

FIGURE 5.9 – Vraisemblance sur Ω_{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.

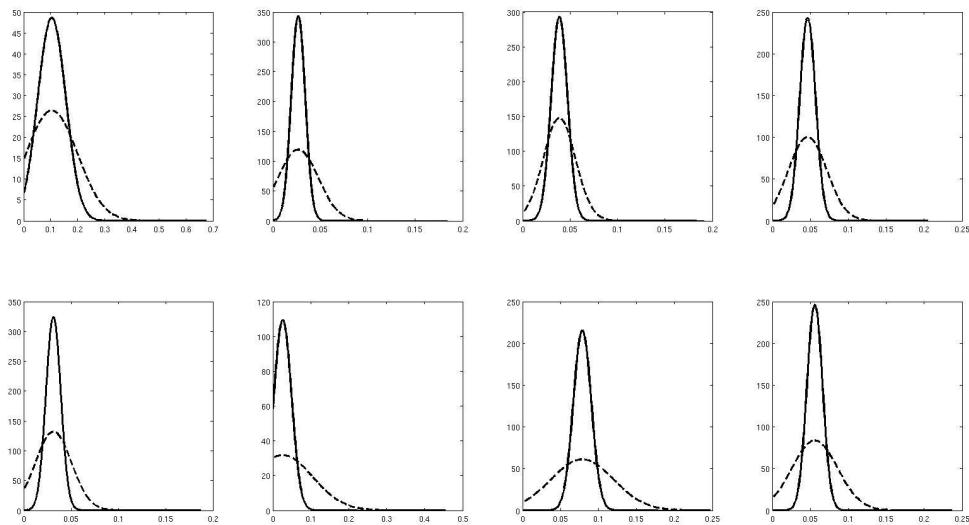


FIGURE 5.10 – Densités estimées pour différents points du spectrogramme à partir de la règle du pouce (trait plein) et du maximum de vraisemblance (pointillés). Les distributions estimées à partir du maximum de vraisemblance et de la règle du pouce sont différentes au niveau de leurs queues de distribution. 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 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 Ω_{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.

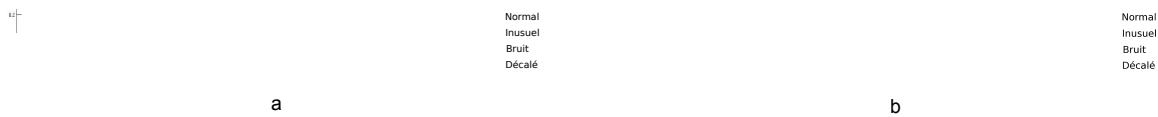


FIGURE 5.11 – Taux de détection en fonction des seuils de détection des différentes classes de points de la base de validation Ω_{Val} à partir des modèles de normalité définis par l'estimation de la distribution avec noyau gaussien en estimant l'échelle par la règle du pouce (a) et par le maximum de vraisemblance (b). Les résultats issus des 2 modèles de normalité sont similaires. Ils montrent un fort taux de bonnes détections 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 Ω_{Test}

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

- à partir des points annotés de la base de test Ω_{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 Ω_{Test}

Le tableau 5.3 fournit les taux de détection des différentes classes de points des patchs dans Ω_{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.

TABLE 5.3 – Taux de détection à partir de la densité estimée par noyau gaussien à partir de différentes données d'apprentissage pour les différentes classes de points des données dans Ω_{Test} et les différentes méthodes de définition de l'échelle du noyau

Classe de points	Taux de détection - règle du pouce	Taux de détection - ML
Classe "normal"	$13.3\% \pm 1.1\%$	$13.3\% \pm 1.1\%$
Classe "inusuel"	$85.1\% \pm 1.3\%$	$83.3\% \pm 1\%$
Classe "bruit"	$3.12\% \pm 0\%$	$3.12\% \pm 0\%$
Classe "décalé"	100%	$99.5\% \pm 1.1\%$

