DOI: 10.1109/TBME.2011.2167971

Automated Prediction of the Apnea-Hypopnea Index from Nocturnal Oximetry Recordings

J. Víctor Marcos, *Student Member*, *IEEE*, Roberto Hornero, *Senior Member*, *IEEE*, Daniel Álvarez, *Student Member*, *IEEE*, Mateo Aboy, *Member, IEEE*, and Félix Del Campo

***Abstract*—Nocturnal polysomnography (PSG) is the gold-standard for sleep apnea-hypopnea syndrome (SAHS) diagnosis. It provides the value of the apnea-hypopnea index (AHI), which is used to evaluate SAHS severity. However, PSG is costly, complex, and time-consuming. We present a novel approach for automatic estimation of the AHI from nocturnal oxygen saturation (SaO_2) recordings and the results of an assessment study designed to characterize its performance. A set of 240 SaO_2 signals was available for the assessment study. The data was divided into training (96 signals) and test (144 signals) sets for model optimization and validation, respectively. Fourteen time-domain and frequency-domain features were used to quantify the effect of SAHS on SaO_2 recordings. Regression analysis was performed to estimate the functional relationship between the extracted features and the AHI. Multiple linear regression (MLR) and multilayer perceptron (MLP) neural networks were evaluated. The MLP algorithm achieved the highest performance with an intraclass correlation coefficient (ICC) of 0.91. The proposed MLP-based method could be used as an accurate and cost-effective method for SAHS diagnosis in the absence of PSG.**

***Index Terms*—Sleep apnea-hypopnea syndrome, oxygen saturation, apnea-hypopnea index, regression analysis, multiple linear regression, multilayer perceptron neural networks.**

I. Introduction

THE sleep apnea-hypopnea syndrome (SAHS) is characterized by repetitive complete (apnea) or partial (hypopnea) collapse of the upper airway during sleep [1]. Apnea events are associated to hypoxemia, heart rate variations and arousals. Epidemiological data support the finding that SAHS may have a role in the initiation or progression of diverse respiratory, cardiovascular, and cerebrovascular diseases [1, 2]. The incidence of SAHS has been estimated at 5% of adults in western countries [2]. Furthermore, previous studies revealed that a high percentage of patients (82% of men and 93% of women) with moderate or severe SAHS might remain undiagnosed [3]. Therefore, early detection and treatment of SAHS are required in order to prevent long-term effects and end-organ damages.

Nowadays, nocturnal polysomnography (PSG) is the gold-standard for SAHS diagnosis [4]. PSG studies are performed in special sleep units and generally involve monitoring several physiological recordings such as electrocardiograms (ECG), electroencephalograms (EEG), electromyograms (EMG), electrooculograms (EOG), airflow signals, respiratory effort, and oxygen saturation (SaO_2) or oximetry [1]. Polysomnographs are usually provided with specific software to assist medical doctors in the interpretation of these signals. Although it may be used during the examination of PSG data, manual analysis performed by a sleep specialist is required for accurate identification of apnea/hypopnea episodes. The number of detected events is divided by the hours of sleep to compute the apnea-hypopnea index (AHI), which is used to assess SAHS severity [4]. However, PSG studies have drawbacks since they are costly, time-consuming, and require subjects to be overnight in a special medical facility [5]. Additionally, the demand for PSG studies is progressively growing as people and clinicians are becoming aware of SAHS whereas the available infrastructure is insufficient to support it [6]. Consequently, research focused on alternative diagnostic methods that overcome some of the limitations associated to PSG has notably increased.

New techniques for simplified SAHS detection have been commonly based on the analysis of a reduced set of data. The utility of clinical and demographic variables [7, 8] as well as ECG [9, 10] has been widely studied. In the context of this problem, SaO_2 signals recorded through nocturnal pulse oximetry are of special interest since they can be easily acquired and enable for portable monitoring [11-18]. Pulse oximetry is a non-invasive technique used to monitor arterial blood oxygen saturation. Oximetry recordings contain essential information about SAHS and play a crucial role to interpret PSG studies. Apneas and hypopneas are usually accompanied by marked desaturation events due to the lack of airflow. As a result, patients with SAHS tend to present unstable SaO_2 signals due to frequent drops in the saturation

Manuscript received May 28, 2011. This work was supported in part by the Ministerio de Ciencia e Innovación and FEDER under Grant TEC 2008-02241, and in part by the grant project from the Consejería de Sanidad de la Junta de Castilla y León GRS 337/A/09. *Asterisk indicates corresponding author*.

∗J.V. Marcos is with the Biomedical Engineering Group, E.T.S.I. Telecomunicación, University of Valladolid, Valladolid 47011, Spain (e-mail: jvmarcos@gmail.com).

R. Hornero and D. Álvarez are with the Biomedical Engineering Group, E.T.S.I. Telecomunicación, University of Valladolid, Valladolid 47011, Spain (e-mail: robhor@tel.uva.es; dalvgon@gmail.com).

M. Aboy is with the Electrical Engineering Department, Oregon Institute of Technology, Portlan, OR, U.S.A. (e-mail: mateoaboy@ieee.org)

F. Del Campo is with the Hospital Universitario Río Hortega, Valladolid 47012, Spain (e-mail: fsas@telefonica.net).

value. A different behavior tends to be observed in healthy patients. Their recordings reflect normal ventilation, which corresponds with a saturation value near 96% and the absence of repetitive abrupt changes in the SaO_2 profile [19].

Several signal processing techniques have been proposed to extract relevant diagnostic information from SaO_2 recordings. Preceding studies showed that spectral and non-linear features can reflect the occurrence of apneas [11, 13, 21]. Moreover, highly accurate diagnostic models based on logistic regression [17] or neural networks [14, 15] have been built from these features. However, these algorithms only provide a categorical decision for each subject (labeled as non-SAHS or SAHS) and no information about SAHS severity is given. Automatic computation of the AHI has been addressed to obtain a more detailed characterization of the patient's state. In addition to clinical and demographic variables [7, 8], SaO_2 signals have been frequently used for this purpose. The linear relationship between the number of desaturation events and the AHI was previously evaluated [22, 23]. Multivariate pattern analysis techniques were also applied in order to deal with several measurements from oximetry simultaneously [12, 24]. Finally, the combination of ECG and SaO_2 features has been also previously studied [9]. However, these previous studies did not evaluate the utility of the estimated AHI to rank SAHS severity and only correlation analysis was performed.

