MULTIVARIATE ENTROPY ANALYSIS WITH DATA-DRIVEN SCALES

M. U. Ahmed[*], *N. Rehman*[*], *D. Looney*[*], *T. M. Rutkowski*[†], *P. Kidmose*[‡], *D. P. Mandic*[*]

[*]Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK
{mosabber.ahmed05, naveed.rehman07, david.looney06, d.mandic}@imperial.ac.uk
[†]University of Tsukuba & RIKEN Brain Science Institute, Japan
tomek@tara.tsukuba.ac.jp
[‡]Aarhus School of Engineering, 8000 Aarhus C, Denmark
pki@iha.dk

ABSTRACT

A data-adaptive algorithm for the entropy-based analysis of structural regularities (complexity) in multivariate signals is proposed. This is achieved by combining multivariate sample entropy with a multivariate extension of empirical mode decomposition, both data-driven multiscale techniques. The proposed analysis across data-adaptive scales makes the approach robust to nonstationarity, a critical issue with information theoretic measures. Simulations on synthetic and real-world physiological data support the approach and validate the hypothesis of increased complexity for unconstrained as compared to constrained (due to e.g. ageing or illness) biological systems.

Index Terms— Multivariate sample entropy, multivariate empirical mode decomposition, multivariate multiscale entropy, dynamical complexity, complexity of physiological data

1. INTRODUCTION

A fundamental goal of statistical signal processing is the estimation of robust descriptors which can fully characterize the underlying generating mechanisms of real-world systems from their observed time series. Examples include complexity, local predictability, irregularity, self-similarity and synchrony. To extend the scope of time-delay embedded reconstruction in order to cater for the dynamics of a system across its different time-scales, Costa *et al.* introduced the multiscale entropy (MSE) that provides a measure of complexity by performing multiple coarse-graining operations on the data and calculating the sample entropy for each scale [1].

While MSE has been successfully applied to distinguish between structural dynamics of different real-world physiological time series based on their complexity [1, 2, 3], the method also gives opportunity for further improvements. Firstly, the algorithm can only cater for univariate time series, which limits its practical use as biological systems are often multivariate in nature. To this end, we have recently introduced multivariate sample entropy [4, 5], which examines complexity both across time and data channels to reveal the extent of the long-range spatio-temporal correlations present in multivariate signals.

A second limitation of the MSE algorithm, inherited also by the multivariate extension [4, 5], is the way in which the data scales are generated. Coarse graining, that is, signal averaging over non-overlapping segments of increasing length, is unsuitable for the analysis of high frequency components and also results in aliasing, causing artifacts. In [6] it was proposed to use a bank of Butterworth filters to circumvent the aliasing problem, however such an approach requires the *a priori* selection of filter parameters, making the analysis critically sensitive to slight changes in experimental conditions. It is therefore desirable to make the complexity analysis data-adaptive, so that it can be conducted across scales that occur naturally, as defined by the data.

Empirical mode decomposition (EMD) is a data-driven method that decomposes a given time series into a set of oscillatory functions (time-scales), known as intrinsic mode functions (IMFs). Unlike projection-based methods, such as those based on Fourier and wavelet theory, EMD obtains the oscillatory modes (scales) adaptively and considers the signal dynamics at the 'local' level, making it a natural choice for generating the data-scales required for entropy-based analysis [7]. We have recently developed a multivariate extension of EMD (MEMD) to process multichannel data [8] which, crucially, aligns the decomposed components from different channels in similar frequency bands, a prerequisite for enabling a scale-by-scale analysis. Recent work [9] has combined the MEMD algorithm with univariate sample entropy estimation, whereby MEMD was used to examine dynamics across multiple trials, it was however limited to single-channel analysis and did not exploit the full potential of MEMD - that for direct multichannel modelling.

In this paper, we propose to use multivariate EMD in combination with the multivariate sample entropy statistic [4, 5] to develop a complete and robust framework for an entropy-based complexity analysis of multivariate data. Due to the mode alignment property of MEMD [10], the scales generated for each data channel are aligned in frequency, which makes the comparison across the data channels meaningful.

2. MULTIVARIATE EMPIRICAL MODE DECOMPOSITION

The empirical mode decomposition (EMD) algorithm was developed as an adaptive approach to time-frequency analysis [11]. The elements (basis functions) of the decomposition, the so-called intrinsic mode functions (IMFs), are by design monocomponent (narrowband) and reflect the underlying intrinsic time-scales within a time series. Unlike coarse graining which is effectively a low pass filter with decreasing bandwidth, its adaptive nature makes it suitable for the analysis of nonlinear and nonstationary data that comprise high frequency signal components. An example of an EMD-defined time scale is given in Fig. 1. Note that both the frequency and amplitude information are well defined locally.

