

Cardiovascular Coupling-Based Classification of Ischemic and Dilated Cardiomyopathy Patients *

Javier Rodriguez, Steffen Schulz, Andreas Voss, and Beatriz F. Giraldo, *Senior Member, IEEE*

Abstract— Cardiovascular diseases are one of the most common causes of death in elderly patients. The etiology of cardiomyopathies is difficult to discern clinically. The objective of this study was to classify cardiomyopathy patients using coupling analysis, through their cardiovascular behavior and the baroreflex response. A total of thirty-eight cardiomyopathy patients (CMP) classified as ischemic (ICM, 25 patients) and dilated (DCM, 13 patients) were analyzed. Thirty elderly control subjects (CON) were used as reference. Their electrocardiographic (ECG) and blood pressure (BP) signals were studied. To characterize the cardiovascular activity, the following temporal series were extracted: beat-to-beat intervals (from the ECG signal), and end-systolic and diastolic blood pressure amplitudes (from the BP signal). Non-linear characterization techniques like high resolution joint symbolic dynamics, segmented Poincaré plot analysis, normalized short-time partial directed coherence, and dual sequence method were used to characterize these time series. The best indices were used to build support vector machine models for classification. The optimal model for ICM versus DCM patients achieved 84.2% accuracy, 76.9% sensitivity, and 88% specificity. When CMP patients and CON subjects were compared, the best model achieved 95.5% accuracy, 97.3% sensitivity, and 93.3% specificity. These results suggest a disfunction in the baroreflex mechanism in cardiomyopathies patients.

I. INTRODUCTION

Heart failure is one of the most common conditions suffered by elderly patients, due to age-related changes of the cardiovascular system. The etiology of these cardiomyopathies is often difficult to discern clinically. Differentiation between ischemic and dilated cardiomyopathy etiology has implications in the therapy, treatment, and prognosis of the patient.

* Research supported in part by the Secretariat of Universities and Research of the Department of Economy and Knowledge of the Government of Catalonia (Consolidated research group GRC 2017 SGR 1770), by CERCA Programme/Generalitat de Catalunya, by the Spanish Ministry of Economy and Competitiveness through project DPI2015-68820-R (MINECO/FEDER).

J. Rodriguez is with Institute for Bioengineering of Catalonia (IBEC), Automatic Control Dept. (ESAI), Universitat Politècnica de Catalunya (UPC), Spain.

A. Voss and S. Schulz are with the Ernst-Abbe-Hochschule Jena, University of Applied Sciences Jena, Institute of Innovative Health Technologies, Carl-Zeiss-Promenade 2, 07745 Jena, Germany (corresponding author to provide phone: 49-3641-205-625; fax: 49-3641-205-626; e-mail: steffen.schulz@eah-jena.de, andreas.voss@eah-jena.de).

B.F. Giraldo is with Automatic Control Dept. (ESAI), the Barcelona East School of Engineering (EEBE), Universitat Politècnica de Catalunya (UPC), Institute for Bioengineering of Catalonia (IBEC) – The Barcelona Institute of Science and Technology, and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Av. Eduard Maristany, 16, 08019, Barcelona, Spain. (beatriz.giraldo@upc.edu).

The cardiovascular interactions are typically complex by nature, and non-linear approaches are more suited to study their dynamics. Several methods have been explored for the quantitative analysis of the cardiovascular system [1, 2]. The study of the cardiovascular couplings can bring information about the baroreflex mechanism, quantifying their direct and indirect relationships under different pathological conditions. Some studies have applied coupling methods successfully on the analysis of similar interactions [3, 4].

In our previous work, we analyzed the cardiovascular interactions in cardiomyopathy patients using different univariate linear and non-linear methods [5-7]. In this study, we propose this analysis using coupling-based non-linear methods. Also, the application of support vector machines method for classification of the patients by their etiology.

II. DATABASE

The noninvasive electrocardiographic (ECG) and blood pressure (BP) signals were recorded from 38 cardiomyopathy patients (CMP) at the Santa Creu i Sant Pau Hospital in Barcelona, Spain (Fig.1.). All patients were studied according to a protocol, previously approved by the Hospital ethics committee. Twenty-five of these patients had a diagnosis of ischemic cardiomyopathy (ICM) while the rest (13) had been diagnosed with dilated cardiomyopathy (DCM) disease. Thirty elderly healthy subjects (68.3 ± 6.4 years) were used as a control group (CON). The clinical information is summarized in Table 1.

