Multivariate Multiscale Dispersion Entropy of Biomedical Times Series

Citation for published version:
Azami, H, Fernandez, A & Escudero, J 2019, 'Multivariate Multiscale Dispersion Entropy of Biomedical Times Series', Entropy, vol. 21, no. 9, 913. https://doi.org/10.3390/e21090913

Digital Object Identifier (DOI):
10.3390/e21090913

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Entropy

Article

Multivariate Multiscale Dispersion Entropy of Biomedical Times Series

Hamed Azami[1,2,*], Alberto Fernández[1], and Javier Escudero[3]

[1] School of Engineering, Institute for Digital Communications, University of Edinburgh, Edinburgh, King's Buildings, EH9 3FB, United Kingdom. (emails: hamed.azami@ed.ac.uk, javier.escudero@ed.ac.uk).
[2] Department of Neurology and Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA.
[3] Departamento de Psiquiatría y Psicología Médica, Universidad Complutense de Madrid, Madrid, Spain. He is also with Laboratorio de Neurociencia Cognitiva y Computacional, Centro de Tecnología Biomédica, Universidad Politecnica de Madrid and Universidad Complutense de Madrid, Madrid, Spain and with the Insitituto de Investigación Sanitaria San Carlos (IdSSC).
* Correspondence: hazami@mgh.harvard.edu

Academic Editor: name
Version September 10, 2019 submitted to Journal Not Specified

Abstract: Due to the non-linearity of numerous physiological recordings, non-linear analysis of multi-channel signals has been extensively used in biomedical engineering and neuroscience. Multivariate multiscale sample entropy (MSE - mvMSE) is a popular non-linear metric to quantify the irregularity of multi-channel time series. However, mvMSE has two main drawbacks: 1) the entropy values obtained by the original algorithm of mvMSE are either undefined or unreliable for short signals (300 sample points); and 2) the computation of mvMSE for signals with a large number of channels requires the storage of a huge number of elements. To deal with these problems and improve the stability of mvMSE, we introduce multivariate multiscale dispersion entropy (MDE - mvMDE), as an extension of our recently developed MDE, to quantify the complexity of multivariate time series. We assess mvMDE, in comparison with the state-of-the-art and most widespread multivariate approaches, namely mvMSE and multivariate multiscale fuzzy entropy (mvMFE), on multi-channel noise signals, bivariate autoregressive processes, and three biomedical datasets. The results show that mvMDE takes into account dependencies in patterns across both the time and spatial domains. The mvMDE, mvMSE, and mvMFE methods are consistent in that they lead to similar conclusions about the underlying physiological conditions. However, the proposed mvMDE discriminates various physiological states of the biomedical recordings better than mvMSE and mvMFE. In addition, for both the short and long time series, the mvMDE-based results are noticeably more stable than the mvMSE- and mvMFE-based ones. For short multivariate time series, mvMDE, unlike mvMSE, does not result in undefined values. Furthermore, mvMDE is faster than mvMFE and mvMSE and also needs to store a considerably smaller number of elements. Due to its ability to detect different kinds of dynamics of multivariate signals, mvMDE has great potential to analyse various signals.

Keywords: Complexity; multivariate multiscale dispersion entropy; multivariate time series; electroencephalogram; magnetoencephalogram.

1. Introduction

Multivariate techniques are needed to analyse data consisting of more than one time series [1–3]. The majority of physiological and pathophysiological activities, and even many non-physiological signals, include interactions between different kinds of single processes. Thus, we expect that parameters or measures with different origins are considered in a multivariate way [1,4]. Furthermore, recent developments in sensor technology enabling routine recordings of multi-channel signals have led to an increasing popularity of this kind of analysis on physiological data [1–3,5,6].

Advances on information theory and non-linear dynamical approaches have recently allowed the study of different kinds of multivariate time series [3,7–9]. Due to the intrinsic non-linearity of diverse physiological and non-physiological processes, non-linear analysis of multivariate time series has been broadly used in biomedical signal processing with the aim of studying the relationship between simultaneously recorded signals [3,7,8].

Multivariate multiscale entropy (mvMSE) as a powerful non-linear measure is based on a combination of multivariate sample entropy (SampEn - mvSE) and the coarse-graining process [8]. mvSE characterizes the likelihood that similar multi-channel embedded patterns, which consider both the time and spatial domains, within a time series will remain similar when the pattern length is increased [3]. mvMSE, by taking into account both the spatial and time domains, shows the complexity of multi-channel signals [8]. Complexity reflects the degree of structural richness of time series [8,10] and is different with that of irregularity or uncertainty defined from classical entropy methods such as SampEn [11], permutation entropy (PerEn) [12], and dispersion entropy (DisEn) [13]. That is to say, neither completely regular or certain nor completely irregular (uncorrelated random) time series are truly complex, since none of them is structurally rich at a global level [8,10,14–16].

The multivariate multiscale entropy-based analysis is interpreted based on: 1) the multivariate time series $\mathbf{X}$ is more complex than the multivariate time series $\mathbf{Y}$, if for the most temporal scales, the mvSE measures for $\mathbf{X}$ are larger than those for $\mathbf{Y}$; 2) a monotonic fall in the multivariate entropy values along the temporal scale factors shows that the signal only includes useful information at the smallest scale factors; and 3) a multivariate signal illustrating long-range correlations and complex creating dynamics is characterized by either a constant mvSE or this demonstrates a monotonic rise in mvSE with the temporal scale factor [8].

Although the mvMSE is a powerful and widely-used method, when applied to short signals, the results may be undefined or unreliable [17]. To alleviate this shortcoming, multivariate multiscale fuzzy entropy (mvMFE) based on multivariate fuzzy entropy (mvFE) and the coarse-graining process was suggested [18]. To decrease the running time of the mvMFE proposed in [18], we have recently proposed an mvMFE with a new fuzzy membership function [17]. Nevertheless, the mvMFE is still slow for real-time applications and may lead to unreliable results for short signals, as shown later.

To overcome the problem of unreliable values for mvMFE and mvMSE, multivariate multiscale PerEn (mvMPE) was proposed [19]. To have more information regarding the amplitude of multi-channel signals, multivariate weighted multiscale PerEn (mvWMPE) has recently been developed [20]. However, both the mvMPE and mvWMPE do not take into account the cross-statistical properties between multiple input channels and do not follow the concept of complexity for some signals such as white Gaussian noise (WGN) and $1/f$ noise [8,14,17].

mvMSE and mvMFE have growing appeal and broad use. They have been successfully used in a number of biomedical and mechanical engineering applications, such as, to characterise electroencephalogram (EEG) signals in Alzheimer's disease (AD) [21,22], to quantitatively distinguish different horizontal oil–water flow patterns [23], to analyze mechanical vibration noise to stimulate the patient's feet while wearing the shoes [24], to analyze the multivariate cardiovascular time series [25], to characterize focal and non-focal EEG time series [17], to analyze the complexity of interbeat interval and interbreath signals [8], and to analyze the postural fluctuations in fallers and non-fallers older adults [26].

However, mvMSE and mvMFE have the following shortcomings: 1) mvMSE and mvMFE values may be unreliable and unstable for short signals (300 sample points); 2) they are not quick enough for real-time applications; and 3) computation of mvMSE and mvMFE of a signal with a large number of channels needs to have large memory space, as shown later. To address these drawbacks and due to the advantages of multiscale dispersion entropy (DispEn - MDE) over the state-over-the-art multiscale entropy techniques in terms of distinguishing different kinds of dynamics of univariate synthetic and real time series and computation time [27–29], we propose four algorithms to extend our recently developed MDE to its multivariate forms, termed multivariate MDE (mvMDE). To evaluate the mvMDE methods, we use both synthetic and real multivariate datasets. Our results indicate that mvMDE is noticeably faster than the existing methods, leads to more stable results, better discriminates different kinds of biomedical time series, does not lead to undefined values for short multivariate time series, and needs to store a considerably smaller number of elements in comparison with mvMSE and mvMFE.

2. Multivariate multiscale dispersion entropy (mvMDE)

In this study, we propose and explore three different alternative implementations of mvMDE until we arrive at a fourth and preferred one. All the mvMDE implementations include two main steps: 1) coarse-graining process for multivariate time series; and 2) multivariate DispEn (mvDE), as an extension of our recently developed DisEn [13]. It is worth noting that for all the mvMDE algorithms, the mapping based on the normal cumulative distribution function (NCDF) used in the calculation of mvDE for the first temporal scale factor is maintained fixed across all scales. In fact, in the mvMDE, μ and σ of the NCDF are respectively set at the average and standard deviation (SD) of the original time series and they remain constant for all temporal scale factors. This fact is similar to r in the mvMSE and mvMFE, setting at a certain percentage (usually 15%) of the SD of the original signal and remaining constant for all scales [8,17].

2.1. Coarse-graining process for multivariate signals

Assume we have a p-channel time series $\mathbf{U} = \{u_{k,b}\}_{k=1,2,\dots,p}^{b=1,2,\dots,L}$ of length L. In the mvMDE algorithms, for each channel, the original signal is first divided into non-overlapping segments of length τ, named scale factor. Next, for each channel, the average of each segment is calculated to derive the coarse-grained signals as follows [8,17]:

$$x_{k,i}{}^{(\tau)} = \frac{1}{\tau} \sum_{b=(i-1)\tau+1}^{i\tau} u_{k,b},\ 1 \le i \le \left\lfloor \frac{L}{\tau} \right\rfloor = N\ ,\ 1 \le k \le p \tag{1}$$

where N denotes the length of the coarse-grained signal. The second step of mvMDE is calculating the mvDE of each coarse-grained signal.

2.2. Background information for the mvDE

We build four diverse alternative implementations of mvDE (mvDE$_{\text{I}}$ to III and mvDE) until we arrive at a preferred (or optimal) one, i.e., mvDE. However, we here present all the simpler alternatives (mvDE$_{\text{I}}$ to mvDE$_{\text{III}}$), since they can still be useful in some settings and allow for clearer comparisons with other current approaches.

2.2.1. mvDE$_{\text{I}}$

The mvDE$_{\text{I}}$ of the multi-channel coarse-grained time series $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,N}$, which is based on the mvMPE algorithm [19], is calculated as follows:

a) First, $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,N}$ are mapped to c classes with integer indices from 1 to c. To this aim, there are a number of linear and nonlinear mapping approaches [30]. The simple linear mapping technique may lead to the problem of assigning the majority of $x_{k,i}$ to limited classes when maximum or minimum values are noticeably larger or smaller than the mean/median value of the image [30]. The weak permanence of DispEn with linear mapping for the characterization of syntactic and real data was illustrated in [13].

