

avec k une des directions de voisinage présentées dans la figure 6.6, $\vec{\mathcal{V}}_{f,N_2}^x$ les voisinages selon les différentes directions associées à la donnée dont x est issu, \mathcal{V}_{f,N_2}^k le voisinage selon la direction k aux coordonnées (f, N_2) , $H_{S\mathcal{V}^k}$ et $H_{\mathcal{V}^k}$ sont les matrices d'échelles utilisées pour estimer respectivement la densité jointe du point étudié avec son voisinage dans la direction k et la densité jointe du voisinage dans la direction k . $\hat{f}_{f,N_2|\mathcal{V}_{f,N_2}^k}^k$ est alors la densité conditionnelle estimée du point par rapport à son voisinage dans la direction k . La densité conditionnelle est estimée de manière identique au chapitre précédent :

$$\hat{f}_{f,N_2|\mathcal{V}_{f,N_2}^k}^k(x|\mathcal{V}_{f,N_2}^{x,k}) = \frac{\hat{f}_{f,N_2,\mathcal{V}_{f,N_2}^k}^k(x, \mathcal{V}_{f,N_2}^{x,k})}{\hat{f}_{\mathcal{V}_{f,N_2}^k}^k(\mathcal{V}_{f,N_2}^k)}.$$

Chaque point du spectrogramme est alors caractérisé par sa vraisemblance par rapport à la densité conditionnelle estimée sur des données normales dans les différentes directions de voisinage. Un point inusuel appartient à une raie atypique et est donc lié à une direction. La vraisemblance de ce point dans la direction de la signature atypique est donc faible.

A la différence de la section précédente, plusieurs distributions doivent être estimées pour chaque point du spectrogramme. Ainsi, afin d'effectuer ces estimations dans un temps raisonnable, nous utilisons la méthode d'estimation de densité par noyau gaussien avec une matrice d'échelle diagonale estimée à partir de la règle du pouce. Le choix d'utiliser une matrice d'échelle pleine dans la section 6.3 venait d'une volonté de tenir compte de l'orientation des dépendances des points dans l'estimation. Cette prise en compte est effectuée dans ce cas directement à partir de la sélection du voisinage.

Le modèle de normalité correspond donc à l'estimation de la densité conditionnelle pour chaque point du spectrogramme par rapport à leurs voisins suivant les différentes directions à partir de la méthode à noyau gaussien appliquée avec des matrices d'échelle diagonales.

6.4.3 Le score de détection

La détection des points atypiques s'effectue à partir de la p-valeur des tests statistiques calculés à partir des distributions estimées. Un test statistique est réalisé au niveau de chaque direction de voisinage opposant l'hypothèse $\mathcal{H}_{0,k}^{f,N_2}$ de normalité du point suivant la direction étudiée à l'hypothèse $\mathcal{H}_{1,k}^{f,N_2}$ d'irrégularité de ce point sur cette même direction. Le calcul de la p-valeur s'effectue de manière analogue à la section précédente (6.7) comme un test d'adéquation par rapport aux distributions des différentes directions.

$$pval_{f,N_2}^{i,K_{\mathcal{V}^k}} = \mathbb{P}_{H_{0,k}^{f,N_2}}(X > S_{f,N_2}^i | \mathcal{V}_{f,N_2}^{i,k}) \approx \mathbb{P}_{\hat{f}_{f,N_2|\mathcal{V}_{f,N_2}^k}^k}(X > S_{f,N_2}^i | \mathcal{V}_{f,N_2}^{i,k}) \quad (6.10)$$

Ainsi, pour chaque point du spectrogramme nous disposons d'un indicateur ϕ_{f,N_2}^i (6.11)

correspondant aux p-valeurs (6.10) des tests statistiques effectués dans les différentes directions.

$$\phi_{f,N_2}^i \left(S_{f,N_2}^i, \vec{\mathcal{V}}_{f,N_2}^i \right) = \left\{ pval_{f,N_2}^{i,K_{\mathcal{V}_k}} \right\}_{k=1,\dots,12} \quad (6.11)$$

Le classifieur correspond alors à la comparaison de la **p-valeur minimale** de cet indicateur par rapport à un seuil de détection (6.12), nous associons cette p-valeur minimale à la direction selon laquelle la signature inusuelle est orientée. Si la p-valeur minimale est supérieure au seuil de détection s , cela signifie que dans toutes les différentes directions, le point peut être considéré comme normal. A l'inverse il suffit qu'elle soit inférieure au seuil de détection sur une seule direction pour que le point soit considéré comme atypique.

$$\tilde{Y}_{f,N_2}^{i,K_{\vec{\mathcal{V}}}} = \begin{cases} 1 & \text{si } \min \phi_{f,N_2}^i \left(S_{f,N_2}^i, \vec{\mathcal{V}}_{f,N_2}^i \right) \leq s \\ 0 & \text{sinon} \end{cases} \quad (6.12)$$

