# Bibliographie


[1] Abdel-Sayed, M., Duclos, D., Faÿ, G., Lacaille, J., and Mougeot, M. Anomaly detection on spectrograms using data-driven and fixed dictionary representations. In *Proceedings of the 24th European Symposium on Artificial Neural Networks* (2016), pp. 417–422.

[2] Abdel-Sayed, M., Duclos, D., Faÿ, G., Lacaille, J., and Mougeot, M. Dictionary comparison for anomaly detection on aircraft engine spectrograms. In *Machine Learning and Data Mining in Pattern Recognition*. Springer, 2016, pp. 362–376.

[3] Abdel-Sayed, M., Duclos, D., Faÿ, G., Lacaille, J., and Mougeot, M. Nmf-based decomposition for anomaly detection applied to vibration analysis. *International Journal of Condition Monitoring 6*, 3 (2016), 73–81.

[4] Aharon, M., Elad, M., and Bruckstein, A. K-svd : An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing 54*, 11 (2006), 4311–4322.

[5] Aharon, M., Elad, M., and Bruckstein, A. M. K-svd and its non-negative variant for dictionary design. In *Optics & Photonics 2005* (2005), International Society for Optics and Photonics, pp. 591411–591411.

[6] Antoni, J. Blind separation of vibration components : Principles and demonstrations. *Mechanical Systems and Signal Processing 19*, 6 (2005), 1166–1180.

[7] Antoni, J. The spectral kurtosis : a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing 20*, 2 (2006), 282–307.

[8] Antoni, J., Bonnardot, F., Raad, A., and El Badaoui, M. Cyclostationarity modeling of rotating machine vibration signals. *Mechanical systems and signal processing 18*, 6 (2004), 1285–1314.

[9] Antoni, J., and Randall, R. B. The spectral kurtosis : application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing 20*, 2 (2006), 308–331.

[10] Ballard, D. H. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition 13*, 2 (1981), 111–122.

[11] Bashtannyk, D. M., and Hyndman, R. J. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis 36*, 3 (2001), 279–298.

[12] Baydar, N., and Ball, A. A comparative study of acoustic and vibration signals in detection of gear failures using wigner–ville distribution. *Mechanical systems and signal processing 15*, 6 (2001), 1091–1107.

[13] Bengio, Y., Courville, A., and Vincent, P. Representation learning : A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence 35*, 8 (2013), 1798–1828.

[14] Benjamini, Y., and Hochberg, Y. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* (1995), 289–300.

[15] Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[16] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning 3*, 1 (2011), 1–122.

[17] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. Lof : identifying density-based local outliers. In *ACM sigmod record* (2000), vol. 29, ACM, pp. 93–104.

[18] Bruna, J., and Mallat, S. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence 35*, 8 (2013), 1872–1886.

[19] Candes, E. J. *Ridgelets : theory and applications*. PhD thesis, Stanford University, 1998.

[20] Candes, E. J., Demanet, L., Donoho, D. L., and Ying, L. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation 5*, 3 (2006), 861–899.

[21] Candès, E. J., and Donoho, D. L. Ridgelets : A key to higher-dimensional intermittency ? *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences 357*, 1760 (1999), 2495–2509.

[22] Candes, E. J., and Donoho, D. L. Curvelets : A surprisingly effective nonadaptive representation for objects with edges. Tech. rep., DTIC Document, 2000.

[23] Cariño-Corrales, J. A., Saucedo-Dorantes, J. J., Zurita-Millán, D., Delgado-Prieto, M., Ortega-Redondo, J. A., Alfredo Osornio-Rios, R., and de Jesus Romero-Troncoso, R. Vibration-based adaptive novelty detection method for monitoring faults in a kinematic chain. *Shock and Vibration 2016* (2016).

[24] Chacón, J. E., and Duong, T. Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test 19*, 2 (2010), 375–398.

[25] Chandola, V., Banerjee, A., and Kumar, V. Outlier detection : A survey. *ACM Computing Surveys* (2007).

[26] Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection : A survey. *ACM computing surveys (CSUR) 41*, 3 (2009), 15.

