35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Entropy-based Complexity of the Cardiovascular Control in Parkinson Disease: Comparison between Binning and K-Nearest-Neighbor Approaches

Alberto Porta, *Member, IEEE*, Vlasta Bari, Tito Bassani, Andrea Marchi, Stefano Tassin, Margherita Canesi, Franca Barbic and Raffaello Furlan

Abstract— **Entropy-based approaches are frequently used to quantify complexity of short-term cardiovascular control from spontaneous beat-to-beat variability of heart period (HP) and systolic arterial pressure (SAP). Among these tools the ones optimizing a critical parameter such as the pattern length are receiving more and more attention. This study compares two entropy-based techniques for the quantification of complexity making use of completely different strategies to optimize the pattern length. Comparison was carried out over HP and SAP variability series recorded from 12 Parkinson's disease (PD) patients without orthostatic hypotension or symptoms of orthostatic intolerance and 12 age-matched healthy control (HC) subjects. Regardless of the method, complexity of cardiovascular control increased in PD group, thus suggesting the early impairment of cardiovascular function.**

I. Introduction

More and more studies suggest that complexity analysis of cardiovascular variabilities can complement more traditional techniques defined in time and frequency domain [1-7]. This result is not surprising because traditional indexes are related to the amount of variability independently of the frequency (i.e. the variance) and/or in specific frequency bands (e.g. respiratory sinus arrhythmia), whereas complexity indexes are linked to the variety of mechanisms simultaneously activated to regulate cardiovascular variables [8]. Some studies suggest that complexity analysis performed over short time series can provide indexes as powerful as those derived from longer sequences undergoing fractal analysis [7], thus enlarging the possibility of using complexity analysis in clinical settings. Among the tools devised to be reliable over short sequences those based on predictability, symbolic analysis and entropy appear the most frequently exploited [1-7].

A. Porta and T. Bassani are with Department of Biomedical Sciences for Health, Galeazzi Orthopedic Institute, University of Milan, Milan, Italy (tel: +39 02 50319976; fax: +39 02 50319979; e-mails: alberto.porta@unimi.it and tito.bassani@libero.it).

V. Bari is with Gruppo Ospedaliero San Donato Foundation, Milan, Italy and Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy (e-mail: vlasta.bari@mail.polimi.it).

A. Marchi is with Department of Emergency, L. Sacco Hospital, Milan, Italy (email: marchi.andrea1@gmail.com).

S. Tassin is with Ecker Technologies Sagl, Switzerland (email: stefano.tassin@ecker-tech.com).

M. Canesi is with Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy (email: margeritag2007@yahoo.it).

F. Barbic and R. Furlan are with Department of Medical Biotechnologies and Translation Medicine, Medical Clinics, Istituto Clinico Humanitas, University of Milan, Rozzano, Milan, Italy (emails: franca.barbic@humanitas.it and raffaello.furlan@unimi.it).

Tools operating over short sequences are fundamentally based on the construction of patterns according to the technique of the delayed coordinates [9]. As a consequence pattern length is a critical parameter for the short-tern complexity analysis. Two main strategies are followed in literature: i) pattern length is arbitrarily assigned to small values, usually decided in relation to the frame length [1-5]; ii) pattern length is chosen on a case-by-case basis according to some optimization criteria [6,7,10,11]. Tools optimizing the pattern length should be theoretically preferred because they are more flexible in presence of between- and within-subject variations, thus possibly reducing the variance of the complexity index and increasing the statistical power of analysis.

The aim of this study is to compare two approaches for the assessment of short-term entropy-based complexity optimizing pattern length: corrected conditional entropy (CCE) [6,10] and k-nearest-neighbor conditional entropy (KNNCE) [11]. These two approaches search for the optimal number of previous samples allowing the most reliable quantification of the information carried by the future sample based on previous values. While CCE exploits a corrective term penalizing long patterns, KNNCE makes use of the natural dispersion of patterns in the embedding space due to noise leading to a gradual increase of information with pattern length. The two methods will be compared over spontaneous heart period (HP) and systolic arterial pressure (SAP) variability series recorded from Parkinson's disease (PD) patients without orthostatic hypotension or symptoms of orthostatic intolerance [12] to assess the ability of entropy-based approaches to detect the early impairment of the cardiovascular regulation.

