Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis

2012 Physiol. Meas. 33 1261

(http://iopscience.iop.org/0967-3334/33/7/1261)

IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **33** (2012) 1261–1275 doi:10.1088/0967-3334/33/7/1261

Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis

G C Gutiérrez-Tobal[1], R Hornero[1], D Álvarez[1], J V Marcos[1] and F del Campo[2]

[1] Biomedical Engineering Group, ETSI de Telecomunicación, University of Valladolid, Paseo Belén 15, 47011, Valladolid, Spain
[2] Hospital Universitario Río Hortega, Servicio de Neumología, c/Dulzaina 2, 47012, Valladolid, Spain

E-mail: gguttob@ribera.tel.uva.es, robhor@tel.uva.es, dalvgon@ribera.tel.uva.es, jvmarcos@gmail.com and fsas@telefonica.net

Received 16 February 2012, accepted for publication 6 June 2012
Published 27 June 2012
Online at stacks.iop.org/PM/33/1261

Abstract

This paper focuses on the analysis of single-channel airflow (AF) signal to help in sleep apnoea–hypopnoea syndrome (SAHS) diagnosis. The respiratory rate variability (RRV) series is derived from AF by measuring time between consecutive breathings. A set of statistical, spectral and nonlinear features are extracted from both signals. Then, the forward stepwise logistic regression (FSLR) procedure is used in order to perform feature selection and classification. Three logistic regression (LR) models are obtained by applying FSLR to features from AF, RRV and both signals simultaneously. The diagnostic performance of single features and LR models is assessed and compared in terms of sensitivity, specificity, accuracy and area under the receiver-operating characteristics curve (AROC). The highest accuracy (82.43%) and AROC (0.903) are reached by the LR model derived from the combination of AF and RRV features. This result suggests that AF and RRV provide useful information to detect SAHS.

Keywords: sleep apnoea–hypopnoea syndrome, airflow, respiratory rate variability, feature extraction, feature selection

1. Introduction

The sleep apnoea–hypopnoea syndrome (SAHS) is characterized by repetitive events of apnoea (complete cessation of breathing) and hypopnoea (significant breathing reduction) during sleep (Flemons *et al* 2003). SAHS has been associated with other diseases such as hypertension, atrial fibrillation, stroke, cardiac failure, aortic dissection and sudden cardiac death (López-Jiménez *et al* 2008). Furthermore, daytime sleepiness caused by SAHS is a risk factor for occupational accidents and motor-vehicle collisions (Lindberg *et al* 2001, Sassani *et al* 2004).

The prevalence of SAHS has been estimated at 1%–5% of adult men and 2% women in western countries. However, studies reported up to 5% of adult population remaining undiagnosed (Young *et al* 2002).

The gold standard for SAHS diagnosis is polysomnography (PSG) (Flemons *et al* 2003). PSG is an overnight test in which many physiological signals are monitored. The apnoea–hypopnoea index (AHI) from PSG is used to characterize its severity (Patil *et al* 2007). Despite its effectiveness, PSG is an expensive and complex test, since it needs the supervision of specialists and a visual inspection of signals to compute AHI. This results in longer waiting lists and increased delay time for a final diagnosis (Flemons *et al* 2004). Therefore, there is a demand of new helping methods of diagnosis capable of overcoming PSG drawbacks (Penzel *et al* 2002). Many studies have focused on analysing a reduced set of signals from overnight PSG. Typically, the diagnostic ability of electrocardiogram (Penzel *et al* 2002), electroencephalogram (Poyares *et al* 2002), airflow (AF) (Nakano *et al* 2007, Han *et al* 2008), and blood oxygen saturation (SpO_2) (Álvarez *et al* 2010) has been evaluated.

The AF waveform is directly affected by the occurrence of respiratory events (Flemons *et al* 2003). Apnoeas are reflected by near-zero values, whereas hypopnoeas cause amplitude reduction. In contrast, clear oscillations are observed for normal breathing periods. Therefore, an intensive analysis of the information from the single-channel AF signal is proposed to help in SAHS detection. In addition to the AF signal, the respiratory rate variability (RRV) series is also analysed. RRV is computed by measuring the time between consecutive breathings in AF, similar to the well-known heart rate variability series (Cysarz *et al* 2008). The normal pattern for RRV also reflects alterations in the presence of SAHS, since sleep apnoea modifies the respiratory oscillation (Cysarz *et al* 2008).

The main purpose of the current study is to evaluate the diagnostic usefulness of AF and RRV series in SAHS detection. In order to characterize SAHS, the extraction of statistical, spectral and nonlinear features from AF and RRV is proposed. Common parameters such as statistical moments have shown to be useful in SAHS detection (Roche *et al* 1999, de Chazal *et al* 2003). Furthermore, frequency analysis has been successfully applied to study different diseases (Casolo *et al* 1991, Penzel *et al* 2002, Poza *et al* 2007). Moreover, nonlinear methods have recently proved high capability to help in SAHS diagnosis (Álvarez *et al* 2006, Hornero *et al* 2007, Morillo *et al* 2009). After feature extraction, a feature selection stage is implemeted. It is carry out by means of the forward stepwise logistic regression (FSLR) methodology (Hosmer and Lemeshow 1999), which has been successfully used in prior studies of SAHS (Álvarez *et al* 2010). The logistic regression (LR) models obtained through the FSLR procedure combine the non-redundant information from the features extracted (Hosmer and Lemeshow 1999). Finally, the diagnostic performance of the single features and the LR models are assessed and compared in terms of sensitivity, specificity, accuracy and area under the receiver-operating characteristic curve (AROC).