[27] Chauris, H., and Nguyen, T. Seismic demigration/migration in the curvelet domain. *Geophysics 73*, 2 (2008), S35–S46.

[28] Chen, S. S., Donoho, D. L., and Saunders, M. A. Atomic decomposition by basis pursuit. *SIAM review 43*, 1 (2001), 129–159.

[29] Chen, S. X. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics 52*, 3 (2000), 471–480.

[30] Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S.-i. *Nonnegative matrix and tensor factorizations : applications to exploratory multi-way data analysis and blind source separation.* John Wiley & Sons, 2009.

[31] Clifton, D. A. *Novelty detection with extreme value theory in jet engine vibration data.* PhD thesis, University of Oxford, 2009.

[32] Clifton, D. A., Hugueny, S., and Tarassenko, L. Novelty detection with multivariate extreme value statistics. *Journal of signal processing systems 65*, 3 (2011), 371–389.

[33] Clifton, D. A., and Tarassenko, L. Novelty detection in jet engine vibration spectra. *International Journal of Condition Monitoring 5*, 2 (2015), 2–7.

[34] Clifton, D. A., Tarassenko, L., McGrogan, N., King, D., King, S., and Anuzis, P. Bayesian extreme value statistics for novelty detection in gas-turbine engines. In *Aerospace Conference, 2008 IEEE* (2008), IEEE, pp. 1–11.

[35] Cunningham, J., and Ghahramani, Z. Linear dimensionality reduction : Survey, insights and generalizations. *Journal of Machine Learning Research 16* (2015), 2859–2900.

[36] Dalal, N., and Triggs, B. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (2005), vol. 1, IEEE, pp. 886–893.

[37] De Haan, L., and Ferreira, A. *Extreme value theory : an introduction.* Springer Science & Business Media, 2007.

[38] Désir, C., Bernard, S., Petitjean, C., and Heutte, L. One class random forests. *Pattern Recognition 46*, 12 (2013), 3490–3506.

[39] Do, M. N., and Vetterli, M. The contourlet transform : an efficient directional multiresolution image representation. *IEEE Transactions on image processing 14*, 12 (2005), 2091–2106.

[40] Donoho, D. L. Orthonormal ridgelets and linear singularities. *SIAM Journal on Mathematical Analysis 31*, 5 (2000), 1062–1099.

[41] Donoho, D. L., and Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences 100*, 5 (2003), 2197–2202.

[42] Donoho, D. L., et al. Wedgelets : Nearly minimax estimation of edges. *The Annals of Statistics 27*, 3 (1999), 859–897.

[43] Donoho, D. L., Tsaig, Y., Drori, I., and Starck, J.-L. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory 58*, 2 (2012), 1094–1121.

[44] Duong, T., and Hazelton, M. L. Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics 32*, 3 (2005), 485–506.

[45] El Badaoui, M., Antoni, J., Guillet, F., Daniere, J., and Velex, P. Use of the moving cepstrum integral to detect and localise tooth spalls in gears. *Mechanical Systems and Signal Processing 15*, 5 (2001), 873–885.

[46] Engan, K., Aase, S. O., and Husoy, J. H. Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on* (1999), vol. 5, IEEE, pp. 2443–2446.

[47] Fadili, J., and Starck, J.-L. Curvelets and ridgelets. In *Encyclopedia of Complexity and Systems Science*. Springer, 2009, pp. 1718–1738.

[48] Fisher, R. A. *The genetical theory of natural selection : a complete variorum edition.* Oxford University Press, 1930.

[49] Friedman, J., Hastie, T., and Tibshirani, R. *The elements of statistical learning*, vol. 1. Springer series in statistics Springer, Berlin, 2001.

[50] Friguet, C. *Impact de la dépendance dans les procédures de tests multiples en grande dimension.* PhD thesis, Agrocampus-Ecole nationale supérieure d'agronomie de rennes, 2010.

[51] Gorodnitsky, I. F., and Rao, B. D. Sparse signal reconstruction from limited data using focuss : A re-weighted minimum norm algorithm. *IEEE Transactions on signal processing 45*, 3 (1997), 600–616.

