



HAL
open science

Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants Compared to Full-Length Long Signals

Airton Borin, Anne Humeau-Heurtier, Luiz Virgílio Silva, Luiz Murta

► **To cite this version:**

Airton Borin, Anne Humeau-Heurtier, Luiz Virgílio Silva, Luiz Murta. Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants Compared to Full-Length Long Signals. *Entropy*, 2021, 23 (12), pp.article id. 1620. 10.3390/e23121620 . hal-03462890

HAL Id: hal-03462890

<https://univ-angers.hal.science/hal-03462890v1>

Submitted on 15 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.





L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants Compared to Full-Length Long Signals

Airton Monte Serrat Borin, Jr. ¹, Anne Humeau-Heurtier ², Luiz Eduardo Virgílio Silva ³
and Luiz Otávio Murta, Jr. ^{4,*}

¹ Federal Institute of Education, Science and Technology of Triangulo Mineiro, Uberaba 38064-790, Brazil; airton@iftm.edu.br

² LARIS—Laboratoire Angevin de Recherche en Ingénierie des Systèmes, University of Angers, 49035 Angers, France; anne.humeau@univ-angers.fr

³ Department of Internal Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto 14049-900, Brazil; luizeduardo@usp.br

⁴ Department of Computing and Mathematics, School of Philosophy, Sciences and Languages of Ribeirão Preto, University of São Paulo, Ribeirão Preto 14040-901, Brazil

* Correspondence: murta@usp.br

Abstract: Multiscale entropy (MSE) analysis is a fundamental approach to access the complexity of a time series by estimating its information creation over a range of temporal scales. However, MSE may not be accurate or valid for short time series. This is why previous studies applied different kinds of algorithm derivations to short-term time series. However, no study has systematically analyzed and compared their reliabilities. This study compares the MSE algorithm variations adapted to short time series on both human and rat heart rate variability (HRV) time series using long-term MSE as reference. The most used variations of MSE are studied: composite MSE (CMSE), refined composite MSE (RCMSE), modified MSE (MMSE), and their fuzzy versions. We also analyze the errors in MSE estimations for a range of incorporated fuzzy exponents. The results show that fuzzy MSE versions—as a function of time series length—present minimal errors compared to the non-fuzzy algorithms. The traditional multiscale entropy algorithm with fuzzy counting (MFE) has similar accuracy to alternative algorithms with better computing performance. For the best accuracy, the findings suggest different fuzzy exponents according to the time series length.

Keywords: multiscale fuzzy entropy; time series



Citation: Borin, A.M.S.J.; Humeau-Heurtier, A.; Virgílio Silva, L.E.; Murta, L.O.J. Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants Compared to Full-Length Long Signals. *Entropy* **2021**, *23*, 1620. <https://doi.org/10.3390/e23121620>

Academic Editor: Geert Verdoolaege

Received: 8 October 2021

Accepted: 28 November 2021

Published: 1 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Complex systems are composed of many agents interacting with each other by nonlinear rules and exhibiting temporal and spatial structures at different scales [1]. Quantifying the complexity level from realizations of the system's dynamics, i.e., time series, is still a challenge. While different interpretations of complexity may be assumed, entropy certainly plays a role in the estimation of time series complexity [2,3].

Most entropy estimators for time series are inspired by the Kolmogorov-Sinai (KS) entropy, e.g., sample entropy [4] and fuzzy entropy [5]. They intend to estimate the rate at which the information grows over time in the system. However, the introduction of multiscale entropy (MSE) [6] was a milestone in the field of complexity analysis since the multiscale aspects of the system's dynamics can now be taken into account. The MSE algorithm is based on a coarse-graining procedure for generating the scaled versions of the original dynamics followed by the calculation of sample entropy for each scaled series.

Although MSE showed itself worthwhile in discriminating different complex dynamics, it introduces bias when dealing with short-term time series. First, the coarse-graining procedure of MSE drastically decreases the time series's length, decreasing the number of

points available for entropy estimation. Second, the sample entropy algorithm is based on similar pattern counting, and short time series may result in a biased or even undefined value of entropy. MSE proves to be inaccurate in short time series analysis [7,8], significantly losing its sensitivity [9]. To overcome these limitations, some approaches propose different coarse-graining procedures and entropy estimation. Composite MSE (CMSE) [10], refined composite MSE (RCMSE) [11], and modified MSE (MMSE) [7,12] are important examples. Multiscale fuzzy entropy (MFE) uses a fuzzy membership function to identify similarities between patterns within time series, avoiding zero counts or numeric instabilities in entropy calculation [13–15].

Some studies also combined the advantages of improved coarse-graining procedures with the fuzzy entropy estimation. Composite and refined composite multiscale fuzzy entropy (CMFE and RCMFE) present a joint approach with CMSE and RCMSE, respectively. Both CMFE and RCMFE were evaluated in the biomedical and non-biomedical contexts [15–18]. In a recent study, we proposed and evaluated the modified multiscale fuzzy entropy (MMFE) for heart rate variability (HRV) analysis [19]. In a systematic comparison, we showed that MMFE is more robust for estimating original MSE than MMSE when the fuzzy parameter is optimized.

This study systematically compares the accuracy of CMSE, RCMSE, MMSE, MFE, CMFE, RCMFE, and MMFE to estimate the original MSE in a short time series. The study considered MSE estimated for full length series as reference to calculate the errors for each method variations. Each method accuracy is analyzed for different lengths of series, and the dependence of the best fuzzy exponent on the series length is reported. We seek to find the most accurate and cost-effective multiscale entropy measure for short time series.

2. Materials and Methods

We handled experiments with long HRV time series obtained from two biological databases (rats and humans) and exhaustively applied CMSE, RCMSE, MMSE, MFE, CMFE, RCMFE, and MMFE to different sizes of segments. The minimum error compared to the original long-term MSE was considered as the maximum accuracy. The corresponding multiscale entropy algorithms are briefly detailed in the following subsections.

For all methods, consider a time series \mathbf{u} with N samples defined as $u(1), u(2), \dots, u(N)$. We define m as the length of vectors (patterns) to be compared and r as the tolerance accepted between corresponding points within the vectors. This tolerance is defined as a percentage of the original time series SD. There are $N - m\delta + 1$ template vectors $\mathbf{x}_m(i)$ for $\{i \mid 1 \leq i \leq N - m\delta + 1\}$, where $\mathbf{x}_m(i) = \{u(i + k\delta) : 0 \leq k \leq m - 1\}$ is a vector with length m and δ is the delay considered between samples.

2.1. Sample Entropy (SampEn) and Fuzzy Entropy (FuzzyEn)

The sample entropy (SampEn) algorithm [4,20] and the fuzzy entropy (FuzzyEn) algorithm [5] calculate the distance between any two vectors as

$$d[\mathbf{x}_m(i), \mathbf{x}_m(j)] = \max_{0 \leq k \leq m-1} |u(i + k\delta) - u(j + k\delta)|, \quad (1)$$

where $j > i + \delta$.

