Effects of the series length on Lempel-Ziv Complexity during sleep

Massimo W. Rivolta[1], Matteo Migliorini[2], Md Aktaruzzaman[1], Roberto Sassi[1] and Anna M. Bianchi[2]

***Abstract*— Lempel-Ziv Complexity (LZC) has been demonstrated to be a powerful complexity measure in several biomedical applications. During sleep, it is still not clear how many samples are required to ensure robustness of its estimate when computed on beat-to-beat interval series (RR). The aims of this study were: i) evaluation of the number of necessary samples in different sleep stages for a reliable estimation of LZC; ii) evaluation of the LZC when considering inter-subject variability; and iii) comparison between LZC and Sample Entropy (SampEn). Both synthetic and real data were employed. In particular, synthetic RR signals were generated by means of AR models fitted on real data.**

The minimum number of samples required by LZC for having no changes in its average value, for both NREM and REM sleep periods, was 10^4 ($p<0.01$) when using a binary quantization. However, LZC can be computed with $N>1000$ when a tolerance of 5% is considered satisfying.

The influence of the inter-subject variability on the LZC was first assessed on model generated data confirming what found ($>10^4$; $p<0.01$) for both NREM and REM stage. However, on real data, without differentiate between sleep stages, the minimum number of samples required was 1.8×10^4.

The linear correlation between LZC and SampEn was computed on a synthetic dataset. We obtained a correlation higher than 0.75 ($p<0.01$) when considering sleep stages separately, and higher than 0.90 ($p<0.01$) when stages were not differentiated.

Summarizing, we suggest to use LZC with the binary quantization and at least 1000 samples when a variation smaller than 5% is considered satisfying, or at least 10^4 for maximal accuracy. The use of more than 2 levels of quantization is not recommended.

I. INTRODUCTION

The Lempel-Ziv Complexity (LZC) for sequences of finite length was proposed by Lempel and Ziv [1] and represents a simple way to measure signal complexity. The variation in complexity of physiological signals has shown to be sensitive to pathological condition in different studies. LZC has been used on electroencephalograms (EEG) to discriminate between wake and sleep conditions in patients under anaesthesia [2] and to compare the EEG background activity in subjects with or without Alzheimer's Disease [3]. Even the complexity of the Autonomic Nervous System (ANS) has been investigated through the computation of this parameter on heart rate variability (HRV) signals. In fact, there are studies that showed how changes in complexity can be referred to ventricular tachycardia and fibrillation [4], or to changes in mood state [5], or how the control of the ANS is modified in pathological conditions, like heart failure or sleep apnoea [6].

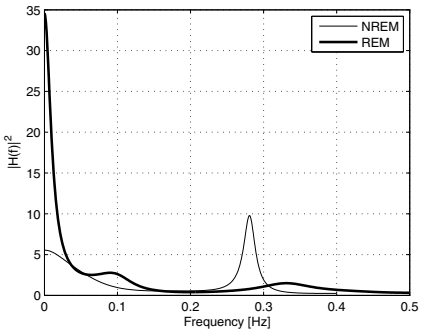

Fig. 1. Squared magnitude of the frequency response of two AR models during REM (bold line) and NREM (light line) sleep stage respectively.

Despite to its proven capability, what LZC can really measure on biomedical signals is still unclear. Aboy *et al.* [7] showed that LZC (when using binary quantization) is dependent on some frequency-related quantities. However, in their simulations, LZC was computed on running window of 10s, totally neglecting the influence of the number of samples and long-term non stationarities on the estimate. To move a step forward, we planned a few synthetic simulations and real data analysis to verify which minimum number of samples should be employed for obtaining a robust estimate on HRV signals expressed as RR series. Moreover, we compared LZC with another complexity measure, *i.e.* SampEn, in order to assess at which extend they are related. We focused our efforts on RR series extracted during sleep because of its optimal signal-to-noise ratio. However, the RR series during sleep can vary significantly with the different sleep stages [8]. For this reason we considered three sleep stages (according to standard sleep labeling [9], [10]): Light Sleep (LS), represented by NREM stage 1 and 2; Deep Sleep (DS), represented by NREM stage 3 and 4; and Rapid-Eye Movements (REM).

Summarizing, the aims of the study were three: i) evaluation of LZC in terms of samples and number of quantization levels; ii) evaluation of LZC when considering inter-subject variability; and iii) comparison of LZC and SampEn.

