



# Non uniform Embedding based on Relevance Analysis with reduced computational complexity: Application to the detection of pathologies from biosignal recordings

Jorge A. Gómez-García<sup>a,\*</sup>, Juan I. Godino-Llorente<sup>a</sup>, Germán Castellanos-Dominguez<sup>b</sup>

<sup>a</sup> Bioengineering and Optoelectronics (ByO) Group, Universidad Politécnica de Madrid, Ctra. Valencia, km. 7, Madrid 28031, Spain

<sup>b</sup> Signal Processing and Recognition (SPR) Group, Universidad Nacional de Colombia, km. 7 vía al Magdalena, Manizales, Colombia

## ARTICLE INFO

### Article history:

Received 23 February 2012

Received in revised form

26 November 2012

Accepted 14 January 2013

Available online 18 October 2013

### Keywords:

Nonlinear dynamics

Non uniform embedding

Time-delay embedding

## ABSTRACT

Nonlinear analysis tools for studying and characterizing the dynamics of physiological signals have gained popularity, mainly because tracking sudden alterations of the inherent complexity of biological processes might be an indicator of altered physiological states.

Typically, in order to perform an analysis with such tools, the physiological variables that describe the biological process under study are used to reconstruct the underlying dynamics of the biological processes. For that goal, a procedure called time-delay or uniform embedding is usually employed. Nonetheless, there is evidence of its inability for dealing with non-stationary signals, as those recorded from many physiological processes.

To handle with such a drawback, this paper evaluates the utility of non-conventional time series reconstruction procedures based on non uniform embedding, applying them to automatic pattern recognition tasks. The paper compares a state of the art non uniform approach with a novel scheme which fuses embedding and feature selection at once, searching for better reconstructions of the dynamics of the system. Moreover, results are also compared with two classic uniform embedding techniques. Thus, the goal is comparing uniform and non uniform reconstruction techniques, including the one proposed in this work, for pattern recognition in biomedical signal processing tasks. Once the state space is reconstructed, the scheme followed characterizes with three classic nonlinear dynamic features (Largest Lyapunov Exponent, Correlation Dimension and Recurrence Period Density Entropy), while classification is carried out by means of a simple  $k$ -nn classifier. In order to test its generalization capabilities, the approach was tested with three different physiological databases (Speech Pathologies, Epilepsy and Heart Murmurs).

In terms of the accuracy obtained to automatically detect the presence of pathologies, and for the three types of biosignals analyzed, the non uniform techniques used in this work lightly outperformed the results obtained using the uniform methods, suggesting their usefulness to characterize non-stationary biomedical signals in pattern recognition applications. On the other hand, in view of the results obtained and its low computational load, the proposed technique suggests its applicability for the applications under study.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The automatic detection of pathologies is an increasingly important issue that has gained popularity in the last few years. The main goal in this field is to develop computer-aided diagnostic systems, enabling an objective assessment, hence helping diagnosis, reducing the evaluation time, and improving the clinical treatment given to each patient [1].

These automatic detection systems require signals recorded from the biological processes under study, which are further analyzed to make decisions about the state of the patient. For this purpose, the analysis by means of nonlinear dynamic tools has become useful for understanding and characterizing biological processes, among other reasons, due to the high nonlinearity, complexity, and non-stationary behavior of biological signals [2], as well as the known relationship between pathophysiological dysfunction and changes in nonlinear dynamic features [3].

In this regard, the nonlinear analysis of time series depends on the reconstruction of the dynamical behavior of the underlying system, so that the evolution of the system is represented in a  $m$ -dimensional space. The most common technique for this purpose is based on the Takens' Time-Delay Embedding Theorem [4], which

\* Corresponding author. Tel.: +34 913367829.

E-mail addresses: [jorge.gomez.garcia@upm.es](mailto:jorge.gomez.garcia@upm.es) (J.A. Gómez-García), [ignacio.godino@upm.es](mailto:ignacio.godino@upm.es) (J.I. Godino-Llorente), [cgcastellanosd@unal.edu.co](mailto:cgcastellanosd@unal.edu.co) (G. Castellanos-Dominguez).

requires the computation of two parameters: the embedding dimension  $m$  and the time lag  $\tau$ , chosen to optimize the spread of the time series without confusing the underlying dynamics in the  $m$ -dimensional space.

Takens' theorem provides a sufficient condition for assessing the embedding dimension  $m$ , suggesting it to be twice plus one larger than the actual dimension of the system. Since that a priori knowledge of the system's properties is not achievable in most of the cases, other approaches are preferred for estimating the embedding dimension, for example, the coherence of embedded data points [5]. In this respect, one of the most popular methods is the *False Nearest Neighbors* (FNN) method, aiming to search for points in time series being neighbors in the embedding space, but that should not be, since their future temporal evolution is too different [6].

On the other hand, and since the Takens' theorem is stated for ideal time series, for which an infinite amount of data is available and no noise is present, the theorem is silent on how to estimate the time delay  $\tau$ , because in the aforementioned ideal scenario any  $\tau$  value might be valid. However, in real world time series, the choice of the time delay should be carried out thoroughly to assure a proper reconstruction. At this point, several methods have been proposed, the *Average Mutual Information* (AMI) criterion [7] being one of the most prominent. Despite the AMI and FNN criteria are the most utilized in the state of the art, some works have stated that instead of focusing on searching  $\tau$  and  $m$  separately, it might be more effective to consider directly the embedding window as it relates them both [8–11].

Although the uniform embedding technique has been widely used for reconstructing attractors providing successful results in several pattern recognition tasks, some issues still arise regarding the analysis of non-stationary biosignals with multiple periodicities. In fact, when choosing  $\tau$ , a short time lag might be optimal for high frequency components, whereas a long lag might be so for low frequency components and modulations. Therefore, a compromise remains inadequate for both time-scales [12]. This fact is crucial for the characterization of complex and non-stationary signals, such as those produced by biological processes. Bearing that in mind, it has been argued that the *non uniform embedding* [12] might deal with such issues, by replacing the single time delay  $\tau$  with a lag vector  $\vec{\tau}$ .

There are some approaches for searching an optimum lag vector  $\vec{\tau}$  in a non uniform embedding reconstruction, though they have been employed mainly for modeling or forecasting purposes. For instance, an early approach is based on the *Minimum Description Length* (MDL) principle [12], which combines embedding and modeling into a single procedure, on which the quality of the embedding is related to the quality of the whole model. Another approach grounded on the False Nearest Neighbors criterion is also presented in [13], which employs the ratio of false neighbors for increasingly adding time delays until that ratio drops to zero. In another approach [14], a feed-forward neural network trained by structural learning is suggested for finding the optimal set  $\vec{\tau}$ . In [15], two statistics are introduced for evaluating the quality of a non uniform embedding reconstruction. The usage of geometrical criterion for selecting the time lag vector is also described in [16], which aims to spread the attractor in the state space as much as possible, such that the magnitude of the attractor's spreading in the state space is a measure of the embedding quality. This approach can be improved using evolutionary algorithms as discussed in [17]. Lastly, an objective function guiding the search for an optimal state-space reconstruction, in terms of noise amplification and minimization of the complexity of reconstruction, is given in [18].

Nevertheless, all the above referred approaches provide either no clue of how to find the best delay vector, or have a high

computational burden (as in most of cases they need to model, to optimize several parameters, or to employ greedy search algorithms), making them unsuitable for pattern recognition tasks.

To overcome the required high computational costs, providing a proper non uniform embedding within the framework of the nonlinear characterization of biosignals, this work proposes a novel method that attempts, at the same time, minimizing the redundancy and maximizing the relevance as part of the feature selection-like schema. To test the feasibility of the proposed schema, several experiments are performed using different biosignal databases. For characterization purposes, the following nonlinear features have been used: the Correlation Dimension ( $d_2$ ), the Largest Lyapunov Exponent ( $\lambda$ ) [6], and Recurrence Period Density Entropy ( $\rho$ ) [19], while for classification, a simple  $k$ -nn classifier is employed. It must be remarked that the main concern of this paper is not to improve the accuracies found in the state of the art for the aforementioned applications (Speech Pathologies, Epilepsy and Heart Murmurs detection), but to investigate the suitability of the non uniform embedding state space reconstruction technique developed for pattern recognition tasks, comparing with the classic uniform and other non uniform embedding techniques.

This paper is organized as follows: Section 2 presents theoretical aspects of the embedding procedures, as well as the Non Uniform embedding proposal. Section 3 presents the methodology and describes the databases used in this study. Section 4 presents the experiments and the results. Finally, Section 5 presents the discussions, and Section 6 the conclusions of this work.

## 2. Theoretical background

Fig. 1 depicts an outline of the system that has been developed to automatically detect pathologies. The system characterizes different types of biosignals by using nonlinear features extracted from a previous reconstruction of the state space using both uniform and non uniform techniques. The different stages of the process are described next.

### 2.1. State space reconstruction

Roughly speaking, the state space has been reconstructed either using uniform or non uniform techniques. Two uniform embedding methods have been used throughout this work, termed as Uniform 1 and Uniform 2; and two non uniform methods have also been used, termed as Non Uniform 1 and Non Uniform 2. The following presents an introduction to these four methods.