2.1.1. SampEn

First, B_i is calculated as the number of matches for the template vector $\mathbf{x}_m(i)$, i.e., the number of vectors $\mathbf{x}_m(j)$ which distances $d[\mathbf{x}_m(i), \mathbf{x}_m(j)]$ are less than or equal to r , for $0 \leq j \leq N - m\delta$. Next, A_i is calculated as the number of matches for the template vector $\mathbf{x}_{m+1}(i)$, i.e., the number of vectors $\mathbf{x}_{m+1}(j)$ which the distances $d[\mathbf{x}_{m+1}(i), \mathbf{x}_{m+1}(j)]$ are less than or equal to r , for $0 \leq j \leq N - m\delta$. Then, $C^m(r)$ and $C^{m+1}(r)$ are computed as

$$C_i^m = \frac{B_i}{N - m\delta - 1}, \quad (2)$$

$$C^m(r) = \frac{1}{N - m\delta} \sum_{i=1}^{N-m\delta} C_i^m(r), \tag{3}$$

and

$$C_i^{m+1} = \frac{A_i}{N - m\delta - 1}, \tag{4}$$

$$C^{m+1}(r) = \frac{1}{N - m\delta} \sum_{i=1}^{N-m\delta} C_i^{m+1}(r). \tag{5}$$

Then, *SampEn* is obtained as the negative logarithm of the conditional probability $C^{m+1}(r)/C^m(r)$, estimated with the parameters m, r , and δ :

$$SampEn(\mathbf{u}, m, r, \delta) = -\ln \frac{C^{m+1}(r)}{C^m(r)} \tag{6}$$

2.1.2. FuzzyEn

FuzzyEn is based on the concept of fuzzy sets [21], defining the similarity levels between vectors by the fuzzy associative (membership) function and the vectors' distances. Vectors $\mathbf{x}_m(i)$ are created similarly to *SampEn*, except by the fact that the mean vector baseline is removed:

$$\mathbf{x}_m(i) = \{u(i + k\delta) - u_0(i) : 0 \leq k \leq m - 1\} \tag{7}$$

where

$$u_0(i) = \frac{1}{m} \sum_{j=0}^{m-1} u(i + j\delta). \tag{8}$$

To calculate the similarity between two vectors, two functions were tested in our work: $\exp(-d_m^n/r)$ and $\exp(-0.6931(d/r)^n)$ [22]. For the first function, we computed

$$B_{ij}^m(r) = \exp(-d_m^n/r) \tag{9}$$

and

$$A_{ij}^m(r) = \exp(-d_{m+1}^n/r), \tag{10}$$

where d is given in Equation (1) and n is the exponent of the fuzzy function. For the second function, $B_{ij}^m(r)$ and $A_{ij}^m(r)$ were computed similarly but using $\exp(-0.6931(d/r)^n)$.

We also define

$$\phi^m(n, r) = \frac{1}{N - m\delta} \sum_{i=1}^{N-m\delta} B_{ij}^m(r). \tag{11}$$

and

$$\phi^{m+1}(n, r) = \frac{1}{N - m\delta} \sum_{i=1}^{N-m\delta} A_{ij}^{m+1}(r), \tag{12}$$

similar to Equations (3) and (5), so *FuzzyEn* for the parameters m, n, δ , and r is calculated by

$$FuzzyEn(\mathbf{u}, m, n, r, \delta) = -\ln \frac{\phi^{m+1}(n, r)}{\phi^m(n, r)}. \tag{13}$$

2.2. Multiscale Entropy (MSE) and Multiscale Fuzzy Entropy (MFE)

In the *MSE* [6,8,23] and *MFE* [13–15] algorithms, the dynamics of a system at different time scales is obtained by a moving average procedure (*coarse-graining* procedure), according to

$$u^\tau(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u(i), \quad 1 \leq j \leq N/\tau. \tag{14}$$

The irregularities in the time series for the scale factor τ are quantified by applying *SampEn* (*FuzzyEn*) for *MSE* (*MFE*) on the coarse-grained time series obtained, with unitary delay ($\delta = 1$), that is

$$MSE(\mathbf{u}, m, r) = SampEn(\mathbf{u}^\tau, m, r, \delta = 1) \tag{15}$$

and

$$MFE(\mathbf{u}, m, n, r) = FuzzyEn(\mathbf{u}^\tau, m, n, r, \delta = 1). \tag{16}$$

2.3. Composite Multiscale Entropy (CMSE) and Composite Multiscale Fuzzy Entropy (CMFE)

In *CMSE* [10] and *CMFE* [15], for each scale factor τ , k coarse-grained time series \mathbf{y}_k^τ are obtained where the elements of the k -th series are defined as

$$y_k^\tau(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} u_i, \quad 1 \leq j \leq \frac{N}{\tau}, \quad 1 \leq k \leq \tau. \tag{17}$$

CMSE and *CMFE* at scale τ are defined as the average entropy obtained from the k series at scale τ , that is

$$CMSE(\mathbf{u}, \tau, m, r) = \frac{1}{\tau} \sum_{k=1}^{\tau} SampEn(\mathbf{y}_k^\tau, m, r, \delta = 1) \tag{18}$$

and

$$CMFE(\mathbf{u}, \tau, m, r) = \frac{1}{\tau} \sum_{k=1}^{\tau} FuzzyEn(\mathbf{y}_k^\tau, m, n, r, \delta = 1). \tag{19}$$

2.4. Refined Composite Multiscale Entropy (RCMSE) and Refined Composite Multiscale Fuzzy Entropy (RCMFE)

The procedure to obtain the coarse-grained series in *RCMSE* [11] and *RCMFE* [24] is the same as for *CMSE* and *CMFE* (see Equation (17)). However, instead of averaging the entropy of each k scaled series for scale τ , entropy is estimated from the average number of matches $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^m$, obtained from all k coarse-grained series.

2.4.1. RCMSE

RCMSE is defined as:

$$RCMSE(\mathbf{u}, \tau, m, r) = -\ln\left(\frac{\bar{n}_{k,\tau}^{m+1}}{\bar{n}_{k,\tau}^m}\right), \tag{20}$$

where $\bar{n}_{k,\tau}^{m+1} = \frac{1}{\tau} \sum_{k=1}^{\tau} n_{k,\tau}^{m+1}$ and $\bar{n}_{k,\tau}^m = \frac{1}{\tau} \sum_{k=1}^{\tau} n_{k,\tau}^m$ are the averages of $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^m$ respectively.

2.4.2. RCMFE

Given $\bar{\phi}_\tau^{m+1}$ and $\bar{\phi}_\tau^m$ as the averages of ϕ^{m+1} and ϕ^m for each k at the scale factor τ , respectively, *RCMFE* is defined by:

$$RCMFE(\mathbf{u}, \tau, m, n, r) = -\left(\ln \frac{\bar{\phi}_{k,\tau}^{m+1}}{\bar{\phi}_{k,\tau}^m}\right), \tag{21}$$

where ϕ^m and ϕ^{m+1} are given by Equations (11) and (12), respectively.

2.5. Modified Multiscale Entropy (MMSE) and Modified Multiscale Fuzzy Entropy (MMFE)

In *MMSE* and *MMFE*, the scaled versions \mathbf{z}^τ of the original time series are created according to the following procedure:

$$z^\tau(i) = \frac{1}{\tau} \sum_{k=i}^{i+\tau-1} u(k), \quad 1 \leq i \leq N - \tau + 1, \quad (22)$$

where τ represents the time scale factor. This procedure is similar to the one adopted in *MSE* (Equation (14)), except for the overlapping in the moving average. In this procedure, the length of the *coarse-grained* time series obtained using the overlapping moving average for a scalar of τ is $N - \tau + 1$, remarkably greater compared to the length of the *coarse-graining* procedure of *MSE* (N/τ).

