



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** de la comunicación de congreso publicada en:
This is an **author produced version** of a paper published in:

Annual International Conference of the IEEE Engineering in Medicine and
Biology Society, 2009. EMBC 2009. IEEE, 2009. 5304 - 5307

DOI: <http://dx.doi.org/10.1109/IEMBS.2009.5333080>

Copyright: © 2009 IEEE

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription

Implicit Wiener Series Analysis of Epileptic Seizure Recordings

Álvaro Barbero^{1*}, Matthias Franz², Wim van Drongelen³,
José R. Dorronsoro¹, Bernhard Schölkopf⁴, Moritz Grosse-Wentrup⁴

Abstract—Implicit Wiener series are a powerful tool to build Volterra representations of time series with any degree of non-linearity. A natural question is then whether higher order representations yield more useful models. In this work we shall study this question for ECoG data channel relationships in epileptic seizure recordings, considering whether quadratic representations yield more accurate classifiers than linear ones. To do so we first show how to derive statistical information on the Volterra coefficient distribution and how to construct seizure classification patterns over that information. As our results illustrate, a quadratic model seems to provide no advantages over a linear one. Nevertheless, we shall also show that the interpretability of the implicit Wiener series provides insights into the inter-channel relationships of the recordings.

I. INTRODUCTION

Epilepsy is a chronic disorder of the brain that affects around 50 million people worldwide, and is characterized by recurrent seizures provoked by excessive electrical discharges in a group of brain cells [1]. Although in some cases patients suffering from epilepsy can be treated with medication, in others neurosurgery needs to be applied, where the region of the brain producing the seizure activity (epileptogenic focus) is removed or separated from the rest of the brain [2]. In order to detect the location of the focus, single photon emission computed tomography (SPECT) can be used if an intravenous injection of a radiopharmaceutical is applied shortly after a seizure onset. However, this is no easy task as usually the seizure is noticed when clinical manifestations appear, which could take place long after the onset of the seizure [3]. A better way of detecting seizure onsets is electroencephalography (EEG) analysis, as alterations in the neuronal activity happen immediately before and during the onset. Unfortunately an experienced electroencephalographer is often required to detect seizure signs in EEG recordings, and so there has been medical interest in developing automatic tools for the analysis and detection of seizure onsets in EEG data.

Previous work in this topic has applied methods from the fields of time series analysis and pattern recognition in an attempt to solve the problem, the former used to obtain relevant features from segments or epochs of raw

EEG data, the latter to classify them as seizure or non-seizure periods. *Webber et al.* [4] extract features as relevant measurements of the raw EEG data and use them as input of a multilayer perceptron, taking its outputs as a discriminative set of features used in a simple rule-based decision method. Another approach from *Gabor et al.* [5] preprocesses the data using a wavelet filter, and then a self-organizing map is used to learn the location of onsets in an unsupervised way, using a vector of values obtained by Fast Fourier Transform (FFT) as explanatory features. A similar method proposed by *Shoeb et al.* [3] applies the wavelet transform to obtain a feature vector and a support vector machine is used as the classification method. The work by *Faul et al.* [6] reviews and evaluates the performance of three seizure detection methods applied in neonatal EEG: analysis of frequency and bandwidth of the main peak in the frequency (FFT) spectrum, analysis of cross-correlations, and examination of the complexity of the EEG data via the use of the Rissanen Minimum Description Length. It is concluded that there is a lack of a reliable detection scheme for clinical use. Finally, in the paper by *Osorio et al.* [7] an algorithm for detecting seizures in electrocorticography (ECoG) data is presented. ECoG is known to provide a better spatial resolution than EEG [8], although it is an invasive method and hence not applied so frequently.

In this work we present a preliminary study on the use of implicit Wiener series to obtain a relevant vector of features from epochs of ECoG data, so as to be able to train a simple classifier with them. The main motivation behind the use of these models is their ability to fit the data to the desired degree of non-linearity, so that the influence of the model complexity in the classification accuracy can be studied. The paper is structured as follows. In Section II we provide a brief introduction to implicit Wiener series. In Section III we present our proposed method. In Section IV we show the results of our experiments with linear and quadratic implicit Wiener series. Finally in Section V we conclude that quadratic models do not provide additional benefits over linear ones for the task of classification, and that the inherent interpretability of these models might prove to be useful to gain insights of the underlying inter-channel relationships.