In the present study, we hypothesize that an accurate estimation of the AHI can be automatically derived from SaO_2 data, providing more useful diagnostic information to practitioners compared with other methods based on a two-class classification approach. We propose modeling SAHS diagnosis as a regression task. Time-domain and frequency-domain features from oximetry data were used to reflect the occurrence of apneas and hypopneas during sleep. An approximation to the functional relationship between the extracted feature pattern and the AHI was derived from multivariate regression analysis. Multiple linear regression (MLR) models and multilayer perceptron (MLP) neural networks were used for this purpose [25]. The overall objective was to evaluate the degree of severity of SAHS (no SAHS, mild-SAHS, moderate-SAHS and severe-SAHS) from the estimated AHI. As a result, a more complete description about SAHS is provided as compared to the two-class classification approach.

II. Subjects and Data

A group of 240 subjects suspected of suffering from SAHS were included in the study. All of them presented typical symptoms such as sleepiness, snoring, or apnea events reported by the subject or a bedmate. For a decision threshold of AHI = 10 h^{-1}, a positive diagnosis of SAHS would correspond to 160 of these subjects while the remaining 80 subjects would be non-SAHS cases.

The initial population was randomly divided into training and test sets. The training set was composed of 96 subjects and was used for model optimization. Once this process is completed, a test set composed of previously unseen samples is required to objectively assess the performance of the estimator. In this assessment study, the oximetry signals from 144 subjects composed the test set. Table I summarizes clinical data of subjects in both sets.

TABLE I
Clinical and Demographic Data for Training and Test Sets

	Training Set	Test Set
Subjects	96	144
Age (years)	52.35 ± 13.76	52.19 ± 13.73
Males (%)	77.08	77.78
BMI (kg/m^2)	29.83 ± 4.17	29.83 ± 4.53
AHI (h^{-1})	24.75 ± 25.19	26.39 ± 26.74

Each subject underwent complete overnight PSG. It was carried out from midnight to 08:00 AM in the Sleep Unit of Hospital Universitario Río Hortega, Valladolid, Spain. The Review Board on Human Studies approved the protocol and subjects gave their consent to participate in the study. Patients were monitored using a polysomnograph (Alice 5, Respironics, Philips Healthcare, The Netherlands). ECG, EEG, chin EMG, EOG, nasal airflow and body position were recorded and stored on a computer. Simultaneously, a Nonin PureSAT pulse oximeter (Nonin Medical Inc., USA) was used to record SaO_2 signals at a sampling frequency (f_s) of 1 Hz. The averaging time was set to 3 seconds, following the recommendations from the American Academy of Sleep Medicine [26]. Oximetry recordings were saved to separate files to be off-line processed. Artifacts represented by drops to zero were removed. Finally, a sleep specialist analyzed the complete set of recordings using the rules proposed by Rechtschaffen and Kales [27] and derived the AHI for each subject.

III. Methods

The proposed method comprises two different stages. In the first one, feature extraction from SaO_2 data is carried out in order to capture the dynamical behavior of the signal. The second stage corresponds to regression analysis, which aims to provide an analytical expression for the AHI as a function of the extracted features.

A. Feature extraction

In the feature extraction phase, information in the SaO_2 recording was summarized into a reduced set of measurements or features. They are defined in order to represent different signal properties related to the degree of SAHS severity.

Prior domain knowledge about the influence of apnea events on SaO_2 dynamics was used to define a set of 14 measurements. According to the domain used for SaO_2 analysis, the extracted features were divided into two groups: time-domain and frequency-domain features.

Time-domain analysis

Marked drops in the amplitude of oximetry signals reflect desaturation events due to apneas. Subjects with low AHI are expected to present SaO_2 tracings with minor oscillations

around 96% during most of the time [19]. In contrast, a high AHI reflects the repetition of apneas, resulting in SaO_2 recordings with marked instability. Conventional statistics and non-linear methods were used to characterize this dynamic behavior in the time domain.

Statistical analysis represents an easy-to-use tool to study SaO_2 signals. The distribution of SaO_2 values tends to reflect different properties depending on the AHI. Mean (μ_S), variance (σ_S), skewness (γ_S) and kurtosis (δ_S) were computed to quantify the central tendency, the degree of dispersion, the asymmetry and the peakedness, respectively, for variable *s* representing the SaO_2 value. These measurements are defined as [28]:

$$\mu_S = \sum_k s_k \, p_S(s_k), \tag{1}$$

$$\sigma_S^2 = \sum_k \left[s_k - \mu_S\right]^2 p_S(s_k), \tag{2}$$

$$\gamma_S = \left(1/\sigma_S^3\right) \sum_k \left[s_k - \mu_S\right]^3 p_S(s_k), \tag{3}$$

$$\delta_S = \left(1/\sigma_S^4\right) \sum_k \left[s_k - \mu_S\right]^4 p_S(s_k), \tag{4}$$

where p_S denotes the probability density function of variable *s*. It was obtained from the relative frequency observed in the sequence of SaO_2 samples $\mathbf{s} = \{s_1, \ldots, s_i, \ldots, s_T\}$.

On the other hand, non-linear analysis of SaO_2 signals by means of approximate entropy (*ApEn*) [29], central tendency measure (*CTM*) [30] and Lempel-Ziv complexity (*LZC*) [31] was performed to measure irregularity, variability and complexity, respectively. As stated by previous studies, these properties are usually more pronounced in oximetry recordings from subjects with higher AHI [13, 21].

