Null or linear-phase filters for the derivation of a new variant of the MSE

E. Grivel
Bordeaux INP
IMS - UMR CNRS 5218
Talence, FRANCE
eric.grivel@ims-bordeaux.fr

B. Berthelot
THALES
Campus Merignac, FRANCE
bastien.berthelot@fr.thalesgroup.com

P. Legrand
Bordeaux University
INRIA/IMB - UMR CNRS 5251
Talence, FRANCE
pierrick.legrand@u-bordeaux.fr

G. Colin
Bordeaux INP
ENSEIRB-MATMECA
Talence, FRANCE
gcolin002@enseirb-matmeca.fr

***Abstract*—The multiscale entropy (MSE) is used in a wide range of applications, especially by physiologists with bio-signals for classification. It consists in estimating the sample entropies of the signal under study and its coarse-grained (CG) versions. The CG process amounts to filtering the signal with an average filter whose order is the scale and then to decimating the filter output by a factor also equal to the scale. In this paper, the novelty stands in the way to get the sequences at different scales, by avoiding distortions that can appear during the decimation step. If a low-pass filter is implemented, the cut-off frequency of which is well suited to the decimation factor, the phase distortions induced by the filter must be attenuated as much as possible. Two ways are considered in this paper: 1/ design a linear-phase finite-impulse-response filter with the window method or the Remez algorithm. 2/ design a zero-phase filter from any infinite impulse response filter. Simulations with white noises and $1/f$ processes are given. The way the sample entropy evolves with the scale is presented. Depending on the filter parameters, one can see how it differs from the evolution obtained with the CG.**

***Index Terms*—MSE, entropies, coarse-grained.**

I. Introduction

In statistical signal processing, a standard processing chain consists in extracting some features from the data that have been collected with sensors. These features make it possible to characterize the set of samples. They can help experts understand some physical phenomena. Thus, the signal power in some frequency bands can be representative of the activity of the sympathetic or parasympathetic nervous system or the activity of the brain. Among the possible features, one may search for an *a priori* model to represent the data by a small set of parameters. Thus, moving average autoregressive (ARMA), fractionally integrated (FI) but also ARFIMA processes can be considered. The first one can be used to model short-memory processes while the others are relevant for long-memory processes. Discrete fractional Gaussian noise (dfGn) as well as discrete fractional Brownian motion (dfBm) can be also of interest. Whatever the model, the model parameters are related to the power spectral density (PSD) and consequently to the correlation function. Other features can be used and are related to the regularity and the auto-similarity property of the signal under study. In that case, the Hurst exponent can be of interest and estimated with different approaches such as the fluctuation analysis, the detrended fluctuation analysis [19] or some variants [3].

In this paper, we focus our attention on another feature, namely the entropy and more particularly the so-called multiscale entropy (MSE) [8]. While Shannon entropy but also Rényi and Sharma-Mittal entropies are widely used in signal processing and can be the core of various subjects of research such the analysis of the entropy rates [11] [10] in the field of information theory, the sample entropy (SE) and the MSE are less popular in our community. It has been mainly used in biomedical applications by physiologists, but also for traffic time series real vibration data, diagnosing rotating machinery faults and financial markets. As various variants of the MSE were proposed, a brief state of the art will be given in the paper so that the reader can have a clear picture of the situation.

The standard version of the MSE is based on the following principle: estimating the SE of the process at various ”scales”. This implies creating coarse-grained (CG) time series. From a signal processing point of view, the signal is filtered by an averaging finite-impulse-response (FIR) causal filter whose order corresponds to the scale. The frequency response of the filter is low-pass, and so all the more as the scale is high. Then the filter output is decimated by a factor equal to the scale. Finally, the sum of the SE of the CG time series is used as a feature. As mentioned by various authors [22] [13] and [26], the decimation may be problematic if the design of the filter is not properly done. Indeed, Shannon theorem should be satisfied for each scale. For this reason, in [22], the FIR filter is replaced with an infinite-impulse-filter (IIR) low-pass Butterworth filter whose cut-off frequency is chosen to *a posteriori* decimate the signal properly. More recently in 2021, the authors in [26] use the same kind of filter[1].

[1]the resulting method is called parallel MSE by the authors.

