Multiscale cross entropy rate as a way to compare signals: application to Gaussian ARMA processes

Eric Grivel
Bordeaux University - INP Bordeaux
IMS lab - UMR CNRS 5218
Talence, France
eric.grivel@ims-bordeaux.fr

Bastien Berthelot
THALES Avionics France
Campus THALES
Merignac, France
bastien.berthelot@fr.thalesgroup.com

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

***Abstract*—The entropy rate of a stochastic process corresponds to the asymptotic difference between the entropies of consecutive sample blocks as their size increases. Widely used in information theory, it also serves as a key marker for signal characterization in classification tasks. Recently, we studied the entropy rate of a signal at different scales using a multiscale approach. The latter generates a set of signals from the original one either (i) by applying a coarse-graining (CG) procedure —where the signal is filtered with an average filter of order equal to the scale and then decimated by a factor equal to the scale— or (ii) by directly decimating the original signal. In this paper, we extend the multiscale framework to the cross-entropy rate, introducing the multiscale cross-entropy rate (MCER). MCER can be defined either as the sum of cross-entropy rates across scales or as a vector storing these values. By applying it to Gaussian ARMA processes, we aim to understand the insights provided by the multiscale procedure and to define the influence of the process parameters on the cross-entropy rate at each scale. To this end, we present the properties of ARMA processes after applying the multiscale procedure, provide analytical expressions for the MCER, and outline a practical method for deriving it. The MCER is a potential alternative to multiscale cross-sample entropy and its variants, which have been used in biomedical applications and finance to quantify joint synchrony between signals.**

***Index Terms*—cross entropy, rate, ARMA processes.**

I. Introduction

Cross entropy (CE) is one of the metrics widely used in information theory where it measures the average number of bits needed to identify an event drawn from the set when the coding scheme used for the set is optimized for an estimated probability distribution p_2 instead of the true distribution p_1. It is also used in deep learning methods as a loss function to be minimized. When dealing with model evaluation in statistics, CE makes it possible to compare statistical models by measuring the likelihood of the data under each model. The lowest CE indicates a better fit to the data.

In information theory, the most popular CE between two probability density functions (pdf) is the Shannon CE. A more general form, the Rényi CE of order α, is defined using the Rényi entropy and Rényi divergence. As α tends to 1, applying L'Hospital rule makes it possible to recover the Shannon CE. To study a dynamical system, with F degrees of freedom, observed every δ seconds and characterized by a "phase space" partitioned into a collection of boxes of size r, Kolmogorov–Sinai (KS) entropy was proposed and corresponds to the Shannon entropy rate, defined as the entropy per unit time, when δ and r tend to 0. Approximations of KS entropy in practical settings led to the approximate entropy [18] and the sample entropy (SE) [20]. Therefore, it is not a surprise if the cross-approximate entropy (CAE) [19] was proposed. It is used to quantify asynchrony between two time series, which refers to the extent to which the stochastic processes exhibit coordinated behaviors. To improve the computational cost, Škorić *et al.* proposed the binarized CAE [33], where the binary encoding results from the value of the difference between two consecutive samples of a signal. Then, the cross-sample entropy (CSE) [20] between two stochastic processes u and v of length N was defined as follows:

$$CSE(u, v, m, r) = -\ln \frac{A^{(m+1)}(u, v, r)}{B^{(m)}(u, v, r)} \tag{1}$$

with $A^{(m+1)}(u, v, r)$ and $B^{(m)}(u, v, r)$ the probabilities that u and v match for $m + 1$ and m points, with a tolerance r.

CSE is known to be a measure of joint synchrony that is more consistent than the CAE. A low CSE is obtained when the time series are highly synchronized, whereas it becomes larger for unsynchronized ones.

Then, variants were proposed such as the modified CSE [29], the modified CSE based on symbolic representation and similarity that is more robust to noise [26], the Kronecker-Delta-based CSE [11], the refined CSE using the cumulative histogram method [22] or the refined CSE based on Freedman-Diaconis rule [3]. The concepts used to derive the permutation entropy (PE) [1] and the fuzzy entropy (FE) [21] were also applied to get the permutation CE [11], the joint PE [30] and the cross FE [27].

Then, the multiscale extensions of some of the above measures were proposed. The principle is inspired by the one used for the multiscale sample entropy (MSE) [4]. It consists in deriving different signals from the original one. Three main approaches have been used:

1/ **the "data-driven" approach** usually consists in applying an empirical mode decomposition (EMD) or a variant [24] and computing the measure with each intrinsic mode functions.

