'hypothèse  $\mathcal{H}_0^{f,N_2}$  où le point aux coordonnées  $(f, N_2)$  est normal à l'hypothèse  $\mathcal{H}_1^{f,N_2}$  où le point de coordonnées  $(f, N_2)$  est inusuel. Ce test correspond à un test d'adéquation à la distribution de normalité du point estimée par  $\hat{f}_{f,N_2}(\cdot | \Omega_{App}^{f,N_2}, h)$ .

$$pval_{f,N_2}^{i,K_N} = \mathbb{P}_{\mathcal{H}_0^{f,N_2}}(X > S_{f,N_2}^i) \approx \mathbb{P}_{\hat{f}_{f,N_2}}(X > S_{f,N_2}^i) \quad (5.10)$$

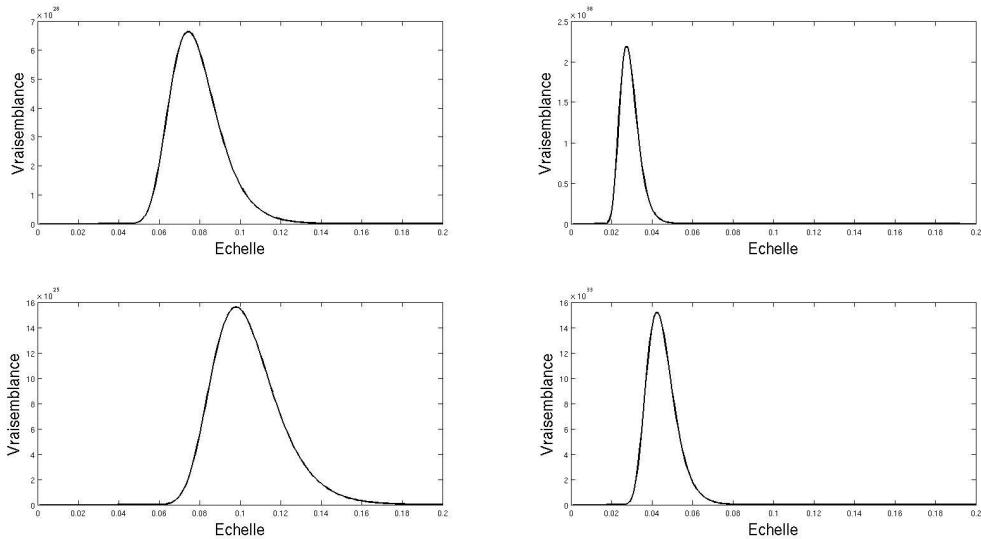


FIGURE 5.9 – Vraisemblance sur  $\Omega_{Val}$  de la distribution estimée en fonction de différentes valeurs de  $h$  et pour différents points du spectrogramme. Pour chaque courbe, le maximum de vraisemblance est atteint en un unique point correspondant en la valeur de  $h$  optimale définie par ce critère.