However, the phase distortions induced by these filters in the pass band should be avoided. The Butterworth filter is neither a linear-phase filter nor a null-phase filter. As a consequence, the above versions address only a part of the problem and introduce another issue, namely the filter-induced phase distortion. Although these methods provide markers that are used by the physiologists and other practitioners, they remain questionable from the signal processing perspective. In this paper, we propose to consider two types of solutions to fix that problem: the first one consists in designing a linear-phase FIR filter. The window method or the Remez algorithm can be used and we will see which method is the most relevant. The second one is based on an IIR-based structure which leads to a null-phase equivalent filter.

The remainder of this paper is organized as follows. Section II deals with the multiscale entropy, its definition, but also a state of the art on its variants. Then, in section III, we present our contribution. This includes theoretical aspects but also illustrations. In terms of methodology and according to what was done before in the literature, when the authors want to compare some variants, synthetic processes such as a white noise and a $1/f$ noise are considered. Then, the authors present how the measure evolves with the scale, based on a certain number of trials.

II. MULTISCALE ENTROPY: DEFINITION, COMMENTS AND VARIANTS

A. Preamble: the sample entropy

The sample entropy ($SampEn$ or SE) was proposed by Richman and Moorman [20]. For a vector[2] $\underline{x}_1$ storing the samples$\{x_n\}_{n=1,...N}$ of the signal x, the SE can be expressed as follows:

$$SE(\underline{x}_1, \tau, m, r) = -\log \frac{\Phi_r^{m+1}\left(\begin{bmatrix} x_1 & x_{1+\tau} \cdots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)}{\Phi_r^{m}\left(\begin{bmatrix} x_1 & x_{1+\tau} \cdots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)} \quad (1)$$

where $\lfloor . \rfloor$ is the floor function. In addition, $\Phi_r^{m}\left(\begin{bmatrix} x_1 & x_{1+\tau} \cdots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)$ is the estimated probability that the distance between two different vectors storing m values of the sample set is smaller than a tolerance level r. Note that m and r are chosen by the user. m is often set at 2 while r depends on the variance of the signal under study. When the scale τ is equal to 1, the original signal is studied and the total number of vector pairs that can be considered is equal to $\frac{(N-m)(N-m-1)}{2}$.

Among the results that have been established, the SE of a white noise and a $1/f$ noise were studied for different scales τ. An analytical expression for the unit-variance zero-mean white noise was obtained [8] [15]:

$$SE(\underline{x}_1, \tau, m, r) = \quad (2)$$
$$-\log\left(\int_{-\infty}^{+\infty} \sqrt{\frac{\tau}{8\pi}} \left[erf\left(\frac{x+r}{\sqrt{\frac{2}{\tau}}} \right) - erf\left(\frac{x-r}{\sqrt{\frac{2}{\tau}}} \right) \right] e^{-\frac{1}{2}x^2} dx \right)$$

[2]The subscript $_1$ refers to the first sample of the set of samples that are considered.

where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp(-t^2)dt$ refers to the error function. Given (2), the SE of a white noise monotonically decreases with τ whereas the entropy of a CG $1/f$ noise is a constant. In addition, the SE is sensitive to short duration signals. Indeed, Richman mentioned that more than 10^m samples are required to get a good estimation of $SE(\underline{x}_1, \tau, m, r)$.

However, there is no clear relationship between the entropy-based regularity and the complexity[3]. Indeed, when the above quantity was directly used with physiologic signals and more particularly heartbeat interval series, larger values were obtained from pathological subjects than healthier ones. The contrary was rather expected by the experts because complexity is related to the ability of living systems to adapt themselves to the situation. Therefore, different scales of the signal were considered, leading to the MSE.

B. Multiscale entropies (MSE): a measure of complexity

Twenty years ago, the MSE was proposed by Costa *et al.* [8] to evaluate the complexity of a signal. It consists in summing the SE of the signal itself but also the SEs of the time-series derived from a CG procedure applied on the signal at different scales. The $\lfloor \frac{N}{\tau} \rfloor$ samples $y_{n,k}^{(\tau)}$ are deduced by computing the following arithmetic mean of the samples of the signal x:

$$y_{n,k}^{(\tau)} = \frac{1}{\tau} \sum_{i=(n-1)\tau+k}^{n\tau+k-1} x_i \qquad \text{for } k = 1, .., \tau \quad (3)$$