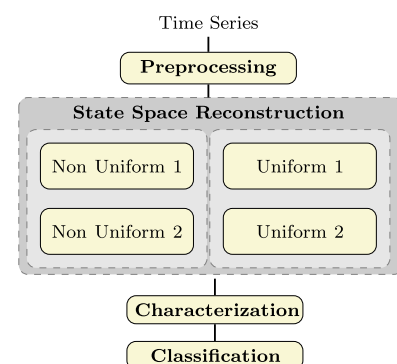


Fig. 1. Diagram of an automatic pathology detection system based on Uniform and Non uniform embeddings.

### 2.1.1. Uniform embedding

Generally, the nonlinear analysis of time series is based on the reconstruction of the system dynamics in a  $m$ -dimensional space, termed *state space*, where  $m$ -dimensional points (*state vectors*) live in. If the system is *dissipative* (on average it contracts as it evolves) the reconstruction aims at finding a closed subset of the state space, called *attractor* of the system, which contains all possible system state vectors and its dynamics. Nonetheless, since a one-to-one reconstruction of the system's original attractor from the underlying time series poses a problem, the Time-Delay Embedding Theorem provides an equivalent attractor to the system's original, holding its same topological properties [4]. The equivalent reconstruction, or *embedding*, defines the state vectors of a time series  $\mathbf{s} = \{x[0], x[1], \dots, x[n]\}$  with length  $n$ , as follows:

$$\vec{\mathbf{x}}[t] = \{x[t], x[t-\tau], \dots, x[t-d_w]\} \quad (1)$$

where the value  $d_w = (m-1)\tau$  is the *embedding window*,  $m$  is the *embedding dimension* or the smallest number of coordinates needed to represent the time series with no overlapping in the state space; and  $\tau$  is the *time lag* (*time delay*) influencing the shape and spread of the reconstructed attractor. Also, since the parameter  $\tau$  remains constant within the whole reconstruction procedure, the resulting embedding is termed *uniform embedding*. Throughout this paper, the classic approach of the uniform embedding employing  $m$  and  $\tau$  is called *Uniform 1*.

An adequate estimation of both embedding parameters,  $m$  and  $\tau$ , is a crucial step to correctly building up the system's attractor. For instance, and assuming a correct choosing of  $m$ , if a small  $\tau$  is employed compared to the timescale of the system, strongly correlated state vectors clustered around the diagonal of the state space will be obtained, i.e., there might be *redundancy* or small information gain. On the opposite, if a large  $\tau$  value is fixed, it may lead to almost independent state vectors spreading all over the state space, losing all the dynamic information contained in the reconstruction, i.e., there is *irrelevance* [6,20].

On the other hand, instead of carrying out each estimation of  $m$  and  $\tau$  separately, other approaches provide a direct computation of the embedding window as it might be more optimal [8–11]. For instance, an approach combining modeling and embedding into a single estimation procedure is discussed in [11]. To this end, the optimal model describing the data is assumed to minimize an information criterion, called *Description Length* (DL). By using local constant models, the description length of the data is then computed for increasing values of  $d_w$ . The minimization of this DL or *Minimum Description Length* (MDL) provides the best embedding window in the DL-sense. Specifically, given a point in the time series  $x[t]$  with reconstructed state vector  $\vec{\mathbf{x}}[t]$ , as well as its successor point,  $x[s+1]$ , where  $s$  is chosen to be the nearest neighbor  $\vec{\mathbf{x}}[s]$  of  $\vec{\mathbf{x}}[t]$ , then an approximation  $x[t+1] \approx x[s+1]$  can be appraised, and thus, the prediction error  $e[\cdot]$  can be determined as follows:

$$e[t+1] = x[t+1] - x[s+1] \quad (2)$$

Thus, the MDL function can be shown to be [11]:

$$DL(\vec{\mathbf{s}}, l_w) \approx \frac{n-l_w}{2} \ln \left( \frac{1}{n-l_w} \sum_{i=l_w+1}^n e^2[i] \right) + \frac{l_w}{2} \left( \frac{1}{l_w} \sum_{t=1}^{l_w} (\vec{\mathbf{s}} - \bar{\mathbf{s}})^2 \right) + l_w + DL(l_w) \quad (3)$$

where  $\bar{\mathbf{s}}$  is the mean of the time series;  $l_w$  is the embedding window to be tested, which varies within the range  $[1, d_w^{\max}]$ , and  $d_w^{\max} = (m-1)\tau$  is an upper limit on the minimization procedure.  $DL(l_w)$  is the description length of the tested window which is  $DL(l_w) = \log(l_w) + \log(\log(l_w)) + \dots$ , where each term is an integer and the last term is zero. Finally, the procedure is summarized as

minimizing the equation (3), by estimating the model prediction error of Eq. (2) for increasing values of  $l_w$ . Lastly, the achieved minimum for a given  $l_w$  will be the optimal embedding window  $d_w$ . From now on, a uniform embedding based on the aforementioned procedure is termed *Uniform 2*.

### 2.1.2. Non uniform embedding

A further improvement might be reached using *non uniform embedding* techniques to characterize nonlinearities. These techniques appraise a *time lag vector*  $\vec{\mathbf{l}} = (l_1, l_2, \dots, d_w)$ , instead of the scalar  $\tau$  to reconstruct the state space. In this manner, Eq. (1) is redefined as follows:

$$\vec{\mathbf{x}}[t] = \{x[t-l_1], x[t-l_2], \dots, x[t-d_w]\} \quad (4)$$

Nevertheless, the non uniform embedding approaches have a high computational burden, restricting their use to small databases, and applications such as modeling and/or predicting time series. Furthermore, for pattern recognition tasks with large databases, the usage of these approaches is not simple.

A suitable non uniform embedding based approach for reducing the computational burden is discussed in [11], which is a direct extension of the estimation procedure of the embedding window based on the MDL. In particular, since the modeling employing the local constant scheme utilizes the model prediction error for selecting the embedding window, the same prediction error might also be suitable for selecting the best lag vector for reconstruction purposes. Since there is a combinatorial explosion in the candidate lag vectors that might be selected, as  $d_w$  increases, the successive inclusion of lags can be considered as a candidate embedding window,  $l_w$ , that is further increased. Particularly, it is assumed that for a  $l_w$ -dimensional embedding, the model includes the lags  $\vec{\mathbf{l}}_w = (l_1, l_2, \dots, l_k)$ , with  $0 \leq l_1 \leq l_2 \leq \dots \leq l_k \leq l_w$ . To determine the lag set for a given  $(l_w+1)$ -dimensional embedding,  $\vec{\mathbf{l}}_{w+1}$ , the performance of the local constant model with lags  $\vec{\mathbf{l}}_p = (l_1, l_2, \dots, l_k, l_w)$  is considered. If  $\vec{\mathbf{l}}_p$  performs better than  $\vec{\mathbf{l}}_w$ , then  $\vec{\mathbf{l}}_p$  is accepted as  $\vec{\mathbf{l}}_{w+1}$ ; otherwise,  $\vec{\mathbf{l}}_w$  is retained as  $\vec{\mathbf{l}}_{w+1}$  [11]. This procedure has to be repeated until reaching a  $l_w^{\max}$  value. From now on this non uniform embedding approach is termed *Non Uniform 1*.

### 2.1.3. Non uniform embedding based on relevance analysis

Since the described non uniform embedding approach (Non Uniform 1) does not consider the whole set of possible relationships among the candidate lag elements, the criterion is sub-optimal to find a lag vector. Therefore, in order to get a more precise searching rather than the mere successive inclusion of lag elements without increasing the computational burden, this work introduces the irrelevance and redundancy of the lags chosen, in a procedure based on a relevance analysis that directly provides an expanded embedding matrix.

For this purpose, an upper limit  $d_w^{\max}$  is chosen with the MDL procedure, and then all lags within  $d_w^{\max}$  are taken into consideration for building the expanded embedding matrix, which constitutes the complete space to be reduced. In this way, the expanded matrix is formed as follows:

$$\mathbf{M} = \begin{bmatrix} x[t] & x[t-1] & \dots & x[t-d_w^{\max}] \\ x[t-1] & x[t-2] & \dots & x[t-(d_w^{\max}+1)] \\ \vdots & \vdots & \ddots & \vdots \\ x[t-(n-d_w^{\max})] & x[t-(n-d_w^{\max}+1)] & \dots & x[t-n] \end{bmatrix} \quad (5)$$

The goal of this expansion is to convert the issue of searching a lag vector into an unsupervised feature selection problem (relevance analysis), where the best  $m$  coordinates in the expanded matrix  $\mathbf{M}$ , represent the lag vector to be found. In addition, and to reduce the complexity of the searching procedure, the embedding dimension  $m$  is fixed in advance. Particularly, the FNN method is employed under the assumption that it is robust enough to clearly distinguish the dynamics of the system.