Le filtrage de voisinage (5.12) est également appliqué pour supprimer la détection de points isolés. Ce filtrage est intéressant dans ce cas de figure particulièrement car la prise en compte du voisinage directionnel entraîne la détection des voisins des points inusuels également.

$$\hat{Y}_{f,N_2}^{i,K_{\vec{\mathcal{V}}}} = \begin{cases} 1 & \text{si } \tilde{Y}_{f,N_2}^{i,K_{\mathcal{V}_k}} = 1 \& \sum_{l \in \mathcal{V}_{f,N_2}} \tilde{Y}_l^{i,K_{\mathcal{V}_k}} \geq 2 \\ 0 & \text{sinon} \end{cases}$$

6.4.4 La caractérisation des signatures inusuelles

La détection d'anomalies prend en compte la p-valeur minimale de l'indicateur (6.11) que nous associons à la direction de la signature inusuelle. Ainsi, pour chaque point nous disposons également de la direction estimée de la signature atypique à laquelle il appartient. Cependant, plusieurs composantes de l'indicateur des p-valeurs peuvent être très faibles numériquement et considérées comme nulles dû à la proximité des différentes directions. Elles ne permettent pas d'estimer la direction mais permettent tout de même de déclarer le point comme inusuel. Ainsi l'estimation de la direction de la signature atypique ne s'effectue pas sur l'indicateur des p-valeurs $\phi_{f,N_2}^i \left(S_{f,N_2}^i, \vec{\mathcal{V}}_{f,N_2}^i \right)$, mais correspond au minimum des densités conditionnelles dans les différentes directions $\varphi_{f,N_2} \left(S_{f,N_2}^i, \vec{\mathcal{V}}_{f,N_2}^i \right)$ (6.9). Cela correspond à sélectionner la direction selon laquelle le point étudié sachant les voisins dans cette même direction est le moins vraisemblable (6.13).

$$\vec{d}_{f,N_2}^i = \operatorname{argmin}_k \left(\varphi_{f,N_2} \left(S_{f,N_2}^i, \vec{\mathcal{V}}_{f,N_2}^i \right) \right) \quad (6.13)$$

Ces directions caractérisent donc les signatures inusuelles car tous les points issus de la même signature atypique et détectés par l'approche possèdent la même direction estimée (ou une des directions voisines) associée à l'orientation de la raie. Cette raie correspond également à une forme

connexe sur le spectrogramme. Une signature inusuelle détectée par notre approche correspond donc à une forme connexe dans l'espace des détections (espace de la dimension du patch dont chaque élément vaut 0 ou 1 suivant que le point est détecté ou non).

Les composantes connexes correspondent à un ensemble de points voisins détectés par notre approche, elles sont donc associées à une signature inusuelle. Nous considérons les directions estimées pour chaque point des composantes connexes comme un indicateur caractérisant les signatures atypiques. Chaque composante est alors caractérisée par un vecteur de dimension égale au nombre de points dans la composante. Chaque élément du vecteur correspond à la direction estimée du point associé. Nous caractérisons alors la signature inusuelle associée à la composante connexe par l'histogramme des différentes directions prises par les points la composant. Le mode de cet histogramme correspond à l'orientation principale de la signature atypique. Seules les composantes connexes contenant suffisant de points sont considérées.

6.4.5 Calibration du seuil de détection

Le seuil de détection est estimé à partir des taux de détection des différentes classes de points annotés des données de validation Ω_{Val} de la base \mathcal{B}_1 pour différentes valeurs de seuil de détection (Figure 6.7). Cette approche permet une très forte détection à plus de 85% des points atypiques ainsi que des points normaux décalés à des seuils de détection très faibles tandis que les autres points normaux restent très faiblement détectés à moins de 10%. Les probabilités conditionnelles sont d'intensités très faibles (principalement dans les directions des signatures inusuelles) entraînant un seuil de détection très faible et proche de 0. Afin d'avoir une meilleure visualisation du seuil de détection, nous représentons sur la figure 6.8 les points détectés sur 4 patchs de la base de validation à différents niveaux de seuils.

