

Table des figures

1	Sources de vibrations normales et anormales.	3
2	Spectrogramme indexé en ordre.	4
3	Illustration des approches multi-class et one-class.	7
1.1	Circulation de l'air lors du fonctionnement du moteur	14
1.2	Signaux vibratoires et tachymétriques acquis sur banc d'essai.	17
1.3	Signaux bruts d'un moteur sans endommagement et avec endommagement	18
1.4	Spectrogrammes avec et sans endommagement rééchantillonnés en N_2	20
1.5	Variabilité de la relation entre le N_1 et le N_2	20
1.6	Annotation manuelle des spectrogrammes par des experts.	22
1.7	Les différents cas de figure d'extraction des zones atypiques (rouge) sur les annotations manuelles (jaune).	24
1.8	Illustration de la base de données et de ses éléments.	25
1.9	Proportion pour chaque point du spectrogramme du nombre de moteurs dont le point étudié appartient à une zone atypique extraite.	26
1.10	Subdivision du spectrogramme en patch.	29
1.11	Extraction du même patch des spectrogrammes de notre base de données avec encadrement des signatures inusuelles.	31
1.12	Histogrammes des intensités vibratoires d'un patch spécifique avec et sans signatures atypiques.	37

1.13	Distance entre les histogrammes des mêmes patches pour différents spectrogrammes.	38
1.14	Projection des histogrammes de représentation sur les 3 premières composantes principales de la ACP.	39
2.1	Processus de détection de nouveautés/anomalies.	42
2.2	Ratio de données normales et atypiques sur les différents patches.	51
2.3	Patches de la base de test utilisés pour présenter les résultats visuels des différentes approches.	52
3.1	Caractérisation d'une forme courbe à différentes échelles.	62
3.2	Atomes du dictionnaire des curvelets à différentes échelles.	65
3.3	Reconstruction d'un patch normal et résidus associés à partir de 100%, 50%, 10%, 5% et 1% des coefficients dans le dictionnaire des curvelets.	67
3.4	Reconstruction d'un patch contenant une signature atypique et résidus associés à partir de 100%, 50%, 10%, 5% et 1% des coefficients dans le dictionnaire des curvelets.	67
3.5	Matrices des distances des coefficients des curvelets et de différence du nombre d'atomes activés entre deux représentations des patches.	68
3.6	Proportion des atomes de curvelets activés sur l'ensemble de la base de données. .	70
3.7	Dimension de $Supp^*$ en fonction de Q et reconstruction de 2 patches inusuels à partir des atomes de $Supp^*$ pour différentes valeurs de Q	72
3.8	Supports définis directement sur la transformée en curvelet et à partir d'ADMM.	74
3.9	Reconstruction à partir des supports de normalité.	75
3.10	Reconstruction de 3 patches à partir du dictionnaire $\mathcal{D}_{Supp^*}^C$ avec les résidus de reconstruction.	77
3.11	Comparaison entre la vérité terrain et la classification à partir du dictionnaire des curvelets.	80
3.12	Histogrammes des résidus définis à partir du modèle de normalité des curvelets. .	81

3.13 Taux de détection des différentes classes de points dans Ω_{Val} en fonction du seuil sur les p-valeurs des tests d'adéquation.	85
3.14 Taux de détection des différentes classes de points dans Ω_{Val} à partir du seuil sur les p-valeurs en fonction du nombre de voisins devant être détectés dans les résidus négatifs.	85
3.15 Détection des points inusuels à partir du seuil sur les p-valeurs.	86
3.16 Taux de détection des différentes classes de points dans Ω_{Val} en fonction du nombre de points n_{max} utilisé pour modéliser la distribution des valeurs extrêmes.	87
3.17 Taux de détection des différentes classes de points dans Ω_{Val} à partir du seuil défini sur la distribution extrême en fonction du nombre de voisins devant être détectés dans les résidus négatifs.	87
3.18 Détection des points inusuels à partir du seuil défini sur la distribution extrême.	88
3.19 Détection des points inusuels sur différents patchs à partir des résidus positifs et négatifs.	92
4.1 Erreur de reconstruction en validation en fonction du rang de la NMF.	103
4.2 Atomes du dictionnaire de la NMF appris sur les données normales pour 2 différents patchs.	105
4.3 Représentation de 2 patchs contenant des signatures inusuelles à partir de \mathcal{D}^{NMF}	106
4.4 Matrices des distances euclidiennes des représentations des patchs dans le dictionnaire \mathcal{D}_j^{NMF}	108
4.5 Comparaison entre la vérité terrain et la classification à partir du dictionnaire issu de la NMF.	110
4.6 Histogrammes des résidus de reconstruction par la NMF	111
4.7 Taux de détection pour les différentes classes de points de la base de validation Ω_{Val} en fonction du seuil sur les p-valeurs.	113
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.	114

