33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Semiparametric Detection of Nonlinear Causal Coupling Using Partial Directed Coherence

Lucas Massaroppe Luiz Antonio Baccalá Koichi Sameshima

Abstract— **Infering causal relationships from observed time series has attracted much recent attention. In cases of nonlinear coupling, adequate inference is often hindered by the need to specify coupling details that call for many parameters and global minimization of nonconvex functions. In this paper we use an example to investigate a new concept, termed here *running entropy mapping*, whereby time series are mapped onto other entropy related time sequences whose analysis via a linear parametric time series methods, such as partial directed coherence, is able to expose the presence of formerly linearly undetectable causal relationships.**

Keywords:Approximate Entropy, Sample Entropy, Granger Causality, Partial Directed Coherence

I. INTRODUCTION

Specially in neuroscience [1], [2], [3], [4], [5], [6], [7], [8], but also in other biomedical applications [9], [10], [11], much recent attention has been paid to methods for infering the relationship between observations that evolve in time. This endeavour has become known as the study of 'connectivity' and is now seen as a pre-requisite for elucidating the brain's inner workings. The reason behind this interest is furthered by the fact that these techniques enable constructing plausible causal explanations for the time evolution of observations in connection to brain states while avoiding often invasive and possibly harmful direct intervention procedures.

Many currently promising techniques somehow ultimately rely explicitly or not on the idea of Granger causality [12]. The reason for this interest, in addition, is the possibility of providing precise measures of information flow [13], [14] which includes the interaction direction as opposed to mere correlation based methods [15].

To date, possibly thanks to their well understood convergence properties, the most sucessfull techniques employ adequately fitted linear multivariate models to simultaneously acquired time series data [16], [17]. It is noteworthy that such techniques have even proved sucessful in detecting some instances of nonlinear interactions given sufficiently high model order and lengthy observations [18].

However, in some cases such as for quadratic coupling (see Sec. III below), linear model approximations fail. Whereas alternatives obviously exist, both nonparametric [19], [20], [21], [22] and parametric [10], [23], they often require many observations for converging. In the parametric case, additional difficulties arise from reliance on 'ad hoc' structural assumptions and from the usually non convex nature of the functionals employed to obtain parameter estimates.

In this paper, to capture the presence of nonlinear interactions whose presence is linearly undetectable, we investigate a new hybrid approach we term *running entropy mapping*. The main idea is to compute a time dependent measure of a time series's complexity over a suitably long running window. This produces an allied time series that portrays how its complexity evolves in time. The next step consists of comparing the resulting mapped time series among themselves via linear multivariate methods.

The rest of this paper is organized as follows: Sec. II describes two entropy running measures and their computation and briefly recaps *partial directed coherence* (PDC) [3] whose use is made in Sec. III to illustrate the effectiveness of the proposal for a simple model. This is followed by a brief discussion and conclusions in Sec. IV.

II. THE METHOD

There are two steps to the method: (a) entropy mapping and (b) linear analysis of the resulting mapped series. Whereas many alternatives exist for the second step, here for definiteness we employ *partial directed coherence* (PDC).

A. Running Entropy Mapping

Consider a time series $x_i(n)$ comprising N sequential observations. Associate it to another time series $\xi_i(n)$ sequence generated from a sliding window $x(n-W+1),\ldots,x(n)$ and constructed so as to reflect some measure of the original time series complexity.

In this paper we examine two such measures: (a) Pincus's *approximate entropy* [24], [25], [26] and (b) Lake's et al. [27] bias corrected *sample entropy* which justifies the *running entropy mapping* terminology adopted here.

In addition to the window length, W, the latter entropies require defining an embedding dimension m and a radius r and consist of counting the odds of sample m length packets in the series that are close to each given such packet to within a distance r.