II. METHODS

A. Dataset

A portion of "The Cyclic Alternating Pattern (CAP) of EEG activity during sleep" dataset [11], [12] was employed in this study. In particular, 13 out of 16 healthy individuals

[1]MWR, AM and RS are with the Dipartimento di Informatica, Università degli Studi di Milano, Crema (26013), Italy massimo.rivolta@unimi.it

[2]MM and AMB are with the Department of Electronic, Information and Biomedical Engineering, Politecnico of Milano, Milano (20133), Italy matteo.migliorini@polimi.it

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

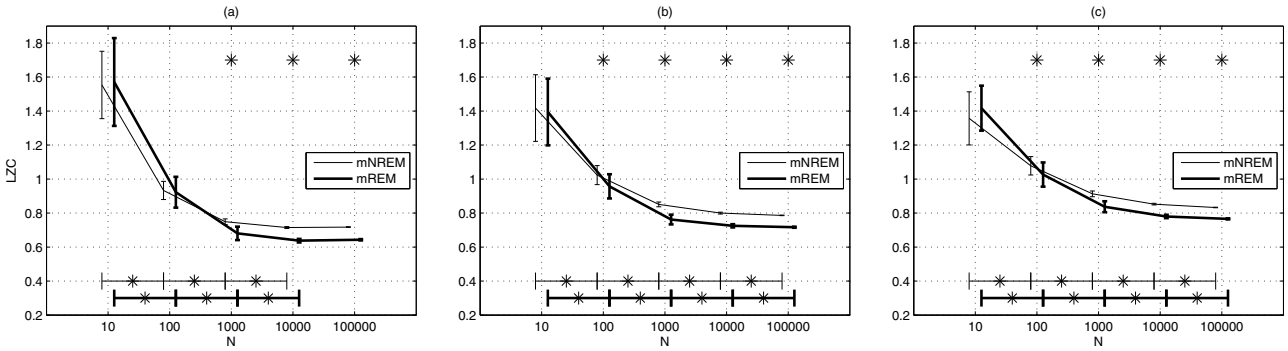

Fig. 2. Mean and standard deviation of LZC as function of the series length N when considering mNREM (light line) and mREM (bold line) and with levels of quantization $L = 2$ (a), $L = 3$ (b) and $L = 4$ (c). $*$ on the horizontal bars refer to the statistical difference in the average estimation between successive series lengths N, and $*$ on the top are used to denote the statistical difference between mNREM and mREM. $*$ refers to p<0.01 of double-tail t-test.

were selected for the following analysis (3 subjects were removed due to the low quality or absence of the ECG).

The ECG was collected for each subject during sleep. Such recordings were used to extract RR series. The sleep stage annotation series was provided with the dataset and employed for segmenting sleep stages.

B. Simulations

1) Evaluation of the series length: The assessment of the number of the samples required by LZC was performed on synthetic series generated by AR models. In particular, two AR models, during NREM (mNREM) and REM (mREM) respectively, were estimated from real data and used for generating synthetic signals. During sleep, the frequency content of RR series varies as function of the sleep stage. Generally, the interaction of ANS is reflected on a predominant power in the low frequency band during REM, and in high frequencies during NREM. Figure 1 shows the squared magnitude of the frequency response of both models. LZC was evaluated on a synthetic dataset composed by 30 independent realizations of the AR process (during REM or NREM) after being quantized with either 2, 3 or 4 levels L (see sec. II-E). The series length N was set to 10, 10^2, 10^3, 10^4 and 10^5 samples.

The evaluation of the number of samples was carried by two procedures. First, the percentage of variation (absolute value), with respect to the average value at 10^5, was computed and the number of samples for having lesser than a prefixed threshold of variation (either 1% or 5%) was found by linear interpolation. Second, a statistical test (double-tail t-test) was employed to compare the average LZC between successive values of N (*e.g.* 10 vs 10^2, 10^2 vs 10^3, etc.). Moreover, mean values of LZC for mNREM were compared with those for mREM to evaluate whether it can distinguish the two populations even before reaching the minimum percentage of variation (double-tail t-test).

2) Inter-subject variability of LZC: The inter-subject variability of LZC was evaluated as function of the series length N. To do that, a specific synthetic dataset was built employing the AR models identified on real series (see sec. II-D). In details, 30 independent realizations were generated for each model selecting an increasing number of samples, and the average LZC value was computed. The percentage of variation was calculated respect to the LZC value computed when the maximum number of samples was employed. In such analysis, we considered LS, DS and REM stages separately. The statistical analysis of sec. II-B.1 were performed (excluding comparison between sleep stages).