TABLE 5.4 – Taux de détection à partir de la densité estimée par noyau gaussien à partir de différentes données d'apprentissage avec application du filtrage de voisinage pour les différentes classes de points des données dans Ω_{Test} et les différentes méthodes de définition de l'échelle du noyau

Classe de points	Taux de détection - règle du pouce	Taux de détection - ML
Classe "normal"	$10.3\% \pm 0\%$	$10.3\% \pm 0\%$
Classe "inusuel"	$75.4\% \pm 1.3\%$	$74.4\% \pm 0\%$
Classe "bruit"	$3.12\% \pm 0\%$	$3.12\% \pm 0\%$
Classe "décalé"	$91.8\% \pm 1.1\%$	$89.7\% \pm 1.8\%$

La méthode permet donc un bon niveau de détection des points atypiques mais également des points décalés, correspondant alors à de fausses détections. Les signatures liées au N_1 sont considérées comme plus atypiques que les points réellement inusuels. Cela provient du fait que ces raies sont d'intensités plus importantes que les signatures inusuelles, et leurs décalages entraînent leurs détections du fait de l'étude ponctuelle. Les points normaux et le bruit sont faiblement détectés conformément aux données de validation. Le filtrage réduit les fausses détections (principalement au niveau des points décalés) ainsi que la détection des points atypiques d'environ 10%. Ces résultats sont à combiner avec les détections visuelles des signatures inusuelles à partir de nos approches car ils correspondent à la détection de quelques points sélectionnés sur le patch uniquement. La détection visuelle permet d'avoir une appréciation du modèle sur l'ensemble du patch et de l'effet du filtrage par voisinage. Les modèles de normalité définis à partir de la règle du pouce et du maximum de vraisemblance sont très proches. Pour cette raison, dans la suite, nous présentons les résultats à partir de la règle du pouce uniquement.

Résultats de détection sur l'ensemble des patchs \mathcal{K}^{128} des spectrogrammes

Nous donnons les résultats de détection sur les différents patchs. Pour cela comme dans la partie précédente, nous calculons l'erreur E_j^i (3.15) correspondant à la proportion de points détectés appartenant à une zone atypique sur un patch contenant des signatures inusuelles et la proportion des points détectés sur les patchs ne contenant pas de zones atypiques. Le tableau 5.6 donne les résultats pour les deux modèles de normalité sur les différents patchs, il est subdivisé en sous-tableaux dont la structure est défini dans le tableau 5.5.

Plusieurs patchs possèdent un pourcentage important de points détectés faisant partie d'une

TABLE 5.5 – Structure des sous-tableaux de résultats pour le modèle défini par la distribution de chaque point estimé par noyau gaussien

Taux de bonnes détections du modèle de normalité défini par la règle du pouce pour les patchs contenant des signatures inusuelles	Taux de points détectés du modèle de normalité défini par la règle du pouce pour les patchs sans signatures inusuelles
Taux de bonnes détections du modèle de normalité défini par la règle du pouce avec filtrage du voisinage pour les patchs contenant des signatures inusuelles	Taux de points détectés du modèle de normalité défini par la règle du pouce avec filtrage pour les patchs sans signatures inusuelles

zone atypique tandis que d'autres possèdent un faible pourcentage. Cette variabilité provient de l'extraction des zones atypiques contenant plusieurs points normaux, ainsi l'intersection de certains patchs avec les zones atypiques ne contient pas de points inusuels. Les forts taux de détection sur les patchs contenant des signatures inusuelles signifient donc que parmi l'ensemble de la base de test sur ce patch, les points détectés se trouvent principalement aux alentours des signatures inusuelles permettant ainsi de les détecter. La proportion de points détectés dans les patchs sans signature atypique est faible comparativement à la taille du patch. La méthode est donc très intéressante pour détecter les points inusuels sur les spectrogrammes.