II. IMPLICIT WIENER SERIES

Implicit Wiener series are based on the theory of the Volterra series [9]. Suppose that a system receives an input function $x(t)$, which depends on the time t , and outputs another function $y(t)$. Then the system can be described as an operator T which transforms x into y : $y = T[x]$. In its

¹ Universidad Autónoma de Madrid - Instituto de Ingeniería del Conocimiento. Authors have been partially supported by Spain's TIN 2007-66862 and Cátedra UAM-IIC en Modelado y Predicción. The first author is kindly supported by the FPU-MEC grant reference AP2006-02285.

² HTWG Konstanz

³ University of Chicago, Department of Pediatrics.

⁴ Max Planck Institute for Biological Cybernetics.

* Corresponding author: alvaro.barbero@uam.es

simplest form T is supposed to have no memory, i.e. the output value of the system at a certain time t_0 is only a function of its input value: $y(t_0) = f(x(t_0))$. If that is the case the Taylor series can be used to represent T . However, if the system does have memory, i.e., the output function y is dependent on past values of x , Volterra series must be used.

The Volterra series representation of a system has the form $y(t) = \sum_{i=0}^{\infty} H_i[x(t)]$, where the H_i are functionals defined as

$$H_n[x(t)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) x(t - \tau_1) \dots x(t - \tau_n) d\tau_1 \dots d\tau_n, \quad (1)$$

with $H_0[x(t)] = h_0$ constant and $h_n(\tau_1, \dots, \tau_n)$ a function that returns 0 if any $\tau_i < 0$. In this way, the Volterra series provide a representation of any possible interaction among input values at different points of time in the past and at any degree of non-linearity. A large and well-understood class of nonlinear systems can be described by using these series. Furthermore, if input and output time series measurements of a system are available, a discrete version of the Volterra functionals can be formulated as

$$H_n[x(t)] = \sum_{i_1=1}^m \dots \sum_{i_n=1}^m h_{n(i_1, \dots, i_n)} x(t)_{i_1} \dots x(t)_{i_n}, \quad (2)$$

where each input pattern $x(t) = (x(t)_1, \dots, x(t)_m)$ represents a set of input values previous to time t . Note that the memory of the approximation is limited to m past values. Here the system can be described by finding the values of the $h_{n(i_1, \dots, i_n)}$ (which are the so called Volterra coefficients) through the minimization of the error obtained by a p degree approximation $\hat{y}(t) = \sum_{i=0}^p H_i[x(t)]$ and the output value $y(t)$ at time t .

Unfortunately, the computation of the h_n is complex. However, an alternative representation known as the Wiener series can alleviate this problem. The Wiener series are a particular case of the Volterra series in which the functionals are forced to be orthogonal and the input is assumed to be white Gaussian noise. In this case it can be shown [9] that there is a procedure, known as the cross-correlation method, which finds a Volterra representation that produces the best approximation in the least squares sense.

An alternative approach to compute these coefficients is presented in [10] and [11] as the implicit Wiener series. It is based on the observation that a Volterra functional can be expressed as $H_n[x] = \eta_n \cdot \phi_n(x)$, where η_n is a vector containing the $h_{n(i_1, \dots, i_n)}$ coefficients and $\phi_n(x)$ is another one including all the possible products of input values from x of degree n : $\phi_0(x) = 1$, $\phi_1(x) = (x_1, \dots, x_m)$, $\phi_2(x) = (x_1^2, x_1x_2, x_2x_1, x_2^2, x_1x_3, \dots, x_m^2)$, ... and so on. In this way the whole Volterra series approximation up to degree p can be written as

$$\sum_{i=0}^p H_i[x(t)] = \eta^{(p)} \cdot \phi^{(p)}(x), \quad (3)$$

where $\eta^{(p)}$ is a vector in which all the η_n up to degree p have been stacked, and $\phi^{(p)}(x)$ a vector in which all the $\phi_n(x)$ up to degree p are contained. Therefore, finding the Volterra coefficients is equivalent to solving a least squares regression problem in the feature space induced by $\phi^{(p)}(x)$, which can be easily done in a kernel setting by defining the kernel function

$$K(x, x')^{(p)} = \phi^{(p)}(x) \cdot \phi^{(p)}(x') = \sum_{n=0}^p (x^T x')^n \quad (4)$$

as shown in [10], and applying one of the available kernel methods for least squares regression. Furthermore, this kind of regression guarantees an optimal solution when the input data is Gaussian, which is one of the base assumptions for the Wiener series. Hence, the obtained coefficients provide in fact the Wiener series representation of the system. Apart from being simpler, this method has been shown to provide better estimations of the coefficients than the cross-correlation method [10].

