

Multiscale Entropy Analysis of Short Signals: The Robustness of Fuzzy Entropy-Based Variants

Airton Monte Serrat Borin Jr. (✉ airton@iftm.edu.br)

Instituto Federal de Educacao Ciencia e Tecnologia do Triangulo Mineiro <https://orcid.org/0000-0001-9303-8291>

Anne Humeau-Heurtier

Université Angers: Universite d'Angers

Luiz Otavio Murta Jr.

USP Campus de Ribeirao Preto: Universidade de Sao Paulo Campus de Ribeirao Preto

Luiz Eduardo Virgilio Silva

USP Campus de Ribeirao Preto: Universidade de Sao Paulo Campus de Ribeirao Preto

Research Article

Keywords: Signals analysis, Fuzzy logic, Information entropy

Posted Date: April 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-361154/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multiscale Entropy Analysis of Short Signals: the Robustness of Fuzzy Entropy-Based Variants

Airton Monte Serrat Borin Jr · Anne Humeau-Heurtier · Luiz Otavio Murta Jr · Luiz Eduardo Virgilio Silva

Received: date / Accepted: date

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. 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 ac-

Airton Monte Serrat Borin Jr.

Federal Institute for Education Science and Technology of Triângulo Mineiro, IFTM, Uberaba, MG, 38064-790, Brazil
Tel.: +55-34-3319-6000
Fax: +55-34-3319-6001
E-mail: airtonr@iftm.edu.br

Anne Humeau-Heurtier

Univ Angers, LARIS - Laboratoire Angevin de Recherche en Ingénierie des Systèmes, 62 avenue Notre-Dame du Lac, Angers 49000, France

Luiz Otavio Murta Jr.

Department of Computing and Mathematics, Ribeirão Preto School of Philosophy, Science and Literature, University of São Paulo, Ribeirão Preto, SP 14049-900 Brazil

Luiz Eduardo Virgilio Silva

Department of Internal Medicine, Ribeirão Preto Medical School, University of São Paulo, SP 14049-900, Brazil

curacy, the findings suggest different fuzzy exponents according to the time series length.

Keywords Signals analysis · Fuzzy logic · Information entropy

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 [4]. 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 [13].

Most entropy estimators for time series are inspired by the Kolmogorov-Sinai (KS) entropy, e.g., sample entropy [16] and fuzzy entropy [7]. They intend to estimate the rate at which the information grows over time in the system. However, the introduction of multiscale entropy (MSE) [8] 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 [10, 18], significantly losing its sensitivity [6]. To overcome these limitations, some approaches propose different coarse-graining procedures and entropy estimation. Composite MSE (CMSE) [20], refined composite MSE (RCMSE) [19], and modified MSE (MMSE) [18] 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 [24, 14, 22].

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 [23] [2]. In a recent study, we proposed and evaluated the modified multiscale fuzzy entropy (MMFE) for heart rate variability (HRV) analysis [5]. 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. Each method's 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. Your text comes here. Separate text sections with

2 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 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 [16, 12] and the fuzzy entropy (FuzzyEn) algorithm [7] 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

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

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

and

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

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

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 :

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

2.1.2 FuzzyEn

The 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:

$$\mathbf{x}_m(i) = \{u(i + k\delta) - u0(i) : 0 \leq k \leq m - 1\} \quad (7)$$

where

$$u0(i) = \frac{1}{m} \sum_{j=0}^{m-1} u(i + j\delta). \quad (8)$$

To calculate the similarity between two vectors, two functions were tested in our work: $\exp(-d_m^n/r)$ and $\exp(-0.6931(d/r)^n)$ [15]. For the first function, we computed

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

and

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

where d is given in Eq. 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

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

and

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

similar to Eq. 3 and Eq. 5, so the FuzzyEn for the parameters m , n , δ and r is calculated by

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

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

In the MSE [8, 10, 9] and MFE [14, 24, 22] algorithms, the dynamics of a system at different time scales is obtained by a moving average procedure (*coarse-graining* procedure), according to

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

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

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

and

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

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

In CMSE [20] and CMFE [23], 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

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

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

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

and

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

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

The procedure to obtain the *coarse-grained* series in RCMSE [19] and RCMFE [1] is the same as for CMSE and CMFE (see Eq. 17). However, instead of averaging the entropy of each k scaled series for scale τ , entropy is estimated from the average number of matches $\bar{n}_{k,\tau}^{m+1}$ and $\bar{n}_{k,\tau}^m$, obtained from all k *coarse-grained* series.

2.4.1 RCMSE

RCMSE is defined as:

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

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, the RCMFE is defined by

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

where ϕ^m and ϕ^{m+1} are given by Eq. 11 and Eq. 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 (Eq. 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 [20] proposes that the *coarse-grained* time series be constructed by Eq. 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 [5] and consists in the application of the same *coarse-graining* procedure as MMSE (Eq. 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 [5, 17]. 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 minutes) at baseline conditions. Computer software (LabChart, ADInstruments, Australia) was used to create RR series from ECG recordings. All RR series were visually inspected for artifacts and corrected when necessary. Since the time series's length varied from 15892 to 32333 points, all RR series was truncated to 15892 points. The second group of HRV series consists of 12 healthy human individuals, obtained from the Physionet MIT-BIH Normal Sinus Rhythm database [11]. 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, intervals between two successive normal beats. Eventually, all RR series was truncated to 15892 samples so that the series of rats and humans have the same length. The recording period of all series runs from 8 am to 10 pm.

3.2 Experiments

We segmented the HRV full series (15892 points) into equal segments of 400, 800, 1200, ..., 15600 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 twenty, i.e. $\tau = 1, 2, \dots, 20$. The embedding dimension and tolerance factor of entropy estimators were set as $m = 2$ and $r = 0.15 \times \text{SD}$ 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.

Moreover, to assess the cost-effectiveness of fuzzy-based MSE variants, we measured each algorithm's average computation time. The analysis was performed in 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.) 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 12000 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) [5]:

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

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 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 Fig. 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.

