

Review

# Multiscale Entropy Analysis of Center-of-Pressure Dynamics in Human Postural Control: Methodological Considerations

Brian J. Gow<sup>1</sup>, Chung-Kang Peng<sup>2</sup>, Peter M. Wayne<sup>1</sup> and Andrew C. Ahn<sup>3,4,\*</sup>

Received: 24 September 2015 ; Accepted: 19 November 2015 ; Published: 30 November 2015  
Academic Editor: Anne Humeau-Heurtier

<sup>1</sup> Osher Center for Integrative Medicine, Brigham and Women's Hospital, Harvard Medical School, 900 Commonwealth Ave., Boston, MA 02215, USA; bgow@partners.org (B.J.G.); pwayne@partners.org (P.M.W.)

<sup>2</sup> Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave., Boston, MA 02215, USA; cpeng@bidmc.harvard.edu

<sup>3</sup> Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave., Boston, MA 02115, USA

<sup>4</sup> Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th St, Charlestown, MA 02129, USA

\* Correspondence: aahn1@mgh.harvard.edu; Tel.: +1-617-754-4677

**Abstract:** Multiscale entropy (MSE) is a widely used metric for characterizing the nonlinear dynamics of physiological processes. Significant variability, however, exists in the methodological approaches to MSE which may ultimately impact results and their interpretations. Using publications focused on balance-related center of pressure (COP) dynamics, we highlight sources of methodological heterogeneity that can impact study findings. Seventeen studies were systematically identified that employed MSE for characterizing COP displacement dynamics. We identified five key methodological procedures that varied significantly between studies: (1) data length; (2) frequencies of the COP dynamics analyzed; (3) sampling rate; (4) point matching tolerance and sequence length; and (5) filtering of displacement changes from drifts, fidgets, and shifts. We discuss strengths and limitations of the various approaches employed and supply flowcharts to assist in the decision making process regarding each of these procedures. Our guidelines are intended to more broadly inform the design and analysis of future studies employing MSE for continuous time series, such as COP.

**Keywords:** multiscale entropy; sample entropy; methodological; center of pressure; systematic review

## 1. Introduction

As highlighted by the emerging fields of systems biology and medicine, health requires the integration—across multiple time and spatial scales—of control systems, feedback loops, and regulatory processes that enable an organism to function and adapt to the demands of everyday life. Within this framework, aging and disease can be viewed as the breakdown of nonlinear feedback loops acting across multiple scales, resulting in a loss of physiological complexity [1]. Physiologic complexity can be estimated using a number of techniques derived from the fields of nonlinear dynamics and statistical physics that quantify the moment-to-moment quality, scaling, and/or correlation properties of dynamic signals [2,3].

One increasingly used entropy-based metric of complexity is multiscale entropy (MSE). MSE characterizes the information content of a signal by quantifying the degree of regularity or predictability over multiple scales of time [4]. MSE has been used to evaluate the relationship between

complexity and health in a number of populations and physiological systems. For example, MSE of heart beat intervals demonstrates a clear loss of complexity with aging, is lower in patients with congestive heart failure, and is predictive of mortality [5]. MSE has also been used to distinguish older adults with atrial fibrillation from healthy controls [6] and differentiate healthy fetuses from fetuses with a pathological condition at birth [7]. Additionally, it has been used in physiological processes as varied as red blood cell flickering, gait dynamics and sleep [6,8–10]. The use of MSE for studying center-of-pressure (COP) dynamics has received a significant amount of attention, particularly in elderly populations where falls are of greater concern [8,11,12].

Despite its promise as a sensitive and novel biomarker of health and disease, few attempts have been made to outline the methodological challenges associated with the calculation of MSE. While current publications on MSE may discuss one or two methodological issues, no publication—to our knowledge—comprehensively covers all the issues presented here. For the MSE-naïve researcher, designing a protocol for the purpose of MSE analysis can be daunting, and this difficulty can be further amplified by the recognition that improper choice of parameters during MSE analysis can lead to ambiguity in complexity signatures between healthy and diseased states [13]. In this paper, we address a number of key issues involved in study design, analysis and interpretation of MSE for physiological signals using COP as a model example. In particular we focus on five methodological issues considered critical for the proper design and analyses of an MSE study: (1) data length; (2) frequency range of analyses; (3) sampling rate; (4) point matching tolerance and sequence length; and (5) filtering. We choose COP because it is distinct from the more commonly analyzed, discrete heartbeat interval; the raw displacement COP data is continuous and potentially plagued by nonstationarities; and the physiologic basis for COP is not as well-defined as that of heart-rate and other physiological processes. A systematic review of publications using MSE to analyze COP was conducted which serves to highlight the existing methodological heterogeneity in key MSE parameters. We start with an overview of MSE since a basic understanding of the technique is required for context in the subsequent sections.

## 2. Overview of Multiscale Entropy

MSE quantifies the degree of irregularity within a system across multiple time scales. The entropy measure used to determine the amount of irregularity at each time scale is called sample entropy (SampEn). SampEn represents the rate of generation of new information and is precisely equal to the negative natural logarithm of the conditional probability that  $m$  consecutive points that repeat themselves, within some tolerance,  $r$ , will again repeat with the addition of the next  $(m + 1)$  point [6]. The tolerance,  $r$ , is often derived by calculating a certain percentage of the time-series standard deviation (SD).

As described by Richman *et al.* [14], the mathematical derivation of SampEn is as follows. For a time-series of length  $N$ ,  $\{u(j) : 1 \leq j \leq N\}$ ,  $N - m + 1$  vectors,  $\mathbf{x}_m(i)$ , are formed for  $\{i | 1 \leq i \leq N - m + 1\}$ , where  $\mathbf{x}_m(i) = \{u(i+k) : 0 \leq k \leq m-1\}$  is the vector of  $m$  data points from  $u(i)$  to  $u(i+m-1)$ . The vectors being compared against  $\mathbf{x}_m(i)$  to assess the number of “repeats” or “matches” are represented by  $\mathbf{x}_m(j)$ . A match is established once the distance between two vectors,  $\mathbf{x}_m(i)$  and  $\mathbf{x}_m(j)$ —defined as the maximum difference of their corresponding scalar components—is less than  $r$ . The vector  $\mathbf{x}_m(i)$  is referred to as the template and in the case of a match the  $\mathbf{x}_m(j)$  is referred to as a template match. This is again repeated for  $\mathbf{x}_{m+1}(i)$  and  $\mathbf{x}_{m+1}(j)$ . This process of matching is illustrated in Figure 1.

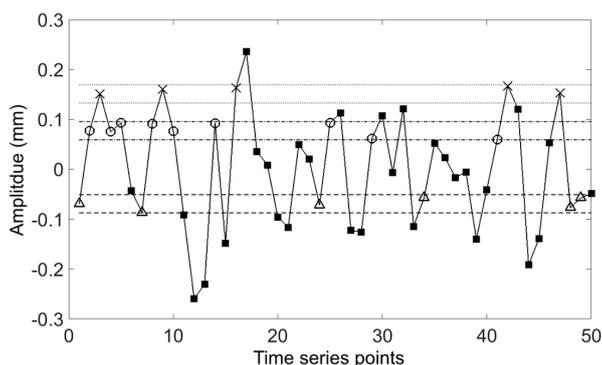
The probability of matches are calculated for each reference vector  $\mathbf{x}_m(i)$  and  $\mathbf{x}_{m+1}(i)$  and represented by  $B_i^m(r)$  and  $A_i^m(r)$ , respectively.  $B_i^m(r)$  equals to  $(N - m - 1)^{-1}$  times the number of vectors,  $\mathbf{x}_m(j)$ , within  $r$  of,  $\mathbf{x}_m(i)$ , and  $A_i^m(r)$  is given by  $(N - m - 1)^{-1}$  times the number of vectors,  $\mathbf{x}_{m+1}(j)$ , within  $r$  of,  $\mathbf{x}_{m+1}(i)$ , when  $j$  ranges from 1 to  $N - m$  and  $j \neq i$ . The restriction  $j \neq i$  assures that self-matches are not counted (*i.e.*, vectors are not compared to themselves). SampEn can then be defined as:

$$S_E(m, r, N) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right) \tag{1}$$

where  $B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r)$  and  $A^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} A_i^m(r)$ .  $B^m(r)$  represents the probability that two sequences will match for  $m$  points and  $A^m(r)$  represents the probability that two sequences will match for  $m + 1$  points across all possible comparisons.

A few points regarding SampEn bear noting. First, by nature of the calculations, SampEn for periodic, regular signals is approximately zero while SampEn is maximal with irregular, random signals. This can be understood with the fact that  $A$  and  $B$  are nearly identical in periodic signals;  $A/B$  is near unity; and thus the logarithm of  $A/B$  approximates to zero. On the other hand, irregular signals have lower probability of matches at  $m + 1$  ( $A$ ) compared to that of matches at  $m$  ( $B$ );  $A/B$  is a low-magnitude fraction; and  $\ln(A/B)$  calculates to a large negative number which is made positive with the negative in Equation (1), ultimately yielding a larger SampEn.

Second, the theoretic basis of sample entropy rests on the *probability* of matches, and the actual calculation is an estimation based on the available samples. Much like the probability of “heads” for a coin approximated by counting the number of heads after a number of trial flips, the estimation of SampEn becomes increasingly susceptible to stochastic effects as the number of trials diminishes in quantity. The confidence in the accuracy of SampEn, thus, diminishes with smaller time series. Longer datasets are considered optimal. However, it may be incorrect to assume that the dynamics remain unchanged over the course of sampled time, particularly for longer time series.



**Figure 1.** Demonstration of SampEn calculation with  $m = 2$ . The dashed line is the tolerance about the first point and highlights matching points to the first point with  $\Delta$  markers. Likewise the dash-dot line and dotted line highlight matches of the second and third points with  $\circ$  and  $\times$  markers respectively. Points which do not match any of the template points are marked by  $\blacksquare$  symbols. SampEn is calculated from the ratio of sequences of length  $m$  and length  $m + 1$  which match  $m$  and  $m + 1$  length templates. The first templates are represented by the first 2 ( $m$ ) and 3 ( $m + 1$ ) points. We observe 2  $\Delta$ - $\circ$  template matches to the  $m$  length template and one  $\Delta$ - $\circ$ - $\times$  template match to the  $m + 1$  length template. The template is then stepped one sample at a time and the process repeated until the end of the waveform is reached. SampEn can then be calculated from the ratio of the total number of  $m + 1$  to  $m$  length template matches. Adapted from [6].