The feature selection problem, carried out in an irrelevance and redundancy analysis basis, is performed with an unsupervised heuristic algorithm termed UmRMR, *Unsupervised Minimal Redundancy/Maximal Relevance* ( $\mu$ ) [21], that defines the *relevance* ( $\text{Rel}(\cdot)$ ) of a column  $\xi^k = \{\xi_1^k, \xi_2^k, \dots, \xi_u^k\}$  in  $\mathbf{M}$ , as the average mutual information among the whole feature set in  $\mathbf{M}$ :

$$\text{Rel}(\xi^k) = \frac{1}{|\mathbf{M}|} \left( H(\xi^k) + \sum_{\xi^l \in \mathbf{M}, k \neq l} I(\xi^k, \xi^l) \right)$$

where  $|\mathbf{M}|$  is the number of columns in  $\mathbf{M}$ ;  $H(\xi^k)$  is the *entropy* of the column, defined as  $H(\xi^k) = -\sum_i p(\xi_i^k) \log(p(\xi_i^k))$ ; and  $I(\xi^k, \xi^l)$  is the mutual information between the columns  $\xi^k$  and  $\xi^l$ , defined as follows:

$$I(\xi^k, \xi^l) = \sum_{i=1}^u P(\xi_i^k, \xi_i^l) \log \left( \frac{P(\xi_i^k, \xi_i^l)}{P(\xi_i^k)P(\xi_i^l)} \right) \quad (6)$$

where  $P(\xi_i^k, \xi_i^l)$  is the joint probability distribution of observing  $\xi_i^k$  and  $\xi_i^l$ ; and  $P(\xi_i^k)$  and  $P(\xi_i^l)$  are marginal probabilities.

In addition, a measure of *redundancy* ( $\text{Red}(\cdot)$ ) in terms of relevance and conditional entropy  $H(\xi^k|\xi^l)$ , is defined as follows:

$$\text{Red}(\xi^k, \xi^l) = \text{Rel}(\xi^k) - \frac{H(\xi^k|\xi^l)}{H(\xi^l)} \text{Rel}(\xi^l) \quad (7)$$

Then, the relevance analysis scheme is stated as a sequential forward search aiming to find a feature set  $\vec{\mathbf{T}} = \{\xi_{s1}, \xi_{s2}, \dots, \xi_{sm}\}$ , containing the  $m$  least redundant and most relevant features among the  $d_w^{\max}$  columns of matrix  $\mathbf{M}$ . Assuming at iteration  $h-1$  that there exists a set  $\mathbf{S}_{h-1}$  of  $h-1$  selected features as well as a set  $\mathbf{U} = \{\mathbf{M} - \mathbf{S}_{h-1}\}$  of unselected features, then, for choosing the  $h$ -th feature using the  $\mu$  principle, the following formula is employed:

$$\mu = \max_{\xi^k \in \mathbf{U}} \left( \text{Rel}(\xi^k) - \frac{1}{h-1} \sum_{\xi^l \in \mathbf{S}_{h-1}} \text{Red}(\xi^k, \xi^l) \right) \quad (8)$$

where the set  $\mathbf{S}$  of selected features is initialized with the most relevant feature in matrix  $\mathbf{M}$  for iteration  $h=1$ .

Fig. 2 presents an outline summarizing the proposed non uniform embedding schema. Additionally, from now on and for the purpose of this work, the proposed non uniform embedding method based on a relevance analysis is termed *Non Uniform 2*.

## 2.2. Characterization

Once the state space is reconstructed, a characterization stage parametrizes the behavior of the signal under analysis. For this task this paper employs the Correlation Dimension and the Largest Lyapunov Exponent features, since they are the most classical nonlinear dynamic features used in the state of the art. In addition, the Recurrence Period Density Entropy has also been used; this is a relatively new feature, which has demonstrated to be useful in several pattern recognition tasks. A brief description of these features is given next:

1. *Correlation dimension*,  $d_2$ , that quantifies the auto-similarity of an embedded time series [6]. For a given time series of length

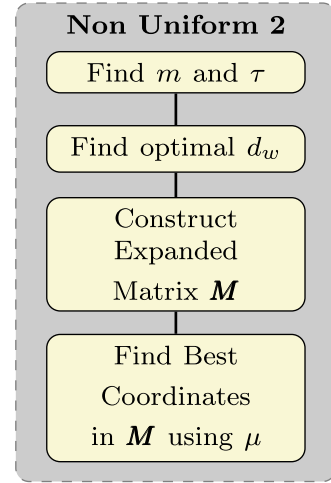


Fig. 2. Diagram of the proposed methodology of non uniform embedding based on relevance analysis for automatic pathology detection of biosignals.

$n$ , the following correlation sum is defined:

$$c = \frac{1}{n} E \left\{ \Theta(\varepsilon - \|\vec{\mathbf{x}}[i] - \vec{\mathbf{x}}[j]\|) : \forall i \neq j, i, j = 1, \dots, n \right\}$$

where  $\Theta$  is the Heaviside function,  $\varepsilon$  is a tolerance measure, and  $\vec{\mathbf{x}}[\cdot]$  are the reconstructed state vectors.  $E\{\cdot\}$  stands for the expectation operator. Since  $\varepsilon \rightarrow 0$ , then  $c = \phi \varepsilon^{d_2}$ , where  $\phi \geq 0$  is a constant. So,  $d_2$  is the estimation of the correlation dimension.

2. *Largest Lyapunov Exponent*,  $\Lambda$ , that measures the divergence of nearby orbits in state space, thus representing the sensitivity to initial conditions as one of the basic attributes of the nonlinear dynamic systems. Value  $\Lambda$  can be estimated as the slope of the function  $Q(\delta)$  exhibiting a robust linear increase within a given range of time evolution,  $\delta$ , as follows [6]:

$$Q(\delta) = E \left\{ \ln \left( \frac{1}{|\mathcal{U}(\vec{\mathbf{x}}[t_0])|} \sum_{\vec{\mathbf{x}}[t] \in \mathcal{U}(\vec{\mathbf{x}}[t_0])} |\vec{\mathbf{x}}[t_0 + \delta] - \vec{\mathbf{x}}[t + \delta]| \right) : \forall t_0 = 1, \dots, N \right\}$$

where  $\vec{\mathbf{x}}[t_0]$  is some initial condition in an appropriate embedding space and  $\mathcal{U}(\vec{\mathbf{x}}[t_0])$  is the neighborhood of  $\vec{\mathbf{x}}[t_0 + \delta]$  with diameter  $\kappa$ . A negative  $\Lambda$  would suggest fixed point attractors. A zero  $\Lambda$  would indicate attractors with stable limit cycles, while positive  $\Lambda$  would suggest noisy signals.

3. *Recurrence Period Density Entropy*,  $\rho$ , that is estimated as the following normalized entropy [19]:

$$\rho = - \frac{T_{\max}}{\ln T_{\max}} E \{ R(T_{t_i}) \ln R(T_{t_i}) : \forall T_{t_i} = 1, \dots, T_{\max} \}$$

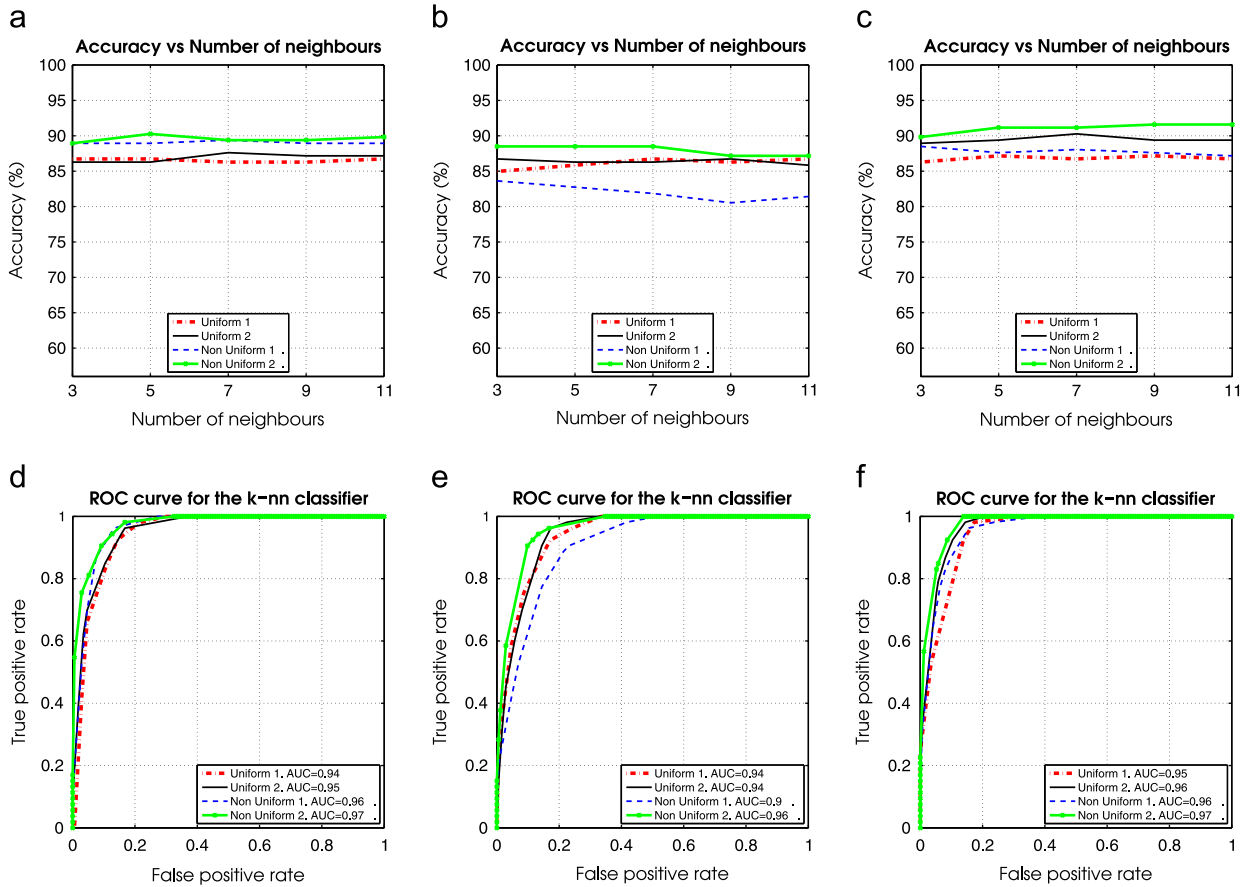
where  $\vec{\mathbf{x}}[t_0]$  is a reference point determined for a given closed ball  $B(\vec{\mathbf{x}}[t_0], \theta)$  with radius  $\theta$  centered in  $\vec{\mathbf{x}}[t_0]$ . As the system evolves, a trajectory is formed from point  $\vec{\mathbf{x}}[t_0]$ . The lasting time during this trajectory for first returning to the ball is called  $t_1$ , serving to define the recurrence time as the difference  $T_{t_0} = t_1 - t_0$ . By finding the recurrence time for all embedded data points, a histogram of recurrence time  $R(T_{t_i})$  is obtained.  $T_{\max}$  equals the maximum recurrence time in the attractor.