[52] Griffaton, J., Picheral, J., and Tenenhaus, A. Enhanced visual analysis of aircraft engines based on spectrograms. *ISMA2014* (2014), 2809–2822.

[53] Hawkins, S., He, H., Williams, G., and Baxter, R. Outlier detection using replicator neural networks. In *International Conference on Data Warehousing and Knowledge Discovery* (2002), Springer, pp. 170–180.

[54] Hayton, P., Utete, S., King, D., King, S., Anuzis, P., and Tarassenko, L. Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences 365*, 1851 (2007), 493–514.

[55] Hazan, A., Lacaille, J., and Madani, K. Extreme value statistics for vibration spectra outlier detection. In *International Conference on Condition Monitoring and Machinery Failure Prevention Technologies* (2012), pp. p–1.

[56] He, Z., Deng, S., and Xu, X. An optimization model for outlier detection in categorical data. *Advances in intelligent computing* (2005), 400–409.

[57] Heidenreich, N.-B., Schindler, A., and Sperlich, S. Bandwidth selection for kernel density estimation : a review of fully automatic selectors. *AStA Advances in Statistical Analysis 97*, 4 (2013), 403–433.

[58] Henderson, D. J., and Parmeter, C. F. Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. *Statistics & Probability Letters 82*, 12 (2012), 2198–2205.

[59] Hochberg, Y. A sharper bonferroni procedure for multiple tests of significance. *Biometrika 75*, 4 (1988), 800–802.

[60] Hodge, V., and Austin, J. A survey of outlier detection methodologies. *Artificial intelligence review 22*, 2 (2004), 85–126.

[61] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* (1998), vol. 454, The Royal Society, pp. 903–995.

[62] Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics 5*, 4 (1996), 315–336.

[63] Klein, R. A method for anomaly detection for nonstationary vibration signatures. In *Annual Conference of the Prognostics and Health Management Society* (2013).

[64] Klein, R., Masad, E., Rudyk, E., and Winkler, I. Bearing diagnostics using image processing methods. *Mechanical Systems and Signal Processing 45*, 1 (2014), 105–113.

[65] Klein, R., Rudyk, E., Masad, E., and Issacharoff, M. Emphasising bearing tones for prognostics. *International Journal of Condition Monitoring 1*, 2 (2011), 73–78.

[66] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[67] Le Pennec, E., and Mallat, S. Sparse geometric image representations with bandelets. *IEEE transactions on image processing 14*, 4 (2005), 423–438.

[68] LeCun, Y., Kavukcuoglu, K., and Farabet, C. Convolutional networks and applications in vision. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on* (2010), IEEE, pp. 253–256.

[69] Lee, D. D., and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature 401*, 6755 (1999), 788–791.

[70] Lee, D. D., and Seung, H. S. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems* (2001), pp. 556–562.

[71] Lehmann, E., Romano, J. P., et al. Generalizations of the familywise error rate. *The Annals of Statistics 33*, 3 (2005), 1138–1154.

[72] Lehmann, E. L., and Romano, J. P. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[73] Lei, Y., Lin, J., He, Z., and Zuo, M. J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing 35*, 1 (2013), 108–126.

[74] Leiva-Murillo, J. M., and Artés-Rodríguez, A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. *Pattern Recognition Letters 33*, 13 (2012), 1717–1724.

[75] Lewicki, M. S., and Sejnowski, T. J. Learning overcomplete representations. *Neural computation 12*, 2 (2000), 337–365.

[76] Lowe, D. G. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (1999), vol. 2, IEEE, pp. 1150–1157.

[77] Ma, J., and Plonka, G. A review of curvelets and recent applications. *IEEE Signal Processing Magazine 27*, 2 (2010), 118–133.

[78] Mailhé, B., Lesage, S., Gribonval, R., Bimbot, F., and Vandergheynst, P. Shift-invariant dictionary learning for sparse representations : extending k-svd. In *Signal Processing Conference, 2008 16th European* (2008), IEEE, pp. 1–5.