A large number of natural processes illustrate a progression from small beginnings that accelerates and approaches a climax over time (e.g., a sigmoid function) [31,32]. When there is not detailed information, a sigmoid function is often used [30,32–34]. The choice of sigmoid function in the context of DispEn was discussed in [30]. We here use NCDF as a well-known sigmoid function like in [13]. Note that using NCDF for each channel also deals with the shortcoming of the amplitude values of each of series $\mathbf{x}_k$ $(k = 1, 2, \ldots, p)$ may be dominated by the components of vectors coming from the time series with the largest amplitudes. The NCDF maps $\mathbf{X}$ into $\mathbf{Y} = \{y_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$ from 0 to 1 as follows:

$$y_{k,i} = \frac{1}{\sigma_k \sqrt{2\pi}} \int_{-\infty}^{x_{k,i}} e^{\frac{-(t-\mu_k)^2}{2\sigma_k^2}} \, \mathrm{d}t \tag{2}$$

where σ_k and μ_k are the SD and mean of time series $\mathbf{x}_k$, respectively. Then, we use a linear algorithm to assign each $y_{k,i}$ to an integer from 1 to c. To do so, for each member of the mapped signal, we use $z_{k,i}^c = \mathrm{round}(c \cdot y_{k,i} + 0.5)$, where $z_{k,i}^c$ denotes the i^{th} member of the classified signal in the k^{th} channel and rounding involves either increasing or decreasing a number to the next digit. Note that, although this part is linear, the whole mapping approach is non-linear because of the use of NCDF.

b) Time series $\mathbf{z}_{k,j}^{m,c}$ are made with embedding dimension m and time delay d according to $\mathbf{z}_{k,j}^{m,c} = \{z_{k,j}^c, z_{k,j+d}^c, + \cdots + z_{k,j+(m-1)d}^c\}$, $j = 1, 2, \ldots, N-(m-1)d$ [13][11][12]. Each time series $\mathbf{z}_{k,j}^{m,c}$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$, where $z_{k,j}^c = v_0$,$z_{k,j+d}^c = v_1$,$\ldots$, $z_{k,j+(m-1)d}^c = v_{m-1}$. The number of possible dispersion patterns that can be assigned to each time series $\mathbf{z}_{k,j}^{m,c}$ is equal to c^m, since the signal has m members and each member can be one of the integers from 1 to c [13].

c) For each channel $1 \le k \le p$ and for each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{j \,\big|\, j \le N-(m-1)d, \mathbf{z}_{k,j}^{m,c} \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)p} \tag{3}$$

where # means cardinality. In fact, $p(\pi_{v_0 \ldots v_{m-1}})$ shows the number of dispersion patterns of $\pi_{v_0 \ldots v_{m-1}}$ that is assigned to $\mathbf{z}_{k,j}^{m,c}$, divided by the total number of embedded signals with embedding dimension m multiplied by the number of channels.

d) Finally, based on the Shannon's definition of entropy, the mvDE$_\mathrm{I}$ is calculated as follows:

$$mvDE_I(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right) \tag{4}$$

In case all possible dispersion patterns have equal probability value, the highest value of mvDE$_\mathrm{I}$ is obtained, which has a value of $ln(c^m)$. In contrast, if there is only one $p(\pi_{v_0 \ldots v_{m-1}})$ different from zero, which demonstrates a completely regular/certain signal, the smallest value of mvDE$_\mathrm{I}$ is obtained. In the algorithm of mvDE$_\mathrm{I}$, we compare Np dispersion patterns of a p-channel signal with c^m potential patterns. Thus, at least $c^m + Np$ elements are stored.

To work with reliable statistics to calculate MDE, it was recommended $c^m < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$ [27]. Since mvDE$_\mathrm{I}$ counts the dispersion patterns for every channel of a multivariate time series, it is suggested $c^m < \left\lfloor \frac{pL}{\tau_{max}} \right\rfloor$. mvDE$_\mathrm{I}$ extracts the dispersion patterns from each of channels regardless of their cross-channel information. Thus, mvDE$_\mathrm{I}$ works appropriately when the components of a multivariate signal are statistically independent. However, the mvDE$_\mathrm{I}$ algorithm, like mvPE [19], does not consider the spatial domain of time series. To overcome this problem, we propose mvDE$_\mathrm{II}$ based on the Taken's theorem [17,35].

2.2.2. mvDE$_\mathrm{II}$

The algorithm of mvDE$_\mathrm{II}$ is as follows:

a) First, like mvDE$_\mathrm{I}$, $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$ are mapped to $\mathbf{Z} = \{z_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$ based on the NCDF.

b) To take into account both the spatial and time domains, multi-channel embedded vectors are generated according to the multivariate embedding theory [35]. The multivariate embedded reconstruction of **Z** is defined as:

$$\begin{aligned} Z_{\mathbf{m}}(j) = [&z_{1,j}, z_{1,j+d_1}, \cdots, z_{1,j+(m_1-1)d_1}, \\ &z_{2,j}, z_{2,j+d_2}, \cdots, z_{2,j+(m_2-1)d_2}, \cdots, \\ &z_{p,j}, z_{p,j+d_p}, \cdots, z_{p,j+(m_p-1)d_p}] \end{aligned} \tag{5}$$

where $\mathbf{m} = [m_1, m_2, \ldots, m_p]$ and $\mathbf{d} = [d_1, d_2, \ldots, d_p]$ denote the embedding dimension and the time lag vectors, respectively. Note that the length of $Z_{\mathbf{m}}(j)$ is $\sum_{k=1}^{p} m_k$. For simplicity, we assume $d_k = d$ and $m_k = m$, that is, all the embedding dimension values and all the delay values are equal.

c) Each series $Z_{\mathbf{m}}(j)$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{mp-1}}$, where $z^c_{1,j} = v_0$, $z^c_{1,j+d} = v_1, \ldots,$ $z_{p,j+(m-1)d} = v_{mp-1}$. The number of possible dispersion patterns that can be assigned to each time series $Z_{\mathbf{m}}(j)$ is equal to c^{mp}, since the signal has mp members and each member can be one of the integers from 1 to c.

d) For each of c^{mp} potential dispersion patterns $\pi_{v_0 \ldots v_{mp-1}}$, relative frequency is obtained based on the DisEn algorithm [13] as follows:

$$p(\pi_{v_0 \ldots v_{mp-1}}) = \frac{\#\{j \,\big|\, j \leq N-(m-1)d, Z_{\mathbf{m}}(j) \text{ has type } \pi_{v_0 \ldots v_{mp-1}}\}}{N-(m-1)d} \tag{6}$$

e) Finally, based on the Shannon's definition of entropy, the mvDE$_{\text{II}}$ is calculated as follows:

$$mvDE_{II}(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^{mp}} p(\pi_{v_0 \ldots v_{mp-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{mp-1}})\right) \tag{7}$$

In the algorithm of mvDE$_{\text{II}}$, at least $c^{mp} + Np$ elements are stored. Thus, when p is large, the algorithm needs huge space of memory to store elements. To work with reliable statistics to calculate mvMDE$_{\text{II}}$, it is recommended $c^{mp} < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. Thus, although mvDE$_{\text{II}}$ deals with both the spatial and time domains, the length of a signal and its number of channels should be very large and small, respectively, to reliably calculate mvDE$_{\text{II}}$ values. To alleviate the problem, we propose mvDE$_{\text{III}}$.

2.2.3. mvDE$_{\text{III}}$

The algorithm of mvDE$_{\text{III}}$ is as follows:

a) First, like the mvDE$_{\text{I}}$ and mvDE$_{\text{II}}$ approaches, $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$ are mapped to $\mathbf{Z} = \{z_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$.

b) Multivariate embedded vectors $\mathbf{Z}_{k,\mathbf{m}}(j)$ with length $m+p-1$ are generated according to the Taken's embedding theorem [35] with p embedding dimension vectors $\mathbf{m}_k = [1, 1, \ldots, m_k, \ldots, 1, 1]$ $(k = 1, \ldots, p)$, where m_k denotes the k^{th} element of $\mathbf{m}$. For simplicity, we assume $m_k = m$ and $d_k = d$.

c) Each series $\mathbf{Z}_{k,\mathbf{m}}(j)$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m+p-2}}$. The number of possible dispersion patterns that can be assigned to each time series $\mathbf{Z}_{k,\mathbf{m}}(j)$ is equal to c^{m+p-1}, since the signal has $m+p-1$ members and each member can be one of the integers from 1 to c [13]. Since we count the number of patterns for each of p different $\mathbf{m}_k$ leading to a considerable increase in the number of dispersion patterns, compared with mvDE$_{\text{II}}$, we have more reliable results for a signal with a small number of samplthan those fore points, as shown later.

d) For each channel $1 \leq k \leq p$ and for each of c^{m+p-1} potential dispersion patterns $\pi_{v_0 \ldots v_{m+p-2}}$, relative frequency is obtained as follows:

$$p(\pi_{v_0 \ldots v_{m+p-2}}) = \frac{\#\{j \,\big|\, j \leq N-(m-1)d, \mathbf{Z}_{k,\mathbf{m}}(j) \text{ has type } \pi_{v_0 \ldots v_{m+p-2}}\}}{(N-(m-1)d)p} \tag{8}$$

e) Finally, based on the Shannon's definition of entropy, the mvDE$_{\text{III}}$ is calculated as follows:

$$mvDE_{III}(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^{m+p-1}} p(\pi_{v_0 \ldots v_{m+p-2}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m+p-2}})\right) \quad (9)$$

mvDE$_{\text{III}}$ assumes embedding dimension 1 for all signals except one, which might limit the potential to explore the dynamics. Moreover, in the algorithm of mvDE$_{\text{III}}$, at least $c^{m+p-1} + Np$ elements are stored. Although this number is noticeably smaller than that for mvDE$_{\text{II}}$, the algorithm still needs to have large memory space for a signal with a large number of channels. To work with reliable statistics to calculate mvMDE$_{\text{III}}$, it is recommended $c^{m+p-1} < \left\lfloor \frac{pL}{\tau_{max}} \right\rfloor$. Therefore, albeit mvDE$_{\text{III}}$ takes into account both the spatial and time domains and needs to smaller number of sample points in comparison with mvDE$_{\text{II}}$, there is a need to have a large enough number of samples and small number of channels. To alleviate these deficiencies, we propose mvDE.

