A comparison of non-linear non-parametric models for epilepsy data

Fumikazu Miwakeichi\textsuperscript{a, *}, Ruben Ramirez-Padron\textsuperscript{b}, Pedro A. Valdes-Sosa\textsuperscript{b}, Tohru Ozaki\textsuperscript{c}

\textsuperscript{a}The Graduate University for Advanced Studies, 4-6-7 Minami Azabu, Minato-ku Tokyo 106-0047, Japan
\textsuperscript{b}Cuban Neuroscience Center, Ave. 25 No. 15202 esquina 152, Apartado 6880, Habana, Cuba
\textsuperscript{c}The Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-ku Tokyo 106-0047, Japan

Received 12 November 1999; accepted 4 May 2000

Abstract

EEG spike and wave (SW) activity has been described through a non-parametric stochastic model estimated by the Nadaraya–Watson (NW) method. In this paper the performance of the NW, the local linear polynomial regression and support vector machines (SVM) methods were compared. The noise-free realizations obtained by the NW and SVM methods reproduced SW better than as reported in previous works. The tuning parameters had to be estimated manually. Adding dynamical noise, only the NW method was capable of generating SW similar to training data. The standard deviation of the dynamical noise was estimated by means of the correlation dimension. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Epilepsy; Time-series analysis; Nadaraya–Watson; Local polynomial regression; Support vector machines; Chaos; Non-linear stochastic system

1. Introduction

Epilepsy is a chronic disorder of the central nervous system characterized by recurrent seizures produced by hyper-synchronous discharges of groups of neurons [1]. Electroencephalograph (EEG) recordings of epileptic seizures have been, for some time, a favorite subject of signal analysis. In particular, one kind of epileptic EEG, the so-called spike and wave (SW) activity, has attracted the attention of many mathematicians and physicists due to its highly non-linear dynamics. With the advent of “chaos theory” there were many initial reports claiming that SW is a chaotic state of brain activity \[2–4\]. Many of these conclusions were based on flawed statistical procedures and a too restrictive dichotomy of systems into only two types: linear stochastic and deterministic chaotic. The
introduction of more adequate statistical control [5] and the recognition that stochastic non-linear systems can exhibit highly complex behavior has led to a new point of view [6]. In contrast to “chaos theory”, this approach views the EEG as the output of a non-linear stochastic system and uses the methods of non-linear time-series analysis to evaluate system dynamics. Central to this way of analyzing the EEG is the explicit construction of a model that is required to be “mimetic” in the sense of producing realistic simulations of the type of EEG under study.

In the case of SW, “chaos theory” has not been able to formulate an explicit time series model capable of reproducing the complex phase relationships that could be easily recognized by any neurologist. In other words, chaos theory has not generated mimetic models. On the other hand, based on the non-linear time series point of view, Hernandez et al. [7] demonstrated that SW data can be adequately described by a non-linear stochastic model fitted by the Nadaraya–Watson (NW) kernel method.

In order to be more specific some notations will be introduced at this point. Given a time series \((x_1, x_2, \ldots, x_N)\), the general formulation of a non-linear stochastic model is

\[
x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-m}) + e_t,
\]

where \(f : \mathbb{R}^m \rightarrow \mathbb{R}\) is a smooth map, which has been designated as the ‘skeleton’ [8]; \(e_t\) represents dynamical noise; and \(m\) is a positive integer, the so-called embedding dimension. In the case in which \(f\) is known except for an unknown vector of parameters \(\theta\) to be estimated, i.e.\(f(x_{t-1}, x_{t-2}, \ldots, x_{t-m}) = f(x_{t-1}, x_{t-2}, \ldots, x_{t-m}, \theta)\), the model is called parametric; otherwise, it is called non-parametric. In this paper, the focus will be on ‘non-parametric’ models. Basically, these models are estimated by overfitted regression functions where some type of regularization is imposed [9,10]. These methods for estimation usually depend on ‘tuning parameters’ that control aspects of the regularization imposed.