Third, the number of matches ( $A$  and  $B$ ) in SampEn is determined by the *cumulative* number of matches found between the possible permutations of vector comparisons. The quotient for  $A$  and  $B$  is subsequently entered into the logarithmic calculations to find SampEn. This approach is inherently different than that taken by approximate entropy (ApEn), the predecessor for SampEn. ApEn relies on determining the probability of matches found for each vector and then entering this probability into the logarithmic function. As a result, a time series with 100 data points would be associated with 99 such probabilities and, by extension, 99 logarithmic calculations for  $m = 2$  (and 98 probabilities and 98 logarithmic calculations for  $m = 3$ ). In stark contrast, SampEn has only one

logarithmic calculation. To obtain ApEn, the logarithmic terms are summed respectively for  $m = 2$  and  $m = 3$ , and the difference of the two sums would equal ApEn. The unintended consequence of the ApEn approach is that smaller time series and highly irregular time series may encounter zero matches which would subsequently yield an undefined ApEn since the logarithm of 0 cannot be calculated. To avoid this issue, self-matches are included to ensure that every logarithmic calculation entails a non-zero positive integer. This naturally biases the ApEn towards a lower entropy value for short and highly irregular time series [14].

MSE is termed “multiscale” because the sample entropy (SampEn) is calculated across multiple time scales ( $\tau$ ). This is achieved through a coarse-graining procedure. At the first scale, the MSE algorithm evaluates SampEn for the time-series at each sampled point. At greater MSE scales, SampEn is computed on coarse-grained versions of the original time-series. The coarse-graining procedure divides the original time-series into non-overlapping windows of length,  $\lambda$ . Within each window the average is taken resulting in a new time-series of length  $N/\lambda$ . This is shown for time scales 2 and 3 in Figure 2. The procedure is repeated until the last time scale is reached [6].

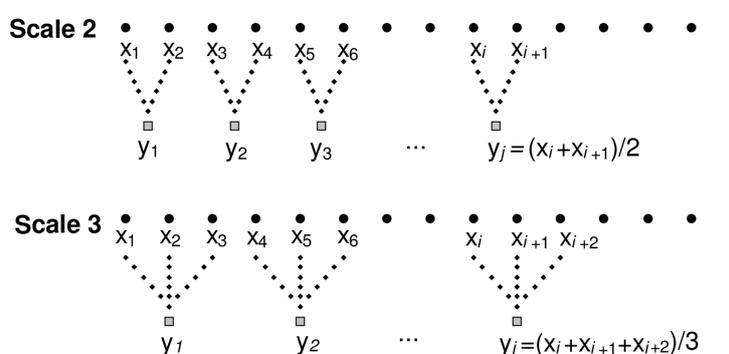


Figure 2. MSE coarse graining procedure example for scales two and three. Adapted from [15].

The MSE output of SampEn *vs.* Scale,  $\tau$ , can be used to calculate a complexity index,  $C_I$ . The  $C_I$  is calculated by taking the area under this curve. A few important points about this composite approach are worth noting. First,  $r$ , the tolerance for matches, remains constant for all scales of the MSE calculation. The  $r$  is determined by the standard deviation of the time series at scale 1 (not coarse grained). Second, MSE assigns a high  $C_I$  to time-series with complex dynamics across all the time scales evaluated. For this reason,  $1/f$  noise is associated with a high  $C_I$  because the SampEn remains relatively constant across time scales. Uncorrelated or white noise, however, is characterized by high irregularity (SampEn) at lower scales but increasingly decreased SampEn at higher scales, ultimately yielding a relatively smaller  $C_I$ . Since  $1/f$  noise is ubiquitous in nature [16], this technique has gained traction in the analysis of physiological signals. Third, the coarse graining procedure, itself, does not necessarily make MSE calculations immune to cross temporal scale effects. Coarse graining is a type of filter that is susceptible to aliasing. Periodicity at a specific frequency (represented by decreased SampEn) can be seen at multiples of the cycle frequency [17].

### 3. Multiscale Entropy of Center of Pressure Dynamics in Human Postural Control: A Systematic Review

Analysis of the center of pressure (COP) during human standing is widely used to characterize postural control and to understand underlying motor control mechanisms during both unperturbed and challenging experimental conditions. Location and dynamics of the COP are typically measured using a force platform. During standing, reaction forces between the body and support surface (*i.e.*, platform) are distributed over the entire contact area. These forces can be summed into a single net force acting at a single point: the center of pressure. COP is not a static measure, and variability in the anteroposterior and mediolateral planes can be characterized using average measures of

displacement (e.g., range, area swept), changes in velocity, or moment-to-moment dynamics. COP dynamics are likely due to complex control process associated with the maintenance of postural control, as well as the inherent noise within the human neuromotor system. COP is widely used to inform the health of the postural control system, and in some populations, is a predictor of instability and falls [18].

### 3.1. Systematic Review Methods

We performed a systematic review of publications using MSE, as defined by Costa *et al.* [15], to analyze COP displacement data. Only this specific version of multiscale entropy was included as part of this review. Variants [19–21] that also use entropy measures across scales were not considered. We completed electronic literature searches using PubMed/MEDLINE, Excerpta Medica Database (Embase), Web of Science™, and Academic Search Premier on 14 May 2015. Combinations of keywords (“Center of Pressure” OR “COP” OR “Postural”) AND (“Multiscale Entropy” OR “Multi-scale Entropy” OR “MSE”) were used as search terms. This returned 92 unique results. Articles were excluded if: (1) They were not written in English; (2) They were not original research; (3) The publication was simply an abstract or a letter; (4) Multiscale entropy was not a primary metric; and (5) Center-of-pressure raw force-plate displacement data were not analyzed. We limited our inclusion to raw displacement analysis of COP data and excluded studies which focused solely on COP velocity data. Analysis of displacement data, unlike that of velocity data, requires special consideration in regards to filtering and the management of nonstationarities.

All manuscripts meeting these inclusion criteria were published in peer-reviewed journals. We included studies with obvious methodological limitations since this review is focused more on MSE methodology rather than the actual quality of the data. All published settings for  $m$ ,  $r$ , data length, sample rate (fs), filtering method, analyzed frequencies and time scales were recorded and tabulated. In many cases, these parameters were not reported (Table 1).

### 3.2. Systematic Review Results

Results of this review show that all of the settings used to analyze COP displacement data with MSE were heterogenous, some much more than others.

The columns for sequence length ( $m$ ) and point matching tolerance ( $r$ ) parameters were relatively consistent across studies. The most common parameter settings for sequence length was  $m = 2$ . The most common setting for the point matching tolerance was  $r = 0.15$ . Of note, a couple studies explicitly evaluated multiple ranges for these parameters [22,23].

In contrast to sequence length and point matching tolerance, settings chosen for time series length, sampling rate, filtering method, frequencies and times scales analyzed, and the number of points remaining at the greatest time scale varied considerably across studies. The length of the time-series in seconds varied greatly across studies, ranging from 7 s to 1800 s. Sampling rate also varied markedly across studies, ranging from 33 Hz to 250 Hz. Studies employed a variety of filtering methods to remove trends outside the frequencies of interest. Empirical Mode Decomposition (EMD) was the most commonly employed technique, in part due to its applicability to nonlinear and nonstationary data [24]. Briefly, EMD decomposes a signal into a set of intrinsic mode functions (IMFs) where each IMF represents a dominant or characteristic frequency with a limited bandwidth. Fourier-based methods were the second most commonly used filtering method. Duarte *et al.*, also explored a number of additional methods for filtering drifts and nonstationarities [22]. The frequencies analyzed also varied greatly across studies. On the low end, frequencies as low as 0.0056 Hz were included. On the high end, frequencies between 7.5 and 60 Hz were analyzed. The MSE scales used for the estimation of the complexity index also varied across studies, with values ranging from well below 10, to greater than 50. Finally, for studies where it was possible to calculate the number of data points remaining at the last MSE scale,  $N_{\tau M}$ , this parameter also varied from 100 to 1800 points. However, only two studies had less than 300 points.