2.5.1. MMSE

The *MMSE* method [10] proposes that the *coarse-grained* time series be constructed by Equation (22) and that the entropy is estimated for each scale factor τ by applying *SampEn* with a time delay equal to τ , that is:

$$MMSE(\mathbf{u}, m, \tau, r) = SampEn(\mathbf{z}^\tau, m, r, \delta = \tau). \quad (23)$$

2.5.2. MMFE

MMFE was recently proposed [19] and consists in the application of the same *coarse-graining* procedure as *MMSE* (Equation (22)) followed by the estimation of entropy using a delayed version of *FuzzyEn*, with a delay equal to τ for each scale. The equation of *MMFE* is given by:

$$MMFE(\mathbf{u}, n, m, \tau, r) = FuzzyEn(\mathbf{z}^\tau, m, n, r, \delta = \tau). \quad (24)$$

3. Dataset and Experiments

3.1. Dataset

Heart rate variability (HRV) series from rats and humans were obtained from previous studies [19,25]. The first group of ECG data was recorded in 18 healthy Wistar rats. The recordings were performed in the Cardiovascular Physiology Laboratory of Ribeirão Preto Medical Schools, University of São Paulo. Briefly, the rats had their ECG recorded for approximately 1 h (40 to 80 min) at baseline conditions. Computer software (LabChart, ADInstruments, Sydney, Australia) was used to create RR series from ECG recordings, sampled at 2 kHz. All RR series were visually inspected for artifacts and corrected when necessary. Since the time series's length varied from 15,892 to 32,333 points, all RR series were truncated to 15,892 points. The second group of HRV series consisted of 12 healthy human individuals, obtained from the Physionet MIT-BIH Normal Sinus Rhythm database digitized at 360 Hz per signal lead [26]. The 12 ECG recordings were selected randomly from the database. The RR series were calculated using the *ann2rr* tool from the WFDB Physionet package, which uses the recordings' beat annotations to calculate the RR intervals. Only normal RR intervals were considered, that is, the intervals between two successive normal beats. Eventually, all RR series were truncated to 15,892 samples so that the series of rats and humans had the same length. The recording period of all series ran from 8 a.m. to 10 p.m.

3.2. Experiments

We segmented the HRV full series (15,892 points) into equal segments of 400, 800, 1200, ..., 15,600 points, with a superimposition of 90% to the previous segment. For each segment, all the variants of *MSE* described above were computed, and the average value over the segments with the same size was taken to represent the whole series. The maximum scale factor assessed was 20, i.e., $\tau = 1, 2, \dots, 20$. The embedding dimension and tolerance factor of entropy estimators were set as $m = 2$ and $r = 0.15 \times SD$ of the

series, respectively. To evaluate each MSE variant's accuracy on the estimation of the original MSE, the mean square error was calculated over all time scales, always taking the original MSE, obtained from the full-length series, as reference. The error was calculated for each series and each segment size, and the mean errors were reported as a function of the segment size.

The mean squared error is obtained by calculating the entropy for time scales from 1 to 20 for each segment of 400, 800, . . . , 15,600 points. If the number of windows is greater than 1, the arithmetic average of each scaling factor is made. For example: for the 400-point segment, we have 385 windows. For each window, we calculated entropy on the 20 time scales. Then, the arithmetic mean of the 385 windows is taken for each scaling factor, so the entropy for scale 1 and the 400-point segment is the entropy average for scale 1 of all 385 windows. The entropy for scale 2 and the segment of 400 points is the average entropy on scale 2 of all 385 windows, so up to scale 20. As a result, we have an average entropy for scales 1, 2, . . . , 20. We then average the mean squared error of these entropy values with the calculated MSE entropy on a scale of 1 to 20.

Moreover, to assess the cost-effectiveness of fuzzy-based MSE variants, we measured each algorithm's average computation time. The analysis was performed on a desktop computer equipped with an Intel Core i7 930@2.8 GHz processor and 16 Gb of RAM. To guarantee the isonomy of the results, all the tests were performed with the MATLAB software (The MathWorks, Inc. Natick, MA, USA) and the `maxNumCompThreads = 1` command so that all methods used a single CPU. The average time consumed to process three randomly selected human HRV series is reported as a segment size function (from 400 to 12,000 points). The fuzzy exponent n adopted in this experiment followed the equation previously found for the choice of the best exponent according to the segment size (x) [19]:

$$n = 0.82 + 0.10 \exp(-3 \times x/10^4). \quad (25)$$

4. Results

The accuracy of CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE, and MFE were evaluated as the error compared to the MSE calculated using the full-length series. Figure 1 shows the accuracy of MFE obtained with both rats and humans HRV series. The descriptive measures of the MSE for both the human and animal datasets can be found in a previous paper [19]. The top left plot shows the mean squared errors as a function of the rat dataset's segment size. We illustrate different error curves for different fuzzy exponents ranging from $n = 0.8$ to $n = 1.5$. The magnification of the errors' curves for short segment sizes are shown at the bottom left corner, and the minimum error for each segment size is shown on the right side of the magnification plot. The mean squared errors are shown as a segment size function for the humans' dataset at the top right corner. We illustrate different error curves for different fuzzy exponents ranging from $n = 0.85$ to $n = 0.92$. One can find the magnification of the errors' curves for short segment sizes at the bottom right corner, as well as the minimum errors obtained for each segment size.

As can be seen in Figure 1, the error for each fuzzy exponent n depends on the segment size (series length), and the optimal n are the ones that provide the lowest errors. For the HRV series from rats, the best exponents increase with the segment size, while it decreases for human HRV series.

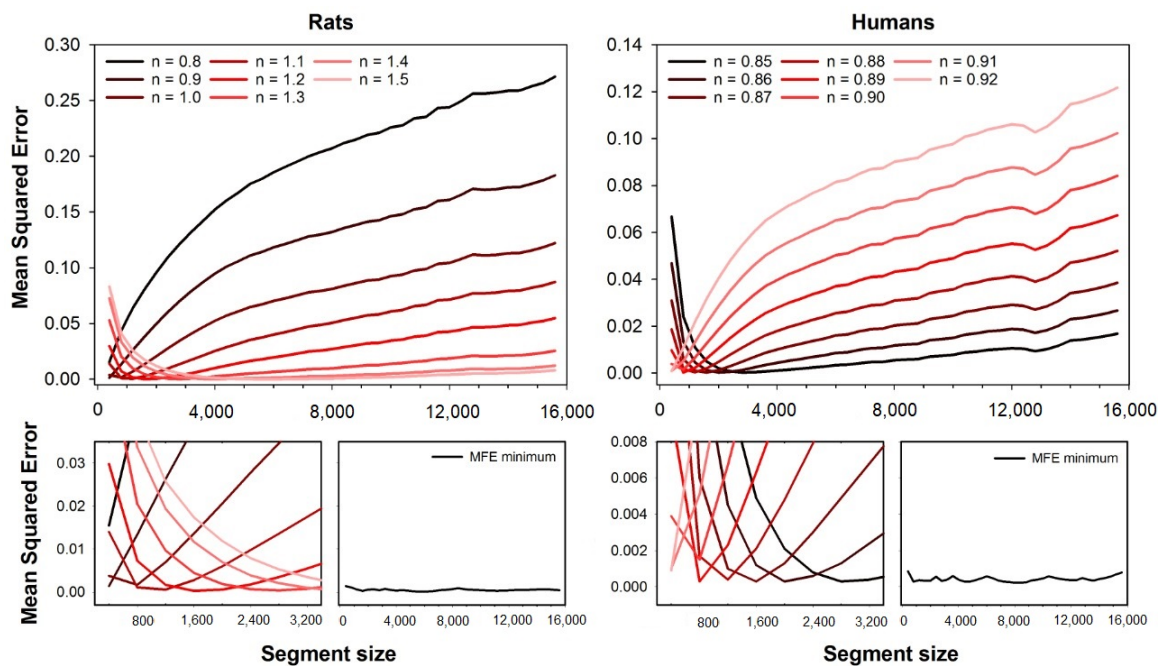


Figure 1. Mean squared errors of MFE as a function of the segment size. The errors were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats) and $n = 0.85$ to $n = 0.92$ (humans). Top plots show all the error curves, while bottom plots show magnification of short segment sizes' errors. One can notice that the fuzzy exponent that gives the best accuracy (lower error) varies according to the segment size. The lowest error for all segment sizes is illustrated in the plots at the magnification plots' right (MFE minimum).