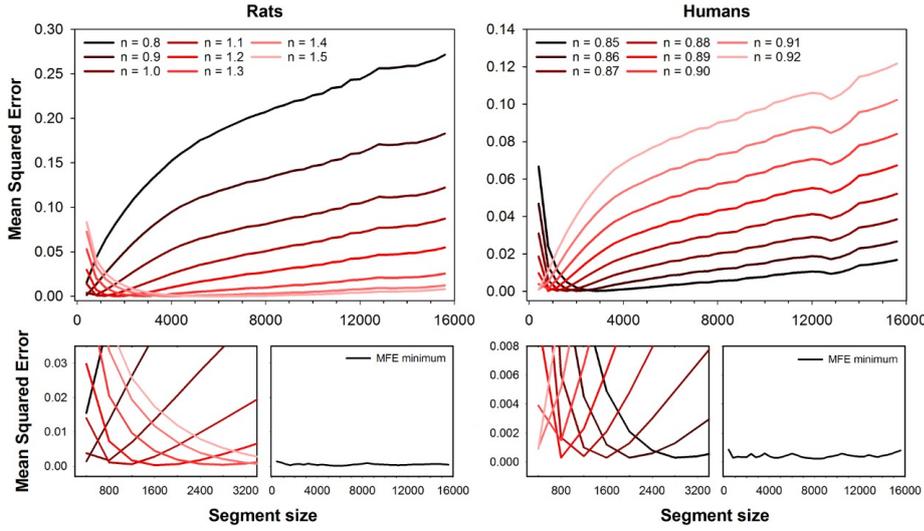


Fig. 1 Mean squared errors of MFE as a function of the segment size. The errors were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats) and $n = 0.85$ to $n = 0.92$ (humans). Top plots show all the error curves, while bottom plots show magnification of short segment sizes' 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 lines) 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 Fig. 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 lines) 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 Fig. 4. Although a similar error plot can be found in a previous study [5], 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

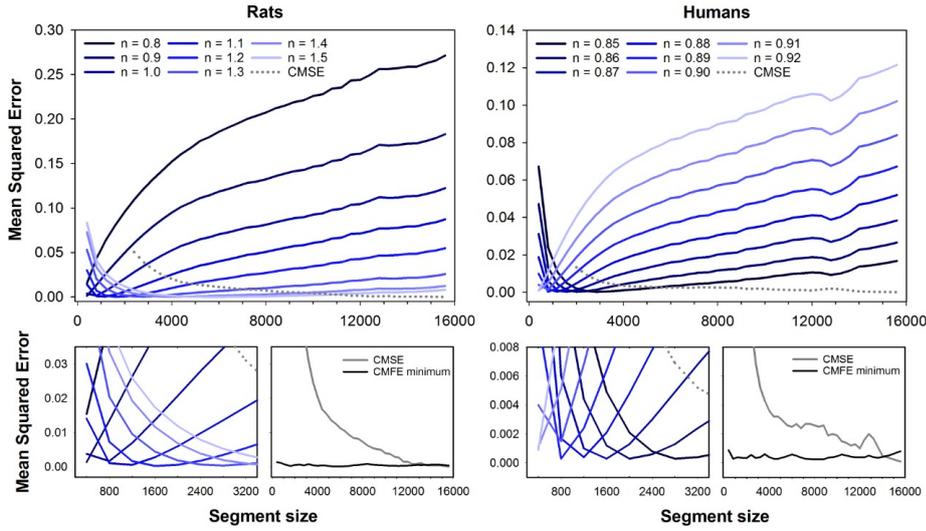


Fig. 2 Mean squared errors of CMFE and CMSE as a function of the segment size. The errors of CMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). Top plots show the full error curves of CMFE (solid lines) and CMSE (dashed 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 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.

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 5 shows the comparison of errors from all the multiscale variants studied, i.e., CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE, and MFE. For fuzzy entropy-based methods, only the minimum error is shown, calculated with each segment size's optimal fuzzy exponent. As shown in the figure, all fuzzy entropy-based variants have lower errors than any sample entropy-based variant, for segments size up to 13000 points.

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 Fig. 6). 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. Notice 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, Fig. 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 showed that the MFE with this alternative fuzzy function also presents dependence on the segment size,

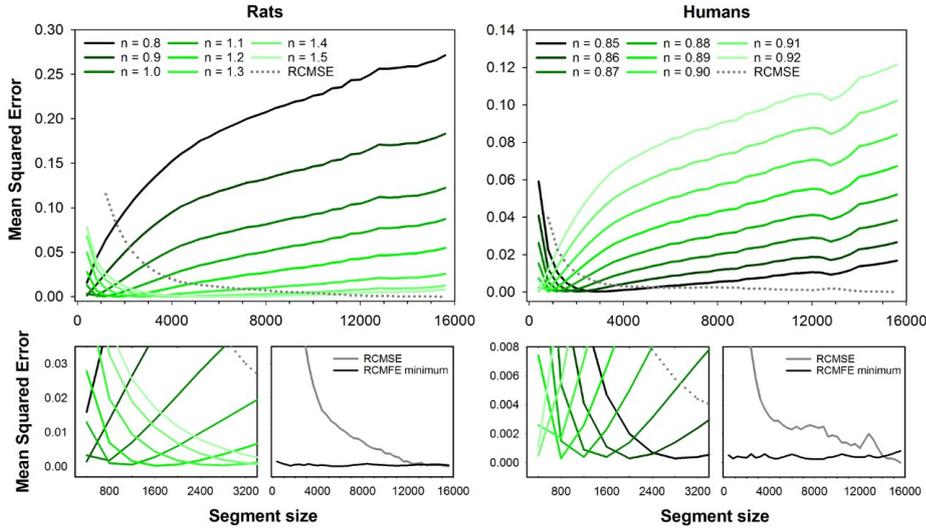


Fig. 3 Mean squared errors of RCMFE and RCMSE as a function of the segment size. The errors of RCMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). Top plots show the full error curves of RCMFE (solid lines) and RCMSE (dashed lines), 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.

similar to the original fuzzy function (Fig. 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 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 12000 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 showed virtually the same computational time (curves are superimposed in Fig. 8).

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

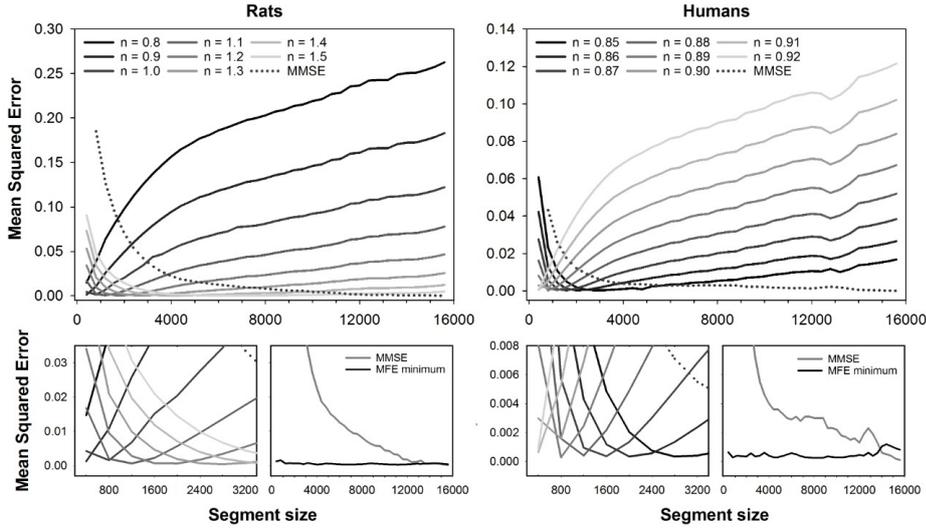


Fig. 4 Mean squared errors of MMFE and MMSE as a function of the segment size. The errors of MMFE were obtained for fuzzy exponents varying in the range $n = 0.8$ to $n = 1.5$ (rats, at left) and $n = 0.85$ to $n = 0.92$ (humans, at right). Top plots show the full error curves of MMFE (solid lines) and 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.