4.9	Taux de détection pour les différentes classes de points de la base de validation Ω_{Val} en fonction du nombre de résidus utilisés pour calibrer la distribution des valeurs extrêmes.	114
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.	118
5.1	Analyse ponctuelle des points du spectrogramme par patch.	126
5.2	Histogrammes des intensités des différents points du spectrogramme.	128
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.	130
5.4	Détection des points atypiques sur des patches contenant des signatures inusuelles par un seuillage sur les p-valeurs des tests statistiques au niveau de chaque point.	131
5.5	Courbes des p-valeurs ordonnées des mêmes patches de différents spectrogrammes.	131
5.6	Détection à partir de la modélisation gamma des points de différents patches avec et sans filtrage.	134
5.7	Résultats des tests statistiques de Mann-Whitney vérifiant l'hypothèse gamma pour chaque point du spectrogramme.	134
5.8	Illustration de l'estimation de densité par noyau.	135
5.9	Vraisemblance sur Ω_{Val} de la distribution estimée en fonction de différentes valeurs de h et pour différents points du spectrogramme.	141
5.10	Densités estimées pour différents points du spectrogramme à partir de la règle du pouce et du maximum de vraisemblance.	141
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.	143
5.12	Détection des points inusuels sur plusieurs patches à partir du modèle de normalité définie par la distribution estimée par noyau avec et sans filtrage de voisinage.	145
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.	149

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.	150
5.15	Détection des points atypiques sur plusieurs patches à partir du modèle de normalité définie par la distribution estimée par noyau gamma.	152
6.1	Corrélation entre les points du patch et leurs voisins respectifs d'ordre 3 au maximum.	156
6.2	Structure du voisinage direct.	161
6.3	Convergence des p-valeurs par rapport au nombre de points utilisés pour estimer la distribution conditionnelle.	166
6.4	Taux de détection des différentes classes de points de la base de validation Ω_{Val} à partir du modèle de normalité défini par les distributions conditionnelles.	166
6.5	Détections sur les patches entiers à partir du modèle de normalité défini par les distributions conditionnelles.	169
6.6	Différentes directions linéaires de voisinage possibles en considérant un rayon de voisinage de 3.	170
6.7	Taux de détection des différentes classes de points de base de validation annotée ponctuellement Ω_{Val} en fonction du seuil de détection.	174
6.8	Détection des points inusuels sur 4 différents patches de Ω_{Val} pour différentes valeurs de seuil.	175
6.9	Détection des points inusuels sur un patch sans signature inusuelle avec caractérisation des composantes connexes.	177
6.10	Détection des points inusuels sur un patch contenant un premier type de signatures inusuelles avec caractérisation des composantes connexes.	178
6.11	Détection des points inusuels sur un patch contenant un second type de signatures inusuelles avec caractérisation des composantes connexes.	178
6.12	Détection des points inusuels sur un patch contenant un troisième type de signatures inusuelles avec caractérisation des composantes connexes.	179
6.13	Détection des points inusuels sur un patch contenant un quatrième type de signatures inusuelles avec caractérisation des composantes connexes.	179