2/ **the CG method** was initially used to get robust information of a dynamical system. It consists in mapping the signal into symbol sequences. Different mapping strategies can be considered, based on operators such as min, max, and average. The average is commonly used for the multiscale entropy measures. In that case, from the signal processing point of view, the CG approach based on averaging amounts to filtering the signal by an averaging causal finite-impulse-response (FIR) filter whose order is the scale s. The FIR is defined by $h_{CG,t} = \frac{1}{s}$ for $t = 0, ..., s-1$, leading to the transfer function:

$$H_{CG}(z) = \frac{1}{s}\sum_{j=0}^{s-1} z^{-j} \tag{2}$$

The filter output is then decimated by a factor equal to s. This means that one sample out of s is kept.

Remark 1: a discrete fractional Gaussian noise (dfGn)[1] remains a dfGn with the same Hurst exponent H after CG. If σ^2 is the variance of the dfGn, the variance of the CG sequence at scale s is equal to $\sigma^2 s^{2H-2}$. This is the case of a white noise with $H = 0.5$.

Remark 2: at each scale s, only the decimated signal starting by the 1^{st} sample of the original signal and whose length is equal to $\lfloor \frac{N}{s} \rfloor$, where $\lfloor . \rfloor$ is the floor function, is considered. However, in the composite multiscale procedure, the decimated signals starting by the other $s-1$ first samples of the original signal are also considered. In that case, the measures are calculated on the resulting s decimated sequences and averaged [25].

3/ **the time-shift (TS) procedure** consists in directly decimating the original signal by a factor s. In that case, the resulting s sequences, each one starting by one of the first s samples of the original signal, are the results of the multiscale procedure.

Given the above concepts, the CSE was combined with the CG to get the multiscale cross sample entropy (MSCE) [28]. To reduce the probability of undefined entropy especially for short-duration signals, some variants were developed such as the asymmetric multiscale cross-sample entropy that also takes into account asymmetry [29], the composite multiscale cross-sample entropy and a refined version [31]. However, other combinations of a multiscale approach and a type of CE can be derived. The reader can refer to the review paper written by Jamin *et al.* [14]. Since then, the composite multiscale partial CSE has been proposed to quantify the intrinsic similarity of two time series linearly affected by a common external factor [15]. In [12], the authors have derived a multiscale cross fuzzy entropy. Finally, some authors recently suggested combining the MSCE with the horizontal visibility graph mapping time series to complex networks [16] [32].

The multiscale procedure makes it possible to create a set of signals from a single one, thereby generating a set of markers for signal classification. Consequently, these multiscale-based measures have been applied in many applications: In finance with time series in stock markets. In engineering to evaluate coupling between sensors. In biomedical applications to analyze the link between complexity, age and physical state of a population by using handlebar angle and speed time series recorded from a bike simulator [13] or to study the diagnosis of epilepsy [16].

Numerous articles have been published on the use of signal processing methods with real data, but their theoretical contributions often remained limited, and the explainability of the results was frequently underdeveloped. To address this, researchers have turned to synthetic stochastic processes such as white noise, 1/f noise, and autoregressive (AR) or moving average (MA) processes. Thus, the analytical expression of the SE for a white noise was derived by Costa [4]. In addition, the SE of a 1/f noise is a constant. In 2018, the analytical expressions of the multiscale PE applied to a dfGn was derived [5]. In 2019, using [2], the case of the 1^{st}-order AR and 1^{st}-order MA processes were addressed [6].

Given all this context information, let us go back to Shannon entropy and its rate which corresponds to the difference between the measures computed for $k+1$ and k variates when k tends to infinity. It means that the measure can be asymptotically seen as an affine function whose slope is the rate. Recently, we suggested combining this rate with CG [10] and analyzed the behaviour of the resulting multiscale entropy rate for 1^{st}-order AR and 1^{st}-order MA processes as well as dfGn. In this new paper, we propose to derive the multiscale cross-entropy rate (MCER) by combining the cross-entropy rate and a multiscale procedure based on CG or TS. Our purpose is to identify some properties of ARMA processes after the multiscale procedure and to study the MCER when dealing with Gaussian ARMA processes. We will see that it allows us to better understand the influences of the ARMA parameters and the scale on the rates and that MCER can be used to characterize stochastic process for classification.

The remainder of this paper is organized as follows: in section II, we recall the main properties of the ARMA processes and present some properties of the ARMA processes after the multiscale procedure. In section III, we introduce the concept of MCER by providing an analytical expression dealing with Gaussian ARMA processes. Illustrations are then given.

