

Les deux méthodes de définition d'échelle donnent des résultats relativement proches. Les pentes des détections des points normaux et du bruit (toutes deux concernant des points normaux) sont plus faibles que pour le cas gaussien (il en est de même pour les points inusuels et les points décalés). Nous prenons donc un seuil de détection plus important pour le modèle défini dans cette section avec $s = 0.15$. Sur la base de validation ce seuil permet un bon niveau de détection des points inusuels d'environ 80% mais également des points décalés comme les méthodes précédentes.

5.5.4 Résultats sur la base de test Ω_{Test}

Les temps de calcul des modèles définis à partir du noyau gamma sont trop longs. Nous présentons uniquement les résultats sur les données tests Ω_{Test} du patch spécifique de la base \mathcal{B}_1 sous la forme du taux de détection des points annotés et de la représentation visuelle des points détectés sur l'intégralité du patch. Nous ne donnons pas les résultats sur l'ensemble des patches à partir de l'erreur E_j^i .

Résultats de classification des points annotés de Ω_{Test}

Le tableau 5.7 présente les taux de détections des différentes classes de points de la base \mathcal{B}_1 à partir du modèle de normalité défini par le noyau gamma pour les deux méthodes d'estimation de l'échelle. Le modèle est défini pour plusieurs ensembles d'apprentissage. Le noyau gamma apporte une bonne caractérisation des distributions ponctuelles des points entraînant une très bonne détection des points atypique de plus de 90%. Les points décalés sont également fortement détectés. Plusieurs points normaux sont également détectés à tort par l'approche, cette fausse détection peut être réduite à partir du filtrage de voisinage. Les effets du filtrage de voisinage sont observables sur la représentation visuelle sur l'ensemble du patch des résultats de détection (Figure 5.15). Les 2 méthodes de définition d'échelle donnent des résultats proches, la définition du seuil à partir du maximum de vraisemblance est légèrement supérieure sur les points annotés. Ces résultats de détection sont légèrement supérieurs au modèle de normalité défini à partir du noyau gaussien, cependant ce dernier est 10 fois plus rapide que le modèle défini à partir du noyau gamma tout en donnant des résultats proches.

TABLE 5.7 – Taux de détection à partir de la densité estimée par noyau gamma 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"	18.5% \pm 1.1%	25.1% \pm 1.1%
Classe "inusuel"	90.7% \pm 0%	96.3% \pm 1.3%
Classe "bruit"	6.25% \pm 0%	6.25% \pm 0%
Classe "décalé"	98.5% \pm 1.4%	97.9% \pm 1.1%