2. Subjects and signals

2.1. Subjects under study

In this study, 148 subjects suspected of suffering from SAHS were involved (79% males and 21% females). The recordings were obtained in the sleep unit of Hospital Universitario Río Hortega in Valladolid, Spain. All subjects presented common symptoms such as daytime hypersomnolence, loud snoring, nocturnal choking and awakenings or referred apnoeic events. The subjects were free from any medication which could influence the respiratory centre. Neither patients suffering from hypothyroidism (two out of the total subjects) nor those

Table 1. Demographic and clinical data of the population under study. Data are presented as mean ± SD or *n* (%). SAHS-positive: subjects with sleep apnoea–hypopnoea syndrome; SAHS-negative: subjects without sleep apnoea–hypopnoea syndrome; BMI: body mass index; time: recording time; AHI: apnoea–hypopnoea index.

	All subjects	SAHS-positive	SAHS-negative
Subjects (n)	148	100 (67.6%)	48 (32.4%)
Age (years)	50.87 ± 11.68	51.89 ± 11.41	48.75 ± 12.07
Males (n)	117 (79.0%)	85 (85.0%)	32 (66.7%)
BMI (kg m^{-2})	29.1 ± 4.6	29.9 ± 4.7	27.6 ± 4.9
Time (h)	7.24 ± 0.38	7.23 ± 0.36	7.27 ± 0.43
AHI (events/h)	–	32.9 ± 24.3	4.0 ± 2.4

suffering from chronic obstructive pulmonary disease (COPD) (six out of the total subjects) were excluded. Physicians considered 100 subjects affected (positive) and 48 not affected (negative) by SAHS. The AHI threshold for a positive diagnosis was 10 events/h at least. Apnoea was defined as the cessation of AF for 10 s or more. Hypopnoea was defined as a minimum of 30% of amplitude reduction for at least 10 s accompanied by a 4% or more decrease in the saturation of haemoglobin. The Review Board on Human Studies accepted the protocol, and all subjects gave their informed consent to participate in the study. Demographic and clinical data of the participants are summarized in table 1.

2.2. AF and RRV signals

The AF recordings were obtained from overnight PSG (Alice 5, Respironics, Philips Healthcare, the Netherlands). The sensor used to register AF was a thermistor (Pro-Tech, Respironics, Philips Healthcare, the Netherlands) and the sampled rate was 10 Hz. Previous to the automatic analysis, a visual inspection of the signals was carried out to assess their quality. Four recordings were excluded due to prolonged malfunction of the thermistor. Thus, the remaining 148 AF recordings were entirely analysed.

A peak detection algorithm was implemented to locate inspiratory onsets in AF signal (Korten and Haddad 1989). Then, RRV was computed by measuring the time between consecutive locations (Cysarz *et al* 2008).

Figure 1(a) shows an example of the AF signal and figure 1(b) shows the corresponding RRV signal. The first 34 s of the AF signal corresponds to a normal breathing pattern. Consequently, the time between breathings remains around 4.2 s in the RRV signal. Then, a hypopnoea is shown in the AF signal which is reflected by a decrease in the RRV signal amplitude. Finally, since the AF normal breathing pattern is recovered, the time between breathings begins to increase.

3. Methods

The proposed methodology started with a spectral analysis of AF and RRV recordings to determine those frequency bands associated with SAHS. Afterwards, spectral, nonlinear and statistical features were extracted from AF and RRV. Then several LR models were obtained by means of the FSLR method. Finally, diagnostic performance of single features and LR models was assessed.

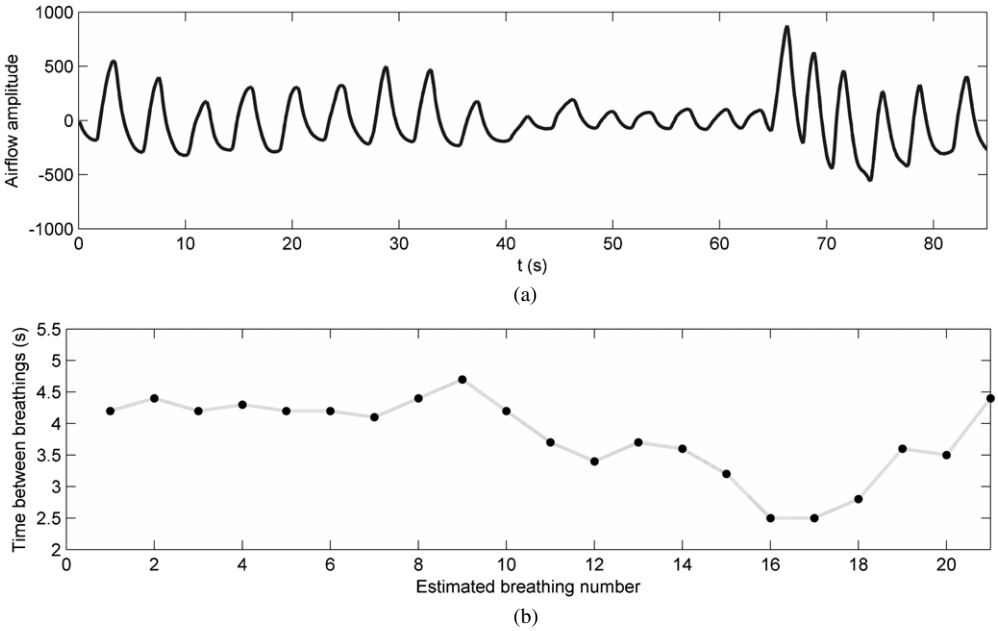

Figure 1. Normal breathing pattern followed by hypopnoea event in (a) AF signal and (b) corresponding RRV signal.

