POSTPRINT VERSION. The final version is published here :
Caza-Szoka, M. et Massicotte, D. (2022). Windowing Compensation in Fourier Based Surrogate Analysis and Application to EEG Signal Classification. IEEE Transactions on Instrumentation and Measurement (Vol. 71, p. 1-11).
doi: 10.1109/TIM.2022.3149325

Windowing Compensation in Fourier Based Surrogate Analysis and Application to EEG Signal Classification

Manouane Caza-Szoka and Daniel Massicotte, *Senior Member*, *IEEE*

***Abstract*— This paper shows how adding a second step of windowing after each phase randomization can reduce the False Rejection Rate in Fourier based Surrogate Analysis. Windowing techniques reduce the discontinuities at the boundaries of the periodically extended data sequence in Fourier Series. However, they add a time domain non-stationarity which affects the Surrogate Analysis. This effect is particularly problematic for short low-pass signals. Applying the same window to the surrogate data allows having the same non-stationarity. The method is tested on order 1 autoregressive process null hypothesis by Monte-Carlo simulations. Previous methods were not able to yield good performances for left-sided and right-sided tests at the same time, even less with bilateral tests. It is shown that the new method is conservative for unilateral tests as well as bilateral tests. In order to show that the proposed windowing method can be useful in real context, in this extended paper, it was applied for an EEG diagnostic problem. A dataset comprising the EEG measurements of 15 subject distributed in three groups: attention-deficit disorder primarily hyperactive-impulsive (ADHD), attention-deficit disorder primarily inattentive (ADD); and anxiety with attentional fragility (ANX) was used. Both statistical and machine learning (Naïve Bayesian) approaches were considered. The Mean Short Windowed SA (MSWSA) was used as a signal feature and its performances was studied with respect to the windowing systems. The main findings were that (i) the MSWSA feature has less variability for ADD than for ADHD or ANX, (ii) the proposed windowing method reduces bias and non-normality of the SA feature, (iii) with the proposed method and a naïve Bayesian classifier, a 93% success rate of discriminating ADD from ADHD and ANX was achieved with leave-one-out cross-validation, and (iv) the new feature could not have yielded interesting results without the proposed windowing system.**

***Index Terms*—ADD, ADHD, EEG, Fractal Dimension, Nonlinear Analysis, Nonlinear Dynamics, Surrogate Data, Windowing Techniques.**

I. Introduction

Surrogate Analysis (SA), a method initially developed for the testing of nonlinearity [1], has been recently applied for signal classification [2] on long time series. This paper shows how the mean SA over small windows can also be used for classification on Electroencephalogram (EEG) signals [3].

This paper is an extended version of [4]. In [4], the windowing compensation technique useful for short time series was described and analyzed on synthetic data representing a null hypothesis. In this extended paper, the proposed compensation is applied to a real EEG dataset. Often, features and algorithms which are interesting for the analysis of one type of biosignal can have applications to other types [5]. Also, different biosignals frequently have relations between them [6]. Hence, although applied on a specific problem, the proposed feature could have applications on multiple biosignal measurements such as EEG [7, 8], surface Electromyography (EMG) [9, 10], or Electrocardiography (ECG) [11, 12].

The SA is a hypothesis test aimed at assessing the nonlinear nature of a signal [1]. It has been applied to a wide variety of domains [13], notably in the study of brain activity [14, 15]. More recently, it has been used as a feature for Low-Back-Pain diagnostic with electromyogram (EMG) sensors [2] and to study the non-randomness of the phase spectrum in the Fourier domain [11]. It has also been extended for distinguishing between non-linear and non-stationary data [16].

Multiple versions of the SA have been developed but can generally be categorized in two groups: Fourier based and auto-regressive moving-average process (ARMA) based. Both approaches were well described [17]. By far, the Fourier method has been the most popular, including its derivative such as the amplitude adjusted Fourier transform (AAFT) and its iterative version [18]. It is a preferred approach because of its simplicity since it does not require any model selection step. However, the Fourier based SA is well known for its sensitivity to signals artifacts. An important example of such artefacts is the impact of limited amount of data. One solution to this artefact is the method of matching ends. It was analysed in [19] that matching the ends gave reasonable performances for reducing the effect of the "periodicity mismatch" (difference between the first and last data point) when the number of data points available is high. Another solution is to apply windowing techniques [20] which have

Submitted the September 14th, 2021. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2015-03674.

M. Caza-Szoka and D. Massicotte are with the Université du Québec à Trois-Rivières, Department of Electrical and Computer Engineering, Laboratoire des signaux et systèmes intégrés, Chaire de recherche sur les signaux et l'intelligence des systèmes haute performance, 3351, Boul. des Forges, Trois Rivières, Québec, Canada, G9A 5H7.
(e-mail: manouane.caza-szoka@uqtr.ca and daniel.massicotte@uqtr.ca)

the advantage of not requiring a variable data length or initial point selection. This makes it easier to implement or to compare results between different time series, especially when the data length is small. Also, non-linear methods can be biased by the data length [21]. The windowing techniques effect can be described in the frequency domain as well as in the time domain. When analyzed in the frequency domain, the main interest of the windowing methods is to allow a trade-off between the main lobe's size and the sidelobes' amplitude, i.e. between resolution and dynamic range. With respect to the SA, the frequency domain effect of the windowing has simply been described as "additional bias" [18]. From a time domain perspective, they reduce the discontinuities at the boundaries of the periodically extended data sequence in Fourier Series, i.e. the periodicity mismatch. In [1, 22], the windowing is applied in the generation of the surrogate data. However, these surrogates are compared to the *unwindowed* original series. This leads to time domain non-stationarity of the surrogate signal which is not present in the original signal. This non-stationarity can produce a bias on the nonlinear feature of the surrogate series. The results showed the trade-off between reducing the periodicity mismatch (called "sampling gap in [22]) and avoiding the spectral leakage. Indeed, it indicated that using the windowing techniques is useful up to a certain data length. Above, the windowing worsens the bias. Moreover, even in the best cases, it remained largely optimistic, yielding between 10 and 25% of False Rejection Rate (FRR) when a 5% rate would be expected. Nonetheless, the application of windowing techniques is important: without it, the FRR were shown to be over 30%.

A different approach would be to compare the surrogate to the windowed version of the original signal. However, the windowing process adds time domain non-stationarity. The windowing of a stationary ARMA process makes its variance changing from sample to sample. It has been shown that the SA may be very sensitive to non-stationarity [16, 23]. This effect has even been exploited notably in [24]. The non-stationarity caused by the windowing in the original data will not be present in the surrogate series. Clearly, the windowing adds a bias. The biases in SA make either the left or right sided test over-optimistic. Also, it makes the bilateral tests always over-optimistic. The bilateral tests are important when the type of nonlinearity that might be present is unknown.

Until now, windowing methods for short time series has been evaluated on synthetic data representing a null hypothesis and has yet to be applied on real measurements.

This paper presents a method for compensating the non-stationarity caused by the windowing in SA. The method consists of keeping the window on the original series and applying the same window on every surrogate data series. This adds the same non-stationarity to the surrogate series as in the original series. The analysis is conducted by Monte-Carlo simulations on an order 1 autoregressive process (AR(1)) as in [22], but considering the unilateral and bilateral tests. Although the added windowing replaces the periodicity mismatch artifact by a frequency smoothing artifact, it will be shown that the overall effect is conservative.

A feature including the described windowing scheme was developed and tested on real data. The dataset was composed EEG measurements of three groups of five subjects (15 in total). The classes were (i) attention-deficit disorder primarily hyperactive-impulsive (ADHD), (ii) attention-deficit disorder primarily inattentive (ADD), and (iii) anxiety with attentional fragility (ANX). Let's note that no control group was present.