## 3. Experimental set-up

### 3.1. Outline of nonlinear analysis methodology

The methodology followed embraces the next four stages (Fig. 1): (a) preprocessing, (b) state space reconstruction, (c) characterization, and (d) classification.





**Fig. 3.** Performance for a  $k$ -nn classifier and Experiment I. (a), (b) and (c) illustrate the classification accuracy for the NDA features,  $\rho$  and Fusion sets respectively; while (d), (e) and (f) present the ROC curve, for NDA,  $\rho$  and Fusion sets, using the  $k$  parameter which produces the highest classification accuracy.

During the *Preprocessing* stage, and to preserve the testing conditions as even as possible, a  $[-1, 1]$  normalization is performed to keep the dynamic range of all the recordings constant.

Within the *State Space Reconstruction* framework, the four aforementioned state space reconstruction schemes are to be tested. For clarity purposes, these are summarized next:

- *Uniform 1* is the classical uniform embedding schema, for which parameter  $\tau$  is calculated by means of the AMI procedure and  $m$  by using the FNN criterion.
- *Uniform 2* employs the embedding dimension  $m$  as given by the FNN criterion, while  $\tau$  is calculated as  $\tau = d_w / (m - 1)$ , where  $d_w$  is as given by the MDL procedure. This experiment is introduced for comparing with the non uniform approaches. It is intended to verify whether the embedding window remains as the only important parameter for reconstruction purposes, or if on contrary a directed search of a lag vector within an embedding window might improve the classification accuracy.
- *Non Uniform 1* is the computational feasible Non Uniform schema in the MDL basis.
- *Non Uniform 2* is the proposed non uniform approach, based on a relevance analysis as shown in Fig. 1.

Once the state space reconstruction is accomplished, and also for comparison purposes, three feature sets were estimated in the *Characterization* stage: The first set, termed NDA, is composed by classical nonlinear dynamic analysis features ( $d_2$  and  $\Lambda$ ); the second set,  $\rho$ , uses the Recurrence Period Entropy; and the third set, called *Fusion* throughout this paper, uses both NDA and  $\rho$  sets

in conjunction. A zero-one normalization was applied to the resulting feature matrix.

Lastly, the *Classification* stage was carried out using a simple  $k$ -nn classifier with the parameter  $k$  ranging from 3 to 11. The rationale to use a simple  $k$ -nn classifier is because the main goal of this work is testing the characterization capabilities of the non uniform embedding schema in pattern recognition applications, rather than improving the classification stage. Thus, using more complex classifiers might obscure the improvements of the methodology on stages different than the state space reconstruction.

In order to minimize the bias of the results, a 11-fold cross-validation methodology was employed. To this end, the database was divided in 11 parts, called folds. 1 fold was chosen for testing, while the remaining 10 were used for training the classifier. This procedure was repeated 11 times, until having chosen all the single folds for testing. The classification accuracy at the *Equal Error Rate* (EER) point, the *sensitivity* (SE) and the *specificity* (SP) were calculated as figures of merit; as well as ROC curves, and estimations of the Area under the ROC curves (AUC) [22].

Sensitivity and specificity are defined as follows:

$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP}$$

where *True Positive*—TP or correctly identified positive instances; *False Positive*—FP or incorrectly identified positive instances; *True Negative*—TN or correctly identified negative instances; and *False Negative*—FN or incorrectly identified negative instances.

Additionally, the ROC curve plots the false positive rate (or 1-SP) vs. true positive rate (or SE) at different threshold levels. The point in the curve on which the false positive rate and the false

negative rate (1-SE) are equal is the EER, and can be used to summarize the performance of the system.

Finally, for the multiclass problem, the employed figure of merit was the *precision*, which for the class  $i$ ,  $PR(i)$ , is defined as follows:

$$PR(i) = \frac{TP(i)}{TP(i) + FP(i)}$$

where  $TP(i)$  and  $FP(i)$  are respectively the true positives and false positives referred to class  $i$ .

**Table 1**

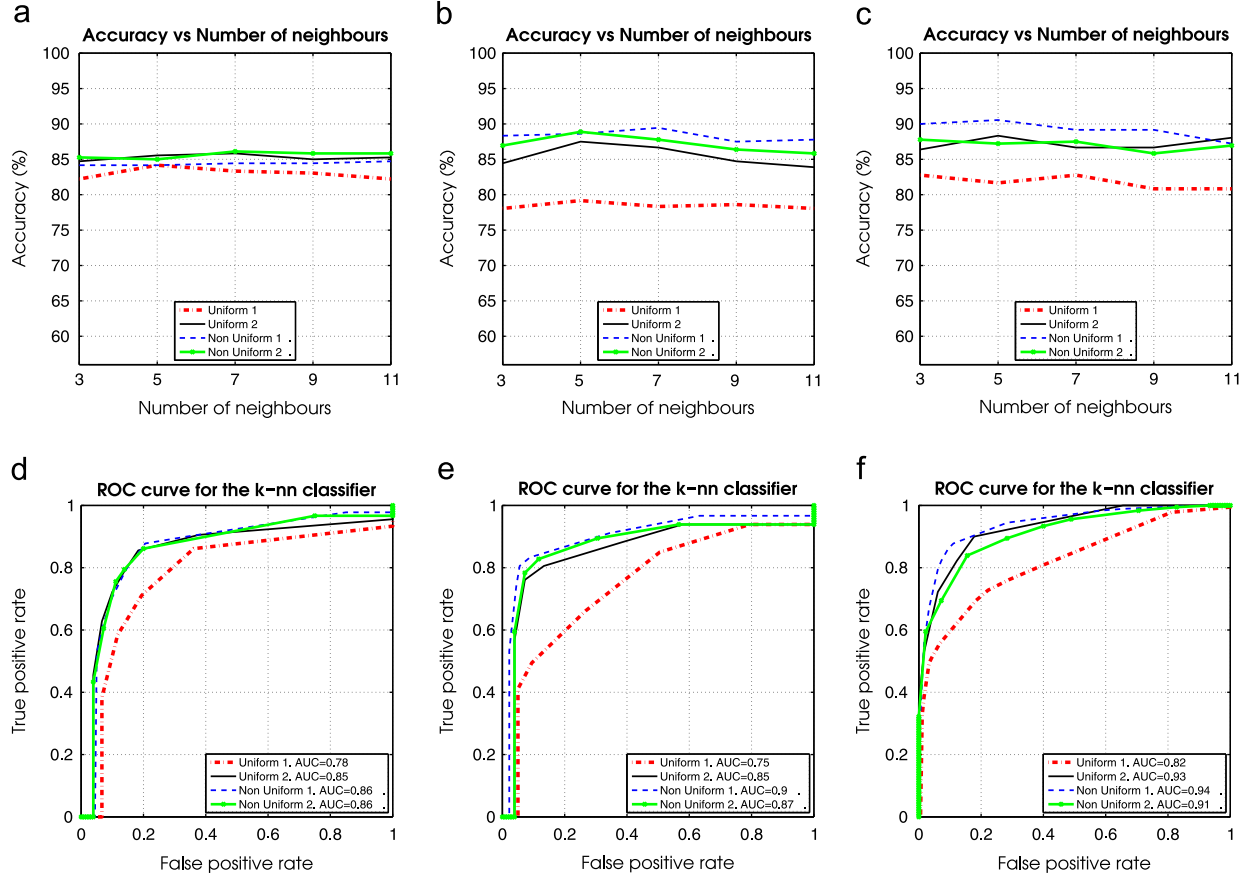
Classification Accuracy, Sensitivity and Specificity for NDA,  $\rho$  and *Fusion* sets, and the four state space reconstruction techniques. Experiment I.