II. About Gaussian real ARMA processes

A. Definitions and properties

Let us define the t^{th} sample x_t of a Gaussian zero-mean (p,q)ARMA process:

$$x_t = -\sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=0}^{q} b_j u_{t-j} \tag{3}$$

where $\{a_j\}_{j=1,...,p}$ and $\{b_j\}_{j=0,...,q}$ respectively denote the AR and MA parameters with $b_0 = 1$. Moreover, the driving process u_t is a zero-mean white Gaussian sequence with variance σ_u^2. When the ARMA parameters and σ_u^2 do not vary over time, the ARMA process is wide-sense stationary.

The ARMA process is also a source-filter process: given (3), x_t is the output of a linear filter whereas the input is the driving process u_t . The corresponding transfer function is characterized by its q zeros $\{z_l\}_{l=1,...,q}$ and its p poles $\{p_l\}_{l=1,...,p}$:

$$H(z) = \frac{\prod_{l=1}^{q}(1 - z_l z^{-1})}{\prod_{l=1}^{p}(1 - p_l z^{-1})} \tag{4}$$

Poles with a modulus close to 1 create resonances in the the power spectral density (PSD). In contrast, zeros generate spectral rejections. When a zero has its modulus equal to 1,

[1] If H is the Hurst exponent, its normalized covariance function is defined for $k > 0$ as $\bar{r}_k = \frac{1}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right)$.

the PSD is null at the normalized angular frequency equal to the argument of the zero.

The pdf of a vector storing k consecutive samples of the ARMA process is characterized by its covariance matrix Q_k. Q_k is non singular even if the PSD is null at some frequencies. However, the infinite-size Toeplitz covariance matrix is not invertible if the transfer function has unit zeros.

Moreover, the elements of Q_k corresponds to a value of the covariance function $r_{x,\tau}$. Given (3), it can be shown that $r_{x,\tau}$ depends on both the AR and the MA parameters when $|\tau| \leq q$. Moreover, $r_{x,\tau}$ for $\tau \geq q+1$ can be expressed as follows:

$$r_{x,\tau} = -\sum_{j=1}^{p} a_j r_{x,\tau-j} \quad (5)$$

Conversely if (5) is satisfied for $\tau \geq q+1$, the AR part can be deduced as well as the order q.

Thus, for a real $(2,2)$ARMA process, by introducing $\beta_1 = 1 - a_1^2 - a_2^2$ $\beta_2 = 1 + a_2$, and:

$$\gamma_1 = \frac{1 + b_1^2 + b_2^2 - 2a_1b_1 - 2a_2b_2 - 2a_1b_1b_2 + 2a_1^2b_2}{\beta_1} \quad (6)$$

$$\gamma_2 = \frac{2a_1a_2(b_1 + b_1b_2 - a_1b_2)}{\beta_1\beta_2} \text{ and } \gamma_3 = 1 + \frac{2a_1^2a_2}{\beta_1\beta_2}$$

one has for $\tau \geq 3$:

$$\frac{r_{x,0}}{\sigma_u^2} = \frac{\gamma_1 + \gamma_2}{\gamma_3}, r_{x,1} = -\frac{a_1}{1+a_2} r_{x,0} + \frac{b_1 + b_1b_2 - a_1b_2}{\beta_2}\sigma_u^2 \quad (7)$$

$$r_{x,2} = -a_1 r_{x,1} - a_2 r_{x,0} + b_2\sigma_u^2, r_{x,\tau} = -a_1 r_{x,\tau-1} - a_2 r_{x,\tau-2}$$

Note that the covariance function of a real $(1,1)$ARMA process and a 1^{st}-order AR process can be easily deduced from (6) by setting some ARMA parameters to 0.

When the moduli of the zeros are smaller than 1, the ARMA process is called minimum-phase. Starting from a non-minimum-phase ARMA process, the minimum-phase ARMA process can be deduced by replacing the zeros $\{z_l\}_{l=1,\dots,m\leq q}$ outside the unit-circle in the z-plane by $1/z_l^*$ for $l = 1, \dots, m$ to get the transfer function $H_{min}(z)$. Then, the variance of the driving process becomes:

$$\sigma_{u,min}^2 = \sigma_u^2 \prod_{l=1}^{q} K_l \text{ with } K_l = \begin{cases} |z_l|^2 \text{ for the replaced zeros} \\ 1 \text{ otherwise} \end{cases} \quad (8)$$

By using different methods such as the matrix determinant lemma, we can show that:

$$\lim_{k\to+\infty} \frac{|Q_{k+1}|}{|Q_k|} = \sigma_{u,min}^2 \quad (9)$$

Considering a minimum-phase ARMA process is useful to obtain the above limit of the ratio of the covariance matrices, but also to define the inverse filter associated with a minimum-phase ARMA filter. The latter defined by $\frac{1}{\sigma_u}H^{-1}(z)$, whose poles have their moduli smaller than 1, is necessarily stable.