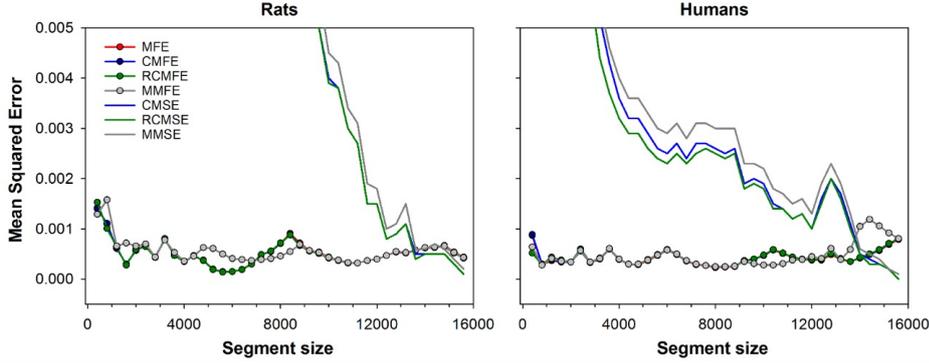


Fig. 5 Mean squared error for all considered approaches, i.e. CMSE, RCMSE, MMSE, CMFE, RCMFE, MMFE and MFE. For fuzzy entropy-based approaches, only the minimum error is shown, obtained with the optimal fuzzy exponent for each segment size. Results are shown for both rats (left) and human (right) database. Notice that the errors are calculated regarding the MSE of full-length time-series, i.e. 15892 beats.

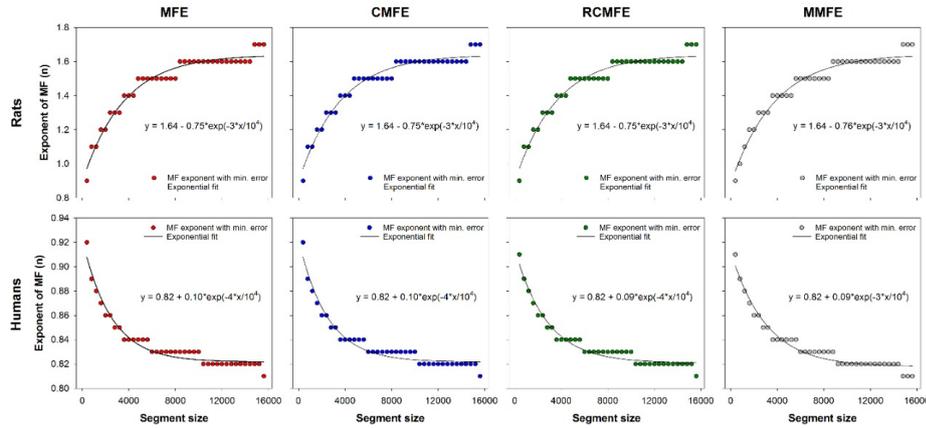


Fig. 6 Best fuzzy exponents found for each segment size in both rats and humans HRV datasets. The solid lines represent the best fitting functions, i.e. a decreasing exponential function $y = 1.64 - 0.75 \times \exp(-3 \times x/10^4)$ for rats and an increasing exponential $y = 0.82 - 0.10 \times \exp(-4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

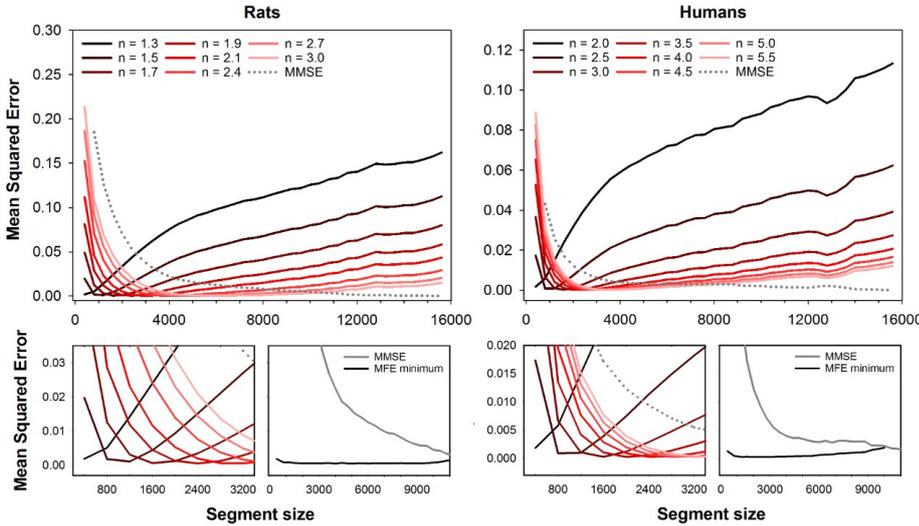


Fig. 7 Mean squared errors of MFE as a function of the segment size, obtained with alternative fuzzy function, i.e., $\exp(-0.6931 \times (d/r)^n)$. The errors are illustrated for the fuzzy exponent varying in the range $n = 1.3$ to $n = 3.0$ (rats) and $n = 2.0$ to $n = 5.5$ (humans). Top plots show 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, 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).