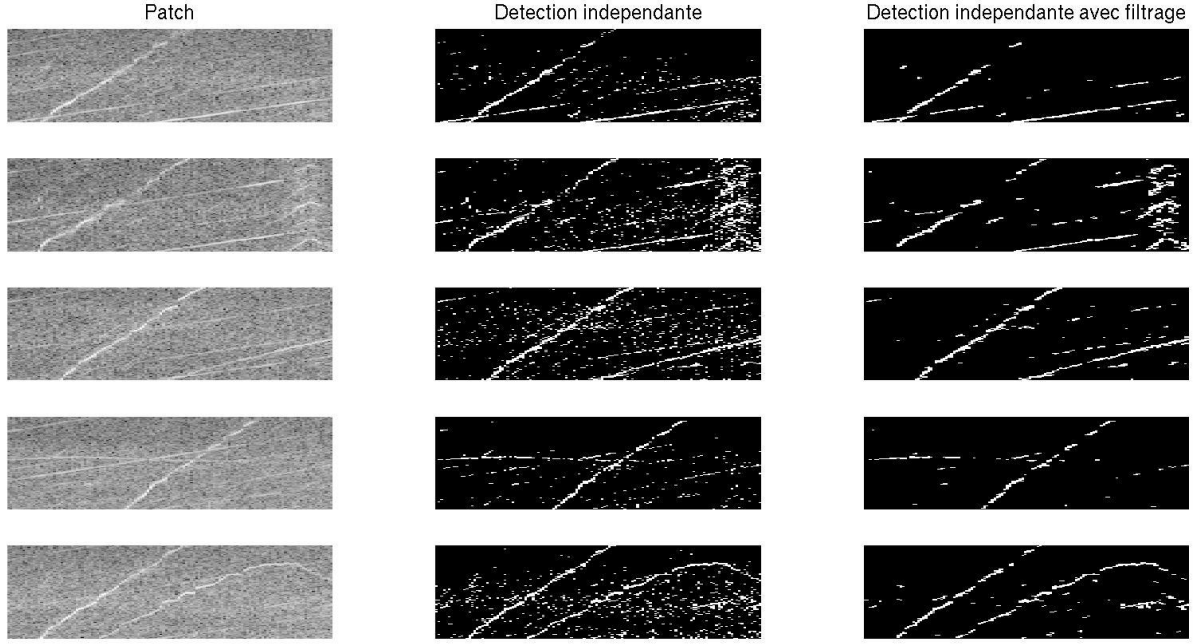


FIGURE 5.15 – Détection des points atypiques sur plusieurs patches avec et sans (1ère ligne) signature atypique à partir du modèle de normalité définie par la distribution estimée par noyau gamma avec (3ème colonne) et sans (2ème colonne) filtrage de voisinage. Les signatures atypiques sont mises en évidence sur les patches. Les signatures liées au N_1 et au N_2 de fortes intensités sont également détectés. Le filtrage de voisinage permet de réduire les fausses détections.

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

La figure 5.15 présente l'intégralité des points détectés pour plusieurs patches, le premier ne contenant pas de signature inusuelle et les suivants contenant différents types de signatures atypiques (Figure 2.3). Les résultats sont présentés avec et sans filtrage de voisinage. Les signatures sont mises en évidence dans l'application du modèle sans filtrage sur les patches contenant des signatures inusuelles. Cependant plusieurs points normaux sont également détectés par l'approche, principalement les points normaux décalés illustrant la sensibilité de cette dernière aux positions des signatures. L'utilisation du filtrage de voisinage permet de réduire efficacement ces fausses détections tout en conservant la détection visuelle des signatures atypiques, permettant ainsi une meilleure mise en évidence des points inusuels. Cependant, comme nous pouvons le voir aussi bien sur le patch normal que les patches inusuels, cette approche entraîne la détection des points normaux issus des signatures liées au N_2 . Elle reste donc très sensible à l'intensité vibratoire et peut entraîner des détections aux niveaux de raies normales de fortes intensités.

La caractérisation de la normalité à partir de la distribution de chaque point estimée avec des noyaux gamma permet une détection intéressante des signatures atypiques présentes sur les patch pour les mettre en évidence. Cela entraîne également la détection des points sur les signatures vibratoires liées au N_1 et au N_2 , montrant une forte sensibilité aux intensités vibratoires et à leurs positions. L'application du filtrage de voisinage réduit le taux de fausses détections visuellement.

5.6 Conclusions

Dans ce chapitre, nous avons présenté la mise en place d'une détection ponctuelle des points des spectrogrammes à travers un modèle de normalité caractérisant chacun de ces points de manière indépendante. Ce modèle de normalité est défini pour chaque point à travers la distribution des points normaux. Bien que le modèle soit défini de manière ponctuelle, nous avons gardé la structure en patch afin de présenter les résultats.

La détection d'anomalies s'effectue à travers de tests statistiques d'adéquation par rapport au modèle de normalité. Cela entraîne donc un nombre de tests égal au nombre de points du spectrogramme et donc des problématiques de tests multiples. Cette problématique est résolue dans la littérature en considérant l'ensemble des tests réalisés globalement comme un unique ensemble à contrôler au niveau α et non unitairement. Les tests multiples sont cependant trop conservateurs et ont tendance à ne pas rejeter l'hypothèse de normalité entraînant une forte proportion de non détections. Cet effet est contradictoire avec le but recherché de détecter toutes les signatures atypiques présentes sur les spectrogrammes. Nous avons donc écarté l'approche basée sur les tests multiples et avons opté sur l'utilisation d'un filtrage basé sur les voisinages des points pour limiter les fausses détections. Un point est réellement détecté uniquement si 2 de ses voisins directs sont détectés également. Ainsi les points isolés ne sont pas détectés.