Various approaches were proposed to estimate the ARMA parameter: the Yule-Walker equations are often used to get the AR parameters [17]. Inner-outer factorization can be considered to estimate the MA parameters. The prediction error method (PEM) is also often used to get the minimum-phase ARMA process [23]. This method is known to be asymptotically unbiased and efficient in the Gaussian case, but its computational cost may be high.

B. Our 1^{st} contribution: properties of ARMA processes after a multiscale procedure at scale s

If the above properties are well known, let us now present some properties of the ARMA processes after a multiscale procedure at the scale s.

<u>After the CG multiscale procedure with a factor s:</u>

Let x, y and w be respectively the original signal, the filtered one and the decimated filtered one at scale s. The determination of the orders of CG ARMA processes is not necessarily obvious with the exception of q^{th}-order MA processes, 1^{st}-order AR processes or ARMA processes of order $(1,q)$. Given (2), (4), (5) and (6), we can conclude that:

1/ Let x be a q^{th}-order MA process. y can be seen as the output of the filtering of a white noise, whose transfer function is $H_{CG}(z) \times \prod_{l=1}^{q}(1 - z_l z^{-1})$. So, it can be seen as a $q+s-1^{th}$-order MA process. Therefore, after decimation by a factor s, w is a MA process of order $\lfloor \frac{q+s-1}{s} \rfloor$.

2/ Let x be an ARMA process of orders $(1,q)$, with $q \geq 0$ and whose AR parameter is a_1. y corresponds to an ARMA process of order $(1, q+s-1)$, and whose AR parameter is still a_1. The proof is once again based on the expression of the transfer function $H_{CG}(z) \times \frac{\prod_{l=1}^{q}(1-z_l z^{-1})}{1-a_1 z^{-1}}$ of the equivalent filter. Due to (5), its correlation function for a lag $\tau \geq q+s$ hence satisfies:

$$r_{y,\tau} = -a_1 r_{y,\tau-1} \quad (10)$$

After the decimation by a factor s, the correlation function $r_{w,\tau}$ of the decimated filtered signal w is equal to $r_{y,s\tau}$. Therefore, the following recurrence relation between $r_{w,\tau}$ and $r_{w,\tau-1}$ is obtained for $\tau \geq \lceil \frac{q+s}{s} \rceil + 1$, with $\lceil . \rceil$ the ceil function:

$$r_{w,\tau} = (-a_1)^s r_{w,\tau-1} \quad (11)$$

This means that the $(1,q)$ARMA process after the CG procedure is an ARMA process of order $(1, \lceil \frac{q+s}{s} \rceil)$ whose AR parameter is equal to $-(-a_1)^s$.

<u>After the TS procedure with a factor s:</u>

1/ A q^{th}-order MA process after a direct decimation by a factor s becomes another MA process of order $\lfloor \frac{q}{s} \rfloor$. Thus, when $q = 1$ and for any s, one gets a white noise.

2/ Following the same above reasoning, an $(1,q)$ARMA process after a decimation by a factor s leads to an ARMA process of order $(1, \lceil \frac{q-1}{s} \rceil)$ whose AR parameter is equal to $-(-a_1)^s$.

III. Multiscale cross entropy rate

A. Expressions of the cross entropy, the entropy and the KL divergence and their rates in the Gaussian case

Given k consecutive samples of two stochastic processes characterized by their pdf, the cross entropy is defined as:

$$H_k^{(1,2)} = -\int p_1(x_{1:k}) \ln(p_2(x_{1:k})) dx_{1:k} = H_k^{(1)} + KL_k^{(1,2)} \quad (12)$$

where $H_k^{(1)} = -\int p_1(x_{1:k}) \ln p_1(x_{1:k}) dx_{1:k}$ is the Shannon entropy, $KL_k^{(1,2)} = \int p_1(x_{1:k}) \ln\left(\frac{p_1(x_{1:k})}{p_2(x_{1:k})}\right) \mathrm{d}x_{1:k}$ is the KL divergence and $x_{1:k}$ denotes the vector of k consecutive samples. It inherently accounts for both the intrinsic randomness of the first signal through its entropy and the mismatch between the two pdf, as measured by the KL divergence.

