A public benchmark for denoising and detection methods

Juan M. Miramont[1], Rémi Bardenet[2], Pierre Chainais[2], François Auger[1]

[1]Nantes Université, Institut de Recherche en Énergie Électrique de Nantes Atlantique (IREENA, UR 4642),
F-44600 Saint-Nazaire, France

[2]Univ. Lille, CNRS, Centrale Lille, UMR9189–CRIStAL
F-59651 Villeneuve d'Ascq, France

`juan.miramont@univ-nantes.fr`, `remi.bardenet@gmail.com`,
`pierre.chainais@centralelille.fr`, `francois.auger@univ-nantes.fr`

Résumé – Dans cet article, un benchmark en libre accès pour les méthodes de débruitage et de détection de signaux est présenté. L'objectif de cet outil est d'établir un cadre commun pour comparer les méthodes, sur la base d'un grand nombre de signaux et d'outils d'interprétation des résultats. Les auteurs de nouvelles méthodes pourraient bénéficier d'un moyen de comparer leur approche avec les précédentes, ainsi que de découvrir les avantages et les inconvénients de leurs propres méthodes. Deux exemples d'utilisation sont fournis, comparant les méthodes basées sur les zéros du spectrogramme aux méthodes basées sur les maxima locaux du spectrogramme, et estimant la puissance statistique des tests de détection.

Abstract – In this article, a public benchmark for signal denoising and signal detection methods is introduced. The aim of this tool is to establish a common framework to compare methods, on the basis of a large number of signals and tools for interpreting the results. Authors of new methods could benefit from having a way to compare their approach with previous ones, as well as discover advantages and disadvantages of their own methods. Two examples of use are provided, comparing methods based on zeros of the spectrogram against methods based on the local maxima of the spectrogram, and estimating the statistical power of detection tests.

1 Introduction

In this article, a framework for benchmarking methods of signal detection and denoising is introduced. Its goal is to provide users with a tool for easily making meaningful comparisons between existing methods and their own approaches, finding out their strengths or room for improvement. This is done by applying the methods to the same set of signals and using a common tool for results interpretation. This benchmark is based on *Python*, and can be accessed through a public repository [1], where users can upload their own method for automatic testing. Interested users can download the test bench to locally perform more personalized tests.

In Sec. 2, signal detection and denoising are defined, two key tasks in signal processing [2,5]. The benchmark is outlined in Sec. 3. Sec. 4 illustrates applications of the benchmark for power estimation of detection tests and comparing methods for denoising. Finally, some conclusions are drawn in Sec. 6.

2 Signal detection and denoising

Let $x(t) = s(t) + \xi(t)$ be a continuous real- or complex-valued signal of $t \in \mathbb{R}$, where s is the noiseless signal, and ξ is the added noise.

Detection. This task aims at discovering the presence of a signal $s(t)$ immersed in noise [3,4]. What is *signal* and what is *noise* depends on the application, but usually one can consider the signal to be a quantity of interest that exhibits some *organization*. In contrast, noise is usually regarded as a disorganized or *random* fluctuation that includes all the other influences on $x(t)$ that are not of interest [2].

The detection task can be formalized as a hypothesis test [2, 8], trying to reject the hypothesis $H_0 : x = \xi$ of observing pure noise. The significance level of the test is one minus the probability, under H_0, to reject H_0.

To assess the power of the test, one might further consider an alternative model

$$\left\{ \begin{array}{l} H_0 : x = \xi \\ H_1 : x = s + \xi \end{array} \right. , \qquad (1)$$

for s some fixed signal. The power is then the probability, under H_1, to (correctly) reject H_0.

Denoising. The aim of a denoising method will be, for a given $x(t)$, to approximate $s(t)$ by some $\tilde{s}(t)$ *as well as possible*, according to some performance metric. A common such metric is the *quality reconstruction factor* (QRF), defined as [6] :

$$\text{QRF} := 10 \log_{10} \left(\frac{\|s\|_2^2}{\|s - \tilde{s}\|_2^2} \right) \text{ dB}, \qquad (2)$$

which in turn can be seen as the SNR of the denoised signal, obtained by estimating the remanent noise as the difference bet-
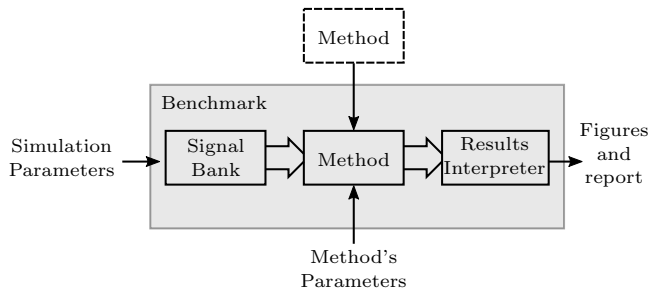
FIGURE 1 – Block diagram of the proposed benchmark. Inward arrows represent the user inputs.