Whereas the proposal of the latter filtering of $x_i(n)$ is a quite general one, the choice of the latter entropy measures is justified by their fast convergence in terms of number of observed points as $W << N$ and their reasonable reported immunity to the presence of additive noise [28]. An added advantage is their asymptotic gaussian behaviour consistent with the statistical tests [29] adopted herein.

Lucas Massaroppe and Luiz Antonio Baccalá are with Escola Politécnica, Department of Telecommunications and Control Engineering, University of São Paulo, Av. Prof. Luciano Gualberto, travessa 3, São Paulo, Brazil, 05508 – 900. baccala@lcs.poli.usp.br, lucasmassaroppe@gmail.com

Koichi Sameshima is with Faculdade de Medicina, Department of Radiology, University of São Paulo, Av. Dr. Enéas de Carvalho Aguiar, 255, São Paulo, Brazil, 01246 – 903. ksameshi@usp.br

The next step consists of applying causality analysis relating mapped $\xi_i(n)$ time series among themselves rather than the original $x_i(n)$ series.

B. PDC Analysis

Partial directed coherence was introduced [3] as a means of exposing linear Granger causal relationships in the frequency domain. When relating K simultaneously observed time series PDC is given by:

$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\sum_{l=1}^{K} |\bar{A}_{lj}(f)|^2}}. \tag{1}$$

where

$$\bar{A}_{ij}(f) = \begin{cases} 1 - \sum_{l=1}^{p} a_{ij}(l) e^{-\mathbf{j}2\pi f l}, & \text{if } i = j \\ -\sum_{l=1}^{p} a_{ij}(l) e^{-\mathbf{j}2\pi f l}, & \text{otherwise} \end{cases} \tag{2}$$

for $\mathbf{j} = \sqrt{-1}$ and where $a_{ij}(l)$ are the coefficients of an adequately fit multivaritate autoregressive model which in the present proposal relates the associated entropy time series rather than the original observations.

III. SIMULATION RESULTS

To examine the proposed approach, consider the following model describing a linear stochastically fed oscillator

$$\begin{cases} x_1(n) = 2R\cos(.2\pi)x_1(n-1) - R^2 x_1(n-2) + w_1(n), \\ x_2(n) = -.9x_2(n-1) + \beta x_1^2(n-1) + w_2(n), \end{cases} \tag{3}$$

that is quadratically connected to a low pass system filter whose connectivity strength is gauged through β. Both $w_i(n)$ were taken as gaussian zero mean mutually uncorrelated white driving processes. The simulations used a sharp resonance, i.e. $R = .99$.

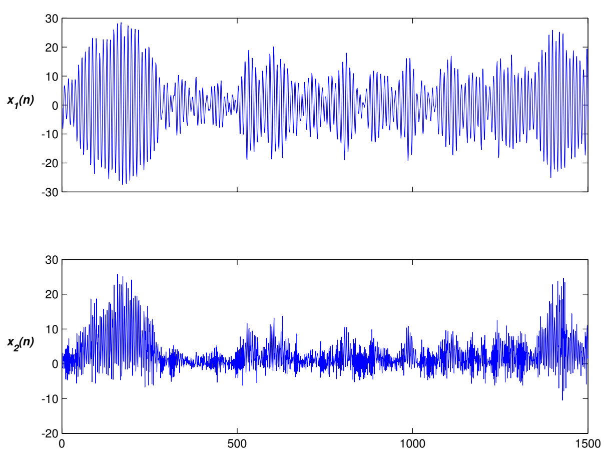

Fig. 1. A realization of **(a)** $x_1(n)$ and **(b)** $x_2(n)$ from the model in Eq. 3 comprizing 1,500 time samples ($R = .99$ and $\beta = .05$).