The recorded data was acquired with the Portapres-system and the Porti 16-biosignal amplifier for 15 minutes, at a sample frequency of 1600 Hz, with the patient in a supine position.

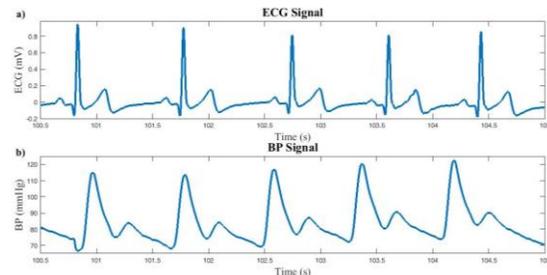


Figure 1. Excerpt of a) ECG signal, b) BP signal

III. METHODS

In the ECG and BP signals the linear trend were removed. Afterwards, the customized pre-processing tools were used to reduce artifacts and spikes, and the outliers were removed.

A. Signal processing

From the ECG, a custom algorithm was used to obtain the beat-to-beat interval time series (*BBI*) by the calculation of the time between two consecutive R peaks. The BP signal was used to extract the diastolic (*DBP*) and systolic blood pressure (*SBP*) time series by the calculation of the minimum and maximum values of the BP in each heartbeat, respectively (Fig. 2.). The time series were then resampled to 1 Hz. Ventricular premature beats and artifacts were interpolated using an adaptive variance estimation algorithm.

TABLE I. CLINICAL INFORMATION

	<i>ICM</i>	<i>DCM</i>
Patients	25	13
Age [years]	65 ± 10.3	62 ± 11.5
Weight [kg]	78 ± 14.8	75 ± 17.4
BMI	28 ± 3.9	28 ± 6.1
NYHA	2 ± 0.4	2 ± 0.3
LVDD [mm]	59 ± 7.6	62 ± 4
AD	45 ± 6.6	45 ± 3.9
ProBNP	1530 ± 1389.2	2290 ± 4407.5
LVEF [%]	35 ± 11.5	34 ± 7.2

ICM = Ischemic Cardiomyopathy; DCM = Dilated Cardiomyopathy; BMI = Body Mass Index; NYHA = New York Heart Association functional classification; LVDD = Left Ventricular Diastolic Dysfunction; AD = Auricular Diameter; ProBNP = Brain Natriuretic Peptide; LVEF = Left Ventricular Ejection Fraction

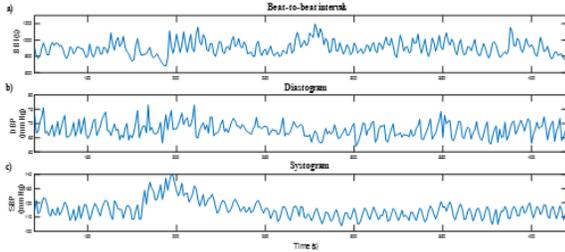


Figure 2. Example of a) *BBI*, b) *DBP*, and c) *SBP* time series

B. Characterization

High resolution joint symbolic dynamics, segmented Poincaré plot analysis, normalized short-time partial directed coherence, dual sequence method, and different time and frequency domain indices were applied to characterize the *BBI*, *SBP* and *DBP* time series.

– High resolution joint symbolic dynamics

Joint symbolic dynamics (JSD) studies the dynamic of a process by the means of symbols [8]. Given \mathbf{X} as a sample vector that contains the information of the *BBI*, *DBP* and *SBP* time series, defined by (1)

$$\mathbf{X} = \{[x_n^{BBI}, x_n^{SBP}, x_n^{DBP}]^T\}_{n=0,1,\dots,N} \text{ with } \mathbf{X} \in \mathbb{R}. \quad (1)$$

The increments between two consecutive values of the time series are coded as ‘1’ while the decrements and equilibriums are coded as ‘0’.

