Channel-Consistent Forewarning of Epileptic Events from Scalp EEG

Lee M. Hively* and Vladimir A. Protopopescu

***Abstract*—Phase-space dissimilarity measures (PSDM) have been recently proposed to provide forewarning of impending epileptic events from scalp electroencephalographic (EEG) for eventual ambulatory settings. Despite high noise in scalp EEG, PSDM yield consistently superior performance over traditional nonlinear indicators, such as Kolmogorov entropy, Lyapunov exponents, and correlation dimension. However, blind application of PSDM may result in channel inconsistency, whereby multiple datasets from the same patient yield conflicting forewarning indications in the same channel. This paper presents a first attempt to solve this problem.**

***Index Terms*—Dynamical systems, epileptic seizure forewarning, nonlinear analysis, phase-space dissimilarity measures.**

I. Introduction

HANS BERGER discovered electrical activity from the cerebrum in 1923 as electroencephalographic (EEG) signals. These signals have several distinct waves of different amplitudes and frequencies that characterize various processes, such as sleep, rest, wakefulness, pathologies, etc. Specific EEG patterns convey a standard of normality, while deviations from this standard indicate abnormality. The interpretation of time-serial EEG data is severely hindered by the lack of an adequate model of the central nervous system that is consistent with these observations. In 1970, Hjorth introduced three parameters (activity, mobility, and completeness) to describe and quantify the EEG signal in the spatio-temporal domain [1]. Soon thereafter, significant advances in nonlinear and chaotic dynamics [2]–[6] (and references therein) suggested a new framework for analyzing and possibly modeling brain activity as a nonlinear dynamical system. The seminal work of Babloyantz *et al.* [7] seemed to indicate that EEG data have deterministic features that are intertwined with noise to produce a wide variety of behaviors, including chaotic behavior. Although a few dissenting opinions have challenged this perspective [8], our working hypothesis for epileptic phenomena is that indeed the brain behaves as a reasonably low-dimensional dynamical system with regimes varying between (quasi-) periodic and completely irregular (chaotic). Thus, interpretation of biomedical (in particular, EEG) data within this framework becomes both legitimate and fruitful, with global aspects of brain dynamics being quantified by traditional nonlinear measures such as Lyapunov exponents, Kolmogorov entropy, and correlation dimension [2]–[6]. While these descriptors discriminate adequately between clear-cut regular and chaotic dynamics, they are not sufficiently sensitive to distinguish between slightly different chaotic regimes, especially when data are limited and/or noisy. Most brain dynamics during the transition from normal activity to the seizure event and postseizure recovery fall into the latter regime, making robust and timely forewarning of epileptic events an outstanding medical challenge.

Epilepsy afflicts about three-million people in the U.S. alone. Epilepsy can be effectively treated in many instances, and many patients are indeed under constant medication. However, constant medication frequently has severe side effects that may be more debilitating than the seizures themselves. Moreover, 10%–30% of the cases cannot be controlled by medication. In addition, some extreme epileptic events require immediate medical intervention to avoid sudden unexplained death in epilepsy (SUDEP), which is characterized by fatal cardiac rhythm disturbances and injuries during seizures. In general though, while most seizures are not life threatening or even serious medical events, they represent an unpredictable source of social nuisance, disruption, and embarrassment. For some patients with a history of only partial seizures, a reliable forewarning device may offer a new treatment paradigm of constant monitoring rather than continuous medication. Such forewarning would allow the patient to interrupt hazardous activity, lie down in a quiet place, undergo the seizure, and then return to normal activity. Other timely preventive steps include taking medication, requesting emergency responders, or contacting the physician. For most patients though, we expect that a seizure forewarning device would more likely be used to complement medications, rather than replacing it.

Successful seizure forewarning requires a detailed examination of the brain dynamics. Le Van Quyen *et al.* [9] review recent research on nonlinear analysis of brain signals that precede epileptic events. Lehnertz *et al.* [10] detect preevent indications in certain nonlinear measures of neural complexity, such as correlation dimension, interdependence, and mean phase coherence. Savit *et al.* [11] use a measure closely related to correlation integrals that detect changes tens of minutes prior to

Manuscript received March 27, 2002; revised November 8, 2002. This work was supported in part by Nicolet Biomedical Inc., Madison, WI (now ViaSys Healthcare Inc.) under a cooperative research and development agreement with the Oak Ridge National Laboratory (ORNL). This work was also supported in part by the U.S. Department of Energy, Office of Energy Management; by the U.S. Department of Energy, Office of Basic Energy Sciences; and by the U.S. Department of Energy, Office of Science, Laboratory Technology Research Program. ORNL is managed for the United States Department of Energy by UT-Battelle, LLC, under Contract DE-AC05-00OR22725. *Asterisk indicates corresponding author.*

*L. M. Hively is with the Oak Ridge National Laboratory, P.O. Box 2008, Bldg. 6011, MS-6415, Oak Ridge, TN 37831 USA (e-mail: hivelylm@ornl.gov).

V. A. Protopopescu is with Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA.

Digital Object Identifier 10.1109/TBME.2003.810693

U.S. Government work not protected by U.S. copyright.

an event. Jerger *et al.* [12] compare seven linear and nonlinear measures and find changes occurring one to two minutes before the event. Osorio *et al.* [13] show that the correlation integral and the correlation dimension depend on the EEG frequency and amplitude, implying that changes in the former are due to trivial changes in the latter. Sunderam *et al.* [14] chemically induce seizures in rats and use a stochastic Markov chain model to predict the interictal duration, T_i, from the seizure duration, T_s, and vice-versa. Litt *et al.* [15] show that the energy in EEG increases as the seizure approaches. These results are mostly based on *intracranial* EEG.

We propose the use of *scalp* EEG, which is much less invasive than intracranial data. While various approaches for seizure forewarning based on intracranial EEG have met with measurable degrees of success, scalp EEG has resisted previous efforts due to: 1) attenuation of the meaningful signal through bone and soft tissue and 2) high noise contamination, due to eye blinks and other muscular artifacts. Our initial analysis of scalp EEG used traditional nonlinear measures (TNM), such as Kolmogorov entropy, correlation dimension, and Lyapunov exponents, and yielded inconsistent detection and forewarning of the epileptic events [16]. Those results indicated that detection of meaningful information in attenuated, noisy, artifact-infested signals requires more sensitive and discriminating measures. Our more recent work showed by direct comparison that phase-space dissimilarity measures (PSDM) have consistently better sensitivity and discrimination power for forewarning of epileptic events than TNM [17]–[22]. The reason for this improvement is rather simple: discrimination by TNM is based on a difference of averages, while discrimination via PSDM is based on averaging the absolute value of pointwise differences. Our most recent work [23] yielded very encouraging results for event forewarning that is independent of patients' age or sex, event onset time, preevent activity, awake- versus asleep-state basecase, and—to a large extent—of data quality. However, our detailed analysis found inconsistent forewarning in the same channel across multiple datasets from the same patient [23]. This inconsistency arose because in our previous studies only total trues were maximized [17]–[23].