To compute *ApEn*, patterns $\mathbf{s}_i$ composed of *m* consecutive samples from the original sequence **s** are obtained. For a given pattern, the regularity (frequency) $C_r^m(i)$ is expressed as:

$$C_r^m(i) = \frac{N^m(i)}{T - m + 1}, \tag{5}$$

where $N^m(i)$ denotes the number of patterns $\mathbf{s}_j$ of length *m* to a distance less or equal than *r* from $\mathbf{s}_i$. The *ApEn* is defined as the following ratio [29]:

$$ApEn(m,r) = \lim_{T\to\infty}\left[\frac{1}{T-m+1}\sum_{i=1}^{T-m+1}\ln C_r^m(i) - \frac{1}{T-m^*+1}\sum_{i=1}^{T-m^*+1}\ln C_r^{m^*}(i)\right], \tag{6}$$

where $m^* = m + 1$. *ApEn* expresses the logarithmic likelihood that runs of patterns that are close remain close on subsequent incremental comparisons.

CTM is obtained from second-order difference scatter plots representing ($s_{k+2} - s_{k+1}$) vs. ($s_{k+1} - s_k$). A circular region of radius ρ_{CTM} is defined around the origin to compute *CTM*. The number of points that fall in this region is counted and divided by the total number of points [30]:

$$CTM = \frac{\sum_{k=1}^{T-2}\delta_k}{T-2}, \tag{7}$$

where δ_k = 1 if the *k*th point is inside the circle and 0 otherwise.

LZC is a non-parametric measure of complexity in a one-dimensional signal. It is related to the number of distinct substrings and the rate of their recurrence along a given sequence. To compute *LZC*, the original signal **s** must be transformed into a two-symbol sequence $\mathbf{p} = \{p_1, \ldots, p_i, \ldots, p_T\}$. Each SaO_2 sample is compared with the median value of the samples to perform the transformation. Then, the sequence **p** is scanned from left to right and the complexity counter $c(T)$ is increased by one unit every time a new subsequence of consecutive characters is encountered. The value of *LZC* is given by [31]:

$$LZC = \frac{c(T)}{b(T)}, \tag{8}$$

where $b(T)$ is a normalization factor. It is given by:

$$b(T) = \frac{T}{\log_2 T}. \tag{9}$$

Oximetry recordings are generally non-stationary. Thus, each time-domain feature was computed by dividing the signal into 512-sample epochs, computing the value of the feature for each epoch and averaging over all the epochs. Several design parameters must be adjusted for the proposed non-linear methods. They were set to the optimum values proposed in previous studies [13, 17]. In the case of *ApEn*, the sequence length *m* was set to 1 while the optimum width of the tolerance window *r* was fixed at 0.25 times the standard deviation of the samples in each signal epoch [13]. To compute *CTM*, a radius ρ_{CTM} = 1 was selected as optimum [17]. Finally, *LZC* was computed by converting SaO_2 samples in each epoch into a 0−1 sequence. Each sample was compared with the median value from the epoch to transform the data [17].

Frequency-domain analysis

Previous studies have shown the influence of repeated apnea events on the spectral properties of SaO_2 signals. Specifically, it has been found that signal power associated with frequency components between 0.010 and 0.033 Hz

tends to increase in subjects with AHI [11]. Apneas originate phase-lagged changes in SaO_2 signals. Their duration usually ranges from 30 seconds to 2 minutes, including the awakening response after the event. Patients suffering from SAHS may have several consecutive episodes of apneas or hypopneas. Thus, the repetition of these events will be produced at a rate between 30 seconds and 2 minutes, which correspond with the frequency values mentioned before. Thus, high-power components in this range denote fluctuations in oximetry recordings due to periods with repetitive apneas.

The non-parametric Welch's method was used to compute the power spectral density (PSD) of oximetry recordings [32]. The original series **s** was divided into M overlapping sequences of length L by applying a window function $\mathbf{v} = \{v_1, \ldots, v_i, \ldots, v_L\}$. The modified periodogram was computed for each of them by using the Fast Fourier Transform (FFT):

$$\hat{P}_m(f_k) = \frac{1}{LU}\left|\sum_{i=0}^{L-1} s_i v_i e^{-\frac{j2\pi ki}{L}}\right|^2 \tag{10}$$

where

$$U = \frac{1}{L}\sum_{i=0}^{L-1}|v_i|^2 . \tag{11}$$

The estimation of the PSD was obtained as the average of the periodograms:

$$PSD(f_k) \simeq \hat{P}(f_k) = \frac{1}{M}\sum_{m=0}^{M-1}\hat{P}_m(f_k). \tag{12}$$

A 512-sample Hanning window and 50% overlapping were applied to estimate the PSD of SaO_2 signals using the Welch's method. The length of the FFT for each signal segment was set to 1024 samples.

Initially, statistical analysis was carried out in order to characterize the spectral properties of the signal. The variable representing the frequency component (f) was considered. The normalized PSD was used as its probability density function (p_F). Mean (μ_F), variance (σ_F), skewness (γ_F) and kurtosis (δ_F) were computed according to the expressions in (1), (2), (3) and (4), respectively.

In order to reflect the incidence of apnea events, three additional features were derived from the PSD function: the total power of the SaO_2 signal (S_T), the power in the band between 0.010 and 0.033 Hz (S_B), and the most significant frequency component in that band (PA). They are given by the following expressions:

$$S_T = \sum_{f=0}^{f_s/2} PSD(f), \tag{13}$$

$$S_B = \sum_{f=0.010}^{0.033} PSD(f), \tag{14}$$

$$PA = \max_{PSD(f)}\{PSD(f)\},\ f \in [0.010, 0.033]\,(Hz). \tag{15}$$

Prior to regression analysis, each of the extracted features was normalized to have a zero mean and unit variance distribution in order to avoid differences between their magnitudes.

B. *Regression analysis*

Regression techniques were used to estimate the function relating the AHI with the set of SaO_2 features. A one-dimensional continuous variable (t) was used to model the AHI value (target variable). The extracted features were grouped into a pattern $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ representing the multivariate independent variable. The approximation is built from a finite training set D composed of N input-output independent pairs $\{(\mathbf{x}_n, t_n)\}_{n=1,\ldots,N}$. Training samples are assumed to satisfy the following condition:

$$t_n = h(\mathbf{x}_n) + \varepsilon_n , \tag{16}$$

where $\mathbf{x}_n$ is known, $h(\cdot)$ is the true function and ε_n is an additive stochastic component (noise) characterized by a zero-mean Gaussian distribution [25].