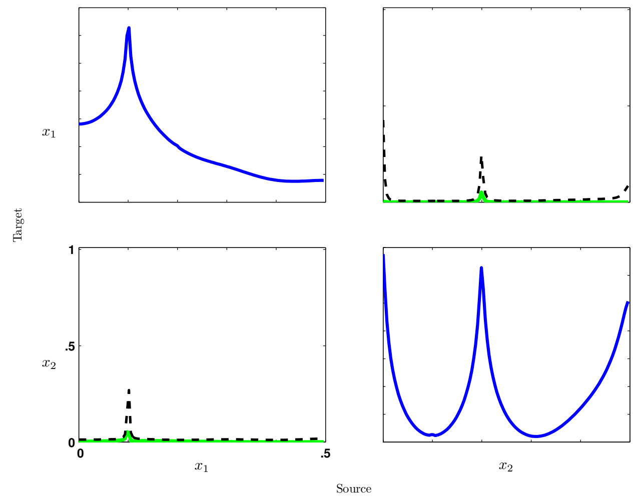

Fig. 2. The computed PDC, in standard form (see text), for the data in Fig. 1 showing that the existing nonlinear influence from $x_1(n)$ to $x_2(n)$ is not captured.

A sample run (1,500 data points) of such vector process is shown on Fig. 1 whose model led to the PDC portrayed in Fig. 2 where no causality can be detected at 5% as the estimates are below the dashed line threshold [29]. In Fig. 2, the usual matrix convention [3] of portraying PDC is adopted. The graphs along the main diagonal represent the series power spectral (arbitrary unit log scale) whereas the counter diagonal portrays PDCs, i.e. the spectral connectivity representations where the bottom left graph corresponds to the $x_1(n) \to x_2(n)$ connection and the upper right graph to $x_2(n) \to x_1(n)$. Similar conventions apply to all PDC graphs used herein.

A. Approximate Entropy

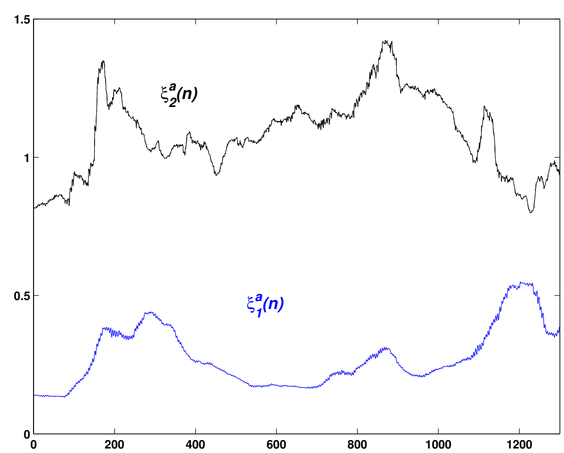

Fig. 3. Approximate running entropy series computed for the data in Fig. 1 using $W = 150$, $m = 1$ and $r = .15$.

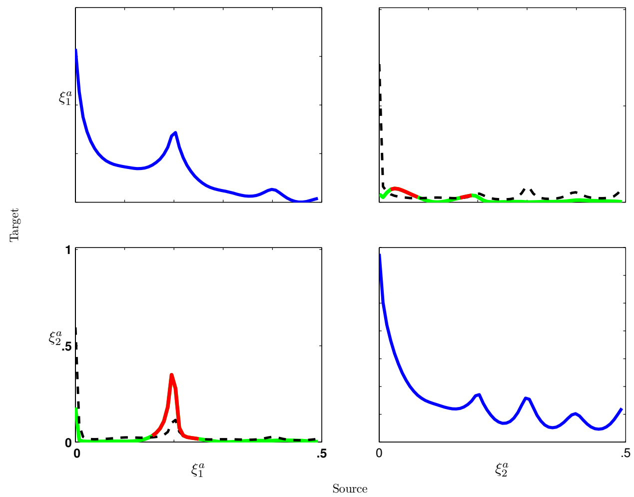

Fig. 4. PDC results for the traces in Fig. 3 in standard form as described in the text showing correctly inferred directionality, i.e. $PDC_{1\to 2}$ (left bottom graph) is above the dashed threshold, whereas $PDC_{2\to 1}$ is mostly below threshold.