Channel-consistent total trues are a much more stringent measure of forewarning performance, which this present work addresses by: 1) formulating a quantitative measure of channel consistency in both true positives (TPs) and true negatives (TNs) for multiple datasets from each of several patients; 2) measuring forewarning performance in terms of channel-consistent total trues; 3) developing a methodology to maximize this performance measure; and 4) showing that these improvements raise the channel-consistent total trues substantially. The following sections discuss the phase-space methodology, channel-consistent forewarning, and the results of our analysis. The last section presents our conclusions.

II. Phase-Space Methodology

We begin this section with an overview of the PSDM. The long stream of time-serial archival scalp EEG data is first divided into contiguous, nonoverlapping windows (cutsets) of N data points. Each channel of every cutset is analyzed as follows. The artifact signal is removed with a novel zero-phase quadratic filter. Then by standard reconstruction of the dynamics [2], the artifact-filtered data is converted into a phase space (PS) representation, which in turn is transformed to a statistical distribution function (DF). Our underlying assumption is that the complex, high-dimensional brain dynamics evolve mostly over a bounded, low-dimensional region of the PS, called an "attractor." Thus, for sufficiently long datasets, the DF captures a statistical representation of the brain activity in the form of an invariant measure on the attractor. If dynamical conditions change, so will the attractor itself as mirrored by the invariant DF on the attractor. The DFs from the first ten cutsets are used as base cases to represent nonseizure dynamics. Our earlier work found that a few of these base case cutsets may be atypical, causing a bias in the detection of condition change [22], [23]. The base case DFs were exhaustively compared to one another in pair-wise fashion to identify outliers, which were removed from the analysis as unrepresentative of the nonseizure state [22], [23]. As discussed below, the present work finds that this removal of outliers is unnecessary. Next, each test case DF is compared to all of the base case DFs to obtain an average condition change between the base- and test-case states. A forewarning indication of an epileptic event (TP) is obtained in any one channel when all four PSDM exceed a threshold for a specified number of successive occurrences. No forewarning in a nonevent dataset corresponds to a TN. Finally, channel-consistency (the focus of the present work) requires that TP and TN indications occur in the same channel(s) across multiple datasets from the same patient. This complete analysis is repeated for every dataset. We use all of the datasets as a training set, which clearly limits the strength of our conclusions. However, the alternative of splitting 30 datasets into equally sized training and test sets would have provided only 15 datasets for the channel-consistency training analysis, which we deemed to be insufficient on statistical grounds. The remainder of this section provides methodological details for all of these steps.

EEG data were acquired as part of standard epilepsy monitoring of each patient for several days to two weeks under standard clinical protocols. Recordings came from 32-channel Biomedical Monitoring Systems Inc. instruments (Nicolet-BMSI, Madison, Wisconsin) with 19 scalp electrodes in the International 10–20 systems of placement as referenced to the ear on the opposing hemisphere. Each channel of scalp potential was amplified separately, band-pass filtered between 0.5–99 Hz, and digitized at 250 Hz. The data were archived on VHS tapes, which limited the maximum record length to slightly more than eight hours. The attending physicians independently selected the data for our analysis from this archive. Their sole selection criterion was that the data were "representative" of the patients' medical condition, without regard to the presence (or absence) of artifacts or other noise. In addition, we required that each dataset be greater than one hour long to assure adequate data for our analysis. We also required that event datasets include a baseline period of at least two hours, as well as the entire event (ictal) segment. The physician's characterization of this data covered only the time interval corresponding to the dataset. Analysis of this selected data represents only a fraction of the total monitoring periods,

and in principle limits the generality of our results. Analysis of the full monitoring periods is beyond the scope of this work.

We retrospectively analyzed 60 scalp EEG datasets: 40 with at least one electrographic temporal lobe event, plus 20 without epileptic events as controls. The attending physician characterized each dataset, as follows: patient's sex and age; start and stop time of the data recording; onset time of the seizure event; the type of seizure; the active electrode; and the patient's activity at the start of the recording and immediately prior to the event. Our earlier work [23, Table 1] summarizes the dataset characterizations, and will not be repeated here. The data were obtained from 41 different patients with ages between 4 and 57 years. Half of the patients have only a single dataset. The other 30 datasets are from 11 different patients with multiple datasets: seven patients with two datasets, one patient with three datasets, two patients with four datasets, and one patient with five datasets. Our analysis begins at the start of each dataset, and ends with the last full cutset in the dataset or with the cutset that includes the first event (whichever comes first). The dataset lengths range between 5016 s (1 h and 23 min) and 29 656 s (8 h and 14 min), with an average length of 15 668 s (4 h and 21 min). The cumulative period of the analysis is 940 104 s (261 h and 8 min). We analyze all 19 EEG channels in each dataset.

We remove eye blink artifacts from scalp EEG with a zero-phase quadratic filter [16] that is more efficient than conventional linear filters. This filter uses a moving window of data points, e_i, with the same number of data points, w, on either side of a central point. We fit a quadratic curve in the least-squares sense over this window, taking the central point of the fit as the low-frequency artifact, f_i. The residual value, $g_i = e_i - f_i$, has essentially no low-frequency artifact activity. All subsequent analysis uses artifact-filtered data.

We next convert each artifact-filtered value into a symbolized form, $0 \leq s_i, \leq S-1$ that is one of S different integers. We use contiguous, nonoverlapping partitions to obtain these symbols, as follows: $s_i = 0$ for $g_{\min} \leq g_i < g_1$; $s_i = 1$ for $g_1 \leq g_i < g_2$; $\ldots$; $s_i = \mathrm{S}-2$ for $g_{S-2} \leq g_i < g_{S-1}$; and $s_i = \mathrm{S}-1$ for $g_{S-1} \leq g_i \leq g_{\max}$. Here, $g_{\min}$ and $g_{\max}$ denote the minimum and maximum values of g_i, respectively, over the base case data. One specific approach creates symbols with a uniform distribution of partitions between the minimum and maximum in signal amplitude (uniform symbols)

$$0 \leq s_i = INT\left[S \cdot \frac{(g_i - g_{min})}{(g_{\max} - g_{\min})}\right] \leq S-1. \tag{1}$$

The function INT converts a decimal number to the closest lower integer. To maintain S distinct symbols, we require that $s_i = S-1$, when $g_i = g_{\max}$. Alternatively, one can use equiprobable symbols, by ordering all N base case data points from the smallest to largest value. The first N/S of these ordered data values correspond to the first symbol, 0, by choosing an appropriate value for g_1. Ordered data values $(N/S)+1$ through $2N/S$ correspond to the second symbol, 1, and so on up to the last symbol, $S-1$. For example, construction of 20 equiprobable symbols from 20 000 data points takes the first 1000 points in rank order as the first symbol, 0. The second 1000 points in rank order are the second symbol, 1, and so on up to the last symbol, 19. By definition, equiprobable symbols have nonuniform partitions in signal amplitude and have the advantage that dynamical structure arises *only* from the PS reconstruction (see below). Moreover, large negative and/or positive values of g_i have little effect on equiprobable symbolization, but significantly change the partitions for uniform symbols. Consequently, we find that equiprobable symbols provide better discrimination of condition change than uniform symbols, as discussed in the results section.