**Table 1.** Systematic review of publications using MSE to analyze center-of-pressure displacement time-series.

Public Ation	Study Design [MSE Measure(s)]	No. Subjects	Time-Series Length (s)	Time Scales Analyzed	$N_{\tau M}$ : Points $\tau_{max}^{\Delta}$	Frequencies Analyzed (Hz) [Dissimilarity Comparison $\Omega$ ]	fs: Sampling Rate (Hz)	SC <sub>H</sub> : Samp./Cycle Highest Freq. $\delta$	m: Sequence Length	r: Point Matching Tolerance	Filtering	Key MSE Related Findings $\omega$
Costa <i>et al.</i> (2007) [8]	3 groups (young (Y), healthy elderly (HE), fallers (F)) and pre-, post- stochastic resonance (SR) exposure. [MSE-CI]	Pre-exposure: Y = 15 HE = 22 F = 22 Post-exposure: Y = 15 HE = 12	30	1–6	300 (~17 <sup>m</sup> )	- [No]	60	-	2	0.15	EMD	<ul style="list-style-type: none"> <li>F had lower MSE-CI than Y and HE.</li> <li>HE showed increased MSE-CI w/SR but not Y.</li> </ul>
Duarte <i>et al.</i> (2008) [22]	2 groups (healthy-young (HY) and healthy-older (HO)). [MSE-CI]	HY = 14 HO = 14	1800	1–50	720 (~27 <sup>m</sup> )	>0.0056 [No]	20	-	2	0.2	Custom	<ul style="list-style-type: none"> <li>HO showed higher MSE-CI than HY.</li> </ul>
Kang <i>et al.</i> (2009) [25]	3 groups (nonfrail (NF), prefrail (PF), frail(F)) under 2 conditions (single-task and dual-task (DT)). [MSE-CI]	NF = 291 PF = 209 F = 50	30	2–8	900 (30 <sup>m</sup> )	7.5–60 [No]	240	2	2	0.15	EMD	<ul style="list-style-type: none"> <li>MSE-CI lower under DT in all groups.</li> <li>MSE-CI associated with frailty status.</li> </ul>
Manor <i>et al.</i> (2010) [11]	3 impaired groups (visual, somatosensory, both (combined)), a control group and 2 exposures (single-task and dual-task (DT)). [MSE-CI]	Contol = 299 Visual = 81 Somatosensory = 49 Combined = 25	30	2–8	900 (30 <sup>m</sup> )	- [No]	240	-	-	-	EMD	<ul style="list-style-type: none"> <li>MSE-CI lower under DT.</li> <li>MSE-CI different between all groups.</li> </ul>
Gruber <i>et al.</i> (2011) [26]	2 groups with adolescent idiopathic scoliosis (AIS) (pre-bracing (PB) and pre-operative (PO)), and a healthy control (CON) group. [MSE-CI]	Control = 10 PB = 18 PO = 18	7	1–20	-	- [No]	-	-	2	0.15 <sup>0</sup>	LPF 20 Hz	<ul style="list-style-type: none"> <li>MSE-CI showed differences for CON vs. AIS, CON vs. PB, CON vs. PO, PB vs. PO.</li> </ul>
Kirchner <i>et al.</i> (2012) [27]	Single group under 2 conditions (dual-task (DT) and quiet-standing (BT)). [MSE-CI]	16	30, 60, 300	1–6, 1–12, 1–60	100 (10 <sup>m</sup> )	- [No]	20	-	2	0.15	Custom	<ul style="list-style-type: none"> <li>MSE-CI differed between BT and DT at 300 s.</li> </ul>
Jiang <i>et al.</i> (2011) [23]	Multiple studies: <ul style="list-style-type: none"> <li>Young vs. Elderly (YE) under 2 conditions (single-task (ST) and dual-task (DT))</li> <li>Eyes-Open vs. Eyes-Closed (EO/EC)</li> <li>Pre- and Post- wearing vibratory (V) insoles. [MSE-CI]</li> </ul>	YE: Young = 15 Elderly = 13 EO/EC: 16 V: Young = 16 Elderly = 26	60	1–7	343 (~19 <sup>m</sup> )	1.5–3 [Yes]	40	13.3	2 *	0.15 *	EMD	<ul style="list-style-type: none"> <li>YE: MSE-CI differed for DT vs. ST within both groups and between groups under both ST and DT.</li> <li>EO/EC: MSE-CI differed between EO and EC.</li> <li>V: MSE-CI differed between pre- and post- V in Elderly.</li> </ul>
Wei <i>et al.</i> (2012) [12]	Single elderly group with 2 exposures (pre- and post-wearing vibratory (V) insoles). [MSE-CI]	26	60	-	-	- [Yes]	31.25	-	2	0.2	EMD	<ul style="list-style-type: none"> <li>No significant differences in MSE-CI.</li> </ul>
Huang <i>et al.</i> (2013) [28]	Single group and variable platform (rigid platform (R) and water pad (W)), variable force-plate (custom, AMTI) each under eyes-open (EO) and eyes-closed (EC) conditions. [MSE-CI]	20	60	-	-	<2 [No]	50	-	-	-	EMD	<ul style="list-style-type: none"> <li>MSE-CI differed between R and W under both EO and EC, across both platforms.</li> </ul>

Table 1. Cont.

Public Aition	Study Design [MSE Measure(s)]	No. Subjects	Time-Series Length (s)	Time Scales Analyzed	$N_{\tau M}$ : Points $\tau$ max $\Delta$	Frequencies Analyzed (Hz) [Dissimilarity Comparison $\Omega$ ]	fs: Sampling Rate (Hz)	SC <sub>H</sub> : Samp./Cycle Highest Freq. $\delta$	m: Sequence Length	r: Point Matching Tolerance	Filtering	Key MSE Related Findings $\omega$
Manor <i>et al.</i> (2013) [29]	Single group exposed to 24 weeks of Tai Chi. [MSE-CI]	25	30	1–5	300 (~17 <sup>m</sup> )	3.125–12.5 [No]	50	4	2	0.15	EMD	• MSE-CI increased with exposure to Tai Chi.
Fournier <i>et al.</i> (2014) [30]	Group of children with Autism Spectrum Disorder (ASD) relative to controls. [MSE-CI]	Controls = 17 ASD = 16	20	1–20	360 (~19 <sup>m</sup> )	<20 [No]	360	18	2	0.2	LPF 20 Hz	• ASD showed lower MSE-CI than controls.
Chen <i>et al.</i> (2014) [31]	Single elderly group exposed to a Resistance Training program. [MSE-CI]	24	60	-	-	- [No]	-	-	-	-	EMD	• No significant change in MSE-CI.
Pau <i>et al.</i> (2014) [32]	2 groups (part-time (retained) and full-time (career) firefighters) pre- and post- a physical task (stressor). [MSE-CI]	Retained = 13 Career = 13	30	1–8	123 (~11 <sup>m</sup> )	<18 [No]	33	1.8	2	0.15	LPF 18 Hz	• Change in MSE-CI was smaller in career <i>vs.</i> retained firefighters after the stressor.
Wayne <i>et al.</i> (2014) [33]	Multiple studies: • Cross-sectional (X-Sec) with two groups (Tai Chi Experts (E) and Naives (N)) under eyes-open (EO) and eyes-closed (EC) conditions. • Longitudinal (LGT) with two groups (randomized to Tai Chi (TC) or Usual Care(UC) under EO and EC conditions. [MSE-CI]	X-Sec: E = 27 N = 60 LGT: TC = 31 UC = 29	50	X-Sec: AP = 1–25, ML = 1–35 LGT: AP = 2–31, ML = 1–39	X-Sec: AP = 500 (~22 <sup>m</sup> ), ML = 357 (~19 <sup>m</sup> ) LGT: AP = 403 (~20 <sup>m</sup> ), ML = 320 (~18 <sup>m</sup> )	X-Sec: AP = 1.3–17.6, ML = 0.5–17.9 LGT: AP = 0.6–3.2, ML = 0.3–8 [Yes]	250	X-Sec: AP = 14.2, ML = 14.0 LGT: AP = 39.1, ML = 31.3	2	0.15	EEMD	• MSE-CI differed between E and N under EO and EC.
Yeh <i>et al.</i> (2014) [34]	3 groups (young, dizzy (w/ vestibular hypofunction), healthy-elderly) exposed to the sensory organization test (SOT). [MSE-CI]	Young = 23 Dizzy = 19 Elderly = 9	-	1–20	-	- [No]	100	-	-	-	-	• MSE-CI analysis showed differences between groups which varied by SOT condition.
Decker <i>et al.</i> (2015) [35]	2 groups (postmenopausal women of lower (L) physical function, those of normal or subnormal (N) physical function) under eyes-open (EO) and eyes-closed (EC) conditions. [MSE-CI]	N = 32 L = 94	51.2	1–6	341 (~18 <sup>m</sup> )	- [No]	40	-	2	0.15	EMD	• MSE-CI did not differ by group or exposure.
Zhou <i>et al.</i> (2015) [36]	Single group under single-task (ST) and dual-task (DT) conditions and 2 exposures (pre- and post- transcranial direct current stimulation (tDCS)(real or sham)). [MSE-CI]	20	60	3–8	1800 (~42 <sup>m</sup> )	- [No]	240	-	2	0.15	EEMD	• tDCS reduced the dual-task cost of MSE-CI.

MSE-CI—is the complexity index which is determined by taking the area under the curve of sample entropy *vs.* time scales; LPF—low pass filter; EMD—empirical mode decomposition; EEMD—ensemble empirical mode decomposition; AP—anterioposterior direction; ML—mediolateral direction;  $\Delta$ —Number of points remaining at last time scale;  $\Omega$ —A dissimilarity comparison is a statistical analysis between a healthier (or otherwise disparate group) and the study group at baseline to determine which frequencies best distinguish the groups;  $\delta$  indicates the number of samples per cycle at the first scale for the highest frequency component;  $\omega$ —Only statistically significant differences are presented;  $\sim$ reported on  $m = 2$  and  $r = 0.2$  but also did additional analysis to check if it changed their result. Used  $m = 1$  to 5 and  $r = 0.1, 0.15, 0.25$  and 0.3. Additionally to account for outliers they tried using a fixed point matching criteria ( $r = 0.2 \times 1$ );  $\theta$  reported that  $0.15 \times SD$  was used for  $r$  but later report that an absolute value of 0.001 was used; \* used  $m = 2, 3$  and  $r = 0.1, 0.15, 0.2, 0.25$  and 0.3, choose  $m = 2$  and  $r = 0.15$  since it maximized the difference between young and elderly at the frequencies analyzed.

#### 4. Methodological Considerations

The first three subsections (4.1–4.3) listed below should be considered prior to protocol development. This ensures that the protocol for data collection is designed appropriately for MSE analysis. All subsections (4.1–4.6) herein discuss methodological considerations applicable to the actual data analysis.

##### 4.1. Determining Required Data Length

Because SampEn is ultimately a probabilistic calculation, SampEn requires a minimum number of points to obtain an accurate estimate of matching probability. The confidence in the accuracy of SampEn is diminished greatly when the number of matches is low. This can occur with shorter data (due to short data acquisition or substantial coarse graining at higher MSE scales), highly irregular time series, tight tolerance window (small  $r$ ), or data with trends or drifts. Of these factors, data length becomes a universally unavoidable issue for all finite time series since coarse graining for ascending MSE scales ultimately generates a time series too short for reliable MSE analyses.

To address this methodological issue, an important first step is to determine the minimum number of points required for SampEn. Estimates for the minimum number of SampEn points are sometimes based on theoretical calculations for ApEn which suggest that  $10^m$  points should be sufficient, although  $20^m$ – $30^m$  points would be preferable for an accurate estimate [27,37]. However ApEn estimates based on simulated random time-series show increasing effects of self-matching bias with a smaller number of points [14]. In comparison, due to the exclusion of self-matches, SampEn is not susceptible to such biases and is generally considered more robust to shorter time series. However, it is noted in Richman *et al.* [14] that the confidence intervals for simulated random time-series at a length of  $10^m$  remain quite large for SampEn, therefore we recommend that between  $14^m$  and  $23^m$  points be present at the last MSE scale analyzed. As denoted by the  $N_{\tau M}$  column in Table 1, the majority of the reviewed studies satisfied this criterion with 300 data points ( $17^m$  with  $m = 2$ ) used for analyses at the last scale. Determining the number of points at the last MSE scale is done by multiplying the sample rate times the data length and then dividing by the largest MSE scale,  $N_{\tau M} = (f_s \times t) / \tau_M$ . One outlier was the Gruber study where the acquired data totaled 7 s. Although the  $N_{\tau M}$  for this study could not be determined due to lack of reporting for  $f_s$ , it is unlikely that sufficient data points at physiologically relevant frequencies could be extracted from such a short acquisition window.