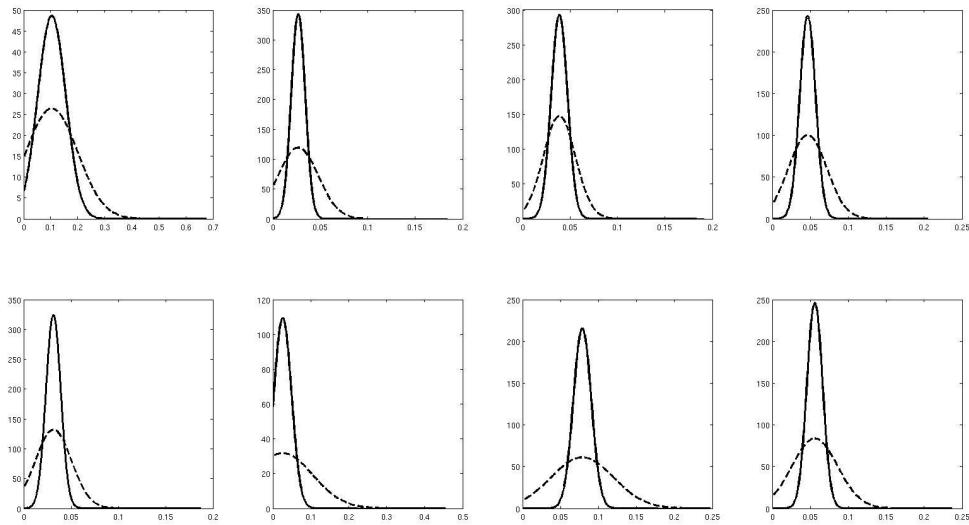


FIGURE 5.10 – Densités estimées pour différents points du spectrogramme à partir de la règle du pouce (trait plein) et du maximum de vraisemblance (pointillés). Les distributions estimées à partir du maximum de vraisemblance et de la règle du pouce sont différentes au niveau de leurs queues de distribution. Les distributions acquises à partir du maximum de vraisemblance sont plus lisses. Les distributions des différents points sont également différentes conformant l'intérêt d'un modèle de normalité pour chaque point des spectrogrammes.

$X$  est une variable aléatoire suivant la loi selon laquelle la probabilité est considérée. Il faut donc définir le calcul de p-valeur pour la distribution estimée par noyau.

$$\begin{aligned}
\mathbb{P}_{\hat{f}_{f,N_2}}(X > S_{f,N_2}^i) &= \int_{S_{f,N_2}^i}^{+\infty} \hat{f}_{f,N_2}(x|\Omega_{App}^{f,N_2}, h) dx \\
&= \int_{S_{f,N_2}^i}^{+\infty} \frac{1}{\text{card}(\Omega_{App}^{f,N_2})h} \sum_{j \in \Omega_{App}^{f,N_2}} K_{\mathcal{N}}\left(\frac{x - S_{f,N_2}^j}{h}\right) dx \\
&= \frac{1}{\text{card}(\Omega_{App}^{f,N_2})} \sum_{j \in \Omega_{App}^{f,N_2}} \int_{S_{f,N_2}^i}^{+\infty} \frac{1}{h} K_{\mathcal{N}}\left(\frac{x - S_{f,N_2}^j}{h}\right) dx \\
&= \frac{1}{\text{card}(\Omega_{App}^{f,N_2})} \sum_{j \in \Omega_{App}^{f,N_2}} \mathbb{P}_{\mathcal{N}(S_{f,N_2}^j, h^2)}(X > S_{f,N_2}^i) \\
&= \frac{1}{\text{card}(\Omega_{App}^{f,N_2})} \sum_{j \in \Omega_{App}^{f,N_2}} \left(1 - F_{\mathcal{N}(S_{f,N_2}^j, h^2)}(S_{f,N_2}^i)\right)
\end{aligned}$$

car  $\frac{1}{h} K_{\mathcal{N}}\left(\frac{x - S_{f,N_2}^j}{h}\right)$  correspond à la densité d'une gaussienne  $\mathcal{N}(S_{f,N_2}^j, h^2)$  de moyenne  $S_{f,N_2}^j$  et de variance  $h^2$ .

La p-valeur de la distribution estimée par noyau correspond alors à la moyenne des p-valeurs issues de chaque noyau centré en un des points d'apprentissage de la densité. L'utilisation du noyau gaussien (ou de noyau symétrique) permet d'avoir une formule fermée pour le calcul de la p-valeur. La décision  $\tilde{Y}_{f,N_2}^{i,K_{\mathcal{N}}}$  (5.11) est alors effectuée par un seuil  $s$  sur ces p-valeurs. Un filtrage par rapport aux détections sur le voisinage direct  $\mathcal{V}_{f,N_2}$  est également utilisé pour réduire les fausses détections (5.12) de manière analogue à l'approche par établie à partir des distributions paramétriques par des lois Gamma.

$$\tilde{Y}_{f,N_2}^{i,K_{\mathcal{N}}} = \mathbb{1}\left\{pval_{f,N_2}^{i,K_{\mathcal{N}}} \leq s\right\} \quad (5.11)$$

$$\hat{Y}_{f,N_2}^{i,K_{\mathcal{N}}} = \begin{cases} 1 & \text{si } \tilde{Y}_{f,N_2}^{i,K_{\mathcal{N}}} = 1 \text{ \& } \sum_{k \in \mathcal{V}_{f,N_2}} \tilde{Y}_k^{i,K_{\mathcal{N}}} \geq 2 \\ 0 & \text{sinon} \end{cases} \quad (5.12)$$