Les signatures inusuelles restent détectables pour toutes les valeurs de seuil étudiées malgré leurs très faibles valeurs. Nous pouvons remarquer que les valeurs de seuil de 10^{-6} et 10^{-8} entraînent de nombreuses détections de points normaux. La décision pour un seuil de 10^{-15} est intéressante mais ce seuil est numériquement très faible pouvant manquer certaines signatures. Nous avons donc opté pour un seuil plus tolérant malgré un nombre plus important de fausses détections dont la majorité pourra être évitée à partir du filtrage de voisinage.

$$s = 10^{-10}$$

6.4.6 Résultats sur la base de test Ω_{Test}

Taux de détection des points annotés

Le tableau 6.2 présente les taux de détection des différentes classes de points de la base de test Ω_{Test} de \mathcal{B}_1 à partir de notre approche avec et sans filtrage. Nous remarquons un important

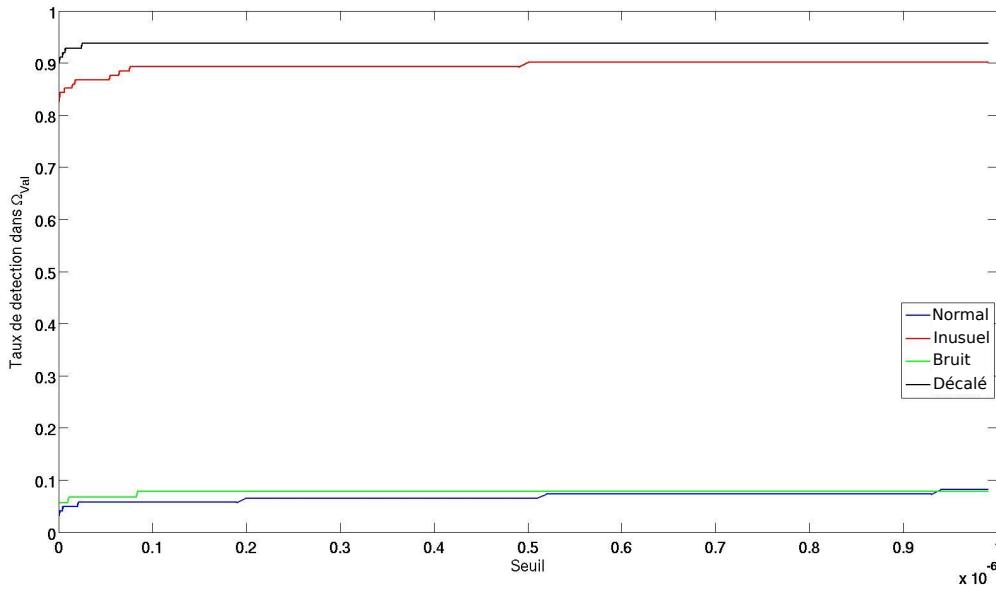


FIGURE 6.7 – Taux de détection des différentes classes de points de base de validation annotée ponctuellement Ω_{Val} en fonction du seuil de détection. Les points atypiques (rouge) et les points normaux décalés (noir) sont très fortement détectés à plus de 85% pour des valeurs de seuils très faibles tandis que les points normaux (bleu) et le bruit (vert) sont très faiblement détectés à moins de 10%. Nous choisissons donc un seuil de détection très faible.

taux de détection avec et sans filtrage des points atypiques attestant des bonnes performances du modèle de normalité basé sur la distribution conditionnelle des points en fonction des voisins dans différentes directions. Les points normaux et le bruit sont faiblement détectés, mais comme pour les autres approches, les points normaux décalés sur les spectrogrammes restent très fortement détectés. L'utilisation du filtrage de voisinage réduit légèrement les taux de détection mais son effet reste plus visible sur les représentations visuelles des résultats.

TABLE 6.2 – Taux de détection calculés à partir du modèle de normalité défini par les densités conditionnelles dans les différentes direction estimées à partir de différentes données d'apprentissage avec et sans filtrage pour les différentes classes de points des données dans Ω_{Test}

Classe de points	Détection sans filtrage	Détection avec filtrage
Classe "normal"	$10.3\% \pm 0\%$	$10.3\% \pm 0\%$
Classe "inusuel"	$88.4\% \pm 2.3\%$	$82.8\% \pm 2.7\%$
Classe "bruit"	$3.8\% \pm 1.4\%$	$3.1\% \pm 0\%$
Classe "décalé"	$93.3\% \pm 2.6\%$	$91.8\% \pm 2.3\%$

Résultats visuels de détection

Les figures 6.9, 6.10, 6.11, 6.12 et 6.13 présentent les résultats de détection à partir du modèle de normalité défini par la densité conditionnelle directionnelle sur tout le patch spécifique de la base \mathcal{B}_1 pour 5 différents spectrogrammes avec et sans application du filtrage de voisinage. Nous présentons également pour chacun d'entre eux les 4 composantes connexes les plus importantes