PS reconstruction [2] converts the data into a geometric form via time-delay vectors, $y(i) = [s_i, s_{i+\lambda}, \ldots, s_{i+(d-1)\lambda}]$, which unfold the underlying dynamics. Typically, the delay, λ, and system dimensionality, d, are related by $M_1 = (d-1)\lambda$, where M_1 the first minimum in the mutual information function. For a given d, we set $\lambda = \mathrm{INT}[0.5 + M_1/(d-1)]$ to obtain an integer value for the lag when M_1 is not evenly divisible by $d-1$. Symbolization divides the PS into S^d bins. The DF is a discretized density on the attractor, which is obtained by counting the number of points that occur in each PS bin. We denote the population of the ith bin of the DF, Q_i, for the base case, and R_i for a test case, respectively. We compare the test case to the base case by measuring the difference between Q_i with R_i [17]–[23] as

$$\chi^2 = \sum_i \frac{(Q_i - R_i)^2}{(Q_i + R_i)} \tag{2}$$

$$L = \sum_i |Q_i - R_i| \tag{3}$$

where the summations run over all of the populated PS cells. These measures account for changes in the geometry, shape, and visitation frequency of the attractor, and are somewhat complementary. The χ^2 measure is one of the most powerful, robust, and widely used comparison between observed and expected frequencies. In this context, χ^2 is a *relative* measure [20] of dissimilarity, rather than a distance in the mathematical sense or an unbiased statistic for accepting or rejecting a null statistical hypothesis. The L_1 distance is the natural metric for DFs by its direct relation to the total invariant measure on the attractor and defines a bona fide distance. Consistent calculations obviously require the same number of points in both the base case and test case DFs, identically sampled; otherwise the DFs must be properly rescaled.

We extend the PS reconstruction by connecting successive PS points for the dynamical flow, $y(i) \rightarrow y(i+1)$. Thus, we obtain [22] a discrete representation of the process flow [4] in the form of a $2d$-dimensional vector, $Y(i) = [y(i), y(i+1)]$, that is formed by adjoining two successive vectors from the d-dimensional reconstructed PS. We call $Y(i)$ the connected PS (CPS). As before, Q and R denote the CPS DFs for the base case and test case, respectively. We then define the PSDM between these two CPS DFs via the L_1-distance and χ^2 statistic, as before [22], [23]

$$\chi_c^2 = \sum_{ij} \frac{(Q_{ij} - R_{ij})^2}{(Q_{ij} + R_{ij})} \tag{4}$$

$$L_c = \sum_{ij} |Q_{ij} - R_{ij}|. \tag{5}$$

The subscript c denotes measures of the CPS DF. The first subscript in (4) and (5) is a label for the initial PS state, $y(i)$. The

second subscript in these equations is a label for the sequel PS state, $y(i+1)$. The CPS measures have higher discriminating power than their nonconnected counterparts. Indeed, we can prove [22] that the (C)PS measures defined in (2)–(5) satisfy the following inequalities: $\chi^2 \leq L$, $\chi_c^2 \leq L_c$, $L \leq L_c$, and $\chi^2 \leq \chi_c^2$.

We use the first B nonoverlapping, contiguous windows of N points (cutsets) from each channel of each dataset as base cases. Our earlier work found that a few of these base case cutsets may be atypical, causing a bias in the detection of condition change [22], [23]. The base case DFs were exhaustively compared to one another in pair-wise fashion to obtain the mean base-case dissimilarity, $\underline{V}$, and a corresponding standard deviation, σ Here, V denotes each dissimilarity measure from the set, $V = \{L, L_c, \chi^2, \text{ and } \chi_c^2\}$. A chi-squared statistical test was used to identify outlier cutsets, which were removed from the analysis as unrepresentative of the nonseizure state [22], [23]. A test of this approach for the present analysis yields 54 channel-consistent-total trues over all 60 datasets without outlier removal, versus 44 channel-consistent-total trues with outlier removal for the best parameters in the results section. Consequently, all of the below analysis is done without removal of outlier cutsets.

The disparate range and variability of these measures are difficult to interpret, so we need a consistent means of comparison. To this end, we renormalized the dissimilarity measures [20], [22], [23], as described below. We compared each of the B base case cutsets to each (ith) test case cutset, and obtained the corresponding average dissimilarity value, V_i, of the ith cutset for each dissimilarity measure. As before, V denotes the various dissimilarity measures from the set, $V = \{L, L_c, \chi^2, \text{ and } \chi_c^2\}$. The mean value, $\underline{V}$, and the standard deviation, σ, of the dissimilarity measure V are calculated using the base case cutsets, as discussed above. The renormalized dissimilarity is the number of standard deviations, by which the test case deviates from the base case mean: $U(V) = |V_i - \underline{V}|/\sigma$. This renormalized dissimilarity is used to test for statistically significant change in the time-serial dynamics, as discussed in Section III.

III. Channel-Consistent Forewarning

Once the renormalized measures for the test and base cases have been obtained, an arbitrary threshold, U_C, is set for each renormalized measure U to distinguish between normal (base) and possibly abnormal (test) regimes. The choice of a reasonable threshold is critical for obtaining robust, accurate, and timely results. A valid forewarning is indicated by a specified number, N_{OCC}, of sequential occurrences $U \geq U_C$, within a preset forewarning window [18]–[23]. These events are interpreted as true positive (TP). The complete statistics also accounts for the number of true negatives, TN, false positives (FPs), and false negatives (FNs), defined with respect to the same preset forewarning window, as shown in Fig. 1. The horizontal axis represents time, t. The thick vertical line at T_{SZ} denotes the seizure onset time. The thin vertical lines delimit the forewarning-time window, $T_1 \leq t \leq T_2 < T_{SZ}$. For illustration, a "reasonable" forewarning window is set with the limits: $T_1 = T_{SZ} - 60$ min and $T_2 = T_{SZ} - 1$ min.

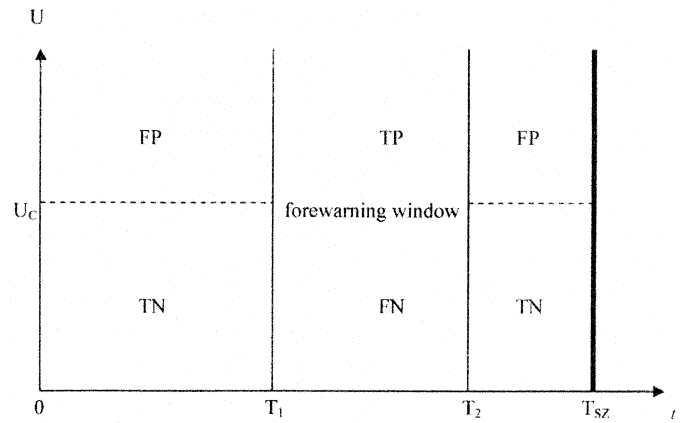

Fig. 1. Illustration of the forewarning window and threshold for TPs, TNs, FPs, and FNs; see text for details.

The latter time was chosen, based on input from a physician collaborator that even with one minute of forewarning, he could do useful things to help the patient medically [24]. The vertical axis corresponds to a renormalized dissimilarity measure, U, as discussed above. The horizontal dashed line (—) shows the threshold, U_C. We define a *forewarning time in one channel*, T_{FW}, as that time when the number of simultaneous indications, N_{SIM}, among the four dissimilarity measures exceeds some minimum value. The best elimination of FPs occurs for a value of $N_{SIM} = 4$. Analysis starts at $t = 0$, and proceeds forward in time until the first forewarning occurs, as defined above. The algorithm then obtains the forewarning statistics by the following ordered sequence of logical tests for each channel:

FP	forewarning at any time, when no seizure occurs, or forewarning with $T_{FW} < T_1$, or $T_{FW} > T_2$;
TP	forewarning with $T_1 \leq T_{FW} \leq T_2$;
TN	no forewarning, when no seizure occurs;
FN	no forewarning for $t \leq T_{SZ}$.