3.1. Definition of spectral bands of interest

The bands of interest were defined as the frequency regions of power spectral density (PSD) in which the highest statistical differences between SAHS-positive and SAHS-negative populations were found. PSD of recordings was estimated by means of a non-parametric Welch method (Welch 1967). This method divides the signals into M overlapping segments of length L. Then, a smooth time window $w[n]$ is applied, and the modified periodogram of each windowed segment $v_L[n]$ is computed by means of the discrete Fourier transform (DFT) $V[f]$ (Welch 1967):

$$\hat{P}[f] = \frac{|V[f]|^2}{f_s \mathrm{LU}}, \tag{1}$$

where f_s is the sample rate:

$$V[f] = \sum_{n=0}^{N-1} v_L[n]\, \mathrm{e}^{-j(2\pi k/N)n}, \tag{2}$$

and

$$U = \frac{1}{M} \sum_{n=0}^{M-1} |w(n)|^2. \tag{3}$$

Finally, the average of all DFTs is calculated to obtain the PSD function. A 2048-sample Hamming window, with 50% overlap and 4096-point DFTs, was used to compute the PSD of AF and RRV recordings. A cubic spline interpolation (resampling at 10 Hz) was applied to RRV series before the computation of the PSDs.

The representations of the joint PSDs of each SAHS-positive and SAHS-negative populations were obtained. The median of the PSD values at each frequency component

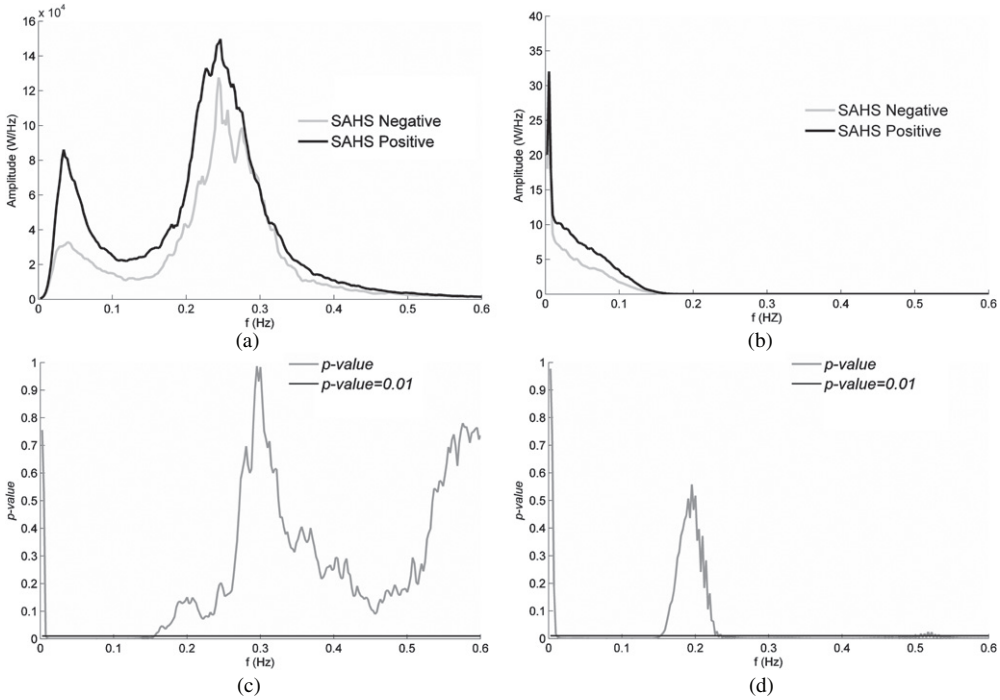

Figure 2. Spectral bands of interest of AF and RRV. AF (a) and RRV (b) median-based representations of PSDs from SAHS-positive (black line) and SAHS-negative (grey line) populations. AF (c) and RRV (d) whole spectrum *p*-value versus frequency representations (solid line).

was applied due to its robustness to outliers. Figures 2(a) and (b) show this median-based representation for AF and RRV, respectively. The *p*-value from the Kruskal–Wallis test was used to find statistically significant differences along frequencies (*p*-value <0.01). Figures 2(c) and (d) display the *p*-value versus frequency representations for AF and RRV, respectively. Figure 2(c) only shows significant differences in the very low frequency band (0.002–0.151 Hz). This agrees with figure 2(a), which displays the greatest qualitative differences in the same band of the spectrums of AF signals. The spectral band of interest was that with the highest statistically significant differences, i.e. 0.022–0.059 Hz. Moreover, the plot in figure 2(d) indicates significant differences in most of the frequency components of the RRV spectrum. However, the highest differences were also found at the very low band. This is also consistent with the corresponding median-based representation of the PSD (figure 2(b)). The spectral band of interest was 0.095–0.132 Hz.

3.2. Feature extraction

3.2.1. Statistical moments. The distributions of AF and RRV values are expected to differ from SAHS-positive and SAHS-negative populations. In order to typify the statistical behaviour of these distributions, the first-to-fourth statistical moments (Mt_1–Mt_4) were obtained from time series. The arithmetic mean (Mt_1), standard deviation (SD) (Mt_2), skewness (Mt_3) and kurtosis (Mt_4) quantify the central tendency, dispersion, asymmetry and peakedness of data, respectively.

3.2.2. Spectral features. The reccurrent nature of apnoeic events can be characterized by means of spectral analysis. Seven spectral features were extracted from the frequency bands of interest and the same features were obtained from the full PSDs.