Figure 2 shows the mean squared errors for CMSE (dashed line) and CMFE (solid lines). Error curves for CMFE are illustrated for fuzzy exponents ranging from $n = 0.8$ to $n = 1.5$ (rats) and $n = 0.85$ to $n = 0.92$ (humans), similar to Figure 1. At the bottom plots, one can see the magnifications of error curves for short segment sizes and the minimum errors for CMFE (black line) compared to CMSE (gray line). Similar to the results with MFE, the errors are dependent on the segment size. The best exponents increase with the segment size for rats while it decreases for human HRV series.

Figure 3 presents the error curves for RCMSE (dashed line) and RCMFE (solid lines). Error curves are illustrated for the fuzzy exponents ranging from $n = 0.8$ to $n = 1.5$ (rats) and $n = 0.85$ to $n = 0.92$ (humans). Magnification of the error curves and the minimum errors for all segment sizes are shown at the bottom plots. Like MFE and CMFE, the errors of RCMFE are dependent on the segment size, and the best exponents increase with the segment size for rats, while it decreases for human HRV series.

The error curves obtained with MMFE and MMSE are shown in Figure 4. Although a similar error plot can be found in a previous study [19], here we expanded the range of exponents evaluated to calculate the minimum MMFE error. Error curves are illustrated for the fuzzy exponents ranging from $n = 0.8$ to $n = 1.5$ (rats) and $n = 0.85$ to $n = 0.92$ (humans). Magnification of the error curves, together with plots of the minimum MMFE and MMSE errors for all segment sizes, are shown at bottom plots. Like all the other fuzzy-based MSE variants, the errors of MMFE depend on the segment size, and the best exponents increase with the segment size for rats, while it decreases for human HRV series.

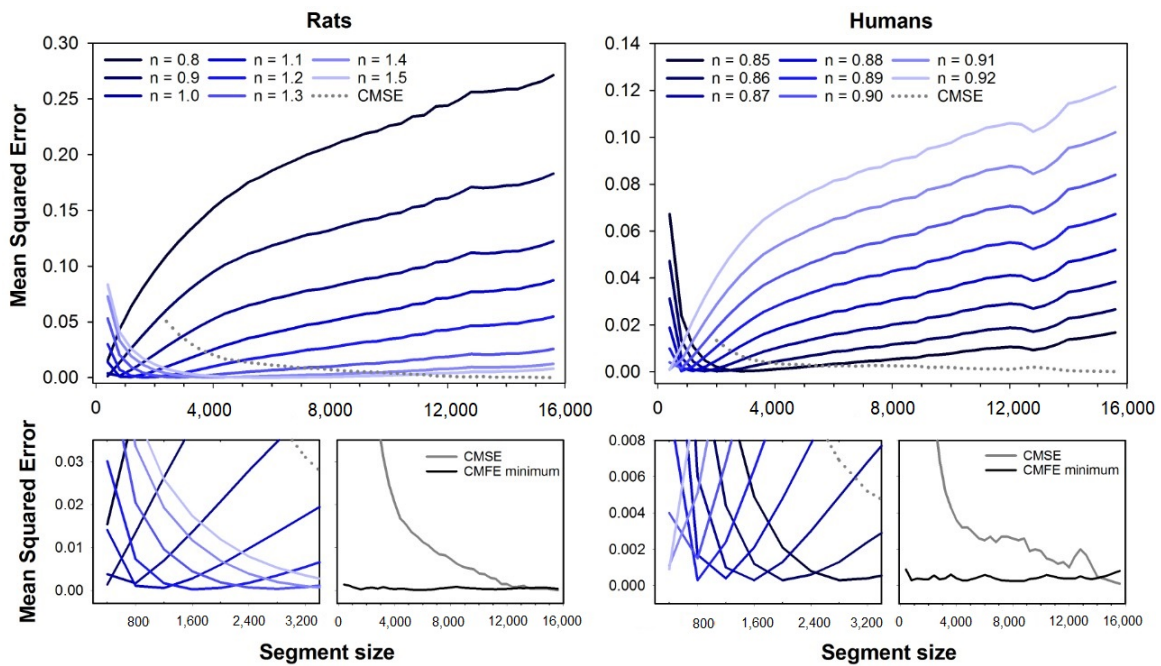


Figure 2. Mean squared errors of CMFE and CMSE as a function of the segment size. The errors of CMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). **Top** plots show the full error curves of CMFE (solid lines) and CMSE (dashed line), while **bottom** plots show a magnification of the errors for short segment sizes. One can notice that the fuzzy exponent that gives the best accuracy (lower error) for CMFE varies according to the segment size. The lowest error for CMFE (black lines) and CMSE (gray lines) for all segment sizes are illustrated in the plots at the magnification plots' right.

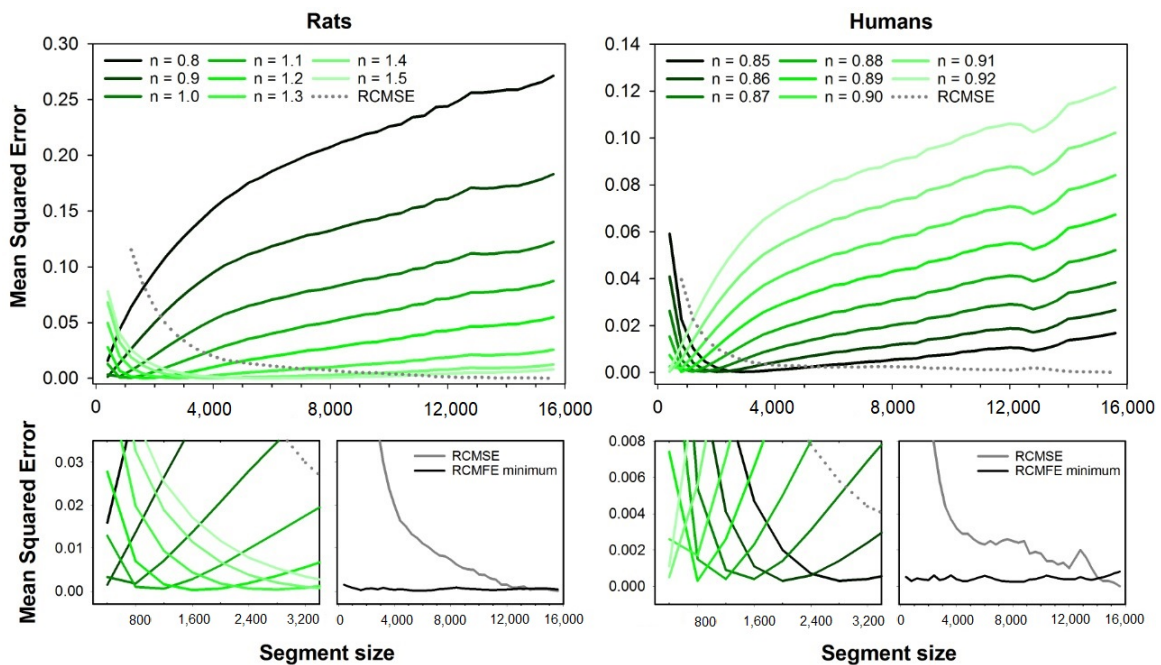


Figure 3. Mean squared errors of RCMFE and RCMSE as a function of the segment size. The errors of RCMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). **Top** plots show the full error curves of RCMFE (solid lines) and RCMSE (dashed line), while **bottom** plots show magnification of short segment sizes' errors. One can notice that the fuzzy exponent that gives the best accuracy (lower error) for RCMFE varies according to the segment size. The lowest error for RCMFE (black lines) and RCMSE (gray lines) for all segment sizes are illustrated in the plots at the magnification plots' right.