We denote the ith dataset as TP if at least one channel shows forewarning in the desired window, $T_1 \leq T_{FW} \leq T_2$. This assignment is equivalent to $TP_i = 1$ for the below summations. A TN dataset shows no forewarning in at least one channel when no seizure occurs, corresponding to $TN_i = 1$ for the below summations. The total true rate is $T = \Sigma_i(TP_i + TN_i)/60$, where the sum runs over all datasets. The corresponding total false rate is $F = 1 - T$. This approach allows selection of an appropriate channel for subsequent forewarning, consistent with the previous medical characterization.

Improvement in the channel-consistent total-true rate is carried out by maximizing an objective function that measures a combination of the total true rate for any one channel, *as well as channel consistency*. To quantify channel consistency, we introduce the following notation and definitions:

i	dataset number;
j	channel number with forewarning $(1 \leq j \leq 19)$;
k	patient number;
$M(k)$	number of datasets for the kth patient;
P	number of patients with multiple datasets;
TN_{ijk}	one for TN indication;
	zero for false negative indication;

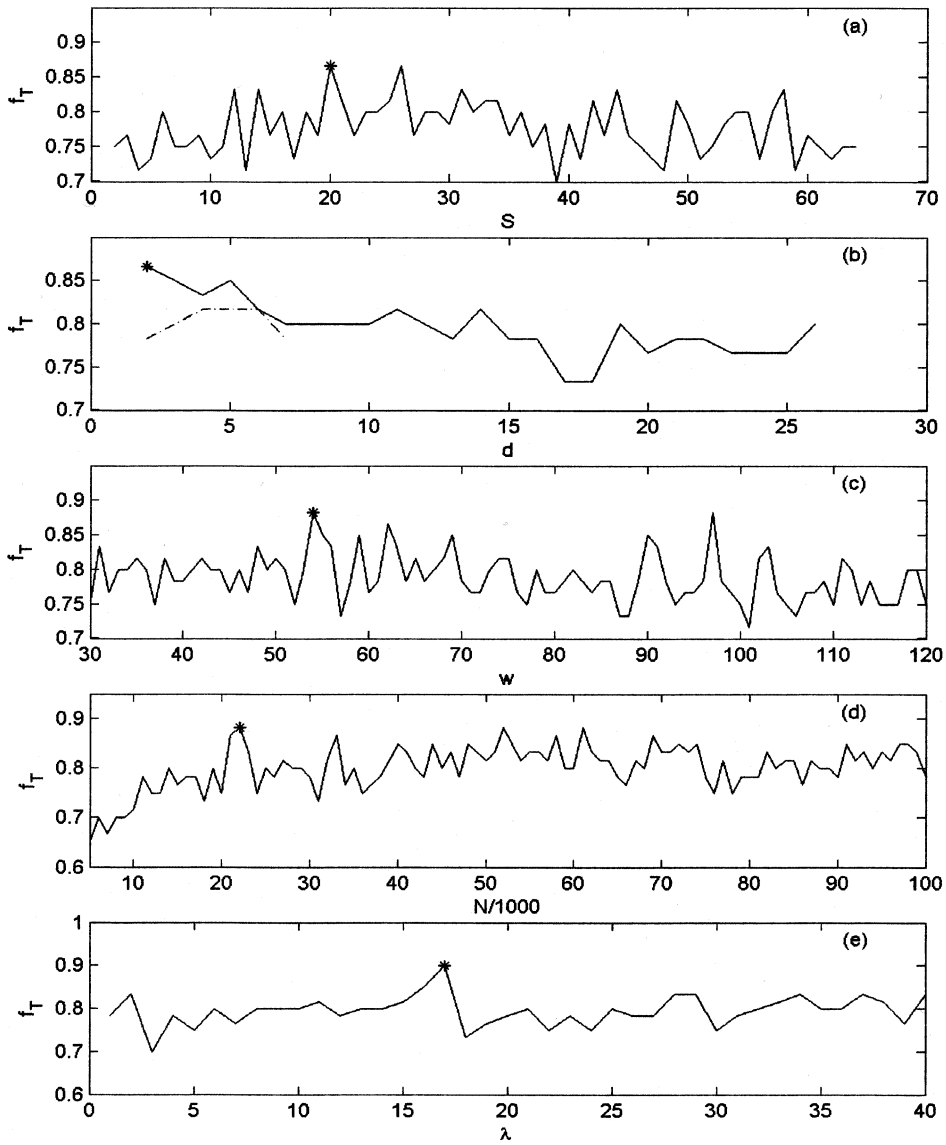

Fig. 2. Variation of the channel-consistent total-true value, f_T: (a) f_T versus S (number of PS symbols) for $d = 2$, $w = 62$, $N = 22\,000$; (b) largest f_T versus d (number of PS dimensions) for $w = 62$, $N = 22\,000$ using equiprobable symbols (solid curve) and uniform symbols (dash-dot curve); (c) f_T versus w (half width of the artifact-filter window in time steps) for $d = 2$, $S = 20$, $N = 22\,000$; (d) f_T versus $N/1000$ (thousands of data points in each cutset) for $d = 2$, $S = 20$, and $w = 54$; and (e) f_T versus λ (time delay for PS reconstruction in time steps) for $d = 2$, $S = 20$, $w = 54$, $N = 22\,000$. The chosen maximum is denoted by an asterisk $(*)$ in each subplot. See text for discussion.

TP_{ijk} one for a TP indication;
zero for a FP indication.

The subscript, "ijk" denotes the jth channel of the ith dataset for the kth patient. The total-true rate for the jth channel of the kth patient is $T_{jk} = \Sigma_i[TP_{ijk}+TN_{ijk}]$, by summing over $i = 1$ to $M(k)$. The occurrence of more than one TP(s) and/or TN(s) in the jth channel is indicated by $T_{jk} \geq 2$, while $T_{jk} \leq 1$ means that the jth channel provides no consistency with other datasets for the same patient. Then, we define the channel overlap as

$$c_k = \max(T_{jk}), \text{ for } T_{jk} \geq 2 \text{ and } k \text{ fixed}$$
$$= 0, \text{ for } T_{jk} \leq 1.$$

The channel-consistent total-true rate is, $f_T = [\Sigma_k c_k]/[\Sigma_k M(k)]$, where the index, k, sums over all P patients, weighting each dataset equally. (If we had defined the channel-consistent total-true rate as $[\Sigma_k \max(T_{jk})/M(k)]/P$, then patients with only one dataset would have been improperly weighted the same as patients with several datasets). We computed the renormalized PSDM for each dataset, and then exhaustively searched over N_{OCC} and U_C to find the largest f_T value.

IV. Results

We searched over seven parameters to maximize f_T. Our experience [20]–[23] shows a huge reduction in the computational effort by varying one parameter at a time while holding the others fixed. Indeed, a sequence of single parameter searches avoids the dimensionality curse of an exhaustive search at the expense of a suboptimal, but quite acceptable result. After determination of these parameters, the forewarning analysis is much faster than the data-recording time. We present the results of this search next.