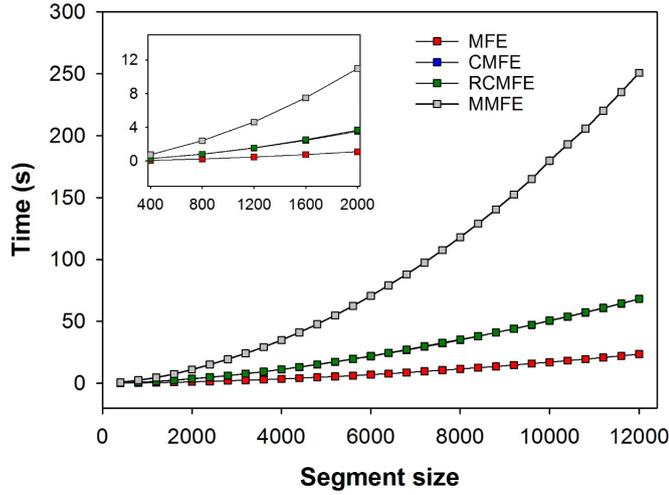


Fig. 8 Computation time of the fuzzy entropy-based variants of MSE, i.e., MFE, CMFE, RCMFE, and MMFE. The plot shows the average time consumed (in seconds) to calculate the variants of MSE for three HRV series of humans up to segments of 12000 points. Of note, CMFE and RCMFE take virtually the same time to be computed, and therefore, these curves are superimposed.

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 lower 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 [5]. 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 Fig. 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 Fig. 6 can be used for the choice of the optimal fuzzy exponents in the dataset evaluated here.

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., $\exp(-0.6931 \times (d/r)^n)$ [15]. 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 [3].

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 shorter as 400 points.

6 Conclusion

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 differ-

ent datasets. Further studies are necessary to determine the optimal fuzzy exponents in datasets with pathological signals.

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

Declarations

Funding

Not applicable.

Conflict of interest

The authors declare that they have no conflict of interest.

Availability of data and material

Data from rats can be obtained upon request to Dr. Rubens Fazan Jr (rfazan@usp.br). Data from humans can be obtained from Physionet (www.physionet.org).

Code availability

The codes can be provided by the authors upon request.

Authors' contributions

Not applicable.

Ethics approval

Experiments with rats were approved by the Committee of Ethics in Animal Research of the School of Medicine of Ribeirão Preto, University of São Paulo, SP, Brazil (Protocol No. 016/2013-1).

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Azami, H., Escudero, J.: Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings. *Biomedical Signal Processing and Control* **23**, 28–41 (2016)
2. Azami, H., Fernández, A., Escudero, J.: Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis. *Medical & biological engineering & computing* **55**(11), 2037–2052 (2017)
3. 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* **7**, 104833–104847 (2019)
4. Boccara, N.: Cellular automata. *Modeling Complex Systems* pp. 191–273 (2004)
5. Borin Jr, A.M.S., Silva, L.E.V., Murta Jr, L.O.: Modified multiscale fuzzy entropy: A robust method for short-term physiologic signals. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30**(8), 083135 (2020)
6. 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* **16**(7), 4032–4043 (2014)
7. Chen, W., Wang, Z., Xie, H., Yu, W.: Characterization of surface emg signal based on fuzzy entropy. *IEEE Transactions on neural systems and rehabilitation engineering* **15**(2), 266–272 (2007)
8. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of complex physiologic time series. *Physical review letters* **89**(6), 068102 (2002)
9. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of biological signals. *Physical review E* **71**(2), 021906 (2005)
10. Costa, M., Peng, C.K., Goldberger, A.L., Hausdorff, J.M.: Multiscale entropy analysis of human gait dynamics. *Physica A: Statistical Mechanics and its applications* **330**(1-2), 53–60 (2003)
11. 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* **101**(23), e215–e220 (2000). DOI 10.1161/01.CIR.101.23.e215
12. Govindan, R., Wilson, J., Eswaran, H., Lowery, C., Preißl, H.: Revisiting sample entropy analysis. *Physica A: Statistical Mechanics and its Applications* **376**, 158–164 (2007)
13. Klamut, J., Kutner, R., Struzik, Z.R.: Towards a universal measure of complexity. arXiv preprint arXiv:2006.01900 (2020)
14. 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. *Journal of Sound and Vibration* **360**, 277–299 (2016). DOI 10.1016/j.jsv.2015.09.016. URL <https://doi.org/10.1016/j.jsv.2015.09.016>
15. 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* **18**(4), 129 (2016)
16. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* **278**(6), H2039–H2049 (2000)
17. Silva, L.E.V., Lataro, R.M., Castania, J.A., da Silva, C.A.A., Valencia, J.F., Murta Jr, L.O., Salgado, H.C., Fazan Jr, R., Porta, A.: Multiscale entropy analysis of heart rate variability in heart failure, hypertensive, and sinoaortic-denervated rats: classical and refined approaches. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **311**(1), R150–R156 (2016)

18. Wu, S.D., Wu, C.W., Lee, K.Y., Lin, S.G.: Modified multiscale entropy for short-term time series analysis. *Physica A: Statistical Mechanics and its Applications* **392**(23), 5865–5873 (2013)
19. 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. *Physics Letters A* **378**(20), 1369–1374 (2014)
20. Wu, S.D., Wu, C.W., Lin, S.G., Wang, C.C., Lee, K.Y.: Time series analysis using composite multiscale entropy. *Entropy* **15**(3), 1069–1084 (2013)
21. Zadeh, L.A.: Information and control. *Fuzzy sets* **8**(3), 338–353 (1965)
22. Zheng, J., Pan, H., Cheng, J.: Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mechanical Systems and Signal Processing* **85**, 746–759 (2017). DOI 10.1016/j.ymssp.2016.09.010. URL <https://doi.org/10.1016/j.ymssp.2016.09.010>
23. Zheng, J., Pan, H., Cheng, J.: Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mechanical Systems and Signal Processing* **85**, 746–759 (2017)
24. Zheng, J.D., Chen, M., Cheng, J.S., Yang, Y.: Multiscale fuzzy entropy and its application in rolling bearing fault diagnosis. *J. Vib. Eng* **27**, 145–151 (2014)