2.3. Multivariate dispersion entropy (mvDE)

The mvDE algorithm is as follows:

a) First, like mvDE$_{\text{I}}$ to III, the multivariate signal $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\ldots,p}^{i=1,2,\ldots,N}$ is mapped to c classes with integer indices from 1 to c.

b) Like mvDE$_{\text{II}}$, to consider both the spatial and time domains, multivariate embedded vectors $Z_{\mathbf{m}}(j), 1 \leq j \leq N-(m-1)d$ are created based on the Taken's embedding theorem [35]. For simplicity, we assume $d_k = d$ and $m_k = m$.

c) For every $Z_{\mathbf{m}}(j)$, all combinations of the $\sum_{k=1}^{p} m_k$ elements in $Z_{\mathbf{m}}(j)$ taken m at a time, termed $\phi_q(j)$ $(q = 1, \ldots \binom{mp}{m})$, are created. The number of the combinations is equal to $\binom{mp}{m}$. Therefore, for all channels, we have $(N-(m-1)d)\binom{mp}{m}$ dispersion patterns.

d) For each $1 \leq q \leq \binom{mp}{m}$ and for each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$p(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{j \,|\, j \leq N-(m-1)d, \phi_q(j) \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)\binom{mp}{m}} \quad (10)$$

e) Finally, based on the Shannon's definition of entropy, the mvDE is calculated as follows:

$$mvDE(\mathbf{X}, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(p(\pi_{v_0 \ldots v_{m-1}})\right) \quad (11)$$

In fact, mvDE explores all combinations of patterns of length m within an mp-dimensional embedding vector. In the mvDE algorithm, at least $c^m + Np$ elements are stored. This number is noticeably smaller than those for mvDE$_{\text{II}}$ to III, leading to more stable results for signals with a short length and a large number of samples. As the number of patterns obtained by the mvMDE method is $(N-(m-1)d)\binom{mp}{m}$, it is suggested $c^m < \left\lfloor \frac{L\binom{mp}{m}}{\tau_{max}} \right\rfloor$ to work with reliable statistics. It is worth mentioning that if the order of channels in a multi-channel time series changes, although the assignment to each dispersion pattern obtained by the mvMDE-based methods may change, the entropy value will stay the same.

2.4. Parameters of the mvMDE, mvMSE, and mvMFE methods

In addition to the maximum scale factor τ_{max} described before, there are three other parameters for the mvMDE methods, including the embedding dimension vector $\mathbf{m}$, number of classes c, and time delay vector $\mathbf{d}$. Although some information with regard to the frequency of signals may be ignored for $d_k > 1$, it is better to set $d_k > 1$ for oversampled time series. However, like previous studies about multivariate entropy methods [2,8], we set $d_k = 1$ for simplicity. Nevertheless, when the sampling frequency is considerably larger than the highest frequency component of a time series, the first minimum or zero crossing of the autocorrelation function or mutual information can be utilized

Table 1. Ability to deal with spatial domain and characterization of short signals (300 sample points), typical number of elements to be stored, and typical number of samples needed for each of the mvSE, mvFE, and mvDE algorithms for a p-channel signal with length N sample points.

Methods	Spatial domain	Short signals	Typical number of elements stored	Typical number of samples
mvSE [3]	yes	undefined	$\binom{Np}{2} + Np(pm+1)$	$10^m < N$
mvFE [17]	yes	unreliable	$\binom{Np}{2} + Np(pm+1)$	$10^m < N$
mvPE [19] and mvWPE [20]	no	reliable	$m! + Np$	$m! < N$
mvDE$_{\mathrm{I}}$	no	reliable	$c^m + Np$	$\frac{c^m}{p} < N$
mvDE$_{\mathrm{II}}$	yes	unreliable	$c^{mp} + Np$	$c^{mp} < N$
mvDE$_{\mathrm{III}}$	yes	unreliable	$c^{m+p-1} + Np$	$\frac{c^{m+p-1}}{p} < N$
mvDE	yes	reliable	$c^m + Np$	$\frac{c^m}{\binom{mp}{m}} < N$

for the selection of an appropriate time delay [36]. We need $1 < c$ to keep away the trivial case of having only one dispersion pattern. For simplicity, we use $c = 5$ and $m_k = 2$ for all signals used in this study, although the range $2 < c < 9$ leads to similar findings. For more information about c, m_k, and d_k, please refer to [13,30].

In this study, d_k, m_k, and r for the mvMSE and mvMFE were respectively set as 1, 2, and 0.15 of the SD of the original time series following recommendations in [8,17]. The maximum scale factor for mvMSE and mvMFE also follows [8,17]. In the algorithm of mvSE and mvFE, at least $\binom{Np}{2} + Np(pm+1)$ elements are stored (the mvSE code available at http://www.commsp.ee.ic.ac.uk/~mandic/research/Complexity_Stuff.htm). Matlab codes of mvMFE and mvMSE are available at http://dx.doi.org/10.7488/ds/1432. Overall, the characteristics and limitations of the mvSE, mvFE, and mvDE algorithms for a p-channel signal with length N are summarized in Table I.

3. Evaluation signals

In this section, the descriptions of correlated and uncorrelated noise signals, bivariate autoregressive (BAR) process, and real time series used in this study are given.

3.1. Synthetic signals

The irregularity of multivariate $1/f$ noise is lower than multivariate WGN, whereas the complexity of the former is higher than the latter [8,14,17]. Thus, $1/f$ noise and WGN signals have been commonly used to assess the multivariate multiscale entropy techniques [8,17,37]. For more information about the algorithms used for multivariate $1/f$ noise and WGN, please refer to [8,17].

To understand the behaviour of the mvMDE methods on uncorrelated WGN and $1/f$ noise, we first generated a trivariate time series, where originally all three data channels were realization of mutually independent $1/f$ noise. Then, we gradually decreased the number of data channels representing $1/f$ noise (from 3 to 0) and at the same time, increased the number of variates representing independent WGN (from 0 to 3) [37]. The number of channels was always three.

To create correlated bivariate noise time series, we first generated a bivariate uncorrelated random time series $\mathbf{H}$. Afterwards, $\mathbf{H}$ was multiplied with the standard deviation (hereafter, sigma) and then, the value of the mean (hereafter, mu) was added. Next, $\mathbf{H}$ was multiplied by the upper triangular matrix $\mathbf{L}$ obtained from the Cholesky decomposition of a defined correlation matrix $\mathbf{R}$ (which is positive and symmetric) to set the correlation. Here, we set $\mathbf{R} = \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1 \end{bmatrix}$ according to [8,17]. An in-depth study on the effect of correlated and uncorrelated $1/f$ noise and WGN on multiscale entropy approaches can be found in [8,10].

Based on the fact that the larger the order of an autoregressive process, the more complex the AR process [8], we evaluate the mvMDE, mvMSE, and mvMFE methods on a BAR(α) process with the maximum lag α describing the evolution of a set of two variables as a linear function of their past values according to:

$$\mathbf{y}_n = \mathbf{e}_n + \sum_{\gamma=1}^{\alpha} \mathbf{y}_{n-\gamma}\mathbf{A}_\gamma \tag{12}$$

where $\mathbf{y}_n = \{y_n(1), y_n(2)\}$ is the n^{th} sample of a bidimensional time series, $\mathbf{A}_\gamma$ denotes the 2×2 matrix of parameters corresponding to lag order γ, and $\mathbf{e}_n$ is the 2×1 vector of error terms assumed to be WGN[38].

3.2. Real biomedical datasets

1) Dataset of Stride Interval Fluctuations: To investigate the ability of the proposed mvMDE methods to reveal the long-range correlations and dynamics of multivariate signals, the stride interval recordings are used [2,39]. The time series were recorded from ten young, healthy men. Mean age was 21.7 years, changing from 18 to 29 years. Height and weight were 1.77 $\pm$ 0.08 meters (mean $\pm$ SD) and 71.8 $\pm$ 10.7 kg (mean $\pm$ SD), respectively. All ten participants provided informed written consent walking for 1 hour at slow, 1 hour at normal, and 1 hour at fast paces and also walking a metronome set to each subject's mean stride interval. Three walking paces were considered as different variables from the same system. In this way, we expect to be able to discriminate between the metronomically-paced and self-spaced walking. For further information, please refer to [39].

2) Dataset of Focal and Non-focal Brain Activity: The ability of the mvMDE methods, in comparison with mvMFE and mvMSE, to differentiate focal from non-focal recordings is evaluated using a publicly-available EEG dataset [40]. The dataset includes 5 patients and, for each patient, there are 750 focal and 750 non-focal bivariate signals. The length of each recording was 20 s with sampling frequency of 512 Hz (10240 sample points). Further information can be found in [40]. Before computing the aforementioned methods, all recordings were digitally filtered employing an FIR band-pass filter with cut-off frequencies at 0.5 Hz and 40 Hz.