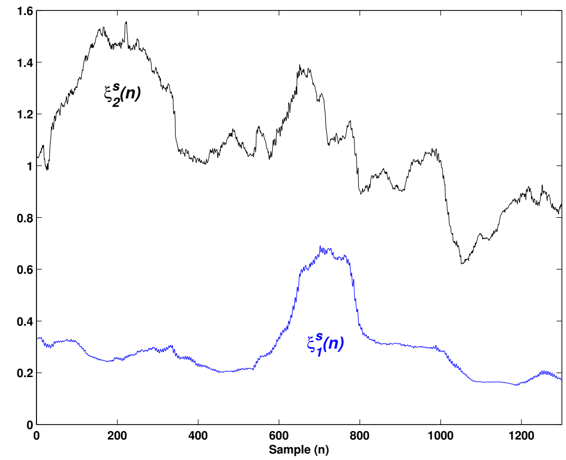

Fig. 5. Running sample entropy reconstruction from the data in Fig. 1.

Reconstruction of the allied approximate entropy time series $[\xi_1^a(n)\ \xi_2^a(n)]^T$ (using $W = 150, r = 0.15$, $h = 1$ and $m = 1$) is shown on Fig. 3 and its associated computed PDC on Fig. 4 where significant $\xi_1^a(n) \to \xi_2^a(n)$ is present above threshold and correctly infers interaction direction whereas no significant interaction happens in the reverse direction.

	TP	FP	TP	FP
$R = 0{,}99$, $h = 1$	$\beta = 0{,}05$		$\beta = 0{,}10$	
$m = 2$, $W = 150$, $r = 0.10$	82.69	24.32	98.64	82.45
$m = 1$, $W = 150$, $r = 0.15$	93.92	2.38	99.98	5.51
$m = 2$, $W = 200$, $r = 0.10$	95.88	45.05	99.87	87.59
$m = 1$, $W = 200$, $r = 0.10$	95.60	3.21	100.00	5.84

TABLE I

APPROXIMATE ENTROPY RESULTS FOR 10,000 TRIALS AS A FUNCTION OF COUPLING STRENGTH (β) AND SPACE RECONSTRUCTION PARAMETERS. THE TP LABEL REFERS TO THE PERCENTAGE OF TIMES $\xi_1^a(n) \to \xi_2^a(n)$ IS CORRECTLY DETECTED WHILE THE FP LABEL REFERS TO THE RATE OF REVERSE INCORRECTLY DETECTED CONNECTIONS ($\xi_2^a(n) \to \xi_1^a(n)$).

To assess method robustness and its dependence on entropy space parameters , 10,000 realizations of the process in Eq. (3) were used in obtaining the values of Table I where the crucial nature of reconstruction parameter choice becomes apparent, the best results happen for $m = 1$, $W = 150$, $h = 1$ and $r = 0.15$ even for small β and are to within the expected 5% test significance.

B. Sample Entropy

Similar results are obtained using the sample entropy (reconstructed $[\xi_1^s(n)\ \xi_2^s(n)]^T$ series on Fig. 5 for $W = 150$, $h = 1$ and $m = 1$) and its allied PDC (Fig. 6) where again connectivity is correctly inferred at 5% (see also Table II).

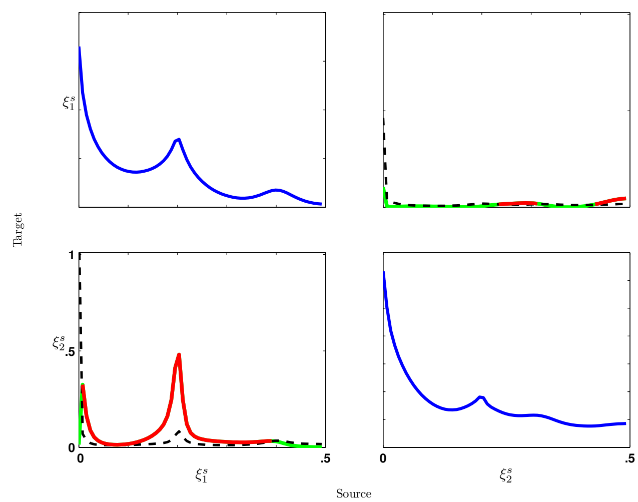

Fig. 6. Standard form PDC (see text) between the traces in Fig. 5 show correct directional connectivity inference.