Ideally, as much data as possible should be acquired but constraints arise from a subject's capacity to sustain such testing for long durations. For COP, fatigue can emerge generating altered dynamics and transient effects such as shifts, fidgets, or drifts. These changes can produce dynamics that no longer become the state for which the investigators were originally intending to evaluate. Duarte's study [22], for instance, acquired testing for 30 min in both young and older subjects, and this duration may have potentially caused transient effects (*i.e.*, nonstationarities as discussed below) that ultimately change the MSE results. The study, however, was interested in very low physiological frequencies and may not have had other options.

The length of testing is therefore dictated by the subject's capacity to maintain a specific dynamic and by the lowest physiologic frequency of interest. When this lowest physiologic frequency of interest is known, the required length of the time series to be collected in a study can be determined. This should be done such that the lowest frequency component included is not clearly oversampled. To achieve this, a simple formula based on the number of points remaining at the last coarse-grained time scale,  $N_{\tau M}$ , can be used:

$$t = \frac{N_{\tau M}}{2 \times (m + 1) \times f_L} \quad (2)$$

where  $f_L$  is the lowest frequency component of interest and  $m$  is the sequence length. This will result in  $2 \times (m + 1)$  samples per cycle for the lowest frequency component at the last time scale. A minimum number of oscillations are required to accurately characterize the information at some low frequency,

so we use an example to check that this is reasonable when using this formula. If we take  $N_{\tau_M} = 300$ ,  $m = 2$  and  $f_L = 0.5$  Hz, we observe that we need our time series to be 100 s in length. This would result in 50 oscillations of this low frequency component which is reasonable.

In the end, the confidence interval for a SampEn calculation may not be dictated by data length alone and can be influenced by other factors such as increased signal regularity or higher tolerance  $r$ . As a result, the extent by which an investigator can coarse grain the time series (for higher MSE scale analyses) can be determined not merely by data length alone but also by an additional stability analysis process. Stability is established by observing consistent trends in SampEn with increasing MSE scales. However, if there are significant deviations or erratic patterns (e.g., an increase with a subsequent decrease or *vice versa*) in consecutive SampEn values as MSE scales increase, then this would suggest that the SampEn calculations are now susceptible to stochastic effects and thus unreliable. This stability analysis can be evaluated within-subjects and across subjects to observe overall patterns. An arbitrary value of a  $\pm 0.1$  change in SampEn from scale  $\tau-1$  to  $\tau$  followed by a change in the opposite direction of  $\pm 0.1$  can be used to determine the last stable scale. When this test fails analysis should stop at  $\tau-1$ , the scale where the instability begins.

Of note, new variants of MSE, such as Modified MSE, have been created to overcome these limitations seen with short time-series [38]. A review of a number of other variants on MSE, each of which has their own strengths is provided in Humeau–Heurtier [39].

#### 4.2. Range of Frequencies for Analysis

For continuous time series such as COP data, each MSE scale represents a time frequency: smaller scales correspond to higher frequencies while larger scales correspond to lower frequencies. For certain discrete data such as heart interbeat intervals, on the other hand, this frequential correlation is not nearly as straightforward since each MSE scale corresponds to an approximate average of vacillating interbeat periods at varying degrees of coarse graining. The MSE analyses of continuous COP data are therefore more conducive to physiological interpretations based on the frequency represented at each scale: a SampEn value for a 1 Hz time series would reveal information about the amount of irregularity of the time series at 1 Hz, and so on.

The frequency range on which to focus MSE analyses is constrained by two factors: (1) physiological considerations and (2) the limits set by the granularity and length of the data. In an ideal world, SampEn values would impart information about a well-described physiological mechanism operating at the analyzed frequency. However, unlike heart rate, the physiological basis for COP is not well understood. This lack of clarity may account for the wide range of frequencies (e.g., from 0.0056 to 60 Hz) analyzed by the studies summarized in Table 1.

Several tactics have been adopted to deal with this issue. One approach is to recruit healthy control groups which, in the case of COP studies, have been largely composed of healthy young individuals. Comparisons are subsequently made at each frequency range to determine which frequencies differed statistically between a disordered condition and the healthy young, and these frequencies are then examined to identify the effects of a specific intervention. Jiang *et al.* [23], for instance, selected the frequencies of interest based on the dissimilarity of the  $C_1$  between elderly and young subjects at baseline. Once the intervention—vibratory insoles—were applied, the  $C_1$  in elderly were re-evaluated at those pre-identified frequencies and were found to increase making their  $C_1$  similar to the  $C_1$  in healthy young. This led to the conclusion that vibratory insoles applied to the elderly people might be able to improve their postural stability [23]. Other studies, such as Wei *et al.* [12] and Wayne *et al.* [33], have similarly used young as healthy controls to identify the frequencies to analyze.

When statistical comparisons with a “healthy” group is not feasible, assumptions must be made about which physiological frequencies are clinically relevant. Some assumptions are based on physiological feasibility: for example, frequencies above 20 Hz are likely too rapid to affect balance-related processes or neurophysiological dynamics at the whole-body level. Other

assumptions are premised on the distribution of spectral power. For instance, the preponderance of spectral power (95%) of quiet standing COP exists at frequencies lower than 1 Hz [22]. However, the majority of the studies in our systematic review examined frequencies greater than 1 Hz and some identified statistically significant results.

The other constraint limiting the range of analyzed frequencies is the granularity and length of data. The highest frequency,  $f_H$ , feasible for analysis is set by the sample rate at which the data is collected. As discussed below, accurate characterization of a physiological process at a specific frequency requires sufficient granularity, and we recommend having at least five points per cycle for  $f_H$  represented at the first MSE scale. Therefore, MSE analyses at 10 Hz frequency should be performed on data with at least a 50 Hz sampling rate ( $f_s$ ). When the data are acquired at very high sample rates, this guideline may permit analyses of data at frequencies that are too high to be physiologically realistic. In this case,  $f_H$  should be selected based on physiologic feasibility.

The lowest frequency,  $f_L$ , feasible for analysis is limited by the length of time over which data is acquired. The shorter the time series, the less one can evaluate the lower frequencies. To determine the lowest frequency which should be included in the analysis based on the time series length one can rearrange Equation (2):

$$f_L = \frac{N_{\tau M}}{2 \times (m + 1) \times t} \quad (3)$$

This will set the lowest frequency which should be included in the analysis (*i.e.*, the lowest frequency IMF or the cutoff frequency of a high-pass filter) such that at the last time scale it will have  $2 \times (m + 1)$  samples per cycle. For example if we performed EMD on a time-series of length  $t = 40$  s which resulted in one of the low frequency IMFs having a characteristic frequency of 0.5 Hz. We need to understand whether that IMF should be included in the analysis or if instead it should be eliminated since it will be oversampled even at the last time scale. Using Equation (3) with  $N_{\tau M} = 300$  we observe that the minimum  $f_L$  should not be lower than 1.25 Hz. Therefore the 0.5 Hz IMF should not be included since it will always be oversampled. We include  $m$  in the denominator because for larger sequence lengths we do not want to detrend the signal to much. Since we are looking at longer sequences the relevant information (oscillations) will be prevalent at lower frequencies. We would like to emphasize that Equations (2) and (3) are merely guidelines to help setup a study and analysis such that meaningful information can be garnered from a dataset. There are a number of other important factors to consider when determining how long of a time-series to collect and which frequencies to analyze; subject fatigue, protocol limitations (inability to collect long data; BOLD-fMRI), and which frequencies are physiologically meaningful. While complexity generally persists across multiple time scales in some cases there may be a valid physiological reason for not analyzing below a particular frequency.

#### 4.3. Appropriate Sample Rate ( $f_s$ )

To adequately capture the dynamics of a specific frequency of interest, a minimum number of samples are required per cycle or period. Traditionally, in engineering, the Nyquist criterion mandates that the sampling rate be twice the frequency to be evaluated. In our systematic review, some researchers choose  $f_s$  such that the number of samples per cycle at the highest frequency,  $f_H$ , was 2 to 4. While this satisfies the Nyquist criterion for sampling, the more conservative approach would recommend at least five samples per cycle ( $5 \times$  the highest frequency) since evaluation of sinusoidal waveforms with sampling at less than 5 samples per cycle results in mean amplitude errors greater than 5% [40]. To fully capture the information contained in the highest frequency component ( $f_H$ ) it is recommended to set  $f_s$  such that there are at least five samples/cycle for  $f_H$  at the first MSE scale. The counter-situation is when the sampling is obtained at a much higher rate.

As stated previously, experimental data obtained at a sampling rate much greater than that required by the Nyquist theorem could lead to analyses of processes that are not relevant to the system of interest. In this oversampled case, matching would occur at smaller time intervals and

thereby fail to assess the dynamics at the frequencies of interest. For example, assuming that no physiological process in COP operates at frequencies greater than 20 Hz, sampling our signal with  $f_s = 1$  kHz and working with  $m = 2$  would lead MSE analyses to characterize dynamics at frequencies which are too high to be physiologically relevant. With this sampling frequency, a 20 Hz cycle would be associated with 50 samples and the MSE analyses utilizing two or three-sample sequences would deal with dynamics that are much greater than 20 Hz. There is the option to increase  $m$  to around 50 to ultimately include data encompassing 20 Hz or lower frequencies, however, this would introduce other unintended and undesired effects—namely, decreased number of matches and diminished confidence in SampEn (to be explained in Section 4.4). In these cases, down-sampling of the data prior to data analyses would be recommended.

#### 4.4. Sequence Length ( $m$ ), and Point Matching Tolerance ( $r$ )

The selection of  $m$  and  $r$  is driven by two overarching factors: (1) maximizing the accuracy and confidence in the SampEn values obtained at each MSE scale and (2) optimizing the ability to distinguish any real, salient features in the dataset. In principle, the accuracy and confidence of the entropy estimate improve as the numbers of matches of length  $m$  and  $m + 1$  increase. The number of matches can be increased by choosing small  $m$  (short templates) and large  $r$  (wide tolerance). However, a larger  $r$  will result in a conditional property (A/B) of 1 and thus a SampEn of zero for nearly all stationary time series, thereby limiting one's ability to discriminate between various time series. On the other hand,  $r$  must be large enough to avoid the influence of noise and to simultaneously increase probability of matches to ensure that confidence in SampEn is adequate [27].

