

# A Novel Tool for Sequential Fusion of Nonlinear Features: A Sleep Psychology Application

Mo Chen<sup>1</sup>, David Sommer<sup>2</sup>, Su Lee Goh<sup>1</sup>, Temujin Gautama<sup>3</sup>  
Dragan Obradovic<sup>4</sup>, Martin Golz<sup>2</sup>, Mary Morrell<sup>5</sup>, Hui Wang<sup>4</sup> and Danilo Mandic<sup>1</sup>

**Abstract**—A framework for automated scoring of sleep stages during afternoon naps of healthy humans is introduced. This is achieved by sequential fusion of nonlinear features extracted from three physiological channels: the electroencephalogram (EEG), electrooculogram (EOG) and respiratory trace (RES). These features are generated by means of the recently introduced “Delay Vector Variance” (DVV) method which examines local predictability of a signal in phase space. The analysis is accompanied by a set of comprehensive simulations, supporting the approach.

## I. INTRODUCTION

The goal of the data fusion is to combine data collected from different sensors to make better use of the available information and achieve improved performance which could not be achieved by the use of only a single sensor only [1][2].

Data fusion can occur at a low, intermediate or high level of abstraction, which corresponds to the:- i) **Observation/measurement level**, which contains vectors of measurements [3]; ii) **Feature level**, which seeks features from time/frequency models (FFT, (N)ARMA [4], particle/Kalman filter [5], information flow [6]); iii) **Decision level**, where the classes are mapped onto decisions (SVM [7], kernel ICA).

One of the main problems associated with data fusion for real-world applications is related to combining the information coming from heterogeneous sensors, acquired at different sampling rates and at different time scales. Data/sensor fusion approaches dealing with combining data from homogeneous sensors are normally based either in the time domain, or in some transform domain, for instance on features coming from the frequency representation of signals, their time-frequency, or state-space features [8]. Notice that in this framework we deal with multivariate and multimodal processes, for which either there are no precise mathematical relationships, or if they exist they are too complex. Such is

the case with the segmenting different stages of human sleep, e.g., sleep stage scoring, to automate scoring of sleep stages is a major challenge.

In general, there are two standard approaches to combine multiple electroencephalogram (EEG) and electrooculogram (EOG) signals. In the first approach, or so-called “Raw Data Fusion”, the sensor data are merged without prior preprocessing or dimensionality reduction. Despite its simplicity, the major disadvantage here is the potentially vast amount of data to be handled. In the second approach, the so-called “Feature Fusion”, features extracted from signals coming from different sources and/or extracted by different methods are fused [9][10]. In our experiments, the state space features obtained from the “Delay Vector Variance” (DVV) [11] method from different sensors are investigated to show whether such a fusion of “nonlinear” features yields improvement in automated sleep stage scoring over the standard approaches using only one of the signals.

The recently introduced DVV approach is a signal modality detection method based on the local predictability in the state space. The virtue of the DVV approach is that it can show both qualitatively and quantitatively whether the linear, nonlinear, deterministic or stochastic nature of a signal has undergone a modality change or not. This way, the DVV methodology represents a complement to the widely used linear Power Spectral Density (PSD) estimation. Notice that the estimation of nonlinearity associated with the DVV method is intimately related to non-Gaussianity.

Hence, we also set ourselves to investigate whether and how this information, which cannot be estimated by standard methods focusing on frequency domain, contributes to the discrimination ability, and if so, to estimate its importance level, as compared to the PSD based discrimination. The purpose of this paper is therefore to provide a theoretical and computational framework for DVV functioning as sequential data fusion tool and to show whether such a fusion of signals obtained from different sensors has the potential in multivariate and multimodal signal processing.

The paper is organised in the following manner. In Section II, some background is given and the DVV method is described and examples are given to illustrate the merits of the method. In Section III, the framework of utilising DVV method as a sequential data fusion tool to perform automated sleep stage scoring is described and the experiments are conducted in several steps. In Section IV, the main results presented in the paper are summarised.