	TP	FP	TP	FP
$R = 0{,}99$, $h = 1$	$\beta = 0{,}05$		$\beta = 0{,}10$	
$m = 2$, $W = 150$, $r = 0{,}10$	94,67	5,05	99,99	3,37
$m = 1$, $W = 150$, $r = 0{,}15$	93,96	2,36	100,00	11,68
$m = 2$, $W = 200$, $r = 0{,}10$	98,68	9,39	99,97	6,84
$m = 1$, $W = 200$, $r = 0{,}10$	93,90	3,40	100,00	13,25

TABLE II

SAMPLE ENTROPY RESULTS FOR 10,000 TRIALS AS A FUNCTION OF COUPLING STRENGTH (β) AND SPACE RECONSTRUCTION PARAMETERS. THE TP LABEL REFERS TO THE PERCENTAGE OF TIMES $\xi_1^s(n) \to \xi_2^s(n)$ IS CORRECTLY DETECTED WHILE REFERS TO THE RATE OF REVERSE INCORRECTLY DETECTED CONNECTIONS ($\xi_2^s(n) \to \xi_1^s(n)$).

IV. DISCUSSION AND FUTURE WORK

Though in many ways still preliminary, the present results point to the potential of using suitable transformations of time series to still infer causality by fitting linear models between the resulting transformed series in those cases where the causal coupling between the original time series is in principle not even approximately detectable via linear vector autoregressions.

The basic idea presented herein is that of using the fluctuations in entropy measures to gauge how complexity flows from one time series to another. The extensive simulations portray how critical phase space reconstruction is for the process to work properly thus giving rise to the new problem of optimal (W, m, r) parameter choice in the present context.

As perhaps expected, sample entropy proves slightly superior by generating a lower false positive rate.

It is interesting to note that testing many of the mapped running entropy series resulted in the presence of significant cointegration between traces which passed specific Granger causality tests at rates comparable to the ones presented here.

The present study case points to the interest in studying the present methodology further specially in cases of models comprizing larger dimensions. Exploratory investigation of further examples is under way. The study of alternative running maps is also in progress.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the FAPESP Grant 2005/56464-9 (CInAPCe). L.M to FAPESP Grant $2009/04397-7$. L.A.B. to CNPq Grant 304404/2009-8.

REFERENCES

[1] M. Kaminski and K. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological Cybernetics*, vol. 65, pp. 203–210, 1991.

[2] L. A. Baccalá, K. Sameshima, G. Ballester, A. C. do Valle, and C. Timo-Iara, "Studying the interaction between brain structures via directed coherence and granger causality," *Applied Signal Processing*, vol. 5, pp. 40–48, 1998.

[3] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[4] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biological Cybernetics*, vol. 83, pp. 35–45, 2000.

[5] L. Astolfi, F. Cincotti, D. Mattia, M. Marciani, L. Baccala, F. Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. Edgar, G. Miller, B. He, and F. Babiloni, "Comparison of different cortical connectivity estimators for high-resolution eeg recordings," *Human Brain Mapping*, vol. 28, no. 2, pp. 143–157, 2007.

[6] L. Astolfi, F. Cincotti, D. Mattia, M. Marciani, L. Baccala, F. De Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, and F. Babiloni, "Assessing cortical functional connectivity by partial directed coherence: Simulations and application to real data," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 9, pp. 1802–1812, 2006.

[7] K. Blinowska, R. Kus, M. Kaminski, and J. Janiszewska, "Transmission of brain activity during cognitive task," *Brain Topography*, vol. 23, no. 2, pp. 205–213, 2010.