Given an arbitrary initial point \((x_{t_0-1}, x_{t_0-2}, \ldots, x_{t_0-m})\), simulated realizations of the model are obtained by generating the noise \(e_t\) and applying the recurrence relation (1) for \(t = t_0 + 1, t_0 + 2, \ldots\), starting from the initial point. If in this recurrence we set the noise equal to zero, noise-free realizations (NFR) are obtained, i.e. they are trajectories generated by the skeleton of the model. They can be thought of as long-term predictions derived from the model. In contrast, the one-step-ahead prediction \(\hat{x}_t\) of each point \(x_t\), on the basis of the past data \((x_{t-1}, x_{t-2}, \ldots, x_{t-m})\), is defined by applying Eq. (1) without noise.

Several parametric and non-parametric methods of fitting the map \(f\) have been considered to describe SW activity [6,11,12]; for example, threshold autoregression, NW, radial basis functions, etc. But only NW has been successful by the mimetic criterion given above [7]. However, in the case of NW method all the original data are necessary for estimating \(f(x)\) at any point \(x\), which is inconvenient for many practical purposes. Moreover, its skeleton generated a noise-free realization with a larger inter-spike interval than those of the actual data. This suggests it is possible to improve on this type of method.

In this paper, we study how to improve the performance of the NW method, and also explore two alternative methods for describing SW data: local linear polynomial regression (LLPR) and support vector machines (SVM). These methods have been shown to be very useful for describing other kinds of complex data. The SVM method has the additional advantage of reducing the number of data points needed in the specification of the map \(f\). Specifically, we carry out a comparison between these methods that includes the following aspects: determination of the attractors of the
estimated skeletons; evaluation of stability of the skeletons; estimation of the number of attractors of the skeletons in a region encompassing actual data; and behavior of the realizations when applying dynamical noise. A comparison between automatic and manual techniques for tuning parameters was also carried out.

2. Methods

In the following subsections the NW, LLPR and SVM methods are briefly introduced. They are used in next sections to estimate the skeletons of the stochastic models for SW activity. Details about computational implementation are given in the final subsection.

2.1. Nadaraya–Watson kernel estimator

NW kernel estimator is a well-known method for non-parametric function fitting [13,14]. According to this method, the estimate of $f$ in (1), at a point $(z_{t-1}, z_{t-2}, \ldots, z_{t-m})$ of the state space, is obtained as a weighted average of all the data $(x_1, x_2, \ldots, x_N)$. Specifically,

$$
\hat{f}(z_{t-1}, z_{t-2}, \ldots, z_{t-m}) = \frac{\sum_{i=1}^{N} x_i \prod_{j=1}^{m} K(|z_{t-j} - x_{i-j}|, h)}{\sum_{i=1}^{N} \prod_{j=1}^{m} K(|z_{t-j} - x_{i-j}|, h)}.
$$

(2)

Here, $K(|z-x|, h)$ is the kernel function, and the tuning parameter $h$ is some positive real number (width of the kernel).

2.2. Local linear polynomial regression

LLPR is a particular case of local polynomial regression (LPR) [9,15,16]. In general, the LPR method approximates the function $f$ in (1), in a neighborhood of each point $x_0$ of the state space, as a local (multivariate) polynomial, i.e.,

$$
f(x) = p^x(x - x_0, \theta(x_0));
$$

where $\theta = \theta(x_0)$ is the vector of coefficients of the $n$-degree polynomial $p$.

To estimate the coefficients, the following least-squares problem is solved:

$$
\hat{\theta}(x_0) = \arg \min_{\theta \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} [y_i - p^x(x_i - x_0, \theta)]^2 K \left( \frac{x_i - x_0}{h} \right) \right\},
$$

(3)

where the $y_i$s are dependent observed real values, and the $x_i$s are independent observed values lying in the state space, and $K$ is some kernel function. The value of $f$ at $x_0$ is estimated by the first component of the coefficient vector $\theta$, i.e., $\hat{f}(x_0) = \hat{\theta}(x_0) \equiv p^x(0, \hat{\theta})$.