Peak amplitude (*PA*) is the maximum of PSD in a given frequency interval. Band power (*BP*) represents the spectral power of a region. Both of them are conventional parameters and can be computed as follows:

$$PA = \max_{\mathrm{PSD}}\{\mathrm{PSD}(f)\}, \quad f(Hz) \in [f_i, f_N], \quad i = 1, 2, \ldots, N, \tag{4}$$

$$BP = \sum_{f_i=f_1}^{f_N} \mathrm{PSD}(f_i), \quad i = 1, 2, \ldots, N, \tag{5}$$

where N is the number of points in the band and f_i are the frequency components of the spectrum. Figures 2(a) and (b) indicate that higher *PA* and *BP* values are expected for the SAHS-positive population.

The Wootters distance (*WD*) is a disequilibrium measurement (Wootters 1981). This parameter requires the PSD to be normalized (PSD_n) in order to consider it as a probability density function (pdf). It is possible to measure the distance between the pdf and the uniform distribution (Wootters 1981):

$$WD = \arccos\left\{\sum_{f_i=f_1}^{f_2} \sqrt{\mathrm{PSD}_n(f)} \cdot \sqrt{1/N}\right\}, \tag{6}$$

with f_1 and f_2 being the limits of the frequency range where *WD* is applied and N is the number of the corresponding PSD_n points. If PSD_n is equal to a uniform distribution along frequencies (as in white noise), then *WD* will be equal to zero. Moreover, if the normalized spectrum is condensed into a narrow frequency band (as in a sum of sinusoids), *WD* reaches the highest values. According to figures 2(a) and (b), differences between *WD* values from both populations are expected.

Finally, first-to-fourth statistical moments (Mf_1–Mf_4) of the amplitude values of PSDs were also obtained. Differences between the distributions of PSD values are reflected by these parameters.

3.2.3. Nonlinear features. The high recurrence of apnoeas and hypopnoeas in SAHS-positive subjects modifies the corresponding AF and RRV waveforms. Since central tendency measure (*CTM*), Lempel–Ziv complexity (*LZC*) and approximate entropy (*ApEn*) are applied in time domain, it is expected that these parameters can reflect the differences in the variability, complexity and irregularity of time series from populations.

The *CTM* quantifies the degree of variability or chaos in a time series (Cohen *et al* 1996). *CTM* is based on the plots of the first-order differences representing $x[n+2]-x[n+1]$ versus $x[n+1]-x[n]$, where $x[n]$ are the time serie values (Abásolo *et al* 2006). It is calculated by counting the points that fall within a circle of radius ρ around the origin and dividing it by the total number of points (Cohen *et al* 1996):

$$CTM = \frac{1}{N-2}\sum_{n=1}^{n-2} \delta(n), \tag{7}$$

where

$$\delta(n) = \begin{cases} 1 & \text{if } \{(x[n+2]-x[n+1])^2 + (x[n+1]-x[n])^2\}^{1/2} < \rho \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

with N being the points of the time series. *CTM* achieves values between 0 and 1, reaching values closer to 1 when a given series is less variable (values more concentrated around centre) and closer to 0 when it has more variability. Radius ρ has to be selected experimentally, depending on the character of signals (Cohen *et al* 1996). A method based on p-value was used to select ρ (Hornero *et al* 1999). First, *CTM* of time series was computed by fixing several radii. Then a statistical significance test was applied to select the ρ which ensured the most significant differences between populations, i.e. the lowest p-value. In this study, two radii were used: $\rho_1 = 31$ for the AF signal and $\rho_2 = 6.61$ for the RRV signal.

The complexity of finite sequences can be estimated by means of *LZC* (Lempel and Ziv 1976). Larger values of the parameter correspond to higher complexity in these sequences (Zhang *et al* 2001). The first step in *LZC* estimation is to convert the time series into finite sequences of simbols, $s(i)$ (Zhang *et al* 2001, Abásolo *et al* 2006). Binary sequences have been commonly proposed. Due to its robustness to outliers, we assumed the median value as the threshold to assign a simbol to each value of time series. Once the sequence is obtained, it is scanned from left to right, and a complexity counter $c(n)$ is increased every time a new subsequence of consecutive characters is encountered (Zhang *et al* 2001). Finally, $c(n)$ is normalized to make the method independent of the length of sequences:

$$LZC = \frac{c(n)}{b(n)}, \tag{9}$$

where

$$b(n) = \frac{n}{\log_\alpha(n)}, \tag{10}$$

and $\alpha = 2$ since the sequence is binary.

ApEn is an irregularity measure in time series which was originally developed to be applied over short and noisy data sets (Pincus 1991). *ApEn* can assess both dominant and subordinates patterns in data for which other methods cannot make the feature recognition easy (Pincus 2001). *ApEn* has two user-specified parameters: a length m and a tolerance window r. Theoretically, the *ApEn* is defined as

$$ApEn(m, r) = \lim_{N\to\infty} [\phi^m(r) - \phi^{m+1}(r)], \tag{11}$$

where N is the total number of points of the original time series and $\phi^m(r)$ is the average of the logarithmic likelihood patterns of length m that are repeated along the original sequence. The tolerance parameter r is used to determine the similarity between patterns. Since N is finite, the *ApEn* is commonly applied as the statistic (Pincus 1991):

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r). \tag{12}$$

Larger values of *ApEn* correspond to more irregularity in the data (Pincus 2001). Despite their influence in the *ApEn* outcome, there are no guidelines to optimize the m and r values (Hornero *et al* 2005). Thus, $m = 1$, $m = 2$ and $r = 0.1$, 0.15, 0.2, 0.25 times the SD of the original data sequence have been proposed as input parameters. These values produce good statistical reproducibility of *ApEn* for time series of length $N \geqslant 60$ (Pincus 2001). In order to choose between these values, the p-value-based methodology previously described was used, and $m = 1$ and $r = 0.25$ SD were selected for both AF and RRV signals.