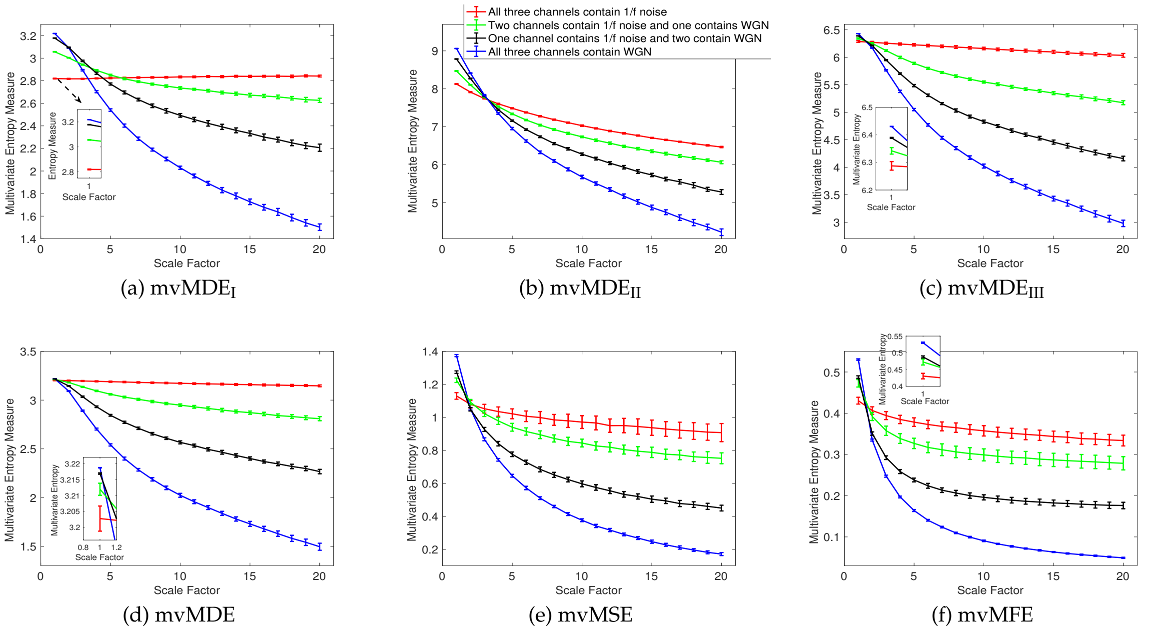
3) Surface MEG Recordings in Alzheimer's Disease: We analysed resting state MEG time series recorded with a 148-channel whole-head magnetometer. All 62 participants agreed for the research, which was approved by the local ethics committee. To screen the cognitive status, a mini-mental state examination (MMSE) was done. There were 36 AD patients (age = 74.06 $\pm$ 6.95 years, all data given as mean $\pm$ SD, and MMSE score = 18.06 $\pm$ 3.36) and 26 controls (age = 71.77 $\pm$ 6.38 years, and MMSE score = 28.88 $\pm$ 1.18). The difference in age between two groups was not significant (p-value = 0.1911, Student's t-test) [41]. The distribution of MEG sensors is shown in Fig. 2 in [41]. For each participant, five minutes of MEG resting state activity were recorded at a sampling frequency of 169.5 Hz. The signals were divided into 10 s segments (1695 samples) and visually inspected using an automated thresholding procedure to discard epochs noticeably contaminated with artifacts. All recordings were digitally band-pass filtered with a Hamming window FIR filter of order 200 and cut-off frequencies at 1.5 Hz and 40 Hz. For more information, please see [41].

4. Results and discussions

4.1. Synthetic signals

4.1.1. Uncorrelated white Gaussian and $1/f$ noises

We first apply the proposed and existing methods to 40 independent realizations of uncorrelated trivariate WGN and $1/f$ noise, described in Section 3. The number of sample points for each of the $1/f$ noise and WGN signals were 15000. mvMSE and mvMFE are based on conditional entropy [2,8,17]. On the other hand, mvMDE is based on the Shannon's entropy definition applied to dispersion patterns. This means that the methods work on slightly different principles. However, the comparison of mvMDE with mvMSE and mvMFE is meaningful because the latter two are the most common multivariate entropy algorithms and MDE has been shown to have similar behaviour to MSE when



(a) $\mathrm{mvMDE_I}$ (b) $\mathrm{mvMDE_{II}}$ (c) $\mathrm{mvMDE_{III}}$

(d) mvMDE (e) mvMSE (f) mvMFE

Figure 1. Mean value and SD of the results using (a) $\mathrm{mvMDE_I}$, (b) $\mathrm{mvMDE_{II}}$, (c) $\mathrm{mvMDE_{III}}$, (d) mvMDE, (e) mvMSE, and (f) mvMFE computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with length 15000 sample points.

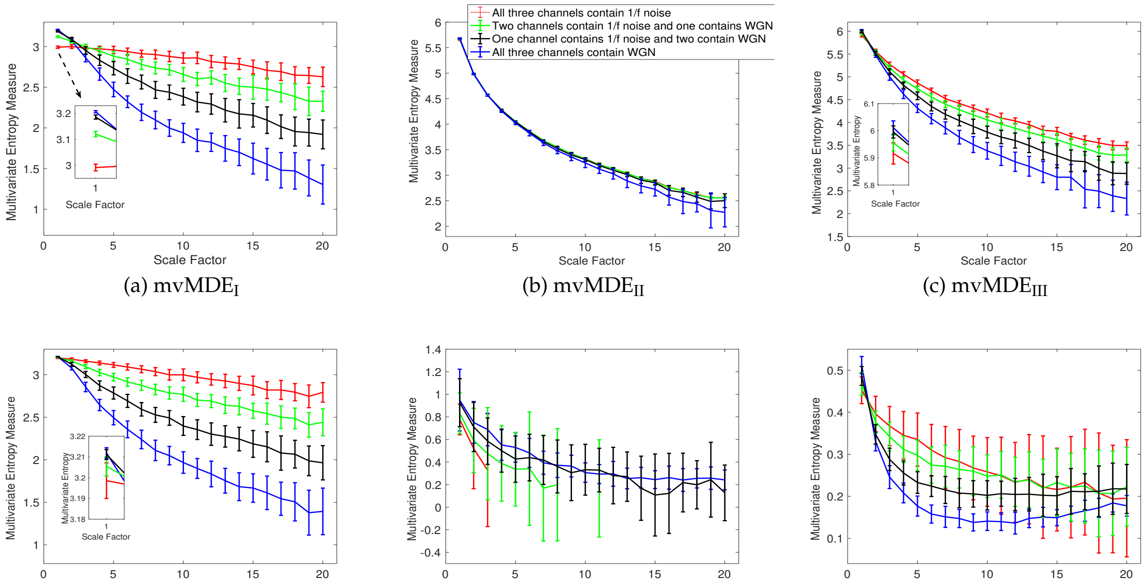
analysing real and synthetic signals [27]. Thus, we compare the mvMDE methods with mvMSE and mvMFE. The average and SD of the results for $\mathrm{mvMDE_I}$, $\mathrm{mvMDE_{II}}$, $\mathrm{mvMDE_{III}}$, mvMDE, mvMSE, and mvMFE are depicted in Fig. 1(a) to 1(f), respectively. Using all the existing and proposed methods, the entropy values of trivariate WGN signals are higher than those of the other trivariate time series at low scale factors. However, the entropy values for the coarse-grained trivariate $1/f$ noise signals stay almost constant or decrease slowly along the temporal scale factor, while the entropy values for the coarse-grained WGN signal monotonically decreases with the increase of scale factors. When the length of WGN signals, obtained by the coarse-graining process, decreases (i.e., the scale factor increases), the mean value of inside each signal converges to a constant value and the SD becomes smaller. Therefore, no new structures are revealed at higher temporal scales. This demonstrates a multivariate WGN time series has information only in small temporal scale factors. In contrast, for trivariate $1/f$ noise signals, the mean value of the fluctuations inside each signal does not converge to a constant value.

For all the methods, the higher the number of variates representing $1/f$ noise, the higher complexity the trivariate signal, in agreement with the fact that multivariate $1/f$ noise is structurally more complex than multivariate WGN [8,14,17]. Here, for multivariate $1/f$ noise and WGN, τ_{max} was 20 for mvMDE, according to Section II.

To compare the results obtained by the mvMDE, mvMSE, and mvMFE methods, we used the coefficient of variation (CV). CV, as a measure of relative variability, is defined as the SD divided by the mean of a time series. We use such a metric as the SDs of time series may increase or decrease proportionally to the mean. We investigate the results obtained by uncorrelated noise signals at scale factor 10, as a trade-off between short and long scale factors. As can be seen in Table II, the smallest CV values for uncorrelated trivariate $1/f$ noise, an uncorrelated combination of bivariate $1/f$ noise and univariate WGN, an uncorrelated combination of bivariate WGN and univariate $1/f$ noise, and trivariate WGN are achieved by mvMDE, $\mathrm{mvMDE_{II}}$, $\mathrm{mvMDE_{II}}$, and $\mathrm{mvMDE_I}$, respectively. Overall, the smallest CV values for trivariate $1/f$ noise and WGN profiles are reached by the mvMDE methods, showing the superiority of the mvMDE methods over mvMSE and mvMFE in terms of stability of results.

Table 2. CV values of the proposed and existing multivariate multiscale entropy-based analyses at scale factor 10 for the uncorrelated trivariate $1/f$ noise and WGN.

Time series	mvMDE_I	mvMDE_II	mvMDE_III	mvMDE	mvMSE	mvMFE
All three channels contain $1/f$ noise	0.0028	0.0025	0.0037	0.0022	0.0405	0.0355
Two channels contain $1/f$ noise and one contains WGN	0.0042	0.0032	0.0036	0.0044	0.0283	0.0274
One channel contains $1/f$ noise and two contain WGN	0.0066	0.0052	0.0058	0.0061	0.0305	0.0292
All three channels contain WGN	0.0072	0.0080	0.0092	0.0101	0.0232	0.0211



(a) mvMDE_I (b) mvMDE_II (c) mvMDE_III

(d) mvMDE (e) mvMSE (f) mvMFE

Figure 2. Mean value and SD of the results obtained by (a) mvMDE_I, (b) mvMDE_II, (c) mvMDE_III, (d) mvMDE, (e) mvMSE, and (f) mvMFE computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with length 300 sample points.

To assess the ability of the mvMDE methods to characterize short signals in comparison with mvMFE and mvMSE, we use trivariate $1/f$ noise and WGN with length of 300 sample points. The results for the mvMDE, mvMSE, and mvMFE approaches at temporal scales 1 to 20 are depicted in Fig. 2(a) to 2(f), respectively. The results show that only mvMDE_I is able to distinguish these four different kinds of noise signals at scale factor 1. For the higher temporal scale factors, mvMDE_I and mvMDE distinguish these time series, showing a limitation of mvMDE for the discrimination of white from $1/f$ noise at lower scale factors and also the importance of considering higher temporal scales for the mvMDE technique. As can be seen in Fig. 2(a) and 2(d), the mvMDE_I and mvMDE methods better discriminate different dynamics of the noise signals. However, the mvMSE values are undefined at higher scale factors. It is worth mentioning that we compared mvMDE with the original algorithms of mvMSE and mvMFE. However, more recent studies on entropy estimation of short physiological signals provided methods to deal with this issue [17,42].

Although the mvMFE- and mvMDE_II-based values are defined at all scale factors, they cannot distinguish the dynamics of different noise signals. The profiles obtained by mvMDE_III are more distinguishable than mvMDE_II, as mentioned that mvMDE_III needs a smaller number of sample points. Nevertheless, the profiles obtained by mvMDE_III have overlaps at several scale factors. Overall, the results show the superiority of mvMDE_I and mvMDE over mvMDE_II, mvMDE_III, mvMSE, and mvMFE for short uncorrelated signals.