ween $\tilde{s}(t)$ and the true signal $s(t)$. Other metrics include the mean squared error or its square root. A plethora of methods for disentangling signal and noise are available [2, 8, 13]. More sophisticated methods, such as the ones described in Sec. 4, behave as *adaptive* filters, whose frequency response is modified considering the frequency contents of the signal [2, 6].

3   A framework for comparing methods

The benchmark proposed in this work is summarized in Fig. 1, where it can be seen that the inputs given by the user are a series of simulation parameters, one or several methods to be evaluated and, eventually, different sets of parameters for the method. The simulation parameters include : 1) The task (either denoising or detection), 2) The number of temporal samples of the signals, 3) The Signal-to-Noise ratio (in dB) : a tuple of values that determines the levels of noise used to contaminate the signals, 4) The number of repetitions, i.e. how many noise realizations are used for each signal.

Once this is set, a number of noisy signals are generated in the "Signal Bank" block, according to the given parameters, and fed to the methods. The output is then analyzed by the "Results Interpreter" block, a toolbox that computes the performance metrics and generates the output figures and files.

A bank of signals.   Noisy sythetic signals provide the means for a quantitative evaluation of a technique's performance, since both the noiseless version of the signal and the noise are known. Moreover, synthetic signals can be designed to pose specific challenges to the methods, which is relevant when the techniques to be benchmarked are based on a *model* of the signal or the noise. Therefore, one possible usage of the benchmark is assessing a method's performance when the signal under study deviates from the model. With this in mind, more than 20 synthetic signals with different time-frequency structures are offered to the user.

Interpreting the results.   Analyzing the amount of data generated during the benchmarking process can be a cumbersome process. For this reason, a toolbox capable of generating a series of tables, figures and files for further post-processing by the user is provided. As an example, the QRF plots displayed later in Fig. 4 were produced this way.

4   Denoising and detection with spectrograms

The spectrogram.   The spectrogram $S_x(t, \nu)$ of a signal $x(t)$ is the squared modulus of its short-time Fourier Transform

$$V_x(t, \nu) = \int_{\mathbb{R}} x(u)g(u - t)^* e^{-i2\pi\nu u} du, \qquad (3)$$

where $g^*(t)$ is the complex conjugate of the the analysis window $g(t)$. We take a Gaussian window $g(t) = 2^{1/4} e^{-\pi t^2}$ throughout this work.

Detection based on zeros of the spectrogram.   The zeros of the spectrogram of white noise seem to be homogeneously distributed in the time-frequency (TF) plane. This distribution has been shown to correspond to that of the zeros of the planar random Gaussian analytic function [8]. When a signal is present, such distribution is locally perturbed by the presence of a ridge, which "pushes away" the zeros, creating larger-than-expected holes in the regular pattern of zeros corresponding to pure noise. A *Monte Carlo envelope test* [8] can be then devised as a detection test based on holes in the pattern of zeros, as explained in the following. Considering the zeros of the spectrogram as a point process $\mathcal{X}$ in $\mathbb{C}$, one can characterize it by means of a functional statistic such as the *empty space function* $F : r \mapsto \mathbb{P}(\mathcal{X} \cap B(0, r) \neq \emptyset)$. There are several estimators for F in the spatial statistics literature. Here we pick the "border correction" estimate [16]. Fix a level of significance $1 - \alpha$. Fix $k \leq m$ two integers such that $\alpha = k/(m + 1)$. Generate m realizations of noise, take their spectrogram, extract their sets of zeros, and estimate the corresponding empty space function to obtain $\hat{F}_1, \ldots, \hat{F}_m$. Similarly, denote the estimated empty space function on the data to be analyzed by $\hat{F}_{\text{data}}$. Form the average $\hat{F}_0$ of $\hat{F}_1, \ldots, \hat{F}_m$ and $\hat{F}_{\text{data}}$. Now fix $r_{\max} > 0$ and compute

$$\left\| (\hat{F} - \hat{F}_0) \mathbb{1}_{[0, r_{\max}]} \right\|_p \qquad (4)$$

for $\hat{F} = \hat{F}_1, \ldots, \hat{F}_m$ or $\hat{F}_{\text{data}}$, where $\|\cdot\|_p$ is either the 2-norm or the supremum norm, and call the output $t_1, \ldots, t_m, t_{\text{data}}$. Rank $t_1, \ldots, t_m$ as $t_{(1)} \geq \cdots \geq t_{(m)}$. If $t_{\text{data}} \geq t_{(k)}$, then reject H_0. Intuitively, a large value of t_{data} among the values obtained by sampling under H_0 is suspicious. A symmetry argument [15] guarantees that this detection test has significance level $1 - \alpha$.