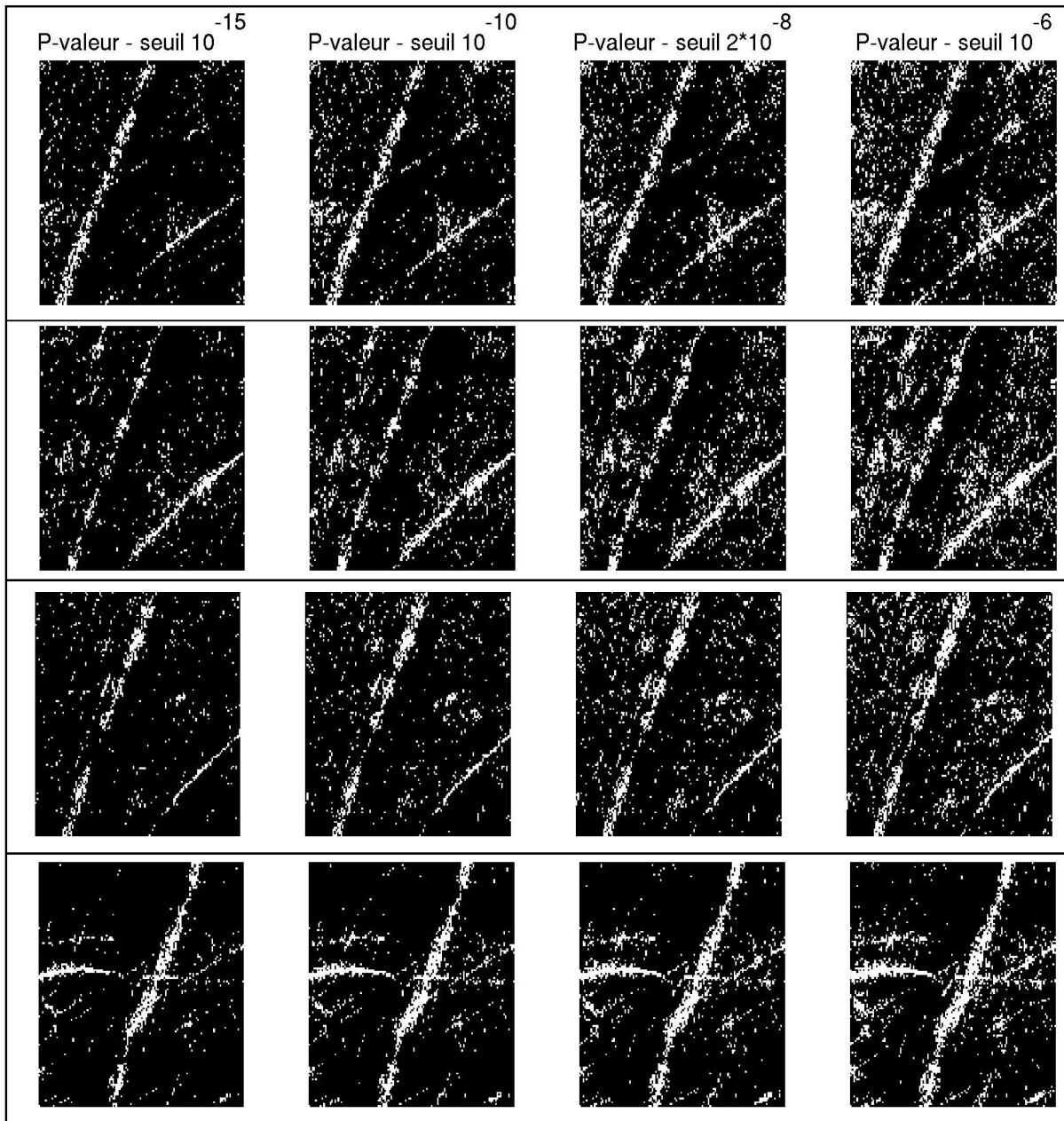


FIGURE 6.8 – Détection des points inusuels sur 4 différents patchs de Ω_{Val} pour différentes valeurs de seuil. Les signatures atypiques sont détectées pour les différentes valeurs de seuils sur les 4 patchs, la détection à 10^{-15} est intéressante. Nous avons choisi un seuil de 10^{-10} plus tolérant pour assurer la détection des signatures atypiques. Les seuils de 2×10^{-8} et 10^{-6} entraînent trop de fausses détections.

(celles contenant le plus de points) dans l'espace des détections avec les histogrammes des directions pour chacune de ces composantes connexes. La figure 6.9 concerne un patch normal sans signature inusuelle et les figures 6.10, 6.11, 6.12, 6.13 contiennent différentes signatures atypiques (Figure 2.3).

Le premier élément à constater sur ces résultats concerne les effets du filtrage. L'approche sans application du filtrage entraîne un certain nombre de fausses détections qui sont fortement réduites par son application permettant de ne laisser principalement que les signatures atypiques et celles décalées liées au N_1 . Le filtrage de voisinage permet donc de diminuer le nombre de fausses détections tout en laissant intact les détections des signatures inusuelles. Les différentes signatures atypiques sont bien mises en évidence dans les différents patchs contenant ces signatures. Nous pouvons constater que ces signatures sont visuellement très soulignées du fait de la détection des points voisins autour de ces dernières. Il s'agit tout de même de fausses détections, pouvant masquer les formes des signatures atypiques (Figure 6.10). Caractériser la normalité à partir de la densité conditionnelle de chaque point par rapport à leur voisinage dans différentes directions est donc une approche pertinente pour détecter les signatures atypiques.