Figures

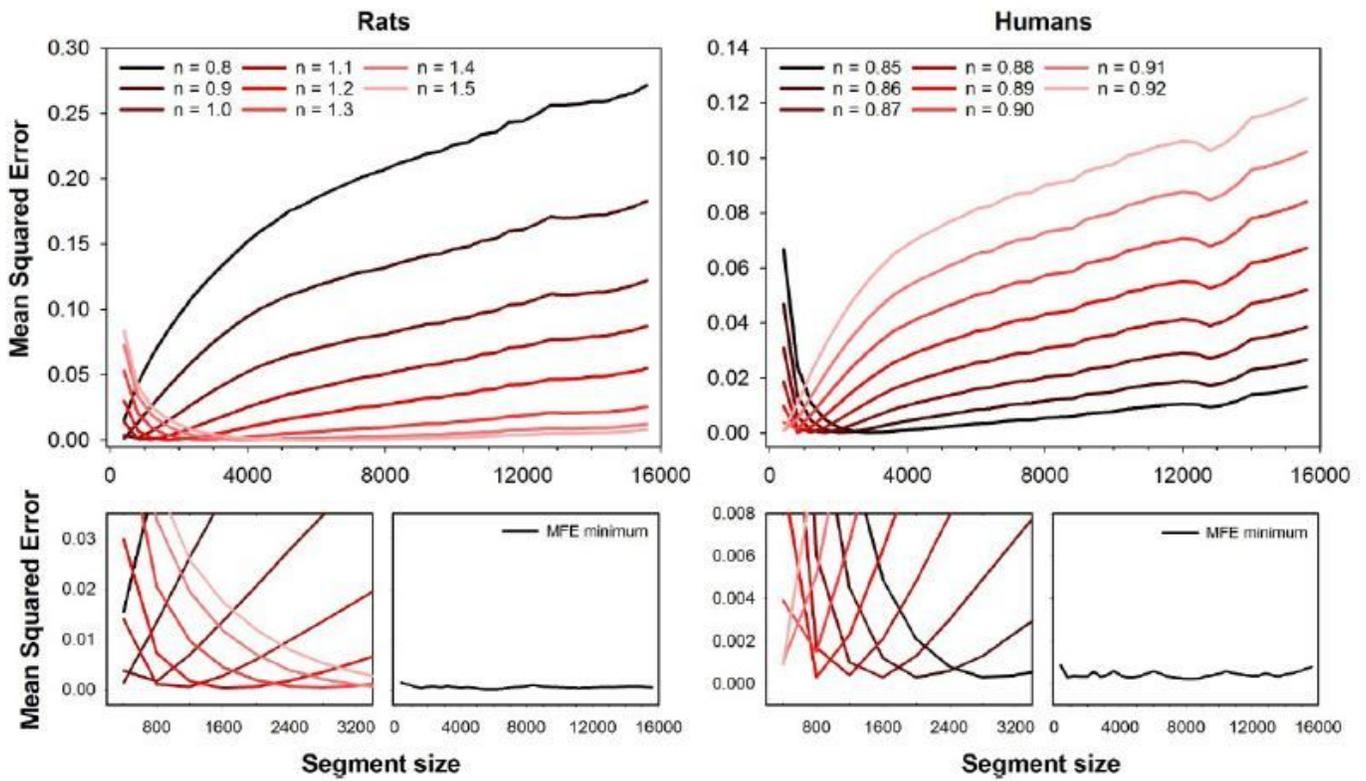


Figure 1

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).

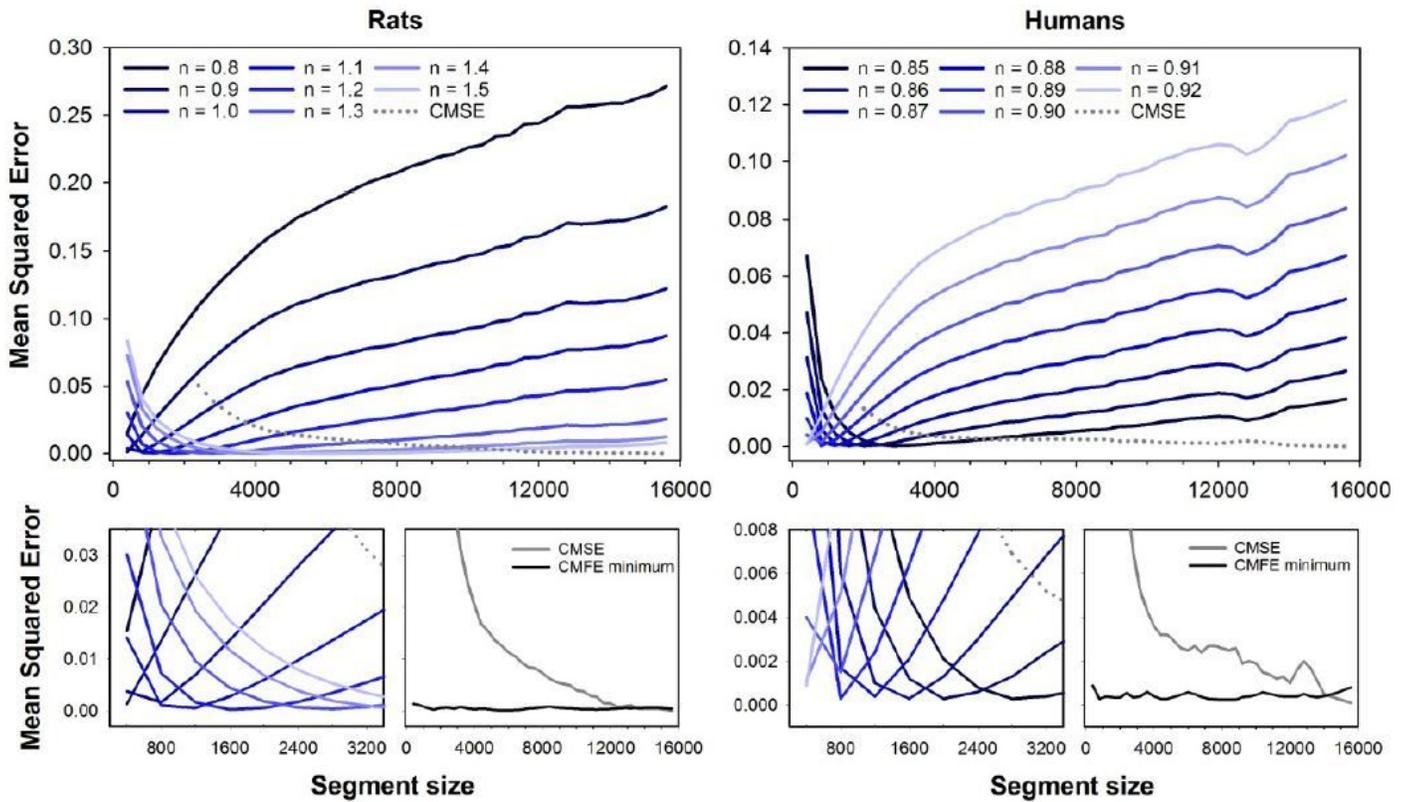


Figure 2

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 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 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.