4.1.2. Computational Time

Table 3. Computational time of the mvMSE, mvMFE, and mvMDE algorithms with $\tau_{max} = 10$.

Number of channels and samples	mvMSE	mvMFE	mvMDE_I	mvMDE_II	mvMDE_III	mvMDE
2 channels and 1,000 samples	0.051 s	0.066 s	0.014 s	0.023 s	0.026 s	0.020 s
2 channels and 3,000 samples	0.237 s	0.296 s	0.035 s	0.057 s	0.068 s	0.052 s
2 channels and 10,000 samples	1.821 s	2.016 s	0.111 s	0.190 s	0.223 s	0.181 s
5 channels and 1,000 samples	0.209 s	0.223 s	0.028 s	43.096 s	0.490 s	0.050 s
5 channels and 3,000 samples	1.129 s	1.204 s	0.080 s	82.246 s	1.137 s	0.137 s
5 channels and 10,000 samples	9.432 s	9.801 s	0.260 s	218.553 s	3.343 s	0.491 s
8 channels and 1,000 samples	0.489 s	0.501 s	0.042 s	out of memory error	65.560 s	0.086 s
8 channels and 3,000 samples	2.973 s	2.906 s	0.124 s	out of memory error	150.122 s	0.243 s
8 channels and 10,000 samples	27.993 s	25.951 s	0.398 s	out of memory error	363.752 s	0.824 s

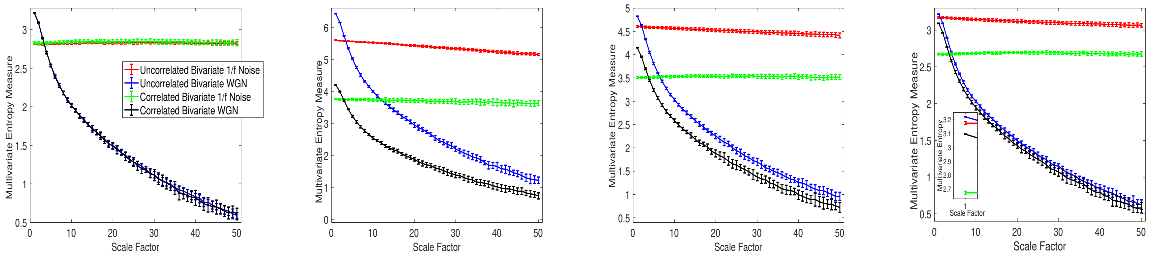
To evaluate the computational time of mvMSE, mvMFE, mvMDE_I to III, and mvMDE, we use uncorrelated multivariate WGN time series with different lengths, changing from 100 to 10,000 sample points, and different number of channels, changing from 2 to 8. The results are depicted in Table III. The simulations have been carried out using a PC with Intel (R) Core (TM) i7-7820X CPU, 3.6 GHz and 16-GB RAM by MATLAB R2018b. The results show that the computation times for mvMSE and mvMFE are close. The slowest algorithm is mvMDE_II, while the fastest ones are mvMDE_I and mvMDE, in that order. For an 8-channel signal with 10,000 samples, using mvMDE_II, the array exceeded the memory available. Overall, in terms of computation time and memory space, mvMDE outperforms the other methods that take into account both the time and spatial domains. We used the mvMSE code provoided in [8] and the mvMDE, mvMSE, and mvMFE Matlab codes have not been optimized.

4.1.3. Correlated white Gaussian and $1/f$ noises

Univariate multiscale entropy approaches only consider every data channel separately and fail to take into account the cross-channel information of multivariate time series [8]. Uncorrelated multi-channel WGN has less structural complexity and more irregularity compared with multi-channel $1/f$ noise. To assess the ability of the existing and proposed multivariate entropy methods to reveal the dynamics across the channels, we created 40 independent realizations of different combinations of bivariate $1/f$ noise and WGN time series with length 20,000 (according to [8,17]), making the channels correlated. Fig. 3(a) to 3(d) respectively show the results obtained using the mvMDE_I, mvMDE_II, mvMDE_III, and mvMDE to model both the within- and cross-channel properties in multivariate signals.

mvMDE_I cannot discriminate the correlated from uncorrelated WGN or $1/f$ noise. This fact is revealed in Fig. 3 (a). Therefore, mvMDE_I should only be used when the components of a multi-channel time series are statistically independent. Multivariate multiscale entropy-based methods at scale factor 1 show the irregularity of multi-channel signals [8]. The mvMDE_II, mvMDE_III, and mvMDE values at scale 1 show that the uncorrelated WGN is the most irregular and unpredictable time series in agreement with [10], while the most irregular signals using mvMFE and mvMSE are the correlated WGN [8,17], in contrast with the fact that correlated multi-channel WGN signals are more predictable and regular than uncorrelated WGN ones [10,27]. Although mvMDE was able to distinguish all four different kinds of noises at the small scale factors, there are some overlaps between the results for the correlated and uncorrelated bivariate WGN time series at the high scale factors showing the importance both low and high temporal scale factors in mvMDE.

The correlated bivariate $1/f$ noise is the most complex signal using the mvMDE_II, mvMDE_III, and mvMDE. The second most complex signal is the uncorrelated bivariate $1/f$ noise, as can be seen in Fig. 3. The decreases of the uncorrelated bivariate WGN profiles using mvMDE_II, mvMDE_III, and mvMDE are the largest, evidencing the fact that the uncorrelated WGN is the least complex time series. These facts are also in agreement with the previous studies [8,14,17]. Therefore, as desired, the mvMDE_II, mvMDE_III, and mvMDE deal with both the cross- and within-channel correlations.



(a) mvMDE_{I} (b) mvMDE_{II} (c) $\text{mvMDE}_{\text{III}}$ (d) mvMDE

Figure 3. Mean value and SD of the results obtained by (a) mvMDE_{I}, (b) mvMDE_{II}, (c) $\text{mvMDE}_{\text{III}}$, and (d) mvMDE computed from 40 different correlated and uncorrelated bivariate WGN and $1/f$ noise time series with length 20,000 sample points.

4.1.4. Bivariate AR processes

The ability of the mvMDE methods to characterize multivariate AR processes is further evaluated using combinations of BAR(1), BAR(3), and BAR(5) with $\mathbf{A}_{\gamma_1} = \begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$, $\mathbf{A}_{\gamma_2} = \begin{bmatrix} 0.10 & 0.10 \\ 0.10 & 0.10 \end{bmatrix}$, and $\mathbf{A}_{\gamma_3} = \begin{bmatrix} 0.15 & 0.15 \\ 0.15 & 0.15 \end{bmatrix}$. The results obtained by the mvMDE_{I}, mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, and mvMDE methods are shown in Fig. 4. As expected, when the lag order increases, the complexity of the corresponding time series using the mvMDE approaches increases, in agreement with the fact that a larger lag order denotes a more complex time series [8]. As the elements of $\mathbf{A}_{\gamma_1}$ are smaller than those of $\mathbf{A}_{\gamma_2}$ and $\mathbf{A}_{\gamma_3}$, the behaviour of the profiles obtained by the mvMDE methods are more similar to the results for WGN (see Fig. 1). In fact, the smaller the elements of $\mathbf{A}_{\gamma}$, the less complex the BAR, leading to lower entropy values at higher scale factors.

In order to investigate the dependence of the mvMDE methods on the sensitivity to changes in the signals, we generated BAR(3) with length of 10000 sample points and sampling frequency of 150 Hz that $\mathbf{A}_{\gamma}$ linearly changes from $\begin{bmatrix} 0.17 & 0 \\ 0 & 0.17 \end{bmatrix}$ to $\begin{bmatrix} 0.17 & 0.17 \\ 0.17 & 0.17 \end{bmatrix}$. In fact, the elements of the diagonal of $\mathbf{A}$ are constant and those of anti-diagonal linearly increase from 0 to 0.17, leading to more complex series. We moved a bivariate window - termed temporal window - with length 2000 samples and 20% overlap along this BAR(3) signal. The entropy of each bivariate temproal window is caculated. The results, depicted in Fig. 5 show that when the time window is occupied at the beginning of the BAR(3) ($\mathbf{A} = \begin{bmatrix} 0.17 & 0 \\ 0 & 0.17 \end{bmatrix}$), the mvMDE_{I}, mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, and mvMDE values at higher scale factors are the smallest, showing the least complexity of BAR(3) in lower temporal windows, while their corresponding entropy values in the end of BAR(3) process ($\mathbf{A} = \begin{bmatrix} 0.17 & 0.17 \\ 0.17 & 0.17 \end{bmatrix}$) are the largest. It is worth noting that as described before, mvMDE_{II} needs a larger number of sample points to appropriately characterize the dynamics of signals. This fact can be observed in Fig. 5, showing mvMDE_{II} is the least able to distinguish such changes.

4.2. Real biomedical datasets

Discrimination of aged and diseased individuals' from control or healthy subjects' time series is a long-lasting challenge in the physiological complexity literature [8,10,17]. To this end, we use the mvMDE methods, in comparison with mvMFE as an improved version of mvMSE [17], to detect different types of dynamical variability of multivariate recordings of three physiological datasets. Of note is that we do not use the mvMDE_{I} for biomedical signals, because it does not take into account both the spatial and time domains at the same time.