In the case $k = 1$ it is called local linear polynomial regression (LLPR), and (3) becomes

$$
\{\hat{f}(x_0), \hat{i}(x_0)\} = \arg \min_{f(x_0), i(x_0)} \left\{ \sum_{i=1}^{n} [y_i - f(x_0) - j(x_0)(x_i - x_0)]^2 K \left( \frac{x_i - x_0}{h} \right) \right\}.
$$

Here $\hat{j}(x_0)$ provides an estimate of the vector of first derivatives (gradient) at the point $x_0$.

For $k = 0$ the LPR method reduces to the NW method described in Section 2.1.
2.3. Support vector machines

The SVM method was proposed by Vapnik [17] on the basis of the Structural Risk Minimization Principle [18]. It was initially designed to solve pattern recognition problems [19]; but it was later applied to function estimation problems [20]. The estimated function is a linear expansion in terms of functions defined on a certain subset of the data (support vectors), and the final number of coefficients in such expansion does not depend on the dimensionality of the space of input variables. These two properties make SVM an especially useful technique for dealing with very large data sets in a high-dimensional space.

The data for solving the regression problem will be represented in this subsection as \( \{X, y\} = \{x_i, y_i\}_{i=1}^n \), where \( n = N - m \), \( x_i \in \mathbb{R}^m \) are the rows of \( X \), and \( y_i \in \mathbb{R} \). The SVM method seeks a function \( \hat{f} \) such that

\[
\hat{f} = \arg \min_{g \in H} R[g] = C \sum_{i=1}^n |y_i - g(x_i)|_e + \frac{1}{2} \|g\|_H^2,
\]

(4)

where

\[
|x|_e = \begin{cases} 
0 & \text{if } |x| < \varepsilon, \\
|x| - \varepsilon & \text{otherwise}
\end{cases}
\]

is the so-called \( \varepsilon \)-insensitive loss function, \( C \) is a positive number, and \( H \) is a reproducing Kernel Hilbert space [21] with norm \( \| \|_H \). It can be proved that the function \( \hat{f} \) can be expressed by

\[
\hat{f}(x) = \sum_{i=1}^\infty \omega_i \phi_i(x) + b,
\]

where the set of functions \( \{\phi_i(x)\}_{i=1}^\infty \) is a basis of \( H \).

The following quadratic programming problem is equivalent to (4):

\[
\min_{x \in \mathbb{R}^p} \left\{ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (z_i^* - z_i)(z_j^* - z_j)K(x_i, x_j) - \sum_{i=1}^n z_i^*(y_i - \varepsilon) + \sum_{i=1}^n z_i(y_i + \varepsilon) \right\}
\]

(5)

with the constraints

\[
0 \leq z_i, z_i^* \leq C, \quad i = 1, \ldots, n,
\]

\[
\sum_{i=1}^n (z_i^* - z_i) = 0,
\]

where \( K(x_i, x_j) \) is the reproducing kernel of \( H \). The function \( f(x) \) is then expressed as

\[
f(x) = \sum_{i=1}^n (z_i^* - z_i)K(x, x_i) + b.
\]

It can be demonstrated that the solution of this problem leads to several coefficients \( \beta_i = (z_i^* - z_i) \) equal to zero, and so the data points associated with them are not involved in the last expression. The remaining data points are called support vectors and they contain all the information needed for the approximation of the function \( f \). Notice that if the number of support vectors is small, the calculation time of the method will be reduced.
Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NW</td>
<td>3.12</td>
</tr>
<tr>
<td>LLPR</td>
<td>5.27</td>
</tr>
<tr>
<td>SVM (119 support vectors)</td>
<td>7.33</td>
</tr>
<tr>
<td>SVM (all vectors)</td>
<td>83.8</td>
</tr>
</tbody>
</table>

2.4. Software implementation issues

All methods were implemented as functions in MATLAB 5.3.1 running on Windows NT 4.0. The NW method is very easy to implement and may be efficiently programmed using vectorized programming techniques in MATLAB directly. The LLPR method requires a regression at each function estimate and therefore is computationally demanding. It was found necessary to implement mex files for this method that used C++ optimized code linked to MATLAB for the critical computational steps. In the case of SVM the critical step was the solution of the very large quadratic programming (QP) problem stated in (5). It was found that the optimization functions of MATLAB were too slow and inaccurate for our purposes. Other standard QP packages, such as NAG’s Toolbox and LSSOL, shared similar handicaps or were unsuitable for large QP problems. Finally, an interior point routine, named BPMPD [22], was found to achieve adequate accuracy and speed. The time required for calculating long-term predictions of 1000 points on a Pentium II system (366 MHz) are shown in Table 1.