Le modèle de normalité a été défini dans un premier temps de manière paramétrique à partir de la distribution gamma. Cependant, cela supposait une distribution unimodale des points ce qui n'est pas nécessairement le cas dans une base de données volumineuse. Nous avons donc opté pour une caractérisation non-paramétrique sans a priori de la distribution des points du spectrogramme à partir de l'estimation de la densité par noyau. Deux noyaux ont été étudiés lors du développement, un noyau gaussien permettant une facilité de calcul mais ayant également une propriété de symétrie engendrant une densité non nulle pour les valeurs négatives, un noyau gamma permettant de pallier cela mais complexifiant les calculs et entraînant une impossibilité de mise à jour efficace du modèle avec de nouvelles données normales. Les temps de calcul des modèles basés sur les noyaux gammas sont 10 fois plus longs que ceux à partir des noyaux gaussiens pour des résultats proches. Il est donc plus pertinent d'utiliser le noyau gaussien dans un cadre industriel. Les modèles développés permettent une détection très efficace des points atypiques potentiellement présents sur les patches et donc des signatures inusuelles. Cependant cette approche est fortement sensible à l'intensité des points et au décalage de ces derniers entraînant de fausses détections (faux positifs) de ces points. Il serait plus pertinent de définir le seuil de normalité sur chacun des points séparément cependant cela requiert une annotation en classes de l'ensemble des points du spectrogramme.

La méthode de détection développée dans ce chapitre considère d'abord les points comme indépendants puis tient compte du voisinage des points détectés via le filtrage. Dans la suite, nous ajoutons la dépendance des points voisins dans la définition du modèle de normalité en le définissant d'emblée par la distribution conditionnelle des points en fonction de leurs voisinages.

Chapitre 6

Analyse ponctuelle conditionnellement au voisinage des spectrogrammes

6.1 Considération des points voisins

6.1.1 Dépendance par rapport au voisinage des points

Dans ce chapitre, nous décidons d'inclure la dépendance entre les points directement dans le modèle de normalité à travers la définition de la distribution des points et de leur voisinage. Il ne semble pas pertinent d'utiliser la dépendance du point par rapport à l'ensemble du reste du spectrogramme ou du patch. La figure 6.1 présente la corrélation entre les différents points d'un patch et leur voisins de rayon maximal 3 (soient 48 voisins dont la distance en norme l_∞ entre le point considéré et les voisins est au plus 3). Les différents voisins ont été ordonnés selon leurs distances en norme infinie par rapport au point, les premiers (ordre 1) sont les voisins de la première couronne autour du point, les suivants (ordre 2) correspondent à la seconde couronne et les derniers (ordre 3) à la troisième couronne. Nous pouvons remarquer que les corrélations sont plus importantes pour les voisins d'ordre 1. Certains voisins éloignés d'ordre 2 et 3 présentent des corrélations significatives avec le point considéré, cependant la plupart des points présents dans ces couronnes ont des corrélations proches de 0. Nous pouvons donc supposer que plus nous nous éloignons du point considéré plus les corrélations avec la majorité des points éloignés deviennent nulles. Il n'est donc pas utile de considérer des voisins trop éloignés du point étudié dans la définition du modèle. De plus, en augmentant le rayon de voisinage, le nombre de voisins à considérer augmente fortement pouvant entraîner des complications de calcul dû au nombre important de variables. Il est donc important de limiter ce rayon de voisinage.

Nous remarquons également que ces corrélations calculées pour tous les points du patch ne font pas apparaître de corrélation négative entre les points. Cela conforte l'idée selon laquelle les points du spectrogramme peuvent interagir entre eux de manière positive mais pas négative.

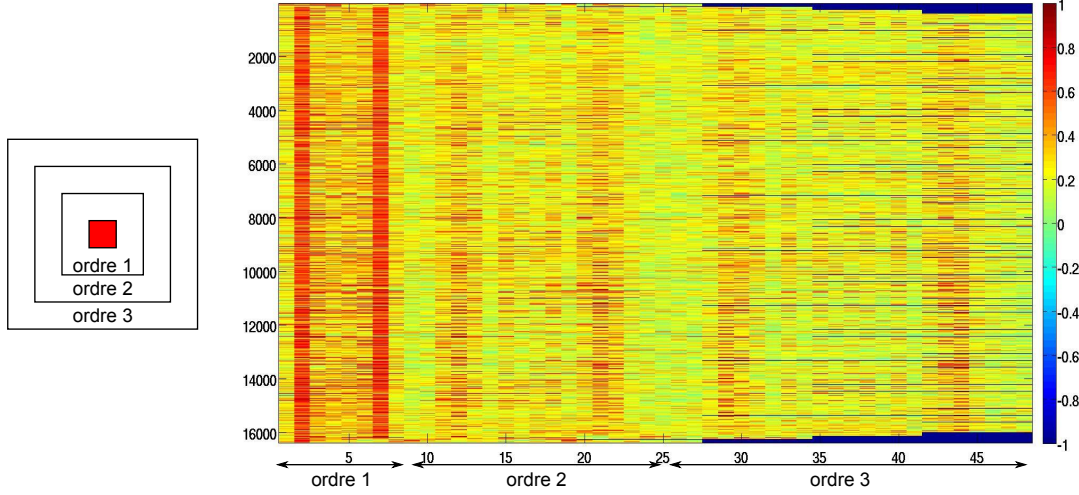


FIGURE 6.1 – Corrélation entre les points du patch et leurs voisins respectifs d'ordre 3 au maximum. Schéma de la signification des voisins des différents ordres. Les corrélations diminuent lorsque la distance entre les points augmente. Les majorité des points voisins d'ordre 3 ont des corrélations proches de 0 avec le point étudié.