[8] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI." *Hum. Brain. Mapp.*, vol. 30, no. 2, pp. 452–461, 2009.

[9] L. Faes, A. Porta, R. Cucino, S. Cerutti, R. Antolini, and G. Nollo, "Causal transfer function analysis to describe closed loop interactions between cardiovascular and cardiorespiratory variability signals," *Biol. Cybern.*, vol. 90, pp. 390–399, 2004.

[10] L. Faes, G. Nollo, and K. H. Chon, "Linear and nonlinear parametric model identification to assess granger causality in short-term cardiovascular interactions," *Computers in Cardiology*, vol. 35, pp. 793–796, 2008.

[11] A. Fujita, P. Severino, J. R. Sato, and S. Miyano, "Granger causality in systems biology: modeling gene networks in time series microarray data using vector autoregressive models," in *Proceedings of the Advances in bioinformatics and computational biology, and 5th Brazilian conference on Bioinformatics*, ser. BSB'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 13–24.

[12] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[13] D. Y. Takahashi, L. A. Baccalá, and K. Sameshima, "Frequency domain connectivity: an information theoretic perspective," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires*, IEEE. IEEE Press, 2010, pp. 1726–1729.

[14] ——, "Information theoretic interpretation of frequency domain connectivity measures," *Biological Cybernetics*, vol. 103, pp. 463–469, 2010.

[15] L. A. Baccalá and K. Sameshima, "Overcoming the limitations of correlation analysis for many simultaneously processed neural structures," *Progress in Brain Research*, vol. 130, no. Advances in Neural Population Coding, pp. 33–47, 2001.

[16] H. Lütkepohl, *New introduction to multiple time series analysis*. Berlin: Springer, 2005.

[17] A. Schlögl, "A comparison of multivariate autoregressive estimators," *Signal Processing*, vol. 86, no. 9, pp. 2426 – 2429, 2006, special Section: Signal Processing in UWB Communications.

[18] B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, H. Lücking, R. Dahlhaus, and J. Timmer, "Testing for directed influences among neural signals using partial directed coherence," *Journal of Neuroscience Methods*, vol. 152, pp. 210–219, 2005.

[19] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.

[20] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear granger causality," *Physical Review Letters*, vol. 100, no. 14, p. 144103, 2008.

[21] M. Paluš, V. Komárek, Z. Hrncír, and K. Sterbová, "Synchronization as adjustment of information rates: Detection from bivariate time series," *Phys. Rev. E*, vol. 63, no. 4, p. 046211, 2001.

[22] M. Paluš and A. Stefanovska, "Direction of coupling from phases of interacting oscillators: An information-theoretic approach," *Physical Review E*, vol. 67, no. 5, pp. 1–4, 2003.

[23] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, pp. 1–37, 2005.

[24] S. M. Pincus, "Approximate entropy as an irregularity measure for financial data," *Econometric Reviews*, vol. 27, no. 4–6, pp. 329–362, 2008.

[25] ——, "Approximate entropy as a measure of irregularity for psychiatric serial metrics," *Bipolar Disorders*, vol. 8, no. 5p1, pp. 430–440, 2006.

[26] ——, "Approximate entropy (apen) as a complexity measure," *Chaos*, vol. 5, no. 1, pp. 110–117, 1995.

[27] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology — Regulatory, Integrative and Comparative Physiology*, vol. 283, pp. R789–R797, 2002.

[28] S. M. Pincus and W.-M. Huang, "Approximate entropy: Statistical properties and applications," *Communications in Statistics — Theory and Methods*, vol. 21, no. 11, pp. 3061–3077, 1992.

[29] D. Y. Takahashi, L. A. Baccalá, and K. Sameshima, "Connectivity inference via partial directed coherence: asymptotic results," *Journal of Applied Statistics*, vol. 34, pp. 1259–1273, 2007.