A much more quantitative approach to seeking a value of  $r$  was advocated by Lake *et al.*, in 2002 [41]. In this study, Lake derived the variance,  $\sigma_{CP}$ , of the conditional probability (CP) of A/B where CP represents the probability of a match of length  $m + 1$  given there is a match of length  $m$ :

$$\sigma_{CP}^2 = \frac{CP(1-CP)}{B} + \frac{1}{B^2}(K_A - K_B(CP)^2) \quad (4)$$

where B is again the number of template matches of length  $m$ ,  $K_A$  is the number of overlapping pairs of matching templates of length  $m + 1$  and  $K_B$  is the number of overlapping pairs of matching templates of length  $m$ . Selection of the value  $r$  is then determined by maximizing the following quantity:

$$\max \left( \frac{\sigma_{CP}}{CP}, \frac{\sigma_{CP}}{-\log(CP)CP} \right) \quad (5)$$

which is the maximum of the relative error of SampEn and of the CP estimate, respectively.

To identify the optimal value for sequence length  $m$ , a number of techniques have been utilized. Selecting the appropriate value for  $m$  has its basis on the fact that  $m$  determines where the information content is being assessed. Since SampEn is essentially a marker of how much new information is generated, it is important to ensure that the template matches for  $m$  and  $m + 1$  are within the vicinity of where the important dynamics are present.

To identify the template lengths associated with sufficient information content and thus the optimal range of  $m$ , Lake *et al.* [41] employed an autoregressive model while Chen *et al.* [42] instead utilized a mutual information method and false nearest neighbor (FNN) technique which is more appropriate for nonlinear time series. These considerations, although applicable to SampEn analyses of raw time-series, are largely negated by the process of coarse-graining and the utilization of multiple scales in MSE. As a result, the choice of  $m$  is relatively arbitrary for MSE but becomes more a function of data logistics:  $m = 2$  is superior to  $m = 1$  since it allows more detailed reconstruction of the joint probabilistic dynamics while  $m > 2$  is unfavorable due to the requirement of larger data lengths [42].

Numerous studies have taken more of an empirical approach to this issue by observing the effects of varying  $m$  and  $r$  on the calculated MSE results. In our systematic review, Duarte *et al.* [22] and Jiang *et al.* [23] performed such evaluations and have concluded that while absolute changes

in complexity values are observed, relative changes were insignificant [9,22]. Indeed, according to Duarte *et al.*, the relative results remain generally consistent when  $r$  is swept between 10% and 30% [22], suggesting that this range should be sufficient for most data sets. Similarly, Pincus *et al.*, has found that entropy analyses produce statistically reliable and reproducible results with  $m = 2$  and  $r = 10\%–25\%$  and an appropriate data length [37]. Our systematic review reveals that the selection of  $m$  and  $r$  are relatively consistent across studies: sequence length  $m$  is typically 2 and point matching tolerance  $r$  is either 15% or 20%.

#### 4.5. Filtering

Filtering raw data is a critical pre-processing step for MSE analysis. General trends and low frequency drifts, in particular, can lead to diminished sequence matching and incorrectly ascertained increase in irregularity manifesting as a higher SampEn. Moreover, the infrequent sequence matching corresponds to a widened confidence interval for the derived SampEn values. Nonstationarities at higher frequencies may also have unpredictable effects on the calculated SampEn values. To remove such effects, Empirical Mode Decomposition (EMD) is the technique most commonly used as demonstrated in Table 1.

EMD is well-suited for decomposition of nonlinear, nonstationary physiologic signals and possesses advantages over Fourier and wavelet analysis because it employs a fully adaptive approach derived by means of a sifting process [8]. Unlike Fourier or wavelet methods, there are no a priori assumptions about the nature of a signal and it does not rely on a specific basis (e.g., sinusoidal or Haar wavelet function) for decomposing the signal. Fourier based filtering of nonlinear, nonstationary signals can produce undesired artifacts in the outputted signal. After decomposition by EMD the resulting IMFs can be recombined in various permutations, representing a range of characteristic frequencies which are a subset of the original signals bandwidth. This resulting signal can then be analyzed with methods such as MSE.

EMD is not without its limitations as it is susceptible to mode-mixing and end-effect issues. Mode mixing occurs when an oscillation at a particular frequency is not fully isolated to a single IMF but rather leaked to adjacent IMFs. Ensemble Empirical Mode Decomposition (EEMD) minimizes mode mixing through the implementation of noise-assisted sifting [24]. End effects represent errors that occur at the beginning or end of an IMF due to the EMD process. To enable proper decomposition of the edges of a time-series, values must be appended at the boundaries in an appropriate manner. Improper additions or extensions can lead to unwanted distortions. A detailed review of EEMD which addresses mode-mixing and end-effect issues can be found in Wu and Huang [24].

Generally, removal of nonstationarities that have characteristics well outside the frequencies of interest is not difficult to accomplish through the use of EMD or other filtering methods. However sudden, transient movements, such as shifts and fidgets, can also cause nonstationarities with predominant frequencies within the frequencies of interest since they are simply larger versions of the complex postural sway adjustments seen on a regular basis. Due to this overlap in frequencies, these particular nonstationarities may commonly persist despite the filtering step. A more detailed discussion on technique (Fourier-based, wavelet, EMD) selection for the filtering of biomedical signals can be found in Fonseca-Pinto [43].

##### 4.5.1. Nonstationarities within Frequency Band of Interest

The presence of a single large nonstationarity can generate significant changes in the calculated tolerance window,  $r$ . As noted previously,  $r$  is directly proportional to the signal's standard deviation (at Scale 1) and importantly is established thereafter for all scales. A large spike or extrema can increase  $r$ , increase the number of template matches, and thus decrease the overall complexity index  $C_I$ . As a consequence, a lower  $C_I$  can be paradoxically construed as either increased regularity or larger presence of spurious extremas. This phenomenon is depicted in Figure 3 where the large nonstationarity starting at 34 s, potentially due to the subject shifting, greatly increases the standard

deviation of the signal and therefore the value of  $r$ . We explore how this nonstationarity affects the SampEn calculation at scale 15. In Figure 4a the nonstationarity is included while in Figure 4b it is excluded. As shown, this results in more template matches in the former case and therefore lower sample entropy. The absolute difference in SampEn at this coarse grained level is substantial:  $|0.872-1.902| = 1.03$ .

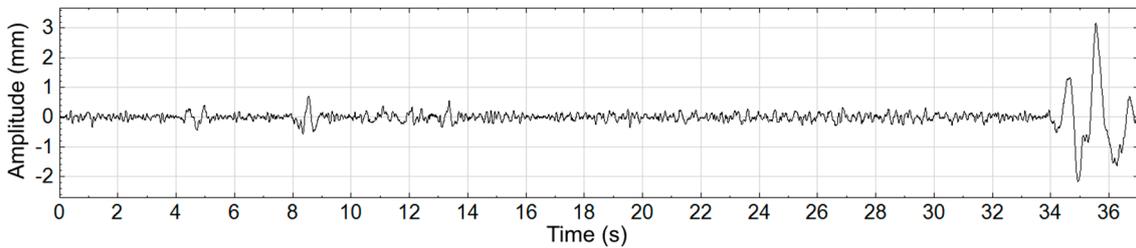


Figure 3. Center-of-Pressure waveform shown with a large nonstationarity between time 34 s and 37 s.

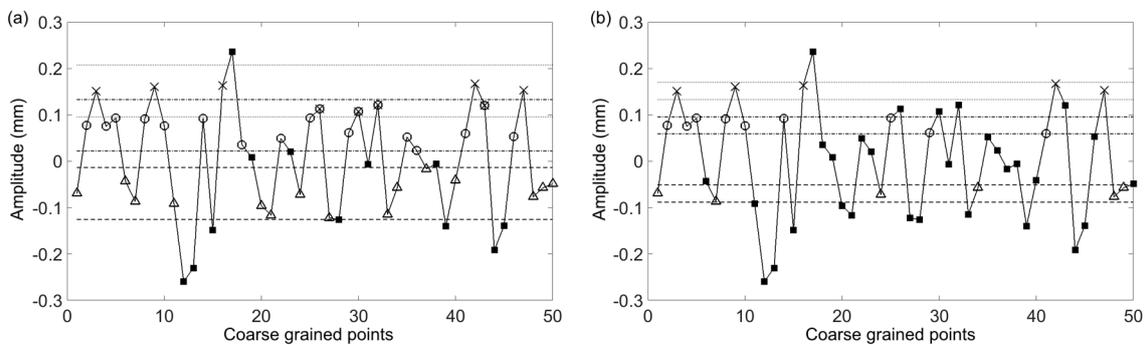


Figure 4. The first 50 coarse-grained points from Figure 3 with  $\tau = 15$ . The straight (dashed, dotted, dashed-dotted) lines represent the point matching tolerance,  $r$ , based on a standard deviation which includes the nonstationarity (a) and does not include the nonstationarity (b). Two sequence template matches are represented by  $\Delta$ - $\circ$  vectors which are comprised of matches to the first ( $\Delta$ ) and second ( $\circ$ ) points from the first template. Three sequence template matches are represented by  $\Delta$ - $\circ$ - $\times$ , where the next point ( $\times$ ) matches the third point from the template. Points which do not match any template points are represented by  $\blacksquare$  symbols. In (a) due to the large nonstationarity, the calculated standard deviation is large enough to cause overlap between the tolerance about the 2nd (dash-dot line) and 3rd points in the template sequence. This results in certain points matching both the 2nd and 3rd points as indicated by markers with both an  $\circ$  and an  $\times$ . Because of the wide tolerance the complexity index will be less than what it would be without the large nonstationarity. In (b) it is observed that exclusion of the nonstationarity results in tighter tolerance about the template sequence points. In turn this will result in a larger complexity index. Adapted from [6].