[79] Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., and Bach, F. R. Supervised dictionary learning. In *Advances in neural information processing systems* (2009), pp. 1033–1040.

[80] Mallat, S. *A wavelet tour of signal processing.* Academic press, 1999.

[81] Mallat, S., and Peyré, G. A review of bandlet methods for geometrical image representation. *Numerical Algorithms 44*, 3 (2007), 205–234.

[82] Mallat, S. G., and Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing 41*, 12 (1993), 3397–3415.

[83] Markou, M., and Singh, S. Novelty detection : a review-part 1 : statistical approaches. *Signal processing 83*, 12 (2003), 2481–2497.

[84] Markou, M., and Singh, S. Novelty detection : a review-part 2 : neural network based approaches. *Signal processing 83*, 12 (2003), 2499–2521.

[85] Nachar, N. The mann-whitney u : A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology 4*, 1 (2008), 13–20.

[86] Ojala, T., Pietikäinen, M., and Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition 29*, 1 (1996), 51–59.

[87] Olshausen, B. A., and Field, D. J. Sparse coding with an overcomplete basis set : A strategy employed by v1 ? *Vision research 37*, 23 (1997), 3311–3325.

[88] Parikh, N., Boyd, S., et al. Proximal algorithms. *Foundations and Trends® in Optimization 1*, 3 (2014), 127–239.

[89] Parzen, E. On estimation of a probability density function and mode. *The annals of mathematical statistics 33*, 3 (1962), 1065–1076.

[90] Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. A review of novelty detection. *Signal Processing 99* (2014), 215–249.

[91] Randall, R. B. *Vibration-based condition monitoring : industrial, aerospace and automative application.* John Wiley & Sons, 2011.

[92] Randall, R. B., Antoni, J., and Chobsaard, S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mechanical systems and signal processing 15*, 5 (2001), 945–962.

[93] Rapin, J., Bobin, J., Larue, A., and Starck, J.-L. Nmf with sparse regularizations in transformed domains. *SIAM Journal on Imaging Sciences 7*, 4 (2014), 2020–2047.

[94] Romano, J. P., Shaikh, A. M., et al. On stepdown control of the false discovery proportion. In *Optimality*. Institute of Mathematical Statistics, 2006, pp. 33–50.

[95] Romano, J. P., and Wolf, M. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*, 469 (2005), 94–108.

[96] Roquain, E. *Contributions to multiple testing theory for high-dimensional data.* PhD thesis, Université Pierre et Marie Curie, 2015.

[97] Rubinstein, R., Bruckstein, A. M., and Elad, M. Dictionaries for sparse representation modeling. *Proceedings of the IEEE 98*, 6 (2010), 1045–1057.

[98] Rubinstein, R., Faktor, T., and Elad, M. K-svd dictionary-learning for the analysis sparse model. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (2012), IEEE, pp. 5405–5408.

[99] Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., Platt, J. C., et al. Support vector method for novelty detection. In *NIPS* (1999), vol. 12, pp. 582–588.

[100] Scott, D. W., and Terrell, G. R. Biased and unbiased cross-validation in density estimation. *Journal of the american Statistical association 82*, 400 (1987), 1131–1146.

[101] Shaffer, J. P. Multiple hypothesis testing. *Annual review of psychology 46*, 1 (1995), 561–584.

[102] Sheather, S. J., et al. Density estimation. *Statistical Science 19*, 4 (2004), 588–597.

[103] Sheather, S. J., and Jones, M. C. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* (1991), 683–690.

[104] Silverman, B. W. *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.

[105] Starck, J.-L., Candès, E. J., and Donoho, D. L. The curvelet transform for image denoising. *IEEE Transactions on image processing 11*, 6 (2002), 670–684.

[106] Storey, J. D. The positive false discovery rate : a bayesian interpretation and the q-value. *Annals of statistics* (2003), 2013–2035.

[107] Tax, D. M., and Duin, R. P. Support vector data description. *Machine learning 54*, 1 (2004), 45–66.