3. Data acquisition

The same data as in [7] were used for comparing the methods in an objective manner. The EEG recording with spike and wave activity was obtained (at the F3 derivation) from an 11 years old female patient suffering from generalized epilepsy, with absence of seizures. A Medicod-03M digital EEG system [23] was employed for data acquisition. Amplifier bandpass was 0.3–30 Hz. The sampling frequency was 200 Hz. The data segment for analysis (631 points of data length) was selected by visual inspection as being approximately stationary. It contains SW complex around 3 Hz.

Typical SW activity can be shown as a signal containing periodical spikes separated by slow waves. Fig. 1 shows the SW data used in this paper and the corresponding trajectory in a two-dimensional state space (Lorenz plot). From the Lorenz plot the two different shape components of SW are obviously distinguished: the elliptical shape at the periphery corresponds to the spike activity, and the trajectory like a rolled wire in the center of the graph is associated with the slow-wave activity.

4. Estimation of tuning parameters

The determination of suitable values for the tuning parameters is the centerpiece of non parametric modeling. The parameters were estimated both automatically and manually. The outcomes of both
approaches are shown in the following subsections. Nevertheless, some important details about our estimation process should be provided first:

1. A Gaussian radial basis function (RBF) kernel

\[
K(z, x) = \exp \left( -\frac{(z-x)^2}{h} \right)
\]

was used for the three methods under consideration.

2. Because of data set is the same as that in [6] the embedding dimension for comparison of these methods was fixed to 8, as was reported in that paper.

3. In order to evaluate the goodness of fit of all the methods, the recording was split into two parts. The first part (usually called training data) was employed to estimate the parameters for all the methods. The second part (test data) was utilized for testing quality of prediction in the case of SVM method. The length of training data was fixed to 400 \((M = 400)\).

4.1. Automatic estimation of tuning parameters

All the automatic procedures for estimating tuning parameters already described in literature are essentially based on the optimization of one-step-ahead predictions, even when we are looking for good long-term predictions. Two of the most well-known procedures were used here.

A cross-validation (CV) technique [10,24] was considered for NW and LLPR. Assume again that training data are represented as \(\{X, y\} = \{x_i, y_i\}_{i=1}^n\), where \(y_i = x_i \in \mathbb{R}\) are the values of the time series, \(x_i = (x_{i-1}, \ldots, x_{i-m}) \in \mathbb{R}^m\) (the rows of \(X\)) are the points in state space. The CV principle is applied in the following way: each pair \((x_i, y_i)\) is sequentially removed from the complete sample, the method is trained using the remaining data \(\{x_1, y_1, x_2, y_2, \ldots, x_{i-1}, y_{i-1}, x_{i+1}, y_{i+1}, \ldots, x_n, y_n\}\), and then \(f(x_i)\) is obtained. The average square of the resulting residuals is the cross-validation error (CVE). The optimal values for the tuning parameters are obtained by the minimization of the CVE over some set of parameter values.