Embedding	Accuracy (%)	SE	SP	AUC
NDA				
Uniform 1	86.73 $\pm$ 4.42	0.85	0.92	0.94
Uniform 2	87.61 $\pm$ 4.29	0.88	0.87	0.95
Non Uniform 1	89.38 $\pm$ 4.01	0.90	0.89	0.96
Non Uniform 2	90.27 $\pm$ 3.86	0.90	0.82	0.97
$\rho$				
Uniform 1	86.73 $\pm$ 4.42	0.87	0.85	0.94
Uniform 2	86.73 $\pm$ 4.42	0.86	0.89	0.94
Non Uniform 1	83.63 $\pm$ 4.82	0.86	0.77	0.90
Non Uniform 2	88.50 $\pm$ 4.16	0.87	0.92	0.96
<i>Fusion</i>				
Uniform 1	87.17 $\pm$ 4.36	0.87	0.87	0.94
Uniform 2	90.27 $\pm$ 3.86	0.90	0.92	0.96
Non Uniform 1	88.50 $\pm$ 4.16	0.89	0.87	0.96
Non Uniform 2	91.59 $\pm$ 3.62	0.91	0.92	0.97

### 3.2. Databases

With the aim to evaluate the generalization capabilities of the proposed methods, three databases of biosignal recordings representing different physiological variables were used. They are briefly presented next:

- *Kay-Elementrics database*: The MEEI voice disorders database [23] contains the registers of the sustained phonation of the /ah/ vowel from patients with a variety of voice pathologies. The registers were previously edited to remove the beginning and ending of each utterance, removing the onset and offset effects in these parts of each utterance. A subset of 173 registers of pathological and 53 normal speakers has been taken according to those enumerated in [24]. As the recordings have different sampling frequencies, all the files were down-sampled to 25 kHz before further processing [25].
- *Heart murmurs database*: The database contains heart sound signals recorded using an electronic stethoscope with a sampling frequency of 44.1 kHz and 16 bits per sample. It is made up of 148 de-identified adult subjects, 50 normal, and 98 exhibiting cardiac murmurs caused by valve disorders (aortic stenosis, mitral regurgitation, etc.). Recordings labeled as normal and those labeled as murmur were separated, extracting 360 individual beats (180 for each class). Those recordings were chosen after a visual and audible inspection by cardiologists, selecting the ones without artifacts and other types of noise that can impair the performance of the algorithms [26].
- *Epilepsy database*: It was collected by the Epilepsy Clinic at the University of Bonn [27], and consists of five sets (denoted A–E), each one composed of 100 segments of EEG channels. Sets A



**Fig. 4.** Performance for a  $k$ -nn classifier and Experiment II. (a), (b) and (c) illustrate the classification accuracy for the NDA,  $\rho$  and *Fusion* sets respectively; while (d), (e) and (f) present the ROC curve, for NDA,  $\rho$  and *Fusion* sets, using the  $k$  parameter which produces the highest classification accuracy.

and B are segments taken from surface EEG recordings (scalp) in five healthy individuals, in waking state with eyes open (A) and eyes closed (B). The sets C, D and E are originated from EEG recordings of pre-surgical diagnosis. Segments in set D were recorded from epileptogenic area, and the set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during interictal intervals, the set E contains only ictal activity.

**Table 2**  
Classification accuracy, sensitivity and specificity for NDA,  $\rho$  and Fusion sets, and the four state space reconstruction techniques. Experiment II.

Embedding	Accuracy (%)	SE	SP	AUC
NDA				
Uniform 1	84.17 $\pm$ 3.77	0.82	0.86	0.78
Uniform 2	85.83 $\pm$ 3.60	0.86	0.86	0.85
Non Uniform 1	84.72 $\pm$ 3.71	0.84	0.85	0.86
Non Uniform 2	86.11 $\pm$ 3.57	0.86	0.86	0.86
$\rho$				
Uniform 1	79.17 $\pm$ 4.19	0.73	0.85	0.75
Uniform 2	87.50 $\pm$ 3.41	0.87	0.88	0.85
Non Uniform 1	89.44 $\pm$ 3.17	0.91	0.88	0.89
Non Uniform 2	88.89 $\pm$ 3.24	0.88	0.89	0.86
Fusion				
Uniform 1	82.78 $\pm$ 3.90	0.79	0.86	0.82
Uniform 2	88.33 $\pm$ 3.31	0.87	0.90	0.93
Non Uniform 1	90.56 $\pm$ 3.021	0.91	0.91	0.94
Non Uniform 2	87.78 $\pm$ 3.38	0.84	0.91	0.91

## 4. Results

Four different experiments are proposed for testing the methodology of Fig. 1.

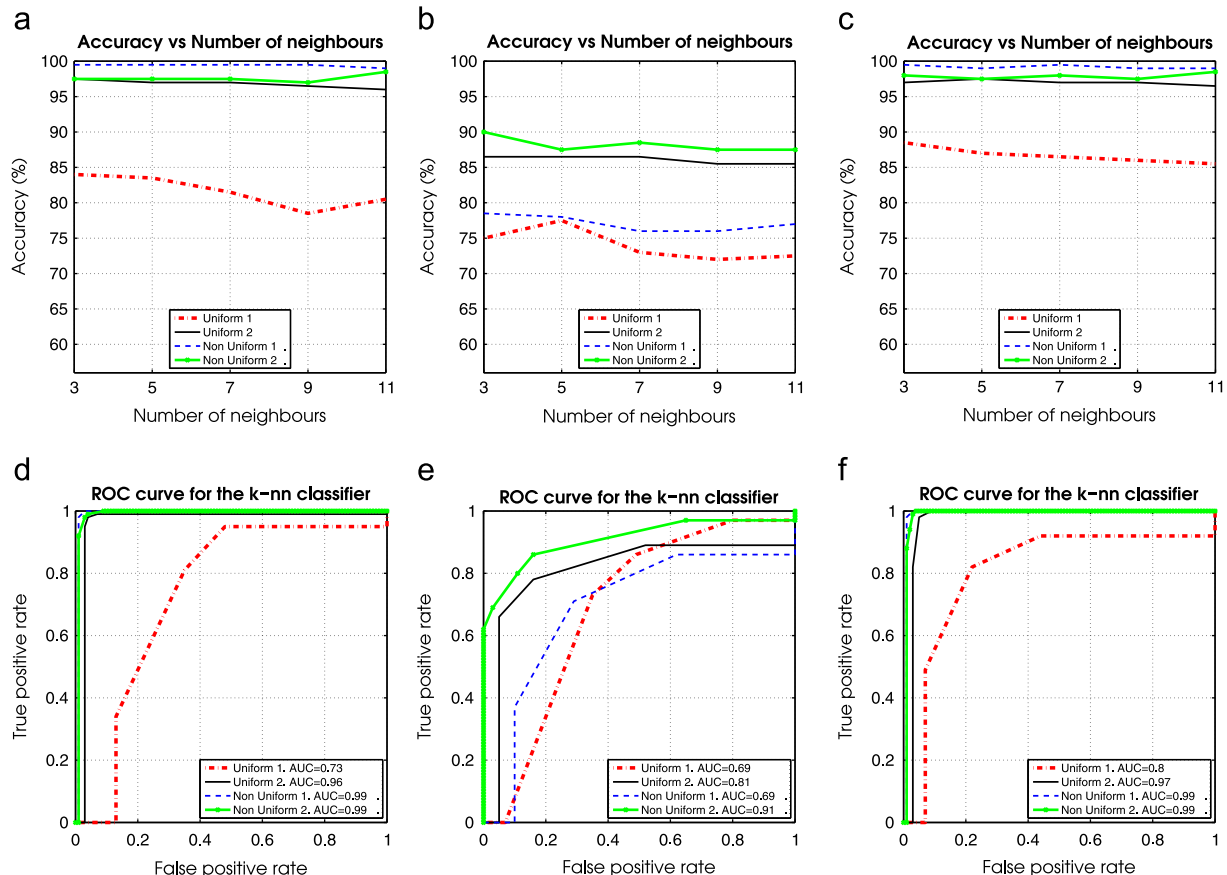
### 4.1. Experiment I

This experiment uses the Kay-Elementrics database. As an additional preprocessing procedure, a short time analysis is performed using 55 ms, with 50% overlapped square windows as suggested in [28]. For this experiment Fig. 3 shows the classification performance using a  $k$ -nn classifier with parameters varying from 3 to 11. The figure also shows the ROC curve for the best  $k$  configuration. Table 1 summarizes the best results.