3.3. Feature selection

The features described in the previous subsections measure different properties of AF and RRV. The information contained in these parameters may be complementary. For simultaneous analysis of these features, several LR models were obtained. The method used to automatically

select the features was FSLR which was proposed by Hosmer and Lemeshow (1999). This procedure was applied to the features from AF, RRV and both signals (AF-RRV).

3.3.1. Forward stepwise logistic regression (FSLR). A regression-based method was used to describe the relationship between a response variable (outcome) and the explanatory variables. In this study, the response is a dichotomous variable codifying the diagnosis of a subject ('0' non-affected, and '1' affected by SAHS), and the explanatory variables are the features explained previously. For this outcome variable, the LR model has become the standard method of analysis:

$$\pi(x) = \frac{e^{\beta_0 + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}}}{1 + e^{\beta_0 + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}}}, \tag{13}$$

where $\pi(x)$ values range between 0 and 1, and can be interpreted as the probability of membership to the SAHS-positive population. β_0 is a constant for each model and $\boldsymbol{\beta}$ is a vector with coefficients for each component of $\mathbf{x}$. Both β_0 and $\boldsymbol{\beta}$ are estimated according to the maximum-likelihood criterion (Hosmer and Lemeshow 1999).

The more variables included in a LR model (higher dimensionality), the more dependent the model becomes on the observed data due to overfitting. Thus, feature selection was used in order to obtain models with higher capability of generalization. The FSLR procedure has been proposed for this purpose. It checks the relevance of features, including or excluding them according to a fixed decision rule. In this work, the decision rule chosen has been the p-value of the likelihood ratio (Hosmer and Lemeshow 1999). FSLR is characterized by a forward selection followed by the backward elimination of variables at each step.

3.4. Statistical analysis

The non-parametric Kruskal–Wallis test was used to assess the differences between the SAHS-positive and the SAHS-negative populations, with a p-value < 0.01 considered as significant. To ensure statistical validity of results, a leave-one-out cross-validation approach was applied. Sensitivity (percentage of SAHS-positive subjects correctly diagnosed), specificity (percentage of SAHS-negative subjects correctly diagnosed) and accuracy (proportion of total subjects under study correctly classified) were computed by averaging all results from the cross-validation process. Additionally, the AROC was computed to quantify the diagnostic performance of a given method (Zweig and Campbell 1993).

4. Results

4.1. Diagnostic performance of single features

A total of 21 features were extracted from each of the two series. Table 2 summarizes the measurements (mean $\pm$ SD) obtained for each feature in SAHS-positive and SAHS-negative populations. The p-value from the non-parametric Kruskal–Wallis significance test is also shown.

In the case of the AF signal, six out of seven spectral features obtained from the band of interest showed statistically significant differences between populations (p-value < 0.01). Only one out of seven (Mf_3) spectral features computed from the full PSD also showed p-value < 0.01. Neither the statistical moments in time domain nor the nonlinear features achieved statistically significant differences. In contrast, two statistical moments (Mt_2 and Mt_3) and two nonlinear features (*CTM* and *LZC*) obtained from RRV showed p-value < 0.01.

Table 2. Average values (mean ± SD) for the features extracted from AF and RRV signals in the SAHS-positive and the SAHS-negative populations. M_1–M_4: statistical moments of recordings in time domain; *CTM*: central tendency measure; *LZC*: Lempel—Ziv complexity; *ApEn*: approximate entropy; Mf_1–Mf_4: statistical moments obtained from full PSDs; *PA*: maximum of the full PSDs; *BP*: total spectral power; *WD*: Wootters distance; Mf_{1b}–Mf_{4b}: statistical moments obtained from the spectral band of interest (AF: 0.022–0.059 Hz, RRV: 0.095–0.132 Hz.); PA_b: maximum of *PSD* at the spectral band of interest; BP_b: spectral power at the band of interest; WD_b: Wootters distance at the spectral band of interest.