III. EXTRACTING DISCRIMINATIVE FEATURES

Implicit Wiener series are designed to provide a characterization of the input-output behaviour of a system, and hence they are not suitable for a direct application to the classification of a given ECoG recording segment as seizure or non-seizure. On the other hand, if we believe channel activity during a seizure onset to be noticeably different from the one observed during non-seizure periods, we might reasonably expect to find different implicit Wiener series in seizure and non-seizure situations and, hence, different Volterra coefficients too; moreover, the Volterra coefficients may turn out to be discriminative features of a patient's state. However, we must find first some way to turn coefficient values into classification patterns; we discuss next a way to do so.

ECoG data are generally presented as an ensemble of time series where each one represents the brain activity of one electrode or channel. In order to apply implicit Wiener series, we consider pairwise relationships among channels, that is, we build models taking the recordings from one channel as input signal and the recordings from another one as output signal. In other words, we will build first temporal patterns of the form $[X_{t-K-D+1} \dots X_{t-K}; Y_t]$ with D being the amount of memory considered for the approximation and X and Y the input and output channels respectively. Note that a delay K is introduced in the output values as the activity from the input channel should take some time to propagate to the output one.

Once the whole recordings have been transformed into multidimensional temporal patterns, if we consider a recording interval containing L such patterns (we implicitly assume stationarity in the interval), we can build the implicit Wiener series by solving a least squares regression problem on this interval. In our experiments we have done so using a Gaussian Process [12] with an inhomogeneous polynomial kernel $K(x_1, x_2) = (1 + x_1 \cdot x_2)^p$, which is a special case

of (4) [11]. We have built the models following a two step strategy in which we first fit the model parameters using Geisser’s surrogate predictive probability method [12] and then fine-tune the model by means of leave-one-out Mean Squared Error estimates.

If we compute the Wiener series over different intervals, it turns out that the Volterra coefficients are highly dependent on the concrete recording segment used. This suggests that statistical information on the coefficient structure may be useful for discrimination purposes. To get a coefficient sample, we first separate the recorded data into different large epochs which we label as ictal or non-ictal according to the state of the patient when the epoch was recorded. We consider for each epoch E_i a number of consecutive and partially overlapping intervals upon which we build concrete Wiener series for selected pairs of nodes; these models will capture the effect of one node on the other. More precisely, we work with intervals I_t of length l and build the next interval I_{t+1} by removing the first s samples from I_t and adding s new ones at its end (see algorithm 1 for more details). In this way we get for epoch E_i a number N_i of Volterra coefficients samples upon which we compute their mean, variance, skewness, kurtosis, min and max. Thus, each epoch E_i yields a classification pattern (C_i, y_i) with C_i the statistical parameters computed over E_i and y_i the label associated to the patient’s state during the epoch. We have used these classification patterns in our experiments, which we report next.

Algorithm 1 Volterra coefficients distributions extraction

```

1: Given a preprocessed data  $\mathcal{X}$ ...
2: for each epoch of  $L$  samples in  $\mathcal{X}$  do
3:    $train$  = the first  $l$  samples of  $L$ .
4:   while there are still at least  $l$  samples in  $train$  do
5:     Train the implicit Wiener series for each pairwise relationship
       using  $train$ , and compute their Volterra coefficients, stacking
       them all in  $vol$ .
6:     Add  $vol$  to an array of estimations  $V$ .
7:     Discard the first  $s$  samples of  $train$ , and add to  $train$  the next
        $s$  samples from  $L$  (if any).
8:   end while
9:   Compute the statistical measurements of the distribution of the
       Volterra coefficients by using the estimates in  $V$ , and build a
       classification pattern  $p$  using these as input values and the state of
        $L$  (ictal or non-ictal) as output value (class).
10:  Add  $p$  to a pattern matrix  $P$ .
11: end for
12: Return  $P$ .
```

IV. EXPERIMENTAL RESULTS

As mentioned before, one of the advantages of the implicit Wiener series is that we can obtain higher order, non-linear Volterra representations by only changing the degree of the polynomial kernel used. In this section we study whether higher order representations can better capture channel interaction. A simple way of putting this to test is to check whether seizure classifiers built over higher order models have a better performance.