Non Uniform 2 slightly outperforms the other three methods. As shown in Table 1 the differences are subtle but are maintained in both sets (approximately 4 absolute points in comparison with Uniform 1 and the NDA set, and 3 with respect to the  $\rho$  set). On the other hand, Non Uniform 1 presents a good performance in the NDA set, but the worst in the  $\rho$  set, while both uniform techniques present an alike performance for both tested sets. The Fusion set provides slight improvements in all tested embeddings, but maintaining the same tendency of the NDA set.

### 4.2. Experiment II

This experiment uses the database of heart murmurs. Since each record in the database is composed by a single heart beat, no additional windowing was used. Fig. 4 shows the classification performance for the Experiment II, using a  $k$ -nn classifier with



**Fig. 5.** Performance for a  $k$ -nn classifier and Experiment III. (a), (b) and (c) illustrate the classification accuracy for the NDA,  $\rho$  and Fusion sets respectively; while (d), (e) and (f) present the ROC curve, for NDA,  $\rho$  and Fusion sets, respectively, using the  $k$  parameter which produces the highest classification accuracy.

parameters varying from 3 to 11, as well as the ROC curve obtained for the best configuration of the  $k$  parameter. Table 2 summarizes the best results.

The performance of Uniform 2, Non Uniform 1 and Non Uniform 2 is alike, with subtle improvements given by the Non Uniform 1 technique. On the contrary, Uniform 1 performed the worst, with differences as high as 10 absolute points compared to the technique which performed the best, in the  $\rho$  set.

#### 4.3. Experiment III

This experiment employs the Epilepsy Database, however only the Ictal (E) and Normal (A) sets are considered. The classification performance for the Experiment III, using a  $k$ -nn classifier with parameters varying from 3 to 11 is shown in Fig. 5, as well as the ROC curve for the best  $k$  parameter. Table 3 summarizes the best results.

In this particular experiment there is a significant difference between the classical Uniform 1 embedding technique and the other three embedding procedures. In the NDA set, and in comparison with Non Uniform 1, the difference raised up to 15 absolute points; while in the  $\rho$  set it was up to 12 absolute points in comparison with Non Uniform 2. Non Uniform 2 performed better than both uniform techniques, however being followed by

Uniform 2. In the *Fusion* set, both Non Uniform 1 and 2 performed well, however, being closely followed by Uniform 2.

#### 4.4. Experiment IV

This is the only multiclass experiment carried out in this paper. The Epilepsy database is employed to discriminate between three classes: Normal (A), Interictal (D), and Ictal (E) sets. Classification results are shown in Fig. 6. In this case, and due to the multiclass nature of the problem, ROC curves are not plotted. Fig. 6 (c) presents the classification accuracy for the NDA,  $\rho$  and *Fusion* set, while Table 4 summarizes the performance for the best  $k$  parameter in the  $k$ -nn classifier for the *Fusion* set.

In spite of the low classification accuracy obtained for the  $\rho$  set, the non uniform embedding based techniques provided light improvements in comparison with their uniform counterparts. The same holds for the NDA feature set where both non uniform techniques performed remarkably well. Also, in both cases, the differences between Uniform 1 and Uniform 2 must be pointed out.

On the other hand, when using the *Fusion* set, a better performance is perceived, specially for the non uniform embedding techniques compared to the uniform ones. In the particular case of Uniform 1, the differences were up to 18 absolute points in comparison with the Non Uniform 2.

**Table 3**

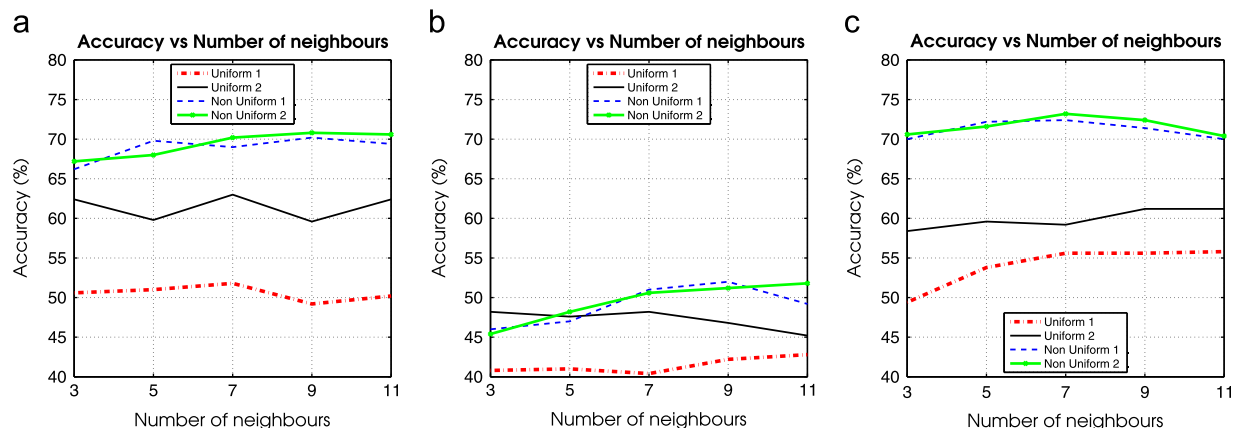
Classification accuracy, sensitivity and specificity for NDA,  $\rho$ , and *Fusion* sets, and the four state space reconstruction techniques. Experiment III.

Embedding	Accuracy (%)	SE	SP	AUC
NDA				
Uniform 1	84.00 $\pm$ 5.08	0.87	0.81	0.73
Uniform 2	97.50 $\pm$ 2.16	0.97	0.98	0.96
Non Uniform 1	99.50 $\pm$ 0	0.99	1.00	0.99
Non Uniform 2	98.50 $\pm$ 0	0.98	0.99	0.99
$\rho$				
Uniform 1	77.50 $\pm$ 5.78	0.82	0.73	0.69
Uniform 2	86.50 $\pm$ 4.73	0.84	0.89	0.81
Non Uniform 1	78.50 $\pm$ 5.69	0.71	0.86	0.69
Non Uniform 2	90.00 $\pm$ 4.10	0.89	0.91	0.91
<i>Fusion</i>				
Uniform 1	88.50 $\pm$ 4.42	0.93	0.84	0.80
Uniform 2	97.50 $\pm$ 2.61	0.97	0.98	0.97
Non Uniform 1	99.50 $\pm$ 0	0.99	1	0.99
Non Uniform 2	98.50 $\pm$ 0	0.98	0.91	0.99

**Table 4**

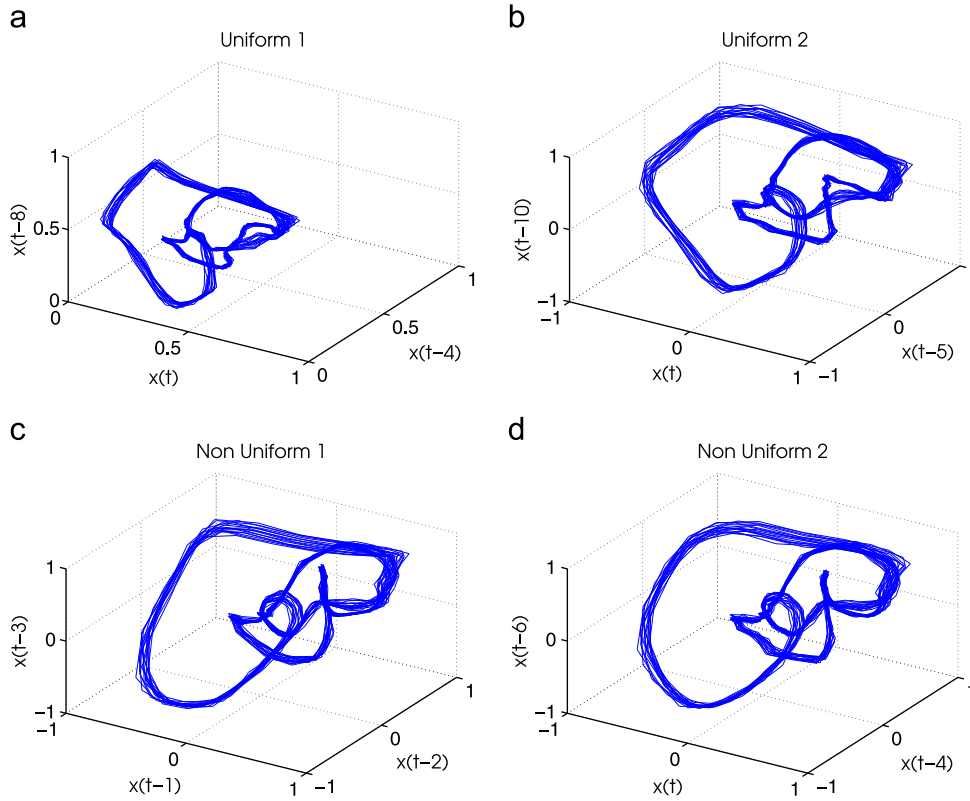
Classification accuracy and precision for NDA and  $\rho$  working in conjunction, and using the four state space reconstruction techniques. Experiment IV.