6.1.2 Modèle de normalité par rapport au voisinage

Le modèle de normalité correspond à la distribution estimée de chaque point du spectrogramme sachant la distribution des points voisins. La classification des points s'effectue comme précédemment par rapport à la p-valeur du test d'adéquation par rapport à la distribution, cela équivaut à déterminer un classifieur binaire sur le point étudié avec comme variables explicatives les points voisins \mathcal{V}_{f,N_2} .

$$\hat{Y}_{f,N_2} = g(S_{f,N_2} | S_{\mathcal{V}_{f,N_2}}) = g(S_{f,N_2} | \mathcal{V}_{f,N_2})$$

avec g le classifieur et \mathcal{V}_{f,N_2} les coordonnées des voisins du point (f, N_2) avec un rayon de voisinage à définir et par abus de langage, pour simplifier les notations, les intensités vibratoires à ces mêmes coordonnées (donc $\mathcal{V}_{f,N_2} \equiv S_{\mathcal{V}_{f,N_2}}$). Les critères d'erreur sont les mêmes que ceux définis dans les chapitres précédents.

Nous sommes toujours dans un cadre de détection de nouveautés, cette fonction g est donc définie à partir de la distribution conditionnelle de normalité du point par rapport à son voisinage. Différents paramètres sont à prendre en compte dans l'estimation de cette distribution ainsi que dans la structure du voisinage. En effet, cette dernière ne correspond pas nécessairement à un cercle autour du point mais peut provenir d'information a priori.

L'intérêt d'étudier les probabilités conditionnelles en fonction du voisinage dans un cadre de détection de nouveautés provient de la structure de nos signatures atypiques. Ces dernières correspondent à des raies sur les spectrogrammes, ainsi plusieurs points voisins des spectrogrammes sont donc à détecter. La distribution conditionnelle des points est estimée une nouvelle fois par des méthodes d'estimation de densité par noyau [11, 62].

6.2 Estimation de densité conditionnelle par noyau

6.2.1 L'estimation de densité par noyau dans un cadre multidimensionnel

La formule générale d'estimation de densité par noyau est proche en d dimensions de la formule en 1 dimension, la principale différence provient de l'utilisation d'une matrice pour définir le paramètre d'échelle. Pour un jeu de données $X = (X_1, \dots, X_n)$ avec $X_i = (X_i^1, \dots, X_i^d)$, l'estimation de la densité multidimensionnelle de ces données au point $x = (x_1, \dots, x_d)$ est donnée par :

$$\hat{f}_{X,H}(x = (x_1, \dots, x_d)) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\det(H)^{\frac{1}{2}}} K\left(H^{-\frac{1}{2}}(x - X_i)\right), \quad (6.1)$$

avec K le noyau multidimensionnel et H la matrice d'échelle symétrique définie positive.

La méthode est similaire à son homologue unidimensionnel, chaque observation du jeu de données multidimensionnelles intervient dans l'estimation de la densité en tout point $x \in \mathbb{R}^d$ avec des contributions plus ou moins importantes suivant la distance entre ce points et les observations.

Si le noyau K respecte les propriétés de densité (intégrale du noyau unitaire), de positivité, de symétrie, de moment d'ordre 2 fini alors l'estimateur est consistant en tout point x et converge en probabilité vers la densité à estimer. Le biais et la variance de cet estimateur sont donnés par :

$$\begin{aligned} \text{biais}(\hat{f}_{X,H}(x)) &= \frac{1}{2} \mu_2(K) \text{tr}(H^T \mathcal{H}_f(x) H), \\ \text{var}(\hat{f}_{X,H}(x)) &= \frac{1}{n \det(H)} \|K\|^2 f(x), \end{aligned}$$

avec \mathcal{H}_f la hessienne de la densité f à estimer et $\mu_2(K) = \int x x^T K(x) dx$.