Different approaches have been attempted to deal with this dilemma. Some researchers have used a fixed standard deviation [22] irrespective of the time-series variability. Although this approach removes sensitivity to nonstationarities, it is also less adaptive to the variability in amplitudes seen across subjects. Subjects who exhibit larger amplitudes will generally be associated with a greater complexity,  $C_I$ , and *vice versa*. Other researchers have chosen to remove the nonstationarity from the time-series [22]. Conceptually, removal of such nonstationarities—which may occur frequently in certain cases—constitutes removal of information which may be an important aspect of the systems dynamics. For this reason, we seek to preserve the intrinsic structure of the signal as much as it is feasible. Lastly, some have used a median absolute deviation (MAD) in place the signal’s standard deviation [44]. The MAD is computed by taking the median of the absolute deviations between the data’s median. This approach is again less sensitive to large nonstationarities but due to the inherent

difference between MAD (based on median of absolute differences) and standard deviation (based on variance, which is the average of squared differences), the comparative relationship between MAD-based MSE and the traditional MSE algorithm is unclear. One possible solution to this issue of nonstationarities is proposed here.

#### 4.5.2. Windowed Standard Deviation MSE

An alternative method for determining the point matching tolerance  $r$  is the windowed standard deviation—herein referred to as windowed-MSE or WMSE. In this approach, the standard deviation is calculated for a fixed width window as it is stepped across the time-series as opposed to calculating standard deviation for the entire time-series. The window is stepped one sample at a time until the end of the time series, and in the process generating N-n standard deviation calculations, where  $n$  is the window width. The median value for all N-n standard deviations is then determined and subsequently used to calculate the point matching tolerance  $r$ , which is subsequently applied to all scales.

The window width  $n$  should be set such that each window provides a reasonable estimate of the population standard deviation. For a normally distributed time series with no outliers, to be 95% confident that the error between the window and population standard deviation is less than 10%, the window width must be 240 samples. The confidence interval for the window standard deviation estimate of the population standard deviation,  $\sigma$ , can be calculated using:

$$\left[ \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}}} \right] \tag{6}$$

where  $n$  is the window width,  $s$  is the sample (or window) standard deviation,  $\chi^2$  is the chi-squared distribution for a given significance level,  $\alpha$ , with degrees of freedom,  $n - 1$  [45]. With the estimate of 240 samples as provided by Equation (6) with an arbitrary  $s$  and  $\alpha$  of 95%, we can be reasonably confident that each of our windows is providing an accurate estimate of the entire time-series (population) standard deviation.

For time-series without existing nonstationarities, WMSE produces results very similar to that of the traditional MSE approach, since standard deviation remains the means by which the point matching tolerance  $r$  is calculated. For time-series with sporadic nonstationarities, WMSE deemphasizes the nonstationarities and yields a larger  $C_1$  relative to the traditional MSE method, as should be expected.

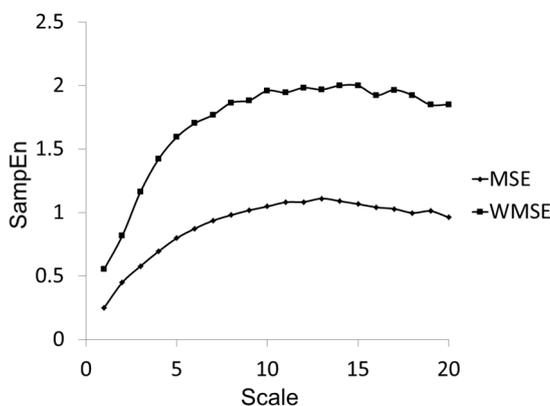


Figure 5. Multiscale Entropy and Windowed Multiscale Entropy calculations for the waveform shown in Figure 3 with a large nonstationarity.

Figure 5 provides an example of the difference in the SampEn Vs  $\tau$  curves for MSE and WMSE calculations using the time-series depicted in Figure 3 with a large nonstationarity. Table 2 highlights the details of the standard deviation result for this waveform using MSE and WMSE. It is evident that the inclusion of the nonstationarity in the time series (0–37 s) generates a much different standard deviation as compared to that seen with exclusion of the nonstationarity (0–34 s) when determined by the traditional algorithm (MSE column). However the standard deviation derived using the WMSE approach is much more robust against the effects of including the nonstationarity (WMSE column). Since the standard deviation as calculated by WMSE is smaller in the nonstationarity case, we see higher sample entropies at a given scale in Figure 5 for the WMSE curve.

**Table 2.** Differences in the standard deviation result between MSE and WMSE for the waveform shown in Figure 3. The results are shown including the nonstationarity (with) at the end and not including it (without).

Nonstationarity	MSE	WMSE
With (0–37 s)	0.3718	0.1214
Without (0–34 s)	0.1234	0.1214

## 5. Conclusions

This systematic review has revealed significant heterogeneity in the way MSE is applied to COP displacement data. Part of the heterogeneity arises from the lack of clarity regarding the methodological challenges involved in MSE-based analyses. We recommend that prior to testing, future studies should consider establishing these important factors: the minimal amount of time for data collection, the physiological frequencies to evaluate, the inclusion of healthy controls, and sampling rate for data acquisition. Once the data is collected, the researchers must then decide how the data should be filtered, what values  $m$  and  $r$  should be assigned, and how to address the nonstationarities that persist despite the filtering process. These recommendations are summarized in flowcharts in Appendix A.

As MSE increases in popularity, modifications of the MSE methodological algorithm will likely arise with corresponding changes in the way the parameters are assigned. Already, different variants of MSE have been published, and this review does not include them due to their sheer number, their limited employment to COP studies, and—at present—their lack of mature development. Nevertheless, many of the methodological challenges discussed here still apply, and this paper intends to help researchers understand how to properly design their studies and to analyze their data using MSE. Further discussions about these methodological issues should hopefully enhance consistency across studies in both reporting and possibly methodology for MSE analyses of COP data and other continuous real-world time series. In turn, accurate and consistent results for the MSE assessment of physiological signals will help determine whether MSE gains more traction as a clinical biomarker. The concept of complexity and health is still novel in the clinical setting but could become an important part of patient diagnoses in the future.

**Acknowledgments:** The authors thank Madalena D. Costa for helpful discussions regarding nonstationarities which led to the idea for computing sample entropy with an  $r$  value dependent on the time series' local (windowed) standard deviation. Andrew Ahn's work was made possible through the generous support from The Institute for Integrative Health. Peter Wayne and Brian Gow's work was made possible by grant number R21 AT005501-01A1 from the National Center for Complementary and Alternative Medicine (NCCAM: <http://nccam.nih.gov/>) at the National Institutes of Health (NIH: <http://www.nih.gov/>), and from grant number UL1 RR025758 supporting the Harvard Clinical and Translational Science Center, from the National Center for Research Resources (NCRR: <http://www.nih.gov/about/almanac/organization/NCRR.htm>). Chung-Kang Peng's work was supported by grant (NSC 102-2911-I-008-001) from the Ministry of Science and Technology of Taiwan.

**Author Contributions:** Brian J. Gow, Chung-Kang Peng, Peter M. Wayne and Andrew C. Ahn have read and approved the final manuscript.

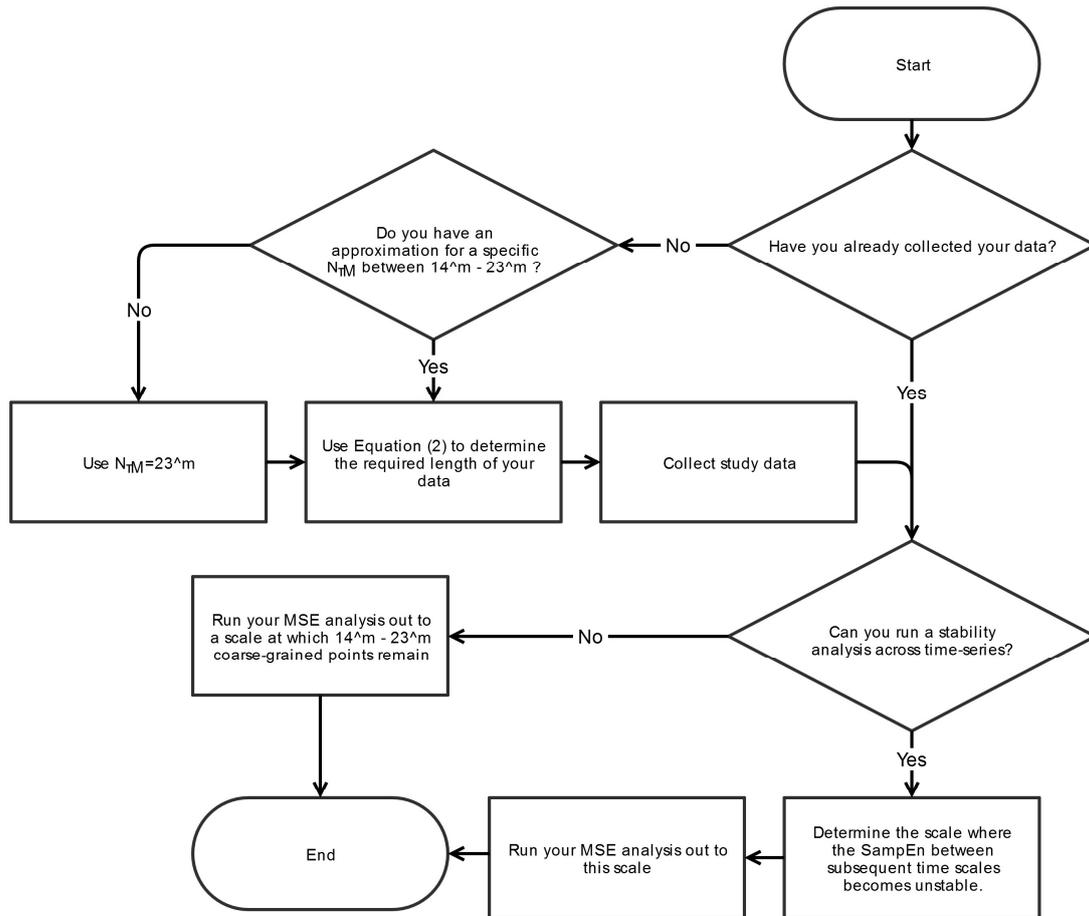
**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix**