	AF signal			RRV signal		
Feature	SAHS-positive	SAHS-negative	p-value	SAHS-positive	SAHS-negative	p-value
Mt_1	0.06 ± 0.21	0.04 ± 0.11	$p > 0.01$	3.66 ± 0.51	3.64 ± 0.50	$p > 0.01$
Mt_2	190.38 ± 80.30	179.87 ± 91.42	$p > 0.01$	1.04 ± 0.33	0.85 ± 0.28	$\mathbf{p < 0.01}$
Mt_3	0.28 ± 0.24	0.28 ± 0.31	$p > 0.01$	0.81 ± 1.39	0.03 ± 1.11	$\mathbf{p < 0.01}$
Mt_4	8.25 ± 15.83	11.74 ± 21.40	$p > 0.01$	12.9 ± 12.7	9.9 ± 8.0	$p > 0.01$
CTM	0.635 ± 0.184	0.628 ± 0.185	$p > 0.01$	0.989 ± 0.017	0.998 ± 0.002	$\mathbf{p < 0.01}$
LZC	0.279 ± 0.029	0.283 ± 0.027	$p > 0.01$	0.992 ± 0.035	0.975 ± 0.037	$\mathbf{p < 0.01}$
ApEn	0.412 ± 0.073	0.435 ± 0.074	$p > 0.01$	1.46 ± 0.072	1.44 ± 0.075	$p > 0.01$
Mf_1	$8.4 \times 10^3 \pm 8.0 \times 10^3$	$7.9 \times 10^3 \pm 9.4 \times 10^3$	$p > 0.01$	0.22 ± 0.14	0.15 ± 0.10	$\mathbf{p < 0.01}$
Mf_2	$4.7 \times 10^4 \pm 5.1 \times 10^4$	$4.9 \times 10^4 \pm 6.410^4$	$p > 0.01$	1.69 ± 1.03	1.40 ± 1.04	$p > 0.01$
Mf_3	8.02 ± 1.77	8.94 ± 2.19	$\mathbf{p < 0.01}$	12.58 ± 4.28	17.51 ± 4.51	$\mathbf{p < 0.01}$
Mf_4	80.09 ± 35.18	96.98 ± 46.25	$p > 0.01$	232.60 ± 158.77	418.02 ± 184.65	$\mathbf{p < 0.01}$
PA	$59.9 \times 10^4 \pm 68.2 \times 10^4$	$68.5 \times 10^4 \pm 96.7 \times 10^4$	$p > 0.01$	37.55 ± 23.52	39.60 ± 32.36	$p > 0.01$
BP	$1.71 \times 10^7 \pm 1.64 \times 10^7$	$1.63 \times 10^7 \pm 1.93 \times 10^7$	$p > 0.01$	451.15 ± 285.91	307.2 ± 205.49	$\mathbf{p < 0.01}$
WD	0.798 ± 0.022	0.808 ± 0.018	$p > 0.01$	0.904 ± 0.009	0.908 ± 0.008	$\mathbf{p < 0.01}$
Mf_{1b}	$9.9 \times 10^4 \pm 12.9 \times 10^4$	$3.4 \times 10^4 \pm 2.3 \times 10^4$	$\mathbf{p < 0.01}$	2.96 ± 2.53	1.39 ± 0.91	$\mathbf{p < 0.01}$
Mf_{2b}	$39.6 \times 10^3 \pm 10.5 \times 10^3$	$7.1 \times 10^3 \pm 0.8 \times 10^3$	$\mathbf{p < 0.01}$	0.98 ± 0.86	0.56 ± 0.42	$\mathbf{p < 0.01}$
Mf_{3b}	0.042 ± 0.689	-0.451 ± 0.634	$\mathbf{p < 0.01}$	0.16 ± 0.39	0.23 ± 0.38	$p > 0.01$
Mf_{4b}	2.402 ± 0.905	2.675 ± 1.026	$p > 0.01$	2.0 ± 0.43	1.99 ± 0.40	$p > 0.01$
PA_b	$16.8 \times 10^4 \pm 30.7 \times 10^4$	$4.4 \times 10^4 \pm 3.710^4$	$\mathbf{p < 0.01}$	4.61 ± 3.72	2.34 ± 1.48	$\mathbf{p < 0.01}$
BP_b	$1.1 \times 10^5 \pm 20.4 \times 10^5$	$5.3 \times 10^5 \pm 3.5 \times 10^5$	$p < 0.01$	49.1 ± 41.6	23.3 ± 15.1	$\mathbf{p < 0.01}$
WD_b	0.109 ± 0.0608	0.063 ± 0.0372	$p < 0.01$	0.14 ± 0.07	0.16 ± 0.08	$p > 0.01$

Table 3. Results from the diagnostic assessment of single features extracted from AF and RRV recordings, derived from the leave-one-out cross-validation procedure. Sen.: sensitivity; Spe.: specificity; Acc.: accuracy; AROC: area under receiver-operating characteristics curve. AROCs > 0.800 are in bold.

	AF signal				RRV signal			
Feature	Sen.(%)	Spe.(%)	Acc.(%)	AROC	Sen.(%)	Spe.(%)	Acc.(%)	AROC
Mt_1	49.00	58.33	52.03	0.520	48.00	52.08	49.39	0.538
Mt_2	53.00	50.00	52.03	0.555	65.00	62.5	64.19	0.669
Mt_3	59.00	37.50	52.03	0.543	60.00	77.08	65.54	0.704
Mt_4	60.00	50.00	56.76	0.537	73.00	47.92	64.86	0.595
CTM	54.00	47.92	52.03	0.510	68.00	75.00	70.27	**0.800**
LZC	57.00	58.33	57.43	0.553	62.00	56.25	60.14	0.654
ApEn	61.00	56.25	59.49	0.585	52.00	50.00	51.35	0.577
Mf_1	58.00	50.00	55.41	0.554	65.00	62.50	64.19	0.676
Mf_2	50.00	47.92	49.32	0.502	61.00	52.08	58.11	0.618
Mf_3	68.00	56.25	64.19	0.634	77.00	68.75	74.32	**0.809**
Mf_4	63.00	56.25	60.81	0.612	79.00	70.83	76.35	**0.807**
PA	48.00	56.25	50.68	0.513	56.00	41.67	51.35	0.528
BP	58.00	50.00	55.41	0.561	65.00	62.50	64.19	0.676
WD	69.00	54.47	64.19	0.631	64.00	56.25	61.49	0.633
Mf_{1b}	71.00	83.33	75.00	**0.826**	61.00	79.17	66.89	0.745
Mf_{2b}	74.00	87.50	78.38	**0.851**	61.00	62.50	61.49	0.702
Mf_{3b}	58.00	68.75	61.49	0.676	50.00	52.08	50.68	0.559
Mf_{4b}	63.00	56.25	60.81	0.581	47.00	52.08	48.65	0.508
PA_b	74.00	83.33	77.03	**0.840**	59.00	66.67	61.49	0.745
BP_b	71.00	83.33	75.00	**0.838**	62.00	79.17	67.57	0.756
WD_b	65.00	70.83	66.89	0.797	61.00	47.92	56.76	0.567

Table 4. Results from the diagnostic assessment of the LR models, derived from the leave-one-out cross-validation procedure. Sen.: sensitivity; Spe.: specificity; Acc.: accuracy; AROC: area under receiver-operating characteristics curve. The number of features introduced as independent variables at each model are in parentheses. AROCs>0.800 are in bold.