The CV technique suggested \(h = 217.6931\) and \(799.0824\) for NW and LLPR, respectively. With these tuning parameters neither method generated SW activity (Fig. 2).
For the SVM method the test is carried on by the minimization of a generalization error $e$, defined by

$$e(f) = \sqrt{\frac{1}{N-M} \sum_{i=M+1}^{N} \frac{(x_i - f(x_{i-1}, x_{i-2}, \ldots, x_{i-m}))^2}{\sigma^2}},$$

evaluated on a set of values of the tuning parameters [25]. In this expression, $\sigma^2$ is the variance of the whole time series. The values of the tuning parameters $(C, \varepsilon, h)$ were sought in the three-dimensional grid defined by $C = (1, 10, 100, 500, 1000, 2000, 5000, 10,000, 100,000)$; $\varepsilon = (20, 23, 26, \ldots, 89)$, and $h = (100, 150, 200, \ldots, 2200)$. If $C$ is fixed, the error depends mostly on $h$, because the variation of the error with $\varepsilon$ is much less. Even more, for the region where the minimum values are located the minimization surface became almost flat. Maintaining $C$ as a large number (more than 1000) the surface described by the error is rather flat with respect to both $\varepsilon$ and $h$. Thus, there is no unique minimum point (Fig. 3).

A question arises here: are all the points on the flat region related to values of the tuning parameters with good prediction properties? Several values lying in this region were tested, but the corresponding noise-free realizations did not resemble the SW activity.

These results show that automatic procedures based on one-step-ahead predictions are not effective for long-term prediction of SW data.

4.2. Manual estimation of tuning parameters

Using the automatic estimation techniques mentioned above, the SW activity could not be generated. So, we tried an alternative approach: to manually tune the parameters through trial and error. For discriminating suitable values of the tuning parameters, the attractors of the noise-free realizations must satisfy the following condition: they must resemble SW morphologically (reproducing the average interval between spikes); given any initial point on the training data. Periodic attractors are preferred, for the simplicity of the model. In order to construct the noise-free realizations, 30 initial points (seeds) were selected at random from the training data.
A set of 100 equidistant values of $h$ between 10 and 1000 was considered for the NW and LLPR methods. In the case of NW method, SW activity is clearly reconstructed for $h = 20$ and 30; moreover, these values of $h$ result in inter-spike distances similar to the actual data (see Fig. 4 for the case $h = 30$). The LLPR method never generated SW activity at all.

In the case of SVM we found noise-free realizations resembling SW activity for $C = 1000$ and several values of $\varepsilon$ and $h$. This result is illustrated in Fig. 5 by means of the Lorenz plots. The promising regions were studied more in detail (Fig. 6). There are several values of the tuning parameters for which the method displays periodic behavior similar to actual data. We selected those values of the tuning parameters associated with the minimum number of support vectors. The minimum number of support vectors obtained was 119 (30.4% of the whole training data set) (Fig. 7); corresponding to $C = 1000$, $\varepsilon = 73$, $h = 1044$ (Fig. 8).
For the three methods, the usefulness of long-term prediction of tuning parameters optimized for one-step-ahead prediction was investigated. Previous sections show that the values of the tuning parameters selected by automatic procedures are not always effective for long-term prediction. Now in addition it was obtained that the value of a tuning parameter suitable for long-term prediction is not necessarily effective for obtaining good one-step-ahead prediction. Fig. 9 shows one-step-ahead prediction generated by NW on the test data set (for $h = 30$). It is impossible for the method to predict the spikes. Table 2 summarizes this section.

5. Stability and steady states of the skeletons

Considering the stationary behavior of the data recorded, stability is a very desirable property of the models. To investigate the stability of the estimated skeletons various noise-free realizations were generated, starting from several initial points near to the original seeds. Five values of distance from the seeds (10, 20, 30, 40, 50) were considered. Two initial points were selected randomly for each combination of seed and distance; i.e., 60 initial points were obtained at a certain distance of the seeds. For each of them, the noise-free realization generated by both the NW method ($h = 20$ and 30) and SVM method ($C = 1000$, $\alpha = 73$, $h = 1044$), converges to the same attractor than that generated by the method starting from the corresponding seed. Thus, the skeletons estimated by the two methods are stable for these tuning parameters.
Fig. 6. More in-detail Lorenz plot for SVM, containing estimated parameters ($C = 1000$).

Fig. 7. Obtained support vectors ($C = 1000$, $\varepsilon = 73$, $h = 1044$): (a) actual data; (b) support vectors.