In its original formulation, the EMD algorithm is univariate, that is, it can only process single channel data. To cater for the multi-channel nature of real-world systems, the multivariate empirical mode decomposition (MEMD) algorithm [8] has been recently introduced, which is a generic extension of EMD to process arbitrary numbers of data channels simultaneously. A key feature of the algorithm is the way in which the local mean of the signal is generated. MEMD operates by taking multiple signal projections[1], and estimating the envelopes for each, the average of which defines the local mean. The local mean is recursively subtracted from the signal to obtain the high frequency 'detail' until IMF conditions are satisfied (zero mean, equal number of local minima and maxima) - defining the so called sifting process by which the first IMF is obtained. This process is repeated to extract the next IMFs until the residual signal only contains the low-frequency trend. Thus, for a given p-variate time series, $\mathbf{x}(t)$, its J IMFs are given by

$$\mathbf{x}(t) = \sum_{j=1}^{J} \mathbf{c}_j(t) \tag{1}$$

where, for instance, symbols $\mathbf{c}_1(t), \mathbf{c}_2(t), \ldots,$ denote the first extracted IMFs, which define the high frequency time scales, and the lower index IMFs, $\mathbf{c}_{J-1}(t), \mathbf{c}_J(t)$, which define the low frequency time scales. The advantages offered by MEMD over univariate (single-channel) EMD are:

1. Direct processing of multichannel data via MEMD produces the same number of IMFs for all data channels allowing their comparison at each scale, independently.
2. MEMD automatically *aligns* common scales, present across multiple channels, in multivariate IMFs; a desirable property hard to achieve by applying univariate EMD channel-wise on multivariate data.

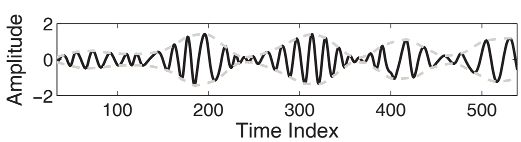


Fig. 1. Data scale obtained using EMD.

3. MULTIVARIATE SAMPLE ENTROPY

To calculate multivariate sample entropy (MSampEn) [4, 5], recall from multivariate embedding theory, that for a p-variate time series $\{x_{k,i}\}_{i=1}^{N}, k = 1, 2, \ldots, p$, observed through p measurement functions $h_k(y_i)$, the multivariate embedded reconstruction is based on the composite delay vector

$$X_m(i) = [x_{1,i}, x_{1,i+\tau_1}, \ldots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, x_{2,i+\tau_2}, \ldots, x_{2,i+(m_2-1)\tau_2}, \ldots, x_{p,i}, x_{p,i+\tau_p}, \ldots, x_{p,i+(m_p-1)\tau_p}],$$

where $\mathbf{M} = [m_1, m_2, \ldots, m_p] \in \mathbb{R}^p$ is the embedding vector, $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_p]$ the time lag vector, the composite delay vector $X_m(i) \in \mathbb{R}^m$ (where $m = \sum_{k=1}^{p} m_k$); the multivariate sample entropy (MSampEn) is calculated in Algorithm 1.

[1]The direction vectors for an $n-$sphere (an extension of the ordinary sphere to an arbitrary dimension) are generated by low-discrepancy quasi-Monte Carlo sequences [8].

Algorithm 1 Multivariate sample entropy (MSampEn)

1: Form $(N - \delta)$ composite delay vectors $X_m(i) \in \mathbb{R}^m$, where $i = 1, 2, \ldots, N - \delta$ and $\delta = \max\{\mathbf{M}\} \times \max\{\boldsymbol{\tau}\}$ and define the distance between any two vectors $X_m(i)$ and $X_m(j)$ as the maximum norm;
2: For a given composite delay vector $X_m(i)$ and a threshold r, count the number of instances P_i for which $d[X_m(i), X_m(j)] \leq r$, $j \neq i$, then calculate the frequency of occurrence, $B_i^m(r) = \frac{1}{N-\delta-1} P_i$, and define $B^m(r) = \frac{1}{N-\delta} \sum_{i=1}^{N-\delta} B_i^m(r)$;
3: Increase $m_k \to (m_k + 1)$ for a specific variable k, keeping the dimension of the other variables unchanged. Thus, a total of $p \times (N - \delta)$ vectors $X_{m+1}(i)$ in $\mathbb{R}^{m+1}$ are obtained;
4: For a given $X_{m+1}(i)$, calculate the number of vectors Q_i, such that $d[X_{m+1}(i), X_{m+1}(j)] \leq r$, where $j \neq i$, then calculate the frequency of occurrence, $B_i^{m+1}(r) = \frac{1}{p(N-\delta)-1} Q_i$, and define $B^{m+1}(r) = \frac{1}{p(N-\delta)} \sum_{i=1}^{p(N-\delta)} B_i^{m+1}(r)$;
5: Finally, for a tolerance level r, estimate $MSampEn$ as