Denoising using maxima and zeros.   The spectrogram can be thought of as a TF distribution of the energy of the signal [2]. Based on this, a classic paradigm for TF filtering consists in looking for maxima or high values in $S_x(t, \nu)$, assuming that more energy is concentrated where there is more influence of the signal [7], and then identifying a region of the plane around these maxima that we shall denote here as the *signal domain* $\mathcal{D}_s$. After this, one can obtain an estimation $\tilde{s}(t)$ by Fourier inversion, that is,

$$\tilde{s}(t) = \iint_{\mathbb{R}^2} V_x(u, \nu) \mathbb{1}_{\mathcal{D}_s}(u, \nu) g(u - t) e^{i2\pi\nu t} du \, d\nu \qquad (5)$$

where $\mathbb{1}_{\mathcal{D}_s}$ is a (typically smoothed) indicator function of $\mathcal{D}_s$, acting as a 1/0 mask.

In contrast to the paradigm of informative large values of the spectrogram, it was also shown that its zeros or *minimum values* were also relevant points that could be used for noise filtering [13], mainly because adding a signal to noise does not change the number of zeros, but forces their reorganization so that they tend to surround the signal domain [2, 13].

The proposed benchmark can be used to compare denoising methods that are based in exploiting the structure of the spectrogram. Hereafter we shall describe four methods that estimate $\mathcal{D}_s$, which could be broadly separated between those based on large values and those based on the zeros of the spectrogram.

Hard-thresholding. This technique consists in removing the values of the STFT, the modulus of which is below a certain threshold [10, 11]. Considering $\xi(t)$ as a realization of complex analytic white Gaussian noise with variance σ^2, the variance of the real and imaginary parts of $V_\xi(t, \nu)$ are given by $\sigma^2 \|g\|_2^2$ where $\|g\|_2$ is the 2-norm of the analysis window. Then one can establish a threshold given by $3\sigma\|g\|_2$, where $\sigma\|g\|_2$ is empirically estimated as [11]:

$$\sigma\|g\|_2 \approx \frac{\operatorname{median}\left\{|\operatorname{Re}\{V_x(t, \nu)\}|\right\}}{0.6745}. \tag{6}$$

Contour-based basins of attractions. This method [14] is based on the fact that the ridges generated by the modes of the signal are attractors of the reassignment vector [7, 9]. One can then find the ridges by looking for the points where the reassignment vector changes its direction [14]. This is also useful to identify the signal domain, which is constituted by all the so-called *basins of attractions* [9]. These latter are formed by the points in the TF plane that would be relocated to a certain ridge following the reassignment method. One thus hopes to extract one basin per ridge.

Delaunay triangulation of zeros. With the purpose of identifying $\mathcal{D}_s$, it was proposed in [13] to compute a Delaunay triangulation of the zeros of the spectrogram. Because of the displacement of zeros caused by the presence of a signal, the triangles along the signal domain are more stretched than those generated only by noise. This makes possible to identify those triangles, by computing the length of the edges and selecting only those of them whose edges' length surpass a certain value $e_{\max}$, in this work 1.85.

Empty spaces. Similar in principle to the previous method, one forms $\mathcal{D}_s$ here by aggregating Euclidean balls of radius $r_0/2$ that do not contain any zero [8]. We use here a manually tuned choice of $r_0 = 0.9$, but to determine r_0, one might actually use known theoretical properties of the zeros of the spectrogram of white noise [8], or even extract r_0 from a preliminary detection test, like the one described above for signal detection. One could then identify the radius $r_0/2$ by finding, after H_0 has been rejected, the smallest $r_{\max}$ in (4) that would have led to rejection.