Thus, $y_{n,1}^{(1)} = x_n$. Applying the CG on the signal x with a factor τ amounts to constructing the vector $\underline{y}_k^{(\tau)}$ by stacking the values $\{y_{n,k}^{(\tau)}\}$ with $n \in \{1, ..., \lfloor \frac{N}{\tau} \rfloor\}$. From a signal processing point of view, this amounts to doing the following two steps:

1. **Applying on the signal x a causal averaging filter**

Its impulse response is given by $h_n^{(\tau)} = \frac{1}{\tau}$ for $n = 0, ..., \tau - 1$ and zero elsewhere. Its frequency response is given for the normalized frequency $f \in [-\frac{1}{2}, \frac{1}{2}[$ by:

$$|H^{(\tau)}(f)| = \begin{cases} \left| \frac{\sin(\pi f \tau)}{\tau \sin(\pi f)} \right| \text{ if } f \neq 0 \\ 1 \text{ if } f = 0 \end{cases} \quad (4)$$

It is a low-pass filter which totally rejects the normalized frequencies $\frac{k}{\tau} \in [-\frac{1}{2}, \frac{1}{2}[$ with k positive or negative integers. Due to the symmetry of the impulse filter, the filtering has the advantage of having a linear phase, leading to a constant group delay and no phase distortion of the signal in the band pass. Only the steady state is considered in the following.

Remark: The choice of the CG based on (3) is not explicitly motivated in the pioneering works [8]. It is true that CG is often considered when dealing with dfGn processes (because the normalized correlation function does not change with the scale) and multifractal analysis.

2. **Decimating the filter output by a factor τ**

Due to the decimation, the frequency components of the signal whose normalized frequencies are above $\frac{1}{2\tau}$ will be a source

[3]The complexity measure aims to represent the irregularities in time series and makes it possible to indicate to what extent it is possible to recover all the information from partial information.

of aliasing. As the filters defined above have their main lobes between $-\frac{1}{\tau}$ and $\frac{1}{\tau}$ and the side lobes are not necessarily much attenuated, the filters are not well suited to be anti-aliasing filters.

Given the above two steps, the SE is computed for each scale and the MSE is defined by the sum $\sum_{\tau=1}^{\tau_{max}} SE\left(\underline{y}_1^{(\tau)}, 1, m, r\right)$:

$$MSE(\underline{x}_1, \tau_{max}, m, r) = -\sum_{\tau=1}^{\tau_{max}} \log \frac{\Phi_r^{m+1}(\underline{y}_1^{(\tau)})}{\Phi_r^m(\underline{y}_1^{(\tau)})} \quad (5)$$

where τ_{max} is chosen by the user. It should be noted that only $\underline{y}_k^{(\tau)}$ with $k = 1$ is used. In addition, one of the problems of the MSE algorithm is the use of the same tolerance level r for all scales [18]. Therefore, some variants were proposed.

C. Some existing variants

In 2009 and 2021, the refined MSE [22] and the parallel MSE [26] were based on the same idea: replacing the FIR filter by an IIR low-pass Butterworth filter whose cut-off frequency is chosen to properly downsample the signal. However, this filter still introduces phase distorsions.

In 2013, the composite MSE consisted in using the τ sequences that can be defined after the decimation step [24]:

$$cMSE(\underline{x}_1, \tau_{max}, m, r) = \sum_{\tau=1}^{\tau_{max}} cSE(\underline{x}_1, \tau, m, r) \quad (6)$$

where $cSE(\underline{x}_1, \tau, m, r) = \frac{1}{\tau}\sum_{k=1}^{\tau} SE\left(\underline{y}_k^{(\tau)}, 1, m, r\right)$ corresponds to the average of the sample entropies computed on the τ sequences obtained from the filtering and the decimation. Averaging reduces the variance of the estimated sample entropy values. Moreover, using (5), (6) can be rewritten as:

$$cMSE(\underline{x}_1, \tau_{max}, m, r) = -\sum_{\tau=1}^{\tau_{max}} \log \frac{(\prod_{k=1}^{\tau} \Phi_r^{m+1}(\underline{y}_k^{(\tau)}))^{\frac{1}{\tau}}}{(\prod_{k=1}^{\tau} \Phi_r^{m}(\underline{y}_k^{(\tau)}))^{\frac{1}{\tau}}} \quad (7)$$

In [4], what is called "short-term MSE" is also described by (7). It amounts to computing the log of geometric means. The refined composite multiscale entropy (rcMSE) proposed in [25] rather consists in computing the log of the means:

$$rcMSE(\underline{x}_1, \tau_{max}, m, r) = -\sum_{\tau=1}^{\tau_{max}} \log \frac{\sum_{k=1}^{\tau} \Phi_r^{m+1}(\underline{y}_k^{(\tau)})}{\sum_{k=1}^{\tau} \Phi_r^{m}(\underline{y}_k^{(\tau)})} \quad (8)$$

Note that the geometric mean is less sensitive than the arithmetic mean to the highest values of a set.