<sup>1</sup>Mo Chen, Sulee Goh and Danilo Mandic are with Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London, SW7 2BT, U.K. {mo.chen, sulee.goh, d.mandic}@imperial.ac.uk

<sup>2</sup>David Sommer and Martin Golz are with Department of Computer Science, University of Applied Science Schmalkaden, 98574 Schmalkaden, Germany {dave, golz}@fh-sm.de

<sup>3</sup>Temujin Gautama is with Philips Leuven, Interleuvenlaan 80, B-3001 Leuven, Belgium temujin.gautama@philips.com

<sup>4</sup>Dragan Obradovic and Hui Wang are with Siemens AG, Corporate Technology, Information and Communications, Otto-Hahn-Ring 6, D-81739 Munich, Germany {Dragan.Obradovic, hui.wang}@siemens.com

<sup>5</sup>Mary Marrell is with Royal Brompton Hospital, London, U.K. m.morrell@imperial.ac.uk

## II. “DELAY VECTOR VARIANCE” METHOD

In the following paper, we will refer to four important properties<sup>1</sup> of a signal [11], defined as follows:-

- i) Linear (strict definition) – A linear signal is generated by a linear time-invariant system, driven by white Gaussian noise;
- ii) Linear (commonly adopted) – Definition i) is relaxed somewhat by allowing the distribution of the signal to deviate from the Gaussian one, which can be interpreted as a linear signal from i), measured by a static, monotonic, and possibly nonlinear observation function;
- iii) Nonlinear – A signal that cannot be generated in the above way is considered nonlinear;
- iv) Deterministic (predictable) – A signal is considered deterministic if it can be precisely described by a set of equations;
- v) Stochastic – A signal that is not deterministic.

Several methods for detecting nonlinear nature of a signal have been proposed over the past few years, which include the “*Deterministic versus Stochastic*” (DVS) plots [15], the Correlation Exponent, and “ $\delta$ - $\varepsilon$ ” method [16]. The recently introduced DVV method [11] is shown to be particularly well suited for applications in the signal processing context, since it simultaneously examines both the nonlinear/linear and deterministic/stochastic nature of a signal. This method is based on the local predictability of a signal in the phase space and can be summarised as follows: For an given embedding dimension  $m$ :

- Generate delay vectors (DVs):  
 $\mathbf{x}(k) = [x_{k-m}, \dots, x_{k-1}]^T$  and the corresponding target  $x_k$ ,
- The mean  $\mu_d$  and standard deviation  $\sigma_d$  are computed over all pairwise Euclidean distances between DVs,  $\|\mathbf{x}(i) - \mathbf{x}(j)\|$  ( $i \neq j$ ),
- The sets  $\Omega_k(r_d)$  are generated such that  $\Omega_k(r_d) = \{\mathbf{x}(i) \mid \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_d\}$ , *i.e.*, sets which consist of all DVs that lie closer to  $\mathbf{x}(k)$  than a certain distance  $r_d$ , taken from the interval  $[\max\{0, \mu_d - n_d\sigma_d\}; \mu_d + n_d\sigma_d]$ , where  $n_d$  is a parameter controlling the span over which to perform the DVV analysis,
- For every set  $\Omega_k(r_d)$ , the variance of the corresponding targets,  $\sigma_k^2(r_d)$ , is computed. The average over all sets  $\Omega_k(r_d)$ , normalised by the variance of the time series,  $\sigma_x^2$ , yields the ‘target variance’,  $\sigma^{*2}(r_d)$ :

$$\sigma^{*2}(r_d) = \frac{1}{N} \sum_{k=1}^N \sigma_k^2(r_d) \quad (1)$$

As  $r_d$  increases, the target variance smoothly converges to unity. This is because all DVs start to belong to the

<sup>1</sup>These definitions are from physics literature, and might be different from what we are commonly familiar with. For more information, please refer to [12][13][14].

same universal set, and the variance of targets is equal to the variance of the time series.