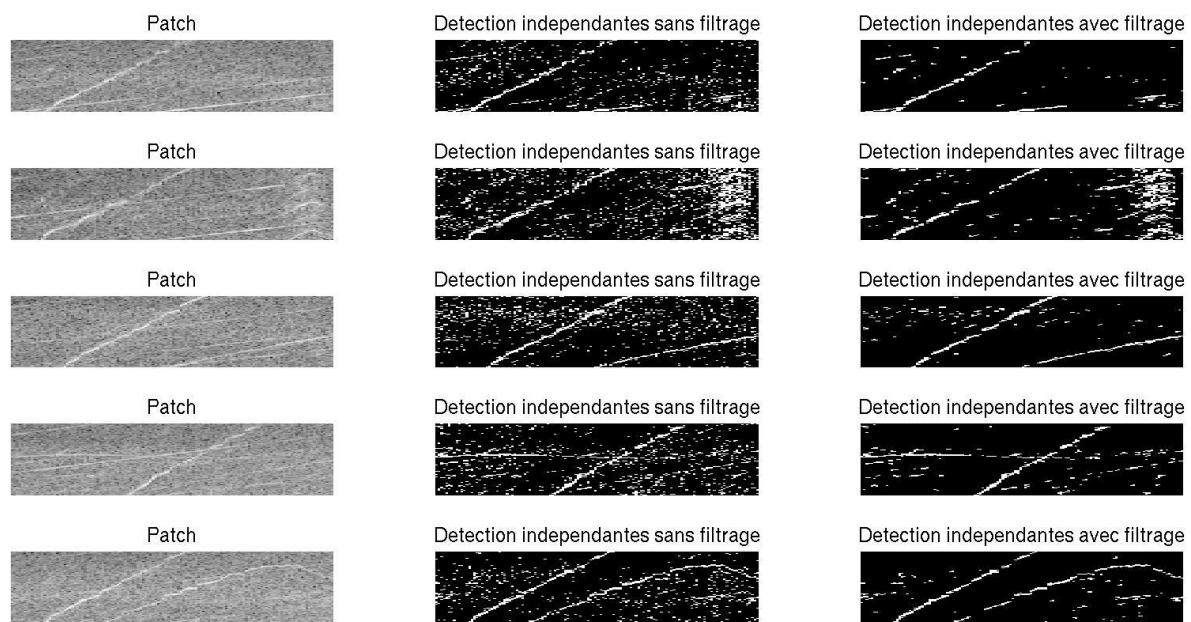


FIGURE 5.12 – Détection des points inusuels sur plusieurs patchs à partir du modèle de normalité définie par la distribution estimée par noyau avec (3ème colonne) et sans (2ème colonne) filtrage de voisinage. Le premier patch (1ère ligne) correspond à un patch sans signature inusuelle, les autres contiennent différentes signatures inusuelles.

TABLE 5.6 – Taux de détection sur l’ensemble des patchs de la base de test à partir du modèle de normalité défini par l’estimation de la distribution des points par noyau avec et sans filtrage par le voisinage.