Regression techniques define a mapping function $y(\mathbf{x},\mathbf{w})$ that represents an approximation to $h(\cdot)$, where **w** denotes a set of model adaptive parameters or weights. According to the maximum likelihood principle, these weights must be chosen in order to minimize the sum-of-squares error (E_D) between the actual and estimated AHI for patterns in D [25]:

$$E_D = \frac{1}{2}\sum_{n=1}^{N}\left[y(\mathbf{x}_n, \mathbf{w}) - t_n\right]^2 . \tag{17}$$

As a result, the output of the model approximates the conditional average of the target data, which is known as the regression of t conditioned on **x** [25]:

$$y(\mathbf{x}, \mathbf{w}^*) = E[t|\mathbf{x}], \tag{18}$$

where $\mathbf{w}^*$ denotes the set of model parameters that minimizes the sum-of-squares error function. In this study, the performance of two regression techniques was analyzed: MLR and MLP networks.

Multiple linear regression

MLR models assume a linear expression for the regression function. Thus, the mapping implemented by the algorithm takes the form [28]:

$$y(\mathbf{x},\mathbf{w}) = w_0 + w_1 x_1 + \cdots + w_d x_d = \mathbf{w}^T \mathbf{x}, \quad (19)$$

where $\mathbf{w} = (w_0, w_1, \ldots, w_d)^T$ are the adaptive parameters and $\mathbf{x} = (1, x_1, x_2, \ldots, x_d)$. Model optimization according to sum-of-squares error minimization yields the following solution [28]:

$$\mathbf{w} = \mathbf{X}^+ \mathbf{t}, \quad (20)$$

where rows of matrix $\mathbf{X}$ are training patterns $\mathbf{x}_n$, $\mathbf{X}^+$ is its pseudoinverse matrix and vector $\mathbf{t} = (t_1, t_2, \ldots, t_N)^T$ contains the target values corresponding to the training patterns.

Multilayer perceptron networks

MLP networks are models for expressing knowledge using a connectionist paradigm inspired in the human brain. They are composed of multiple simple units or neurons known as perceptrons, which are characterized by an activation function $g_t(\cdot)$ [33]. Perceptrons are arranged in several interconnected layers. Each network connection between two of them is associated with a network adaptive parameter or weight (w_{ij}). The response of the network to the input pattern is provided by units in the final layer (output layer). The remaining network layers are referred to as hidden layers [33].

Typically, MLP networks with a single hidden layer composed of non-linear perceptrons (i.e., with a non-linear activation function) are implemented since they are capable of universal approximation [34]. The number of units in this layer must be determined by the designer. The configuration of the output layer depends on the specifications of the problem. The proposed regression task aims to approximate a one-dimensional continuous variable representing the AHI. Thus, a single output unit with a linear activation function is required [35]. Accordingly, the network output is given by:

$$y(\mathbf{x},\mathbf{w}) = \sum_{j=1}^{N_H} \left[w_j g_t \left(\sum_{i=1}^{d} w_{ij} x_i + b_j \right) + b_0 \right], \quad (21)$$

where $\mathbf{w}$ is the weight vector composed of all the adaptive parameters (weights and biases) in the network, N_H is the number of hidden units, w_j is the weight connecting hidden unit h_j with the output unit, b_0 is the bias associated with the output unit, w_{ij} is the weight connecting the input feature i with hidden unit h_j and b_j is its associated bias.

Weights are adjusted from samples in the training set during the training or learning process. The aim is to infer the statistical properties of the problem into the network. According to the maximum likelihood principle, weights are chosen in order to minimize the sum-of-squares error function. Second-order non-linear optimization algorithms are used for this purpose [25].

Weight decay regularization can be applied to control network complexity and increase generalization capability. As stated by the bias-variance trade-off, networks with a large number of adaptive parameters (compared to the size of the training set) may overfit the data [25, 33]. Weight decay favors small weights (smooth mappings) by adding a penalty term to the error function E_D. It is equal to the sum of the squares of the network weights [25]:

$$E_T = E_D + \upsilon \sum_i w_i^2 = \frac{1}{2} \sum_{n=1}^{N} \left[y(\mathbf{x}_n,\mathbf{w}) - t_n \right]^2 + \upsilon \sum_i w_i^2, \quad (22)$$

where υ is known as the regularization parameter.

IV. Results

A. Intraclass correlation coefficient

Regression methods were evaluated using the intraclass correlation coefficient (ICC), which is a measure of reliability between observers [36]. The model (2,1) defined by Shrout and Fleiss [36] for ICC was considered since it takes into account both the random and systematic errors. The ICC ranges from −1 to 1. A negative value indicates that more differences are observed within (error in the approximation) than between subjects. ICC values close to one reflect good reliability of the algorithm.

B. Design and optimization from the training set

The MLR model has a unique solution given the training set D [28]. Consequently, no design is required. Table II shows the coefficients ($w_1, \ldots, w_d$) associated to each of the input features according to the MLR equation in (19). The additional bias term was $w_0 = 25.75\ \mathrm{h}^{-1}$.

On the other hand, MLP models require a thorough design to achieve high generalization performance. According to the bias-variance trade-off, both excessively simple and complex models will lead to poor generalization due to underfitting and overfitting, respectively [25]. Therefore, model selection is required in order to find the optimum network complexity. It is related to the number and magnitude of network weights. Thus, complexity is influenced by the number of hidden units (N_H) and the regularization parameter (υ). The performance of several network configurations was compared by varying these parameters. A wide range of values was defined for them in order to analyze their effect on generalization ability: N_H was varied from 2 to 50 units while υ values between 0.01 and 100 were evaluated. For each network configuration, the ICC was computed using leave-one-out cross-validation from data in the training set.