II. Conditional Entropy Evaluation

A. CCE

The CCE was proposed as an entropy-based method for the evaluation of complexity of a time series $x=\{x(i), i=1,\ldots,N\}$, where i is the sample counter and N is the series length [6,10]. It was devised originally to avoid the a priori assignment of the pattern length L. This setting is usually necessary to assess approximate entropy [1] and sample entropy [2]. Full details about CCE definition can be found in [6,10]. Briefly, defined as $x_L(i)=(x(i),x(i-1),\ldots,x(i-L+1))$ the ordered sequence of L consecutive samples formed by the most recent value x(i) and by the previous L-1 samples, $x_{L-1}(i-1)=(x(i-1),\ldots,x(i-L+1))$, the approach is based on the evaluation of conditional entropy (CE) measuring the

average amount of information carried by x(i) when $x_{L-1}(i-1)$ is given. CE is calculated as

$$CE(L) = -\sum p(x_{L-1}(i-1)) \cdot SE(x / x_{L-1}(i-1)) \quad (1)$$

where $p(x_L(i))$ is the probability of the pattern $x_L(i)$ and $SE(x/x_{L-1}(i-1))$ is the Shannon entropy of the conditional distribution of x given $x_{L-1}(i-1)$ computed as

$$SE(x / x_{L-1}(i-1)) = \sum p(x(i) / x_{L-1}(i-1)) \cdot \log(p(x(i) / x_{L-1}(i-1))) \quad (2)$$

where $p(x(i)/x_{L-1}(i-1))$ is the conditional probability of x(i) given $x_{L-1}(i-1)$ and log is natural logarithm. CE(1) is set as the SE of x

$$SE(x) = -\sum p(x(i)) \cdot \log(p(x(i))) \quad (3)$$

measuring the average amount of information carried by x when no previous samples are given. Uniform quantization over a fixed number of bins, ξ, allows the estimation of $p(x_L(i))$ and $p(x(i)/x_{L-1}(i-1))$ [13]. Indeed, since uniform quantization imposes a partition of the L-dimensional embedding space into ξ^L cells of side $(\max(x)-\min(x))/\xi$, where max(x) and min(x) represent maximum and minimum of x, $p(x_L(i))$ and $p(x(i)/x_{L-1}(i-1))$ can be estimated as the sample frequencies of $x_L(i)$ and x(i) given $x_{L-1}(i-1)$ [13].

Unfortunately, when assessed over a limited amount of data, the reliability of the conditional distribution of x(i) given $x_{L-1}(i-1)$ decreases with L. This degradation is the consequence of the increasing sparseness of the patterns in the L-dimensional embedding space. This sparseness is due to gradual spread of the patterns with L inducing a continuing reduction of the number of patterns in the cells of the partition, thus producing a bias toward the artificial decrease of CE with L. In order to prevent the artificial decline of CE a corrective term was added to the CE estimate, thus leading to the definition of CCE [6,10]. The corrective term tackles the loss of reliability of CE with L by substituting the trend toward a null complexity with a drift toward the maximum complexity quantified by SE(x). As a consequence of its design, CCE exhibits a minimum and this minimum is taken as complexity index (CI_{CCE}) [6,10]. Normalized CI_{CCE} (NCI_{CCE}) is obtained by dividing CI_{CCE} by SE(x) [6,10], thus obtaining an index ranging from 0 to 1, where 0 indicates null complexity and maximum regularity and 1 indicates the opposite situation.