Let us now deduce their expressions for zero-mean Gaussian processes. To this end, given the pdf of the i^{th} real random Gaussian column vector $x_{1:k}$ of size k and covariance matrix $Q_{k,i}$ defined for $i = 1, 2$ by:

$$p_i(x_{1:k}) = \frac{1}{(\sqrt{2\pi})^k |Q_{k,i}|^{1/2}} \exp\big(-\frac{1}{2} x_{1:k}^T Q_{k,i}^{-1} x_{1:k}\big) \quad (13)$$

and after mathematical development similar to those presented in [8], the entropy, the KL and the cross entropy become:

$$\begin{cases} H_k^{(1)} & = \frac{k}{2}\ln(2\pi) + \frac{1}{2}\Big(\ln|Q_{k,1}| + \text{Tr}(Q_{k,1}^{-1}Q_{k,1})\Big) \\ & = \frac{k}{2}(1+\ln(2\pi)) + \frac{1}{2}\ln|Q_{k,1}| \\ KL_k^{(1,2)} & = \frac{1}{2}\Big(\text{Tr}(Q_{k,2}^{-1}Q_{k,1}) - k - \ln\frac{|Q_{k,1}|}{|Q_{k,2}|}\Big) \\ H_k^{(1,2)} & = \frac{1}{2}\Big(k\ln(2\pi) + \text{Tr}(Q_{k,2}^{-1}Q_{k,1}) + \ln|Q_{k,2}|\Big) \end{cases} \quad (14)$$

Let us now express the increment of the cross entropy, $\Delta H_k^{(1,2)} = H_{k+1}^{(1,2)} - H_k^{(1,2)}$ as well as the increments of the entropy and the KL. Then, let us look at their limits when k tends to infinity. Given (14), one has:

$$\begin{cases} \Delta H_k^{(1)} = \frac{1}{2}\big(1+\ln(2\pi) + \ln\frac{|Q_{k+1,1}|}{|Q_{k,1}|}\big) \\ \Delta KL_k^{(1,2)} = \frac{1}{2}\big(P_k^{(1,2)} - 1 - \ln\frac{|Q_{k+1,1}||Q_{k,2}|}{Q_{k,1}||Q_{k+1,2}|}\big) \\ \Delta H_k^{(1,2)} = \frac{1}{2}\big(\ln(2\pi) + P_k^{(1,2)} + \ln\frac{|Q_{k+1,2}|}{|Q_{k,2}|}\big) \end{cases} \quad (15)$$

with $P_k^{(1,2)} = \text{Tr}(Q_{k+1,2}^{-1}Q_{k+1,1}) - \text{Tr}(Q_{k,2}^{-1}Q_{k,1})$.
When k tends to infinity, by using (9), one gets the rates:

$$\begin{cases} \Delta H^{(1)} = \frac{1}{2}\big(1+\ln(2\pi)+\ln\sigma_{u,min,1}^2\big) \\ \Delta KL^{(1,2)} = \frac{1}{2}\big(P^{(1,2)} - 1 - \ln\frac{\sigma_{u,min,1}^2}{\sigma_{u,min,2}^2}\big) \\ \Delta H^{(1,2)} = \frac{1}{2}\big(\ln(2\pi) + P^{(1,2)} + \ln\sigma_{u,min,2}^2\big) \end{cases} \quad (16)$$

where $\sigma_{u,min,i}^2$ is the variance of the driving process of the minimum-phase ARMA process associated with the i^{th} process and $P^{(1,2)}$ is the power of the 1^{st} process filtered by the inverse filter associated with the 2^{nd} ARMA process. This is the limit of $P_k^{(1,2)}$ when k tends to infinity [8]. The cross entropy rate $\Delta H^{(1,2)}$ depends on $\sigma_{u,min,2}^2$ and the power $P^{(1,2)}$. The latter can be discriminative, but it may sometimes have small values with respect to the other terms or may become large if the 2^{nd} ARMA process has a zero whose modulus tends to 1 and that is not shared with the 1^{st} ARMA process.
Remark: Another expression could be considered based on Szegö theorem. Indeed, one has:

$$\lim_{k\to+\infty} \ln\frac{|Q_{k+1,i}|}{|Q_{k,i}|} = \frac{1}{2\pi}\int_{\pi}^{\pi} \log S_{i,\theta}d\theta = \sigma_{u,min,i}^2 \quad (17)$$

where $S_{i,\theta}$ denotes the power spectral density (PSD) of the i^{th} process that could be estimated using a periodogram [7].