### 5.4.3 Calibration des seuils

Ce seuil correspond au niveau des tests statistiques unitaires (5.10) mis en place pour chaque point du spectrogramme. Le seuil défini sur les p-valeurs est identique pour chaque point par manque de données annotées ponctuellement. Nous calculons le taux de détection des différentes classes de point des données de validation  $\Omega_{Val}$  de la base  $\mathcal{B}_1$  (Figure 5.11) pour différentes valeurs de seuils. Ce seuil est estimé pour les deux méthodes de définition d'échelle  $h$  du modèle de normalité. Les résultats obtenus à partir des deux modèles de normalité sont similaires, nous

définissons un seuil de détection  $s$  sur les p-valeurs identique pour les deux méthodes avec :

$$s = 0.07$$

Ce seuil permet la détection d'environ 85% des points atypiques de la base de validation, malgré une détection d'environ 95% des points décalés provenant des signatures vibratoires liées au  $N_1$ . Les points normaux et le bruit sont faiblement détectés. Notre approche a pour but de mettre en évidence les signatures inusuelles, pour cela le seuil de détection ne doit pas entraîner un grand nombre de fausses détections pour éviter de masquer les signatures inusuelles sur le résultat final.

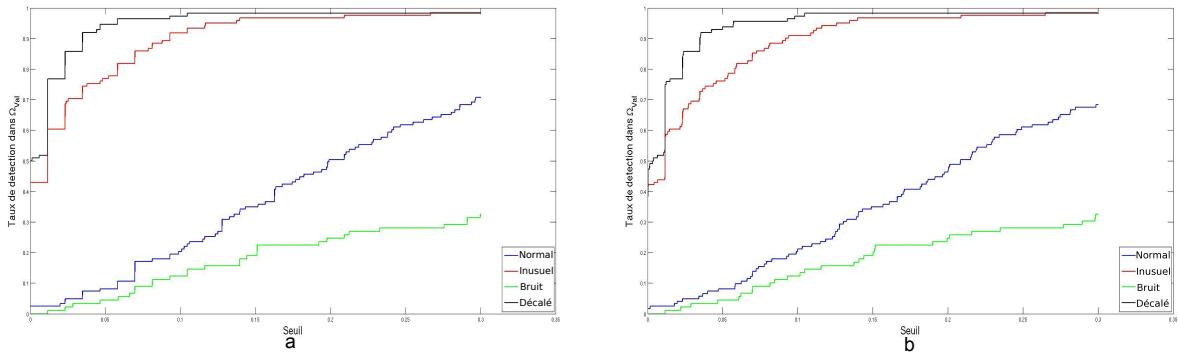


FIGURE 5.11 – Taux de détection en fonction des seuils de détection des différentes classes de points de la base de validation  $\Omega_{Val}$  à partir des modèles de normalité définis par l'estimation de la distribution avec noyau gaussien en estimant l'échelle par la règle du pouce (a) et par le maximum de vraisemblance (b). Les résultats issus des 2 modèles de normalité sont similaires. Ils montrent un fort taux de bonnes detections des points inusuels pour des petites valeurs de seuils, les points normaux décalés sont également fortement détectés. Les points normaux liés aux raies  $N_2$  et le bruit sont faiblement détectés pour ces valeurs de seuils, nous sélectionnons donc un seuil de détection au niveau de ces dernières.

#### 5.4.4 Résultats sur la base de test $\Omega_{Test}$

Comme dans la partie précédente, les résultats sont présentés sous 3 formats :

- à partir des points annotés de la base de test  $\Omega_{Test}$  de  $\mathcal{B}_1$ ,
- à partir du taux de points détectés appartenant aux zones atypiques extraites sur tous les patchs,
- à partir d'une visualisation des points détectés sur le patch.

#### Résultats de classification des points annotés de $\Omega_{Test}$

Le tableau 5.3 fournit les taux de détection des différentes classes de points des patchs dans  $\Omega_{Test}$  issus de la base  $\mathcal{B}_1$  sans application du filtre pour différentes données d'apprentissage de la distribution. Le tableau 5.4 fournit les mêmes résultats avec l'application du filtre de voisinage.