The high resolution joint symbolic dynamics (HRJSD) uses 3 symbols after setting a threshold l [3]. The increments are then coded as ‘2’, the decrements as ‘0’ and the equilibrium states as ‘1’. Using this rule \mathbf{X} can be transformed into the symbolic vector \mathbf{S} as shown in (2):

$$\mathbf{S} = \{[s_n^{BBI}, s_n^{SBP}, s_n^{DBP}]^T\}_{n=0,1,\dots,N} \text{ with } \mathbf{S} \in \{0,1\}$$

$$s_n^Z = \begin{cases} 0: (x_{n+1}^Z - x_n^Z) \leq 0 \\ 1: (x_{n+1}^Z - x_n^Z) > 0 \end{cases}, \quad (2)$$

being Z each one of the time series (*BBI*, *SBP*, *DBP*). Sequences of symbols were considered words of length 3 transforming the vector \mathbf{S} into the vector matrix \mathbf{W} . The new space is composed of combinations of 27 different possible types of words (from 000 to 222) to a total of 729 indices. The indices were analyzed by their probability of occurrence. All word types were grouped into 8 pattern families transforming the vector matrix \mathbf{W} into a vector matrix family (\mathbf{Wf}). These indices were grouped and analyzed according to their family description (Table II).

TABLE II. DESCRIPTION OF PATTERN FAMILIES IN HRJSD METHOD

Family	Description
E0	No variation of 3 successive ‘0’ symbols (‘000’)
E1	No variation of 3 successive ‘1’ symbols (‘111’)
E2	No variation of 3 successive ‘2’ symbols (‘222’)
LU1	Low increasing behavior (‘122’, ‘022’, ‘112’, ‘221’, ‘220’, ‘211’, ‘121’, ‘212’)
LD1	Low decreasing behavior (‘011’, ‘001’, ‘002’, ‘110’, ‘100’, ‘200’, ‘010’, ‘101’)
LA1	Fast alternant behavior (‘020’, ‘202’)
P	Alternant peak-like behavior (‘120’, ‘201’, ‘210’)
V	Alternant valley-like behavior (‘021’, ‘102’, ‘012’)

– Segmented Poincaré plot analysis

Poincaré plot analysis (PPA) is a useful technique to analyze the self-similarity of a system by plotting the data to a higher dimensional state space. The graph is obtained by plotting the present values of the temporal series against the next consecutive values of the series. The typical representation is an elongated scatterplot with all the points whose values are near the mean placed towards the center.

The indices of long- and short- term variability can be evaluated by calculating the standard deviation across the perpendicular axis (SD1) and the longitudinal direction (SD2), respectively [6]. SD1 and SD2 are given by (3)

$$SD1 = \sqrt{\text{Var}\left(\frac{X(n) - X(n+1)}{\sqrt{2}}\right)}; \quad SD2 = \sqrt{\text{Var}\left(\frac{X(n) + X(n+1)}{\sqrt{2}}\right)}. \quad (3)$$

Segmented Poincaré plot analysis (SPPA) is an enhanced version of the PPA obtained by rotating the scatter points α degrees ($\alpha = 45^\circ$) around the main focus plot, defined by (4)

$$\begin{bmatrix} X'(n) \\ X'(n+1) \\ z' \end{bmatrix} = \begin{bmatrix} \bar{X}(n) \\ \bar{X}(n+1) \\ z \end{bmatrix} + \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} X(n) - \bar{X}(n) \\ X(n+1) - \bar{X}(n+1) \\ z \end{bmatrix}. \quad (4)$$

The x and y axes represent the $X'(n)$ and $X'(n+1)$ values rotated by the z' axis. Afterwards, a 12 x 12 rectangular grid is drawn on the plot [9]. The size of the rectangles is adapted as a function of the SD1 (row) and SD2 (column) values. The plot is analyzed considering the probability of appearance of the points in each rectangle normalized by their total number.

– Normalized short-time partial directed coherence

The directed coherence method allows the evaluation of whether two systems are functionally connected. The assessment of the relative structural relationships between the

systems by decomposing their interactions into “feedback” and “feedforward” aspects. The normalized short-time partial directed coherence (NSTPDC) determines the either direct or indirect causality between the systems, detecting levels and directions in multivariate and complex dynamic systems [10].