Increasing the distance from the seeds the basin of attraction of the attractors generated by both methods could be explored. Only the NW method generated noise-free realizations that converge to another kind of attractor (specifically, a point attractor). Table 3 shows the percentage of noise-free realizations that resemble SW activity for different values of $h$ and distance. For $h = 30$ the method
generated SW activity for larger distances than for $h = 20$, so the former value of $h$ is selected as the best value for reconstruction of SW dynamics. In addition, 500 initial points were randomly selected from the region $G = \{ (z_1, z_2, \ldots, z_m) : z_i \in [-2000, 2000], i = 1, 2, \ldots, m \}$. For these points the NW method always generates trajectories that vanished to a point attractor with zero value. Thus, it seems that only initial points near actual data can generate SW activity. In case of SVM the attractors remain similar to that shown in Fig. 8; that is, they resembled the SW activity. This result suggests that the basin of attraction of this attractor is larger than that of the NW method, and probably it contains all $G$. 

Fig. 8. SW reconstructed by SVM with manually estimated parameters: (a) noise-free realization ($C = 1000$, $\varepsilon = 73$, $h = 1044$); (b) corresponding Lorenz plot (limit cycle).

Fig. 9. One-step ahead prediction of NW for testing data (dotted line is actual data; solid line is prediction): (a) $h = 217.6731$ (adjusted by CV); (b) $h = 30$ (manually estimated).
Table 2
Parameters selection and corresponding attractors

<table>
<thead>
<tr>
<th>Methods</th>
<th>AP</th>
<th>Values of tuning parameters from AP</th>
<th>Type of attractor of the skeleton from AP</th>
<th>Values of tuning parameters from MP</th>
<th>Type of attractor of the skeleton from MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NW</td>
<td>CV</td>
<td>$h = 217.6931$</td>
<td>point attractor</td>
<td>$h = 20$</td>
<td>Limit cycle similar to actual data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>or $h = 30$</td>
<td></td>
</tr>
<tr>
<td>LLPR</td>
<td>CV</td>
<td>$h = 799.0824$</td>
<td>SW was not generated</td>
<td>Impossible to obtain</td>
<td>SW was never generated</td>
</tr>
<tr>
<td>SVM</td>
<td>MGE</td>
<td>Impossible to obtain</td>
<td>SW was not generated</td>
<td>$C = 1000$</td>
<td>Limit cycle similar to actual data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$e = 73$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$h = 1044$</td>
<td></td>
</tr>
</tbody>
</table>

*AP: automatic procedure; MP: manual procedure; CV: cross-validation technique; MGE: minimization of a generalization error.

Table 3
Rates of appearance of SW-like attractors for NW, using starting points at distinct distances from the seeds (the numbers are in percentages)

<table>
<thead>
<tr>
<th>Distances</th>
<th>75</th>
<th>100</th>
<th>125</th>
<th>150</th>
<th>175</th>
<th>200</th>
<th>225</th>
<th>250</th>
<th>275</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 20$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>36.67</td>
<td>26.67</td>
<td>25</td>
<td>18.33</td>
<td>16.67</td>
<td>8.33</td>
</tr>
<tr>
<td>$h = 30$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>68.33</td>
<td>53.33</td>
<td>48.33</td>
</tr>
</tbody>
</table>

6. Reconstructing the SW dynamics

In the previous sections it was demonstrated that the models fitted by NW and SVM methods generate noise-free realizations which are regular oscillations similar to SW activity. However, the actual SW activity is not a regular oscillation, in the sense that the amplitude of the spikes is variable and the slow waves are noisy. These deterministic models are completely inadequate to describe real SW activity. Therefore, it is valuable to study the effect of the dynamical noise.