In the modified MSE [24], the averaging filter is still applied. Instead of decimating the filter output, the sample entropies are directly computed on y. According to the authors, this has the advantage of increasing the number of vector pairs.

The concept of multiscale was also extended to other entropy metrics such as the permutation entropy [2]. This led to the family of multiscale permutation entropies [21] which includes the refined composite multiscale permutation entropy [14], the improved multiscale permutation entropy [1], the multiscale fractional-order permutation entropy and its weighted version [6] or the fractional multiscale phase permutation entropy [23]. Starting from the fuzzy entropy [7], the multiscale fuzzy sample entropy was derived as well as the composite and refined composite multiscale fuzzy entropy [27]. In 2015, the generalized multiscale entropy was proposed in [9]. Instead of computing the mean of the signal in the first step of the MSE approach, a higher-order moment (for instance the variance) is considered. When dealing with the variance, one obtains the so-called MSEσ^2. As the length of CG series decreases when the scale increases, the estimation of the sample entropy may not be accurate. For this reason, the refined MSEσ^2 was proposed in [16], following the same methodology as in the rcMSE. As for the solution proposed by Hu *et al.* [12], it is based on the decomposition of the signal into intrinsic mode functions (IMFs) by using the multivariate empirical mode decomposition (MEMD). As the MEMD essentially acts as dyadic-filter banks, the authors suggest summing a certain number of IMFs in order to get different sets of data whose spectrum tends to exhibit lower and lower frequencies. The number of IMFs that are considered corresponds to the scale. Then, the SE is computed on each signal. This leads to the adaptive multiscale entropy. In [17], the authors suggest applying the CG procedure, where the sample average is replaced with the median value. Then, the authors define the corresponding sign time series. When the difference between two consecutive samples is positive, the value is equal to 1. Otherwise, it is equal to 0. As a consequence, the number of possible patterns is finite and the authors use both the sample entropy and the second entropy based on the sign time series. This method is called multiscale symbolic entropy analysis. In the hierarchical entropy [15], the signal is decomposed into two sequences: the low-pass filtered one obtained by the averaging filter and the high-pass filtered one obtained by the difference filter. Then, each resulting sequence is decomposed into two sequences: a low-pass and a high-pass one. This leads to a "hierarchical" decomposition of the signal. Then, the SE is computed for each component. In [5], the Bubble entropy leads to time-shift multiscale bubble entropy.