B. KNNCE

As in the case of CCE, KNNCE was proposed [11] to allow entropy-based complexity analysis without a priori fixing the pattern length. At difference with CCE, KNNCE does not exploit a corrective term to compensate the bias of the CE estimate. The KNNCE is computed as

$$KNNCE(L) = \frac{1}{N-L+1} \sum_{i=L}^{N} SE(x / x_{L-1}(i-1)) \quad (4).$$

This approach takes advantage of a completely different strategy to assess $SE(x/x_{L-1}(i-1)$ compared to the uniform quantization procedure exploited by CCE. Indeed, the distribution of x given $x_{L-1}(i-1)$ was built by considering the next value of the k patterns nearest to $x_{L-1}(i-1)$ regardless of their actual distance from it. The distance was computed according to the Euclidean norm. This strategy prevents the degradation of the reliability of the estimated conditional distributions with L given that k values are always considered regardless of L. The final course of KNNCE is the balance between two opposite tendencies: i) if past values are helpful to reduce the uncertainty associated to future samples, KNNCE decreases with L; ii) in presence of noise patterns progressively take apart with L in the multidimensional embedding space and, consequently, uncertainty in predicting future values raises with L. The two opposite tendencies result in a minimum in presence of repetitive patterns. The value of KNNCE at the minimum is taken as CI_{KNNCE} [11]. Normalized CI_{KNNCE} (NCI_{KNNCE}) is obtained by dividing CI_{KNNCE} by SE(x) [11].

III. Experimental Protocol and Data Analysis

A. Experimental Protocol

Data belongs to a database designed to monitor the early impairment of the cardiovascular control in PD patients via spectral analysis of spontaneous HP and SAP variabilities [12]. We studied 12 PD patients without orthostatic hypotension or symptoms of orthostatic intolerance (age range: 55-79 years, median 65 years, 8 males) and 12 healthy control (HC) subjects matched by age and gender with PD group (age range: 58-72 years, median 67 years, 7 males). PD patents (2-4 Hoehn-Yhar scale) were at the best of their habitual pharmacological treatment. ECG (lead II) and noninvasive arterial pressure (Finapress 2300, Ohmeda, Englewood, Colorado, USA) were recorded. Sampling rate was 300 Hz. Signals were recorded with the subject lying on the tilt table, gently fasten by two belts, and with both the feet touching the footrest of the tilt table. Recordings of 10 minutes were obtained at baseline (B) and during 75° head-up tilt (T75). Subjects were breathing spontaneously but were not allowed to talk. All the subjects gave their written informed consent. The study is in keeping with the principles of the Declaration of Helsinki for medical research involving humans. The human research and ethical review boards of the University of Milan, Italy, approved the protocol.

B. Time Series Extraction and Data Analysis

After detecting the QRS complex on the ECG and locating the R-apex using parabolic interpolation, the temporal distance between two consecutive R parabolic apexes was computed and utilized as an approximation of HP. The maximum of arterial pressure inside HP was defined as SAP. The occurrences of QRS and SAP peaks were carefully checked to avoid erroneous detections or missed beats. If isolated ectopic beats affected HP and SAP values, these measures were linearly interpolated using the closest values unaffected by ectopic beats. HP and SAP were extracted on a beat-to-beat basis. The series were linearly detrended. Sequences of 256 consecutive measurements were randomly selected inside each experimental condition, thus focusing on short-term cardiovascular regulation [12]. If evident nonstationarities, such as changes of the mean and/or variance, were present despite the linear detrending, the

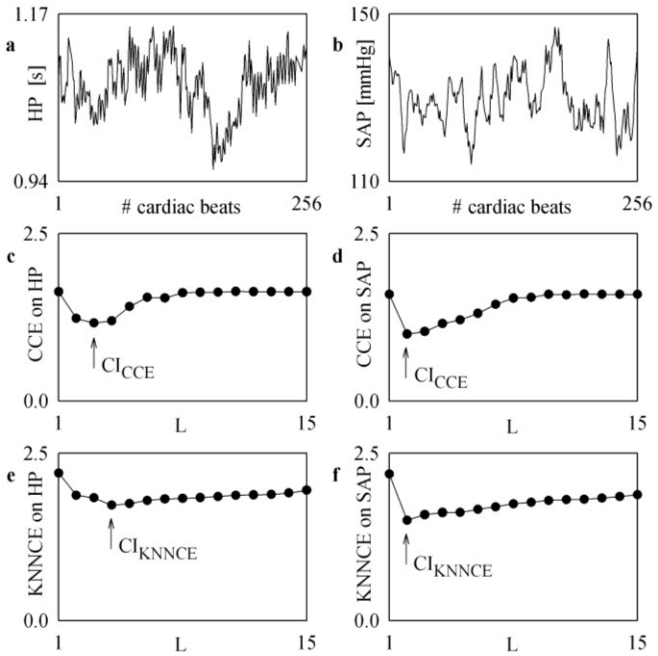

Fig.1. Examples of HP and SAP series are shown in (a) and (b) respectively. CCE functions computed over (a) and (b) are depicted in (c) and (d) respectively, while KNNCE functions in (e) and (f) respectively.