<i>NA</i>	7.33% \pm 0.09%	<i>NA</i>	7.47% \pm 0.03%	0% \pm 0%	6.44% \pm 0.05%
<i>NA</i>	6.37% \pm 0.08%	<i>NA</i>	5.76% \pm 0.02%	0% \pm 0%	4.29 \pm 0.04%
<i>NA</i>	9.32% \pm 0.15%	<i>NA</i>	9.63% \pm 0.08%	73.3% \pm 0.5%	7.44% \pm 0.06%
<i>NA</i>	8.29% \pm 0.14%	<i>NA</i>	7.21% \pm 0.06%	72.5% \pm 0.9%	5.02% \pm 0.04%
23.1% \pm 0.1%	9.09% \pm 0.12%	3.1% \pm 0.2%	10.3% \pm 0.01%	22.5% \pm 0.4%	7.64% \pm 0.06%
23.7% \pm 0.4%	7.88% \pm 0.12%	4.3% \pm 0.3%	7.34% \pm 0.07%	21.0% \pm 0.4%	5.09% \pm 0.06
43.1% \pm 0.3%	8.1% \pm 0.09%	27.6% \pm 0.3%	9.48% \pm 0.08%	34.1% \pm 0.4%	7.37% \pm 0.15%
45.2% \pm 0.4%	7.0% \pm 0.08%	27.5% \pm 0.3%	6.35% \pm 0.06%	32.5% \pm 0.6%	3.84% \pm 0.11%
30.5% \pm 0.1%	9.75% \pm 0.11%	32.9% \pm 0.1%	12.5% \pm 0.26%	41.8% \pm 0.3%	10.35% \pm 0.1%
31.4% \pm 0.2%	8.69% \pm 0.09%	32.1% \pm 0.3%	9.03 \pm 0.23%	42.4% \pm 0.4%	6.15% \pm 0.08%
55.4% \pm 0.1%	8.41% \pm 0.11%	50.5% \pm 0.1%	9.4% \pm 0.19%	75.3% \pm 0.1%	8.3% \pm 0.31%
55.2% \pm 0.1%	7.25% \pm 0.1%	47.6% \pm 0.2%	5.64% \pm 0.15%	79.0% \pm 0.3%	4.75% \pm 0.2%
59.3% \pm 0.3%	7.1% \pm 0.15%	49.3% \pm 0.3%	6.97% \pm 0.04	60.4% \pm 0.1%	6.49% \pm 0.11%
59.3% \pm 0.4%	6.1% \pm 0.13%	46.3% \pm 0.5%	3.77% \pm 0.04%	64.3% \pm 0.2%	3.3% \pm 0.06%
28.2% \pm 2.3%	5.25% \pm 0.14%	45.3% \pm 0.1%	5.59% \pm 0.22%	74.2% \pm 0.2%	5.47% \pm 0.1%
25.4% \pm 0.6%	4.53% \pm 0.13%	48.0% \pm 0.7%	3.39% \pm 0.17%	70.5% \pm 0.4%	2.62% \pm 0.07%
6.1% \pm 0.1%	7.66% \pm 0.16%	25.2% \pm 0.5%	8.03% \pm 0.14%	47.2% \pm 0.3%	7.59% \pm 0.1%
6.3% \pm 0%	6.72% \pm 0.15%	22.2% \pm 0.5%	5.13% \pm 0.11%	53.3% \pm 0.9%	4.11% \pm 0.08%
<i>NA</i>	7.13% \pm 0.33%	<i>NA</i>	7.95% \pm 0.03%	35.1% \pm 0.1%	8.69% \pm 0.24%
<i>NA</i>	6.25% \pm 0.03%	<i>NA</i>	5.2% \pm 0.01%	31.8 \pm 0.6%	5.18% \pm 0.02%
12.8% \pm 0.1%	6.7% \pm 0.29%	<i>NA</i>	6.55% \pm 0.09%	23.8% \pm 0.1%	7.86% \pm 0.27%
13.3% \pm 0.1%	5.79% \pm 0.27%	<i>NA</i>	4.06% \pm 0.07%	21.7% \pm 0.3%	4.76% \pm 0.22%
6.9% \pm 0.1%	7.72% \pm 0.26%	13.5% \pm 0.1%	8.0% \pm 0.15%	13.2% \pm 0.3%	9.05% \pm 0.14%
5.7% \pm 0.2%	6.79% \pm 0.26%	12.7% \pm 0.6%	5.02% \pm 0.11%	10.0% \pm 1%	5.36% \pm 0.1%
8.0% \pm 0.2%	7.24% \pm 0.16%	20.4% \pm 0.3%	7.47% \pm 0.17%	36.4% \pm 0.6%	7.83% \pm 0.14%
7.4% \pm 0.4%	6.2% \pm 0.15%	20.9% \pm 0.8%	4.7% \pm 0.13%	31.8% \pm 0.6%	4.45% \pm 0.11%
<i>NA</i>	6.94% \pm 0.21%	2.2% \pm 0.2%	7.34% \pm 0.04%	51.7% \pm 1.1%	7.48% \pm 0.4%
<i>NA</i>	5.92% \pm 0.19%	1.0% \pm 0.5%	4.44% \pm 0.03%	66.8% \pm 0.9%	4.13% \pm 0.28%
<i>NA</i>	7.48% \pm 0.12%	41.7% \pm 1.2%	9.88% \pm 0.26%	50.2% \pm 0.5%	11.7% \pm 0.42%
<i>NA</i>	6.71% \pm 0.11%	47.3% \pm 0.4%	7.82% \pm 0.21%	58.4% \pm 0.5%	8.7% \pm 0.36%
20.0% \pm 0.1%	7.74% \pm 0.23%	31.5% \pm 0.2%	9.59% \pm 0.17%	40.3% \pm 0.2%	9.3% \pm 0.17%
20.1% \pm 0.3%	7.06% \pm 0.22%	34.4% \pm 0.3%	7.35% \pm 0.16%	49.1% \pm 0.1%	6.21% \pm 0.13%
83.6% \pm 0.1%	7.51% \pm 0.07%	30.0% + 0.3%	7.56% \pm 0.1%	14.5% \pm 0.1%	7.73% \pm 0.28%
84.5% \pm 0.1%	6.57% \pm 0.06%	37.0% \pm 0.5%	5.19% \pm 0.07%	16.4% \pm 0.3%	4.82% \pm 0.21%
22.9% \pm 0.2%	8.02% \pm 0.16%	3.0% \pm 0.1%	7.32% \pm 0.26%	27.2% \pm 0.2%	7.59% \pm 0.27%
22.8% \pm 0.5%	7.06% \pm 0.14%	4.1% \pm 0.1%	4.72% \pm 0.2%	28.4% \pm 0.1%	4.52% \pm 0.21%