In the case of NW, the dynamical noise $e$ in (1) is modeled as a zero mean normal variable with standard deviation $\sigma$. A bootstrapping technique [26] was first applied to estimate $\sigma$, leading to $\sigma = 1.6164$. Using this value, the simulated time series was almost the same as the noise-free realization, which suggests that this is too small a value for $\sigma$. The standard deviation was then adjusted on the basis of the correlation dimension (D2) [27]. D2 of the actual data was estimated as 4.11, so the appropriate standard deviation should be that allowing the simulation of data with D2 near 4.11. The standard deviation of the dynamical noise was increased from 10 to 60 (using
a step size of 5), and simulated realizations were generated 10 times for each standard deviation. Fig. 10 shows the average D2 against $\sigma$. At standard deviation 25, the average D2 was 4.019, the value closest to the D2 of the actual data. Fig. 11 shows simulated realization obtained by applying dynamical noise with this standard deviation, and the corresponding Lorenz plot. Morphologically, it is very similar to the actual data. So, it can be stated that a standard deviation of 25 is suitable for modeling the dynamical noise.

In the case of SVM, the $\varepsilon$-insensitive noise has been suggested as a dynamical noise [28]. The associated probability density function is

$$p(x) = \frac{1}{2(1 + \varepsilon)} \exp(-|x|_\varepsilon).$$

Fig. 12 shows simulated realization with this dynamical noise (with $\varepsilon = 73$). The morphology of SW activity is destroyed (sometimes very atypical oscillations appears). Besides $\varepsilon$-insensitive noise, zero normal noise and uniform noise were also attempted. Though the standard deviation was kept
at a small value, the same negative results were obtained. It can be said that SVM is very sensitive to dynamical noise; therefore, it is not a useful method for realistic simulations.

7. Discussion

The main purpose of this paper is to compare the description of SW activity by three non-linear non-parametric stochastic methods. Points of interest are both methodological as well as about the conclusions that may be made about spike and wave activity itself.

From the methodological side there are three points to be made:

1. As stated before non-parametric methods are just over-fitted models that are regularized, the degree of regularization depending on a “tuning parameter” traditionally adjusted by automatic procedures. Two particular automatic procedures are the ones used in this paper: the cross-validation technique [10,24] for NW and LLPR, and the minimization of a generalization error [25] for SVM (though the CV criterion has been demonstrated to be approximately equivalent to many others [29]). The values of the tuning parameters provided by these automatic procedures were not effective for generating noise-free realizations of SW activity. The reason for this may be that current methods for automatic procedures are based on one-step-ahead predictions while what is of interest is the long-term prediction. We suggest that until new methods are developed tuning parameter selection be manual. In concordance with this conclusion we shall report only the results with manually obtained tuning parameters.

2. The first target in modeling was to determine which types of models allowed stable reconstructions of the deterministic attractor of SW activity. The LLPR was very unstable. This negative result is probably due to overfitting, since it is well known that the LLPR method performs poorly in regions of space that are sparse in data. Using manually estimated values of the tuning parameters, NW and SVM were very stable for the deterministic models (skeletons), resulting in periodic NFR that mimic SW activity.

3. The second target in modeling was to produce mimetic stochastic models for SW. However, the now popular SVM method did not produce stable mimetic responses for the stochastic case,
suggesting that NW method is to be preferred due to its simplicity, speed and stability both for deterministic and stochastic situations. An interesting point related to point 1 above, is the introduction of a measure independent of the one-step-ahead errors in order to determine the size (variance) of the contribution of the noise to SW activity. This was the correlation dimension.

The results obtained are of some interest in terms of the physiology of the SW. Hernandez et al. [7] showed that SW activity could be modeled as the output of a nonlinear system that produces a limit cycle that is subjected to random perturbations. Thus it was shown that a plausible alternative to the chaotic interpretations of SW activity is compatible with the data. The results reported here strengthen this conclusion. Limit cycle behavior occurs for the NFR of a different model than NW — namely radial basis SVM autoregression. This indicates that the identification of this type of deterministic attractor in SW is not overly model dependent but rather a fundamental property of the cortico-thalamic circuits responsible for its generation.

However, we should emphasize a point already made in [6]: only when a physiologically based neural model is constructed for a given type of EEG activity will it be possible to attempt to identify the underlying neural dynamics. In order to explore such model driven approaches the data driven exploratory techniques outlined in this paper may be of use.