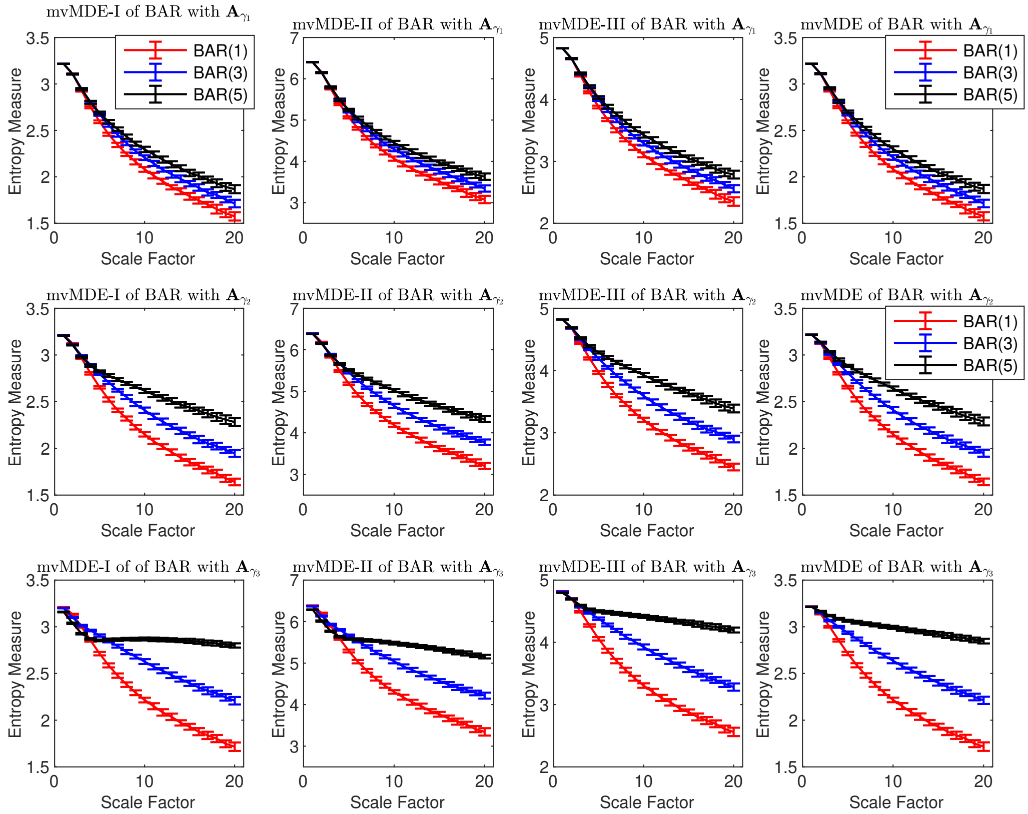


Figure 4. Mean and SD values of the results using mvMDE_{I}, mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, and mvMDE computed from 40 different BAR(1), BAR(3), and BAR(5) time series with $\mathbf{A}_{\gamma_1}$ (first row), $\mathbf{A}_{\gamma_2}$ (second row), and $\mathbf{A}_{\gamma_3}$ (third row).

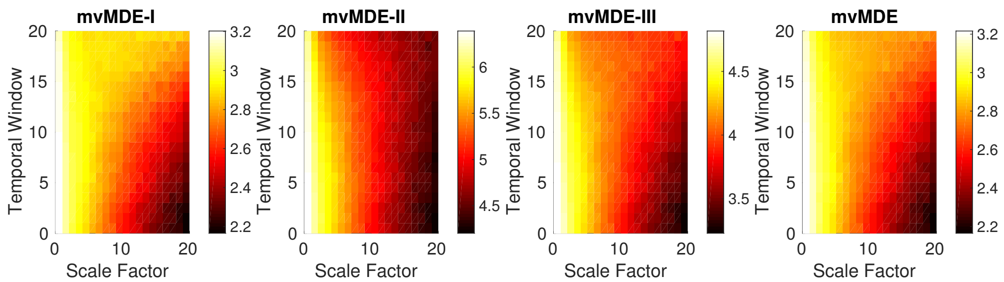
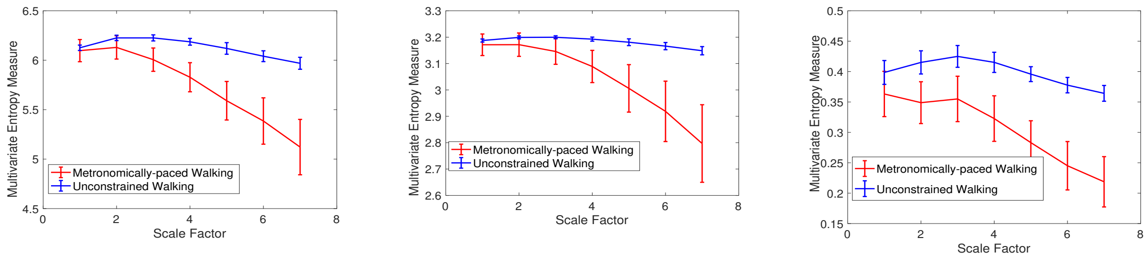


Figure 5. Results obtained by the mvMDE methods using a bivariate temporal window with length 2000 sample points moving along the BAR(3) signal, which the elements of anti-diagonal of the matrix **A** linearly increase from 0 to 0.17, leading to more complex series.



(a) $\mathrm{mvMDE_{III}}$ (b) mvMDE (c) mvMFE

Figure 6. Mean value and SD of the results using (a) $\mathrm{mvMDE_{III}}$, (b) mvMDE, and (c) mvMFE for self-paced vs. metronomically-paced stride interval fluctuations.

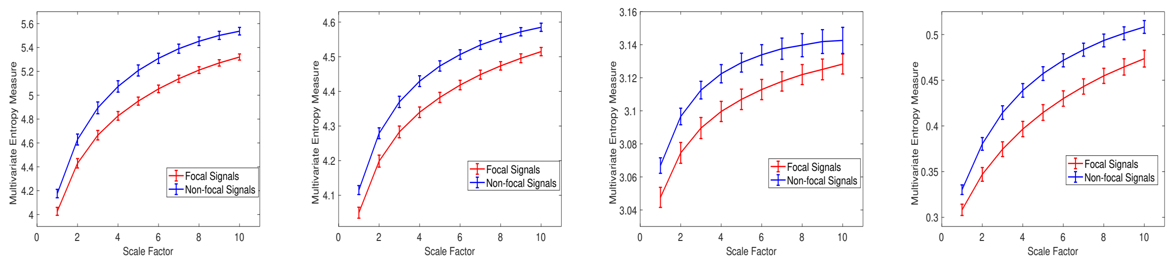
Table 4. CV values of the entropy results at scale factor 4 using $\mathrm{mvMDE_{III}}$, mvMDE, and mvMFE for self-paced walk (SPW) vs. metronomically-paced walk (MPW).

Stride interval fluctuations	mvMFE	$\mathrm{mvMDE_{III}}$	mvMDE
Self-paced walk	0.040	0.005	0.002
Metronomically-paced walk	0.116	0.025	0.019

1) Dataset of Stride Interval Fluctuations: For the self-paced versus metronomically-paced stride interval fluctuations, the results obtained by the $\mathrm{mvMDE_{III}}$, mvMDE, and mvMFE, respectively depicted in Fig. 6(a), (b), and (c), show that the self-paced unconstrained walk's fluctuations have more complexity and greater long-range correlations than the metronomically-paced walk's series, in agreement with those reportred in [2]. We did not use $\mathrm{mvMDE_{II}}$, as the signals do not follow the typical number of samples required for $\mathrm{mvMDE_{II}}$. To compare the results, the CV values for both the metronomically- and self-paced walk (MPW and SPW) at scale factor 4, as a trade-off between the long and short scales, are shown in Table IV. The CV values for the $\mathrm{mvMDE_{III}}$- and mvMDE-based profiles are smaller than those for mvMFE, showing the superiority of the proposed methods over mvMFE in terms of the stability of results. The smallest CV values are achieved by the mvMDE.

2) Dataset of Focal and Non-focal Brain Activity: For the focal and non-focal EEG recordings, the results obtained by $\mathrm{mvMDE_{II}}$, $\mathrm{mvMDE_{III}}$, mvMDE, and mvMFE, respectively depicted in Fig. 7(a), (b), (c), and (d), show that the focal time series are less complex than the non-focal ones, in agreement with previous studies [40][43]. The CV values for the focal- and non-focal-based results at scale 6 are shown in Table V. All the mvMDE-based CV values are smaller than those using mvMFE, showing more stability of the results obtained by the proposed methods. Moreover, the CV values for mvMDE are smaller than those for $\mathrm{mvMDE_{III}}$, and the latter ones are smaller than those for $\mathrm{mvMDE_{II}}$, suggesting that the mvMDE leads to more stable profiles.

3) Surface MEG Recordings in Alzheimer's Disease: To assess the ability of mvMDE, in comparison with mvMFE, we applied the methods to the 148-channel MEG signals to discriminate AD patients from



(a) $\mathrm{mvMDE_{II}}$ (b) $\mathrm{mvMDE_{III}}$ (c) mvMDE (d) mvMFE

Figure 7. Mean value and SD of the results using (a) $\mathrm{mvMDE_{II}}$, (b) $\mathrm{mvMDE_{III}}$, (c) mvMDE, and (d) mvMFE for focal vs. non-focal time series.

Table 5. CV values of the entropy results at scale factor 6 using $mvMDE_{II}$, $mvMDE_{III}$, mvMDE, mvMSE, and mvMFE for focal vs. non-focal EEG recordings.

Signals	mvMSE	mvMFE	$mvMDE_{II}$	$mvMDE_{III}$	mvMDE
focal EEGs	0.019	0.019	0.006	0.003	0.002
Non-focal EEGs	0.021	0.015	0.008	0.003	0.002

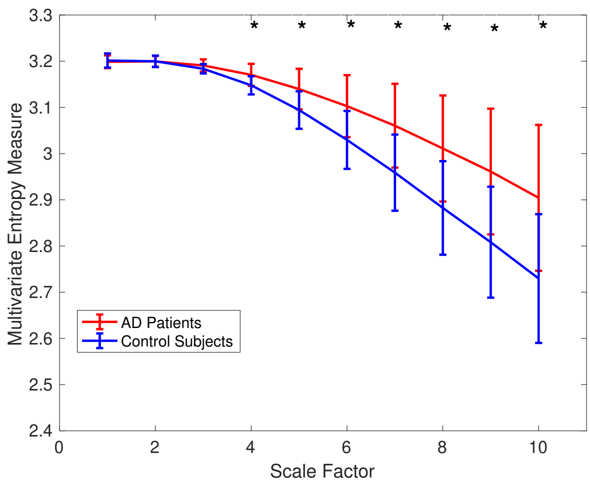


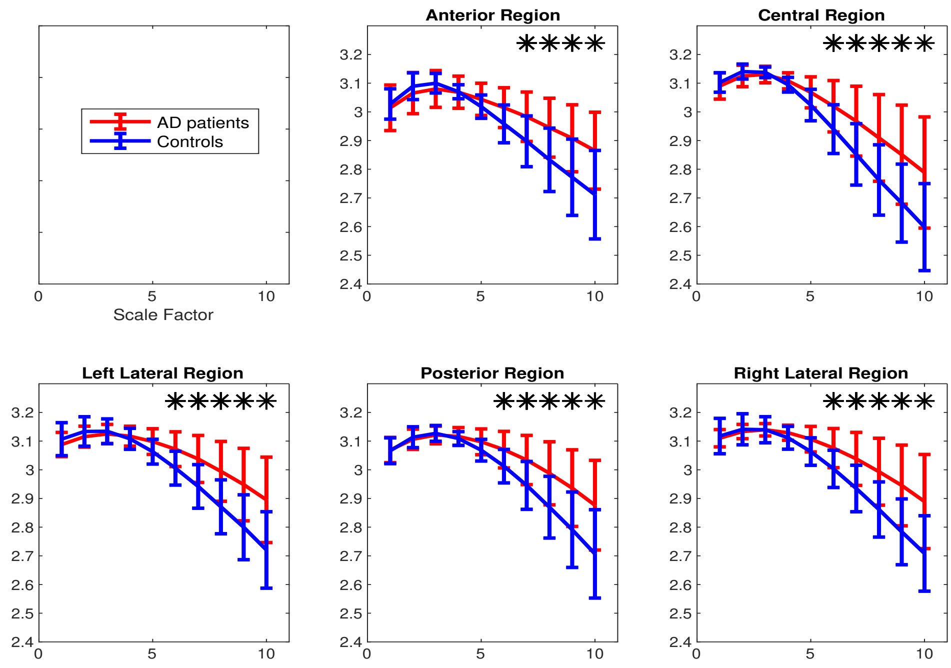
Figure 8. Mean value and SD of the results obtained by mvMDE computed from 36 AD patients versus 26 elderly controls for all the 148 channels. Red and blue respectively indicate AD patients and controls. The scales with p-values smaller than 0.001 are shown with *.