random selection was carried out again. NCI_{CCE} and NCI_{KNNCE} were calculated from both HP and SAP series. NCI_{CCE} was computed with ξ=6 [6,10] and NCI_{KNNCE} with k=30 [11].

C. Statistical Analysis

We performed paired t-test to check the significance of the difference between NCI_{CCE} (or NCI_{KNNCE}) assessed from HP and SAP series regardless of the experimental condition and groups. Unpaired t-test was utilized to check the significance of the difference between NCI_{CCE} (or NCI_{KNNCE}) computed over HC subjects and PD patients regardless of the type of series and experimental condition. Two way repeated measures analysis of variance (one factor repetition, Holm-Sidak test for multiple comparisons) was performed to assess the effect of T75 on NCI_{CCE} (or NCI_{KNNCE}) within the same population (HC or PD group) and the difference between groups within the same experimental condition (B or T75). A $p<0.05$ was always considered as significant.

IV. Results

Figure 1 shows an example of HP (Fig.1a) and SAP (Fig.1b) series derived from a HC subject. CCE functions, computed from SAP and HP series (Figs.1c,d), exhibited a deep minimum: it was detected at L=3 in Fig.1c and at L=2 in Fig.2d. KNNCE functions (Figs.1e,f) showed a clear minimum as well. When assessed over HP series (Fig.1e) the minimum was found at L=4, while it remained at L=2 over SAP series (Fig.1f).

Figure 2 shows box-and-whiskers plots reporting 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the normalized minimum of the CCE and KNNCE (i.e. NCI_{CCE} and NCI_{KNNCE}) as a function of the series (Figs.2a,b) and population (Figs.2c,d). In Figs.2a,b data are pooled together regardless of the experimental condition and population,

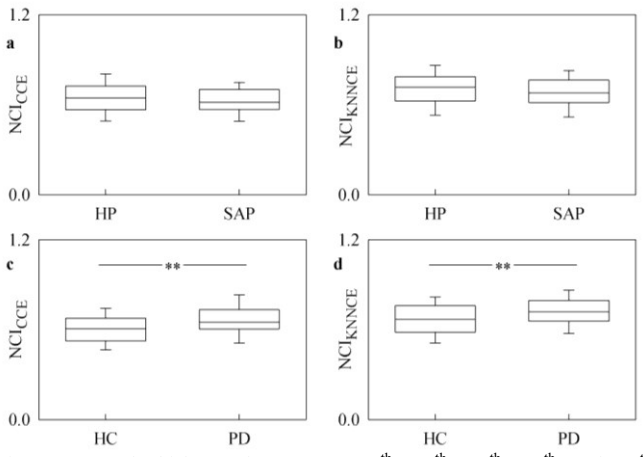

Fig.2. Box-and-whiskers plots report 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of NCI_{CCE} and NCI_{KNNCE} assessed as a function of the series regardless of the experimental condition and population in (a) and (b) respectively and regardless of the experimental condition and series in (c) and (d) respectively. The symbol ** indicates $p<0.01$.

whereas in Figs.2c,d regardless of the series and experimental condition. Complexity of cardiovascular control as assessed from NCI_{CCE} computed from HP series was similar to that of the vascular control as inferred from NCI_{CCE} calculated from SAP series (Fig.2a). The same finding held in case of NCI_{KNNCE} (Fig.2b). PD patients were characterized by a higher complexity of cardiovascular control (Fig.2c) and this conclusion did not depend on the method (i.e. parallel changes were observable in Fig.2d).