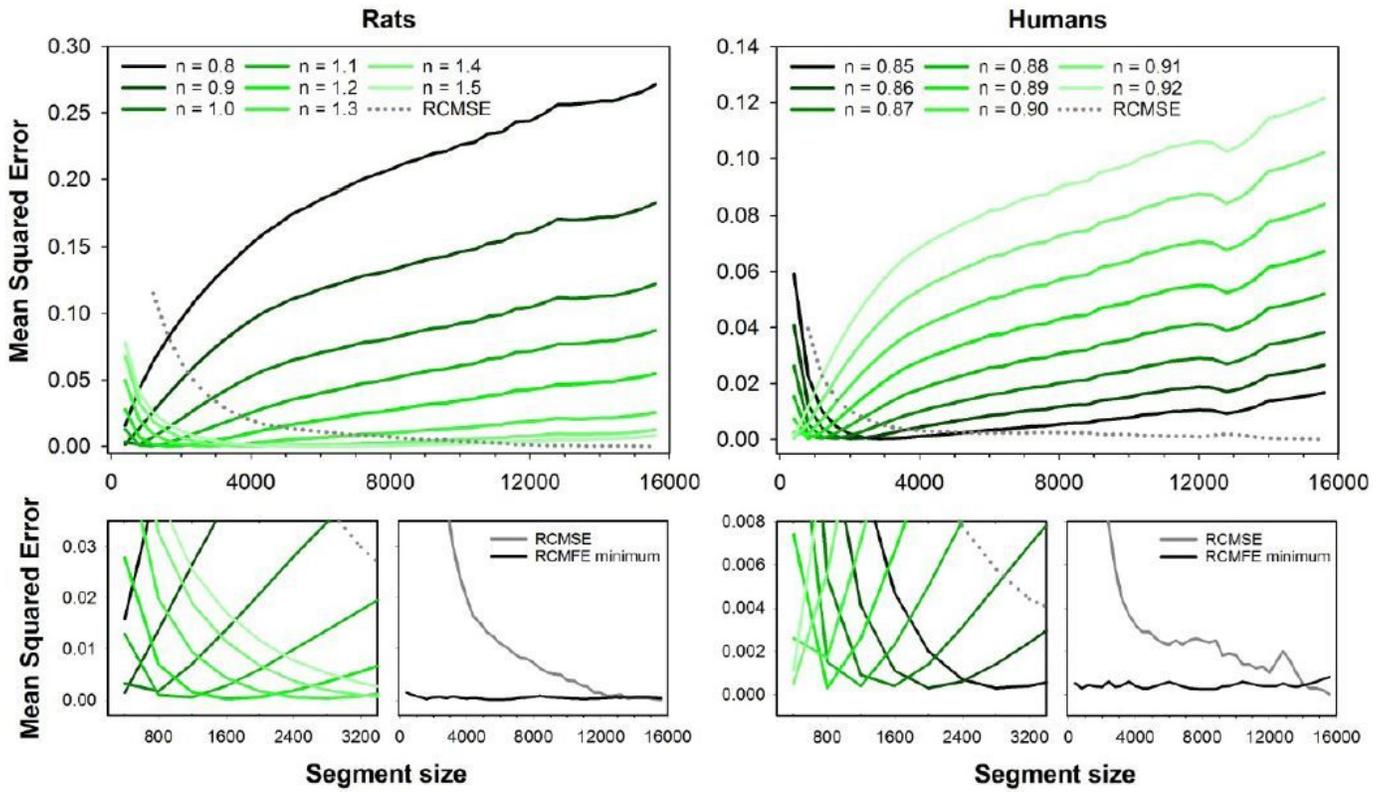


Figure 3

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 lines), 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.

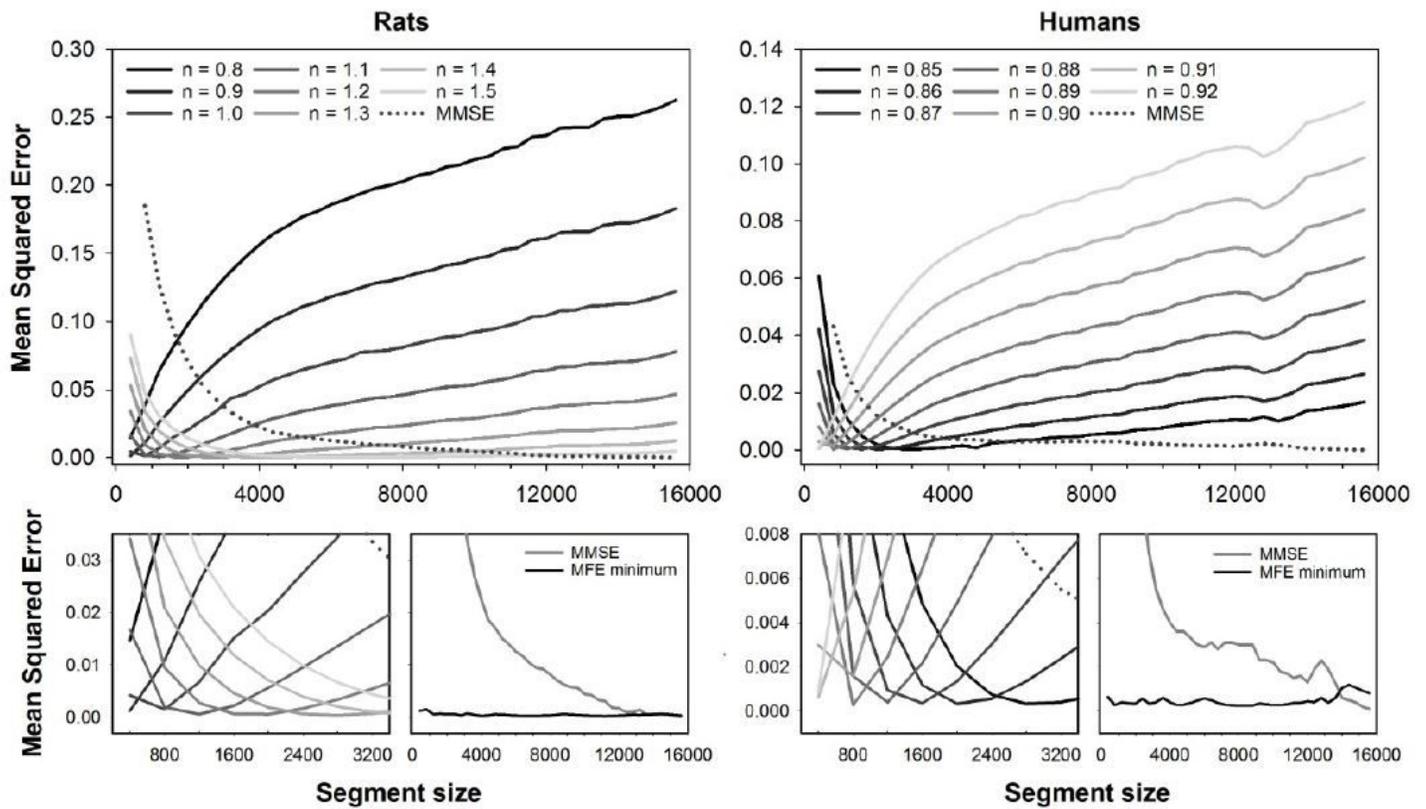


Figure 4

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.