We used intracranially recorded seizure data from the Pediatric Epilepsy Center at The University of Chicago. They

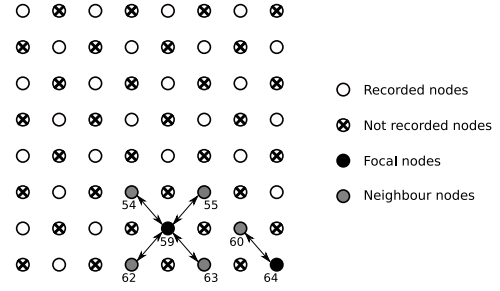


Fig. 1. Layout of the right frontal grid electrodes, containing the two focal nodes. The relationships modelled involved those focal nodes and their neighbours, and are represented as directed arrows. The numbers under the nodes stand for their corresponding channel in the data provided.

consist of 15 hours and 20 minutes of continuous EEG and ECoG recordings of a single patient suffering 7 seizure onsets. It is known that there are two focal nodes, i.e., nodes originating the seizures, and that they are both located in two of the ECoG nodes belonging to the right frontal grid (RFG), namely those numbered 59 and 64. A layout of this grid is depicted in Fig. 1. In our experiments we only analysed ECoG channels that correspond to focal nodes or nodes adjacent to them. As it can be seen in the figure, node 59 has four adjacent nodes for which recordings were available, while node 64 has only one adjacent recording node. In other words, there were five 5 focal node neighbours, i.e., 10 node pairs.

A sixth order forward-reverse Butterworth filter letting pass frequencies between 2 Hz and 50 Hz was applied over the whole data to remove recording artifacts. The parameters used for the pattern construction were $D = 10$, $K = 10$. We worked with epochs of $L = 5000$ samples, intervals of length $l = 100$ and changing $s = 10$ samples when moving from one interval to the next. This resulted in 2292 classification patterns, only 85 of them belonging to seizure onsets. Although under-sampling could have tackled this class imbalance, we have used the whole dataset to avoid an increase in the number of false positives in our results, as in a real scenario the same class ratio would be present.

Many methods are available to perform the classification task; we shall work here with CART trees [13] as they are fairly simple and easily provide an interpretation of the resulting model. Moreover, since a CART tree automatically selects a small subset of explanatory features that it finds representative, we may expect that the variables selected will be those which show larger differences in the Volterra coefficient distribution of seizure and pre-seizure states. We can exploit this information by identifying the electrode pairs associated to the selected coefficients and determining which relationships have significantly different behavior in seizure and non-seizure states.

In an initial exploratory experiment we worked with linear polynomials using all the data available from the first seizure onset to the last one. The resulting CART tree (shown in Fig. 2) is quite small, containing only two variables which correspond to the mean of a Volterra coefficient linearly

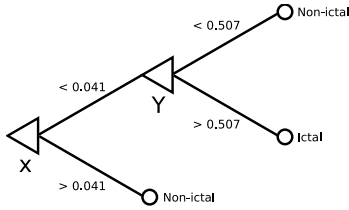


Fig. 2. A depiction of the CART tree model obtained when using the whole data as training set. Only two variables, denoted here as X and Y , are used, which correspond to the mean and standard deviation of two Volterra coefficients representing particular relationships between nodes.

linking node 63 to node 59 (neighbour to focal) and to the standard deviation of a coefficient linking node 59 to 62 (focal to neighbour). Applying this classifier over the whole dataset results in a 1.55 % misclassification rate, with 0 % false positives and 1.40 % false negatives. Notice, however, that the same data have been used for training and testing, so these accuracies may not be representative of a general performance. On the other hand, the two extracted features clearly have discriminative power and the relationship between node 59 and nodes 62 and 63 seems to greatly depend on a patient's state.