Box-and-whiskers plots reporting 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of NCI_{CCE} and NCI_{KNNCE} as a function of the experimental condition are shown in Figs.3,4. Indexes shown in Fig.3 were assessed over HP series, whereas those in Fig.4 over SAP series. Some tendencies are clearly visible in Fig.3a: i) $NCI_{CCE,HP}$ decreased during T75 in HC subjects; ii) the $NCI_{CCE,HP}$ decrease during T75 was less evident in PD patients; iii) in PD patients in both experimental conditions $NCI_{CCE,HP}$ tended to be larger. Trends of $NCI_{KNNCE,HP}$ (Fig.3b) were similar to those of $NCI_{CCE,HP}$ (Fig.3a). The tendency of complexity to increase in PD group was corroborated by the findings shown in Fig.4a. Indeed, at B $NCI_{CCE,SAP}$ was significantly larger in PD patients than in HC subjects. Similar between-group changes were observed during T75 but modifications did not reach the significance level. $NCI_{KNNCE,SAP}$ showed similar trends but this index was less powerful than $NCI_{CCE,SAP}$: indeed, the within-condition difference between groups was never significant.

V. Discussion

The study compares two entropy-based methods that do not need to fix the length of the conditioning pattern. Both the methods exploit a minimization procedure to find out the optimal length of the conditioning pattern. The CCE function exhibits a minimum as a result of the compensation between two opposite tendencies: the better ability of longer conditioning patterns to predict future samples, thus reducing CE with L, and the increase of the corrective term accounting for the loss of reliability of estimated conditional distributions as a function of the pattern length L [6,10]. The

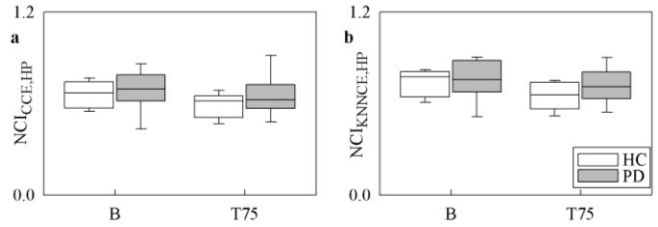

Fig.3. NCI_{CCE} (a) and NCI_{KNNCE} (b) assessed over HP series as a function of the experimental condition in HC and PD groups.

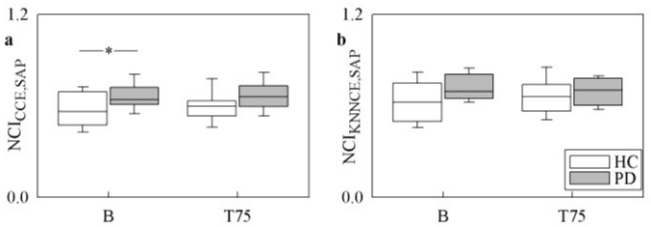

Fig.4. NCI_{CCE} (a) and NCI_{KNNCE} (b) assessed over SAP series as a function of the experimental condition in HC and PD groups. The symbol * indicates $p<0.05$.

minimum of the KNNCE function is the consequence of the balance between the ability of a higher dimensional embedding space to resolve the ambiguities of the system dynamics, thus improving prediction and limiting entropy rate, and the increasing sparseness of the k nearest patterns with L due to the presence of noise leading to flatter and flatter conditional distributions as a function of L [11]. The KNNCE function has an important theoretical advantage compared to CCE: the reliability of the conditional distributions does not degrade with the pattern length L because they are constructed based on k samples regardless of L [11]. This study compares these two approaches in PD patients without orthostatic hypotension or symptoms of orthostatic intolerance. In this pathological group the early impairment of the cardiovascular control was observed using traditional spectral analysis [12]. A maneuver inducing a sympathetic activation (i.e. T75) was exploited to challenge cardiovascular control [12].

The main finding of this study is that CCE provides results comparable to KNNCE, thus suggesting that a binning procedure is equivalent to a k-nearest-neighbor approach. This result supports the conclusion that CCE counteracts the loss of reliability of the estimated conditional distributions with L [6,10].