To illustrate the meaning of ‘signal nature’, consider a linear benchmark signal (AR(4)), given by [4]

$$x(k) = 1.79 x(k-1) - 1.85 x(k-2) + 1.27 x(k-3) - 0.41 x(k-4) + n(k) \quad (2)$$

and a nonlinear benchmark signal (a Narendra Model Three realisation), given by [17]

$$z(k) = \frac{z(k-1)}{1 + z^2(k-1)} + r^3(k)$$

$$r(k) = 1.79 r(k-1) - 1.85 r(k-2) + 1.27 r(k-3) - 0.41 r(k-4) + n(k) \quad (3)$$

where  $\{n(k)\}$  is white Gaussian noise  $n(k) \in \mathcal{N}(0, 1)$ .

As a result of the standardisation of the distance axis, the resulting ‘DVV plot’ (target variance,  $\sigma^{*2}(r_d)$  as a function of the standardised<sup>2</sup>,  $\frac{r_d - \mu_d}{\sigma_d}$  distance), are easy to interpret, as illustrated in Figure 1(a) and Figure 1(b). The minimal target variance, *e.g.*, the lowest point of the curve, is a measure for the amount of noise which is present in the time series. The presence of a strong deterministic component will lead to small target variances for small spans. At the extreme right, the DVV plots smoothly converge to unity, since for maximum spans, all DVs belong to the same set, and the variance of the targets is equal to the variance of the time series.

In the following step, the linear or nonlinear nature of the time series is examined by performing DVV analyses on both the original and a number<sup>3</sup> of surrogate time series<sup>4</sup>, using the optimal embedding dimension of the original time series. Due to the standardisation of the distance axis, these plots can be conveniently combined within a *scatter diagram*, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, as illustrated by Figure 1(a), the DVV scatter diagram coincides with the bisector line, and the original time series is judged to be linear, as shown in Figure 1(c) (for the linear signal (2)). If not, as illustrated by Figure 1(b), the DVV scatter diagram will deviate from the bisector line and the original time series is judged to be nonlinear, as depicted in Figure 1(d) (for the nonlinear signal (3)).

For a more detailed explanation of how DVV method operates, please refer to Appendix I.

## III. CASE STUDY: AUTOMATIC SLEEP-STAGE SCORING

Before we continue to introduce the novel framework for automated sleep stage scoring, there is a need for us to first describe the physiology data used in our experiments.

<sup>2</sup>Note that we use the term ‘standardised’ in the statistical sense, namely as having zero mean and unit variance.

<sup>3</sup>In all of our simulations, we choose to generate 25 surrogate each time as increasing the number of surrogate will not improve experiment results but increase the computational complexity.

<sup>4</sup>Surrogate data are nonparametric linear versions of the original data. For more detailed information on surrogate data, please refer to [12][13].

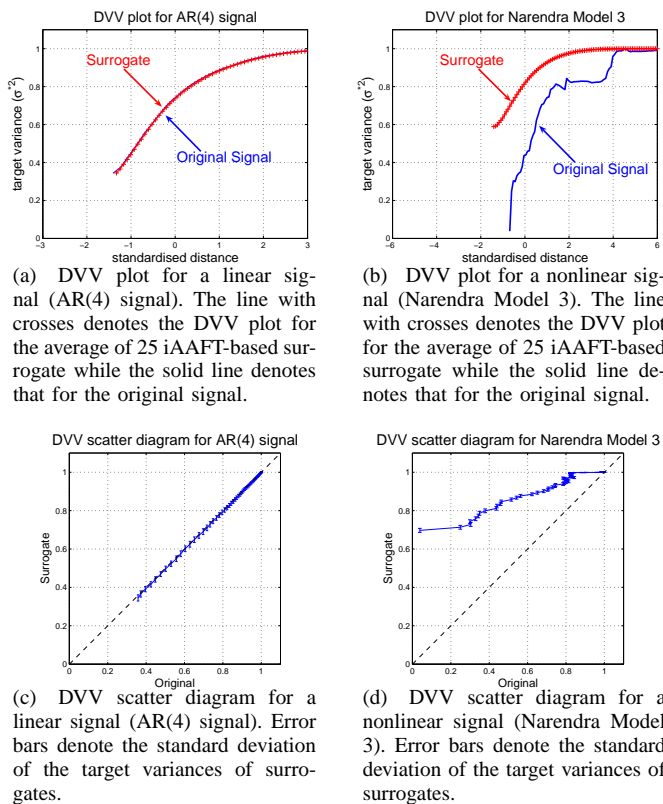


Fig. 1. Nonlinear and deterministic nature of signals.