EEG signals have been an interesting kind of signal for the study of nonlinear signal processing. For the SA, they have been an application example from the first proposition of the method in [1]. Arguably, the EEG signals are where the SA has been used the most [25-31]. EEG signals are therefore relevant signals to test the proposed windowing method.

To our knowledge, there are no studies for EEG classification that attempts to address these three conditions together. However, the ADHD vs control classification has an extensive literature. In [3], it is reported that the state-of-the-art methods can achieve between 90 and 98% of classification success. Although some of the studies reported made use of data augmentation methods [3] or multiple trials [32] which can overestimate the performances, there were clearly achieving a clinically useful level of classification. Even if near perfect classification is attained, it is for relatively simple classification problems compared to, let's say, automatically detect any condition reported in the DSM-5 [33]. There is therefore still a lot of room for research of interesting EEG features.

Although linear methods have been used to classify EEG signals, the best results are obtained by nonlinear analysis [34]. However, it is well known that linear aspects of the signal have strong or even dominant effect on nonlinear features. It is with this fact in mind that SA was developed. Since the publication of the SA method, continuous research has been made toward understanding the nonlinear aspect of EEG [35-38]. However, the SA was only used in order to justify the use of nonlinear feature (e.g. [39]) or interpreting the results (e.g. [37]). In the study of depression, [39] used both linear and nonlinear features, having better results with the nonlinear ones. SA scores based on FD and the Largest Lyapunov Exponent were obtained separately for a patient group and a normal group. However, the analysis ended with the confirmation that the signals were nonlinear. In [40], the Sample Entropy (SampEn), a nonlinear feature, along with its SA counterpart was tested for EEG based Alzheimer's Disease (AD) detection. The SA score used was equivalent to the absolute value of the score considered in this paper. The results showed that there was a clear difference between the Sampled Entropy of the AD group and the control group, but there was no difference found for the SA score. However, only long time series were used. The SA with Katz' FD has been successfully applied in [2] in the context of Low-Back-Pain diagnostic with surface Electromyography (EMG) sensor array. It applied the SA on very large data series. In summary, there is no use of the SA as a signal feature in EEG literature and there is no short window SA application on real data.

The experimental validation of the proposed method is carried with the analysis of an EEG dataset. A new feature is proposed: The Mean Short-Windowed SA (MSWSA) based on Higuchi's FD. Since the dataset is small, Naïve Bayesian classifier will be used to reduce the number of meta-

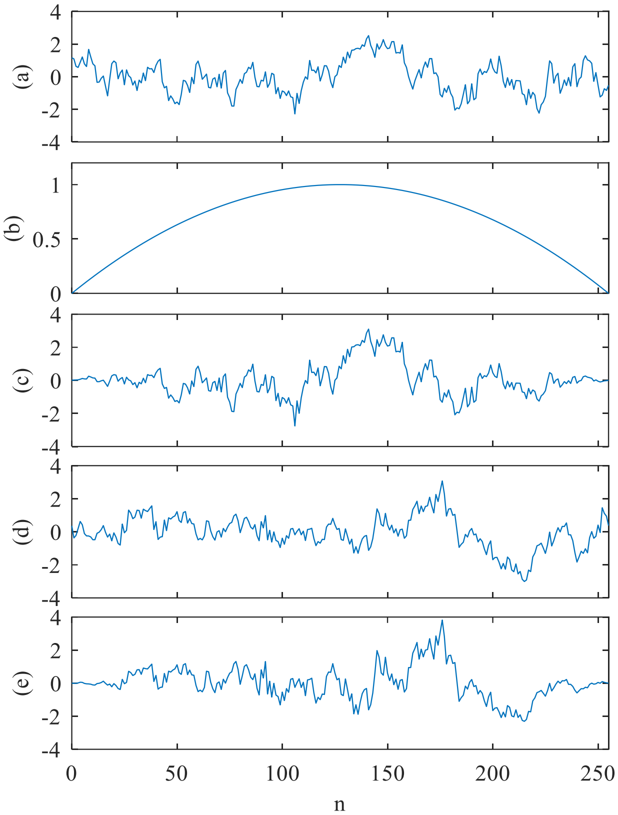


Fig. 1. Time domain representation of an example of 256 data points of an AR(1) signal with α_1 = 0.9 and its processed versions by the proposed system. The signal variables are given with respect to the proposed system shown in Fig. 2 (d). The original signal x (a), the Welch window (b), the original windowed signal x_w (c), a surrogate of the windowed signal s (d) and the windowed surrogate signal s_w (e) are presented.

parameters. The new feature will be examined from both a statistical and a machine learning point-of-view.

The main findings were that (i) the MSWSA feature has less variability for ADD than for ADHD or ANX, (ii) the proposed windowing method reduces bias and non-normality of the SA feature, (iii) with the proposed method and a naïve Bayesian classifier, a 93% success rate of discriminating ADD from ADHD and ANX was achieved with leave-one-out cross-validation, and (iv) the new feature could not have yielded interesting results without the proposed windowing system.

The paper is organized as follows: Section 2 describes the basic numerical methods testing methodology. Section 3 reviews the classical windowing method for the SA found in the literature and presents the proposed approach. Section 4 describes the EEG experimental data and the presentation of the new MSWSA feature. The results of the simulations are reported in Section 5 while those of the EEG experimentation are found in Section 6. The discussion of both the numerical simulation and the EEG experimentation are regrouped in Section 7. Finally, the conclusions are drawn in Section 8.

II. Numerical Methods

This section describes the conditions and results of the numerical experiments done by Monte-Carlo simulations.

2.1. Nonlinear Method: Higuchi's Fractal Dimension

The nonlinear method used in this paper is the FD calculated by the Higuchi's method [41]. A similar use of FD in the context of SA was notably used in [42] for magnetoencephalography (MEG) signals. The FD is easily calculated, even with a low number of data points. The Higuchi's method for obtaining the FD is based on the calculation of the signal's absolute length L with a subsampling factor k:

$$L_m(k) = \sum_{i=1}^{\frac{N}{k}-1} \left| x(ik+m) - x\big((i-1)\cdot k + m\big)\right| / k \quad (1)$$

These is averaged for m, the starting point:

$$L(k) = \sum_{m=1}^{k} L_m(k)\,/k \quad (2)$$

The Higuchi's FD is the slope of the logarithm of L with respect to the logarithm of k.

The method necessitates the selection of the time intervals (k). For speed, we only used k from 1 to 5.

2.2. Window Method

In [22], different windows were tested. In this paper different systems are tested. Hence, the best window proposed in [22], the Welch Window [43], is used for the numerical experiment. (Another window is used for the experimental section). The Welch Window is a parabola centered at N/2, as shown in Fig. 1 (b):

$$w_{Welch}(n) = 1 - \left(\frac{2n-(N+1)}{N-1}\right)^2 \quad (3)$$

The $w(n)$ are the window's tap coefficients.

2.3. Test Signal

The AR(1) process is used for the tests. It follows the relation:

$$x(n) = \alpha_1 x(n-1) + e(n) \quad (4)$$

where x is the signal, α_1 is the process parameter controlling the cut-off frequency and e is an independent identically distributed random variable called process noise. In this paper, the noise has a normal distribution. This process was used in [22] with α_1 = 0.995. This choice was appropriate to highlight the windowing effect. The periodicity mismatch in FFT is much more obvious in low pass signals. We used the same α_1 in most of this paper analysis. However, we also show results with α_1 =0.9, which clearly reduce the periodicity mismatch problem and give an advantage to the System 1, with no windowing. An AR(1) process with α_1 =0.9 is shown in Fig. 2 (d). As in [22], the first data point were removed in order to attain the steady state regime. The transient regime can be seen as non-stationary, which can affect the SA. We used an extra 2000 data points in order to have a relative impact of 5x10^{-5} in the case where α_1 = 0.995.