Findings confirm that complexity of cardiovascular control is increased in PD patients as previously detected using a model-based predictability approach [7]. The increase of complexity of cardiovascular control is already evident at B, thus suggesting that it is not necessary to challenge cardiovascular regulation to distinguish HC subjects from PD patients. However, it appears that the statistical power of entropy-based model-free indexes of complexity is remarkably smaller than that of the model-based predictability ones [7]: indeed, the within-group differences due to modification of the experimental condition and between-group variations within the same experimental condition, clearly visible in Figs.3,4, do not reach the same level of significance as reported in [7]. As a consequence a model-based predictability approach to the quantification of complexity of cardiovascular control seems to be preferable to model-free entropy-based techniques.

VI. Conclusion

Entropy-based model-free analysis confirmed the increase of complexity of cardiovascular control observed through a model-based predictability tool in PD and the capability of complexity analysis to detect the early impartment of cardiovascular control well before that evident signs of autonomic dysregulation become manifest.

References

[1] G.M. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos*, vol. 5, pp. 110-117, 1995.

[2] J.S. Richman and J.R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol*, 278 H2039-H2049, 2000.

[3] N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, and A. Voss, "Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates," Phys Rev E, vol. 61, pp. 733-739, 2000.

[4] S. Schulz, M. Koschke, K.-J. Bar, and A. Voss, "The altered complexity of cardiovascular regulation in depressed patients," *Physiol Meas*, vol. 31, pp. 303-321, 2010.

[5] M. Javorka, Z. Trunkvalterova, I. Tonhajzerova, J. Javorkova, K. Javorka, and M. Baumert, "Short-term heart rate complexity is reduced in patients with type 1 diabetes mellitus," *Clin Neurophysiol*, vol. 119, pp. 1071-1081, 2008.

[6] A. Porta, S. Guzzetti, N. Montano, M. Pagani, V.K. Somers, A. Malliani, G. Baselli, and S. Cerutti, "Information domain analysis of cardiovascular variability signals: evaluation of regularity, synchronisation and co-ordination," *Med Biol Eng Comput*, vol. 38, pp. 180-188, 2000.

[7] A. Porta, P. Castiglioni, M. Di Rienzo, V. Bari, T. Bassani, A. Marchi, A.C.M. Takahashi, E. Tobaldini, N. Montano, A.M. Catai, F. Barbic, R. Furlan, A. Cividjian, and L. Quintin, "Short-term complexity indexes of heart period and systolic arterial pressure variabilities provide complementary information," *J Appl Physiol*, vol. 113. pp. 1810-1820, 2012.

[8] A. Porta, M. di Rienzo, N. Wessel, and J. Kurths, "Addressing the complexity of cardiovascular regulation," *Phil Trans R Soc A*, vol. 367, pp. 1215-1218, 2009.

[9] F. Takens, "Detecting strange attractors in fluid turbulence," in *Dynamical Systems and Turbulence*, D. Rand and L.S. Young Eds, Berlin: Springer, 1981, pp. 366-381.

[10] A Porta, G Baselli, D Liberati, N Montano, C Cogliati, T Gnecchi-Ruscone, A Malliani, and S. Cerutti, "Measuring regularity by means of a corrected conditional entropy in sympathetic outflow," *Biol Cybern*, vol. 78, pp. 71-78, 1998.

[11] A. Porta, P. Castiglioni, V. Bari, T. Bassani, A. Marchi, A. Cividjian, L. Quintin, and M. Di Rienzo, "K-nearest-neighbor conditional entropy approach for the assessment of short-term complexity of cardiovascular control," *Physiol Meas*, vol. 34, pp. 17-33, 2013.

[12] F. Barbic, F .Perego, M. Canesi, M. Gianni, S. Biagiotti, G. Costantino, G. Pezzoli, A. Porta, A. Malliani, and R. Furlan, "Early abnormalities of vascular and cardiac autonomic control in Parkinson's disease without orthostatic hypotension," *Hypertension*, vol. 49, pp. 120-126, 2007.

[13] A. Porta, L. Faes, M. Masé, G. D'Addio, G.D. Pinna, R. Maestri, N. Montano, R. Furlan, S. Guzzetti, G. Nollo, and A. Malliani, "An integrated approach based on uniform quantization for the evaluation of complexity of short-term heart period variability: application to 24h Holter recordings in healthy and heart failure humans," *Chaos*, vol. 17, 015117, 2007.