Les éléments détectés sur le patch normal sont également très intéressants. Nous pouvons voir que le nombre de points détectés sur ce patch est nettement inférieur aux autres données étant donné qu'il ne possède pas de signature inusuelle. Seule la raie liée au N_1 est cataloguée comme inusuelle. Certaines sous zones de petite taille sont également détectées. Elles constituent une infime partie du patch étudié. Nous pouvons également voir que les raies liées au N_2 ne sont pas détectées par l'approche et sont donc bien caractérisées par le modèle de normalité.

Les plus grandes composantes connexes sur les différents patchs concernent les signatures liées au N_1 et les signatures atypiques. Ces différentes signatures peuvent se composer d'une ou plusieurs composantes. Il s'agit donc bien des éléments détectés par l'approche. Logiquement, les composantes connexes détectées sont les mêmes pour les détections avec et sans filtrage, ce qui permet d'affirmer l'inaction du filtrage sur les signatures inusuelles. Chaque composante connexe est associée à un histogramme de la même couleur dont les bins correspondent aux numéros des directions présentés en figure 6.6. Les histogrammes associés à ces composantes possèdent des modes au niveau de la direction de la signature, cependant la détection des points voisins entraînent la considération d'autres directions également. Pour le patch normal (Figure 6.9), seule la plus grande composante contenant le plus de points est intéressante (les autres ne contenant pas suffisamment de points pour caractériser les signatures). L'histogramme associé à cette composante connexe possède un mode correspondant exactement à la direction de la signature. La détection des signatures inusuelles au niveau du second patch (Figure 6.10) concerne de nombreux points voisins également masquant les directions des signatures. L'histogramme de la composante au niveau de la raie N_1 (jaune) possède un mode au niveau de la direction de la signature. Les composantes des autres patchs possèdent des histogrammes dont les modes sont proches de l'orientation de la signature. La direction des points détectés par l'approche sur les patchs permet bien de caractériser plus finement les signatures atypiques avec la détection des points et la direction de la signature à laquelle ils appartiennent.

Le modèle de normalité basé sur la densité conditionnelle de chaque point en fonction des voisins dans différentes directions est donc efficace pour détecter les points atypiques sur les spectrogrammes à partir de la p-valeur minimale malgré la détection des points voisins masquant certaines signatures. Les densités associées à ces p-valeurs permettent d'estimer la direction de la signature inusuelle. L'étude des composantes connexes dans l'espace de détections permet d'obtenir un histogramme des directions caractérisant les signatures atypiques qui pourrait permettre de labéliser les signatures à partir d'un clustering.