B. Our 2nd contribution: multiscale cross entropy rate

Let us define the MCER (or multiscale profile) as a vector of the CERs computed on the original signals and the ones obtained by the multiscale procedure (CG or TS) up to the maximum scale τ_{max} selected by the practitioner. As recalled in [9], the phenomenon of overlapping spectra generated during the decimation step should not be *a priori* attractive as it introduces aliasing In our case, this will be a way to create diversity. Moreover, a zero on a unit circle (corresponding to a null PSD at the normalized angular frequency equal to the argument of the zero) should disappear at a certain scale due the spectrum overlapping, meaning that the resulting cross entropy rate would be finite.

C. Illustrations

$\Delta H^{(1,2)}(s)$ and $P^{(1,2)}(s)$ respectively denote the cross entropy rate and the power of the 1^{st} process filtered by the inverse filter associated with the 2^{nd} ARMA process at scale s. Note that $\Delta H^{(1,2)}(1) = \Delta H^{(1,2)}$ and $P^{(1,2)}(1) = P^{(1,2)}$.

1) Toy example: two white noises with variance $\sigma_{u,1}^2$ and $\sigma_{u,2}^2$: When using the TS multiscale procedure, as the noise variances remain unchanged and correspond to the minimum-phase case, one gets the following rates:

$$\begin{cases} \Delta H^{(1)}(s) = \frac{1}{2}\big(1+\ln(2\pi)+\ln\sigma_{u,1}^2\big) = \Delta H^{(1)} \\ \Delta KL^{(1,2)}(s) = \frac{1}{2}\big(\frac{\sigma_{u,1}^2}{\sigma_{u,2}^2} - 1 - \ln\frac{\sigma_{u,1}^2}{\sigma_{u,2}^2}\big) = \Delta KL^{(1,2)} \\ \Delta H^{(1,2)}(s) = \frac{1}{2}\big(\ln(2\pi) + \frac{\sigma_{u_1}^2}{\sigma_{u_2}^2} + \ln\sigma_{u,2}^2\big) = \Delta H^{(1,2)} \end{cases} \quad (18)$$

whereas the CG multiscale procedure leads to:

$$\begin{cases} \Delta H^{(1)}(s) = \frac{1}{2}\big(1+\ln(2\pi)+\ln\frac{\sigma_{u,1}^2}{s}\big) = \Delta H^{(1)} - \frac{\ln s}{2} \\ \Delta KL^{(1,2)}(s) = \frac{1}{2}\big(\frac{\sigma_{u_1}^2}{\sigma_{u_2}^2} - 1 - \ln\frac{\sigma_{u,1}^2}{\sigma_{u,2}^2}\big) = \Delta KL^{(1,2)} \\ \Delta H^{(1,2)}(s) = \frac{1}{2}\big(\ln(2\pi) + \frac{\sigma_{u,1}^2}{\sigma_{u,2}^2} + \ln\frac{\sigma_{u,2}^2}{s}\big) = \Delta H^{(1,2)} - \frac{\ln s}{2} \end{cases} \quad (19)$$

Therefore, the multiscale procedure does not bring in much.

2) Illustration with two 1^{st}-order AR processes: In this example, the AR processes are defined by their AR parameter $a_{1,i}$ and driving-process variance $\sigma_{u,i}^2$, with $i = 1, 2$. Given section II. B, let us first give some comments.
After the TS multiscale procedure with a factor s:
Both 1^{st}-order AR processes remain 1^{st}-order AR processes, but the AR parameters are now equal to $-(-a_{1,i})^s$, with $i = 1, 2$. The pole modulus decreases with s. When $a_{1,i}$ is positive, the decimated process exhibits a pick either in low or in a high frequency depending on s. The variance of the driving process can be easily deduced from (6) with $a_{2,i} = b_{1,i} = b_{2,i} = 0$. It is equal to $\frac{1-a_{1,i}^{2s}}{1-a_{1,i}^2}\sigma_{u,i}^2$. The latter increases with s.
Moreover, at each scale s, (6) can be used to get the power $P^{(1,2)}(s)$. It corresponds to the correlation function for a lag $\tau = 0$ when considering a (1,1)ARMA process defined by its AR parameter $-(-a_{1,1})^s$, its MA parameter $-(-a_{1,2})^s$ and its driving-process variance $\frac{1-a_{1,1}^{2s}}{1-a_{1,1}^2}\frac{1-a_{1,2}^2}{1-a_{1,2}^{2s}}\frac{\sigma_{u,1}^2}{\sigma_{u,2}^2}$.
Nevertheless, using (6), it can be shown that $P^{(1,2)}(s)$ and consequently the cross entropy rate associated with the set of AR processes defined by $a_{1,1}$ and $a_{1,2}$ and the cross entropy rate of the set of AR processes defined by $-a_{1,1}$ and $-a_{1,2}$ are the same for any scale s. The resulting MCER profiles are hence the same in that case.
After the CG multiscale procedure with a factor s:
Both 1^{st}-order AR processes become $(1, 1)$ARMA processes. At each scale s, the power $P^{(1,2)}(s)$ corresponds to the correlation function for a lag $\tau = 0$ when considering a (2,2)ARMA process. Its expression is given in (6) but requires the values of the ARMA parameters. Due to the low-pass filtering in the