3) Relationship between LZC and SampEn: The relationship between LZC and SampEn was evaluated on a synthetic dataset (30 independent realizations) generated by the same AR models of sec II-B.2. LZC was computed on series with a length of 2.3×10^4. SampEn was determined by the procedure described in [13] in which its theoretical value, *i.e.* the value when $N \to \infty$, was computed analytically knowing the coefficients of the AR model. The parameters of SampEn, *i.e.* m and r, were set to 1 and 0.2 (after power normalization) respectively. The linear correlation between LZC and SampEn was computed considering: i) groups LS, DS and REM separately; and ii) no groups. It is worth noting that LZC and SampEn were computed on different dataset because we were interested in the average correlation, not dependent on the specific realization.

C. Real data analysis

Each RR series was first preprocessed (sec. II-D), quantized (sec. II-E) and then, LZC was computed on windows with an increasing length, up to reaching a maximum value of 2.3×10^4 samples. The standard deviation of LZC, after subtraction of LZC computed at 2.3×10^4, was employed for assessing the variability at each series length N.

The number of samples for having a reduction of factor 10 of the standard deviation with respect to that at N=1000 was considered acceptable as a robust estimate of LZC.

D. Preprocessing and data modelling

AR models of fixed order 9 [6] were fitted on consecutive windows of 400 RR samples (overlap 0%) previously edited for artifact elimination. Each real RR series was high pass filtered by means of a median filter of 200 samples and afterwards, those RR intervals outside 3 times the interquartile range were removed from the series. Model fitting was

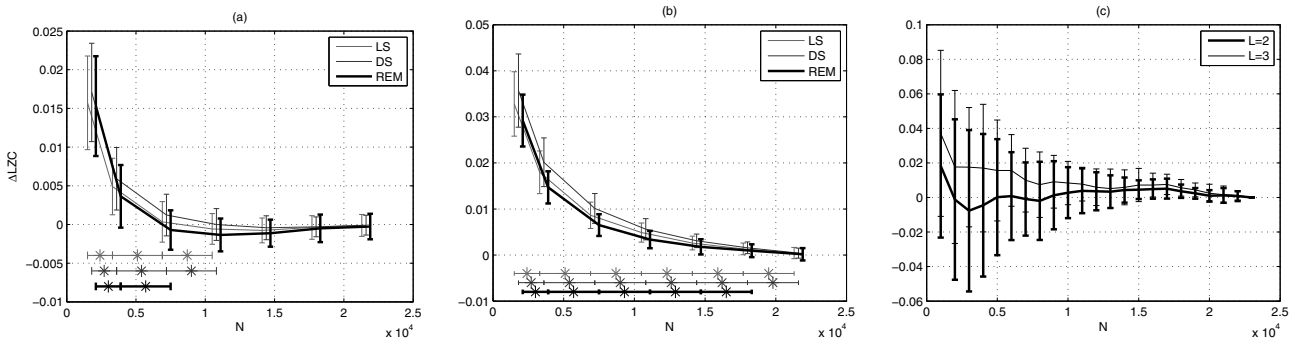

Fig. 3. Inter-subject variability of LZC. Simulated data: mean and standard deviation of LZC (along models) as function of the series length N is reported in the first two panels when considering L=2 (a) and L=3 (b). Real data: mean and standard deviation of LZC (along subjects) as function of the series length N is reported in panel (c). ΔLZC=LZC(N)-LZC(2.3×10^4). $*$ refers p$<$0.01 of a non-paired double-tail t-test evaluated on consecutive N values.

performed separately for each subjects within a specific sleep stage employing the Yule-Walker equations. Those models which obtained a prediction error variance higher than 0.05 s^2 were excluded from the following analysis. A variable number of models was thus available depending on subject and sleep stage (272 LS, 219 DS and 169 REM).

E. Series quantization

The quantization was performed by dividing the amplitude of each RR series in intervals and labeling them with a different symbol. A binary quantization is normally employed in the computation of LZC and the median value of the series is used as the quantization threshold whose divides the series distribution at half [7], [3]. When more than two levels were required, intervals were built using percentiles (being maximum and minimum value of the series too much dependent on artifacts). Each interval was defined by percentiles equally far from each other, *i.e.* intervals having the same probability. For example, when setting L=2 the median value was used as threshold, with L=3 the 33^{th} and the 66^{th} percentiles were employed, and so on.