Résultats visuels de détection sur la base de test Ω_{Test}

La figure 5.12 présente les points détectés à partir du modèle de normalité défini à partir de la règle du pouce avec et sans filtrage sur un patchs spécifique issu de la subdivision \mathcal{K}^{128} pour différents spectrogrammes, un dont le patch spécifique est normal (1ère ligne) et 4 spectrogrammes dont les patchs spécifiques contiennent différentes signatures atypiques. Les signatures inusuelles, ainsi que plusieurs points normaux, sont détectées sur les patchs. L'application du filtrage permet de visuellement réduire considérablement le nombre de points faussement détectés tout en conservant les points inusuels.

Le modèle de normalité pour chaque point du spectrogramme est défini sans aucun apriori à partir d'une approche non-paramétrique d'estimation de la densité par un noyau gaussien. La normalité de chaque point est alors caractérisée par cette distribution apprise et vérifiée à partir d'un test d'adéquation de ce point par rapport à la distribution estimée. Le modèle proposé permet la détection des points atypiques mais est bien plus sensible à l'intensité et la position des différents points que ne l'étaient les méthodes basées sur les dictionnaires entraînant une détection plus importante des signatures normales décalées. L'application du filtrage de voisinage est pertinente et permet de réduire les fausses détections.

5.5 Modèle de normalité défini à partir de l'estimation de densité par noyau gamma

Le choix du noyau gaussien permettait d'effectuer les calculs de manière simple. Cependant bien que le noyau n'influence pas fortement la distribution, ce dernier donne une densité aux valeurs négatives tandis que les intensités vibratoires appartiennent à $[0, +\infty[$. L'utilisation du noyau gaussien entraîne donc un biais à l'origine entraînant une perte de la densité à droite. Cela peut entraîner la détection de points normaux. Il faut donc utiliser un noyau n'allouant pas de poids aux valeurs négatives. Aucun noyau symétrique ne respecte ce critère par définition. Les distributions des points des spectrogrammes possèdent des allures proches des lois gamma (Figure 5.2), l'utilisation du noyau gamma pour l'estimation de la densité semble donc pertinent respectant ainsi la positivité des points des spectrogrammes.