CG, the issue in the TS multiscale procedure (where the cross-entropy rate was identical for the sets of AR processes defined by $a_{1,1}$ and $a_{1,2}$, and by $-a_{1,1}$ and $-a_{1,2}$) is now fixed.

Let us now provide 4 examples of the MCER profile with CG when $N = 15000$ samples of each AR process are available. $\sigma^2_{u,1} = \sigma^2_{u,2} = 100$. The PEM approach [23] is used to estimate the model parameters at each scale s. $P^{(1,2)}(s)$ is then computed by using (6) with the estimated parameters. It should be noted that at each scale s, $P^{(1,2)}(s)$ could be also estimated by first applying the inverse filtering defined with the estimated parameters obtained with the PEM method and then estimating the filter-output power. This can be the method used in practice for any process.

Each column in Table I defines the MCER averaged on 100 realizations of both AR processes. In theory, $\Delta H^{(1,2)}(1)$ should be equal to 6.8465 for cases 1 and 3, and equal to 5.9274 for cases 2 and 4. As shown in Table I, in practice, using the other scales makes it possible to differentiate the cases.

Scale s	Case 1 $a_{1,1} = 0.7$ $a_{1,2} = -0.8$	Case 2 $a_{1,1} = -0.8$ $a_{1,2} = 0.7$	Case 3 $a_{1,1} = -0.7$ $a_{1,2} = 0.8$	Case 4 $a_{1,1} = 0.8$ $a_{1,2} = -0.7$
1	$6.84(\pm 0.23)$	$5.94(\pm 0.09)$	$6.79(\pm 0.29)$	$5.90(\pm 0.21)$
2	$3.65(\pm 0.01)$	$6.46(\pm 0.12)$	$5.62(\pm 0.07)$	$3.49(\pm 0.01)$
3	$3.60(\pm 0.01)$	$10.34(\pm 0.36)$	$8.81(\pm 0.31)$	$3.53(\pm 0.01)$
4	$3.49(\pm 0.01)$	$11.40(\pm 0.44)$	$8.04(\pm 0.22)$	$3.34(\pm 0.01)$

TABLE I
AVERAGE VALUES OF THE CER, $\Delta H^{(1,2)}(s)$, AT EACH SCALE

The MCER based on a CG procedure is more relevant than the MCER based on time-shift multiscale procedure.

IV. CONCLUSIONS AND PERSPECTIVES

The multiscale cross-entropy rate is an alternative to the existing multiscale cross-entropies. Its applications to ARMA processes has the advantage of better understanding what the multiscale procedure can bring because analytical expressions can be derived. As a perspective, we plan to compare it with the other existing multiscale cross-entropies, address the non Gaussian case and use it with real biosignals.

REFERENCES

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

[2] C. Bandt and B. Shiha. Order patterns in time series. *Journal of time series analysis*, 28, (5):646–665, 2007.

[3] J. E. Contreras-Reyes and A. Brito. Refined cross-sample entropy based on freedman-diaconis rule: Application to foreign exchange time series. *Journal of Applied and Comput. Mechanics*, 8, (3):1005–1013, 2022.

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

[5] A. D'Avalos, M. Jabloun, P Ravier, and O. Buttelli. Theoretical study of multiscale permutation entropy on finite-length fractional Gaussian noise. *Proceedings of EUSIPCO*, 2018.

[6] A. D´avalos, M. Jabloun, P. Ravier, and O. Buttelli. Multiscale permutation entropy: Statistical characterization on autoregressive and moving average processes. *Proceedings of EUSIPCO*, 2019.

[7] A. Feutrill and M. A. Roughan. Review of Shannon and differential entropy rate estimation. *Entropy*, 23:1046, 2021.

[8] E. Grivel. Expressions and properties of the Shannon, Rényi, Sharma-Mittal and Tsallis cross-entropies for Gaussian ARMA processes. *Proceedings of EUSIPCO*, 2022.