If X and Y are two time series, the coupling factor (CF) describes the strength and direction of a coupling formed by both series. CF is obtained by evaluating the relation between the mean values of Y coupled with X , and vice versa. This factor is then normalized to a set of values according to a normalization factor (NF), defined by (5)

$$NF = \{-2, -1, 0, 1, 2\}. \quad (5)$$

If NF takes values -2 or 2 (-2 denotes Y as the driver and X as the target variable) the coupling is considered strongly unidirectional, values equal to -1 or 1 indicate bidirectional coupling (similar influence in both directions), and $NF = 0$ corresponds to no coupling.

– Dual sequence method

Dual sequence method (DSM) permits the estimation of the spontaneous baroreflex sensitivity. A fluctuation is defined considering variations in $BBI \geq 5\text{ms}$, $DBP \geq 1\text{ mmHg}$ and $SBP \geq 1\text{ mmHg}$. The highest slope of every sequence is taken for linear regression. The slopes of the regression lines between SBP and BBI sequences were taken as an index for local spontaneous baroreflex sensitivity and calculated for every subject analyzed.

The DSM analyses two different types of BBI response:

- Bradycardic fluctuations (bslope): when an increase in SBP causes an increase in BBI .
- Tachycardic fluctuations (tslope): when a decrease in SBP causes a decrease in BBI .

Additionally, several time and frequency domain indices were calculated from the time series.

The cardiovascular activity was characterized considering three different couplings:

- The coupling formed between the BBI and DBP (cd)
- The coupling formed between BBI and SBP (cs)
- The coupling formed between DBP and SBP (ds)

Table III shows the description of these indices.

C. Classification

The Support vector machines (SVM) algorithm was used to build classification models. The objective of the method is to find a higher dimensional space where the classes can be more easily separated than in the original space. The vectors that define the hyperplane are called support vectors. This method allows classification in data where it is not originally possible using linear classifiers [11].

If $X = \{x_1, \dots, x_L\}$, $x \in \mathbb{R}$ is a given set of data vectors and $Y = \{y_1, \dots, y_L\}$ their corresponding labels. The SVM function, defined as a linear discriminant function, known as hyperplane, is given by (6)

$$f(x) = wz + b = \sum_i^L \alpha_i y_i K_\sigma(x_i y_i) + b, \quad (6)$$

where w is the normal vector of the hyperplane, and $K_\sigma(x_i y_i)$ the kernel function that shapes the hyperplane allowing its optimization. The indices α_i and b defines the efficiency of

the classifier on optimal values, and σ is a penalization term. The kernels considered were the Gaussian, Laplace and ANOVA [6]. The model was then optimized by maximizing the margin while minimizing the training error.

TABLE III. INDICES EXTRACTED AND THEIR DESCRIPTION

Index	Description
HRJSD _{xy} _F _x -F _y	Probability of occurrence of the F _x and F _y word families from the x-y coupling
SPPA _{xy} _row_n-m	Probability of occurrence of the Row n-m from the x-y coupling
SPPA _{xy} _col_n-m	Probability of occurrence of the Column n-m from the x-y coupling
NSTPDC _{xy} _NF	Normalized coupling factor from the x-y coupling
NSTPDC _{xy} _A _x → _y	Coupling strength of x-y from x to y
NSTPDC _{xy} _A _y → _x	Coupling strength of x-y from y to x
bslope	Bradycardic fluctuations
tslope	Tachycardic fluctuations
mean_Sx	Mean value of the S _x time series
sd_Sx	Standard deviation of the S _x time series
meanNN_Sx	Mean value of the S _x normalized time series
sdNN_Sx	Standard deviation of the normalized S _x time series

x and y represent the couplings cd , cs and ds ; F represents the word families: E0, E1, E2, LU1, LD1, LA1, P and V; n and m represent the number of row and column in the PPA based indices, respectively; Sx represents the time series: BBI , SBP and DBP

Using Lagrange multipliers method, a dual formulation can be obtained by [12] (7)

$$\min P(w, b) = \frac{1}{2} \|w_m z\|^2 + C \sum_i K_1 [y_i f(x_i)], \quad (7)$$

where C is a penalty index. When the value of C increases the penalty of the errors is stronger and the decision boundary narrows [12]. The features were scaled and normalized (zero mean and unit variance) in order to avoid biases.