Le noyau gaussien multidimensionnel est le plus utilisé dans la littérature pour ses propriétés intéressantes de symétrie et de facilité de calcul.

$$K_{N(\mu, \Sigma)}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

La matrice H de l'échelle correspond au paramètre de la méthode d'estimation de densité multivariée par noyau. La forme de la matrice est le premier élément à définir. Elle peut être diagonale avec (h_1, \dots, h_d) les éléments diagonaux, donc de la forme :

$$H = \begin{pmatrix} h_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & h_d \end{pmatrix}.$$

Dans ce cas, le noyau symétrique peut s'écrire sous une forme multiplicative à partir de (6.1),

où chaque terme du produit correspond à un noyau unidimensionnel appliqué à la dimension correspondante [112],

$$K_H(x) = K_{h_1}(x_1) \times K_{h_2}(x_2) \times \dots \times K_{h_d}(x_d).$$

Le cas où tous les h_i sont égaux correspond à une sphère. Il est également possible de considérer une matrice d'échelle H pleine,

$$H = \begin{pmatrix} h_{11} & \dots & h_{1d} \\ \vdots & \ddots & \vdots \\ h_{d1} & \dots & h_{dd} \end{pmatrix}.$$

Dans ce cas, le noyau possédera une orientation dépendante de la matrice H . Ce cadre n'a pas d'analogie avec le cas univarié. Cette approche correspond à préorienter les données et utiliser un noyau diagonal. Une fois la forme de la matrice définie, il reste à estimer les différentes valeurs de cette dernière.

6.2.2 Calibration de la matrice d'échelle

L'estimation de la matrice d'échelle H est plus complexe que pour son homologue univarié, ce qui l'a rendu bien moins intéressante que son homologue durant plusieurs années. La structure de la matrice d'échelle H étant définie, différentes approches ont été mises en place pour définir les valeurs. Les différentes mesures de l'erreur d'estimation sont identiques au cas unidimensionnel tel que MISE et son approximation asymptotique AMISE. Il s'agit des deux grandeurs les plus couramment utilisées pour déterminer la matrice d'échelle H .

Il est possible d'utiliser la règle du pouce de manière analogue au cas univarié [104] en prenant une matrice d'échelle dépendante de la dimension d , du nombre d'observations n pour l'estimation de la densité et de la matrice de covariance estimée sur les données $\hat{\Sigma}$,

$$\hat{H} = n^{-\frac{1}{d+4}} \hat{\Sigma}^{\frac{1}{2}} \quad (6.2)$$

Une seconde estimation de H basée sur la règle du pouce est donnée dans [58] dans un cadre gaussien entraînant cependant un lissage trop important de la distribution.

$$\hat{H} = \left(\frac{4}{n(d+2)} \right)^{\frac{1}{d+4}} \hat{\Sigma}^{\frac{1}{2}}.$$

Des méthodes plus sophistiquées similaires à l'estimation de densité par noyau en 1 dimension ont également été mises en place. Elles se regroupent suivant 2 différentes approches, les méthodes plug-in et les méthodes de validation croisée. La première approche permet de trouver une estimation H en approximant les termes dépendant de la densité à estimer dans AMISE [112, 24]. La seconde approche consiste à appliquer la méthode de validation croisée "leave one out" de manière analogue au cas univarié pour trouver le minimum des fonctionnelles MISE ou

AMISE [44]. Une autre catégorie d'approche consiste à utiliser le maximum de vraisemblance pour trouver la matrice H optimale sur les données [119, 74].

6.2.3 L'estimation de la densité conditionnelle par noyau

L'estimation de la densité conditionnelle par noyau provient de l'application de l'estimation par noyau multidimensionnel à la formule de base de la densité conditionnelle [62]

$$f_{Y|X} = \frac{f_{XY}}{f_X} \approx \frac{\hat{f}_{XY}}{\hat{f}_X},$$

$f_{Y|X}$, f_{XY} et f_X sont respectivement la densité conditionnelle de $Y|X$, la densité jointe du couple (X, Y) et la densité de la variable aléatoire X .

Il s'agit donc d'utiliser l'estimation de la densité multidimensionnelle par noyau pour estimer aussi bien f_{XY} et f_X . L'estimateur de $f_{Y|X}(Y|X)$ à partir des données $(X, Y) = \{(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}\}_{i=1, \dots, n}$ avec $X_i = (X_i^1, \dots, X_i^d)$ est donc

$$\hat{f}_{Y|X}(y|x) = \frac{\sum_{i=1}^n K_{H_{XY}}((x, y) - (X_i, Y_i))}{\sum_{i=1}^n K_{H_X}(x - X_i)}.$$

$K_{H_{XY}}$ correspond au noyau permettant d'estimer la loi jointe de (X, Y) en dimension $d + 1$ et K_{H_X} correspond au noyau utilisé pour estimer la densité de la variable X en dimension d .