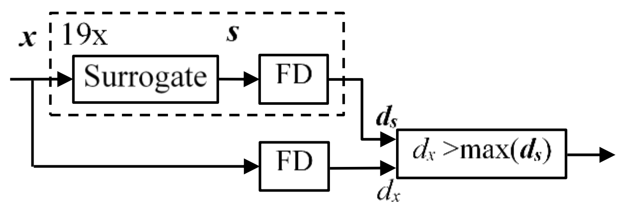
2.4. Performance Measure: The False Rejection Rate

The FRR is used to compare the performances of the different systems. It is often called "Type 1 error." A perfectly fair, or balanced test should give a FRR of exactly

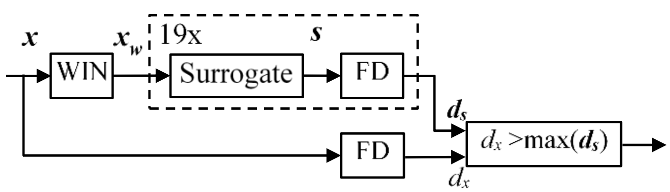
5%. When the FRR is lower, the test is considered conservative while if it is higher, the test is said to be optimistic.

III. Surrogate Analysis and Windowing

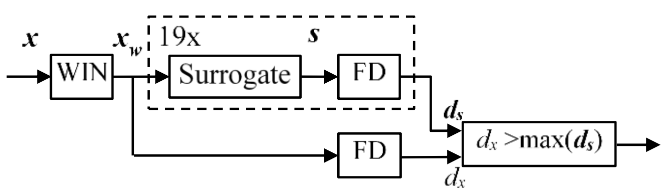
The SA compares a nonlinear feature of a signal to the distribution of the same feature obtained on random signals with identical power spectrum. To do this, surrogate signals respecting the null hypothesis with the same power spectrum as the originals signals must be generated. The most common approach, which is considered here, is the phase randomization in the frequency domain. First, the Fast Fourier Transform (FFT) is applied. The phase is set to a uniform distribution between 0 and 2π, with conjugate symmetry in order to preserve real values. Then, the signal is set back in the time domain by inverse Fourier transform (IFFT). The number of surrogate series that must be generated depends on the desired FRR [19]. For a unilateral test with an aimed 5% FRR, 19 surrogate series are generated. As previously stated, in this paper, the FD is the nonlinear feature considered. When the SA is computed with the FD as its nonlinear feature, the FD is calculated for the original series and every surrogate series. The final stage is a comparison of the FD of the original series to those of the surrogates. A positive result to the test is given if the original series FD is higher than the surrogate series FD when the test is right sided and lower when the test is left sided. For a bilateral test again with an aimed 5% FRR, the number of surrogate series must be increased to 39 and the test will be positive if the original FD is higher of lower than all the surrogate series FD.



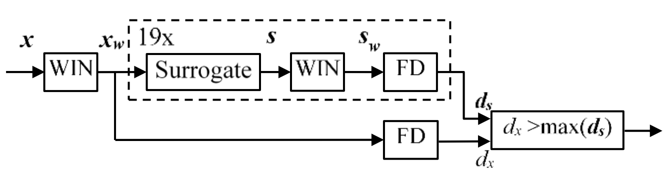
System 1: No Windowing
(a)



System 2: Windowing Applied as the Method of Suzuki
(b)



System 3: Same Spectrum
(c)



System 4: Proposed System
(d)

Fig. 2. Windowing in the surrogate analysis system for right-sided tests. The system with no windowing is presented in (a). The method used in [12] is shown in (b), with the only difference being that the FD is used instead of Local Linear Prediction. The system in (c) is a variation of the previous one in which the FD of the original series is obtained from the windowed original series.

The null hypothesis is that the signal can be produced by a stationary linearly filtered white Gaussian noise. If the data processing produces some artifacts that are not of an ARMA type, these may influence the results.

When windowing is applied on a signal, two effects must be considered:

1) The spectral leakage, explained by the convolution theorem,

2) A non-stationarity in the sample-to-sample variance of the resulting signal.

The non-stationarity is the result of the time-domain multiplication between the original signal, which can be stationary, and the fixed window. The resulting signal will have a variance which will vary from sample-to-sample, with a variance scaling factor of $w^2(n)$.

These two effects can affect the nonlinear feature of the signal. It is important to note that non-stationarity is removed by phase randomization. If some windowing artifacts are present in the original signal but not in the surrogate (or the reverse), a bias is created.

The four systems presented in Fig. 2 show different approaches for applying windowing in SA. The systems are shown for right-sided tests with FRR set at 0.05 (if the test was unbiased). The rest of the present section describes the first three systems used as a basis of comparison and the proposed system.

3.1. System 1: Windowless Surrogate Analysis

The basic SA without windowing is presented in Fig. 2 (a). It is the system mostly used throughout the literature proposed in [1]. It can be interpreted as using a rectangular window. For a data series that respects the null hypothesis, the differences between the original signal and its surrogates include both the spectral leakage and the periodicity mismatch.

3.2. System 2: Method of Suzuki [22]

In Fig. 2 (b), the windowing in applied before the FFT. However, the FD of the original series is calculated on the unwindowed version. Although the periodicity mismatch effect is strongly reduced, the power spectrum differs between the original and surrogate series. Moreover, this difference is identical throughout the surrogate series. On the other hand, all the signals are stationary. The method was possibly first briefly mentioned in the original SA paper [1] but was really analyzed in [22].

3.3. System 3: Same Spectrum

A final reference method is given in Fig. 2 (c). The system is similar to the method in (b), but the FD of the original series is calculated on the windowed version. Therefore, the original and surrogate series have the same power spectrum.

However, while the surrogate series are stationary, the windowed original signal is not. This method has not been found clearly explained in the literature. Although System 3 is not part of the literature, it is included in the numerical tests for the sake of completion, because it shows a situation where the only artifact is the non-stationarity.

3.4. System 4: Proposed Windowing Method for Surrogate Analysis

The proposed system is reported in Fig. 2 (d). The difference with (c) is that a second windowing stage is added to the surrogate series before calculating the FD. The three main aspects of the original and surrogate series are:

1) The spectral leakage is different between the surrogates,

2) The spectral leakage is different between the surrogates and the windowed original series,

3) The non-stationarities are similar.

The stationarity aspect ensures that the bias in the nonlinear method caused by the window are similar. The spectral leakage between the surrogates adds variability in the surrogate series power spectrum. For this reason, it is expected that the test will be more conservative. Examples of the different signals obtained in the System 4 are shown in Fig. 1 (a) (c-e) along with the window used (b). The windowed signal is non-stationary.

IV. Experimental Data and Feature Extraction

In this section, the experimental methods are explained. The description of the measurements and dataset are given, as well as the feature extraction method.

1) EEG Measurements and Dataset

A dataset comprising the EEG measurements of 15 subjects distributed equally in three groups: attention-deficit disorder primarily hyperactive-impulsive (ADHD), attention-deficit disorder primarily inattentive (ADD); and anxiety with attentional fragility (ANX) was used.

In each session, about 8 min of EEG data (International 10-20 system, 19 channels) was collected, with 5 minutes while resting state with the eyes open, followed by 3 minutes resting with the eyes closed. The EEG measurement system had an analog bandpass filter from 0.3 Hz to 70 Hz (4th order Butterworth zero-phase band-pass filter) and a sampling rate of 250 Hz. An Anterior-Posterior Bipolar Montage was obtained by offline calculation, leading to 18 channels. For each channel, 20000 data points with closed eyes were kept.

2) Preprocessing

The preprocessing stage was kept to a minimum in order to leave the nonlinearity as much as possible unaffected. Only a simple notch filter was applied in the frequency domain at 60 and 120 Hz. The use of the frequency domain notch filter allows to keep the phase spectrum intact, an important aspect in nonlinear analysis. There was no separation between the different brain waves.