5.5.1 Le modèle de normalité

Le modèle de normalité issu du noyau gamma (5.13) est similaire à celui défini à partir du noyau gaussien (5.9). Il est défini par la distribution indépendante estimée par noyau gamma sur

les points normaux d'apprentissage Ω_{App}^{f,N_2} pour chaque point (f, N_2) des spectrogrammes.

$$\forall (f, N_2) \quad \hat{f}_{f,N_2}^{K_\Gamma}(x | \Omega_{App}^{f,N_2}, h) = \frac{1}{\text{card}(\Omega_{App}^{f,N_2})} \sum_{i \in \Omega_{App}^{f,N_2}} K_{\rho_h(x), h}(S_{f,N_2}^i) \quad (5.13)$$

Estimation de l'échelle du noyau

Le seul paramètre à déterminer afin de définir intégralement le modèle de normalité est l'échelle h du noyau gamma. Nous calibrons ce paramètre à partir des deux mêmes approches développées pour l'échelle du noyau gaussien. La première méthode de calibration de l'échelle du noyau est la règle du pouce définie de manière identique à la section précédente :

$$\hat{h}_{RP} = 1.06\hat{\sigma} \text{card}(\Omega_{App})^{-\frac{1}{5}},$$

avec $\hat{\sigma}$ l'écart-type estimée sur les données.

La seconde méthode de calibration de l'échelle du noyau correspond au maximum de vraisemblance par rapport à l'échelle h estimée sur les points normaux de la base de validation Ω_{Val}^{f,N_2} ,

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

La vraisemblance en fonction de différentes valeurs de h est présentée sur la figure 5.13 pour différents points des spectrogrammes. Le maximum de la vraisemblance est atteint en un unique point qui correspond à \hat{h}_{max}^Γ . La valeur de l'échelle h du noyau est donc différente pour chaque point du spectrogramme et dépendante des données en ces coordonnées justifiant la définition d'un modèle de normalité pour chaque points des spectrogrammes.

5.5.2 Le score de détection

Le score pour chaque point du spectrogramme est défini à partir d'un test opposant l'hypothèse \mathcal{H}_0^{f,N_2} de normalité du point de coordonnées (f, N_2) à l'hypothèse \mathcal{H}_1^{f,N_2} caractérisant sa singularité. Identiquement au cas gaussien (5.10), ce test correspond à un test d'adéquation à la distribution estimée (5.14). La p-valeur correspondante à chacun de ces tests définit le score de détection permettant de classifier un point comme normal ou atypique.

$$\begin{aligned} pval_{f,N_2}^{i,K_\Gamma} &= \mathbb{P}_{\mathcal{H}_0^{f,N_2}}(X > S_{f,N_2}^i) \approx \mathbb{P}_{f_{f,N_2}^{K_\Gamma}}(X > S_{f,N_2}^i) \\ &\approx \int_{S_{f,N_2}^i}^{+\infty} \hat{f}_{f,N_2}^{K_\Gamma}(x | \Omega_{App}^{f,N_2}, \hat{h}) dx = \frac{1}{\text{card}(\Omega_{App}^{f,N_2})} \sum_{j \in \Omega_{App}^{f,N_2}} \int_{S_{f,N_2}^i}^{+\infty} K_{\rho_{\hat{h}}(x), \hat{h}}(S_{f,N_2}^j) dx \end{aligned} \quad (5.14)$$