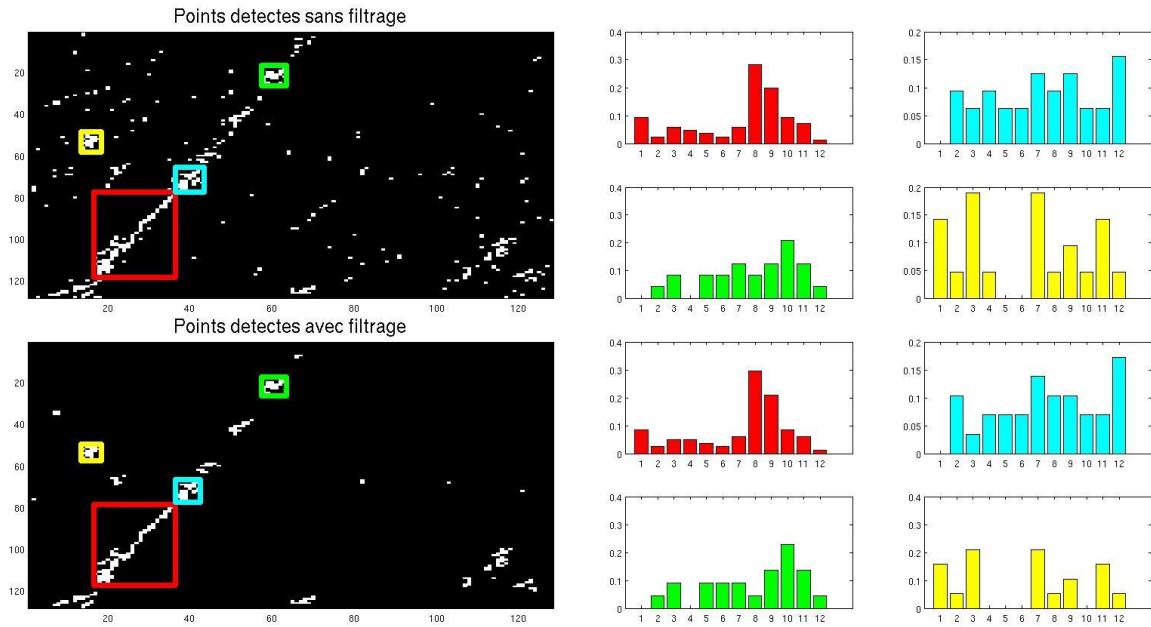


FIGURE 6.9 – Détection des points inusuels sur un patch sans signature inusuelle avec (partie inférieure) et sans (partie supérieure) application du filtrage de voisinage. Les 4 composantes connexes sur les espaces de détections sont extraites (cadres de couleur) et associées à un histogramme de détection (à droite dans la même couleur) pour les 2 images de détections, l'abscisse des histogrammes correspond aux différentes directions de voisinage étudiées. Le filtrage de voisinage (partie inférieure) permet de réduire les mauvaises détections. Les composantes connexes correspondent à des éléments de la raie N_1 , l'histogramme de la composante connexe principale possède un modèle correspondant à l'orientation de la signature.

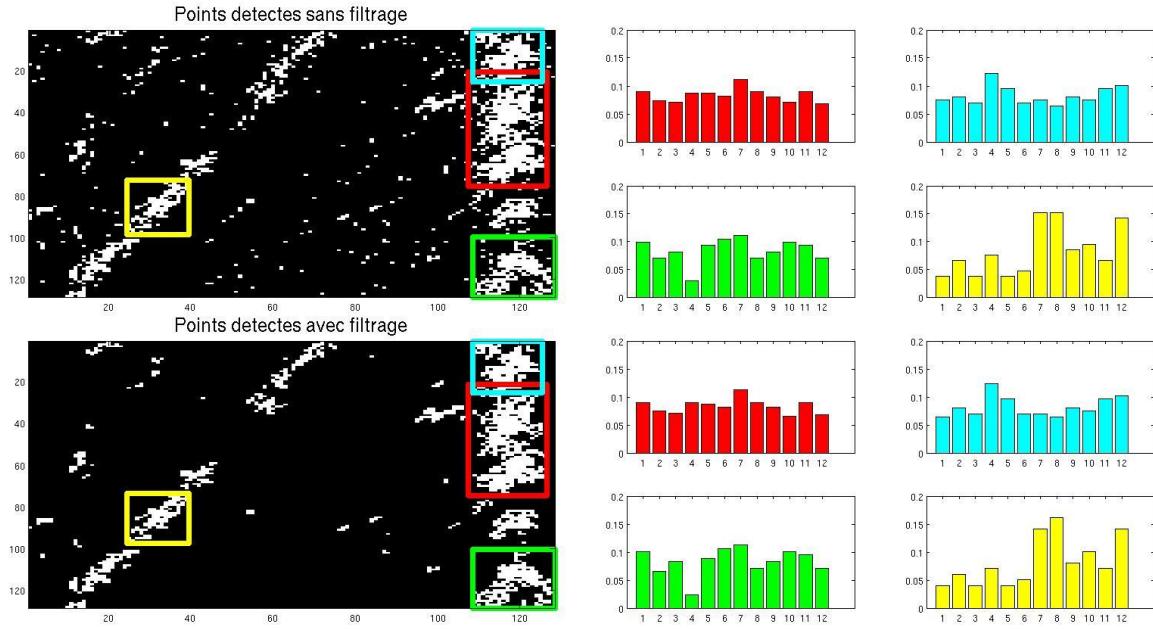


FIGURE 6.10 – Détection des points inusuels sur un patch contenant un premier type de signature inusuelle. La structure est identique à la figure 6.9.