III. RESULTS

A. Simulations

1) Evaluation of the series length: The number of necessary samples required by LZC, *i.e.* for having no statistical significant changes in its average value, was evaluated for two AR models (mNREM and mREM). In particular, when considering mNREM, the 1% of variation was reached when N was 9300 with L=2 (1000 at 5%), 52000 with L=3 (5500 at 5%) and 66100 with L=4 (6800 at 5%). Similar results were obtained when considering mREM (at 1%: 53500 with L=2, 45300 with L=3 and 64900 with L=4; at 5%: 3900 with L=2, 3500 with L=3 and 6300 with L=4). However, the mean value of LZC at N=10^4 was not different to that at N=10^5 only when considering L=2 for both models (fig. 2; p$<$0.01).

2) Inter-subject variability of LZC: The influence of the inter-subject variability on LZC was evaluated. The maximum number of samples required by LZC for having the 1% of variation was 3500 with L=2 and 9000 for L=3 (1100 with L=2 and 1700 with L=3 at 5%). However, even in this case, only for L=2 and when employing more than 10^4, mean values of LZC were not distinguishable (fig. 3a and fig. 3b; p$<$0.01).

3) Relationship between LZC and SampEn: The relationship between LZC and SampEn was evaluted for LS, DP and REM after separate grouping (fig. 4) and when no groups were considered. In the first case, the linear correlation was always higher than 0.75 (p$<$0.01). In particular, when considering LS and REM with L=2, the correlation was higher than 0.90. In the second case, the linear correlation was higher than 0.90 (p$<$0.01) for both L=2 and L=3 (the sample was not balanced). Table I summarizes the relations and the correlations found.

TABLE I

THE LINEAR RELATIONSHIP BETWEEN LZC AND SAMPEN (ALL) IS SHOWN. ALSO, RELATIONS ARE REPORTED AS FUNCTION OF THE SLEEP STAGE AND THE LEVEL OF QUANTIZATION L. LINEAR CORRELATION IS SHOWN IN BRACKETS ($*$ REFERS TO $p<0.01$).

Sleep stage	L=2	L=3
LS	2.20×LZC + 0.52 (0.92 *)	2.17×LZC + 0.38 (0.93 *)
DS	1.52×LZC + 1.02 (0.75 *)	1.65×LZC + 0.81 (0.81 *)
REM	2.25×LZC + 0.44 (0.97 *)	2.26×LZC + 0.28 (0.98 *)
ALL	2.20×LZC + 0.52 (0.90 *)	2.21×LZC + 0.35 (0.93 *)

B. Real data

LZC and its variability were computed on windows with different series length N (fig. 3c). For both L=2 and L=3, the minimum number of samples for having a reduction of factor 10 of the standard deviation with respect to that at N=1000 was 18000.

IV. CONCLUSIONS

In this work, we verified how many samples are required to ensure robustness of LZC estimation when used to measure the complexity of beat-to-beat series. We also assessed the influence of the inter-subject variability when computing LZC. Secondarily, we estimated the linear correlation between LZC and SampEn.

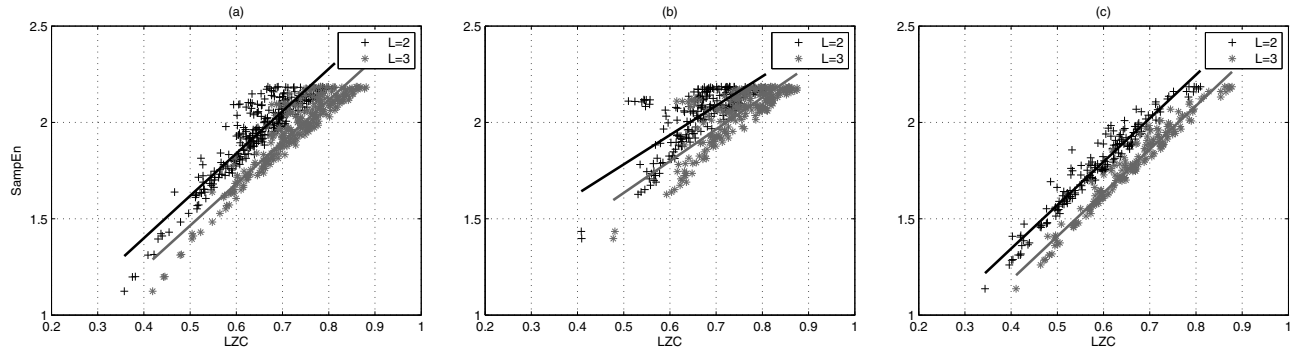

Fig. 4. Scatter plot and linear regression between LZC and SampEn when considering LS (a), DS (b) and REM (c) with L=2 (gray) and L=3 (black).