3) Windowing

The window method used was the Hann window [44], also called Raised-Cosine. It is simply composed of a cosine signal raised to have only positive numbers.

$$w_{Hann}(n) = \frac{1}{2}\left(1 - \cos\left(2\pi \frac{n}{N}\right)\right) \tag{5}$$

The choice of this window was made to eliminate as much as possible periodicity mismatch, including at the multiple derivative levels. The length of the windows was 30 data points (equivalent to 120 ms), with overlaps of 50%, applied over 80 seconds of data. This is applied to every EEG channel. For comparison, the data were also treated with a single window.

4) SA Score

A score-based representation of the SA is used to produce continuous features. The score Z is obtained as follows:

$$Z = \frac{D - \overline{D_S}}{\mathrm{std}(D_S)} \tag{6}$$

where D is the FD of the original series, D_S is the FD of the surrogate series, and std is the standard deviation. It must be emphasized that a similar score was proposed in [1] but included an absolute value. In [2],the score without the absolute value was proposed. While for the nonlinearity detection problem in the simulations, as said earlier, only 19 surrogate series are required for the unilateral test and 39 for the bilateral test for a 95% confidence interval, much more are needed for this continuous version of the SA. Both the mean and standard deviation of the surrogate data series nonlinear feature must be estimated reasonably precisely. For this soft version of the SA, 200 surrogate series were used.

5) Proposed Mean Short-Windowed SA Feature

The MSWSA is proposed as the estimated mean SA score Z evaluated on all the data windows:

$$MSWSA = \frac{1}{N_W} \sum_{n_w} Z_w \tag{7}$$

Here, N_W is the number of data windows and Z_w is the SA score of the w-th data window. The MSWSA is proposed as a signal feature, useful for classification purpose. The rational behind the MSWSA has two main components:

- Using short data windows makes the score principally influenced by short duration phenomenon.
- Using many of these short windows is useful to reduce the variability of the score.

The interesting aspect of the MSWSA with respect to EEG is that it could target phenomena of durations of the same order of magnitude as short Event-Related Potentials (ERP) [45-47]. Hence, it is expected that there would be a certain match between the dynamic of the EEG and the MSWSA.

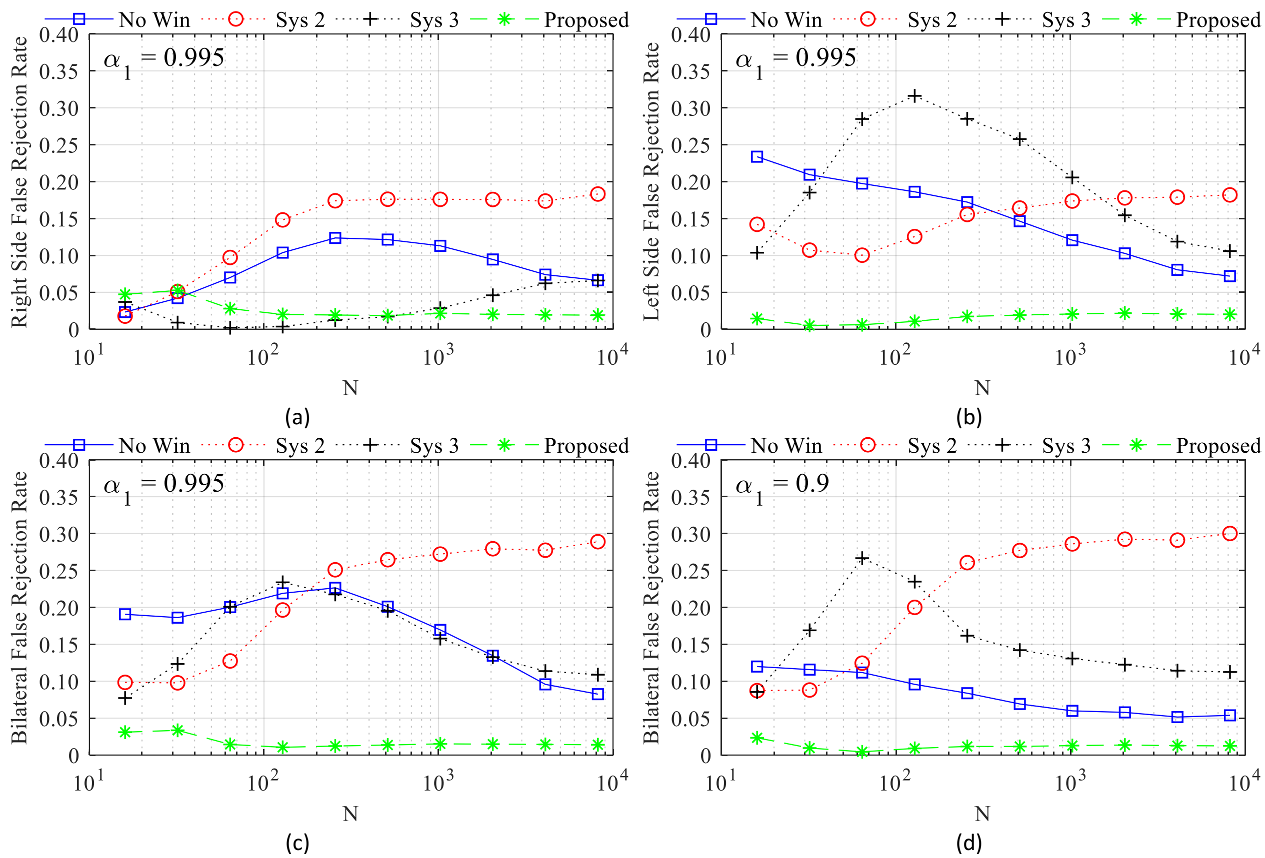


Fig. 3. False Rejection Rates vs the number of data points (window size) of all systems for an AR(1) process with α_1 = 0.995 for (a-c). Right-sided (a), left-sided (b) and bilateral (c) tests are shown. In (d), the process has α_1 = 0.9 and the test is bilateral. The number of tests was 20000 per points. The x axis title *N* stands for number of data point in a window.

V. Simulation Results

Monte-Carlo simulations were conducted to obtain the FRR of the different systems. The number of tests per point in these graphs was 20000. All the simulations were made for windows of size *N* ranging from 16 to 8192. The same signals were tested for every method. This gives a maximum error of 1% at 3 standard deviations when the FRR is of 50% and 0.2% when FRR is at 5%. In the first simulation, α_1 was set to 0.995.

The results of the theoretical Monte-Carlo simulations are shown in Fig. 3 (a-c). The right-sided (a), left-sided (b) and bilateral tests results are shown. In (d), the results for the bilateral test are given for α_1 = 0.9.

A bias in the FD of the surrogate will reduce the FRR of one of the one-sided tests while increasing the FRR of the other. For example, the System 3 shows a very low FRR for the right-sided test in Fig. 3 (a) but very high for the left-sided test. The effect of the bias for bilateral test is harder to predict. When a normal distribution is considered, the bias simply increases the FRR. However, when the distribution is skewed, the FRR can be either lowered or raised (or in some particular cases stayed unaffected.) Adding some Kurtosis effects, it becomes necessary to simply rely on simulations to assess the impact of the different biases. In Fig. 3 (a-c), the System 2 bilateral test has some FRRs higher than for both one-sided tests while the System 3 does not. The relation between the standard deviation of the offsets and standard deviation of the surrogates FDs will impact both one-sided tests, as well as the bilateral test.

The results show that the proposed system is generally very conservative. Since the right-sided tests FRR is higher than for the left-sided tests, there is still a weak bias. The bilateral test FRR is almost always at 1%. The proposed system is the only one which can give conservative results in bilateral tests.