5 Results of simulations

Power determination for detection tests. Determining power requires specifying a signal model. We consider here a linear chirp with 256 time samples. The test was then run for 200
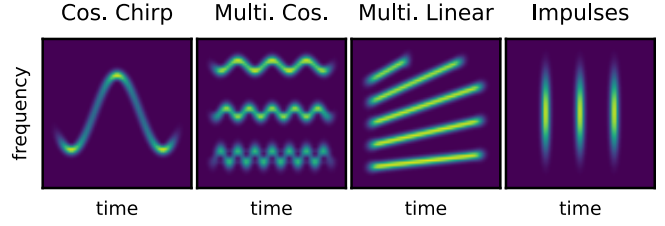

FIGURE 2 – Spectrograms of the signals used in the simulations. First Panel : A single cosine chirp. Second Panel : Multiple cosine chirps. Third Panel : Multiple linear chirps. Fourth Panel : Three impulsive transients.
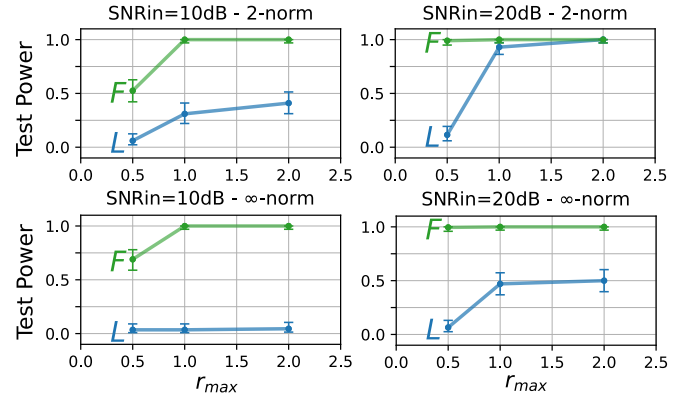

FIGURE 3 – Power estimation for different SNRs, norms and for two functional statistics : the empty space function (F), and the variance-stabilized Ripley function (L) [16]. 95% Clopper-Pearson confidence interval are shown for each value of $r_{\max}$.

independent realizations of white Gaussian noise, using the 2-norm or the supremum norm in (4) and either the empty space function F or the so-called variance-stabilized Ripley function L [16], another popular functional statistic for point processes. The power was estimated for three values of $r_{\max}$ (0.5, 1.0 and 2.0), passed as parameters of the method. Results of this simulation are shown in Fig. 3, where it can be seen that the power increases with the SNR as expected. The F statistic seems to be more sensible for the detection, regardless of the norm chosen, whereas using L results in lower power in almost all the cases, in accordance to the results described in [8].

Comparison of denoising methods. Simulations of the denoising task where carried out using signals with 512 samples, 30 repetitions per experiment, SNRs of 0, 10, 20 and 30 dB, and using real white Gaussian noise. The employed signals were a cosine-like chirp and three multicomponent signals, the spectrograms of which can be seen in Fig. 2. The plots in Fig. 4 show the QRF for the described methods and the chosen SNR values. It is possible to see that the Hard Thresholding (HT) technique shows the best performance for the monocomponent signal and the one with the impulsive transients (rightmost panel of Fig. 2). For the remaining two signals, the performance worsens. This could be mainly because, for the multiple cosine and linear chirps signals (second and third panels from Fig. 2),
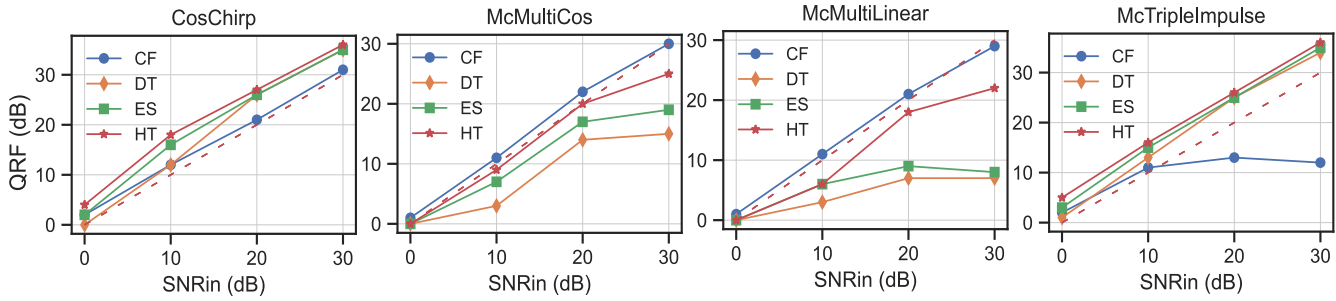
FIGURE 4 – QRF for each signal and method. "CF" : contour-based Filtering, "DT" : delaunay triangulation, "ES" : empty spaces, "HT" : hard-hresholding. "Mc" is for multi-component. The dashed line marks the identity function.
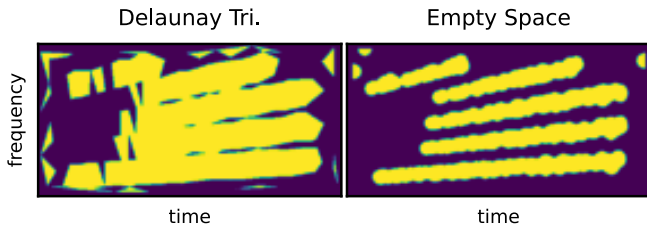