The evolution of ICC is shown in Fig. 1. The performance

TABLE II
Coefficients of the MLR Model Derived from the Training Set

Feature	Coefficient	Feature	Coefficient
μ_S	-2.39	μ_F	14.41
σ_S	-6.74	σ_F	0.95
γ_S	-0.73	γ_F	3.23
δ_S	-5.55	δ_F	0.90
ApEn	-4.54	S_T	13.45
CTM	-11.49	S_B	-27.08
LZC	6.08	*PA*	14.38

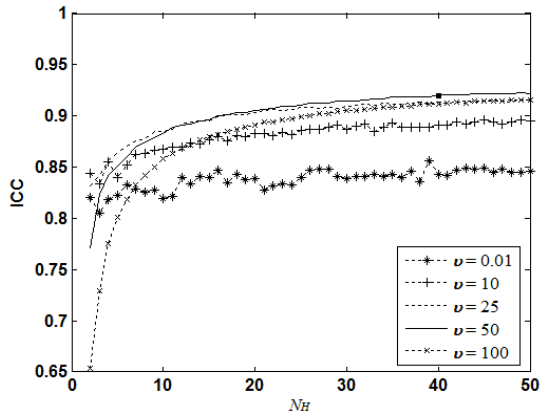

Fig. 1. Influence of the number of hidden nodes (N_H) and the regularization parameter (υ) on generalization ability.

increased as υ varied from 0.01 up to 50, which indicates that configurations with small υ may be affected by overfitting. Setting υ higher than 50 resulted in lower performance due to an excessive reduction of network complexity. Thus, υ was set to 50. For this value, we observed that ICC gradually increased as more hidden nodes were added. However, there was no substantial improvement beyond a given value of N_H, which approximately corresponds to N_H = 40. Therefore, this number of hidden nodes was selected as optimum. Finally, a MLP network with the selected configuration was trained using the complete training set. The scaled conjugate gradient algorithm was used for weight optimization [37].

C. *Performance assessment on the test set*

MLR and MLP algorithms were assessed on the test set. From ICC analysis, the MLP network (ICC = 0.91) outperformed the MLR model (ICC = 0.80). Figure 2 depicts actual versus predicted AHI as well as Bland-Altman plots for MLR and MLP models. Graphs were derived from AHI estimations computed for subjects in the test set. As reflected by the ICC value, graphic representation of the results shows that more accurate AHI estimations were provided by the MLP network. A smaller deviation from the target AHI (dotted line) can be observed for this model. This behavior is also reflected by Bland-Altman analysis. The mean of the differences between actual and predicted AHI is closer to zero for the MLP model. Furthermore, the scatter of the points is substantially higher for the MLR model, as indicated by the value of the endpoints for the 95% confidence interval. Additionally, the ability of these estimators to rank SAHS severity was evaluated. The predicted AHI was used to assign each subject to one of the following categories [1]: non-SAHS (0 $h^{-1} \leq$ AHI $<$ 5 h^{-1}), mild-SAHS (5 $h^{-1} \leq$ AHI $<$ 15 h^{-1}), moderate-SAHS (15 $h^{-1} \leq$ AHI $\leq$ 30 h^{-1}) and severe-SAHS (AHI $>$ 30 h^{-1}). The confusion matrices for MLR and MLP models are shown in Table III. The element (i,j) of the matrix represents the number of times that a class i subject was

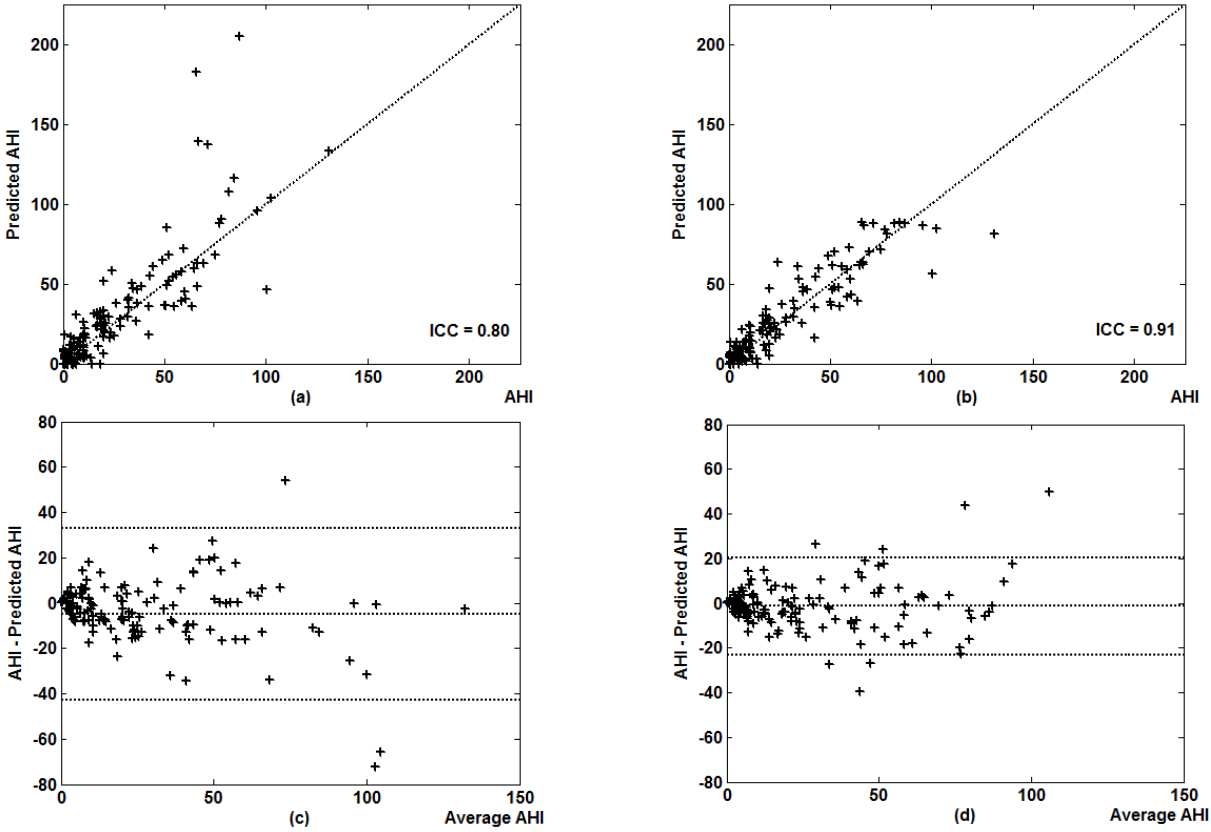

Fig. 2. Predicted versus actual AHI for (a) the MLR model and (b) the MLP network. Bland-Altman plots obtained for (c) the MLR model and (d) the MLP network. Plots were derived from AHI estimations for subjects in the test set.

assigned to class *j* [38]. Both algorithms revealed difficulties in differentiating between non-SAHS and mild-SAHS cases. The MLR algorithm correctly identified one non-SAHS subject more than the MLP network. However, it showed poor diagnostic ability to classify mild-SAHS and moderate-SAHS subjects. The results indicate that the MLP network achieved the highest overall performance.