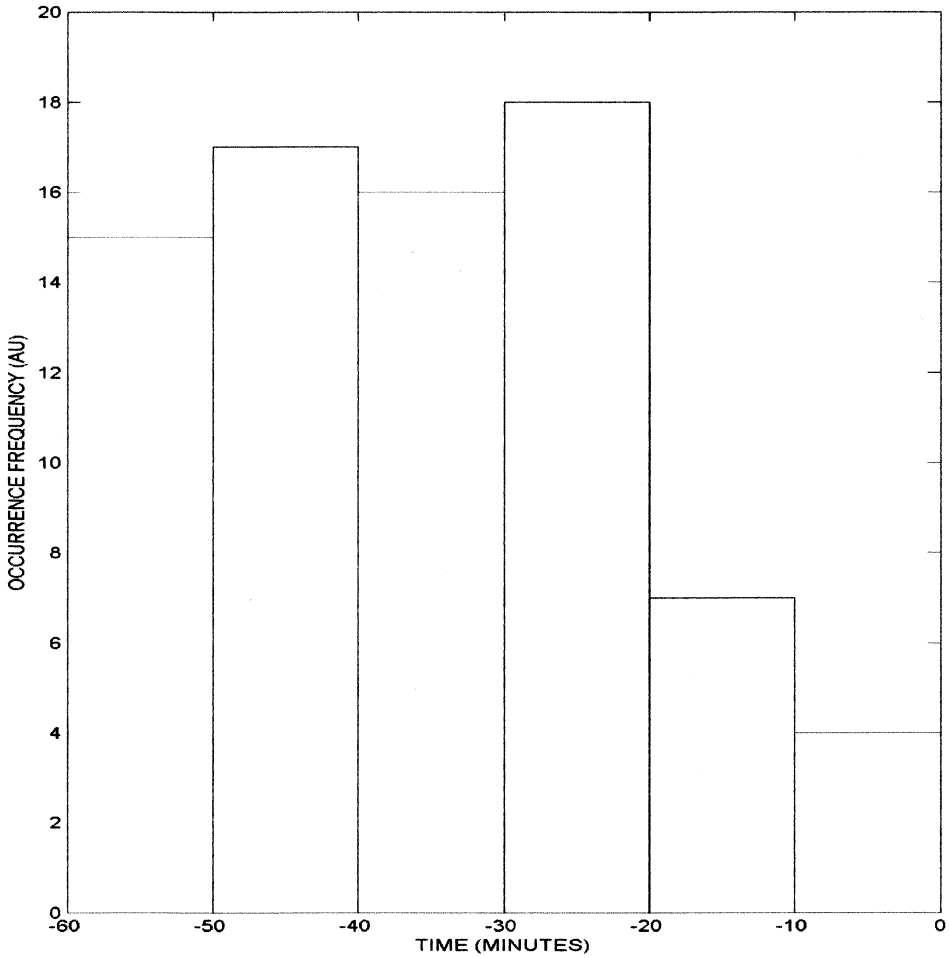

Fig. 3. The occurrence frequency of the channel-consistent forewarning time (in arbitrary units) versus time (in minutes) prior to the seizure event, which occurs at time, $t = 0$.

We first optimized over the parameter pair, $\{S, d\}$, constrained by a computational limit on the numeric labels for the CPS bins using modular arithmetic. This limit arises from the largest double-precision real number that can be distinguished from one unit larger: $S^{2d} \leq 2^{52}$, yielding the limit, $d \leq \mathrm{INT}(26 \ \ln 2/\ln S)$. Two PS symbols $(S = 2)$ limit the search to $2 \leq d \leq 26$; three PS symbols $(S = 3)$ correspond to $2 \leq d \leq 16$; and so forth. Since equiprobable symbols always yield larger values for f_T, we performed the analysis using this symbolization. Fig. 2(a) shows f_T as a noisy function of the number of equiprobable symbols, S, with the number of PS dimensions, $d = 2$. The largest channel-consistent total-true rate is $f_T = 0.8667$ at $S = 20$ and $S = 26$. We next obtained the largest value of f_T versus the number of equiprobable symbols for each value of PS dimension in the range, $2 \leq d \leq 26$. These results are displayed as the solid curve in Fig. 2(b), showing that f_T decreases nonmonotonically from a maximum at $d = 2$ to a minimum at $d = 17$ or 18, and then rises somewhat for still larger values of d. For completeness, Fig. 2(b) also shows similar analysis only for an even number of uniform symbols (dash-dot curve), because the central bin for an odd number of uniform symbols accumulates the vast majority of the PS points, thus degrading the results in every case. Based on these results, further analysis for uniform symbols is unnecessary.

Additional parameters are available for optimization. We set $d = 2$ and $S = 20$ (equiprobable symbols) while varying the half-width of the artifact filter window, w, as shown in Fig. 2(c). We find that f_T is a noisy function of w with a maximum value of $f_T = 0.8833$ at two places. We choose $w = 54$ for the next parameter scan, because a slight trend for larger f_T lies in that region. Fig. 2(d) displays f_T versus the number of points in each cutest, N, in increments of 1000 with the other parameters fixed at $S = 20$, $d = 2$, and $w = 54$. This plot shows that f_T rises nonmonotonically with increasing cutset length to a noisy plateau for $N > 21\,000$. The largest value, $f_T = 0.8833$, occurs three times. We chose $N = 22\,000$, because f_T tends to be larger in that region. All of the above analysis uses a time delay that is based on the first minimum, M_1, in the mutual information function, $\lambda = \mathrm{INT}[0.5 + M_1/(d-1)]$. This lag turns out to be different for every channel of each dataset. Thus, we can consider λ as another parameter in the optimization. Fig. 2(e) illustrates the variation of f_T versus λ, which is set to the same value for every channel of every dataset, yielding a maximum of $f_T = 0.9$ at $\lambda = 17$ with the other parameter values set as follows: $S = 20$, $d = 2$, $w = 54$, and $N = 22\,000$.

We searched over two other parameters for completeness. We varied the number of base cases in the range, $3 \leq B \leq 20$, and obtained a largest $f_T = 0.9$ for $B = 10$. This value is