FIGURE 5 – Extraction masks for Delaunay Triangulation and Empty Spaces methods for multiple linear chirps signal.

the TF plane is more occupied, in proportion, by the signal domain, producing an overestimation of the noise variance, computed as in (6). Zeros-based methods seem to fail in the same cases as the HT method. Fig. 5, where the corresponding 1/0 masks used in (5) are depicted for these two methods, reveals that they fail to extract the areas where the components of the signal get close to each other. This might be due to the fact that the zeros in the spectrogram generated by the interference *between* the signal's components are not as separated, while they need to be for the methods to recognize the expected elongated triangles or empty-zeros zones, respectively. This result illustrates an interesting limitation of the methods based on zeros regarding the separation of the modes. One could deal with this by exploring different values of the parameters that govern those methods, which would be another possible usage of the proposed benchmark.

The contours-based filtering shows a similar performance for all signals except the one with impulses. This was previously described as a limitation of the method [6, 14], and some modifications have been proposed to deal with similar cases [17]. In contrast, zeros-based methods are more suited to deal with impulsive structures since these approaches are independent of the orientation of signal's components [8, 13].

6 Conclusions

A public toolbox for benchmarking signal denoising and detection methods was introduced, along with a brief demonstration of its applications. Using the proposed benchmark, the performances of four denoising methods were compared. Additionally, it was shown how the same benchmark can be used to determine the statistical power of signal detection tests. This benchmark could give signal processing practitioners the possibility to independently contrast novel methods in a common framework, making their results more significant. Future work will include an interface for testing Matlab based methods, increase the number of available signals at the Signal Bank and provide users with different kinds of noises.

Références

[1] https://github.com/jmiramont/benchmark-test

[2] P. Flandrin. *Explorations in time-frequency analysis*. Cambridge University Press, 2018.

[3] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume II, Detection Theory*. Prentice Hall PTR, 1998.

[4] H.L. Van Trees, K.L. Bell and Z. Tian. *Detection Estimation and Modulation Theory, Detection, Estimation, and Filtering Theory*. Wiley, 2013.

[5] S. Klimenko et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors *Phys. Rev. D*, 93(4) :042004, 2016.

[6] S. Meignen, T. Oberlin, P. Depalle, P. Flandrin, and S. McLaughlin. Adaptive multimode signal reconstruction from time–frequency representations. *Philos. Trans. Royal Soc. A*, 374(2065) :20150205, 2016.

[7] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process*, 43(5) :1068–1089, 1995.

[8] R. Bardenet, J. Flamant, and P. Chainais. On the zeros of the spectrogram of white noise. *Appl. Comput. Harmon. Anal.*, 48(2) :682–705, 2018.

[9] E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin. Differential reassignment. *IEEE Signal Process. Lett.*, 4(10) :293–294, 1997.

[10] S. Mallat *A Wavelet Tour of Signal Processing, Third Edition : The Sparse Way,* 3rd ed., Academic Press, 2008.

[11] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.

[12] Patrick Flandrin. *Time-frequency/time-scale analysis*. Academic Press, 1998.

[13] P. Flandrin. Time–frequency filtering based on spectrogram zeros. *IEEE Signal Process. Lett.*, 22(11) :2137–2141, 2015.

[14] Y. Lim, B. Shinn-Cunningham, and T. J. Gardner. Sparse contour representations of sound. *IEEE Signal Process. Lett.*, 19(10) :684–687, 2012.

[15] A. Baddeley, P. J. Diggle, A. Hardegen, T. Lawrence, R. K. Milne and N. Gopalan. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, 84(3) :477–489, 2014.

[16] A. Baddeley, E. Rubak and R. Turner. *Spatial point patterns : methodology and applications with R*. CRC press, 2015.

[17] D.H. Pham and S. Meignen. An adaptive computation of contour representations for mode decomposition In *IEEE Signal Process. Lett.*, 24(11) :1596–1600, 2017.