### A. Data Description

A set of multi-dimensional physiological recordings of five healthy humans during three consecutive afternoon naps were used in our experiment. These data sets are publicly available<sup>5</sup> and have been widely used for automatic sleep-stage scoring purpose [18]. There are three physiological signals for every patient and each nap: the electroencephalogram (EEG), electrooculogram (EOG) and respiratory trace (RES). These three signals are of different length, and carry different dynamic information of the subjects. Manual scoring label by medical experts divides these signals into six classes:

- 1) awake, eye open (W1);
- 2) awake, eye closed (W2);
- 3) sleep stage I (S1);
- 4) sleep stage II (S2);
- 5) no assessment (NA);
- 6) artifact in EOG-signal (AR).

Figure 2 illustrates the time series representation of three signals of the first nap of patient 1 and the manually assigned label for sleep stage. From Figure 2, these three physiological signals have complete different natures, even the sampling rate is not the same. Manually labelling (scoring) all those data sets is a very tedious and error-intolerant process. Therefore, an automated scoring method would be much needed and have been pursued over the past few years [18][19][20].

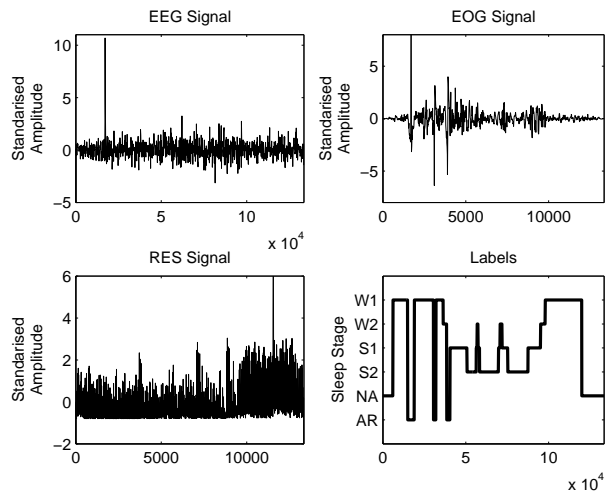


Fig. 2. Physiological data used in the experiments. Top left: Time series representation of EEG signal. Top Right: Time series representation of EOG signal. Bottom Left: Time series representation of RES signal. Bottom Right: Label for sleep stage scores assigned by a medical expert.

### B. Experiment Settings and Simulation Results

Similar to effect of the Fourier transform, the DVV method transfers a signal from time domain into state-space feature domain. In other words, it converts a time series into a small set of data points containing sufficient information to describe the entire dynamics of the original time series. Due to the considerable length of the physiology signals, we divided them into smaller segments using a sliding window with 50% overlapping to avoid the boundary effect. The manually assigned sleep stage labels (by medical experts) are taken as the desired labeling in the training phase. Also, to simplify the experiment, we take the label which occurs most frequently in a window as the correct sleep stage label for that segment window.

The first step of the automated sleep stage scoring process is to extract the feature vectors from each segment window, which is achieved by performing the DVV transform and obtaining the target variance Eq.(1), *e.g.*, DVV values, for each data window. The embedding dimension  $m$ , span  $n_d$  were both set to four, and each initial feature vector was composed of 100 samples, evaluated over the entire span of the standardised distance. As this feature vector still contains redundant information, we perform Principle Component Analysis (PCA) to further reduce the number of dimensions. In fact, we only consider those components that account for 99.9% of the variance, which yields a final set of 10 principle components. Each feature vector is reduced in dimensionality by retaining only the length of the projections onto these 10 principle components. The actual segmentation is then performed by a perceptron, chosen for its robustness and easy implementation without bringing too much additional computation demands. In all simulations, single-layer perceptrons with sigmoid transfer function, followed by winner-take-all step, are trained to converge using the resilient back-

<sup>5</sup><http://www.first.gmd.de/persons/Kohlmorgen.Jens/publications.html>

propagation training algorithm [21].