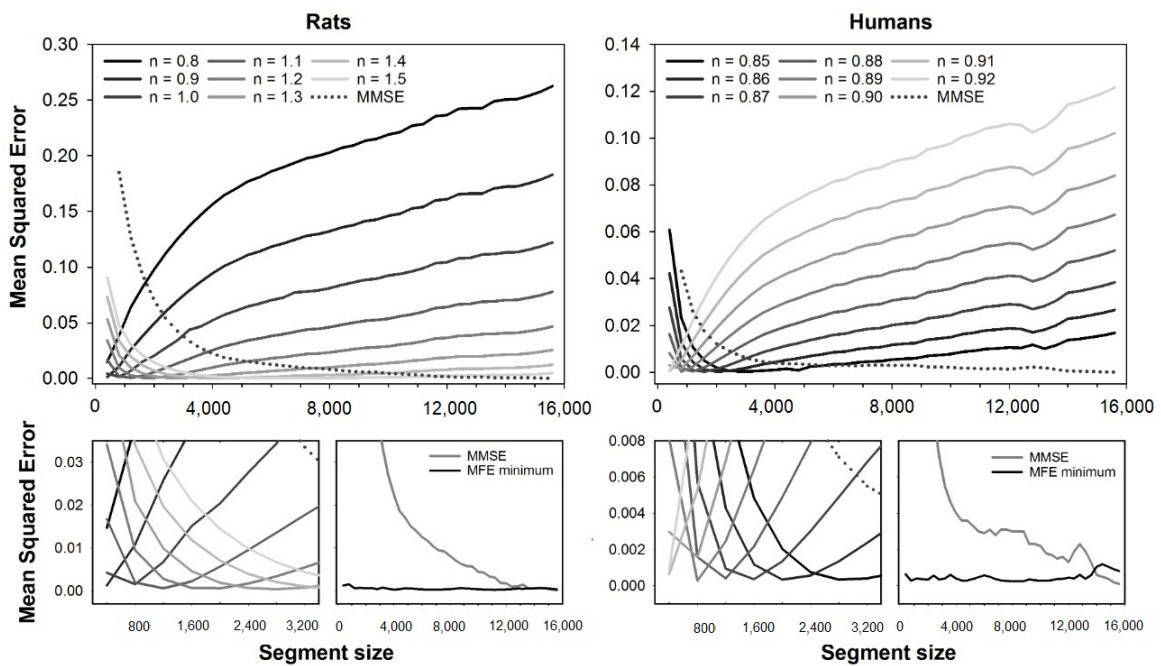


Figure 4. Mean squared errors of MMFE and MMSE as a function of the segment size. The errors of MMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). **Top** plots show the full error curves of MMFE (solid lines) and MMSE (dashed lines), while **bottom** plots show a magnification of the errors for short segment sizes. One can notice that the fuzzy exponent that gives the best accuracy (lower error) for MMFE varies according to the segment size. The lowest error for MMFE (black lines) and MMSE (gray lines) for all segment sizes are illustrated in the plots at the magnification plots' right.

Figure 5 compares errors from all the multiscale variants studied, i.e., CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE, and MFE. The minimum error is shown for fuzzy entropy-based methods, calculated with each segment size's optimal fuzzy exponent. The figure shows that all variants based on diffuse entropy have fewer errors than any variant based on sample entropy. For segments sized up to 13,000 points, the MFE, CMFE, and RCMFE curves are superimposed because these methods have similar results. Mean squared error for all considered approaches, i.e., CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE, and MFE. For fuzzy entropy-based approaches, only the minimum error is shown, obtained with the optimal fuzzy exponent for each segment size. Results are shown for both rats (left) and human (right) database. Note that the errors are calculated regarding the MSE of full-length time-series, i.e., 15,892 beats.

Figure 6 shows the optimal fuzzy exponents of fuzzy-based MSE variants for each segment size. These exponents were utilized to calculate the minimum error curves for multiscale fuzzy entropy-based variants (see Figure 5). The curves were fitted to exponential functions, which can be employed to find the best fuzzy exponent of those datasets according to the time series length. Note that for the HRV series from rats, the optimal fuzzy exponent increases with the series length, decreasing the HRV series of humans.

For the sake of comparison, Figure 7 shows the mean square error of MFE using the alternative fuzzy function $\exp(-0.6931 \times (d/r)^n)$. The errors are illustrated for the fuzzy exponent ranging from $n = 1.3$ to $n = 3.0$ (rats) and from $n = 2.0$ to $n = 5.5$ (humans). Results show that MFE with this alternative fuzzy function also presents dependence on the segment size, similar to the original fuzzy function (Figure 1). However, the values of optimal exponents for each segment size are markedly different from the original fuzzy function, and the best n increases with the segment size for both rats and humans HRV series.

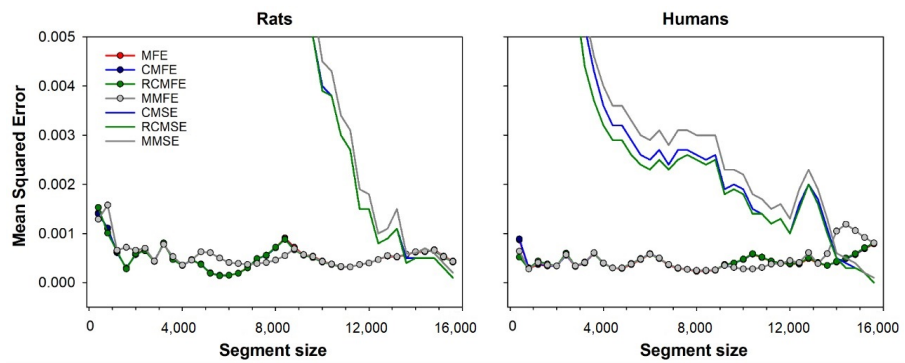


Figure 5. Mean squared error for all considered approaches, i.e., CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE and MFE. For fuzzy entropy-based approaches, only the minimum error is shown, obtained with the optimal fuzzy exponent for each segment size. Results are shown for both rats (left) and human (right) database. Notice that the errors are calculated regarding the MSE of full-length time-series, i.e., 15,892 beats. The figure shows that all variants based on diffuse entropy have fewer errors than any variant based on sample entropy. For segments sized up to 13,000 points, the MFE, CMFE, and RCMFE curves are superimposed because these methods have similar results.