6.14	Détection du patch sans signature inusuelle à partir de la \hat{Y}^{NMF} , $\hat{Y}^{\mathcal{C}}$, $\hat{Y}^{K_{\mathcal{N}}}$, $\hat{Y}^{K_{\nu}}$ et $\hat{Y}^{K_{\vec{\nu}}}$, ainsi que de l'approche mixte \hat{Y}	182
6.15	Détection du patch contenant un premier type de signature inusuelle à partir de la \hat{Y}^{NMF} , $\hat{Y}^{\mathcal{C}}$, $\hat{Y}^{K_{\mathcal{N}}}$, $\hat{Y}^{K_{\nu}}$ et $\hat{Y}^{K_{\vec{\nu}}}$, ainsi que de l'approche mixte \hat{Y}	182
6.16	Détection du patch contenant un second type de signature inusuelle à partir de la \hat{Y}^{NMF} , $\hat{Y}^{\mathcal{C}}$, $\hat{Y}^{K_{\mathcal{N}}}$, $\hat{Y}^{K_{\nu}}$ et $\hat{Y}^{K_{\vec{\nu}}}$, ainsi que de l'approche mixte \hat{Y}	183
6.17	Détection du patch contenant un troisième type de signature inusuelle à partir de la \hat{Y}^{NMF} , $\hat{Y}^{\mathcal{C}}$, $\hat{Y}^{K_{\mathcal{N}}}$, $\hat{Y}^{K_{\nu}}$ et $\hat{Y}^{K_{\vec{\nu}}}$, ainsi que de l'approche mixte \hat{Y}	183
6.18	Détection du patch contenant un quatrième type de signature inusuelle à partir de la \hat{Y}^{NMF} , $\hat{Y}^{\mathcal{C}}$, $\hat{Y}^{K_{\mathcal{N}}}$, $\hat{Y}^{K_{\nu}}$ et $\hat{Y}^{K_{\vec{\nu}}}$, ainsi que de l'approche mixte \hat{Y}	184

Liste des tableaux

3.1	P-valeurs des tests statistiques issus de l'erreur de reconstruction basée sur \mathcal{D}_{Supp}^C du moteur endommagé et des moteurs normaux sur le patch contenant les signatures anormales qui ont permis la détection de l'endommagement et sur un patch ne présentant aucune signature atypique sur l'ensemble de la base de test.	79
3.2	Taux de détection à partir du modèle défini par les curvelets des différentes classes de points de la base de test Ω_{Test} pour différentes données d'apprentissage	89
3.3	Structure des sous-tableaux de résultats pour le modèle défini par le dictionnaire des curvelets	90
3.4	Taux de détection sur l'ensemble des patches de la base de test à partir du modèle de normalité défini à partir des curvelets	91
4.1	P-valeurs des tests statistiques issus de l'erreur de reconstruction du moteur endommagé et des moteurs normaux sur des patches avec et sans signatures anormales.	109
4.2	Taux de détection des différentes classes de points de la base de test Ω_{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	115
4.3	Structure des sous-tableaux de résultats pour le modèle défini par le dictionnaire de la NMF	116
4.4	Taux de détection sur l'ensemble des patches de la base de test à partir du modèle de normalité défini à partir de la NMF	117
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 Ω_{Test} pour différentes données d'apprentissage	133

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.	140
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	144
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	144
5.5	Structure des sous-tableaux de résultats pour le modèle défini par la distribution de chaque point estimé par noyau gaussien	145
5.6	Taux de détection sur l'ensemble des patches 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.	146
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	151
6.1	Taux de détection calculés à partir du modèle de normalité défini par la densité conditionnelle estimée à partir de différentes données d'apprentissage avec matrice d'échelle diagonale et pleine pour les différentes classes de points des données dans Ω_{Test}	168
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}	174
A.1	Performance des tests multiples en fonction des fausses et bonnes détections et non détections pour un seuil de décision t sur les p-valeurs.	210

Annexe A

Les tests multiples

Principe

Avec l'apogée des bases de données, les nombres de variables stockées ont fortement augmenté entraînant de nombreux tests statistiques à effectuer sur les mêmes données. Cependant, plus le nombre de tests augmente, plus la probabilité de commettre une erreur de première espèce (type I) (rejet de l'hypothèse nulle \mathcal{H}_0 à tort) augmente exponentiellement. On considère un test statistique de niveau α de la variable aléatoire X testant l'hypothèse \mathcal{H}_0 contre l'hypothèse \mathcal{H}_1 et la p-valeur $p(X)$ associée à ce test. Sous l'hypothèse \mathcal{H}_0 , la variable aléatoire $p(X)$ suit une loi uniforme sur $[0,1]$.