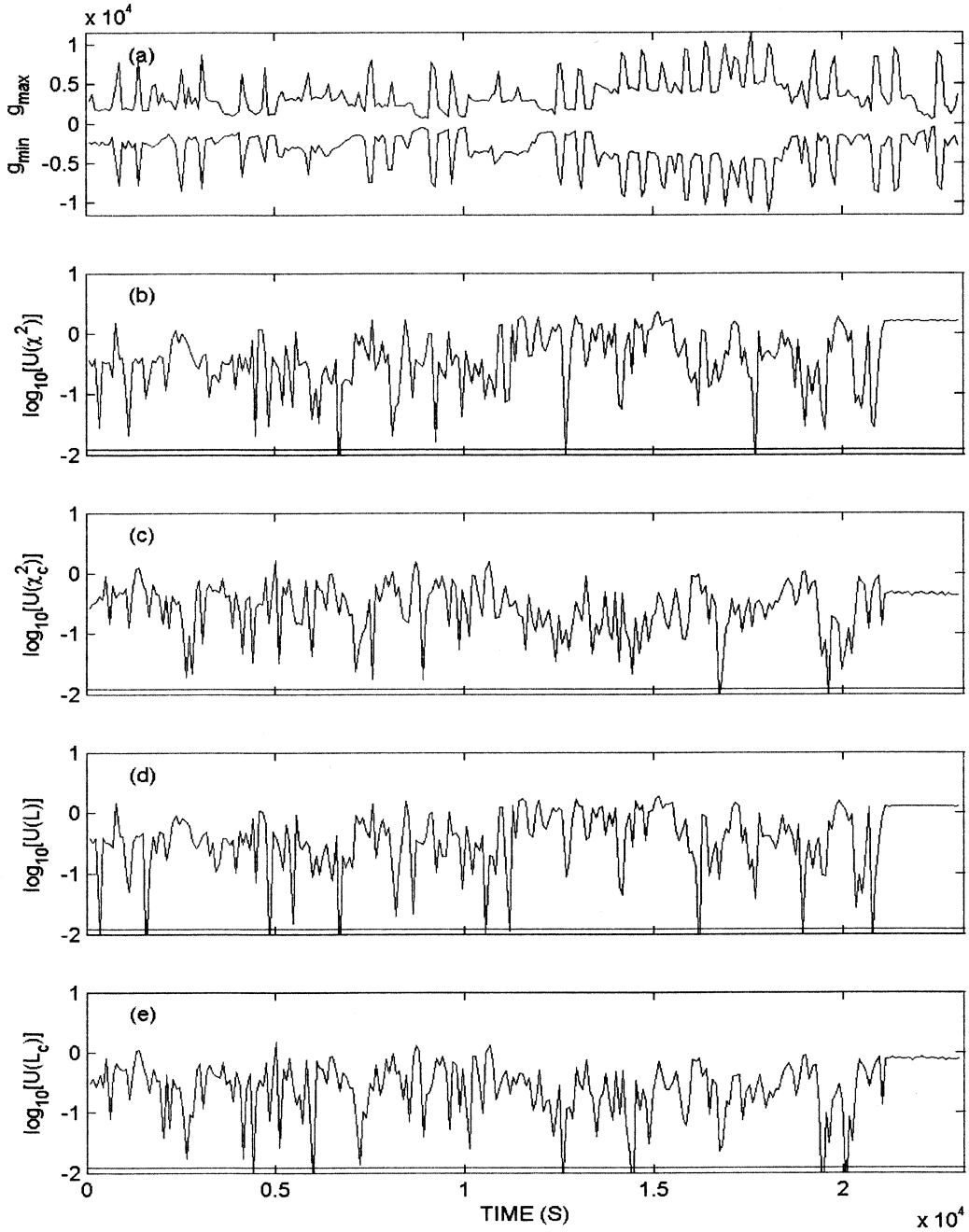

Fig. 4. Linear measures and renormalized PSDM for a TN (channel F8 of dataset 37 for patient 11) versus time: (a) minimum $(g_{\min})$ and maximum $(g_{\max})$ values of the artifact-filtered EEG; (b) $U(\chi^2)$; (c) $U(\chi_c^2)$; (d) $U(L)$; and (e) $U(L_c)$. The indication threshold $(U_C = 0.012)$ is shown as the solid horizontal line at the bottom of (b)–(e); see text for discussion.

the same as before. The above analysis also has an inter-symbol lag, $\mu = 1$, for the connected phase-space, $y(i) \rightarrow y(i+\mu)$, as described previously. We searched over the range, $1 \leq \mu \leq 20$. The maximum is $f_T = 0.9$ at $\mu = 1$, which also is the same as before. A second series of single parameter searches over S, w, and N did not find a larger f_T. Moreover, an exhaustive search over $15 \leq S \leq 25$ and $12 \leq \lambda \leq 22$ did not improve f_T with the other parameters fixed as before. Thus, we obtain the maximum channel-consistent total-true rate of $f_T = 0.9$ for $S = 20$, $d = 2$, $\lambda = 17$, $w = 54$, $N = 22\,000$, $B = 10$, $\mu = 1$. We next discuss this best result in more detail.

This retrospective analysis shows that substantial improvement in the rate of channel-consistent total trues is possible over the previous value of $f_T = 0.68$ [23]. Our best result yields 29 trues for patients with a single dataset: 23 TPs, six TNs, one FP, and no false negatives. This result also includes 25 channel-consistent trues for patients with multiple datasets: 12 TPs, 13 TNs, 4.47 FPs, and 0.53 false negatives. The noninteger values arise from two event datasets from patient 69 that have inconsistent or false results for every channel, for which the falses were averaged across all channels for both datasets. A channel with a FP (negative) in one dataset and a TP in the other dataset was counted as two FPs (negatives). The total number of FPs is 5.47, or 0.021 FPs per hour (mean time between FPs of 47.7 h). The total number of TP datasets is 35 out of 40 (87.5%). Fig. 3 shows the occurrence frequency of the channel-consistent forewarning