controls. Because mvMFE needs to store a huge number of elements for a signal with a large number of channels, mvMFE was not able to simultaneously analyse all 148 time series. However, the results using mvMDE are depicted in Fig. 8. It represents an advantage of mvMDE over mvMFE for signals with a large number of channels. To compare the mvMFE and mvMDE, we applied the methods to five main scalp regions, namely, anterior (17 channels), right (34 channels) and left lateral (34 channels), central (29 channels), and posterior (34 channels) areas, leading to the smaller number of channels to noticeably decrease the number of elements stored by the use of the mvMFE algorithm.

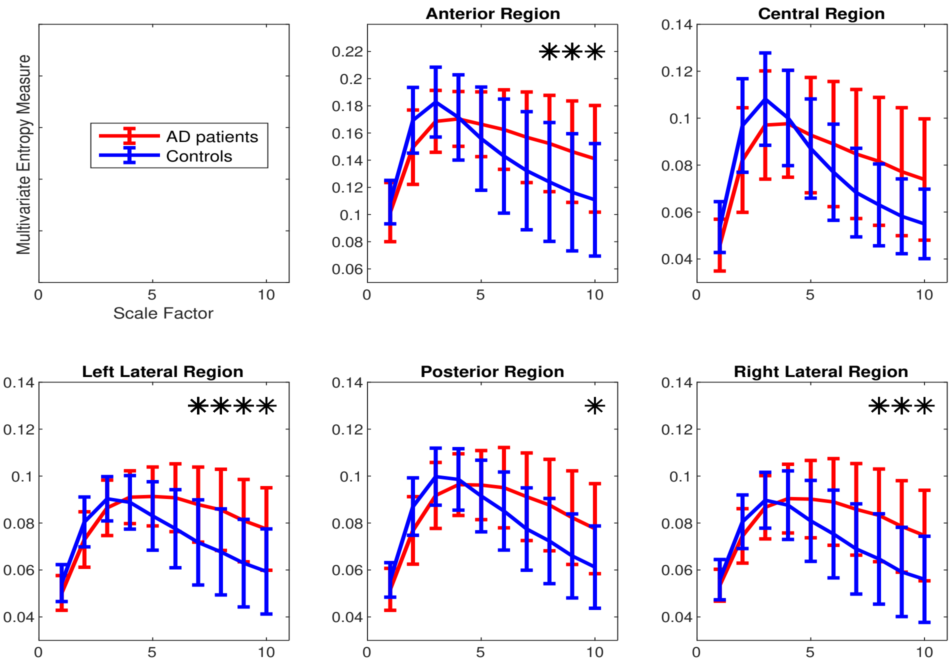
The average and SD of mvMDE and mvMFE values for five regions are respectively shown in Fig. 9(a) and 9(b). As can be seen in Fig. 8 and Fig. 9, the average mvMDE and mvMFE values for AD patients are smaller than those for controls at lower scale factors (short-time scale factors), while at higher scales, the AD subjects' recordings have larger entropy values (long-time scale factors) for both the mvMFE and mvMDE, in agreement with [21,44,45]. Because the larger the number of channels, the smaller the mvMSE and similarly mvMFE values [21], the entropy values for anterior region are larger than those for the other four regions. It is worth noting that we only use mvMDE, as the signals do not follow the typical number of samples required for $mvMDE_{II}$ and $mvMDE_{III}$.

The Mann-Whitney U-test was used to assess the differences between the mvMDE and mvMFE profiles at each temporal scale for AD patients versus controls, because the mvMDE and mvMFE values at each scale factor did not follow a normal distribution. The temporal scales with p-values smaller than 0.001 are shown with * in Fig. 8 and Fig. 9. The p-values show that the mvMDE, compared with the mvMFE, significantly discriminated the controls from subjects with AD at a larger number of scale factors. Moreover, the smallest p-value was achieved by the mvMDE, evidencing the superiority of mvMDE over mvMFE.

The Hedges' g effect size [46] was also used to quantify the differences between the entropy values for the AD patients' vs. healthy controls' MEGs for the five main brain regions [47]. The Hedges' g test shows the difference between the means of two groups, divided by the weighted average of standard



(a) mvMDE



(b) mvMFE

Figure 9. Mean value and SD of the results obtained by (a) mvMDE and (b) mvMFE computed from 36 AD patients versus 26 elderly age-matched controls over five scalp regions. Red and blue indicate AD patients and controls, respectively. The scale factors with p-values smaller than 0.001 are shown with *.

Table 6. Differences between results for AD patients' vs. healthy controls' MEGs obtained by mvMFE and mvMDE for five main brain regions based on the Hedges' g effect size.

Region - method	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6	Scale 7	Scale 8	Scale 9	Scale 10
Anterior - mvMFE	0.36	0.73	0.57	0.04	0.33	0.53	0.63	0.70	0.72	0.73
Central - mvMFE	0.68	0.67	0.49	0.10	0.23	0.48	0.65	0.76	0.79	0.83
Left lateral - mvMFE	0.53	0.64	0.34	0.18	0.60	0.83	0.92	0.98	0.97	0.98
Posterior - mvMFE	0.46	0.72	0.58	0.16	0.30	0.57	0.73	0.78	0.82	0.85
Right lateral - mvMFE	0.30	0.50	0.22	0.18	0.53	0.71	0.84	0.92	0.97	0.95
Anterior - mvMDE	0.18	0.37	0.36	0.03	0.49	0.80	0.95	1.02	1.06	1.04
Central - mvMDE	0.29	0.45	0.29	0.48	0.78	0.88	0.97	1.01	1.03	1.04
Left lateral - mvMDE	0.37	0.40	0.24	0.24	0.77	1.07	1.17	1.20	1.19	1.19
Posterior - mvMDE	0.05	0.19	0.18	0.24	0.67	0.90	1.015	1.05	1.06	1.06
Right lateral - mvMDE	0.15	0.19	0.00	0.51	0.90	1.05	1.14	1.18	1.20	1.16

deviations for these two groups. The differences, illustrated in Table 6, show that the highest effect size is obtained by mvMDE, showing the advantage of this method over mvMFE.

On the whole, the profiles for the real datasets evidence the advantage of mvMDE$_{\mathrm{II}}$, mvMDE$_{\mathrm{III}}$, and mvMDE over mvMFE to discriminate different types of dynamics of multi-channel signals as well as the superiority of mvMDE over mvMFE in terms of ability to discriminate various dynamics of time series, computational time, and memory cost. As mentioned before, mvMPE does not consider the spatial domain. We have also refined the mvMPE [19] on the basis of mvMDE$_{\mathrm{II}}$, mvMDE$_{\mathrm{III}}$, and mvMDE. These approaches have the following advantages over the first version of mvMPE [19]: 1) they take into account both the spatial and time domains; 2) their results were more stable than the mvMPE-based ones; and 3) better distinguished different dynamics of multivariate signals. However, since the mvMDE methods are considerably faster, result in more stable profiles, and lead to larger differences between physiological conditions of recordings, for simplicity, we did not report the mvMPE-based results.

In this article, we proposed four implementations of the mvDE methods combined with the most commonly used coarse-graining process [3,8,17]. The key contribution of this study was introducing the mvDE methods. The alternative coarse-graining processes based on multivariate empirical mode decomposition [2,28,48–50], and FIR filters [28,51], though out of the scope of this paper, can be employed instead of the classical implementation of coarse-graining process used herein.

Our future study will aim at proposing the refined composite mvMDE (RCmvMDE) approaches according to [17]. Moreover, we will explore the mvMDE and RCmvMDE on other physiological and non-physiological time series. The similarity of two multi-channel signals based on mvMDE and cross-entropy [11] can also be developed as future work. An important step in making mvMDE a useful and stable metric is the mapping of the data to discrete set of integers via the normal cumulative distribution. Other mapping functions are available in [30]. The mvMDE method and its univariate form can also be generalized based on Renyi entropy [52].

5. Conclusions

To quantify the complexity of multivariate time series, we built four diverse alternative implementations of mvMDE as further developments of our recently introduced MDE [27]. These insights help towards a comprehensive understanding of four strategies to extend a univariate-based entropy method to its multivariate versions and therefore, provide invaluable information for future studies on multivariate time series. Although mvMDE was the best algorithm in terms of ability to discriminate dynamics of multivariate signals, computational time, and memory cost, the simpler alternatives (mvDE$_{\mathrm{I}}$ to mvDE$_{\mathrm{III}}$) may still be useful in some settings.

We assessed their performance on the correlated and uncorrelated multivariate noise signals, the bivariate AR time series, and three physiological datasets. The results showed the similar behavior of mvMSE-, mvMFE-, and mvMDE-based profiles. However, mvMDE had the following advantages over the existing methods: 1) it was faster than the existing methods; 2) mvMDE, in comparison with mvMSE and mvMFE, resulted in more stable profiles; 3) mvMDE better discriminated different kinds of biomedical signals; 4) for short multivariate time series (300 sample points), mvMDE did not result in undefined values; and 5) mvMDE, compared with mvMSE and mvMFE, needed to store a considerably smaller number of elements.