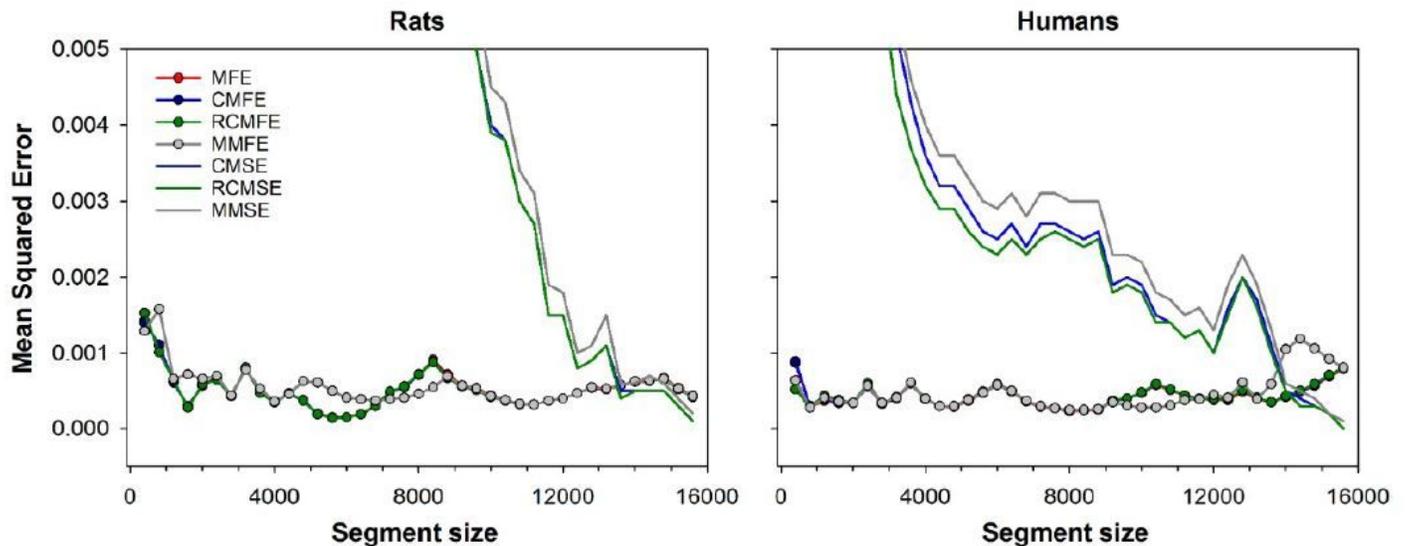


Figure 5

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. 15892 beats.

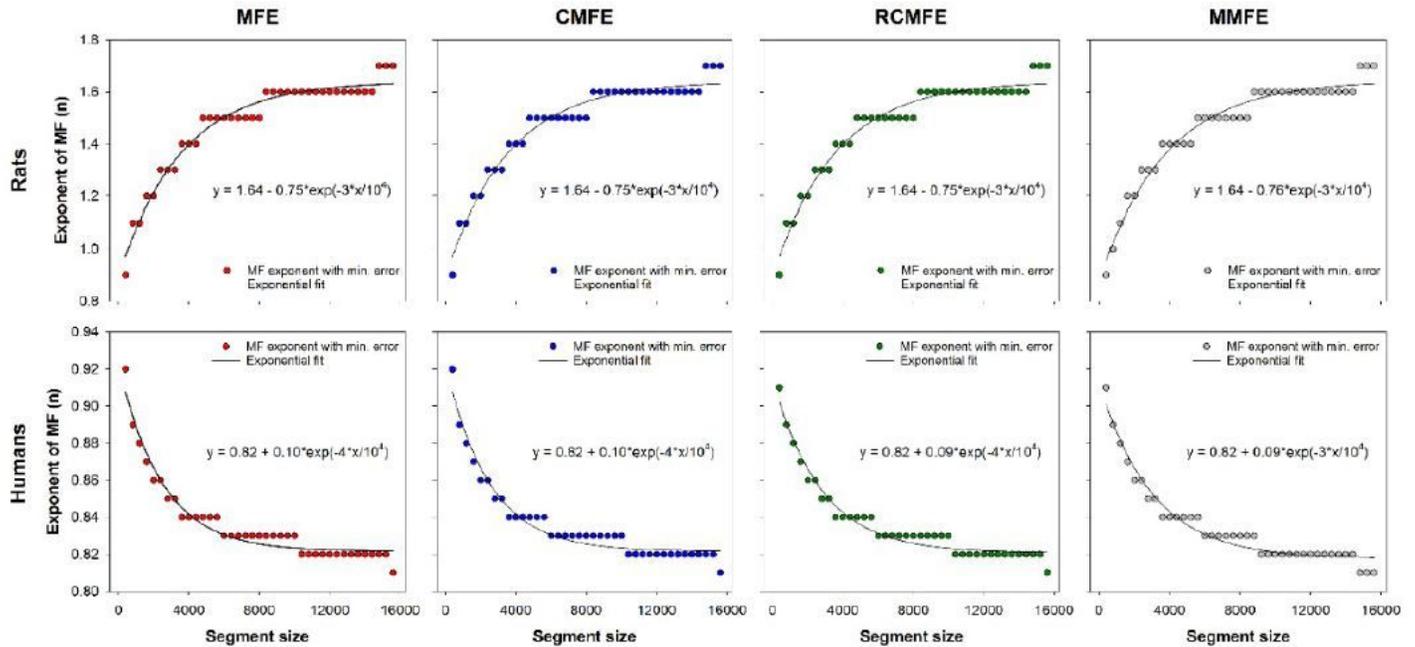


Figure 6

"Please see the Manuscript PDF file for the complete figure caption".