Both regression algorithms were also assessed in a binary classification context in which non-SAHS and SAHS are the two only possible categories. Table IV summarizes the results obtained using an AHI of 5, 10 and 15 h^{-1} as the decision threshold. The MLP network improved the classification capability of the MLR model for all the evaluated thresholds. The highest accuracy of both algorithms was achieved for a decision threshold of 15 h^{-1}, which represents a more conservative definition of SAHS. The MLP network provided a correct decision rate of 93.06% whereas the MLR model achieved 88.89%.

V. Discussion

We proposed a regression approach to model SAHS diagnosis. A novel method to estimate the AHI from SaO_2 recordings was presented. Time-domain and frequency-domain features were used to reflect the dynamic behavior of these signals. Regression analysis was performed to approximate the functional relationship between the extracted features and the AHI. We assessed the performance of MLR and MLP regression models. The MLP algorithm showed the highest capability to estimate the AHI (ICC = 0.91). It improved the performance achieved from linear analysis using MLR (ICC = 0.80).

The results show that our proposed MLP algorithm provides an accurate assessment of the AHI from SaO_2 data. It achieved a sensitivity of 94.87% and a specificity of 90.91% using a decision threshold of 15 h^{-1}. The network was able to identify subjects with moderate and severe SAHS with a high level of accuracy. A total of 45 out of 48 severe-SAHS subjects and 21 out of 30 moderate-SAHS subjects were correctly diagnosed. As expected, most of the diagnostic errors corresponded to border-line patients. The algorithm assigned 14 non-SAHS subjects to the mild-SAHS group. However, the predicted AHI was smaller than 10 h^{-1} for 10 of them. Labeling SAHS-positive patients as non-SAHS is the most relevant diagnostic error since the lack of appropriate treatment could lead to other health complications. Only one subject with AHI $\geq$ 15 h^{-1} (moderate or severe SAHS) was catalogued as non-SAHS by the proposed MLP algorithm. The corresponding AHI was 19.8 h^{-1} and the minimum value of SaO_2 during the night was 91.7%, which indicates that apnea events did not tend to be accompanied by marked desaturations. Additionally, 8 mild-SAHS patients were assigned to the non-SAHS group. The true AHI for 5 of these subjects was smaller than 10 h^{-1}. Therefore, based on the results of this assessment we conclude that our proposed MLP-based algorithm is a reliable method to assess SAHS severity.

Other methods to estimate the AHI from oximetry data have been suggested in the literature. Vázquez et al. [22] proposed the respiratory disturbance index (RDI), which is based on the detection of desaturation events over 4%. The correlation coefficient between RDI and AHI was 0.97. The utility of this index to classify subjects as non-SAHS or SAHS was evaluated. A sensitivity of 97% and a specificity of 80% were reached using AHI $\geq$ 10 h^{-1} to define SAHS. Magalang et al. [12] used several indices from oximetry data to compute the AHI. The oxygen desaturation index over 3% or 4% and the cumulative time spent below different levels of saturation were used as inputs to a multivariate adaptive regression splines (MARS) model. A correlation of 0.84 with the true AHI was achieved. This method provided a sensitivity of 90% and a specificity of 70% using AHI $\geq$ 15 h^{-1} to define SAHS. Similarly, Lin et al. [23] reported a correlation coefficient of 0.92 by counting the number of desaturations over 3%. In addition, other signals and data different to SaO_2 have been analyzed to approximate the AHI. Roche et al. [24] developed a MLR model from the combination of clinical and oximetry features. It achieved a low correlation value with the AHI derived from PSG (0.38). Its diagnostic accuracy was 62%. De Chazal et al. [9] proposed an algorithm to detect apnea epochs from ECG and SaO_2 features. An estimate of the AHI was derived from the epoch-based classification approach. A sensitivity of 95% and a specificity of 83% were reached through the estimated AHI. However, patients with mild SAHS were not considered for testing.

Neural networks have shown to be a powerful tool for regression analysis. As suggested in the present study, other researchers developed neural network-based regression algorithms for AHI estimation. Kirby et al. [8] used 23 clinical variables as inputs to a generalized regression neural network

TABLE III
Diagnostic Results Achieved by MLR and MLP Regression Models

MLR				
True	Predicted			
	No SAHS	Mild	Moderate	Severe
No SAHS	21	11	2	0
Mild	10	11	10	1
Moderate	1	2	17	10
Severe	0	0	3	45
MLP				
True	Predicted			
	No SAHS	Mild	Moderate	Severe
No SAHS	20	14	0	0
Mild	8	18	6	0
Moderate	1	3	21	5
Severe	0	0	3	45

TABLE IV
Diagnostic Results using a Binary Classification Approach

	MLR			MLP		
AHI (h^{-1})	5	10	15	5	10	15
Se (%)	90.00	89.58	96.15	91.82	89.58	94.87
Sp (%)	61.76	77.08	80.30	58.82	81.25	90.91
Acc (%)	83.33	85.42	88.89	84.03	86.81	93.06

Se: sensitivity: Sp: specificity; Acc: accuracy.

(GRNN), which provided a diagnostic accuracy of 91%. El-Solh et al. [7] used a MLP network with clinical and demographic data to determine the AHI. The sensitivity (95%) was significantly higher than the specificity (65%). Nevertheless, these studies did not assess the ability of their methods to rank SAHS severity. Similarly, other previous studies from our research group were focused on SAHS detection from SaO_2 analysis using neural networks. However, a classification approach was proposed. In this context, the output of the algorithm is a categorical variable that indicates the group membership (non-SAHS or SAHS) for the input feature pattern. MLP [14] and RBF [15] classifiers achieved an accuracy of 86% using non-linear features from SaO_2 signals as inputs. Recently, linear classifiers based on discriminant analysis and logistic regression also achieved significant results. They provided an accuracy of 93% [18] and 90% [17], respectively.