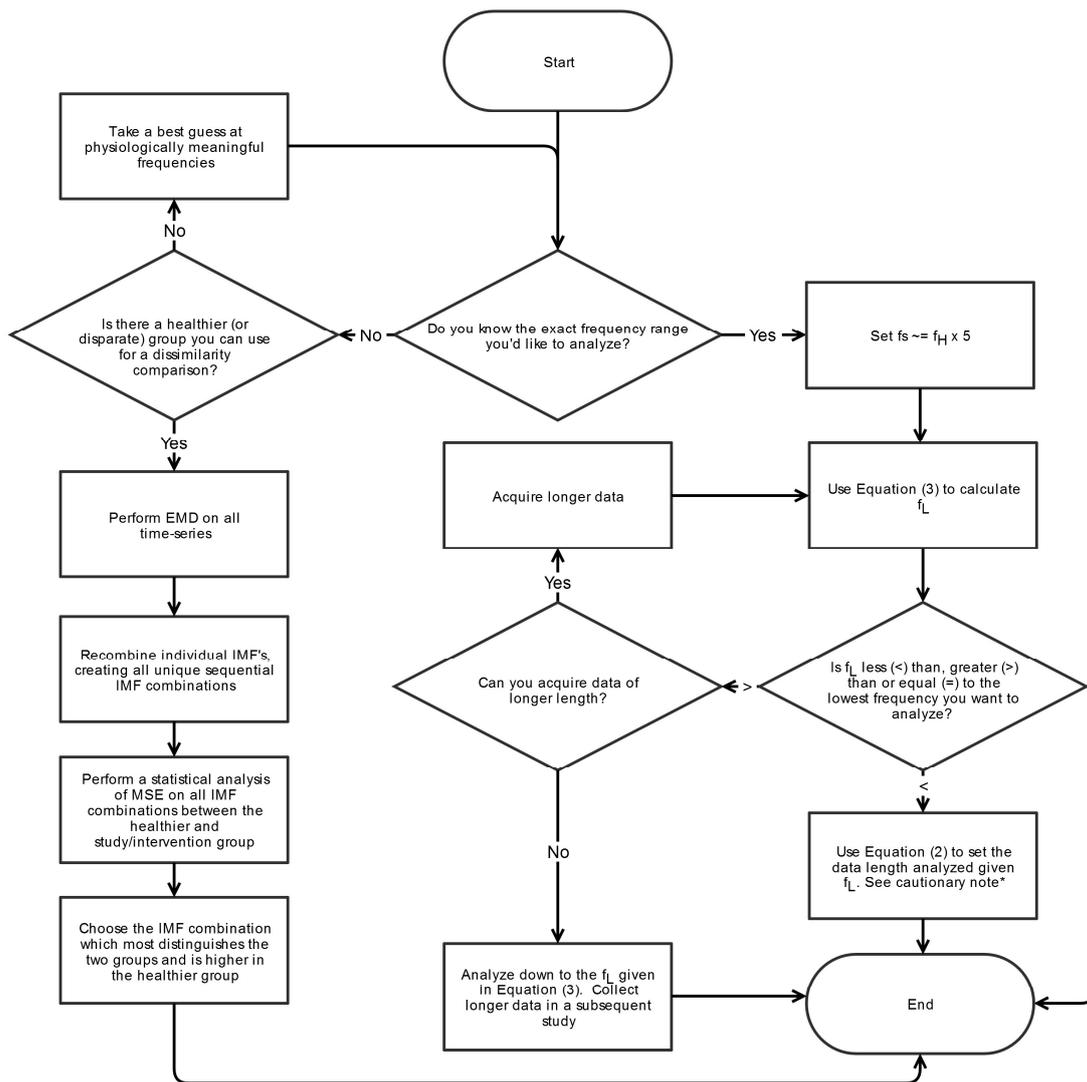
**A. Flowcharts**

The following flowcharts provide a visual representation of our recommendations for each section in 4. These flowcharts can be used when making decisions specific to a given study.

Stability analysis—involves looking for a large change in SampEn from one time scale to the next followed by another large change in the opposite direction between the next two scales.



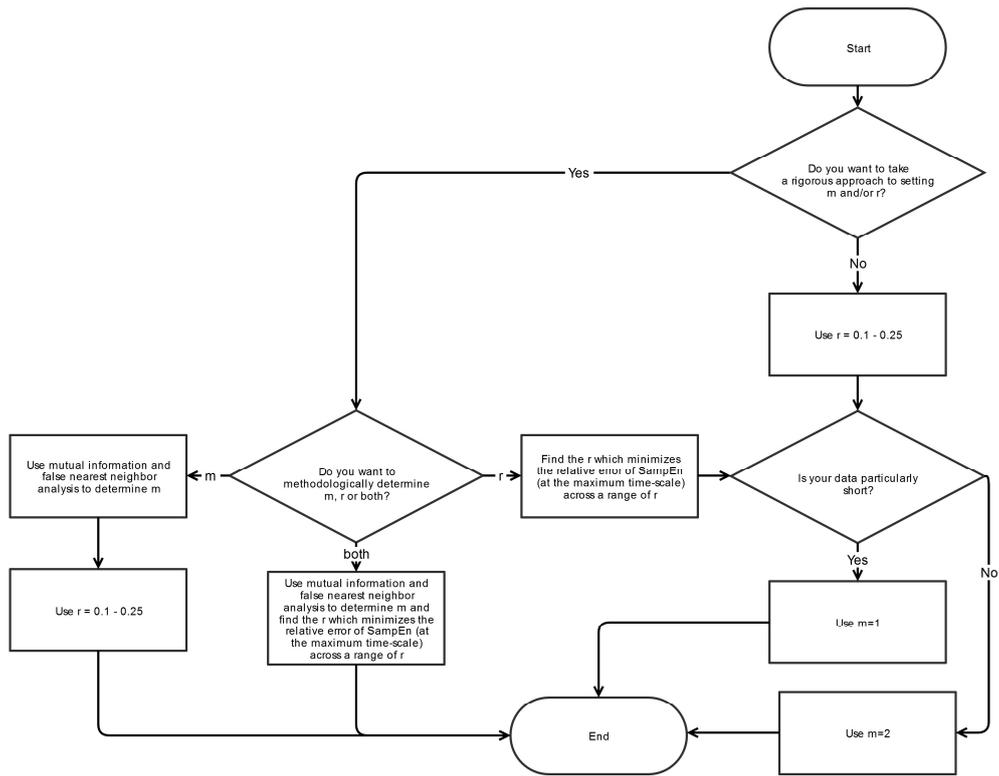
**Figure A1.** Determining Required Data Length.  $N_{\tau M}$ —Number of points remaining at the last time scale which dictates whether the SampEn calculation will be accurate. The ^ symbol represents to the power.



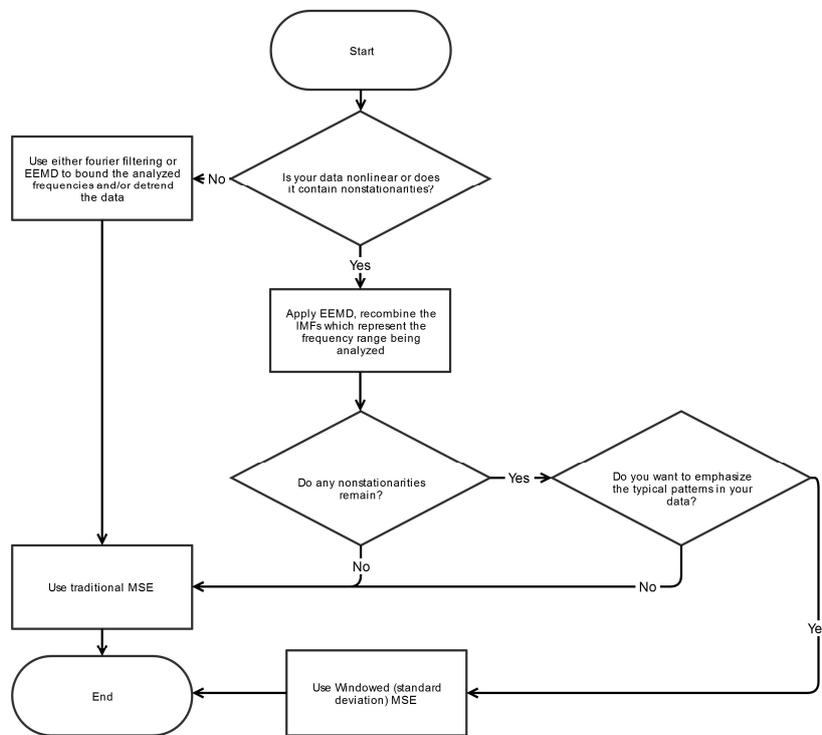
**Figure A2.** Range of Frequencies for Analysis.  $f_s$ —sampling rate;  $f_H$ —highest frequency included in analysis;  $f_L$ —lowest frequency included in analysis; dissimilarity comparison—is a statistical analysis between a healthier (or otherwise disparate group) and the study group at baseline to determine which frequencies best distinguish the groups; \* Cautionary note: Since the premise of MSE is that complexity exists across all scales; excluding available data should only be done when there is a clear reason for doing so.



**Figure A3.** Appropriate sample rate ( $f_s$ ).  $f_s$ —sampling rate;  $f_H$ —highest frequency included in analysis; dissimilarity comparison—is a statistical analysis between a healthier (or otherwise disparate group) and the study group at baseline to determine which frequencies best distinguish the groups.



**Figure A4.** Selection of sequence length ( $m$ ), and point matching tolerance ( $r$ ). Mutual information and False Nearest Neighbor Reference: Chen *et al.* [42]; Minimizing the relative error of SampEn Reference: Lake *et al.* [41].