$$MSampEn(\mathbf{M}, \boldsymbol{\tau}, r, N) = -ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right]. \tag{2}$$

4. MEMD-BASED MULTIVARIATE MSE

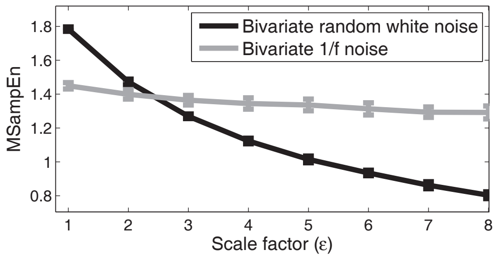
We propose to use MEMD to generate multiple data driven, intrinsic, temporal scales for a given multivariate data, and subsequently perform multivariate entropy analysis on so generated cumulative IMFs (scales). For this cause, fully aligned scales from input multivariate data are first obtained by applying MEMD both *across multiple channels and multiple trials of the input data*. Next, multivariate sample entropy estimates are calculated for the so-defined 'scales' of the multivariate input data to reveal the long-range correlation structure. The proposed algorithm is shown in Algorithm 2.

Remark 1 . *Note that, unlike other EMD/MEMD based sample entropy methods [7] [9] which employ univariate sample entropy, the proposed method is fully multivariate as it calculates multivariate sample entropy estimates, thereby, catering for linear/nonlinear correlations between channels.*

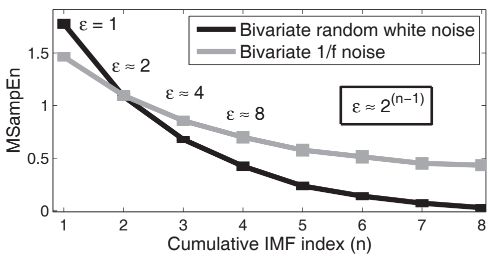
Algorithm 2 MEMD-based multivariate multiscale entropy

1: Generate multiple scales from J IMFs obtained by applying MEMD to a given multivariate time series $\{x_{k,i}\}_{i=1}^{N}$for $k = 1, 2, \ldots, p$, where p denotes the total number of variates (channels) and N represents the total number of samples in each variate which does not change across MEMD-based scales.
2: Define data-driven 'scales' of x as the cumulative sum of IMFs either by $\mathbf{c}_n = \sum_{j=n}^{J} \mathbf{c}_j$ (Approach 1) or by $\mathbf{c}_n = \sum_{j=1}^{J-n+1} \mathbf{c}_j$ (Approach 2), where $n \in [1, J]$ denotes the cumulative IMF index. *Only Approach 1 is used in the sequel*.
3: Calculate and plot multivariate sample entropy measure, given in (2), for each scale n.

To illustrate the performance of the proposed method, it was applied to a synthetically generated bivariate white noise and bivariate $1/f$ noise. The $1/f$ noise possesses long-range correlations and its standard entropy (at scale 1) is lower than that of white noise, however, the $1/f$ noise is structurally complex whereas the bivariate white noise is not, and any complexity measure should be higher for $1/f$ noise at increasing scales. Observe from Fig. 2(b) that though bivariate white noise has higher complexity than $1/f$ noise for the first scale, the complexity becomes lower than $1/f$ noise for higher scales. This example on synthetic data illustrates, that by design, $1/f$ noise is structurally more complex than uncorrelated random noise, a result consistent with standard MSE/MMSE [1, 4, 5] as shown in Fig. 2(a).