Model	Selected features	Sen.(%)	Spe.(%)	Acc.(%)	AROC
AF (21)	WD_b, BP_b, *PA*, Mf_{2b}	84.00	70.83	79.73	**0.889**
RRV (21)	Mf_3, *CTM*	84.00	58.33	75.68	**0.850**
AF-RRV (42)	Mf_3^{RRV}, PA^{AF}, BP_b^{AF}	88.00	70.83	82.42	**0.903**

Furthermore, nine out of 14 RRV spectral features obtained from the full PSD and the band of interest presented statistically significant differences.

The results of the individual diagnostic assessment of the features are shown in table 3. Consistent with the statistical significance test analysis, those features with p-value < 0.01 improved the diagnostic performance of the others. For the AF signal, the highest accuracy (78.38%) and AROC (0.851) were reached by Mf_{2b}. In the case of RRV, the highest accuracy (76.35%) was obtained by Mf_4, whereas Mf_3 achieved the highest AROC (0.809).

4.2. Performance of the FSLR procedure

Table 4 shows the diagnostic results provided by the LR models. The features automatically selected are also specified. The order of appearance of the selected features in the table is the same as the order obtained from the FSLR method.

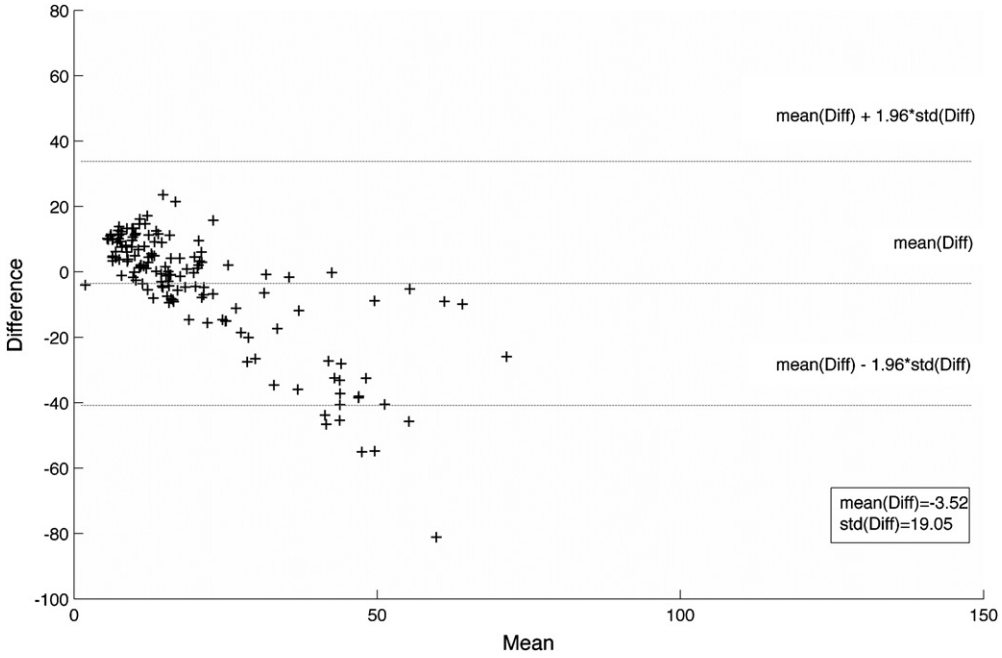

Figure 3. Bland–Altman plot comparing the AHI from MLR_{AF_RRV} with the AHI from PSG.

Table 5. Linear correlation analysis between features selected to the models and BMI/Age. ρ: Pearson's correlation coefficient.

Feature	ρ (BMI)	ρ (Age)
WD_b^{AF}	0.222	0.115
BP_b^{AF}	0.314	0.021
PA^{AF}	0.156	−0.071
Mf_{2b}^{AF}	0.269	−0.003
Mf_3^{RRV}	−0.054	−0.051
CTM^{RRV}	−0.225	−0.262

Four parameters (WD_b, BP_b, Mf_{2b} and PA) were automatically selected by the FSLR procedure when all 21 AF features were introduced as independent variables. The AF model reached 79.73% accuracy and 0.889 AROC. Mf_3 and CTM were automatically selected in the case of the RRV model, obtaining 75.68% accuracy and 0.850 AROC. Finally, the highest sensibility (88.00%), specificity (70.83%), accuracy (82.43%) and AROC (0.903) were achieved by the *AF-RRV* model. Three parameters were automatically selected: Mf_3 from RRV and PA and BP_b from AF. These features were used to obtain a multivariate linear regression model (MLR_{AF_RRV}). The output of MLR_{AF_RRV} (estimated AHI) was compared to the AHI from PSG. The Bland–Altman plot shown in figure 3 was used for this purpose. It displays an overestimation tendency for lower AHI values whereas an underestimation tendency is shown when the AHI becomes higher. Finally, table 5 shows the assessment of linear correlation between all the features automatically selected and the BMI and age. None of the features obtained high values of Pearson's correlation coefficient with BMI or age.