Dans la littérature, les noyaux utilisés sont souvent d'échelle diagonale permettant de les mettre sous leurs formes multiplicatives, les noyaux définis au dessus se mettent donc sous la forme :

$$K_{H_{XY}}(x^1, \dots, x^d, y) = K_{H_Y}(y) \prod_{i=1}^d K_{H_{X^i}}(x^i) \quad K_{H_X}(x^1, \dots, x^d) = \prod_{i=1}^d K_{H_{X^i}}(x^i)$$

avec $K_{H_{X^i}}$ le noyau monodimensionnel avec pour valeur d'échelle $H_X(i, i)$ l'élément diagonale i de la matrice d'échelle. L'estimation de la densité conditionnelle (6.3) se reformule donc dans ce cadre :

$$\begin{aligned} \hat{f}_{Y|X}(y|x) &= \frac{\sum_{i=1}^n K_{H_Y}(y - Y_i) \prod_{j=1}^d K_{H_{X^j}}(x^j - X_i^j)}{\sum_{i=1}^n \prod_{j=1}^d K_{H_{X^j}}(x^j - X_i^j)} \\ &= \sum_{i=1}^n w_i(x|X_i^1, \dots, X_i^d) K_{H_Y}(y - Y_i) \end{aligned} \quad (6.3)$$

avec

$$w_i(x|X_i^1, \dots, X_i^d) = \frac{\prod_{j=1}^d K_{H_{X^j}}(x^j - X_i^j)}{\sum_{i=1}^n \prod_{j=1}^d K_{H_{X^j}}(x^j - X_i^j)} \quad (6.4)$$

Il s'agit donc d'une somme pondérée sur toutes les observations utilisées pour calibrer la distribution du noyau unidimensionnel appliqué à la variable cible Y par rapport à la contribution de

chaque observation w_i des noyaux associés aux variables conditionnelles X .

Le noyau le plus couramment utilisé pour estimer la densité conditionnelle est le noyau gaussien. Informatiquement il permet un calcul rapide de la densité conditionnelle en utilisant directement la densité gaussienne multidimensionnelle. H correspondant alors à la matrice de covariance de la distribution.

6.3 Le modèle de normalité défini à partir du voisinage d'ordre 1

Le modèle correspond à la densité conditionnelle du point en fonction de ses voisins. Il est défini sur les données normales pour chaque point du spectrogramme en considérant l'ensemble des points au voisinage du point étudié défini par une structure. Deux éléments sont donc à prendre en compte dans la définition de ce modèle de normalité :

- les points voisins utilisés comme paramètres de mise en place du modèle, il s'agit de la structure du voisinage prise en compte,
- la structure de la matrice d'échelle et l'estimation de ses coefficients.

6.3.1 La structure du voisinage

La figure 6.1 présente les corrélations entre un point étudié et son voisinage de rayon allant de 1 à 3. Les corrélations entre la majorité des points voisins et le point considéré diminuent en fonction de leurs distances l_∞ . Les intensités vibratoires ne sont pas toutes liées les unes aux autres, une intensité vibratoire à fréquence et régime donnés n'a aucune relation avec les intensités à des fréquences et régimes complètement différents. Il est donc logique de se limiter au voisinage proche (en norme l_∞) du point étudié. Nous ne disposons pas d'un nombre important de données pour estimer la densité de nos données. Cependant l'estimation de cette dernière impose l'estimation de la densité jointe dans un espace de grande dimension qui augmente en fonction du rayon de voisinage. Plus le nombre de variables considérées est important, plus le nombre de données nécessaires pour avoir une estimation fiable de la densité jointe augmente. Le nombre de variables explicatives est de 8 pour un rayon de voisinage de 1, 24 pour un rayon de 2 et 48 pour un rayon de convergence de 3. Ainsi pour faciliter et rendre pertinente l'estimation de la distribution conditionnelle, nous utilisons les voisins du premier ordre pour le modèle de normalité (Figure 6.2). Les variables de rayon de voisinage supérieur ou égal à 2 sont moins corrélées au point étudié (sur plusieurs directions) et apportent un nombre trop important de variables explicatives.