[108] Thevenin, J. le turboréacteur, moteur des avions à réaction. *AAAF (Association Aéronautique et Astronautique de France) 3* (2004).

[109] Thompson, B. B., Marks, R. J., Choi, J. J., El-Sharkawi, M. A., Huang, M.-Y., and Bunje, C. Implicit learning in autoencoder novelty assessment. In *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on* (2002), vol. 3, IEEE, pp. 2878–2883.

[110] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

[111] Timusk, M., Lipsett, M., and Mechefske, C. K. Fault detection using transient machine signals. *Mechanical Systems and Signal Processing 22*, 7 (2008), 1724–1749.

[112] Wand, M. P., and Jones, M. C. Multivariate plug-in bandwidth selection. *Computational Statistics 9*, 2 (1994), 97–116.

[113] Wang, W., and McFadden, P. Early detection of gear failure by vibration analysis–i. interpretation of the time-frequency distribution using image processing techniques. *Mechanical Systems and Signal Processing 7*, 3 (1993), 193–203.

[114] Wang, W., and McFadden, P. Early detection of gear failure by vibration analysis–ii. interpretation of the time-frequency distribution using image processing techniques. *Mechanical Systems and Signal Processing 7*, 3 (1993), 205–215.

[115] Wang, W., and Wong, A. K. Autoregressive model-based gear fault diagnosis. *Transactions-American Society of Mechanical Engineers Journal of Vibration and Acoustics 124*, 2 (2002), 172–179.

[116] Yan, R., Gao, R. X., and Chen, X. Wavelets for fault diagnosis of rotary machines : A review with applications. *Signal Processing 96* (2014), 1–15.

[117] Yang, H., Mathew, J., and Ma, L. Vibration feature extraction techniques for fault diagnosis of rotating machinery : a literature survey. In *Proceedings of the 10th Asia-Pacific Vibration Conference* (2003), pp. 801–807.

[118] Zhan, Y., and Makis, V. A robust diagnostic model for gearboxes subject to vibration monitoring. *Journal of Sound and vibration 290*, 3 (2006), 928–955.

[119] Zhang, X., King, M. L., and Hyndman, R. J. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis 50*, 11 (2006), 3009–3031.

# Table des figures

# Liste des tableaux

# Annexe A

# Les tests multiples

## Principe

Avec l'apogée des bases de données, les nombres de variables stockées ont fortement augmenté entrainant 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 $\alpha$ 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].

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

L'hypothèse $\mathcal{H}_0$ est rejetée à un niveau $\alpha$ 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 $\alpha$ alors qu'elle vérifie l'hypothèse $\mathcal{H}_0$ est donc :

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

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

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

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 entraine 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 $\{p_1, ..., p_m\}$ 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 $\{p_1, ...p_m\}$ 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.

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

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 entrainant 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 à la probabilité que le cardinal de cette intersection (le nombre de faux rejets) soit supérieur ou égal à 1 (A.2).

$$FWER(R) = \mathbb{P}(\text{card}\,(R \cap \mathbf{H}_0) \geq 1) = \mathbb{P}(V(R) \geq 1) = \mathbb{P}\left(\inf_{i \in \mathbf{H}_0} \{p_i(X)\} \leq t_{proc}\right) \tag{A.2}$$

Le seuil de décision $t$ est équivalent au rejet effectué à partir de ce seuil, donc $V(R) = V(t)$.

Ainsi en contrôlant cette probabilité à un niveau $\alpha$, nous nous assurons de commettre 0 fausse détection avec probabilité $1 - \alpha$. La procédure de Bonferroni [59] consiste à définir le seuil $t$ de détection par $\alpha/m$ avec $\alpha$ le niveau souhaité.

$$t_{Bonferroni} = \frac{\alpha}{m}$$

La procédure de Bonferroni permet de contrôler le FWER à un niveau $\alpha$.

$$R_{Bonferroni} = \{i \in \{1, ..., m\} : p_i(X) \leq t_{Bonferroni}\}$$

$$\begin{aligned} FWER(R_{Bonferroni}) &= \mathbb{P}\left(\bigcup_{i \in \mathbf{H}_0} \{p_i(X) \leq t_{Bonferroni}\}\right) \\ &\leq \sum_{i=1}^{m} \mathbb{1}\{i \in \mathbf{H}_0\} \mathbb{P}\left(p_i(X) \leq t_{Bonferroni}\right) = \frac{m_0}{m}\alpha \leq \alpha \end{aligned}$$

Plus le nombre de tests est important, plus le seuil défini sur les p-valeurs sera très faible. Cette procédure est très conservatrice, elle a donc fortement tendance à éviter de rejeter l'hypothèse $\mathcal{H}_0$.