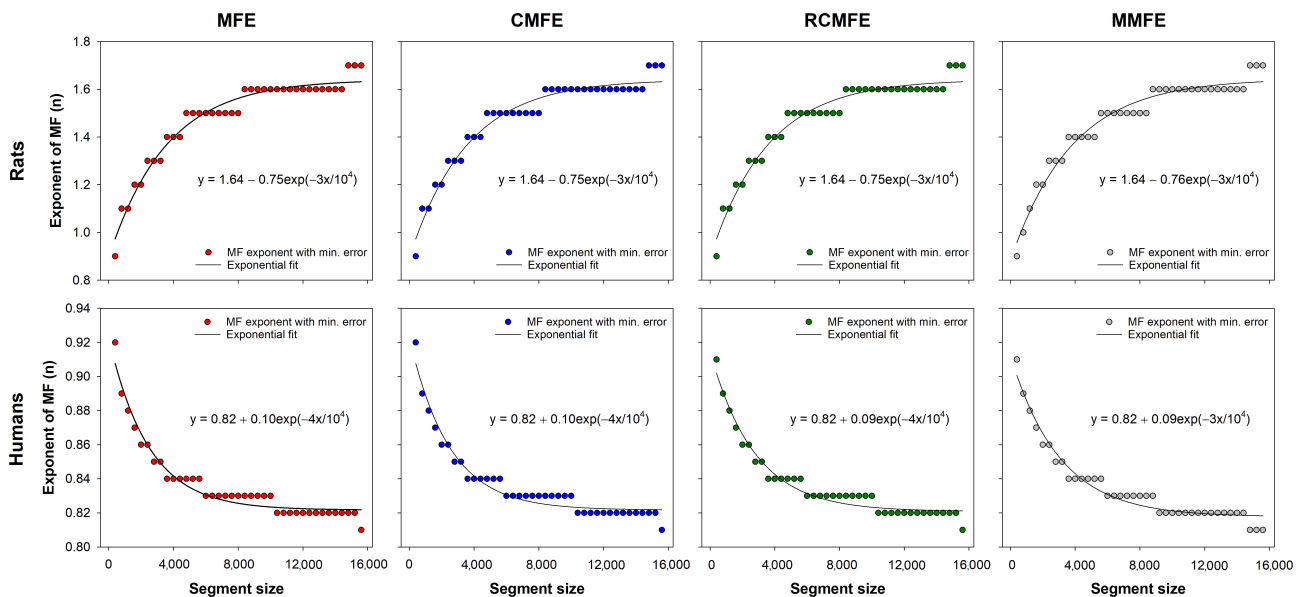


Figure 6. Best fuzzy exponents found for each segment size in both rats and humans HRV datasets. The solid lines represent the best fitting functions, i.e., a decreasing exponential function $y = 1.64 - 0.75 \exp(-3x/10^4)$ for rats and an increasing exponential $y = 0.82 - 0.10 \exp(-4x/10^4)$ for humans. These functions can be used to choose the best fuzzy exponent of these datasets according to the series length.

Figure 8 shows the average time (over three human HRV series) spent calculating the different multiscale fuzzy entropy-based variants at increasing segment sizes (up to 12,000 samples). As expected, the computational time required to run any method increases with the segment size. However, MMFE has the highest computational cost, while MFE showed the lowest. CMFE and RCMFE show virtually the same computational time.

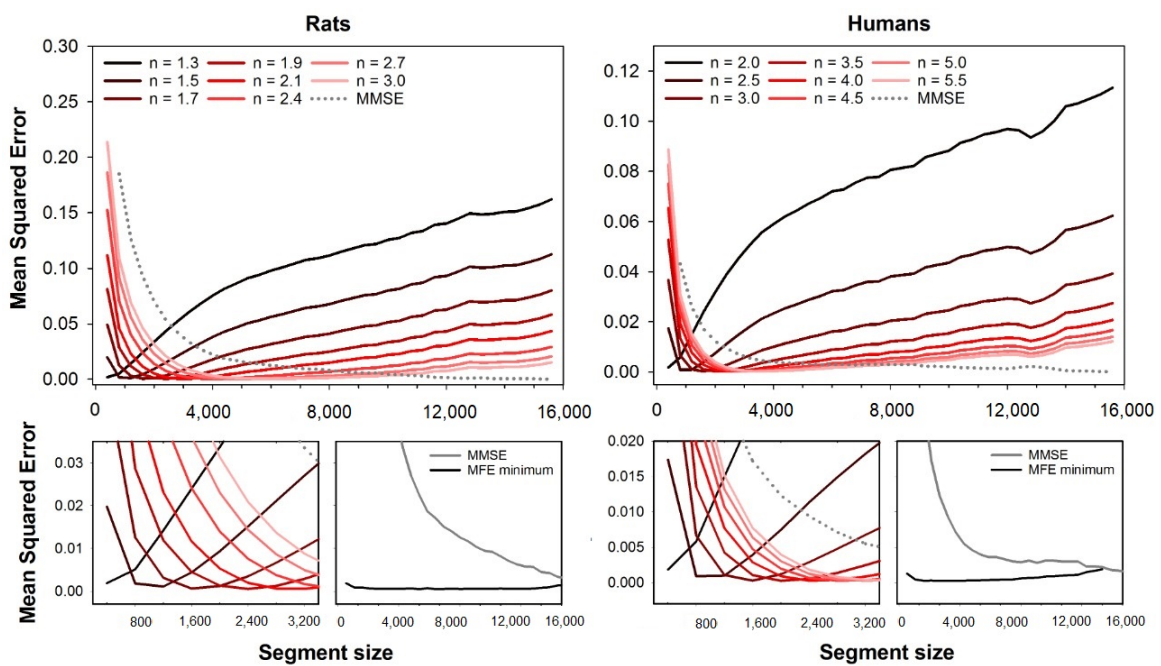


Figure 7. Mean squared errors of MFE as a function of the segment size, obtained with alternative fuzzy function, i.e., $\exp(-0.6931 \times (d/r)^n)$. The errors are illustrated for the fuzzy exponent varying in the range $n = 1.3$ to $n = 3.0$ (rats) and $n = 2.0$ to $n = 5.5$ (humans). The **top** plots show all the error curves, while the **bottom** plots show magnification of short segment sizes' errors. One can notice that the fuzzy exponent that gives the best accuracy (lower error) varies according to the segment size, although the optimal exponents are different from those obtained with the original fuzzy function. The lowest error for all segment sizes is illustrated in the plots at the magnification plots' right (MFE minimum). For reference, MMSE is also illustrated in the plots (gray lines).

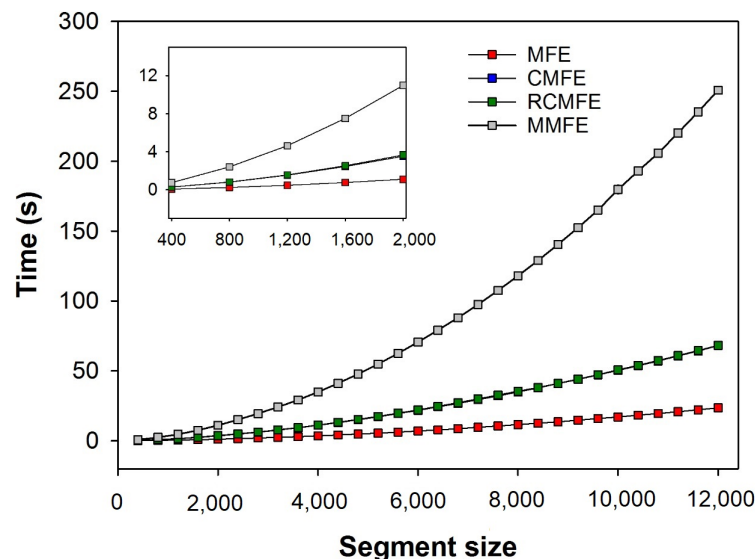


Figure 8. Computation time of the fuzzy entropy-based variants of MSE, i.e., MFE, CMFE, RCMFE, and MMFE. The plot shows the average time consumed (in seconds) to calculate the variants of MSE for 3 HRV series of humans up to segments of 12,000 points. Of note, CMFE and RCMFE take virtually the same time to be computed, and therefore, these curves are superimposed.