As expected, when α_1 = 0.9 (d), the impact of not using windowing (System 1) is much weaker. Also, as the number of data points increase, the windowing problem disappear. The number of data points necessary to remove the need of windowing depends on the bandpass of the signal, controlled by α1 in this case. The System 2 has the same kind of behaviour with FD that it had with Local Linear Prediction as used in [22]. It has better performances than System 1 (no window) when the number of data points is small, but worse when the number is higher. Although the System 3 performs better than System 2 when the number of data points is high, it almost never performs better than the System 1 and System 2 simultaneously.

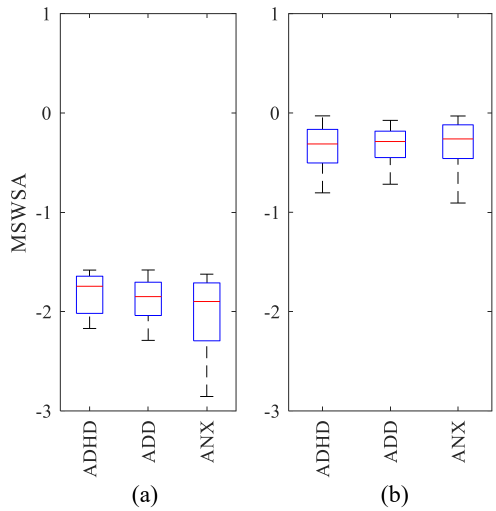


Fig. 4. Boxplot of the MSWSA of **all channels together** separated by classes. System 3 and System 4 results are shown in (a) and (b) respectively.

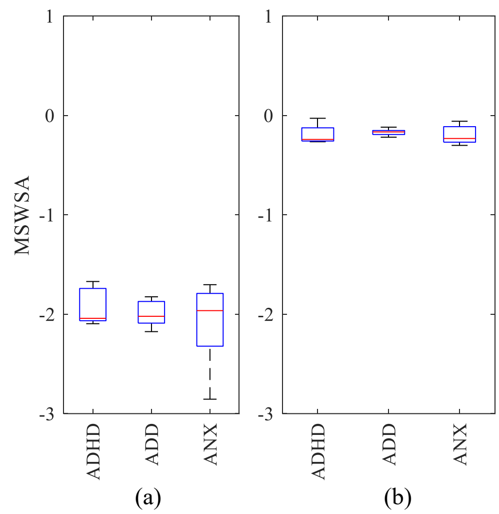


Fig. 5. Example of the boxplot of the MSWSA for a **single channel** (Fp1-F7) separated by classes. System 3 and System 4 results are shown in (a) and (b) respectively.

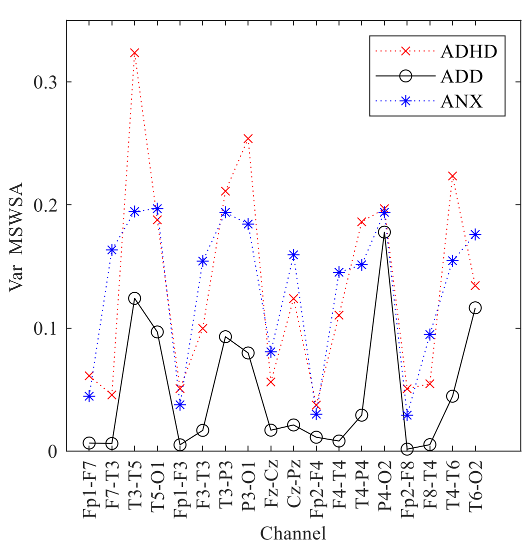


Fig. 6. Variance of the MSWSA for each channel and for the three groups. The System 4 windowing method is used.

TABLE I
PROPORTION OF POSITIVE NORMALITY TEST RESULTS

N Groups	3	3	2	2
Win System	3	4	3	4
KS	0	0	0	0
SW	0.5	0.06	0.61	0.06

TABLE II
NUMBER OF CHANNELS WITH POSITIVE VARIANCE TEST RESULTS AND P-VALUES

System	3	3	4	4
	N Chan	p	N Chan	p
Bartlett	5	0.00155	5	0.00155
Levene Quadratic	0	1.00000	5	0.00155
Levene Absolute	3	0.05813	7	0.00002
Brown Forsythe	2	0.22648	5	0.00155
O'Brien	0	1.00000	5	0.00155
F-Test	5	0.00155	4	0.01087

VI. EXPERIMENTAL RESULTS

The results obtained from the EEG dataset are shown in this section. Both statistical and machine learning approaches were considered.

A. *Statistical Analysis Results*

The statistical analysis of the MSWSA features is caried to verify if there are statistical differences between the different groups and to observe possible biases intervening in the different windowing systems. The statistical analysis is reported for the windowing System 3 and System 4.

A first analysis is done by plotting in Fig. 4 the MSWSA of all channels together, separated by classes. Each boxplot is obtained from 5 subjects with 18 channels, hence 90 MSWSA values. System 3 and System 4 results are shown in (a) and (b) respectively. The main difference is clearly an offset difference of the MSWSA. However, this offset difference is similar for all classes. Also, there are no obvious observable variability variation with the methods.

An example of the boxplot of the MSWSA for a single channel (Fp1-F7) is given in Fig. 5. Again, there is an offset gap. Also, there is reduced variability for the ADD group comparatively to ADHD and ANX groups. The effect was not found to be statistically significant. A similar conclusion could be made on the other channels.

In Fig. 6, the variance of the MSWSA for each channel and for the three groups are presented. The unbiased estimation of the variance was used. It can be seen that the variance of the ADD group is always smaller than that of the ADHD and ANX groups. The lines representation is used not to indicate any particular relation between the points but rather to highlight the fact that the ADD group is different for the variability, i.e., its line is always lower. Such clear results were not found for the mean MSWSA. Also, for System 2 and System 3, the results were less obvious.

With the caveat that the individual channels are not independent and that the hypothesis has been made *à*

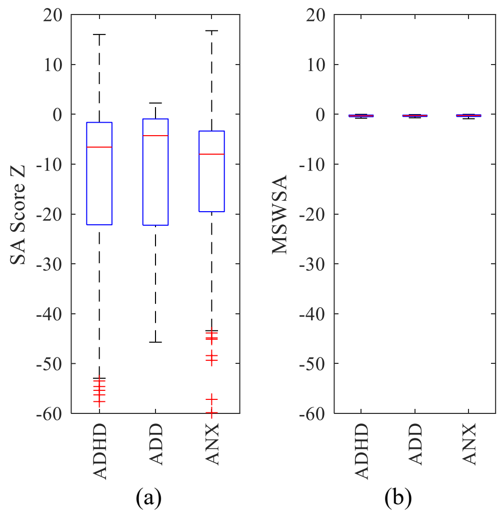


Fig. 7. Comparison of the distribution of the SA score Z for a single long window (a) and the MSWSA (b). The windowing is done by System 4 in both situations. All channels are included. Although an offset is observed, the main difference is the variability of the SA score Z for a single long window which is 50 times greater than the MSWSA.

posteriori, considering that the probability of having randomly 18 points all lower than two other series is 2.5×10^{-9} is still convincing that the results would not likely have occurred by chance only.

It is desired to confirm the variability relation by different method. Hence, hypothesis testing on the difference of variance was used. Normality was first verified. Non-normal data can produce spurious rejection of the null hypothesis.

The normality of the MSWA score was assessed. Results are shown in Table I for System 3 and System 4. Both Kolmogorov-Smirnov (KS) and Shapiro-Wilks (SW) Normality tests were used with $\alpha=0.05$. The tests are applied for every channel and the proportion of positive results are given. The KS test is never sensitive enough to detect non-normality. However, the proportion of channels which are detected as non-normal is much lower for System 4 than System 3. Similar results are obtained whether the ADHD and ANX are regrouped (2 groups) or not (3 groups). The null hypothesis was much more rejected for System 3 than System 4. Hence, variance variability results of System 3 will be more biased. This does not change the conclusion drawn.