Pour relaxer cette forte contrainte il est possible de définir le k-FWER [71] permettant de contrôler les tests multiples en autorisant avec probabilité $1-\alpha$, un nombre de fausses détections d'au plus $k - 1$.

$$k - FWER = \mathbb{P}(\text{card}\,(R \cap \mathbf{H}_0) \geq k) = \mathbb{P}(V(R) \geq k)$$

Dans un cadre indépendant, la majoration de la probabilité de l'union par la somme des probabilités est pertinente, cependant si les tests sont fortement dépendants cette majoration est très forte entrainant un niveau réel nettement inférieur à $\alpha$, dans ce cas il est possible de contrôler le FWER comme un quantile de niveau $\alpha$ de la variable aléatoire $\inf_{i \in \mathbf{H}_0}\{p_i(X)\}$ [95]. Une procédure simple permettant de contrôler cette grandeur est de définir le seuil sur une base de données de validation comme étant la plus petite p-valeur dont l'hypothèse $\mathcal{H}_0^i$ est vraie. Cependant dans notre cas d'étude, cette approche n'est pas pertinente due aux différentes classes des points sous l'hypothèse $\mathcal{H}_0$ de normalité. Les p-valeurs des points normaux de la classe 4 (c'est-à-dire les points décalés dans les différents spectrogrammes) ont le même ordre de grandeur que les signatures inusuelles. Le seuil défini serait alors trop faible et entrainerait le non-rejet de l'hypothèse $\mathcal{H}_0$ pour de nombreux points inusuels. Afin de définir un seuil de cette manière, il est plus pertinent de le fixer uniquement sur les points des classes 1 (normaux) et 3 (bruit), augmentant ainsi la détection au niveau de la classe 4.

La méthode de Romano-Wolf [95] permet de contrôler le FWER à un niveau $\alpha$ dans un cadre dépendant en appliquant la procédure de Bonferroni de manière itérative sur des sous-ensembles de tests . A la première itération, la procédure de Bonferroni est appliquée sur l'ensemble des tests, le sous-ensemble récupéré $R_(1)^c$ (complémentaire des points rejetés) correspond à l'ensemble des tests qui n'ont pas rejetés l'hypothèse $\mathcal{H}_0$, sur lequel on applique une nouvelle fois la procédure de Bonferroni en considérant uniquement ces tests. La procédure continue rejetant de nouveaux tests $R_k$ jusqu'à l'itération $\hat{k}$ où les tests rejetant l'hypothèse $\mathcal{H}_0$ soient les mêmes sur deux itérations successives.

$$R_k = \left\{ i \in \{1, ..., m\} : p_i(X) \leq t_{Bonferroni}^{R_{k-1}^c} \right\}$$

$$\hat{k} = \min \{k \geq 1 : R_{k+1} = R_k\}$$

avec $t_{Bonferroni}^{R_{k-1}^c}$ le seuil de la procédure de Bonferroni défini à partir des tests dans $R_{k-1}^c$.

Il existe également des procédures permettant de contrôler le FWER lorsque les dépendances entre les différents tests sont connues [50].