Overall, we expect the mvMDE approach to play a key role in the assessment of complexity in multivariate time series.

Funding: This research received no external funding.

Acknowledgments: The MATLAB code of the mvMDE techniques will be made publicly-available upon publication.

Conflicts of Interest: The authors declare no conflict of interest.

1. Cerutti, S.; Hoyer, D.; Voss, A. Multiscale, multiorgan and multivariate complexity analyses of cardiovascular regulation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **2009**, *367*, 1337–1358.
2. Ahmed, M.; Rehman, N.; Looney, D.; Rutkowski, T.; Mandic, D. Dynamical complexity of human responses: A multivariate data-adaptive framework. *Bulletin of the Polish Academy of Sciences: Technical Sciences* **2012**, *60*, 433–445.
3. Ahmed, M.U.; Mandic, D.P. Multivariate multiscale entropy analysis. *IEEE Signal Processing Letters* **2012**, *19*, 91–94.
4. Cerutti, S. Multivariate and multiscale analysis of biomedical signals: Towards a comprehensive approach to medical diagnosis. Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on. IEEE, 2012, pp. 1–5.
5. Fernández-Sotos, A.; Martínez-Rodrigo, A.; Moncho-Bogani, J.; Latorre, J.M.; Fernández-Caballero, A. NEURAL CORRELATES OF PHRASE QUADRATURE PERCEPTION IN HARMONIC RHYTHM: AN EEG STUDY (USING A BRAIN-COMPUTER INTERFACE). *International Journal of Neural Systems* **2018**, *28*, 1750054.
6. Spyrou, L.; Martín-Lopez, D.; Valentín, A.; Alarcón, G.; Sanei, S. Detection of intracranial signatures of interictal epileptiform discharges from concurrent scalp EEG. *International Journal of Neural Systems* **2016**, *26*, 1650016.
7. Pereda, E.; Quiroga, R.Q.; Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology* **2005**, *77*, 1–37.
8. Ahmed, M.U.; Mandic, D.P. Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Physical Review E* **2011**, *84*, 061918.
9. Mammone, N.; Bonanno, L.; Salvo, S.D.; Marino, S.; Bramanti, P.; Bramanti, A.; Morabito, F.C. Permutation disalignment index as an indirect, EEG-based, measure of brain connectivity in MCI and AD patients. *International Journal of Neural Systems* **2017**, *27*, 1750020.
10. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Physical review E* **2005**, *71*, 021906.
11. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* **2000**, *278*, H2039–H2049.
12. Bandt, C.; Pompe, B. Permutation entropy: a natural complexity measure for time series. *Physical review letters* **2002**, *88*, 174102.
13. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time series analysis. *IEEE Signal Processing Letters* **2016**, *23*, 610–614.
14. Fogedby, H.C. On the phase space approach to complexity. *Journal of statistical physics* **1992**, *69*, 411–425.
15. Silva, L.E.V.; Cabella, B.C.T.; da Costa Neves, U.P.; Junior, L.O.M. Multiscale entropy-based methods for heart rate variability complexity analysis. *Physica A: Statistical Mechanics and its Applications* **2015**, *422*, 143–152.
16. Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **2015**, *17*, 3110–3123.

17. Azami, H.; Escudero, J. Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals. *Physica A: Statistical Mechanics and its Applications* **2017**, *465*, 261–276.
18. Li, P.; Ji, L.; Yan, C.; Li, K.; Liu, C.; Liu, C. Coupling between short-term heart rate and diastolic period is reduced in heart failure patients as indicated by multivariate entropy analysis. Computing in Cardiology Conference (CinC), 2014. IEEE, 2014, pp. 97–100.
19. Morabito, F.C.; Labate, D.; La Foresta, F.; Bramanti, A.; Morabito, G.; Palamara, I. Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy* **2012**, *14*, 1186–1202.
20. Yin, Y.; Shang, P. Multivariate weighted multiscale permutation entropy for complex time series. *Nonlinear Dynamics* **2017**, pp. 1–16.
21. Azami, H.; Abásolo, D.; Simons, S.; Escudero, J. Univariate and Multivariate Generalized Multiscale Entropy to Characterise EEG Signals in Alzheimer's Disease. *Entropy* **2017**, *19*, 31.
22. Labate, D.; La Foresta, F.; Morabito, G.; Palamara, I.; Morabito, F.C. Entropic measures of EEG complexity in alzheimer's disease through a multivariate multiscale approach. *Sensors Journal, IEEE* **2013**, *13*, 3284–3292.
23. Gao, Z.K.; Ding, M.S.; Geng, H.; Jin, N.D. Multivariate multiscale entropy analysis of horizontal oil–water two-phase flow. *Physica A: Statistical Mechanics and Its Applications* **2015**, *417*, 7–17.
24. Wei, Q.; Liu, D.H.; Wang, K.H.; Liu, Q.; Abbod, M.F.; Jiang, B.C.; Chen, K.P.; Wu, C.; Shieh, J.S. Multivariate multiscale entropy applied to center of pressure signals analysis: an effect of vibration stimulation of shoes. *Entropy* **2012**, *14*, 2157–2172.
25. Zhao, L.; Wei, S.; Tang, H.; Liu, C. Multivariable Fuzzy Measure Entropy Analysis for Heart Rate Variability and Heart Sound Amplitude Variability. *Entropy* **2016**, *18*, 430.
26. Ramdani, S.; Bonnet, V.; Tallon, G.; Lagarde, J.; Bernard, P.L.; Blain, H. Parameters Selection for Bivariate Multiscale Entropy Analysis of Postural Fluctuations in Fallers and Non-Fallers Older Adults. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2016**, *24*, 859–871.
27. Azami, H.; Rostaghi, M.; Abasolo, D.; Escudero, J. Refined Composite Multiscale Dispersion Entropy and its Application to Biomedical Signals. *IEEE Transactions on Biomedical Engineering* **2017**, *64*, 2872–2879.
28. Azami, H.; Escudero, J. Coarse-Graining Approaches in Univariate Multiscale Sample and Dispersion Entropy. *Entropy* **2018**, *20*, 138.
29. Azami, H.; Kinney-lang, E.; Ebied, A.; Fernández, A.; Escudero, J. Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer's disease. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 3182–3185.
30. Azami, H.; Escudero, J. Amplitude-and Fluctuation-Based Dispersion Entropy. *Entropy* **2018**, *20*, 210.
31. Tufféry, S. *Data mining and statistics for decision making*; Vol. 2, Wiley Chichester, 2011.
32. Baranwal, G.; Vidyarthi, D.P. Admission control in cloud computing using game theory. *The Journal of Supercomputing* **2016**, *72*, 317–346.
33. Gibbs, M.N.; MacKay, D.J. Variational Gaussian process classifiers. *IEEE Transactions on Neural Networks* **2000**, *11*, 1458–1464.
34. Duch, W. Uncertainty of data, fuzzy membership functions, and multilayer perceptrons. *IEEE Transactions on Neural Networks* **2005**, *16*, 10–23.
35. Cao, L.; Mees, A.; Judd, K. Dynamics from multivariate time series. *Physica D: Nonlinear Phenomena* **1998**, *121*, 75–88.
36. Kaffashi, F.; Foglyano, R.; Wilson, C.G.; Loparo, K.A. The effect of time delay on approximate & sample entropy calculations. *Physica D: Nonlinear Phenomena* **2008**, *237*, 3069–3074.
37. Humeau-Heurtier, A. Multivariate generalized multiscale entropy analysis. *Entropy* **2016**, *18*, 411.
38. Penny, W.; Roberts, S. Bayesian multivariate autoregressive models with structured priors. *IEE Proceedings-Vision, Image and Signal Processing* **2002**, *149*, 33–41.
39. Hausdorff, J.M.; Purdon, P.L.; Peng, C.; Ladin, Z.; Wei, J.Y.; Goldberger, A.L. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *Journal of applied physiology* **1996**, *80*, 1448–1457.
40. Andrzejak, R.G.; Schindler, K.; Rummel, C. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. *Physical Review E* **2012**, *86*, 046206.

41. Escudero, J.; Sanei, S.; Jarchi, D.; Abásolo, D.; Hornero, R. Regional coherence evaluation in mild cognitive impairment and Alzheimer's disease based on adaptively extracted magnetoencephalogram rhythms. *Physiological measurement* **2011**, *32*, 1163.
42. Lake, D.E.; Moorman, J.R. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *American Journal of Physiology-Heart and Circulatory Physiology* **2010**, *300*, H319–H325.
43. Sharma, R.; Pachori, R.B.; Acharya, U.R. Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals. *Entropy* **2015**, *17*, 669–691.
44. Yang, A.C.; Wang, S.J.; Lai, K.L.; Tsai, C.F.; Yang, C.H.; Hwang, J.P.; Lo, M.T.; Huang, N.E.; Peng, C.K.; Fuh, J.L. Cognitive and neuropsychiatric correlates of EEG dynamic complexity in patients with Alzheimer's disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **2013**, *47*, 52–61.
45. Hornero, R.; Abásolo, D.; Escudero, J.; Gómez, C. Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **2009**, *367*, 317–336.
46. Rosenthal, R.; Cooper, H.; Hedges, L. Parametric measures of effect size. *The handbook of research synthesis* **1994**, *621*, 231–244.
47. Sullivan, G.M.; Feinn, R. Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education* **2012**, *4*, 279–282.
48. Hu, M.; Liang, H. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Transactions on Biomedical Engineering* **2012**, *59*, 12–15.
49. Hu, M.; Liang, H. Perceptual suppression revealed by adaptive multi-scale entropy analysis of local field potential in monkey visual cortex. *International Journal of Neural Systems* **2013**, *23*, 1350005.
50. Tonoyan, Y.; Looney, D.; Mandic, D.P.; Van Hulle, M.M. Discriminating multiple emotional states from EEG using a data-adaptive, multiscale information-theoretic approach. *International Journal of Neural Systems* **2016**, *26*, 1650005.
51. Valencia, J.F.; Porta, A.; Vallverdu, M.; Claria, F.; Baranowski, R.; Orlowska-Baranowska, E.; Caminal, P. Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects. *IEEE Transactions on Biomedical Engineering* **2009**, *56*, 2202–2213.
52. Renner, R.; Wolf, S. Smooth Rényi entropy and applications. International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings. IEEE, 2004, p. 233.