To obtain a more significative accuracy measure, we built 100 training-test subset pairs (each pair containing the same number of patterns as the original dataset) by bootstrap with resampling, obtaining now an average 2.21 % test misclassification rate, with 0.71 % false positives and 1.47 % false negatives. This is clearly better than the 3.71 % *a priori* distribution of seizures. Another interesting fact is that, again, the two most frequent variables selected by the CART procedure over the 100 training subsets are precisely the ones obtained when using the whole dataset for training. This seems to imply some robustness in our method. Moreover, and while these error rates may not be competitive against state-of-the-art seizure classifiers, they show that the statistical information they are built upon clearly has discriminative power and can, therefore, be used to check whether quadratic channel relationships may provide a better explanation of the underlying activity. To do so, we repeated the previous experiments using now quadratic kernels and an augmented interval length $l = 200$. However, working with the same 100 train-test pairs as before, the misclassification rate is now 2.36 %, with 0.82 % false positives and 1.54 % false negatives. In fact, a Wilcoxon rank sum test shows that the misclassification rates derived from linear kernels are significantly better at the 5 % level. Thus, quadratic Volterra coefficient information does not seem to improve the performance of CART trees over that achieved by linear ones.

V. DISCUSSION AND FUTURE WORK

In this work we have studied how implicit Wiener series can be used to analyze seizure ECoG recordings. In particular, implicit Wiener series makes it quite easy to obtain such representations simply by adjusting the non-linearity degree of the approximation and, in principle, higher order

Volterra representations should capture more information on the underlying data. A simple way of checking whether this is true is to work with classifiers acting on patterns derived from different degree models. We have first shown how such patterns can be derived from statistical information on the Volterra coefficients and performed then experiments with CART trees acting over patterns built from linear and quadratic kernels. In our experiments, a quadratic model does not improve the classification accuracies of a linear one, even though the quadratic CART classifier selects Volterra coefficients that correspond to non-linear node relationships. This might mean that, although non-linear relationships among channels exist, they are not relevant for the task of seizure classification.

On the other hand, another advantage of the application of implicit Wiener series to seizure data lies in their ability to pinpoint the most relevant relationships between the individual recording channels, something for which CART trees are quite useful. These results hint that implicit Wiener series might be useful for tasks such as focal node location or the detection of the most relevant channels involved in the appearance of the seizure. In any case, a more thorough study is needed on the application of our procedure to data from different patients and, also, to determine its applicability when used on EEG data instead of ECoG recordings.

REFERENCES

- [1] W. H. Organization, "Epilepsy: aetiology, epidemiology and prognosis," <http://www.who.int/mediacentre/factsheets/fs165/en/>, February 2001.
- [2] "The national society for epilepsy," <http://www.epilepsynse.org.uk>.
- [3] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, T. Treves, and J. Guttag, "Patient-specific seizure onset detection," in *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, USA, September 2004.
- [4] W. R. S. Webber, R. P. Lesser, R. T. Richardson, and K. Wilson, "An approach to seizure detection using an artificial neural network (ann)," *Electroencephalography and clinical Neurophysiology*, no. 116, pp. 250–272, 1996.
- [5] A. J. Gabor, R. R. Leach, and D. F. U., "Automated seizure detection using a self-organizing neural network," *Electroencephalography and clinical Neurophysiology*, no. 99, pp. 257–266, 1996.
- [6] S. Faul, G. Boylan, S. Connolly, L. Marnane, and G. Lightbody, "An evaluation of automated neonatal seizure detection methods," *Clinical Neurophysiology*, no. 116, pp. 1533–1541, 2005.
- [7] I. Osorio, M. G. Frei, J. Giftakis, T. Peters, J. Ingram, M. Turnbull, M. Herzog, M. T. Rise, S. Schaffner, R. A. Wennberg, T. S. Walczak, M. W. Risinger, and C. Ajmone-Marsan, "Performance reassessment of a real-time seizure-detection algorithm on long ecog series," *Epilepsia*, no. 43(12), pp. 1522–1535, 2002.
- [8] E. Asano, C. Juhasz, A. Shah, O. Muzik, D. C. Chugani, J. Shad, S. Sood, and H. T. Chugani, "Origin and propagation of epileptic spasms delineated on electrocorticography," *Epilepsia*, no. 46, pp. 1086–1097, 2005.
- [9] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems*. Malabar: Krieger, 1980.
- [10] M. O. Franz, "Implicit wiener series," MPI für biologische Kybernetik, Tech. Rep.
- [11] M. Franz and B. Schölkopf, "A unifying view of wiener and volterra theory and polynomial kernel regression," *Neural Computation*, vol. 18, pp. 3097–3118, 2006.
- [12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [13] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, ser. Wadsworth Statistics / Probability series. Wadsworth, 1984.