Despite the high performance of these classification algorithms, the regression approach results in a more accurate model to characterize SAHS. Two main advantages can be derived from AHI estimation. First, the model provides information about the severity of SAHS. Second, it is insensitive to the criterion used for a positive diagnosis, i.e., the AHI value used to discriminate between SAHS-negative and SAHS-positive. Typically, different criteria between 5 h^{-1} and 15 h^{-1} are used and there is not a globally accepted standard [4].

The MLP algorithm achieved the highest diagnostic capability for the proposed regression problem. However, several limitations can be pointed out. Neural networks are complex models that require an exhaustive design process in comparison with MLR models. The study results indicate that the main source of errors of the MLP algorithm correspond to non-SAHS and mild-SAHS subjects. Therefore, increasing the number of samples in these groups would be desirable in order to obtain a more detailed description from training data. Future studies should also evaluate the effect of feature selection as a technique to remove redundant information and reduce model complexity. In addition, signal preprocessing for motion artifact reduction should be improved to achieve accurate SaO_2 measurements, resulting in more reliable AHI estimations. Internal average (low-pass) filtering performed by the pulse oximetry equipment is not capable of complete artifact removal. Furthermore, the influence of the averaging time is a relevant factor to be considered. An excessively high value of this parameter may result in underestimated readings of desaturation events associated with apneas, leading to an incorrect representation of SaO_2 dynamics [39]. Another limitation of the study is due to the behavior of SaO_2 signals. In some cases, apneas and hypopneas occurred during sleep may not be accompanied by desaturation events. As a result, the extracted time-domain and frequency-domain features do not reflect the actual AHI, leading to a poor estimation. In order to avoid these situations, information from oximetry data could be combined with other signals such as nasal airflow. However, it may increase the complexity of the data acquisition process and the resulting algorithm.

VI. Conclusion

We proposed an algorithm for automatic estimation of AHI from SaO_2 based on feature extraction of time-domain and frequency-domain characteristics, combined with a MLP-based algorithm. The results of our assessment study show that the proposed MLP-based algorithm outperforms equivalent MLR-based algorithms that use the same input features. Our results indicate a high agreement between actual and predicted AHI (ICC = 0.91).

The proposed algorithm only requires nocturnal SaO_2 recordings as the input, eliminating the need for costly, inconvenient, complex, and time-consuming PSG studies for most subjects. Furthermore, the non-invasive nature of SaO_2 makes it possible to create low-cost portable devices designed for home monitoring that can be used for widespread and cost-effective first-line assessment of SAHS severity.

References

[1] A. Qureshi and R. D. Ballard, “Obstructive sleep apnea,” *J. Allergy Clin. Immunol.*, vol. 112, pp. 643−651, 2003.

[2] T. Young, P. E. Peppard and D. J. Gottlieb, “Epidemiology of obstructive sleep apnea: a population health perspective,” *Am. J. Respir. Crit. Care Med.*, vol. 165, pp. 1217−1239, 2002.

[3] T. Young, L. Evans, L. Finn and M. Palta, “Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women,” *Sleep*, vol. 20, pp. 705–706, 1997.

[4] S.M. Caples, A.S. Gami and V.K. Somers, “Obstructive sleep apnea,” *Ann. Intern. Med.*, vol. 142, pp. 187−197, 2005.

[5] J.A. Bennet and W.J.M. Kinnear, “Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome,” *Thorax*, vol. 54, pp. 958−959, 1999.

[6] W.W. Flemons, N.J. Douglas, S.T. Kuna, D.O. Rodenstein and J. Wheatley, “Access to diagnosis and treatment of patients with sleep apnea,” *Am. J. Respir. Crit. Care Med.*, vol. 169, pp. 668−672, 2004.

[7] A.A. El-Solh, M.J. Mador, E. Ten-Brock, D.W. Shucard, M. Abul-Khoudoud and B.J.B. Grant, “Validity of neural network in sleep apnea,” *Sleep*, vol. 22, pp. 105−111, 1999.

[8] S.D. Kirby, W. Danter, C.F.P. George, T. Francovic, R.P.F. Ruby and K.A. Ferguson, “Neural network prediction of obstructive sleep apnea from clinical criteria,” *Chest*, vol. 116, pp. 409−415, 1999.

[9] P. de Chazal, C. Heneghan and W.T. McNicholas, “Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals,” *Philos. Trans. R. Soc. A*, vol. 367, pp. 369−389, 2009.

[10] M.O. Méndez, A.M. Bianchi, M. Matteucci, S. Cerutti and T. Penzel, “Sleep apnea screening by autoregressive models from a single ECG lead,” *IEEE Trans. Biomed. Eng.*, vol. 56, pp. 2838−2850, 2009.

[11] C. Zamarrón, F. Gude, J. Barcala, J. R. Rodríguez and P. V. Romero, “Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome,” *Chest*, vol. 123, pp. 1567−1576, 2003.

[12] U. J. Magalang, J. Dmochowski, S. Veeramachaneni, A. Draw, M. J. Mador, A. El-Solh and B. J. B. Grant, “Prediction of the apnea-hypopnea index from overnight pulse oximetry,” *Chest*, vol. 124, pp. 1694−1701, 2003.

[13] R. Hornero, D. Álvarez, D. Abásolo, F. Del Campo and C. Zamarrón, “Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome,” *IEEE Trans. Biomed. Eng.*, vol. 54, pp. 107–113, 2007.

[14] J.V. Marcos, R. Hornero, D. Álvarez, F. Del Campo, C. Zamarrón and M. López, “Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry,” *Comput. Meth. Programs Biomed.*, vol. 72, pp. 79−89, 2008.