III. Designing a filtering step well suited to carry out a multiscale analysis

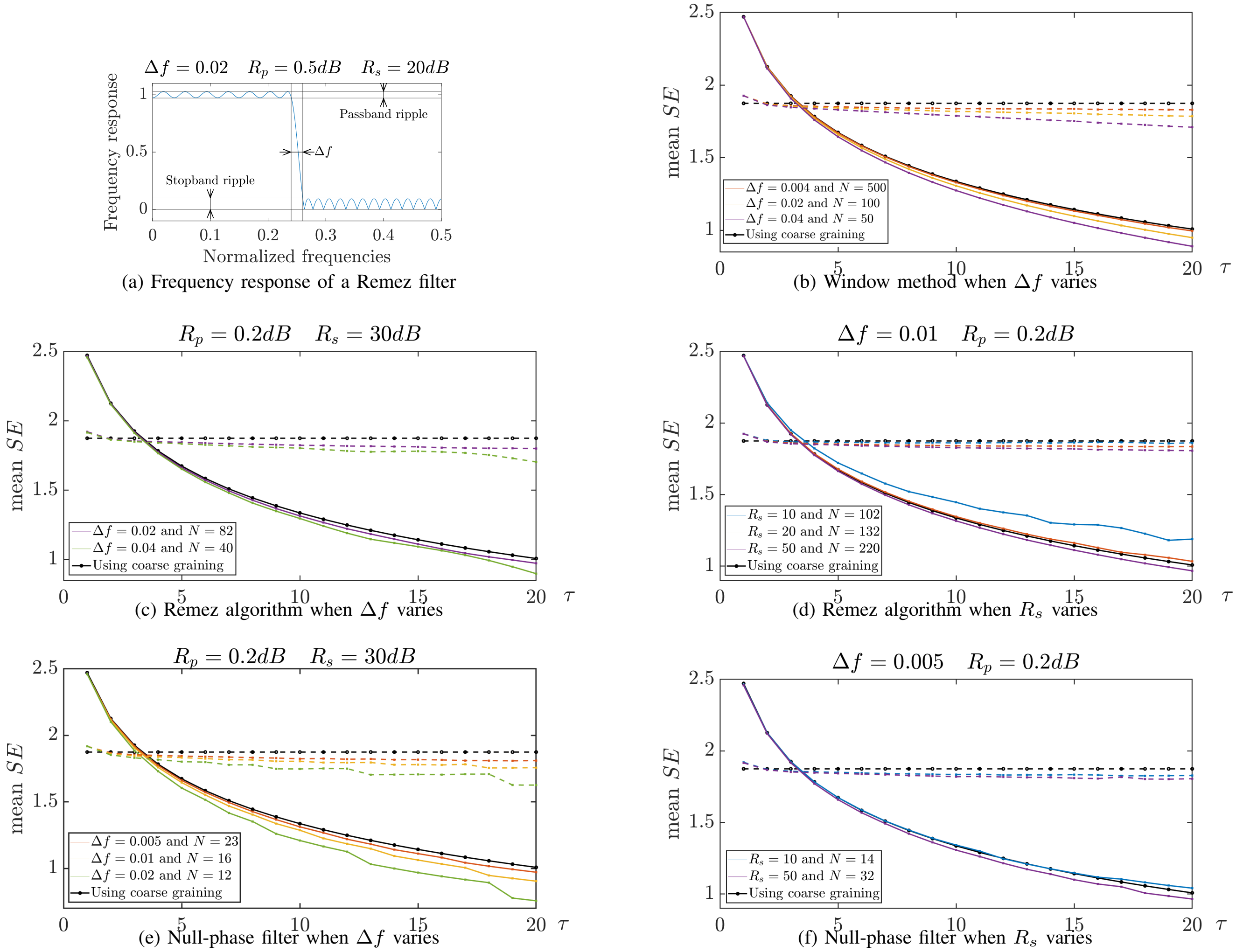
A. Steps of the MSE to follow in order to avoid spectrum overlap and distorsions induced by the filter phase

Given the above state of the art, we now propose our contribution. It consists in deriving a MSE by taking care of the filtering step. It operates with the following steps:

Step 1. Defining the maximum scale τ_{max}.

Step 2. For each scale $\tau \in [2, ..., \tau_{max}]$, selecting a low-pass causal FIR filter, with a linear phase or a null phase and a normalized cutoff frequency that is smaller than $\frac{1}{2\tau}$. This can be done by using the so-called window method for a causal linear-phase FIR filter design or Remez algorithm. An alternative is to design an IIR filter using the bilinear transform and then to design a zero-phase linear filter: two types of processing chains can be considered:

• The signal x_n is filtered by a filter whose real impulse response is h_n, leading to the signal g_n. Then, the time-reversed version g_{-n} is filtered by a filter whose impulse response is still equal to h_n. The output is denoted as r_n.



(a) Frequency response of a Remez filter

(b) Window method when Δf varies

(c) Remez algorithm when Δf varies

(d) Remez algorithm when R_s varies

(e) Null-phase filter when Δf varies

(f) Null-phase filter when R_s varies

Figure 1. Evolution of the SE according to τ for different choices of Δf and R_s. white noise: ——; pink noise: – –

Finally, r_n is time-reversed to get the filtered signal of interest, denoted as y_n.
Let $H(f) = |H(f)| \exp(j\Phi(f)) = H^*(-f)$, one has:

$$\begin{cases} G(f) = H(f)X(f) \\ R(f) = H(f)G(-f) \\ Y(f) = R(-f) = |H(f)|^2 X(f) \end{cases} \tag{9}$$

The Fourier transform of the impulse response of the equivalent filter is $H_{equ}(f) = |H_{equ}(f)| \exp(j\Phi_{equ}(f)) = |H(f)|^2$. It is hence a null-phase filter ($\Phi_{equ}(f) = 0$) characterized by $|H_{equ}(f)| = |H(f)|^2$.