D. Statistical Analysis

A Mann-Whitney non-parametric statistical test was used to evaluate the statistical significance of the indices obtained in the characterization step. An index with a p -value ≤ 0.05 was considered statistically significant. Afterwards, a correlation analysis was performed, discarding indices with high correlation ($\rho > 0.7$) and relative lower significance. The leave-one-out cross-validation method was used to validate the results. The classification results are presented in terms of accuracy (Acc), sensitivity (Sn) and specificity (Sp).

IV. RESULTS

In the characterization step, 298 indices were extracted. To evaluate the results, the following comparison were analyzed:

- Ischemic versus dilated cardiomyopathy patients (ICM vs DCM)
- Cardiomyopathy patients versus control group (CPM vs CON)

When the ICM and DCM patients were compared, six indices presented statistically significant differences. The correlation analysis shown that two of them were correlated with higher significant indices and were discarded. A total of four indices were chosen to build the SVM models.

When the CMP and CON groups were compared ninety-seven indices shown statistical significance. Afterwards, the correlation analysis, forty-seven of these indices were

discarded. Therefore, fifty indices were used during the classification process. According to the results, the most relevant indices for each comparison are shown in tables IV and V, expressed as mean and standard deviation values.

TABLE IV. SIGNIFICANT INDICES FOR ICM VS DCM

Index	ICM (25)	DCM (13)	p-value
SPPAcs_row5	12.8 ± 2	10.7 ± 3.6	0.01
HRJSDcd_E2-E1	0.011 ± 0.01	0.013 ± 0.03	0.03
HRJSDcs_P-V	0.013 ± 0.01	0.016 ± 0.009	0.04
HRJSDcs_V-LA1	0.007 ± 0.007	0.013 ± 0.009	0.02

TABLE V. SIGNIFICANT INDICES FOR CPM VS CON

Index	CMP (38)	CON (30)	p-value
SPPAds_row4	1.8 ± 0.8	2.4 ± 1.2	0.02
SPPAcs_row5	12 ± 2.8	13.2 ± 2.3	0.03
meanNN_SBP	48.8 ± 8.9	86.5 ± 13.8	<0.0001
tslope (ms/mmHg)	8.6 ± 5.1	10.6 ± 5.1	0.04

The HRJSDcd_E2-E1 and HRJSDcs_P-V indices were the optimal choices for the ICM vs DCM SVM classification achieving 84.2% accuracy, 76.9% sensitivity and 88% specificity. The best SVM model for the CMP vs CON comparison was built using the SPPAds_row4 and meanNN_SBP indices, obtaining a 95.5% accuracy, 97.3% sensitivity and 93.3% specificity. The best model was obtained using Laplace kernel in both comparisons. The classification plot and results are shown in Fig. 3 and table VI, respectively.

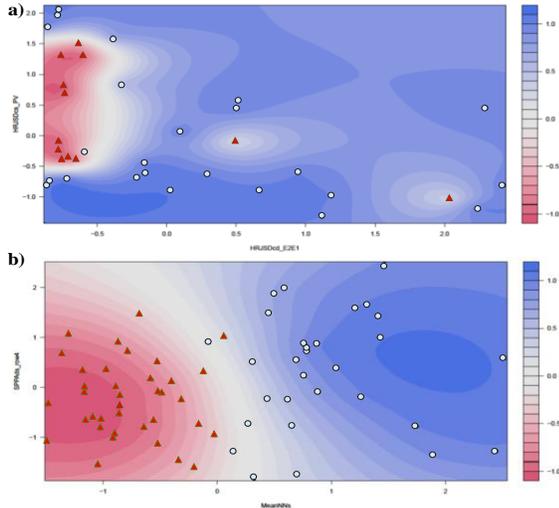


Figure 3. SVM classification results for, a) ICM (○) vs DCM (▲) patients, and b) CMP (○) patients vs CON (▲) subjects

V. DISCUSSION AND CONCLUSION

The cardiovascular regulatory mechanisms of the patients studied were characterized using coupling analysis. The statistically significant indices found were used to classify the patients according to their etiology.