(a) Multivariate sample entropy for scales using coarse graining



(b) Multivariate sample entropy for MEMD-based scales

Fig. 2. MMSE analysis for bivariate white and $1/f$ noise: (a) using coarse graining and (b) MEMD-based scales. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

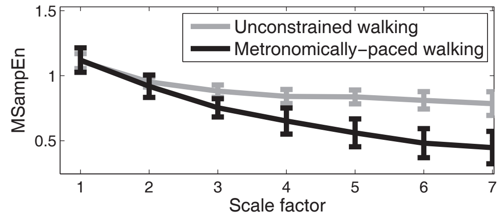
Remark 2 . *A direct comparison is often not possible between the scales of MEMD-based MSE and those of standard MSE as, by design, the frequency ranges of the cumulative IMFs adapts to the data. In the case of white noise, however, the dyadic filter bank property of MEMD is well known [10]. Disregarding elements of coarse graining*[2]*, the averaging operation at scale ϵ is equivalent to low pass filtering with a cutoff frequency (normalised) of $f_c = 0.5/\epsilon$. Thus for the nth cumulative IMF index (Approach 1) of white noise, the equivalent scale factor is given by $\epsilon \approx 2^{n-1}$. For insight, the equivalent scale factors for white noise are shown for cumulative IMF indexes in Fig. 2(b).*

5. SIMULATION RESULTS

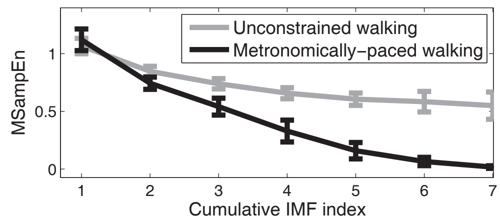
5.1. Gait analysis: complexity change with constraints

Stride intervals from human gait [12] data were analysed with the aim of revealing long-range correlations, a signature that suggests cooperation within the different bodily subsystems at different time scales. Stride interval fluctuations were recorded from ten healthy subjects who walked for 1 hour at normal, slow, and fast paces. The participants were further asked to walk following a metronome which was set to each participant's mean stride interval [12].

[2]The filtering operation equivalent to coarse graining is characterised by a very slow roll-off as well as large sidelobes which introduce aliasing artifacts [6]. The equivalent relationship between scale factor and cumulative IMF index given in the paper assumes a considerably faster roll-off as well as the absence sidelobes.



(a) Multivariate sample entropy for scales using coarse graining



(b) Multivariate sample entropy for MEMD-based scales

Fig. 3. MMSE analysis for self-paced vs metronomically-paced stride interval time series: (a) using coarse graining and (b) MEMD-based scales. The curves represent an average over 10 subjects, and the error bars the SD.

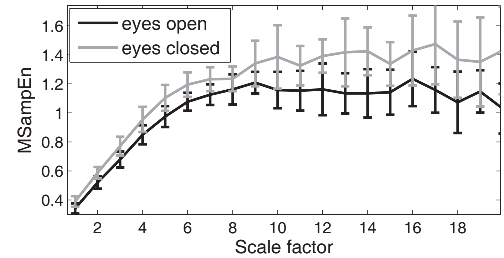
We considered the three walking conditions (from the data available from [12]) as different variables from the same system, and calculated MSampEn for different scales (cumulative IMFs) generated from MEMD and in this way were able to discriminate between the 'self-paced' and 'metronomically-paced' walk. The values of the parameters used to calculate MSampEn were $m_k = 2$, $\tau_k = 1$ and $r = 0.15\times$(standard deviation of the normalized time series) for each data channel; these parameters were chosen on the basis of previous studies indicating good statistical reproducibility for SampEn [3].

Fig. 3(a) shows the results obtained by the standard coarse-graining based MMSE method and Fig. 3(b) for the proposed MEMD-based method. Both methods found that self-paced 'unconstrained' walk has higher complexity, and thereby long-range correlations, than constrained 'metronomically-paced' walk. To evaluate the statistical difference of the entropy statistics of self-paced and metronomically-paced sets, the Student's t-test and the Mann-Whitney U test were applied. Both these tests revealed significant differences ($p < 0.01$) at all IMF-defined scales except the first two for standard coarse graining as well as MEMD-based MMSE method. Note that the first scale corresponds to the raw signal and MSampEn measures cannot discriminate between self-paced and metronomically-paced walk in either method. Moreover, as desired the separation between the MMSE curves of unconstrained and metronomically-paced walk is higher for the MEMD-based method (Fig. 3(b)), as indicated by much smaller error bars. Thus, using cumulative IMFs as data-adaptive scales offers a significant improvement over the coarse-graining based MMSE. These results also support the more general concept of multiscale complexity loss with ageing and disease or when a system is under constraints (metronomically-paced walk), which all reduce the adaptive capacity of biological organization at all levels [13].