• The sequences x_n and x_{-n} are filtered by the same filter whose impulse response is h_n. This leads to two sequences respectively denoted as g_n and r_n. The final output is equal to $g_n + r_{-n}$. In that case, the Fourier transform of the output is $Y(f) = 2|H(f)|^2 \cos(\Phi(f))$. It is a zero-phase filter if $\Phi(f) \in [0, \frac{\pi}{2}[$ in the pass band. As there is no guarantee this is always the case, only the first strategy is considered. In this case, the filter can be either FIR or IIR. For IIR, different filter families exist such as the Butterworth filters whose order can be large due to the constraint of maximally-flat magnitude in the pass band and the Chebyshev filters known to exhibit equiripple in the pass band (type I) or the stop band (type II).

The specifications of the low-pass filter frequency response have to be defined. It consists in choosing the transition bandwidth $\Delta f = f_s - f_p$ where f_s and f_p respectively denote the normalized stop-band corner frequency and the normalized pass-band corner frequency. For the window method and Remez algorithm, f_s and f_p are defined as $\frac{1}{2\tau} \pm \Delta f/2$. For the null-phase filter, f_p and f_s are respectively defined as $\frac{1}{2\tau} + 10^{-3}$ and $f_p + \Delta f$. Moreover, the maximum permissible pass-band loss R_p as well as the stop-band attenuation R_s have to be defined. We will see the influence of their choices in section III-B.

Step 3. Computing the MSE (or one of its variants).

B. Application on white noises and $1/f$ process

In this section, we propose to follow the same methodology as [8] and [22] by providing the way the entropy measure evolves with the scale for white noises and $1/f$ noise. As the approaches we propose depend on the specifications of the low-pass filter, we propose to look at the way the curves vary with one of the specification parameters. They will be compared with the curves obtained the standard MSE [8].

In Fig. 1, N denotes the length of the impulse response for the window method (1.b) and the Remez algorithm (1.c and 1.d) whereas it denotes the order of the type-II Chebyshev filter for the null-phase filter structure (1.e and 1.f). In both cases, the larger N, the larger the computational cost. The steps described in III-A are performed on 50 signals of length 30000. The results are then averaged to obtain the curves.

According to our simulations, when R_p is modified for the design of the IIR filter, it does not change much the shape of the curve. For this reason, we do not present them in Fig. 1. When R_s is taken too small, the linear-phase or null-phase low-pass filter with the normalized cutoff frequency smaller than $\frac{1}{\tau}$ do not filter the frequencies in the stop band enough. Therefore, like the CG-based MSE, spectrum overlapping may occur. The resulting evolution of the entropy measure tends to the one obtained with the CG-based MSE. When the transition bandwidth becomes too wide, the stop band becomes smaller. Some frequencies in the transition band are not more or less attenuated and the resulting entropy diverts from the SE. The proposed results serve as reference curves for this new MSE as the ones presented in [8] and in [22].

IV. Conclusions and perspectives

To replace the standard coarse graining step of the MSE (or its variants) that induces aliasing, two linear-phase FIR filters and a null-phase IIR filter have been used in our variant. Following the same type of analysis as in [8] and [22], the sample entropies of white noise and $1/f$ noise at different scales are studied. As the parameter of the filters can be tuned by practitioner, our simulations confirm that if the stop-band ripple is too small or if the transition bandwidth of the filter is too wide, the anti-aliasing properties of the filter become lesser. As the FIR-filter order may be large, it may be less relevant when the number of samples available is low and we suggest using the null-phase filter structure. Our proposal, based on a linear or zero phase filter, seems moderate from the signal processing point of view, but the resulting MSE could serve a large community of users.

References

[1] H. Azami and J. Escudero. Improved multiscale permutation entropy for biomedical signal analysis:interpretation and application to electroencephalogram recordings. *Biomedical Signal Processing and Control*, 23:28–41, 2016.

[2] C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88:174102, 2002.

[3] B. Berthelot, E. Grivel, and P. Legrand. New variants of DFA based on LOESS and LOWESS methods: generalization of the detrended moving average. *ICASSP*, 2021.