In the first experiment, we analysed the three signals relating to the first nap of the first patient separately. A wide range of different window sizes were used and for each combination of a window size and signal (EEG or EOG or RES), in order to find an optimal window size for the rest of the experiments. The perceptron served as the classifier, and was trained with different initial seeds. The goal of the training was to correctly classify each segment according to the 6 manually assigned labels. The RES signal exhibited the highest median accuracy (75%) for a window size of 450 samples. The EEG and EOG yielded the highest median accuracy with a window size of 4000 and 300 respectively.

Now we provide an extension of the method used in the previous experiment which allows for an analysis on the fused signal of all the three signals of interest. Since these three signals are of different nature, dynamics and sampling rate, it is not possible to fuse them in the time domain. However, the DVV transformation allows for a unified analysis irrespective of the signal length and recording conditions. Therefore, we perform the sequential data fusion by concatenating the respective DVV feature vector for each data window. Next, PCA-based dimensionality reduction was performed as explained above. Training a perceptron with different initial seeds on these components yielded a median classification accuracy of 88%; a significant 10% increase compared to the best among the individual signal analysis. Figure 3 illustrates the labels assigned by the medical expert (left diagram) and by the perceptron (right diagram) for the first nap of patient 1. From the figure, it is clear that these two labels are a close match.

In the third experiment, we take the analysis one step further by investigating the generalisation ability of the classifiers. Perceptrons, trained on signals from one particular nap, are now used to segment the data from other naps. The data set contains all three signals (EEG, EOG and RES) for two more naps of Patient 1. Poor classification performance was achieved with the highest median accuracy 55%. To show the principle, we make a few simplifications to those classes in the following experiment: All the transients (NA and AR) were abandoned, sleep stages S1 & S2 were grouped into one sleep class and the same for wake stage W1 & W2.

With the new setting, we analysed the signals individually. The perceptrons trained on nap 1 of the training set were next used to predict the labels of nap 2 and nap 3. The highest accuracies were achieved on the EEG signal and are summarised in the first row of Table I. Next, for the analysis on fused data, the respective feature vectors were concatenated, and the result is shown as the third row of Table I, a significant improvement in accuracy as compared to the individual signal analysis was obtained. It is interesting to see whether the fused signal carries sufficient information to assign labels correctly. To this end, we repeated the third experiment with a much more powerful classifier, Support Vector Machine (SVM) to perform the classification procedure, which is an almost unbiased estimator of the true

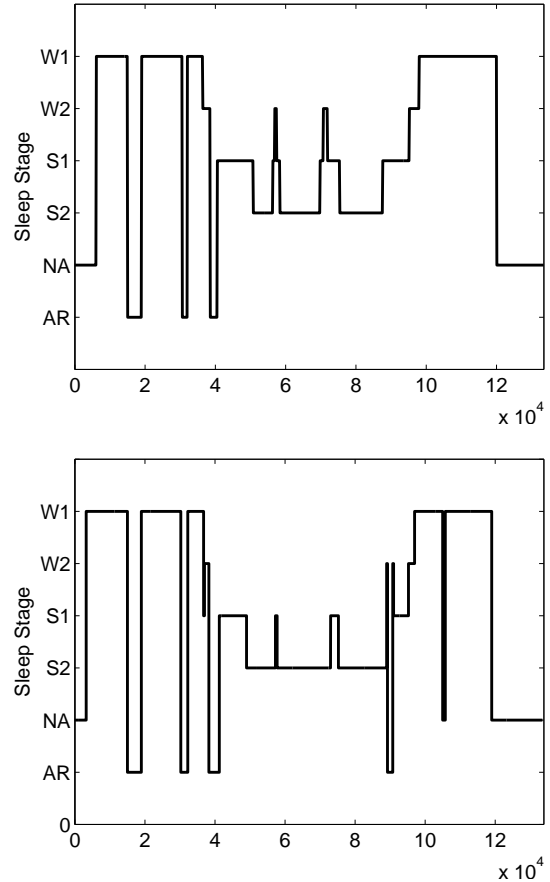


Fig. 3. Labels for sleep stage scoring. Top: True labels assigned by the medical expert. Bottom: Labels assigned by the perceptron