[15] J.V. Marcos, R. Hornero, D. Álvarez, F. Del Campo, M. López and C. Zamarrón, “Radial basis function classifiers to help in the diagnosis of

the obstructive sleep apnoea syndrome from nocturnal oximetry," *Med. Biol. Eng. Comput.*, vol. 46, pp. 323−332, 2008.
[16] D.S. Morillo, J.L. Rojas, L.F. Crespo, A. León and N. Gross, "Poincaré analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome," *Physiol. Meas.*, vol. 30, pp. 405−420, 2009.
[17] D. Álvarez, R. Hornero, J.V. Marcos and F. Del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2816−2824, 2010.
[18] J.V. Marcos, R. Hornero, D. Álvarez, F. Del Campo and M. Aboy, "Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis," *Med. Biol. Eng. Comput.*, vol. 48, pp. 895−902, 2010.
[19] N. Netzer, A. H. Eliasson, C. Netzer and D. A. Kristo, "Overnight pulse oximetry for sleep-disordered breathing in adults: a review," *Chest*, vol. 120, pp. 625-633, 2001.
[20] K. E. Bloch, "Getting the most out of nocturnal pulse oximetry," *Chest*, vol. 124, pp. 1628-1630, 2003.
[21] D. Álvarez, R. Hornero, D. Abásolo, F. Del Campo and C. Zamarrón, "Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection," *Physiol. Meas.*, vol. 27, pp. 399–412, 2006.
[22] J.C. Vázquez, W.H. Tsai, W.W. Flemons, A. Masuda, R. Brant, E. Hadjuk, W.A. Whitelaw and J.E. Remmers, "Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea," *Thorax*, vol. 55, pp. 302−307, 2000.
[23] C.L. Lin, C. Yeh, C.W. Yen, W.H. Hsu and L.W. Hang, "Comparison of the indices of oxyhemoglobin saturation by pulse oximetry in obstructive sleep apnea hypopnea syndrome," *Chest*, vol. 135, pp. 86−93, 2009.
[24] N. Roche, B. Herer, C. Roig and G. Huchon, "Prospective testing of two models based on clinical and oximetric variables for prediction of obstructive sleep apnea," *Chest*, vol. 121, pp. 747−752, 2002.
[25] C.M. Bishop, *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press, 1995.
[26] N.A. Collop, W.Mc. Anderson, B. Boehlecke, D. Claman, R. Goldberg, D.J. Gottlieb, D. Hudgel, M. Sateia and R. Schwab, "Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients," *J. Clin. Sleep Med.*, vol. 3, pp. 737−747, 2007.
[27] A. Rechtschaffen and A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Washington, D.C.: Public Health Service U. S. Government Printing Office, 1968.
[28] J.D. Jobson, *Applied multivariate data analysis. Volume I: regression and experimental design*. New York: Springer, 1991.
[29] S.M. Pincus, "Assesing serial irregularity and its implications for health," *Ann. NY Acad. Sci.*, vol. 954, pp. 245−267, 2001.
[30] M.E. Cohen, D.L. Hudson and P.C. Deedwania, "Applying continuous chaotic modeling to cardiac signals," *IEEE Eng. Med. Biol. Mag.*, vol. 15, pp. 97−102, 1996.
[31] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. 22, pp. 75−81, 1976.
[32] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. Acous. Speech*, vol. 15, pp. 70-73, 1967.
[33] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New Jersey: Prentice Hall, 1999.
[34] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, pp. 251−257, 1991.
[35] I.T. Nabney, *NETLAB: Algorithms for pattern recognition*. Berlin: Springer, 2002.
[36] P.E. Shrout and J.L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychol. Bull.*, vol. 36, pp. 420−428, 1979.
[37] M.F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, pp. 525−533, 1993.
[38] J.R. Parker, "Rank and response combination from confusion matrix data," *Inf. Fusion*, vol. 2, pp. 113−120, 2001.
[39] R. Farré, J.M. Montserrat, M. Rotger, E. Ballester and D. Navajas, "Accuracy of thermistors and thermocouples as flow-measuring devices for detecting hypopneas," *Eur. Resp. J.*, vol. 11, pp. 179-182, 1998.

J. Víctor Marcos (S'07) was born in Valladolid, Spain, in 1981. He received the M.S. degree in telecommunication engineering and the PhD degree from the University of Valladolid (Spain) in 2006 and 2011, respectively. He is a member of the Biomedical Engineering Group in the Department of Signal Theory of this university.
His current research interests include biomedical signal processing using pattern recognition techniques.

Roberto Hornero (M'04) was born in Plasencia, Spain, in 1972. He received the M.S. degree in telecommunication engineering and the Ph.D. degree from the University of Valladolid (Spain), in 1995 and 1998, respectively.
He is a Professor in the Department of Signal Theory, University of Valladolid. He founded the Biomedical Engineering Group in 2004. The research interests of this group are connected with the field of nonlinear dynamics, chaotic theory, and wavelet transform with applications in biomedical signal and image processing.

Daniel Álvarez (S'07) was born in Bembibre, Spain, in 1978. He received the M.S. degree in telecommunication engineering and the PhD degree from the University of Valladolid, in 2005 and 2011, respectively. He is currently a researcher at the Biomedical Engineering Group in the Department of Signal Theory in this university.
His current research interests include nonlinear analysis of biomedical signals.

Mateo Aboy (M'98) was born in Pontevedra, Spain, in 1980. He received the B.S. degree (high honors) in electrical engineering and the B.S. degree (high honors) in physics with a minor in mathematics from Portland State University (PSU), Portland, OR, in 2002, the M.S. degree (*summa cum laude*) in electrical and computer engineering from PSU in 2004, and the Ph.D. degree (*summa cum laude*) from the University of Vigo (Spain) in 2005.
He has a Bar Admission to practice patent law before the United States Patent and Trademark Office and was a Managing Partner and a Patent Agent at Aboy&Associates PC (www.aboypatentlaw.com). Since 2005, he has been with the Oregon Institute of Technology (OIT), Portland. He is currently a tenure Associate Professor and the Department Chair of the Electrical Engineering Department (EERE). His current research interests include practical applications of signal processing.

Félix del Campo was born in Valladolid, Spain, in 1953. He received the degree in medicine and the M.D. degree from the University of Valladolid, Valladolid, Spain, in 1978 and 1994, respectively.
He is currently with the Hospital Clínico Universitario del Río Hortega, Valladolid. He is currently an Adjunct Professor in the Department of Medicine, University of Valladolid. His current research interests include respiratory sleep disorders.