[4] Y. C. Chang, H.T. Wu, H.R. Chen, A. B. Liu, J. J. Yeh, M. T. Lo, J. H. Tsao, C. J. Tang, I. T. Tsai, and C. K. Sun. Application of a modified entropy computational method in assessing the complexity of pulse wave velocity signals in healthy and diabetic subjects. *Entropy*, 16:4032–4043, 2014.

[5] F. Chen, W. Tian, L. Zhang, J. Li, C. Ding, D. Chen, W. Wang, F. Wu, and Wang B. Fault diagnosis of power transformer based on time-shift multiscale bubble entropy and stochastic configuration network. *Entropy*, 24, (8):1135, 2022.

[6] S. Chen, P. Shang, and Y. Wu. Multivariate multiscale fractional order weighted permutation entropy of nonlinear time series. *Physica A*, 515:217–231, 2019.

[7] W. Chen, Z. Wang, H Xie, and W. Yu. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15, (2):266–72, 2007.

[8] M. Costa, A. L. Goldberger, and C. K. Peng. Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, 89:068102, 2002.

[9] M. D. Costa and A. L. Goldberger. Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series. *Entropy*, 17:1197–1203, 2015.

[10] L. Golshania and E. Pasha. Rényi entropy rate for gaussian processes. *Information Sciences*, 180:1486–1491, 2010.

[11] E. Grivel. Expressions and properties of the shannon, rényi, sharma-mittal and tsallis cross-entropies for gaussian arma processes. *EUSIPCO 2022*, 2022.

[12] M. Hu and H. Liang. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Transactions on Biomedical Engineering*, 59:12–15, 2012.

[13] A. Humeau-Heurtier. The multiscale entropy algorithm and its variants: A review. *Entropy*, 17:3110–3123, 2015.

[14] A. Humeau-Heurtier, C.-W. Wu, and S-D Wu. Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence. *IEEE Signal Processing Letters*, 22:2364–2367, 2015.

[15] Y. Jiang, C. K. Peng, and Y. Xu. Hierarchical entropy analysis for biological signals. *Journal of Computational and Applied Mathematics*, 236:728–742, 2011.

[16] Y. Liu, Y. Lin, J. Wanga, and P. Shang. Refined generalized multiscale entropy analysis for physiological signals. *Physica A*, 490:975–985, 2018.

[17] M. T. Lo, Y. C. Chang, C. Lin, H.-W. V. Young, Y.-H. Lin, Y.-L. Ho, C.-K. Peng, and Hu K. Outlier-resilient complexity analysis of heartbeat dynamics. *Scientific reports*, 5:8836, 2015.

[18] V. V. Nikulin and T. Brismar. Comment on multiscale entropy analysis of complex physiologic time series. *Physical Review Letter*, 92:089803, 2004.

[19] C. K. Peng, S. Havlin, J. M. Hausdorf, J. E. Mietus, H. E. Stanley, and A. L. Goldberger. Fractal mechanisms and heart rate dynamics. *Journal of Electrocardiology*, 28:59–64, 1996.

[20] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278:H2039–H2049, 2000.

[21] V. Vakharia, V. K. Gupta, and P. K. Kankar. A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings. *Journal of Vibration and Control*, 21, (16):3123–3131, 2015.

[22] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal. Refined multiscale entropy: application to 24-h holter recordings of heart periodvariability in healthy and aortic stenosis subjects. *IEEE Transactions on Biomedical Engineering*, 56:2202–2213, 2009.

[23] L. Wan, G. Ling, Z. H. Guan, Q. Fan, and Y. H. Tong. Fractional multiscale phase permutation entropy for quantifying the complexity of nonlinear time series. *Physica A*, 600:127506, 2022.

[24] S. D. Wu, C. W. Wu, S. G. Lin, C. C. Wang, and K.Y. Lee. Time series analysis using composite multiscale entropy. *Entropy*, 15:1069–1084, 2013.

[25] S. D. Wu, P. H. Wu, C. W. Wu, J. J. Ding, and C. C. Wang. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy*, 14:1343–1356, 2012.

[26] D. Zhao, S. Liu, S. Cheng, X. Sun, L. Wang, Y. Wei, and H. Zhang. Parallel multi-scale entropy and its application in rolling bearing fault diagnosis. *Measurement*, 168, 108333, 2021.

[27] J. Zheng, H. Pan, and J. Cheng. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mechanical Systems and Signal Processing*, 85:746–759, 2017.