**Figure A5.** Filtering.

## References

1. Lipsitz, L.A.; Goldberger, A.L. Loss of ‘Complexity’ and Aging. *J. Am. Med. Assoc.* **1992**, *267*, 1806–1809. [[CrossRef](#)]
2. Goldberger, A.L.; Giles, F. Filley lecture. Complex Systems. *Proc. Am. Thorac. Soc.* **2006**, *3*, 467–471. [[CrossRef](#)] [[PubMed](#)]
3. Goldberger, A.L.; Amaral, L.A.N.; Hausdorff, J.M.; Ivanov, P.C.; Peng, C.-K.; Stanley, H.E. Fractal dynamics in physiology: Alterations with disease and aging. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2466–2472. [[CrossRef](#)] [[PubMed](#)]
4. Costa, M.; Goldberger, A.; Peng, C.-K. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Phys. Rev. Lett.* **2002**, *89*, 6–9. [[CrossRef](#)] [[PubMed](#)]
5. Norris, P.R.; Anderson, S.M.; Jenkins, J.M.; Williams, A.E.; Morris, J.A., Jr. Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 trauma patients. *Shock* **2008**, *30*, 17–22. [[CrossRef](#)] [[PubMed](#)]
6. Costa, M.; Goldberger, A.; Peng, C.-K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 1–18. [[CrossRef](#)] [[PubMed](#)]
7. Ferrario, M.; Signorini, M.G.; Magenes, G.; Cerutti, S. Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 119–125. [[CrossRef](#)] [[PubMed](#)]
8. Costa, M.; Priplata, A.A.; Lipsitz, L.A.; Wu, Z.; Huang, N.E.; Goldberger, A.L.; Peng, C.-K. Noise and poise: Enhancement of postural complexity in the elderly with a stochastic-resonance-based therapy. *Europhys. Lett.* **2007**, *77*, 68008. [[CrossRef](#)] [[PubMed](#)]
9. Costa, M.; Peng, C.-K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Phys. A Stat. Mech. Its Appl.* **2003**, *330*, 53–60. [[CrossRef](#)]
10. Costa, M.; Ghiran, I.; Peng, C. Complex dynamics of human red blood cell flickering: Alterations with in vivo aging. *Phys. Rev. E* **2008**, *78*, 1–10. [[CrossRef](#)] [[PubMed](#)]
11. Manor, B.; Costa, M.D.; Hu, K.; Newton, E.; Starobinets, O.; Kang, H.G.; Peng, C.K.; Novak, V.; Lipsitz, L.A. Physiological complexity and system adaptability: evidence from postural control dynamics of older adults. *J. Appl. Physiol.* **2010**, *109*, 1786–1791. [[CrossRef](#)] [[PubMed](#)]
12. Wei, Q.; Liu, D.-H.; Wang, K.-H.; Liu, Q.; Abbod, M.; Jiang, B.; 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. [[CrossRef](#)]
13. Thuraisingham, R.A.; Gottwald, G.A. On multiscale entropy analysis for physiological data. *Phys. Stat. Mech. Appl.* **2006**, *366*, 323–332. [[CrossRef](#)]
14. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[PubMed](#)]
15. Costa, M.; Goldberger, A.L.; Peng, C.-K. Multiscale entropy to distinguish physiologic and synthetic RR time series. *Comput. Cardiol.* **2002**, *29*, 137–140. [[PubMed](#)]
16. Musha, T.; Yamamoto, M. 1/f Fluctuations in Biological Systems. In Proceedings of the 19th Annual International IEEE Conference of the Engineering in Medicine and Biology Society, Chicago, IL, USA, 30 October–2 November 1997.
17. Valencia, J.F.; Porta, A.; Vallverdú, M.; Clarià, F.; Baranowski, R.; Orłowska-Baranowska, E.; Caminal, P. Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2202–2213. [[CrossRef](#)] [[PubMed](#)]
18. Shumway-Cook, A.; Woollacott, M.H. *Motor Control: Translating Research Into Clinical Practice*, 3rd ed.; Wolters Kluwer: Philadelphia, PA, USA, 2007.
19. Baltich, J.; von Tscharnner, V.; Zandiyeh, P.; Nigg, B.M. Quantification and reliability of center of pressure movement during balance tasks of varying difficulty. *Gait Posture* **2014**, *40*, 327–332. [[CrossRef](#)] [[PubMed](#)]
20. Baltich, J.; Whittaker, J.; Von Tscharnner, V.; Nettel-Aguirre, A.; Nigg, B.M.; Emery, C. The impact of previous knee injury on force plate and field-based measures of balance. *Clin. Biomech.* **2015**, *30*, 832–838. [[CrossRef](#)] [[PubMed](#)]
21. Hu, M.; Liang, H. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 12–15. [[PubMed](#)]

22. Duarte, M.; Sternad, D. Complexity of human postural control in young and older adults during prolonged standing. *Exp. Brain Res.* **2008**, *191*, 265–276. [[CrossRef](#)] [[PubMed](#)]
23. Jiang, B.C.; Yang, W.; Shieh, J.; Fan, J.S.-Z.; Peng, C.-K. Entropy-based method for COP data analysis. *Theor. Issues Ergon. Sci.* **2013**, *14*, 227–246. [[CrossRef](#)]
24. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
25. Kang, H.G.; Costa, M.D.; Priplata, A.A.; Starobinets, O.V.; Goldberger, A.L.; Peng, C.-K.; Kiely, D.K.; Cupples, L.A.; Lipsitz, L.A. Frailty and the degradation of complex balance dynamics during a dual-task protocol. *J. Gerontol. A Biol. Sci. Med. Sci.* **2009**, *64*, 1304–1311. [[CrossRef](#)] [[PubMed](#)]
26. Gruber, A.H.; Busa, M.A.; Gorton, G.E., III; van Emmerik, R.E.A.; Masso, P.D.; Hamill, J. Time-to-contact and multiscale entropy identify differences in postural control in adolescent idiopathic scoliosis. *Gait Posture* **2011**, *34*, 13–18. [[CrossRef](#)] [[PubMed](#)]
27. Kirchner, M.; Schubert, P.; Schmidtbleicher, D.; Haas, C.T. Evaluation of the temporal structure of postural sway fluctuations based on a comprehensive set of analysis tools. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 4692–4703. [[CrossRef](#)]
28. Huang, C.-W.; Sue, P.-D.; Abbod, M.F.; Jiang, B.C.; Shieh, J.-S. Measuring center of pressure signals to quantify human balance using multivariate multiscale entropy by designing a force platform. *Sensors* **2013**, *13*, 10151–10166. [[CrossRef](#)] [[PubMed](#)]
29. Manor, B.; Lipsitz, L.A.; Wayne, P.M.; Peng, C.-K.; Li, L. Complexity-based measures inform Tai Chi's impact on standing postural control in older adults with peripheral neuropathy. *BMC Complement. Altern. Med.* **2013**, *13*, 87. [[CrossRef](#)] [[PubMed](#)]
30. Fournier, K.A.; Amano, S.; Radonovich, K.J.; Bleser, T.M.; Hass, C.J. Decreased dynamical complexity during quiet stance in children with Autism Spectrum Disorders. *Gait Posture* **2014**, *39*, 420–423. [[CrossRef](#)] [[PubMed](#)]
31. Chen, M.-S.; Jiang, B.C. Resistance Training Exercise Program for Intervention to Enhance Gait Function in Elderly Chronically Ill Patients: Multivariate Multiscale Entropy for Center of Pressure Signal Analysis. *Comput. Math. Methods Med.* **2014**, *2014*, 1–10. [[CrossRef](#)] [[PubMed](#)]
32. Pau, M.; Kim, S.; Nussbaum, M.A. Fatigue-induced balance alterations in a group of Italian career and retained firefighters. *Int. J. Ind. Ergon.* **2014**, *44*, 615–620. [[CrossRef](#)]
33. Wayne, P.M.; Gow, B.J.; Costa, M.D.; Peng, C.-K.; Lipsitz, L.A.; Hausdorff, J.M.; Davis, R.B.; Walsh, J.N.; Lough, M.; Novak, V.; *et al.* Complexity-Based Measures Inform Effects of Tai Chi Training on Standing Postural Control: Cross-Sectional and Randomized Trial Studies. *PLoS ONE* **2014**, *9*, e114731. [[CrossRef](#)] [[PubMed](#)]
34. Yeh, J.-R.; Lo, M.-T.; Chang, F.-L.; Hsu, L.-C. Complexity of human postural control in subjects with unilateral peripheral vestibular hypofunction. *Gait Posture* **2014**, *40*, 581–586. [[CrossRef](#)] [[PubMed](#)]
35. Decker, L.M.; Ramdani, S.; Tallon, G.; Jaussent, A.; Bernard, P.L.; Blain, H. Physical function decline and degradation of postural sway dynamics in asymptomatic sedentary postmenopausal women. *J. Nutr. Health Aging* **2015**, *19*, 348–355. [[CrossRef](#)] [[PubMed](#)]
36. Zhou, D.; Zhou, J.; Chen, H.; Manor, B.; Lin, J.; Zhang, J. Effects of transcranial direct current stimulation (tDCS) on multiscale complexity of dual-task postural control in older adults. *Exp. Brain Res.* **2015**, *233*, 2401–2409. [[CrossRef](#)] [[PubMed](#)]
37. Pincus, S.M.; Goldberger, A.L. Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol.* **1994**, *266*, H1643–H1656. [[PubMed](#)]
38. Wu, S.-D.; Wu, C.-W.; Lee, K.-Y.; Lin, S.-G. Modified multiscale entropy for short-term time series analysis. *Phys. Stat. Mech. Appl.* **2013**, *392*, 5865–5873. [[CrossRef](#)]
39. Humeau-Heurtier, A. The Multiscale Entropy Algorithm and Its Variants: A Review. *Entropy* **2015**, *17*, 3110–3123. [[CrossRef](#)]
40. Nilsson, J.; Panizza, M.; Hallett, M. Principles of digital sampling of a physiologic signal. *Electroencephalogr. Clin. Neurophysiol.* **1993**, *89*, 349–358. [[PubMed](#)]
41. Lake, D.E.; Richman, J.S.; Griffin, M.P.; Moorman, J.R. Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2002**, *283*, R789–R797. [[CrossRef](#)] [[PubMed](#)]
42. Chen, X.; Solomon, I.; Chon, K. Comparison of the use of approximate entropy and sample entropy: applications to neural respiratory signal. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2005**, *4*, 4212–4215. [[PubMed](#)]

43. Fonseca-Pinto, R. A New Tool for Nonstationary and Nonlinear Signals: The Hilbert-Huang Transform in Biomedical Applications. In *Biomedical Engineering, Trends in Electronics, Communications and Software*; Laskovski, A., Ed.; InTech: Rijeka, Croatia, 2011; pp. 481–504.
44. Govindan, R.B.; Wilson, J.D.; Eswaran, H.; Lowery, C.L.; Preißl, H. Revisiting sample entropy analysis. *Phys. Stat. Mech. Appl.* **2007**, *376*, 158–164. [[CrossRef](#)]
45. Devore, J.L. *Probability and Statistics for Engineering and the Sciences*, 8th ed.; Brooks/Cole: Boston, MA, USA, 2012.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).