5. Discussion

In the present study, we adopted a systematic comparison between three variants of MSE (CMSE, RCMSE, MMSE) and their fuzzy-based adaptations (CMFE, RCMFE, MMFE, MFE) to check the accuracy of them to estimate the real MSE for short-term signals.

As expected, all fuzzy-based methods performed superior compared to the algorithms based on sample entropy. Surprisingly, all fuzzy-based methods' accuracy is quite similar, pointing that the use of fuzzy entropy in place of sample entropy seems sufficient to provide optimal estimations of MSE for short-term signals. In other words, the improvements adopted in the coarse-graining for CMFE, RCMFE, and MMFE seem to have little or no effect in fuzzy-based variants, since MFE provided errors in the estimate of original MSE as low as the ones found in CMFE, RCMFE, and MMFE. The replacement of the rigid similarity of SampEn (Heaviside function) by the smooth fuzzy function in FuzzyEn seems to be the most relevant improvement for a good estimation of entropy in short time series. However, one must be aware that optimal fuzzy exponents' choice is crucial to obtain high accuracy.

In a previous study with the same dataset, we showed that MMFE provides better estimates of the original MSE than MMSE when proper choices of the fuzzy exponent n are made [19]. Here, we showed that CMFE, RCMFE, and MFE also have a dependence on n and that the optimal exponents found for MMFE are virtually the same for CMFE, RCMFE, and MFE, as can be seen in Figure 6. Although both the rats and humans datasets represent health conditions, the best exponents for rats increase with series length, while it decreases for humans. This is likely to be a consequence of the different species, but it still has to be investigated together with datasets with pathological HRV series. Nevertheless, the fitting equations provided in Figure 6 can be used for the choice of the optimal fuzzy exponents in the dataset evaluated here. The measure stationarity is a possible issue concerning long-term time-series as previously pointed out [27]. However, all time-series and analyzed segments are supposed to be taken at baseline physiological state. We introduce an illustrative analysis of two time-series data, i.e., one human and one rat, in the appendix to this paper presented in Figure A1.

To check the influence of the fuzzy function on the accuracy of fuzzy-based MSE variants, we calculated the mean squared error of MFE using an alternative fuzzy function, i.e., $\exp(-0.6931 \times (d/r)^n)$ [22]. Interestingly, this alternative function also presented a dependence of the minimum error on the choice of the exponent n . However, the range of optimal n values is markedly different from the ones found for the original fuzzy function, and curiously, the exponent always increases with the series length for both datasets (rats and humans). The extensive evaluation of different fuzzy functions is out of the present study's scope, and one must be aware that changing the fuzzy function requires the search for the optimal fuzzy exponents. For an extensive review on the possible fuzzy functions and their differences, please refer to [28].

For entropy estimators based on a similarity function between patterns (such as SampEn and FuzzyEn), the tolerance factor, r , is commonly defined as a percentage of the signal's SD, making the results comparable within signals with different magnitude. Alternatively, the signal can be normalized to mean zero and SD one, a procedure that has the same effect of multiplying the tolerance factor by the signal SD. However, in FuzzyEn, these two procedures are not always equivalent and depend on the fuzzy function adopted. In the case of $\exp(d^n/r)$ (the fuzzy function adopted in this study), one can notice that the distance between patterns (d) and r are not raised to the same power (except when $n = 1$). Thus, normalizing the series (affecting d) or normalizing r may result in different entropy values. The alternative fuzzy function evaluated with MFE ($\exp(-0.6931 \times (d/r)^n)$) does not show this limitation, since d and r are both raised to n .

The computation cost (time) analysis necessary to calculate all the fuzzy-based variants of MSE showed that MMFE is the most time-consuming. On the other hand, MFE is the fastest algorithm among them. Since all algorithms' accuracy is very similar, the simplicity and computational efficiency of MFE make this algorithm the most attractive to be used for the analysis of short-term signals, which provides good accuracy for HRV series as short as 400 points.

6. Conclusions

In this study, several SampEn- and FuzzyEn-based MSE variants were calculated in short HRV time series, and their accuracy to estimate the actual MSE was investigated. All fuzzy entropy-based algorithms provided better accuracy (given the fuzzy exponent's proper choice) compared to the variants based on SampEn. Moreover, all FuzzyEn-based algorithms evaluated showed similar accuracy. Therefore, since MFE is the most simple and cost-effective algorithm among them, we recommend the use of MFE for the analysis of short-term time series. The results also indicate that different fuzzy functions may provide good accuracies. However, the dependence of the fuzzy exponent (n) to the series length may vary from one function to another and on different datasets. Further studies are necessary to determine the optimal fuzzy exponents in datasets with pathological signals. A possible limitation of the study is the fact of assuming the entropy computed for the whole series compared to shorter segments of itself, although the comparison can reflect accuracy and precision.

Author Contributions: Conceptualization, A.M.S.B.J., A.H.-H., L.E.V.S. and L.O.M.J.; methodology, A.M.S.B.J., A.H.-H., L.E.V.S. and L.O.M.J.; validation, A.M.S.B.J., L.E.V.S. and L.O.M.J.; formal analysis, A.M.S.B.J., L.E.V.S. and L.O.M.J.; investigation, A.M.S.B.J.; data curation, L.E.V.S.; writing—original draft preparation, A.M.S.B.J., A.H.-H., L.E.V.S. and L.O.M.J.; writing—review and editing, A.M.S.B.J., A.H.-H., L.E.V.S. and L.O.M.J.; supervision, L.E.V.S. and L.O.M.J.; project administration, A.M.S.B.J., L.E.V.S. and L.O.M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The rat ECG dataset was approved by the Animal Research Ethics Committee of the Faculty of Medicine of Ribeirão Preto, University of São Paulo, SP, Brazil (Protocol n°016/2013-1). The human ECG dataset under the auspices of the National Center for Research Resources, National Institutes of Health. Key participating centers include Beth Israel Deaconess Medical Center/Harvard Medical School, Boston University's Center for Polymer Studies, Division of Health Sciences and Technology, Harvard University-Massachusetts Institute of Technology, and McGill University's Center for Nonlinear Dynamics in Physiology and Medicine.

Informed Consent Statement: Information on all subjects involved in the study is available at <https://doi.org/10.13026/C2NK5R>.

Data Availability Statement: The rat ECG dataset can be requested from the authors. The human ECG dataset is available at <https://physionet.org/content/nsr2db/1.0.0/> (accessed on 26 November 2021).

Acknowledgments: We thank Rubens Fazan Junior, Helio C. Salgado, and Carlos A. A. Silva for kindly providing part of the data used in this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

SampEn	Sample Entropy
FuzzyEn	Fuzzy Entropy
MSE	Multiscale Entropy
MFE	Multiscale Fuzzy Entropy
CMSE	Composite Multiscale Entropy
CMFE	Composity Multiscale Fuzzy Entropy
RCMSE	Refined Composite Multiscale Entropy
RCMFE	Refined Composite Multiscale Fuzzy Entropy
MMSE	Modified Multiscale Entropy
MMFE	Modified Multiscale Fuzzy Entropy

Appendix A. Fuzzy Entropy Stationarity

In this section, we illustrate the dynamic behavior of calculated fuzzy entropy for short segments over the time for both rat and human examples. One can observe the higher stationarity of fuzzy entropy compared to ample entropy in Figure A1. The chosen subject and animal are typical and representative of the whole datasets for both human and rats, respectively.