## False discovery rate (FDR)

Le FWER contrôle le nombre de fausses détections effectuées par les m tests, la grandeur ne tient pas compte du nombre entier d'hypothèses $\mathcal{H}_0^i$ rejetées par la procédure. $V(R) = 1$ hypothèse rejetée à tort parmi $R = 10$ hypothèses rejetées est plus grave que parmi $R = 100$ hypothèses rejetées. Pour le FWER, il n'y a pas de différence entre ces deux résultats. Le False discovery proportion (FDP) prend en compte cette différence en considérant comme grandeur pour contrôler les tests multiples le ratio entre le nombre de rejets d'hypothèses $\mathcal{H}_0^i$ à tort $V(R)$ et le nombre total $R$ de rejets d'hypothèse $\mathcal{H}_0^i$.

$$FDP(R) = \frac{\text{card}\,(R \cap \mathbf{H}_0)}{\max\,(R, 1)} = \frac{V(R)}{\max\,(R, 1)}$$

Le FDP est une variable aléatoire et ne permet donc pas de définir un niveau pour contrôler les tests multiples, le FDR (A.3) est la grandeur issue du FDP permettant de contrôler les tests multiples à un niveau $\alpha$. Elle correspond à une tolérance moyenne de fausses détections parmi l'ensemble des détections.

$$FDR(R) = \mathbb{E}(FDP(R)) = \begin{cases} \mathbb{E}\left(\frac{V(R)}{R}\right) & \text{si } R > 0 \\ 0 & \text{si } R = 0 \end{cases} \tag{A.3}$$

La procédure de BH [14] définie par le seuil (A.4) permet de contrôler le FDR à un niveau $\alpha$

dans le cas où les tests sont indépendants ou positivement dépendants.

$$t_{BH} = \frac{\alpha \hat{l}}{m} \text{ avec } \hat{l} = \max\left\{ l \in \{0, 1, ..., m\} : p_{(l)} \leq \frac{\alpha l}{m} \right\} \tag{A.4}$$

avec $p_{(l)}$ les p-valeurs $(p_1, ..., p_m)$ ordonnées par ordre croissant.

$$R_{BH} = \{i \in \{1, ..., m\} : p_i(X) \leq t_{BH}\} \Leftrightarrow \{p_{(1)}, ..., p_{(\hat{l})}\}$$

$$FDR(R_{BH}) \leq \frac{m_0}{m}\alpha \leq \alpha$$

Cette procédure consiste à définir comme seuil le dernier point d'intersection entre la courbe des p-valeurs ordonnées et la droite d'équation $l \mapsto \alpha l/m$. Cette procédure est invariante par échelle lorsque le nombre de tests m croit. La procédure BH peut être considérée comme une procédure "step-up" car le seuil est défini comme le point maximal d'intersection, il est également possible de prendre le point d'intersection minimal (le premier point où les 2 courbes se croisent) ou un point intermédiaire. Il existe d'autres procédures permettant de contrôler le FDR [96].

Dans le cadre de nos données de spectrogrammes, les différents points sur les patchs peuvent être considérés comme indépendants ou comme étant positivement dépendants. Une intensité vibratoire à régime et fréquence donnés n'influent pas négativement avec les autres points du voisinage, mais peut impacter positivement les valeurs du voisinage, des points éloignés dans l'espace régime/fréquence peuvent être considérés comme indépendants.

Le fait de contrôler le FDR à un niveau $\alpha$ ne signifie pas que le FDP est lui aussi contrôler à ce même niveau, cela est uniquement le cas asymptotiquement lorsque les dépendances entre les tests sont faibles. Il existe donc des procédures visant à contrôler le FDP [94] via les grandeurs suivantes :

$$\mathbb{P}(FDP(R) > \alpha) \leq \zeta,$$

$$\limsup_{m} \mathbb{P}(FDP(R) > \alpha) \leq \zeta.$$

Une approche différente proposée dans [106] considère une estimation du FDR à partir de la région de rejet, le FDR n'est plus défini à partir du seuil de décision à partir duquel la région de rejet est calculé, la région de rejet est prédéfinie et le niveau du FDR est calculé en conséquence. Pour cela différentes régions de rejet sont testées sur les p-valeurs ordonnées, et une estimation du FDR est apportée pour chacune d'entre elles, il s'agit alors de sélectionner la région apportant une estimation convenable du FDR. Cette approche permet d'estimer une variation du FDR, le positive false discovery rate (pFDR).