Embedding	Accuracy (%)	PR (%)			
		E	D	A	
NDA					
Uniform 1	52.80 $\pm$ 4.38	48.93	48.93	53.97	
Uniform 2	61.60 $\pm$ 4.26	54.46	55.13	71.42	
Non Uniform 1	71.60 $\pm$ 3.95	62.81	69.83	78.50	
Non Uniform 2	71.00 $\pm$ 3.97	61.01	69.02	78.79	
$\rho$					
Uniform 1	43.60 $\pm$ 4.35	19.56	42.45	50.24	
Uniform 2	48.80 $\pm$ 4.38	51.32	42.93	52.50	
Non Uniform 1	51.20 $\pm$ 4.38	44.61	49.54	54.79	
Non Uniform 2	54.20 $\pm$ 4.36	50.00	53.67	56.19	
Fusion					
Uniform 1	55.80 $\pm$ 4.35	58.67	50.50	59.66	
Uniform 2	61.20 $\pm$ 4.27	52.27	53.38	76.13	
Non Uniform 1	72.40 $\pm$ 3.92	62.85	69.64	79.90	
Non Uniform 2	73.20 $\pm$ 3.88	64.76	70.83	79.80	



**Fig. 6.** Performance for a  $k$ -nn classifier and Experiment IV. (a) and (b) illustrate the classification accuracy for the NDA features and  $\rho$  feature respectively, while (c) illustrates classification accuracy for the *Fusion* set.





**Fig. 7.** Attractor reconstruction for a voice frame of the Kay-Elementric database. Using (a) Uniform 1 with embedding parameters  $m=4$ ,  $\tau=4$ ; (b) Uniform 2 with embedding parameters  $m=4$ ,  $\tau=5$ ; (c) Non Uniform 1 with embedding parameters  $\vec{T} = (1, 2, 3, 4, 5, 6, 7, 11, 15)$ ; and (d) Non Uniform 2 with embedding parameters  $\vec{T} = (0, 4, 6, 10)$ .

#### 4.5. Attractor reconstruction

In order to demonstrate the effect of the different reconstruction techniques, Fig. 7 illustrates the influence of the embedding methods for the reconstruction of time series. The figure plots the (projected) state space reconstructions using the four embedding schemes, applying them to a single frame belonging to a normal voice of the Kay-Elementrics database.

The graphics show attractors with slight differences among them suggesting that the reconstruction technique has a clear influence on the shape of the attractor.

## 5. Discussions

This paper studies the usefulness of two non uniform embedding state space reconstruction techniques applied to a nonlinear analysis of biological signals for pattern recognition tasks. The performance of these techniques were compared to more classical uniform procedures following five different experimental scenarios. In this context, the paper proposes a novel non uniform embedding method that converts the issue of finding the best delay vector into a relevance analysis problem, for which a heuristic and computationally inexpensive procedure called UmRMR was employed. The usage of such a procedure is grounded on the need of considering nonlinear relationships between coordinates, and of exploiting the concepts of relevance and redundancy.

Since one of the scopes of the paper is to test out if the non uniform embedding might perform better than its uniform counterpart, a simple yet clear methodology is utilized, focusing on the study of the state space reconstruction, rather than in other additional stages. Therefore, the characterization stage is restricted to using only three features:  $d_2$ ,  $\Lambda$  and  $\rho$ . Other classical features,

such as the Approximate [29] or Sample Entropy [30], have not been considered as they rely on a thoroughly parameter tuning, which might in turn distract from the aims of this paper. In addition, and for classification purposes, a simple  $k$ -nn classifier is employed.

The best results were obtained for the Epilepsy database (Experiments III and IV) using the non uniform embedding techniques. This confirms that these techniques are able to characterize the high non-stationarity of the registers in such database. Non uniform techniques also provided subtle improvements for the Heart Murmurs database (Experiment II), where non-stationarity is also an important issue. On the other hand, the non uniform methods also performed well for Experiment I, even when the windowing procedure converts the frames under analysis into quasi-stationary signals.

In the uniform embedding scenario, the lower performance of the method Uniform 1 compared to Uniform 2 is remarkable. As pointed out in [13], Uniform 1 suffers from extra issues due to the employment of the AMI method for finding the time delay: (1) AMI calculates the probability distribution of the Automutual Information function using histograms, making the choosing of the time lag dependent on the data partition; (2) The method is formulated for two dimensional reconstructions and then extended to the multidimensional case; however, it does not necessarily hold in higher dimensions; and (3) The method chooses the first minimum of the mutual information function, having no obvious reasons for this selection.

All above drawbacks, summed to the intrinsic problems of the uniform embedding, affect the reconstruction of the attractor, thus influencing the characterization, and limiting the discriminative power of the nonlinear dynamics features estimated with the Uniform 1 approach. On the contrary, Uniform 2 performed better than Uniform 1, reinforcing the thesis that finding an embedding window for a state space reconstruction is more appropriate than

**Table 5**  
Computation time for different non uniform embedding methodologies.

Method	Time (s)
<b>Non Uniform 2</b>	1.96
Non Uniform 1 [5]	0.35
Ragulsksis [16]	147
Judd 1 [12] <sup>a</sup>	69,154
Judd 2 [12] <sup>b</sup>	1.31
Boccaletti [32] <sup>c</sup>	> 150 times more costly than [12] <sup>b</sup>
Garcia [13] <sup>c</sup>	> 44,000 times more costly than [12] <sup>b</sup>

<sup>a</sup> Using a brute force approach and radial basis functions.

<sup>b</sup> Using reduced autoregressive models and linear basis functions.

<sup>c</sup> According to [31].

estimating the parameters  $\tau$  and  $m$  separately. However, despite it has been argued that the embedding window remains as the only crucial parameter for a correct reconstruction and characterization of the dynamic invariants (as the correlation dimension [10]), a procedure for choosing lags within a given embedding window could even improve the results as verified by the performance improvements (in most of the cases) of Non Uniform 2 in comparison with Uniform 2.

The obtained results suggest that, in most of the cases under study, the non uniform state space reconstruction performs better than the uniform techniques, suggesting their appropriateness for pattern recognition tasks. Besides, and even if the proposed non uniform embedding schema (Non Uniform 2) is surpassed by other techniques in some punctual cases, it provides the most consistent and sustained performance among all the tested techniques.

But the main advantage of this method is that it is computationally inexpensive, assuring its feasibility for pattern recognition applications. To illustrate this, as a matter of example, finding a single lag vector (also including the time for finding the embedding window) took only 1.96 s. for a single frame of 1375 samples belonging to a normal voice signal of the Kay-Elementrics database. These results were obtained using a PC with an Intel Core i3-540 at 3.00 GHz, with 4 Gb of RAM and running Ubuntu 12.10 and Matlab 7.14 R 2012a. For comparison, the naive Non Uniform 1 [5] method took 0.35 s, while the method in [16] took 147 s. On the other hand, the method presented in [12], using a brute force approach and radial basis models, took 69,154 s. However, considering only linear basis functions, and consequently reduced autoregressive models, the computation time was reduced to 1.31 s. This reduced autoregressive model is also utilized in [31] (with certain changes in the objective function), where it is compared to other non uniform embedding techniques. In particular, it was found that the method in [32] was up to 150 times more expensive than the reduced autoregressive model, while the method in [13] was up to 44,000 times more expensive. Table 5 summarizes these results.

It should be remarked, however, that neither the proposed technique (Non Uniform 2), nor the Non Uniform 1 methodologies are optimal for non uniform embedding attractor reconstruction. Particularly, for the proposed method, one of the principal shortcomings is its dependence on an a priori embedding dimension for choosing the number of lags that must be selected. An improper choice of  $m$  might introduce redundancy in the reconstruction, which might affect the quality of the state space reconstruction. The same holds for the embedding window, and therefore for the initialization of the expanded embedding matrix. On the other hand, the Non Uniform 1 criterion is sub-optimal because, in order to assure its computational feasibility, it only testes the successive inclusion of lag elements within an embedding window, whereas all other combinations of lag elements are not investigated. Having those considerations in mind, it is reasonable to hypothesize that a method tackling these drawbacks might in turn improve the performance.

## 6. Conclusions

This paper investigates the usefulness of two non uniform state space reconstruction techniques for pattern recognition tasks, comparing its performance with the classical uniform embedding in several pathology detection experiments. Besides, a novel and computationally feasible non uniform procedure has been presented to obtain the time lag vector needed to reconstruct the state space. According to the obtained results it is reasonable to conclude that the non uniform embedding techniques emerges as appropriate for automatic pathology detection tasks. The results suggest that the proposed technique is appropriate for non-stationary time series, as those recorded from biological processes, mainly due to its ability to choose the different time lags, and therefore to follow the different dynamics present in the bio-signals. The results also suggest that the method is able to improve the quality of the reconstructions and therefore contribute to a better performance in pattern recognition tasks.