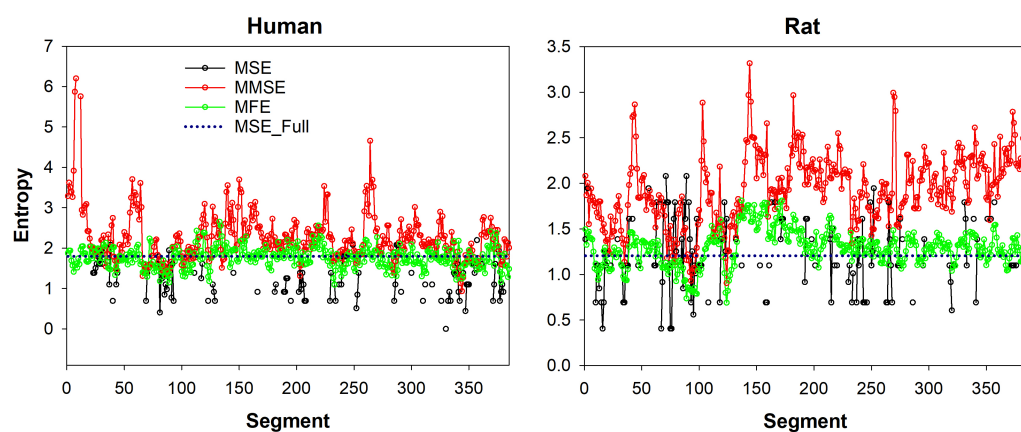


Figure A1. The image on the left shows the calculated entropy value for a human patient on time scale 18 and 400 point windows over time. The image on the right shows the calculated entropy value for a mouse on the 13 time scale and 400 point windows over time.

References

- Boccara, N. Cellular Automata. In *Modeling Complex Systems*; Springer: New York, NY, USA, 2010; pp. 191–273. [\[CrossRef\]](#)
- Klamut, J.; Kutner, R.; Struzik, Z.R. Towards a Universal Measure of Complexity. *arXiv* **2020**, arXiv:2006.01900.
- Delgado-Bonal, A.; Marshak, A. Approximate Entropy and Sample Entropy: A Comprehensive Tutorial. *Entropy* **2019**, *21*, 541. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, W.; Wang, Z.; Xie, H.; Yu, W. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 266–272. [\[CrossRef\]](#) [\[PubMed\]](#)
- Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [\[CrossRef\]](#)
- Wu, S.D.; Wu, C.W.; Lee, K.Y.; Lin, S.G. Modified multiscale entropy for short-term time series analysis. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 5865–5873. [\[CrossRef\]](#)
- Costa, M.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Phys. A Stat. Mech. Appl.* **2003**, *330*, 53–60. [\[CrossRef\]](#)
- Chang, Y.C.; Wu, H.T.; Chen, H.R.; Liu, A.B.; Yeh, J.J.; Lo, M.T.; Tsao, J.H.; Tang, C.J.; Tsai, I.; Sun, C.K.; et al. Application of a modified entropy computational method in assessing the complexity of pulse wave velocity signals in healthy and diabetic subjects. *Entropy* **2014**, *16*, 4032–4043. [\[CrossRef\]](#)
- Wu, S.D.; Wu, C.W.; Lin, S.G.; Wang, C.C.; Lee, K.Y. Time series analysis using composite multiscale entropy. *Entropy* **2013**, *15*, 1069–1084. [\[CrossRef\]](#)
- Wu, S.D.; Wu, C.W.; Lin, S.G.; Lee, K.Y.; Peng, C.K. Analysis of complex time series using refined composite multiscale entropy. *Phys. Lett. A* **2014**, *378*, 1369–1374. [\[CrossRef\]](#)
- Wang, F.; Lu, B.; Kang, X.; Fu, R. Research on driving fatigue alleviation using interesting auditory stimulation based on VMD-MMSE. *Entropy* **2021**, *23*, 1209. [\[CrossRef\]](#)
- Zheng, J.D.; Chen, M.; Cheng, J.S.; Yang, Y. Multiscale fuzzy entropy and its application in rolling bearing fault diagnosis. *J. Vib. Eng.* **2014**, *27*, 145–151.
- Li, Y.; Xu, M.; Wang, R.; Huang, W. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **2016**, *360*, 277–299. [\[CrossRef\]](#)
- Zheng, J.; Pan, H.; Cheng, J. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759. [\[CrossRef\]](#)
- Azami, H.; Fernández, A.; Escudero, J. Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis. *Med. Biol. Eng. Comput.* **2017**, *55*, 2037–2052. [\[CrossRef\]](#)

17. Zheng, J.; Pan, H.; Tong, J.; Liu, Q. Generalized refined composite multiscale fuzzy entropy and multi-cluster feature selection based intelligent fault diagnosis of rolling bearing. *ISA Trans.* **2021**. [[CrossRef](#)] [[PubMed](#)]
18. Tomčala, J. New Fast ApEn and SampEn Entropy Algorithms Implementation and Their Application to Supercomputer Power Consumption. *Entropy* **2020**, *22*, 863. [[CrossRef](#)]
19. Borin, A.M.S., Jr.; Silva, L.E.V.; Murta, L.O., Jr. Modified multiscale fuzzy entropy: A robust method for short-term physiologic signals. *Chaos Interdiscip. J. Nonlinear Sci.* **2020**, *30*, 083135. [[CrossRef](#)] [[PubMed](#)]
20. Govindan, R.; Wilson, J.; Eswaran, H.; Lowery, C.; Preißl, H. Revisiting sample entropy analysis. *Phys. A Stat. Mech. Appl.* **2007**, *376*, 158–164. [[CrossRef](#)]
21. Zadeh, L.A. Information and control. *Fuzzy Sets* **1965**, *8*, 338–353.
22. Mayer, C.; Bachler, M.; Holzinger, A.; Stein, P.K.; Wassertheurer, S. The effect of threshold values and weighting factors on the association between entropy measures and mortality after myocardial infarction in the cardiac arrhythmia suppression trial (CAST). *Entropy* **2016**, *18*, 129. [[CrossRef](#)]
23. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906. [[CrossRef](#)]
24. Azami, H.; Escudero, J. Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings. *Biomed. Signal Process Control* **2016**, *23*, 28–41. [[CrossRef](#)]
25. Silva, L.E.V.; Lataro, R.M.; Castania, J.A.; da Silva, C.A.A.; Valencia, J.F.; Murta, L.O., Jr.; Salgado, H.C.; Fazan, R., Jr.; Porta, A. Multiscale entropy analysis of heart rate variability in heart failure, hypertensive, and sinoaortic-denervated rats: Classical and refined approaches. *Am. J.-Physiol.-Regul. Integr. Comp. Physiol.* **2016**, *311*, R150–R156. [[CrossRef](#)]
26. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)] [[PubMed](#)]
27. Gonçalves, H.; Henriques-Coelho, T.; Rocha, A.P.; Lourenço, A.P.; Leite-Moreira, A.; Bernardes, J. Comparison of different methods of heart rate entropy analysis during acute anoxia superimposed on a chronic rat model of pulmonary hypertension. *Med. Eng. Phys.* **2013**, *35*, 559–568. [[CrossRef](#)]
28. Azami, H.; Li, P.; Arnold, S.E.; Escudero, J.; Humeau-Heurtier, A. Fuzzy entropy metrics for the analysis of biomedical signals: Assessment and comparison. *IEEE Access* **2019**, *7*, 104833–104847. [[CrossRef](#)]