The statistical significance of the difference of variance of the MSWA was tested on the 2 classes problem. Six tests for the difference of variance have been used: Bartlett, Levene Quadratic and Absolute, Brown Forsythe, O'Brien and the F-test. The number of channels where the result was positive are reported in Table II for each test. The p-values associated were obtained by considering the probability of obtaining the number of channels with positive test or higher randomly given a probability of 0.05. Similar results are obtained when 3 groups are considered, but the F-test is not applicable in this case.

TABLE III
CROSS-VALIDATION CLASSIFICATION RESULTS

Total Data	Window Length	Number Windows	Sys 2	Sys 3	Sys 4
20000	20000	1	53%	60%	53%
19995	30	1332	60%	40%	**93%**

B. *Classification Results*

A single long window SA Z score is compared to the proposed MSWA. For both the single long window SA and the MSWSA, the windowing Systems 2, 3 and 4 are tested. The System 1, without windowing, is not tested, for reasons explained in the discussion.

A Naive Bayesian classifier [48] was used for every system. No meta-parameter was necessary to adjust, beside those already mentioned (number of windows, choice of windowing, Fractal Dimension's parameter which was constant for all tests.) Moreover, it is a method well-known for its robustness [49]. The function *fitcnb* of MATLAB 2020b was used to perform the training of the Naive Bayesian classifier. This method is well suited for small data sample sizes.

The classification was performed using only the MSWSA feature. The MSWSA of all the differential channels were used. Hence, 18 features were included.

The performances were obtained by a Leave-One-Out Cross-Validation approach. The results are reported in Table III. Training group performances are not shown because, although useful, they are generally misleading. Only the proposed System 4 using the MSWSA (mean SA over 1332 size 30 windows) showed statistically significant results. The p-value of this result is $p = 0.00049$.

A comparison of the distribution of the SA score Z for a single long window and the MSWSA is shown in Fig. 7 for all channels together. There is no significant difference between the groups. Also, the SA score Z are extremely strong.

VII. DISCUSSION

The simulation results shown in this paper were based on a certain nonlinear method (FD), a particular window (Welch) and a specific type of signal, the AR(1) process with very low pass characteristics. It must be emphasized that every time the SA is used, a careful examination of the null hypothesis should be carried out with the selected nonlinear method and window on signals with similar power spectrum as the data. In other words, the analysis carried out in this paper should be done for any new combination of nonlinear function, window, and signal. The SA should never be used blindly [50].

The notion of nonlinearity in a signal is perhaps vague in the literature. A direct definition of a "nonlinear signal" was found in [51] "Nonlinear Signal: A nonlinear signal is generally defined as the signal generated by the system that does not obey superposition and scaling properties." Another way to understand it can be “a signal for which a sample has nonlinear dependence with respect to past values.”

As it should be reminded in every paper about the SA, the interpretation of the results must be limited to rejecting or not that “a linear, Gaussian, stationary, stochastic dynamical process underlies the data” [52].

The Welch’s window was used to obtain the simulation results because it was the recommended window in [22] while the Hann window was selected for the experimental study in order to completely remove the periodicity mismatch while avoiding the need to adjust window parameters. However, these choices may be improved. Obviously, the optimal window depends strongly on the nonlinear method used and the type of signal analysed. Multiple testing with different windows can be performed if a sufficient number of subjects is available.

As for the choice of the window’s type, multiple trials should be made with more data to optimize the method with respect to the window’s size. Also, multiple window sizes could be used simultaneously to give multiple MSWSA. Also, more fundamentally, well-controlled experiments should be conducted to allow a better understanding of the MSWSA relationship with the EEG signal.

The SA has variants in which the null hypothesis includes a static (memoryless) monotonic nonlinear transform such as Amplitude Adjusted Fourier Transform (AAFT) [1] or Iterated AAFT (IAAFT) [18]. Another approach that considers a data-generating process is linear and Gaussian as null hypothesis was also proposed in [53]. The proposition of this paper can be extended to these variants with the same expected benefits.

The System 1 was not considered in the classification trials. The total absence of windowing leaves artifacts which are themselves influenced by the power spectrum. For example, a discontinuity in a high frequency signal has generally less impact than in a low frequency signal. Among the tried windowing approach, only the proposed System 4 was able to achieve good classification results. Hence, the use of this windowing system was necessary to show that the MSWSA can extract useful information in EEGs. Also, the reduced variability of the ADD group was possible to observe because of the System 4 windowing method. Although it was possible to observe this behavior with the other systems, it was less evident. Moreover, the single large window approach was not able to classify the EEG. The very large variability of the large window score is probably linked to the detection of signals artifacts. The small window approach was therefore an improved step.

The success rate obtained (93%) with the proposed scheme was comparable to the performances reported in the literature (up to 98%) [3]. The method did not require data augmentation technique. In some studies, the data augmentation is made before the separation of the data for cross-validation purpose, and this can lead to biased performances estimation. The number of trials was kept low, with the prior hypothesis that the proposed method would perform better and with the use of the Naïve Bayesian method which does not needs meta-parameters. However, two main weakness can be identified. First, the number of subjects was low. Although the success rate obtained is statistically significant (as stated earlier, p =0.00049), it has a large interval of confidence. With an “add two successes and two failures” correction for the proportion estimation with small number of data, a lower bound of the bilateral 95% confidence interval of the success rate is 60%. Second, the regrouping of ADHD and ANX was decided after observing the variability of the scores of the three groups. This is a multiple comparison problem which arises commonly in data mining. However, compensating the threshold value with a Bonferroni correction with 4 possible class regroupings (including leaving the three classes separated), the success rate reported would still pass a significance test at α= 0.002.

The results shown in Fig. 7 can be interpreted as the single long window SA is extremely sensitive to all kinds of artifacts. Such behavior did not seem to be present when the MSWSA was used.

The stationarity is often tested before using the SA, as was done in [54]. As it is one of the main aspects of the present paper, non-stationarity can strongly affect the result of the SA. However, for really short time series, it is hard to distinguish between nonlinearity and non-stationarity. Moreover, while the goal of understanding the relation between the EEG nonlinear signal characteristic and the physiology (or even a patient’s condition) is desirable, there is still lot of research to be done to achieve this with studies similar to [55].

Meanwhile, algorithms that can do automatic diagnostics are probably much closer. Although the approach presented in this paper could lead to better understanding of the physiology, it was more geared toward giving a new feature useful for EEG classification. In that sense, there is no important difference if the reason for classification performance is caused by nonlinearity or non-stationarity. However, perhaps the separation of nonlinearity and non-stationarity could lead to better classification.

The effect of non-normal distribution on the SA should also be mentioned. It is well-known that the distribution of the signal affects the results of the SA [1]. For this reason, an Amplitude-Adjusted Fourier-Transform (AAFT) and its iterative form (IAAFT) methods are commonly used, compensating for a static nonlinearity (memoryless). Is has also been used for EEG SA testing [39] and synchronization evaluation [55]. Again, in a classification-oriented approach, separating the effect of the distribution with the nonlinearity or non-stationarity is not necessary. Moreover, for very small windows, the relation between the windowing method and the amplitude adjustment is complex and could lead to unexpected results. It should be remembered that to reduce the multiple testing problem, a limited number of features and parameters must be tested.

While [54] draws the attention toward low frequency components (Delta, Theta, Alpha, Beta waves), the use of very short windows could reduce their importance in FD analysis. In fact, the windowing effect transfers a part of the low frequency signal’s energy toward the mean and the FD is not affected by the signal’s mean. Hence, it is possible that the high frequency Gamma waves are emphasized by this approach. Further studies need to be done to confirm the importance of Gamma waves for the presented classification problem.