Additionally, the following considerations have to be taken into account:

- The results achieved suggest that the embedding window is an important parameter for correctly reconstructing and subsequently characterizing the time series dynamics. In that respect, a methodology which looks directly for the embedding window  $d_w$  instead of separating  $m$  and  $\tau$  parameters might produce an improved performance on pattern recognition tasks.
- The results also suggest that despite the performance improvements of a methodology using embedding windows, one encompassed within a non uniform reconstruction might even improve the state space reconstruction.
- The non uniform embedding seems to work better with signals with prominent non-stationarity than those with soft dynamics.
- Despite the usefulness of the proposed method in pattern recognition tasks, there is no clue whether or not the method might be used for modeling or forecasting purposes. For validating the methodology on these applications, a study must be carried out thoroughly, remaining as future work.

The design of a methodology which automatically searches for the best coordinates, without having to fix an a-priori embedding dimension is still open, and remains as future work. Also, the methodology should be verified in other databases and using other features as, for example, those based on entropy.

## Acknowledgment

This research is carried out under Grants: *TEC2009-14123-C04* from the Spanish Ministry of Education; *AL11-P(I+D)-022* and *Ayudas para la realización del doctorado (RR01/2011)* from Universidad Politécnica de Madrid, Spain; and *Grupo de control y procesamiento digital de señales, código: 20501007205* from Universidad Nacional de Colombia.

## References

- [1] J. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, S. Aguilera-Navarro, P. Gómez-Vilda, An integrated tool for the diagnosis of voice disorders, *Med. Eng. Phys.* 28 (3) (2006) 276–289.
- [2] A.L. Goldberger, L.A.N. Amaral, J.M. Hausdorff, P.C. Ivanov, C.-K. Peng, H.E. Stanley, Fractal dynamics in physiology: alterations with disease and aging, *Proc. Natl. Acad. Sci. USA* 99 (Suppl. 1) (2002) 2466–2472, <http://dx.doi.org/10.1073/pnas.012579499>.
- [3] J.-J. Jiang, Y. Zhang, C. McGilligan, Chaos in voice, from modeling to measurement, *J. Voice* 20 (1) (2006) 2–17.
- [4] F. Takens, Detecting strange attractors in turbulence: dynamical systems and turbulence, Warwick 1980 (1981) 366–381.

- [5] M. Small, Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance (World Scientific Series on Nonlinear Science Series a), World Scientific Pub. Co Inc., 2005.
- [6] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd ed., Cambridge University Press, 2004.
- [7] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A* 33 (2) (1986) 1134.
- [8] J. Gibson, J. Doynne Farmer, M. Casdagli, S. Eubank, An analytic approach to practical state space reconstruction, *Phys. D: Nonlinear Phenom.* 57 (1–2) (1992) 1–30.
- [9] D. Kugiumtzis, State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length, *Phys. D: Nonlinear Phenom.* 95 (1996) 13–28.
- [10] H.S. Kim, R. Eykholt, J.D. Salas, Delay time window and plateau onset of the correlation dimension for small data sets, *Phys. Rev. E* 58 (5) (1998) 5676–5682.
- [11] M. Small, C. Tse, Optimal embedding parameters: a modelling paradigm, *Phys. D: Nonlinear Phenom.* 194 (3–4) (2004) 283–296, <http://dx.doi.org/10.1016/j.physd.2004.03.006>.
- [12] K. Judd, A. Mees, Embedding as a modeling problem, *Phys. D: Nonlinear Phenom.* 120 (3–4) (1998) 273–286.
- [13] S. García, J. Almeida, Nearest neighbor embedding with different time delays, *Phys. Rev. E* 71 (3) (2005) 1–4, <http://dx.doi.org/10.1103/PhysRevE.71.037204>.
- [14] Y. Manabe, B. Chakraborty, A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network, *Neurocomputing* 70 (7–9) (2007) 1360–1371, <http://dx.doi.org/10.1016/j.neucom.2006.06.005>.
- [15] L. Pecora, L. Moniz, J. Nichols, T. Carroll, A unified approach to attractor reconstruction, *Complex Dynamics in Physiological Systems: From Heart to Brain*, vol. 20, 2009, pp. 3–19.
- [16] M. Ragulskis, K. Lukoseviciute, Non uniform attractor embedding for time series forecasting by fuzzy inference systems, *Neurocomputing* 72 (10–12) (2009) 2618–2626, <http://dx.doi.org/10.1016/j.neucom.2008.10.010>.
- [17] K. Lukoseviciute, M. Ragulskis, Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems, *Neurocomputing* 73 (10) (2010) 2077–2088.
- [18] L. Uzal, G. Grinblat, P. Verdes, Optimal reconstruction of dynamical systems: a noise amplification approach, *Phys. Rev. E* 84 (1) (2011) 016223.
- [19] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, I.M. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection, *Biomed. Eng. Online* 6 (2007) 23, <http://dx.doi.org/10.1186/1475-925X-6-23>.
- [20] M. Rosenstein, J. Collins, C. De Luca, Reconstruction expansion as a geometry-based framework for choosing proper delay times, *Phys. D: Nonlinear Phenom.* 73 (1–2) (1994) 82–98.
- [21] J. Xu, Adapt the mRMR criterion for unsupervised feature selection, *Adv. Data Mining Appl.* (2010) 111–121.
- [22] T. Fawcett, Roc graphs: notes and practical considerations for researchers, *Mach. Learn.* 31 (2004) 1–38.
- [23] Massachusetts Eye and Ear Infirmary, Voice disorders database, version.1.03 [cd-rom], Kay Elemetrics Corp., Lincoln Park, NJ, 1994.
- [24] V. Parsa, D. Jamieson, Identification of pathological voices using glottal noise measures, *Lang. Hear. Res.* 43 (2) (2000) 469.
- [25] N. Saenz Lechon, J. Godino Llorente, V. Osma Ruiz, P. Gomez vilda, Methodological issues in the development of automatic systems for voice pathology detection, *Biomed. Signal Process. Control* 1 (2) (2006) 120–128, <http://dx.doi.org/10.1016/j.bspc.2006.06.003>.
- [26] E. Delgado-Trejos, A. Quiceno-Manrique, J. Godino-Llorente, M. Blanco-Velasco, G. Castellanos-Dominguez, Digital auscultation analysis for heart murmur detection, *Ann. Biomed. Eng.* 37 (2) (2009) 337–353.
- [27] R.G. Andrzejak, F. Mormann, G. Widman, T. Kreuz, C.E. Elger, K. Lehnertz, Improved spatial characterization of the epileptic brain by focusing on nonlinearity, *Epilepsy Res.* 69 (1) (2006) 30–44, <http://dx.doi.org/10.1016/j.eplepsyres.2005.12.004>.
- [28] J.D. Arias-Londoño, J.I. Godino-Llorente, G. Castellanos-Dominguez, Short time analysis of pathological voices using complexity measures, in: *Proceedings of Third Advanced Voice Function Assessment International Workshop*, 2009.
- [29] S. Pincus, I. Gladstone, A regularity statistic for medical data analysis, *J. Clin. Monit.* (1991) 335–345.
- [30] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol. Heart Circ. Physiol.* 278 (6) (2000) H2039–H2049.
- [31] Y. Hirata, H. Suzuki, K. Aihara, Reconstructing state spaces from multivariate data using variable delays, *Phys. Rev. E* 74 (2) (2006) 026202.
- [32] S. Boccaletti, D.L. Valladares, L.M. Percora, H.P. Geffert, T. Carroll, Reconstruction embedding spaces of coupled dynamical systems from multivariate data, *Phys. Rev. E* 65 (2002) 035204.



**Jorge Andrés Gómez-García** received his Bachelor and Master degrees from Universidad Nacional de Colombia, Manizales in 2008 and 2010 respectively. Currently he is pursuing a Ph.D. degree from Universidad Nacional de Colombia and Universidad Politécnica de Madrid, Spain. His research interests include signal processing, Nonlinear Dynamics Analysis, Complexity Analysis and Information Theoretical Learning with biomedical applications.



**Juan Ignacio Godino-Llorente** was born in Madrid, Spain. He received the M.Sc. degree in communications engineering in 1996 and the Ph.D. degree with honors in Computer Science from the Universidad Politécnica de Madrid, Spain, in 2002. Currently he is a Full Professor at the Universidad Politécnica de Madrid. Also, from 2005 to 2009, he was the Head of the Circuits and Systems Engineering Department, which belongs to the same University. His main research areas are in the fields of biomedical signal and image processing, rehabilitation technologies, e-health, and ambient intelligence. He has led a large amount of projects obtained in competitive calls developed in collaboration with other institutions and with the industry.



**Germán Castellanos-Dominguez** received the undergraduate degree in radiotechnical systems as well as the Ph.D. degree in processing devices and systems from the Moscow Technical University of Communication and Informatics, Moscow, Russia, in 1985 and 1990, respectively. He is currently a Professor in the Department of Electrical, Electronic and Computer Engineering, Universidad Nacional de Colombia, Manizales, where he is also the Director of the Signal Processing and Recognition Group. His current research interests include Subspace learning, Dynamic feature extraction, Biomedical stochastic processes, Time–frequency representations, and statistical methods for short datasets.