8. Conclusions

Non-parametric autoregression is confirmed as a useful model for determining the underlying dynamics of neural systems. Unfortunately, fully automatic procedures are not available for fitting these models for data since, until better methods are developed, it seems to be necessary to manually adjust tuning parameters. It also seems that simple kernel local constant regression models have an edge in terms of simplicity and performance. This work also strengthens the interpretation of epilepsy as the output of a random (non-linear) dynamical system. Noise-free realizations resembling SW activity can be generated by hand tuning the NW and SVM methods. The attractors obtained are limit cycles that resemble SW morphology better than previously reported results, keeping not only the shape but also the correct inter-spike distance. Adding dynamical noise, the NW method was the only one capable of generating SW activity similar to training data; this simulation was also more realistic than former ones. It was shown that the standard deviation of the dynamical noise could be estimated by means of the correlation dimension. However, these results are based on only about 3 s of data from one patient, and need to be confirmed by the application to a broader range of data sets.

9. Summary

EEG spike and wave (SW) activity has been demonstrated to be a highly non-linear phenomenon. Many works suggest the possibility of it being the reflection of deterministic chaotic dynamics, while other papers support an explanation in terms of a non-linear stochastic system. Those who support the first hypothesis have not been able to formulate any chaotic model capable of simulating SW activity. Also their conclusions are based on the use of data analytic techniques that have been
increasingly questioned. On the other hand, the non-linear stochastic point of view has been able to simulate SW activity on the basis of a non-parametric model estimated by the Nadaraya–Watson (NW) method. This paper continues this line of work by comparing the NW approach with two other possible methods of describing non-linear time series that have received wide attention recently. These methods are the local linear polynomial regression (LLPR) and Support vector machines (SVM).

It was demonstrated that noise-free realizations resembling SW activity can be generated by hand tuning the NW and SVM methods. Both methods are stable. The attractors obtained are limit cycles that resemble SW morphology better than as reported in previous works, keeping not only the shape but also the correct inter-spike distance. The tuning parameters had to be estimated manually due to failure of traditional automatic techniques. Adding dynamical noise, the NW method was the only one capable of generating SW activity similar to training data. It was shown that the standard deviation of the dynamical noise could be estimated by means of the correlation dimension.

Acknowledgements

The authors are grateful to the Takahashi Science Foundation for supporting this joint work. Thanks to Prof. Juan Carlos Jimenez (Cuban Neuroscience Center), Prof. Rolando J. Biscay Lirio (Havana University) and Prof. Jose Luis Hernandez (Center for Cybernetics Applied to Medicine, Havana) for their helpful advice. We also want to thank Dr. Vladimir Lopez (Cuban Neuroscience Center) for providing the actual data. The comments of an anonymous referee were very useful in improving this manuscript.

References


Funikazu Miwakeichi is a Ph.D. student at the Graduate University for Advanced Studies, Japan (The Institute of Statistical Mathematics). He received B.S. (Physics) from the Science University of Tokyo in 1996. He also received the M.S. (Material Science) from the Japan Advanced Institute of Science and Technology, Hokuriku. His primary research interests are related to the estimating parametric model of EEG signals in epilepsy through partial linear model and additive model, and special time-series analysis.

Ruben Ramírez-Padron studied Computer Science at the Central University of Las Villas, Cuba. After graduation in 1996 he worked as a computer programmer for the Cuban Ministry of Agriculture at Cienfuegos province. Since 1998 he has been working as a Junior Associated Researcher for the Cuban Neuroscience Center. His current research interests include non-parametric and semi-parametric modeling of epilepsy, time-series analysis and learning machines.


Tohru Ozaki received the B.S. (Mathematics) from the University of Tokyo in 1969. He obtained the degree of Doctor of Science from the Tokyo Institute of Technology in 1981. He is the Director of the Department of Prediction and Control at the Institute of Statistical Mathematics. He is also the Professor of the Department of Statistical Mathematics at the Graduate University of Advanced Studies, Japan. His major research interests include non-linear time-series analysis, stochastic differential equations, non-linear dynamical systems and spatial time-series analysis.