[9] E. Grivel, B. Berthelot, G. Colin, P. Legrand, and V. Ibanez. Benefits of zero-phase or linear phase filters to design multiscale entropy: Theory and application. *Entropy*, 26, (4):332, 2024.

[10] E. Grivel, P. Legrand, and B. Berthelot. Multiscale entropy rate: a variant of the multiscale entropy. *Proceedings of EUSIPCO 2024*, 2024.

[11] J. He, P. Shang, and H. Xiong. Multidimensional scaling analysis of financial time series based on modified cross-sample entropy methods. *Physica A*, 500:210–221, 2018.

[12] J. Huang, J. Liu, H. Gong, and X. Deng. Multimodal loosening detection for threaded fasteners based on multiscale cross fuzzy entropy. *Mechanical Systems and Signal Processing*, 186:109834, 2023.

[13] A. Jamin, G. Duval, C. Annweiler, P. Abraham, and A. Humeau-Heurtier. A novel multiscale cross-entropy method applied to navigation data acquired with a bike simulator. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 733–736, 2019.

[14] A. Jamin and A. Humeau-Heurtier. (multiscale) cross-entropy methods: A review. *Entropy*, 22, (45), 2020.

[15] B Li, G. Han, S. Jiang, and Z. Yu. Composite multiscale partial cross-sample entropy analysis for quantifying intrinsic similarity of two time series affected by common external factors. *Entropy*, 22, (9):1003, 2020.

[16] G. Lin and A. Lin. Modified multiscale sample entropy and cross-sample entropy based on horizontal visibility graph. *Chaos, Solitons and Fractals*, 165:112802, 2022.

[17] M. Najim. *Modeling, estimation and optimal filtering in signal processing*. Wiley, 2010.

[18] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States*, 88, (6):2297–2301, 1991.

[19] T. Pincus, S. Mulligan, A. Iranmanesh, S. Gheorghiu, M. Godschalk, and J. D. Veldhuis. Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. *Proceedings of the National Academy of Sciences, USA*, 93:14100–14105, 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] R. M. Rostaghi, M. M. Khatibi, M. R. Ashory, and H. Azami. uzzy dispersion entropy: A nonlinear measure for signal analysis. *IEEE Transactions on Fuzzy Systems*, 30:3785–3796, 2021.

[22] D. Shang, P. Shang, and Z. Zhang. Efficient synchronization estimation for complex time series using refined cross-sample entropy measure. *Communications in Nonlinear Science and Numerical Simulation*, 94:105556, 2021.

[23] T. Soderstrom and P. Stoica. *System Identification*. Prentice Hall, 1989.

[24] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin. A complete ensemble empirical mode decomposition with adaptive noise. *Proceedings of ICASSP*, pages 4144–4147, 2011.

[25] 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.

[26] Y. Wu, P. Shang, and Y. Li. Multiscale sample entropy and cross-sample entropy based on symbolic representation and similarity of stock markets. *Communications in Nonlinear Science and Numerical Simulation*, 56:49–61, 2018.

[27] H. B. Xie, Y.P. Zheng, J. Y. Guo, and X. Chen. Cross-fuzzy entropy: A new method to test pattern synchrony of bivariate time series. *Information sciences*, 180:1715–1724, 2010.

[28] Y. Yin and P. Shang. Comparison of multiscale methods in the stock markets for detrended cross-correlation analysis and cross-sample entropy. *Fluctuation and Noise Letters*, 13, (3):1450023, 2014.

[29] Y. Yin and P. Shang. Modified cross sample entropy and surrogate data analysis method for financial time series. *Physica A*, 433:17–25, 2015.

[30] Y. Yin, P. Shang, A. C. Ahn, and C.-K. Peng. Multiscale joint permutation entropy for complex time series. *Physica A*, 515:388–402, 2019.

[31] Y. Yin, P. Shang, and G. Feng. Modified multiscale cross-sample entropy for complex time series. *Applied Mathematics and Computation*, 289:98–110, 2016.

[32] Y. Yin, X. Wang, W. Wang, Q. Li, and P. Shang. Multiscale cross-sample entropy based on visibility graph for quantifying time series irreversibility. *Communications in Nonlinear Science and Numerical Simulation*, 124:107308, 2023.

[33] T. Škorić, O. Mohamoud, B. Milovanovic, N. Japundzic-Zigon, and D. Bajic. Binarized cross-approximate entropy in crowdsensing environment. *Computers in Biology and Medicine*, 80:137–147, 2017.