TABLE VI. SIGNIFICANT INDICES FOR ICM VS DCM

Groups	C	σ	Acc (%)	Sn (%)	Sp (%)
ICM vs DCM	2	1	84.2	76.9	88
CMP vs CON	2.5	1	95.5	97.3	93.3

The results shown that ICM patients exhibits lower short-term variability compared to DCM patients. A similar behavior was observed when the patients were compared to the CON group. Additionally, changes in both, the systolic and diastolic activity, translates in changes in heart rate less frequently in ICM patients. The aforementioned suggests that the baroreflex mechanism behaves differently in ICM and DCM patients. The tslope index was significantly lower in CMP patients compared to the CON group, lower values are typically associated with baroreflex dysfunction [13].

In conclusion, the analysis of the cardiovascular coupling allows to find indices that differentiate cardiomyopathy patients by their type of pathology. However, these results should be validated with a dataset of a greater size.

REFERENCES

- [1] H. Fujita, Acharya, R., Sudarshan, V., Ghista, D., Sree, V., Lim Wei Jie, L. and Koh, J., "Sudden cardiac death (SDC) prediction based on nonlinear heart rate variability features and SCD index," *Applied Soft Computing*, vol. 43, pp. 210-519, 2016.
- [2] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal, "Methods derived from nonlinear dynamics for analysing heart rate variability," *Philos Trans A Math Phys Eng Sci*, vol. 367, no. 1887, pp. 277-96, Jan 28 2009.
- [3] S. Schulz, N. Tupaika, S. Berger, J. Haueisen, K. J. Bar, and A. Voss, "Cardiovascular coupling analysis with high-resolution joint symbolic dynamics in patients suffering from acute schizophrenia," *Physiol Meas*, vol. 34, no. 8, pp. 883-901, Aug 2013.
- [4] K. Niizeki and T. Saitoh, "Analysis of cardiorespiratory phase coupling and cardiovascular autonomic responses during food ingestion," *Physiol Behav*, vol. 159, pp. 1-13, May 15 2016.
- [5] A. Arcentales, A. Voss, P. Caminal, A. Bayés-Genís, M. T. Domingo, and B. F. Giraldo, "Characterization of patients with different ventricular ejection fractions using blood pressure signal analysis," in *Computing in Cardiology 2013*, 2013, pp. 795-798.
- [6] J. Rodriguez, A. Voss, P. Caminal, A. Bayes-Genis, and B. F. Giraldo, "Characterization and classification of patients with different levels of cardiac death risk by using Poincare plot analysis," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2017, pp. 1332-1335, Jul 2017.
- [7] B. F. Giraldo, J. Rodriguez, P. Caminal, A. Bayes-Genis, and A. Voss, "Cardiorespiratory and cardiovascular interactions in cardiomyopathy patients using joint symbolic dynamic analysis," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2015, pp. 306-9, 2015.
- [8] M. Baumert, V. Baier, S. Truebner, A. Schirdewan, and A. Voss, "Short- and long-term joint symbolic dynamics of heart rate and blood pressure in dilated cardiomyopathy," *IEEE Trans Biomed Eng*, vol. 52, no. 12, pp. 2112-5, Dec 2005.
- [9] A. Voss, C. Fischer, R. Schroeder, H. R. Figulla, and M. Goernig, "Segmented Poincare plot analysis for risk stratification in patients with dilated cardiomyopathy," *Methods Inf Med*, vol. 49, no. 5, pp. 511-5, 2010.
- [10] S. Schulz, Bär, K., and Voss, A., "Analyses of heart rate, respiration and cardiorespiratory coupling in patients with schizophrenia," *Entropy*, vol. 17, pp. 483-501, 2015.
- [11] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Scholkopf, and G. Ratsch, "Support vector machines and kernels for computational biology," *PLoS Comput Biol*, vol. 4, no. 10, p. e1000173, Oct 2008.
- [12] C. a. V. Cortes, V., "Support vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [13] M. T. La Rovere, G. D. Pinna, and G. Raczak, "Baroreflex sensitivity: measurement and clinical implications," *Ann Noninvasive Electrocardiol*, vol. 13, no. 2, pp. 191-207, Apr 2008.