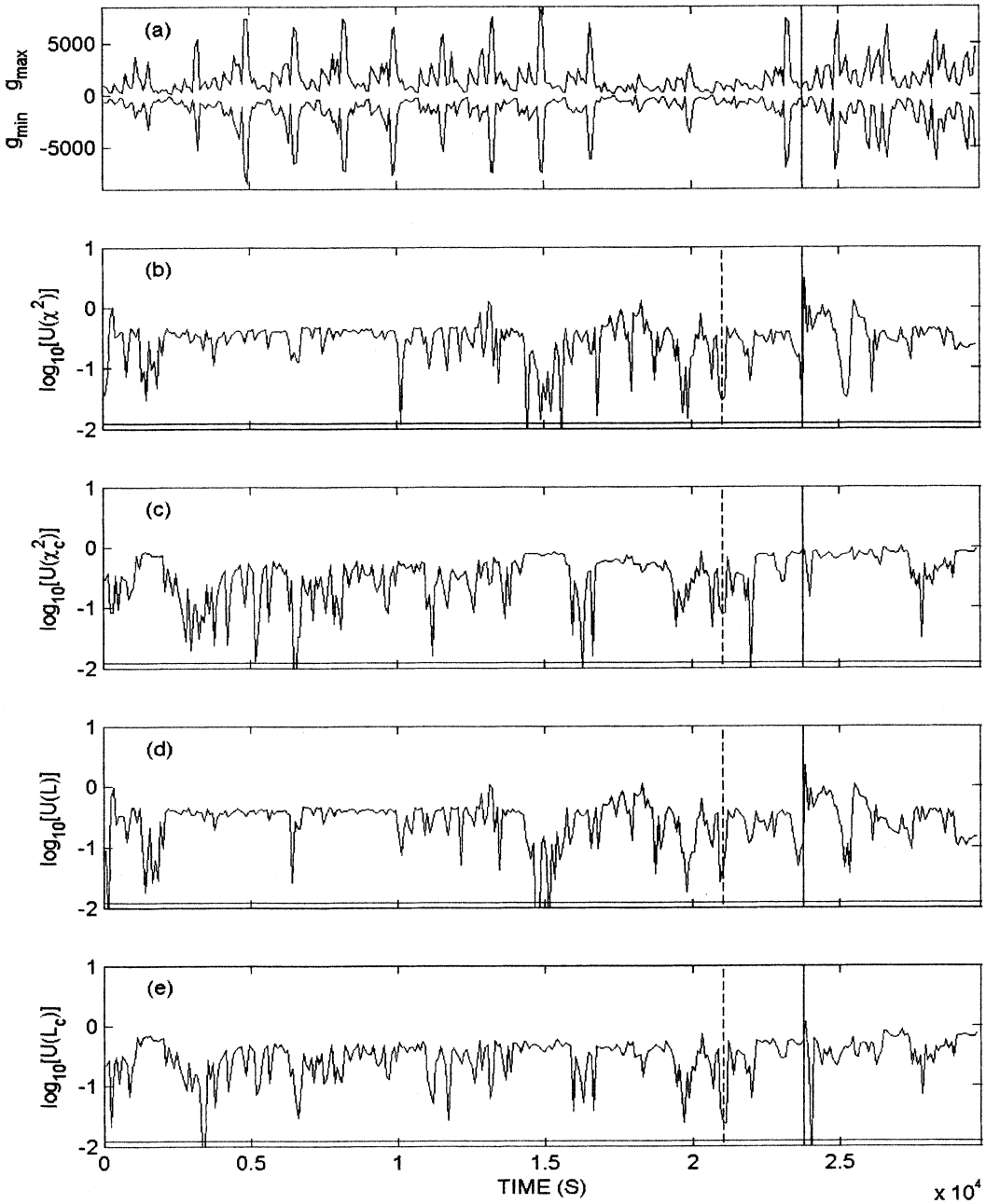

Fig. 5. Linear measures and renormalized PSDM for a TP (channel P4 of dataset 207 for patient 83) versus time: (a) minimum $(g_{\min})$ and maximum $(g_{\max})$ values of the artifact-filtered EEG; (b) $U(\chi^2)$; (c) $U(\chi_c^2)$; (d) $U(L)$; and (e) $U(L_c)$. The indication threshold $(U_C = 0.012)$ is shown as the solid horizontal line at the bottom of (b)–(e). The epileptic event is denoted by the vertical bar at 23 760 s in each subplot. The dashed line (- -) at 21 032 s shows the forewarning indication. See text for discussion.

times, relative to the seizure event onset, in intervals of 600 s (10 min). These forewarning times are consistent with the seizure event forewarning zone, with an inclusive range of 60–3,588 s, an average of 2,095 s, and a standard deviation of 905 s.

Fig. 4 is an example of a TN (no forewarning when no event occurred) with the same parameters as the best result, above. This dataset spans 23 056 s (6 h and 24 min) and is one of four for patient 11, with each of the other three datasets having an epileptic event. Three of the four datasets had a correct indication (TN for dataset 37 and TPs for datasets 125 and 131) in channel F8, which is the channel displayed in this figure. Fig. 4(a) shows many large amplitude EEG extremes (particularly after 16 000 s) that provide no insight. The large values of the PSDM after 21 000 s were diagnosed independently by the physician as a nonseizure event, for which the method correctly shows no forewarning. The infrequent occurrence of all four PSDM [Fig. 4(b)–(e)] below the indication threshold $(U_C = 0.012)$ is the feature that allows the method to avoid a FP indication.

Fig. 5 is an example of a TP (forewarning when an event occurred) with the same parameters as the best result, above. This dataset spans 29 656 s (8 h and 14 min) and is the only dataset for patient 83. This figure shows data for channel P4 with a seizure event at 23 760 s. The many large amplitude EEG ex-

tremes [Fig. 5(a)] provide no forewarning and no indication of the epileptic event. Three of the four PSDM have maximal peaks at the seizure event [Fig. 5(b)–(e)]. The presence of infrequent small-dissimilarity states before 16 000 s avoided a premature (FP) forewarning. The absence of small-dissimilarity states for 54 cutsets after 16 000 s allowed the method to give a forewarning at 21 032 s (2728 s of forewarning time prior to the event).

V. Discussion

As substantiated by the results in Figs. (2)–(5) and in our previous work [17]–[23], we view the present approach as encouraging and worthy of further development despite several limitations, which we discuss next. First, the present analysis uses all of the data as a training set, limiting the strength of our conclusions. However, the alternative of splitting 30 datasets into equally sized training and test sets would have provided only 15 datasets for the channel-consistency training analysis, which would have resulted in weak statistics. Second, we analyze EEG data from a controlled clinical setting, which differs from an uncontrolled, ambulatory environment. Third, we define forewarning as a TP indication in at least one channel, which is a rather weak criterion. Fourth, our results depend on careful adjustment of the analysis parameters for the best rate of TPs plus TNs. Fifth, we analyze only portions of the patient data, as selected by the attending physician, rather than the full monitoring period. Sixth, the present results are limited by our analysis of only 60 datasets (40 with epileptic events). Much more data (hundreds of datasets spanning more than 1000 h) are needed for an optimal choice of the analysis parameters and for a definitive statistical demonstration of the effectiveness of our approach. These data requirements are far beyond our present capabilities, and almost everyone else's, based on published results to date. Seventh, we have not performed prospective analysis of long-term continuous data, which is the acid test for any predictive approach. Clearly, much work remains to address these issues. In this light, we view the importance of the present work as the methodology and search strategy for channel-consistent forewarning, rather than in the specific results.

We anticipate that a practical ambulatory (prospective) forewarning device would involve the following. The analysis parameters would be chosen, based on a statistical analysis of many datasets, and then fixed for subsequent patient monitoring. An individual patient would undergo epilepsy monitoring for several days to two weeks under standard clinical protocols. The complete data record of this monitoring would be analyzed with the PSDM methodology to determine the best channel(s) for forewarning. This analysis would also provide the number of false negatives and the hourly rate of FPs. If this false rate is acceptable, then the patient would be fitted for electrodes in a wig or cap, which would be worn continuously for forewarning. The initial patient setup would assure that the patient is awake during the scalp electrode setup by medical personnel, who would then explain the device and its use to the patient and/or the caregiver while the base case is acquired for several minutes. The analysis of one cutset of data would only proceed after being written to the computer harddrive or memory, and thus would be at least one cutset behind real-time. If statistically based parameters for the forewarning analysis give an excessive FP rate, then the parameters could be set via a patient-specific analysis if at least ten seizure events occurred during clinical monitoring. If the existing base case data yield an excessive false hourly rate, then a new base case could be acquired subsequently, and/or at periodic intervals. This scenario has many uncertainties, based on the present state of our research.

Acknowledgment

The authors would like to acknowledge the provision and characterization of the EEG data by Nicolet Biomedical Inc., Madison, WI (now ViaSys Healthcare Inc.) under a Cooperative Research and Development Agreement with ORNL.