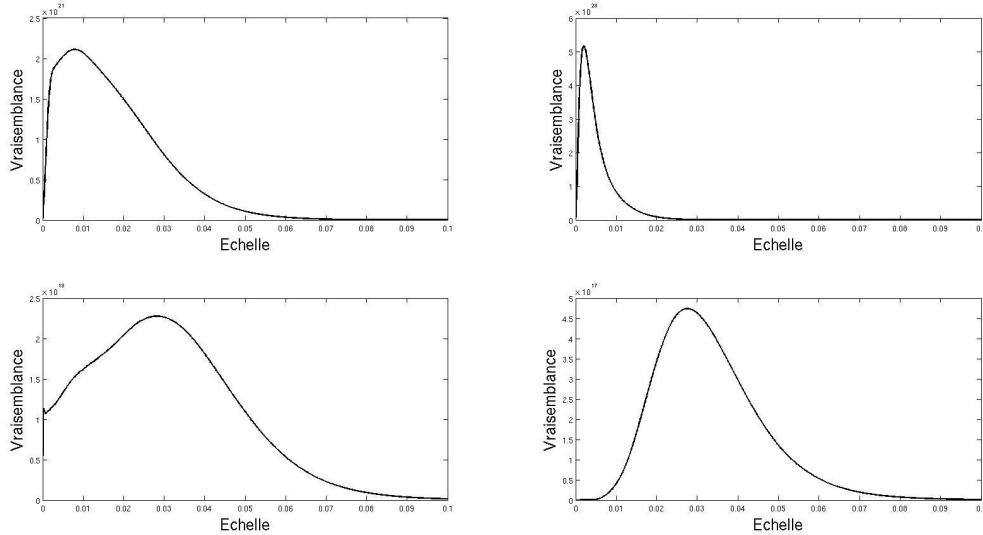


FIGURE 5.13 – Vraisemblance sur Ω_{Val} de la distribution estimée par noyau gamma en fonction de différentes valeurs de h et pour différents points du spectrogramme. Le maximum est atteint en un unique point définissant l'échelle optimale. La valeur du maximum de vraisemblance est différente selon le point étudié, la définition d'un modèle de normalité pour chaque point est donc pertinente.

Le noyau gaussien permettait de se servir de la symétrie afin de considérer le noyau comme la densité d'une gaussienne aussi bien centrée en S_{f,N_2}^j qu'en x . Cela permettait de faciliter le calcul de la p-valeur et de la décomposer comme la moyenne des p-valeurs de chaque noyau centré en la variable S_{f,N_2}^j . Le noyau gamma ne possède pas cette symétrie, entraînant un calcul d'intégrale ne possédant pas de forme fermée. Il n'est pas possible de calculer cette p-valeur par décomposition de la densité estimée en fonction de ses noyaux.

Pour calculer cette p-valeur, nous utilisons la positivité du domaine de définition et estimons $F_{\hat{f}_{f,N_2}^{K_\Gamma}}(S_{f,N_2}^i) = \mathbb{P}_{\hat{f}_{f,N_2}^{K_\Gamma}}(X \leq S_{f,N_2}^i)$, la fonction de répartition de la densité estimée, plutôt que $\mathbb{P}_{\hat{f}_{f,N_2}^{K_\Gamma}}(X > S_{f,N_2}^i)$. Nous utilisons l'approximation de l'intégrale par la méthode des trapèzes (5.15) avec un pas de discrétisation $\xi = 10^{-4}$ équivalant à celui des intensités des spectrogrammes. La densité est ainsi calculée pour toutes les valeurs possibles des spectrogrammes. Nous définissons alors la suite $(x_k)_{k \in \mathbb{N}}$ comme la discrétisation de \mathbb{R}_+ par pas de ξ ($x_{k+1} = x_k + \xi$) sur laquelle la fonction de répartition est estimée.

La méthode des trapèzes ne donne pas une intégrale unitaire sur \mathbb{R}_+ . Afin d'obtenir des p-valeurs bien comprises entre 0 et 1, nous normalisons l'estimation de la fonction de répartition par M , l'estimation de l'intégrale de la densité par la méthode des trapèzes appliquée à l'ensemble de l'espace.