Larger tests should be performed, not only to increase the confidence in method and allow meta parameter optimization, but also to test the performances of other nonlinear features in the MSWSA method. The selection of the nonlinear feature should include those commonly studied for ADHD [56-58], ADD [59] and ANX [60]. Also, only the MSWSA was used for classification to show that the featured contained information useful for classification purpose. However, in a full EEG classification system, there is no reason to use a single kind of feature. Hence, the proposed feature should be joined to other linear and nonlinear features. Also, in such a larger test, attention should be given to the gender and the age distributions of the subjects. Perhaps this information could explain in part the results presented in this paper. However, these subject characteristics could be used as *à priori* information for the classification problem and the subject age and gender distribution are clearly insufficient to produce the kind of results obtained.

The classification problem tackled in this paper did not include a control group. Although this is a limitation of the study, it is not as severe as in treatment studies. There is no placebo effect involved. Also, the problem of distinguishing the three classes is very relevant [61]. Although the classification ended with the regrouping of two classes, the results showed a good step forward in EEG signal classification. Perhaps an even more relevant but challenging future research question would be to have subjects with comorbidities between attention deficit and anxiety [61].

The interpretation of the SA results can include nonlinearity, non-stationarity and particular distribution of the data, as generally reported, but could also include Inter-Frequency-Synchronization as presented in [11].

Finally, the windowing approach was shown successful in an EEG classification context. However, it must be reminded that the method should be adapted to the targeted application signal. Notably, the nonlinear feature and the window type and size must be carefully chosen in every application.

VIII. Conclusion

The aim of this paper was to show that small data length can be used in Surrogate analysis if a windowing method is applied both on the original data series as well as on the surrogate data series. The method proposed, although very conservative, allows to rule out the effect of windowing of the already complicated interpretation of the surrogate analysis. The experimental validation showed that the proposed windowing structured could be used in realistic context. The windowing was shown to be effective when very small windows are used. Therefore, it allowed to create a new feature, the MSWSA, that would otherwise be ineffective. This new feature was tested successfully on an EEG dataset. It showed interesting statistical properties and discrimination capabilities. However, without the new windowing method, the new feature would not have produced significant experimental results.

References

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena,* vol. 58, no. 1-4, pp. 77-94, 1992.

[2] M. Caza-Szoka, D. Massicotte, F. Nougarou, and M. Descarreaux, "Surrogate analysis of fractal dimensions from SEMG sensor array as a predictor of chronic low back pain," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016: IEEE, pp. 6409-6412.

[3] A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin, and M. Elsharnouby, "FOCUS: detecting ADHD patients by an EEG-based serious game," *IEEE Transactions on Instrumentation and Measurement,* vol. 67, no. 7, pp. 1512-1520, 2018.

[4] M. Caza-Szoka and D. Massicotte, "Windowing Compensation in Fourier Based Surrogate Analysis," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2021: IEEE, pp. 1-6.

[5] A. Ravelomanantsoa, H. Rabah, and A. Rouane, "Compressed sensing: A simple deterministic measurement matrix and a fast recovery algorithm," *IEEE Transactions on Instrumentation and Measurement,* vol. 64, no. 12, pp. 3405-3413, 2015.

[6] J. A. Mucarquer, P. Prado, M.-J. Escobar, W. El-Deredy, and M. Zañartu, "Improving EEG muscle artifact removal with an EMG array," *IEEE transactions on instrumentation and measurement,* vol. 69, no. 3, pp. 815-824, 2019.

[7] P. M. Sovilj, S. S. Milovancev, and V. Vujicic, "Digital stochastic measurement of a nonstationary signal with an example of EEG signal measurement," *IEEE Transactions on Instrumentation and Measurement,* vol. 60, no. 9, pp. 3230-3232, 2011.

[8] A. Lay-Ekuakille *et al.*, "Entropy index in quantitative EEG measurement for diagnosis accuracy," *IEEE Transactions on Instrumentation and Measurement,* vol. 63, no. 6, pp. 1440-1450, 2013.

[9] M. Caza-Szoka, D. Massicotte, and F. Nougarou, "Naive Bayesian learning for small training samples: Application on chronic Low Back Pain diagnostic with sEMG sensors," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015: IEEE, pp. 470-475.

[10] S. Thongpanja, A. Phinyomark, F. Quaine, Y. Laurillau, C. Limsakul, and P. Phukpattaranont, "Probability density functions of stationary surface EMG signals in noisy environments," *IEEE Transactions on Instrumentation and Measurement,* vol. 65, no. 7, pp. 1547-1557, 2016.

[11] M. Caza-Szoka and D. Massicotte, "Detection of Non Random Phase Signal in Additive Noise with Surrogate Analysis," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019: IEEE, pp. 1159-1163.

[12] A. Josko, "Discrete wavelet transform in automatic ECG signal analysis," in *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, 2007: IEEE, pp. 1-3.

[13] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 25, no. 9, p. 097610, 2015.

[14] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: review of an emerging field," *Clinical neurophysiology,* vol. 116, no. 10, pp. 2266-2301, 2005.

[15] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E,* vol. 64, no. 6, p. 061907, 2001.

[16] L. Faes, H. Zhao, K. H. Chon, and G. Nollo, "Time-varying surrogate data to assess nonlinearity in nonstationary time series: application to heart rate variability," *IEEE transactions on biomedical engineering,* vol. 56, no. 3, pp. 685-695, 2008.

[17] J. Theiler and D. Prichard, "Constrained-realization Monte-Carlo method for hypothesis testing," *Physica D: Nonlinear Phenomena,* vol. 94, no. 4, pp. 221-235, 1996.

[18] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical review letters,* vol. 77, no. 4, p. 635, 1996.

[19] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D: Nonlinear Phenomena,* vol. 142, no. 3-4, pp. 346-382, 2000.

[20] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE,* vol. 66, no. 1, pp. 51-83, 1978.