References

[1] B. Hjorth, "EEG analysis based on time series properties," *Electroencephal. Clin. Neurol.*, vol. 29, pp. 306–310, 1970.
[2] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.*, vol. 57, pp. 617–655, 1985.
[3] E. A. Jackson, *Perspectives of Nonlinear Dynamics*. Cambridge, U.K.: Cambridge Univ. Press., 1989, vol. 1.
[4] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*. New York: Springer-Verlag, 1996.
[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[6] S. H. Strogatz, *Nonlinear Dynamics and Chaos*. Reading, MA: Addison Wesley, 1994.
[7] A. Babloyantz and A. Destexhe, "Low dimensional chaos in an instance of epilepsy," *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3513–3517, 1986.
[8] V. Gribkov and D. Gribkova D, "Learning dynamics from nonstationary time series: analysis of electroencephalograms," *Phys. Rev. E*, vol. 61, pp. 6538–6545, 2000.
[9] M. Le Van Quyen, J. Martinerie, V. Navarro, M. Baulac, and F. J. Varela, "Characterizing neurodynamic changes before seizures," *J. Clin. Neurophysiol.*, vol. 18, pp. 191–208, May 2001.
[10] K. Lehnertz, R. G. Andrzejak, J. Arnhold, T. Kreuz, F. Mormann, C. Reike, G. Widman, and C. E. Elger, "Nonlinear EEG analysis in epilepsy: its possible use for interictal focus localization, seizure anticipation, and prevention," *J. Clin. Neurophysiol.*, vol. 18, pp. 209–222, May 2001.
[11] R. Savit, D. Li, W. Zhou, and I. Drury, "Understanding dynamical state changes in temporal lobe epilepsy," *J. Clin. Neurophysiol.*, vol. 18, pp. 246–258, May 2001.
[12] K. K. Jerger, T. I. Netoff, J. T. Francis, T. Sauer, L. Pecora, S. L. Weinstein, and S. J. Schiff, "Early seizure detection," *J. Clin. Neurophysiol*, vol. 18, pp. 259–268, May 2001.
[13] I. Osorio, M. A. F. Harrison, Y.-C. Lai, and M. G. Frei, "Observations on the application of the correlation dimension and correlation integral to the prediction of seizures," *J. Clin. Neurophysiol.*, vol. 18, pp. 269–274, May 2001.
[14] S. Sunderam, I. Osorio, M. G. Frei, and J. F. Watkins III, "Stochastic modeling and prediction of experimental seizures in sprague-dawley rats," *J. Clin. Neurophysiol.*, vol. 18, pp. 275–282, May 2001.
[15] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, and G. Vachtsevanos, "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients," *Neuron*, vol. 30, pp. 51–64, Apr. 2001.
[16] L. M. Hively, N. E. Clapp, C. S. Daw, and W. F. Lawkins, "Nonlinear Analysis of EEG for Epileptic Seizures," Oak Ridge Nat. Lab., Oak Ridge, TN, ORNL/TM-12961, Apr. 1995.
[17] P. C. Gailey, L. M. Hively, and V. A. Protopopescu, "Detection of condition change in nonlinear times series using a connected-phase-space PDF," presented at the Amer. Phys. Soc. Conf., Atlanta, GA, Mar. 1999, Paper LC32.06.
[18] L. M. Hively, P. C. Gailey, and V. A. Protopopescu, "Sensitive measures of condition change in EEG data," in *Proc. Workshop Chaos in Brain?*, Mar. 1999, pp. 333–336.
[19] P. C. Gailey, L. M. Hively, and V. A. Protopopescu, "Robust detection of dynamical change in EEG," presented at the *5th Experimental Chaos Conf.*, Orlando, FL, 1999.

[20] L. M. Hively, P. C. Gailey, and V. A. Protopopescu, "Detecting dynamical change in nonlinear time series," *Phys. Lett. A*, vol. 258, pp. 103–114, July 1999.

[21] L. M. Hively, N. E. Clapp, V. A. Protopopescu, J. Joseph, C. E. Merican, and T. Lucht, "Epileptic Seizure Forewarning by Nonlinear Techniques," Oak Ridge National Lab., Oak Ridge, TN, ORNL/TM-2000/333, Nov. 2000.

[22] L. M. Hively, V. A. Protopopescu, and P. C. Gailey, "Timely detection of dynamical change in scalp EEG signals," *Chaos*, vol. 10, pp. 864–875, Dec. 2000.

[23] V. A. Protopopescu, L. M. Hively, and P. C. Gailey, "Epileptic seizure forewarning from scalp EEG," *J. Clin. Neurophysiol.*, vol. 18, pp. 223–245, May 2001.

[24] M. L. Eisenstadt, *Private communication*.

Lee M. Hively received the B.S. degrees in engineering science and in mathematics, and the B.A. degree in general arts and sciences in 1970 from the Pennsylvania State University, University Park; the M.S. degree in physics from University of Illinois (Urbana), in 1971; and the Ph.D. degree in nuclear engineering in 1980 from University of Illinois (Urbana).

From 1971 to 1974, he was a Member of Research Staff at the Western Electric Company's Engineering Research Center, Princeton, NJ, where his work involved noncontact inspection of millimeter waveguide. From 1974–1980, he was a Graduate Research Assistant at University of Illinois, where he pursued research on fusion product loss in axisymmetric tokamaks. From 1980 to 1984, he was employed by General Electric Company under subcontract to the Fusion Engineering Research Center at Oak Ridge National Laboratory (ORNL). His research interests included energetic alpha particle thermalization and losses in tokamak fusion reactors with toroidal field ripple, optimization of toroidal bundle divertors, plasma performance simulation, and electromagnetic railgun launchers. Since 1984, he has been a Research Staff Member at ORNL (Oak Ridge, TN). His work has included: health and safety research, program management as a detailee at U.S. Department of Energy Headquarters (Germantown, MD), fusion product physics, and criticality safety and nuclear shielding for nuclear transportation packages. His present research focuses on nonlinear analysis of time serial data to forewarn of critical events, such as epileptic seizures, cardiac fibrillation, breathing difficulty, syncope, sepsis, machine failure, and tool wear. He has authored seven U.S. Patents and one U.S. Patent pending; he has coauthored one U.S. Patent. He also is an adjunct assistant professor of surgery at the University of Tennessee Medical Center, Knoxville.

Vladimir A. Protopopescu received the Ph.D. degree in mathematical physics from the Institute for Atomic Physics, Bucharest, Romania, in 1976.

Since 1968, he worked successively at the Institute for Atomic Physics, Bucharest; Chalmers University of Technology, Goteborg, Sweden; Yale University, New Haven, CT; and Boston University, Boston, MA. In 1985, he joined the Oak Ridge National Laboratory, where presently he is a member of the Distinguished Research Staff in the Computer Science and Mathematics Division. Since 1988, he has been an Adjunct Professor of Mathematics at the University of Tennessee, Knoxville.

He is a Member of AMS, SIAM, APS, and IAMP. He serves as an Associate Editor for two archival journals, *Mathematical Models and Methods in Applied Sciences* and *Transport Theory and Statistical Physics*, and two book series *Modeling and Simulation in Science, Engineering, and Technology* (Cambridge, MA: Birkhäuser) and *CRC Mathematical Modeling* (Boca Raton, FL: CRC Press). His research interests include mathematical modeling, analysis and optimal control of partial differential equations, dynamical systems and chaos, inverse problems, global optimization, quantum control, and quantum computing. He has published over 160 scientific papers in various areas of applied mathematics and mathematical physics, and coauthored a monograph on boundary value problems in kinetic theory, which—more than 15 years after its publication—continues to be the absolute reference for rigorous treatment of transport problems.

Dr. Protopopescu is the recipient of the 1976 "D. Hurmuzescu" Award of the Romanian Academy for physics and of the 1998 R&D 100 Award on his work on global optimization. He holds two U.S. patents.