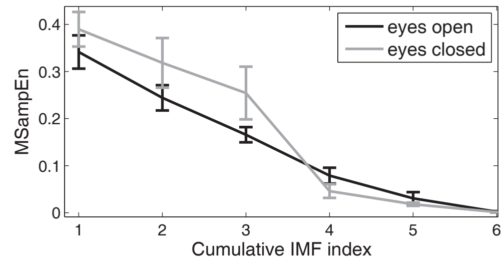
5.2. Structural complexity of different brain states

The proposed algorithm was next applied to multivariate electroencephalogram (EEG) signals to establish whether the differences in multichannel complexity can indicate changes in brain states. As a proof of concept we considered the well understood alpha-attenuation paradigm. Namely, closing of the eyes causes an increase in alpha activity (8 - 12 Hz) in the spectrum of EEG, a well known response that is closely linked to the degree of alertness. Seven 3 s recordings were made for the same subject[3] for the states of 'eyes open' and 'eyes closed'. The recording trials for both states can be represented as a composite $2 \times N_e \times N_{tr}$-variate vector where $N_e = 2$ denotes the number of electrode channels, $N_{tr} = 7$ denotes the number of trials, and the length of the vector equals the sample length of each recording. A total of 9 IMFs (data-scales) were obtained by performing a single operation of MEMD on the composite data vector, in this way ensuring aligned scales both across trials and electrode channels, and different brain states.

The average complexity analysis over the 7 trials is shown in Fig. 4(a) using standard coarse graining-based MMSE and in Fig. 4(b) using MEMD-based MMSE. Note that while separation between the 'eyes closed' and 'eyes open' states of alertness was not possible using standard MMSE (the error bars overlap for every scale in Fig. 4(a)), the adaptive nature of MEMD-based MMSE enabled a clear separation between the states (even the error bars do not overlap for index 3) at scales which correspond to the alpha frequency range - scales 2 and 3 in Fig. 4(b).



(a) Multivariate sample entropy for scales using coarse graining



(b) Multivariate sample entropy for MEMD-based scales

Fig. 4. MMSE analysis for 'eyes open' and 'eyes closed' EEG: (a) using coarse graining and (b) MEMD-based scales. The curves represent an average of 7 trials and the error bars the SD.

6. CONCLUSIONS

By combining adaptive scale estimation with multivariate entropy theory, a robust structural complexity descriptor for multivariate time series has been developed. Unlike standard techniques, the proposed algorithm is both suitable for nonstationary data and can measure complex coupled dynamics within the vector data channels, prerequisites for the analysis of real-world systems which are typically of a multivariate, coupled and noisy nature. Simulations for biological systems illustrate conclusively how the approach can be used to reveal long-range spatio-temporal correlations present in their time series, signatures of the underlying complex signal generating mechanism.

7. REFERENCES

[1] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, pp. 068102, 2002.

[2] R. Hornero, D. Absolo, J. Escudero, and C. Gomez, "Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease," *Phil. Trans. R. Soc. A*, vol. 367, no. 1887, pp. 317–336, 2009.

[3] M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, "Multiscale entropy analysis of human gait dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 330, no. 1-2, pp. 53–60, 2003.

[4] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Phys. Rev. E*, vol. 84, pp. 061918, 2011.

[5] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy analysis," *Signal Processing Letters, IEEE*, vol. 19, no. 2, pp. 91–94, 2012.

[6] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, "Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects," *IEEE Trans. Biomed. Eng*, vol. 56, no. 9, pp. 2202–2213, 2009.

[7] H. Amoud, H. Snoussi, D. Hewson, M. Doussot, and J. Duchene, "Intrinsic mode entropy for nonlinear discriminant analysis," *IEEE Sig. Proc. Letters*, vol. 14, no. 5, pp. 297–300, 2007.

[8] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proc. of the Royal Society A: Mathematical, Physical and Engineering Science*, 2009.

[9] M. Hu and H. Liang, "Adaptive multiscale entropy analysis of multivariate neural data," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 1, pp. 12–15, 2012.

[10] N. Rehman and D. P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2421–2426, may 2011.

[11] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.

[12] J. M. Hausdorff, P. L. Purdon, C.-K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations," *Jrnl of Appl Physiol*, vol. 80, no. 5, pp. 1448–1457, 1996.

[13] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. Suppl 1, pp. 2466–2472, 2002.

[3]The data was recorded according to the 10-20 system using a g.tec g.USBamp from electrodes Cz and POz at 512 Hz. The data was then bandpass filtered (FIR filter) to retain frequencies within the range 1 - 45 Hz.