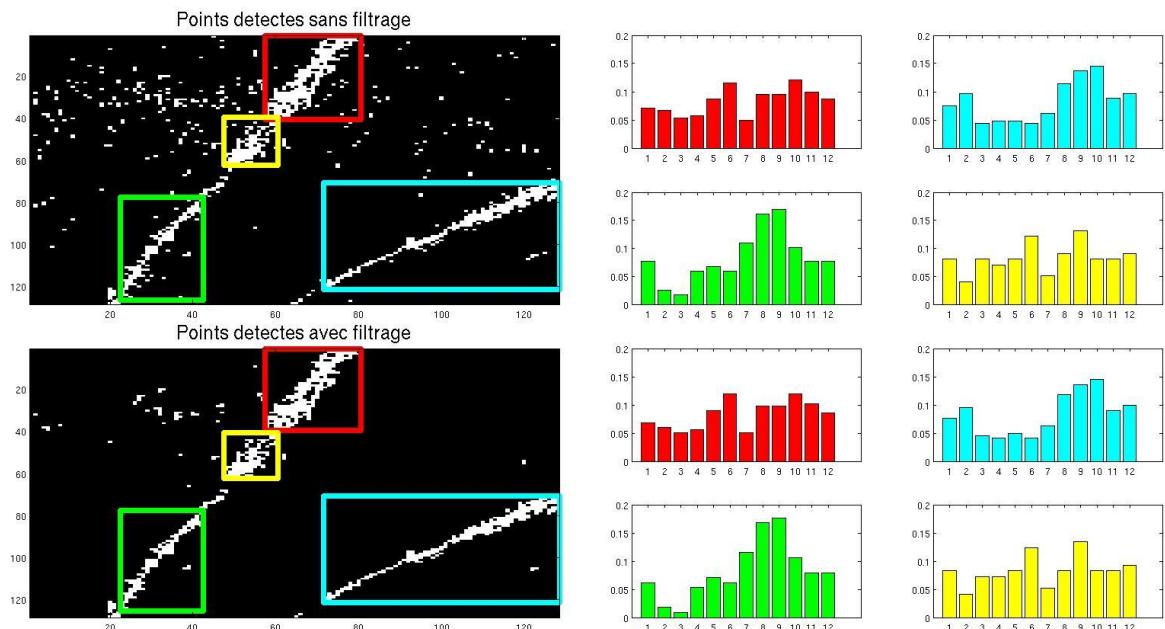


FIGURE 6.11 – Détection des points inusuels sur un patch contenant un second type de signature inusuelle. La structure est identique à la figure 6.9.

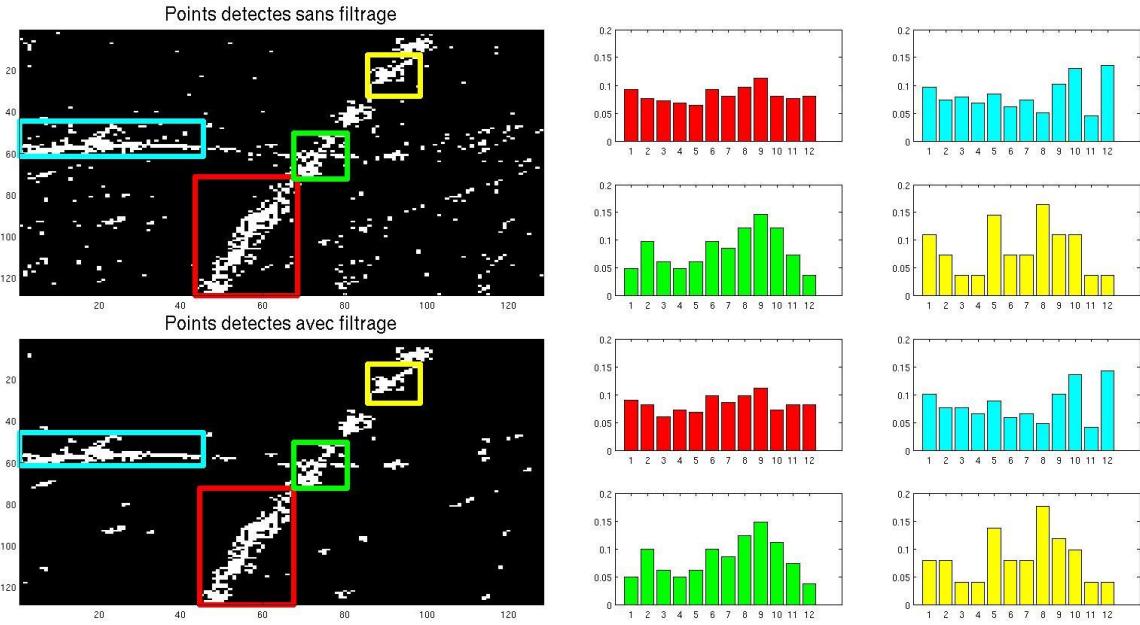


FIGURE 6.12 – Détection des points inusuels sur un patch contenant un troisième type de signature inusuelle. La structure est identique à la figure 6.9.