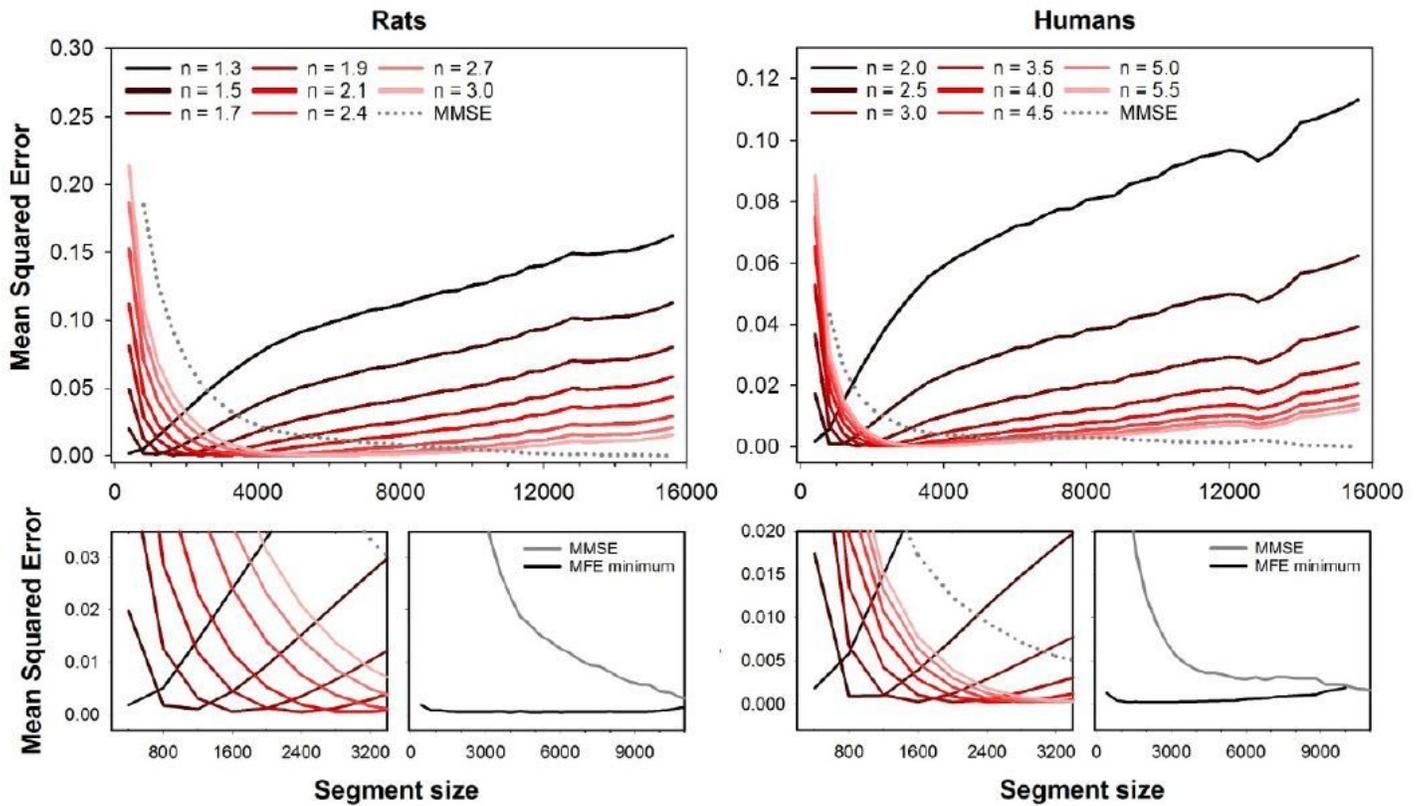


Figure 7

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). 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, 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).

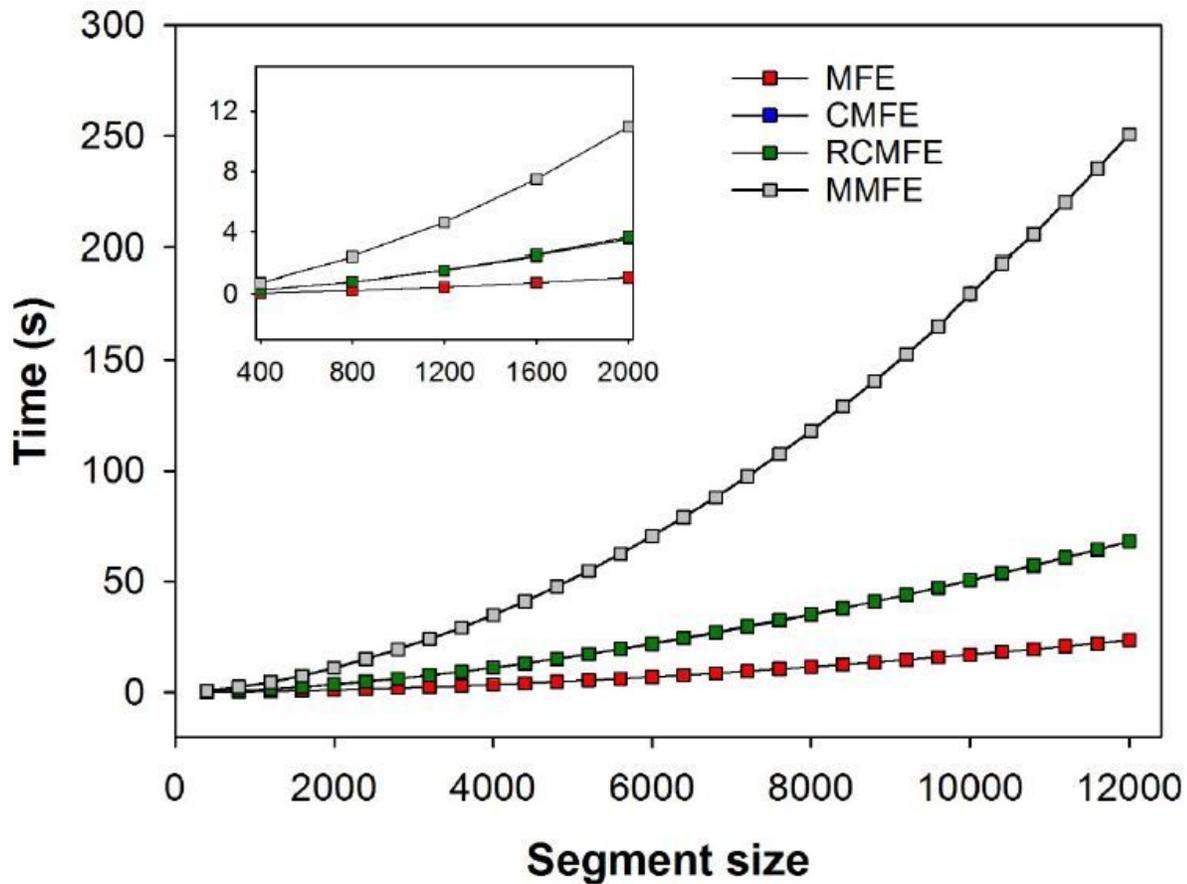


Figure 8

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 three HRV series of humans up to segments of 12000 points. Of note, CMFE and RCMFE take virtually the same time to be computed, and therefore, these curves are superimposed.