[21] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical review A,* vol. 34, no. 3, p. 2427, 1986.
[22] T. Suzuki, T. Ikeguchi, and M. Suzuki, "Effects of data windows on the methods of surrogate data," *Physical Review E,* vol. 71, no. 5, p. 056708, 2005.
[23] J. Timmer, "Power of surrogate data testing with respect to nonstationarity," *Physical Review E,* vol. 58, no. 4, p. 5153, 1998.
[24] R. G. Andrzejak, K. Schindler, and C. Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Physical Review E,* vol. 86, no. 4, p. 046206, 2012.
[25] N. Kannathal, U. R. Acharya, C. M. Lim, and P. Sadasivan, "Characterization of EEG—a comparative study," *Computer methods and Programs in Biomedicine,* vol. 80, no. 1, pp. 17-23, 2005.
[26] W. S. Pritchard, D. W. Duke, and K. K. Krieble, "Dimensional analysis of resting human EEG II: Surrogate-data testing indicates nonlinearity but not low-dimensional chaos," *Psychophysiology,* vol. 32, no. 5, pp. 486-491, 1995.
[27] D. Prichard and J. Theiler, "Generating surrogate data for time series with several simultaneously measured variables," *Physical review letters,* vol. 73, no. 7, p. 951, 1994.
[28] J. Fell, J. Röschke, and C. Schäffner, "Surrogate data analysis of sleep electroencephalograms reveals evidence for nonlinearity," *Biological cybernetics,* vol. 75, no. 1, pp. 85-92, 1996.
[29] Y.-J. Lee, Y.-S. Zhu, Y.-H. Xu, M.-F. Shen, H.-X. Zhang, and N. Thakor, "Detection of non-linearity in the EEG of schizophrenic patients," *Clinical Neurophysiology,* vol. 112, no. 7, pp. 1288-1294, 2001.
[30] Y. Shen, E. Olbrich, P. Achermann, and P. Meier, "Dimensional complexity and spectral properties of the human sleep EEG," *Clinical Neurophysiology,* vol. 114, no. 2, pp. 199-209, 2003.
[31] N. Pradhan and P. Sadasivan, "Relevance of surrogate-data testing in electroencephalogram analysis," *Physical Review E,* vol. 53, no. 3, p. 2684, 1996.
[32] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PloS one,* vol. 8, no. 11, p. e79476, 2013.
[33] F. Edition, "Diagnostic and statistical manual of mental disorders," *Am Psychiatric Assoc,* vol. 21, 2013.
[34] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, "EEG signal analysis: a survey," *Journal of medical systems,* vol. 34, no. 2, pp. 195-212, 2010.
[35] J. Theiler, "On the evidence for low-dimensional chaos in an epileptic electroencephalogram," *Physics Letters A,* vol. 196, no. 1-2, pp. 335-341, 1994.
[36] C. Stam, B. Jelles, H. Achtereekte, S. Rombouts, J. Slaets, and R. Keunen, "Investigation of EEG non-linearity in dementia and Parkinson's disease," *Electroencephalography and clinical neurophysiology,* vol. 95, no. 5, pp. 309-317, 1995.
[37] C. L. Ehlers, J. Havstad, D. Prichard, and J. Theiler, "Low doses of ethanol reduce evidence for nonlinear structure in brain activity," *Journal of Neuroscience,* vol. 18, no. 18, pp. 7474-7486, 1998.
[38] R. Kunhimangalam, P. K. Joseph, and O. Sujith, "Nonlinear analysis of EEG signals: Surrogate data analysis," *IRBM,* vol. 29, no. 4, pp. 239-244, 2008.
[39] S. D. Puthankattil, "Fronto-Temporal Analysis of EEG Signals of Patients with Depression: Characterisation, Nonlinear Dynamics and Surrogate Analysis," in *Electroencephalography-From Basic Research to Clinical Applications*: IntechOpen, 2020.
[40] X. Wang, X. Zhao, F. Li, Q. Lin, and Z. Hu, "Sample entropy and surrogate data analysis for Alzheimer's disease," *Mathematical Biosciences and Engineering,* vol. 16, no. 6, pp. 6892-6906, 2019.
[41] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena,* vol. 31, no. 2, pp. 277-283, 1988.
[42] C. Gómez, Á. Mediavilla, R. Hornero, D. Abásolo, and A. Fernández, "Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients," *Medical engineering & physics,* vol. 31, no. 3, pp. 306-313, 2009.
[43] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics,* vol. 15, no. 2, pp. 70-73, 1967.
[44] A. V. Oppenheim, *Discrete-time signal processing*. Pearson Education India, 1999.
[45] S. Sur and V. K. Sinha, "Event-related potential: An overview," *Industrial psychiatry journal,* vol. 18, no. 1, p. 70, 2009.
[46] S. J. Luck, G. F. Woodman, and E. K. Vogel, "Event-related potential studies of attention," *Trends in cognitive sciences,* vol. 4, no. 11, pp. 432-440, 2000.
[47] S. J. Luck and E. S. Kappenman, *The Oxford handbook of event-related potential components*. Oxford university press, 2011.
[48] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine learning,* vol. 29, no. 2, pp. 103-130, 1997.
[49] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and information systems,* vol. 14, no. 1, pp. 1-37, 2008.
[50] D. Kugiumtzis, "Test your surrogate data before you test for nonlinearity," *Physical Review E,* vol. 60, no. 3, p. 2808, 1999.
[51] T. Balli and R. Palaniappan, "Nonlinear Approach to Brain Signal Modeling," in *Encyclopedia of Information Science and Technology, Second Edition*: IGI Global, 2009, pp. 2834-2839.
[52] J. Timmer, "What can be inferred from surrogate data testing?," *Physical Review Letters,* vol. 85, no. 12, p. 2647, 2000.
[53] S. Giannerini, E. Maasoumi, and E. B. Dagum, "Entropy testing for nonlinear serial dependence in time series," *Biometrika,* vol. 102, no. 3, pp. 661-675, 2015.
[54] L. Päeske *et al.*, "Surrogate data method requires end-matched segmentation of electroencephalographic signals to estimate non-linearity," *Frontiers in physiology,* vol. 9, p. 1350, 2018.
[55] H. Bakhshayesh, S. P. Fitzgibbon, A. S. Janani, T. S. Grummett, and K. J. Pope, "Detecting connectivity in EEG: a comparative study of data-driven effective connectivity measures," *Computers in biology and medicine,* vol. 111, p. 103329, 2019.
[56] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," *Research on Biomedical Engineering,* vol. 36, no. 1, pp. 1-11, 2020.
[57] H. T. Tor *et al.*, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Computer Methods and Programs in Biomedicine,* vol. 200, p. 105941, 2021.
[58] H. Sohn *et al.*, "Linear and non-linear EEG analysis of adolescents with attention-deficit/hyperactivity disorder during a cognitive task," *Clinical Neurophysiology,* vol. 121, no. 11, pp. 1863-1870, 2010.
[59] F. Nazarimehr, S. Jafari, S. M. R. Hashemi Golpayegani, and J. Sprott, "Can Lyapunov exponent predict critical transitions in biological systems?," *Nonlinear Dynamics,* vol. 88, no. 2, pp. 1493-1500, 2017.
[60] P. L. Schoenberg, "Linear and Nonlinear EEG-Based Functional Networks in Anxiety Disorders," *Anxiety Disorders,* pp. 35-59, 2020.
[61] A. Canadian, "Resource Alliance (CADDRA): Canadian ADHD Practice Guideliness, 4.1 Edition," *Toronto, Ontario: CADDRA,* 2020.

Manouane Caza-Szoka received his B.E, M.E. and Ph.D. in electrical engineering from Université du Québec à Trois-Rivière, Canada, in 2010, 2013 and 2020 respectively. He is currently working as a Postdoctoral Fellow. His research interests include biosignal processing, acoustic signal processing and real-time signal processing.

Daniel Massicotte (Senior Member, IEEE) received the B.Sc.A. and M.Sc.A. degrees in electrical engineering and industrial electronics in 1987 and 1990 respectively from the Université du Québec à Trois-Rivières (UQTR), Qc, Canada. He obtained the Ph.D. degree in electrical engineering in 1995 at the École Polytechnique de Montréal, Qc,

Canada. In 1994, he joined the Department of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, where he is currently a Full Professor. He is Founder of the Laboratory of Signal and Systems Integration. Since 2001, he has been Founding President and Chief Technology Officer of Axiocom Inc. He has been Head of the Industrial Electronic Research Group from 2011 to 2018 and Head of the Electrical and Computer Engineering Department, from 2011 to 2020. Head of the Research Chair in Signals and Intelligence of High-Performance Systems. He received the Douglas R. Colton Medal for Research Excellence awarded by the Canadian Microelectronics Corporation, the PMC-Sierra High Speed Networking and Communication Award and the Second place at the Complex Multimedia/Telecom IP Design Contest from Europractice. His research interests include advanced VLSI implementation, digital signal processing for wireless communications, measurement, medical and control problems for linear/nonlinear complex systems. He has proposed many methods based on modern signal and biosignal processing such as machine learning, transform domain, and metaheuristics. He is author/coauthor of more than 200 technical papers in international conferences and journals, as well as of 9 inventions. He was also General Chair of IEEE NEWCAS 2014 and Guest Editor of the Springer Analog Integrated Circuits and Signal Processing for the special issues of NEWCAS 2013.

Dr. Massicotte is also member of the "Ordre des Ingénieurs du Québec", "Groupe de Recherche en Électronique Industrielle" (GREI), and "Microsystems Strategic Alliance of Quebec" (ReSMiQ).