TABLE I  
MEDIAN CLASSIFICATION ACCURACY ON TRAINING AND TEST DATA.

nap	1 (training)	2 (test)	3 (test)	classifier
EEG	92%	73%	79%	perceptron
EEG	93%	76%	81%	SVM
Fused features	98%	77%	83%	perceptron
Fused features	99%	82%	89%	SVM

classification error [22]. With the utilisation of SVM, features obtained from single EEG channel yields improvement over perceptron classifier. The highest classification accuracy was achieved for fused signal, as shown in the bottom row of Table I. This is the demonstration for that the sequential data fusion of DVV features can provide significant improvement in the automated sleep stage scoring.

## IV. CONCLUSIONS

The “Delay Vector Variance” (DVV) method has been employed as a sequential data fusion tool for automated sleeps stage scoring. It has been shown that, in principle, the nonlinear features generated by the DVV method enable the fusion of scoring from multiple and heterogenous sensors, and thereby improvement in the performance. The analysis is supported by extensive simulations on multivariate multimodal physiological recordings and has been shown to yield an close match with the scoring conducted by a medical expert.

### APPENDIX I

#### THE ‘DELAY VECTOR VARIANCE’ (DVV) METHOD

The DVV method can be summarised in the following steps:-

- 1) Generate certain number of iterative amplitude adjust fourier transform (iAAFT) surrogate data of the input signal, *e.g.*, original data;
- 2) Apply DVV method on the original data and obtain the target variance,  $\sigma^{*2}$  and the standerised distance using optimal embedding dimension  $m$ ;
- 3) Apply DVV method onto the group of iAAFT surrogate data using the same parameter settings as the original signal and obtain the the target variance of those surrogate data,  $\{\sigma_{surr,k}^{*2}\}$ ,  $k = 1, \dots, N$  ( $N$  is the total number of the surrogate data);
- 4) Plot  $\sigma^{*2}$  and the mean of  $\{\sigma_{surr,k}^{*2}\}$ , *e.g.*,  $\bar{\sigma}_{surr}^{*2}$  against the standerised distance, and obtain the DVV plot where the minimal target variance indicates the un-predictability;
- 5) Plot  $\sigma^{*2}$  against  $\bar{\sigma}_{surr}^{*2}$ , and obtain the DVV scatter diagram with error bars denoting the standard deviation of  $\{\sigma_{surr,k}^{*2}\}$ , where the nonlinear nature is examined by checking whether DVV scatter diagram lies on the bisector line or not.

This is illustrated in Figure 4.

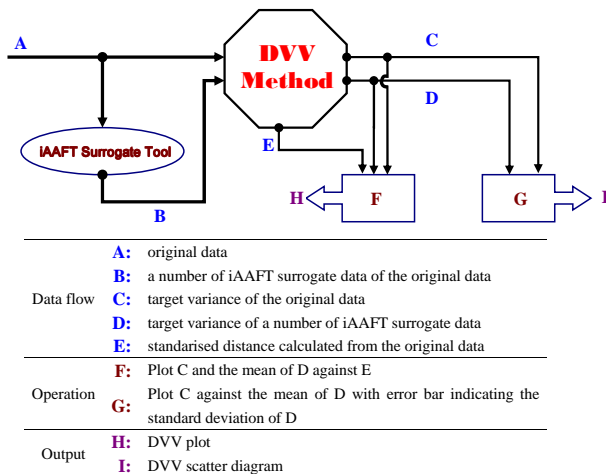


Fig. 4. Block diagram of the DVV method. The table below describes the DVV data flow, operations and output.