During a specific sleep stage, the minimum number of samples required by LZC, *i.e.* for having no changes in its average value, is 10^4 (practically impossible to collect for a single sleep stage) when employing binary quantization (fig. 2; $p<0.01$). Such result is still valid considering the inter-subject variability (fig. 3a and fig. 3b; $p<0.01$). However, a variation smaller than 5% was found when employing $N > 1000$ for both L=2 and L=3.

A number of quatization levels higher than 2 is not recommended because more than 10^5 samples are required (fig. 2). It remains to accurately evaluate whether LZC can discriminate between NREM and REM even before convergence (partially demonstrated; fig. 2; $p<0.01$) and, if its value reflects physiological information.

Furthermore, the quantization method employed converts the RR distribution, whatever it was, into a uniform distribution of symbols. Such method overestimates the number of samples required when $L > 2$. The study of which quantization technique should be used, as a function of the specific application, was postponed for future works.

On real data, due to the presence of non-stationaries, a proper evaluation of the number of samples required by LZC during sleep is not feasible. However, the cyclicity of sleep stages during the night leads to a reduction of the variability of the LZC when increasing the number of samples making possible an empirical evaluation.

Finally, the linear correlation between LZC and SampEn was assessed on a synthetic dataset (tab. I and fig. 4, >0.90; $p<0.01$). However, the quantization procedure was not the same employed for LZC (standard parameter were considered for SampEn). Such comparison was meant to evaluate whether the two measures carry different information. From our results, it seems to be excluded. Therefore, when LZC is required on short series, the methodology proposed by Aktaruzzaman and Sassi [13] could overcome such issue. This result suggests to verify which of these two measures converges more rapidly when long series are available.

Summarizing, we suggest to use LZC with the binary quantization and at least 1000 samples when a variation smaller than 5% is considered satisfying, or at least 10^4 for maximal accuracy. The use of more than 2 levels of quantization is not recommended.

REFERENCES

[1] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Trans Inform Theory*, 22(1):75–81, 1976.

[2] X. S. Zhang, Rob J. Roy, and E. W. Jensen. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans Biomed Eng*, 48(12):1424–1433, 2001.

[3] D. Abásolo, R. Hornero, C. Gómez, M. García, and M. López. Analysis of EEG background activity in alzheimer's disease patients with LempelZiv complexity and central tendency measure. *Med Eng Phys*, 28(4):315–322, 2006.

[4] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang. Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Trans Biomed Eng*, 46(5):548–555, 1999.

[5] M. Migliorini, M. O. Mendez, and A. M. Bianchi. Study of heart rate variability in bipolar disorder: linear and non-linear parameters during sleep. *Front Neuroeng*, 4(22):1–7, 2012.

[6] A. M. Bianchi, M. O. Mendez, M. Ferrario, L. Ferini-Strambi, and S. Cerutti. Long-term correlations and complexity analysis of the heart rate variability signal during sleep: Comparing normal and pathologic subjects. *Methods Inf Med*, 49(5):479–483, 2010.

[7] M. Aboy, R. Hornero, D. Abásolo, and D. Álvarez. Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans Biomed Eng*, 53(11):2282–2288, 2006.

[8] R. Cabiddu, S. Cerutti, G. Viardot, S. Werner, and A. M. Bianchi. Modulation of the sympatho-vagal balance during sleep: Frequency domain study of heart rate variability and respiration. *Front Physiol*, 3(45):1–10, 2012.

[9] A. Kales and A. Rechtschaffen. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Allan Rechtschaffen and Anthony Kales, editors.* U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[10] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, Sairam Parthasarathy, Stuart F Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, M. M. Tangredi, and American Academy of Sleep Medicine. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the american academy of sleep medicine. *J Clin Sleep Med*, 8(5):597–619, 2012.

[11] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (cap) in human sleep. *Sleep Med*, 2(6):537–553, 2001.

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C-K Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[13] M. Aktaruzzaman and R. Sassi. Sample entropy parametric estimation for heart rate variability analysis. In *Comput Cardiol*, pages 429–432, 2013.