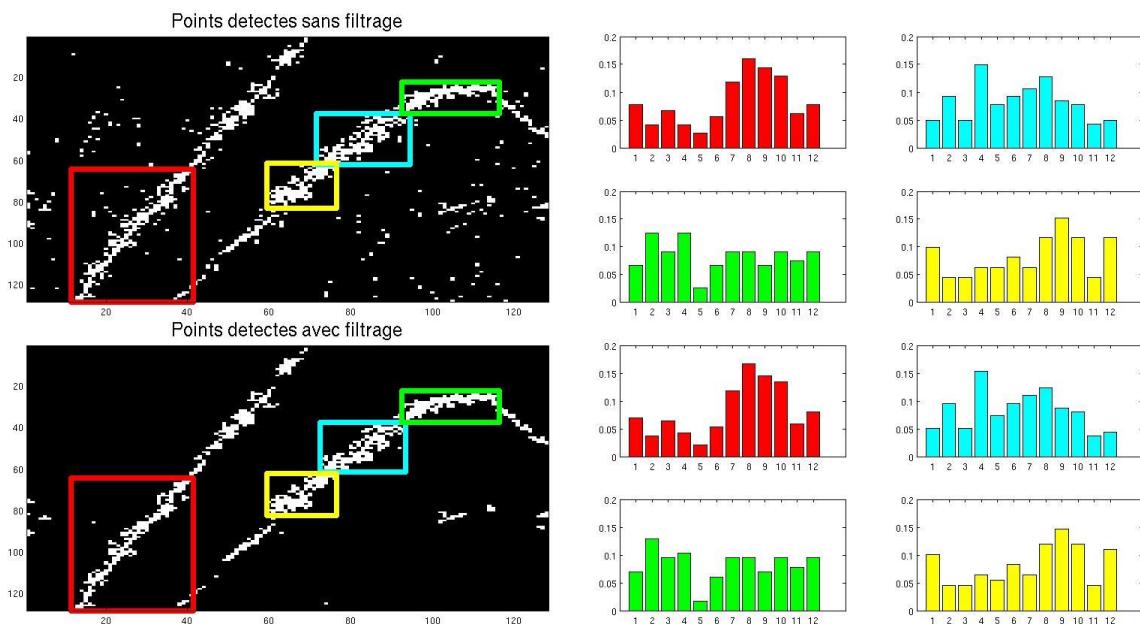


FIGURE 6.13 – Détection des points inusuels sur un patch contenant un quatrième type de signature inusuelle. La structure est identique à la figure 6.9.

6.5 Fusion des différentes approches

6.5.1 Comparaison des approches par dictionnaire et ponctuelles

Les approches par dictionnaire définissent un modèle de normalité globalement sur le patch, tandis que les approches ponctuelles définissent des modèles de normalité localement pour chaque point. Les approches par dictionnaire sont donc plus à même de capturer la structure entière du patch, tandis que les approches ponctuelles sont plus efficaces pour étudier les comportements locaux. Dans l'étude ponctuelle, la détection s'effectue uniquement par comparaison avec les mêmes points sur les patchs normaux entraînant une sensibilité aux intensités vibratoires et à la position des raies. Pour les modèles définis sur le patchs, la détection est effectuée à travers les résidus de reconstruction intra-patchs. Le seuil de détection est différent sur chaque patch entraînant une détection systématique de plusieurs points, y compris pour les patchs sans signature atypique. Les approches ponctuelles définissant un modèle pour chaque point sont plus coûteuses et prennent plus de temps que les approches par patch. La définition du modèle de normalité des approches par dictionnaire est de l'ordre de quelques secondes (entre 5 secondes pour les curvelets et 10 secondes pour la NMF). Pour les approches ponctuelles, la définition est de l'ordre de plusieurs minutes (2 minutes pour les approches indépendantes, 5 minutes pour les approches conditionnelles aux voisins d'ordre 1 et une dizaine de minutes pour les approches conditionnelles directionnelles). Ces temps de calcul restent convenables dans un cadre industriel. La décision pour ces différents modèles s'effectuent en des temps très court.

La comparaison des approches montrent un avantage pour les approches ponctuelles par rapport aux approches par dictionnaire. Elles proposent de bon niveau de détection tout en limitant les fausses détections des points liés au N_2 et du bruit. Les points liés au N_1 restent détectés par les 2 types d'approches. Visuellement les deux approches permettent la détection des signatures atypiques sur les spectrogrammes. La différence provient des points détectés à tort. Les approches ponctuelles détectent à tort des points isolés tandis que les approches par dictionnaire détectent à tort des morceaux de raies. Les résultats numériques sont à nuancer car nous utilisons une base annotée ponctuellement de quelques points afin d'établir les seuils et de présenter les résultats. Il serait plus efficace de définir un seuil sur chaque point séparément afin d'optimiser la détection mais cela nécessite une base de données intégralement annotées ponctuellement afin de donner des résultats généraux.

6.5.2 Fusion des approches par dictionnaire et ponctuelles

Dans la section 4.5.5, les détections des 2 approches par dictionnaire avaient été fusionnées afin d'obtenir des résultats plus robustes et de réduire les fausses détections. Cela était possible car les différentes approches permettaient la détection des signatures inusuelles mais également de différents points normaux. Les approches ponctuelles permettent également la détection des points inusuels. Il est alors possible d'utiliser les différents résultats de détection des différentes