## REFERENCES

- [1] D. Mandic, Su lee Goh, and K. Aihara, “Sequential data fusion via vector spaces: Complex modular neural network approach,” *Proceedings of the IEEE Workshop on Machine Learning for Signal Processing*, pp. 147–151, 2005.
- [2] D. Mandic, D. Obradovic, A. Kuh, T. Adali, U. Trutschel, M. Golz, P. De Wilde, J. Barria, A. Constantinides, and A. Chambers, “Data fusion for modern engineering applications: An overview,” *Proceedings of the International Conference on Artificial Neural Networks, ICANN ’05, Warsaw, Poland*, vol. 2, pp. 715–721, 2005.
- [3] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis, second edn.*, JCambridge University Press, 2004.
- [4] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, John Wiley & Sons, 2001.
- [5] G. Deco and D. Obradovic, *An Information-Theoretic Approach to Neural Networks (Perspectives in Neural Computing)*, Springer-Verlag, 1996.
- [6] Deco G. and B. Schuermann, *Information Dynamics: Foundations and Applications*, Springer-Verlag, 2001.
- [7] C. Zhu and A. Kuh, “Sensor network loc. using pat. rec.,” *Proceeding of HISC*, 2005.
- [8] D. Hall and J. Llinas, “An introduction to multisensor data fusion,” *Proceedings of the IEEE*, 1997.
- [9] P. Van Hese, W. Philips, J. De Koninck, R. Van de Walle, and I. Lemahieu, “Microsleep detection in electrophysiological signals,” *Proceedings of the 1st International Workshop on Biosignal Processing and Classification (BPC 2005)*, Dinesh Kant Kumar and Hugo Gamboa (Eds.), pp. 102–109, 2005.
- [10] D. Sommer, M. Chen, M. Golz, U. Trutschel, and D. Mandic, “Fusion of state space and frequency domain features for improved microsleep detection,” *Proceedings of the International Conference on Artificial Neural Networks, ICANN ’05, Warsaw, Poland*, vol. 2, pp. 753–759, 2005.
- [11] T. Gautama, D. P. Mandic, and M. M. Van Hulle, “The delay vector variance method for detecting determinism and nonlinearity in time series,” *Physica D*, vol. 190, no. 3-4, pp. 167–176, 2004.
- [12] T. Schreiber and A. Schmitz, “Surrogate time series,” *Physica D*, vol. 142, pp. 346–382, 2000.
- [13] T. Gautama, M. M. Van Hulle, and D. P. Mandic, “On the characterisation of the deterministic/stochastic and linear/nonlinear nature of time series,” *Technical Report*, 2004.
- [14] T. Gautama, D. P. Mandic, and M. M. Van Hulle, “Indications of nonlinear structures in brain electrical activity,” *Phys. Rev. E*, vol. 67, pp. 046204-1–046204-5, 2003.
- [15] M. C. Casdagli and A. S. Weigend, “Exploring the continuum between deterministic and stochastic modeling,” in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, Eds., pp. 347–367. Reading, MA: Addison-Wesley, 1994.
- [16] D. Kaplan, “Exceptional events as evidence for determinism,” *Physica D*, vol. 73, no. 1, pp. 38–48, 1994.
- [17] K.S. Narendra and K. Parthasarathy, “Identification and control of dynamical systems using neural networks,” *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.
- [18] J. Kohlmorgen, K. R. Muller, J. Rittweger, and K. Pawelzik, “Identification of nonstationary dynamics in physiological recordings,” *Biol. Cybern.*, vol. 83, pp. 73–84, 2000.
- [19] P. Van Hese, W. Philips, J. De Koninck, R. Van de Walle, and I. Lemahieu, “Automatic detection of sleep stages using the eeg,” *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 2, pp. 1944 – 1947, 2001.
- [20] K. Pauwels, T. Gautama, D. Mandic, and M. M. Van Hulle, “Towards mode detection,” *Proceedings of the 4th International Conference on Recent Advances in Soft Computing, RASC2002, Nottingham, U.K.*, pp. 77–78, 2002.
- [21] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1994.
- [22] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer, New York, 1996.