$$\mathbb{P}_{\hat{f}_{f,N_2}^{K_\Gamma}}(X \leq S_{f,N_2}^i) \approx \frac{1}{M} \sum_{k=1}^{k:x_k=S_{f,N_2}^i} \frac{\hat{f}_{f,N_2}^{K_\Gamma}(x_{k+1} | \Omega_{App}^{f,N_2}, \hat{h}) - \hat{f}_{f,N_2}^{K_\Gamma}(x_k | \Omega_{App}^{f,N_2}, \hat{h})}{2} \xi \quad (5.15)$$

avec

$$M = \sum_{k=1}^{+\infty} \frac{\hat{f}_{f,N_2}^{K_\Gamma}(x_{k+1}|\Omega_{App}^{f,N_2}, \hat{h}) - \hat{f}_{f,N_2}^{K_\Gamma}(x_k|\Omega_{App}^{f,N_2}, \hat{h})}{2} \xi$$

Cette approche de définition du modèle de normalité par le noyau gamma possède un temps de calcul 10 fois plus important que pour le modèle défini à partir du noyau gaussien. Dans le cas gaussien, la simplification des calculs permettait d'accélérer le processus et d'obtenir des résultats dans des temps convenables, cette approche ne permet pas de simplification et entraîne plusieurs estimations pour chaque point du spectrogramme.

Nous avons alors :

$$pval_{f,N_2}^{i,K_\Gamma} \approx \mathbb{P}_{\hat{f}_{f,N_2}^{\Gamma}}(X > S_{f,N_2}^i) \approx 1 - \mathbb{P}_{\hat{f}_{f,N_2}^{\Gamma}}(X \leq S_{f,N_2}^i).$$

La décision de singularité des points s'effectue par la comparaison des p-valeurs à un seuil de détection s (5.16) identique pour tous les points (par manque d'annotations sur tous les points du spectrogramme) avec une application du filtrage de voisinage \mathcal{V}_{f,N_2} identique à (5.12).

$$\tilde{Y}_{f,N_2}^{i,K_\Gamma} = \mathbb{1}\{pval_{f,N_2}^{i,K_\Gamma} \leq s\} \quad (5.16)$$

5.5.3 Calibration des seuils de détection

La figure 5.14 présente les taux des détections à partir des modèles définis par l'estimation de la distribution par noyau gamma des différentes classes de points des données de validation Ω_{Val} de la base \mathcal{B}_1 en fonction de différents seuils de détection. La calibration a été effectuée pour les 2 méthodes de définition de l'échelle.

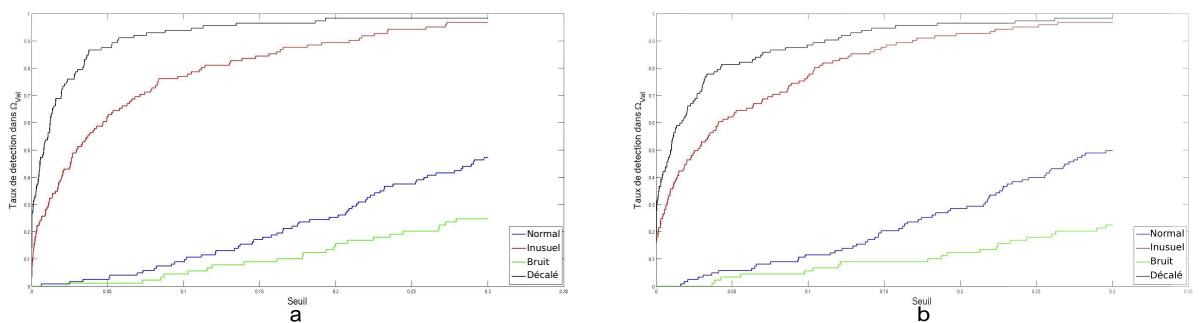


FIGURE 5.14 – Taux de détection en fonction du seuil de détection des différentes classes de points de la base de validation Ω_{Val} à partir des modèles de normalité définis par l'estimation de la distribution avec noyau gamma en estimant l'échelle par la règle du pouce (a) et par le maximum de vraisemblance (b). Les résultats entre les deux modèles sont proches. Les points atypiques (rouge) et les points normaux décalés (noir) sont fortement détectés pour de petites valeurs de seuil tandis que les points normaux (bleu) et le bruit (vert) le sont faiblement. Le seuil de détection est donc défini au niveau de ces valeurs.