$$\mathbb{P}_{\mathcal{H}_0}(p(X) \leq t) = t, \quad \forall t$$

L'hypothèse \mathcal{H}_0 est rejetée à un niveau α lorsque $p(X) \leq \alpha$, ainsi la probabilité de commettre une erreur de type I, c'est-à-dire que la p-valeur soit inférieure à α alors qu'elle vérifie l'hypothèse \mathcal{H}_0 est donc :

$$\mathbb{P}_{\mathcal{H}_0}(p(X) \leq \alpha) = \alpha$$

Prenons maintenant en considération l'application de m tests indépendants de niveau α chacun, $\{\mathcal{H}_0^1, \dots, \mathcal{H}_0^m\}$ contre $\{\mathcal{H}_1^1, \dots, \mathcal{H}_1^m\}$, sur la même donnée avec leurs p-valeurs associés $\{p_1, \dots, p_m\}$. Nous définissons les ensembles $\mathbf{H}_0 = \{i \in \{1, \dots, m\} : \mathcal{H}_0^i \text{ est vraie}\}$ et $\mathbf{H}_1 = \{i \in \{1, \dots, m\} : \mathcal{H}_0^i \text{ est faux}\}$, la probabilité de commettre au moins une erreur de type I correspond alors à :

$$\begin{aligned} \mathbb{P}(\exists i \in \mathbf{H} : p_i(X) \leq \alpha) &= 1 - \mathbb{P}(\forall i \in \mathbf{H}, p_i(X) > \alpha) \\ &= 1 - \prod_{i=1}^m \mathbb{P}_{\mathcal{H}_0^i}(p_i(X) > \alpha) = 1 - (1 - \alpha)^m \xrightarrow{m \rightarrow +\infty} 1. \end{aligned}$$

Cela prouve bien que la probabilité de commettre une erreur de type I augmente exponentiellement avec le nombre de tests. Ainsi considérer chaque test unitairement entraîne une forte

probabilité de commettre des erreurs. Les grandeurs mesurant les risques de ces tests doivent considérer la multiplicité de ces derniers, il s'agit de la théorie des tests multiples [96, 101, 50].

Il existe 4 grandeurs permettant de mesurer les performances des tests multiples en fonction d'un seuil t de décision sur les p-valeurs des tests (Tableau A.1), il s'agit des vraies ou fausses détections et des vraies ou fausses non-détections. $V(t)$ correspond alors au nombre de fausses

TABLE A.1 – Performance des tests multiples en fonction des fausses et bonnes détections et non détections pour un seuil de décision t sur les p-valeurs.

Seuil de p-valeur t	\mathcal{H}_0 acceptée	\mathcal{H}_0 refusée	Total
\mathcal{H}_0 vraie	$U(t)$	$V(t)$	m_0
\mathcal{H}_0 fausse	$T(t)$	$S(t)$	m_1
Total	$W(t)$	$R(t)$	m

détections et $T(t)$ au nombre de non détections, $R(t)$ donne le nombre de détections. Parmi les m tests, m_0 tests vérifient l'hypothèse nulle et m_1 tests la réfutent, les valeurs de m_0 et m_1 ne sont pas connues a priori.

Les tests multiples prennent en compte les différentes p-valeurs des tests unitaires $\{p_1, \dots, p_m\}$ sur lesquelles un seuil tenant compte du nombre de tests est défini pour rejeter ou approuver les hypothèses. Les procédures de tests multiples définissent un ensemble d'indices de tests rejetés par la procédure (A.1), c'est-à-dire l'ensemble des p-valeurs $\{p_1, \dots, p_m\}$ inférieures à un seuil t_{proc} défini par la procédure. Les procédures de tests multiples sont donc entièrement caractérisées par ce seuil. Par abus de langage, nous notons R le nombre de rejets et l'ensemble des indices des p-valeurs des tests dont l'hypothèse \mathcal{H}_0 est rejetée.

$$R = \{i \in \{1, \dots, m\} : p_i(X) \leq t_{proc}\} \quad (\text{A.1})$$

La grandeur mesurée sur un test d'hypothèse unitaire permettant de contrôler le niveau (et donc le seuil sur la p-valeur) est l'erreur de première espèce. Pour les tests multiples, les grandeurs mesurées tiennent compte de cette multiplicité, nous distinguons deux grandeurs contrôlant le niveau des tests multiples et entraînant différentes valeurs de seuil sur les p-valeurs :

- Family-wise error rate (FWER) [59]
- False discovery rate (FDR) [14]

Family-wise error rate (FWER)

Les tests multiples sont caractérisés par les faux rejets de l'hypothèse \mathcal{H}_0 effectués à partir du seuil défini par la procédure. Ces faux rejets correspondent à $R \cap \mathbf{H}_0$, c'est-à-dire l'intersection entre les indices des p-valeurs rejetés R et les indices des tests dont l'hypothèse \mathcal{H}_0^i est vraie.

La grandeur définie par le FWER et permettant de contrôler les tests multiples correspond