5. Discussion and conclusions

The utility of AF signals in SAHS detection was assessed. The information from RRV series, which was derived from AF, was also evaluated. Statistical, spectral and nonlinear features were used to characterize the behaviour of AF and RRV recordings in SAHS-positive and SAHS-negative populations. AF and RRV frequency bands of interest within the very low frequency region were proposed for SAHS detection. Regarding the AF signal, since the normal breathing rate at rest is set close to 15 breaths per minute, i.e. 0.25 Hz. (Farré *et al* 1998), the spectral components at very low frequencies of AF correspond to an abnormal respiratory behaviour. Moreover, the selected spectral band of interest was located below 0.1 Hz. (0.022–0.059 Hz.). Since apnoea and hypopnoea events last 10 s or more (Flemons *et al* 2003), the band is consistent with pathophysiology. However, further analysis is required in order to assess the cause-motivating differences in frequencies higher than 0.1 Hz. The occurrence of larger number of short-time respiratory events in SAHS-positive subjects is proposed as a cause for this behaviour. In the case of RRV spectrum, significant differences were found in most of the spectral components, indicating major changes in time between breathings caused by SAHS. Most of the AF features that achieved significant differences between populations were extracted from the spectral band of interest. In contrast, p-value <0.01 was achieved by statistical, spectral and nonlinear features from RRV. This suggested that RRV signal contains useful information about SAHS in time and frequency domains. The diagnostic performance of the features reinforced the ideas exposed above: several spectral features from the AF band of interest reached higher values of AROC (0.825–0.851) than any other single feature. This confirmed the usefulness of the spectral information contained in AF signals to help in SAHS diagnosis.

None of the features selected for the *AF*, *RRV* and *AF-RRV* models presented high linear correlation with BMI or age of subjects under study. The *AF-RRV* model achieved the highest diagnostic performance (82.42% accuracy and 0.903 AROC). Mf_3^{RRV}, PA^{AF} and BP_b^{AF} were automatically selected, containing information from both *AF* and *RRV* signals. One out of the two hypothyroidism patients (false positive) and none of the COPD patients were misclassified. These results showed that the FSLR procedure improved the diagnostic performance of single features and suggested that information contained in AF and RRV signals could be complementary.

The AF signals obtained from thermistor have been recently analysed to help in SAHS diagnosis. Two hundred and eighty eight subjects participated in a multi-centre study to evaluate a screening device based on the detection of respiratory events (Shochat *et al* 2002). Thus, 86% sensitivity, 57% specificity and 0.81 AROC were achieved in the classification of subjects. The same methodology was applied to a different population, comparing the screening performance of the device to the performance of nocturnal pulsioximetry (Gergely *et al* 2009). The results of the study achieved 71.9% sensitivity and 73.1% specificity using AHI $= 15$ as a cut-off threshold. The *AF-RRV* model obtained in this work improved the diagnostic performance of both studies.

There exists an extensive literature focused on the analysis of AF from nasal pressure (NP) sensor. Most of them aimed to locate respiratory events in AF. Subsequently, a respiratory disturbance index (RDI) is computed in order to assess its diagnostic performance (De Almeida *et al* 2006, Erman *et al* 2007, Nakano *et al* 2007, Grover and Pittman 2008, Wong *et al* 2008, Tonelli *et al* 2009, Chen *et al* 2009, Rofail *et al* 2010). Populations involved in these studies ranged from 25 to 200 subjects (83.5 ± 64.6, mean $\pm$ SD). Sensitivity, specificity and AROC reached ranged 82%–97%, 62%–90% and 0.84–0.98, respectively. The best results achieved in this study are included in these intervals.

Some limitations have to be taken into account. The population under study could be larger, with a more balanced proportion of SAHS-positive and SAHS-negative subjects. Furthermore, all subjects were suspected of having SAHS before PSG test. A control group (subjects without any symptoms) should be analysed in order to assess the universal application of the methodology. It would provide additional information to complete this study. The use of a thermistor to acquire the AF signal, instead of a NP sensor, is al.so a limitation. Measurements from thermistor are only indirectly related to the AF, resulting in the underdetection of hypopnoeas (Farré *et al* 1998). The NP sensor has shown a better performance for detecting obstructive respiratory events (Bahammam 2004). However, it has a roughly quadratic relationship with the flow, causing AF changes to be exaggerated and, consequently, resulting in an overestimation of apnoea events (Bahammam 2004). The American Academy of Sleep Medicine recommends the use of both types of sensors due to these disadvantages (Iber *et al* 2007). The comparison of features extracted from the signals acquired with the two sensors and the joint analysis of the information extracted from them are future goals. Another limitation has to be considered. The variability of AHI from PSG in successive nights is well known (Carlile and Carlile 2008, Levendowski *et al* 2009). However, night-to-night variability is not often assessed due to economics and time limitations (Levendowski *et al* 2009). Repeated sleep studies in successive nights would be necessary to complete the assessment of this methodology. Finally, the use of an AHI threshold = 10 events/h to discriminate SAHS is also a limitation since subjects in the range 5–10 events/h could benefit from the continuous positive airway pressure (CPAP) treatment. CPAP is the most widely used treatment for severe SAHS (Lindberg *et al* 2006, Marshall *et al* 2006). Further work is needed to assess the accuracy of the methodology for screening those patients who would benefit from CPAP.

In summary, AF and RRV signals were analysed. A spectral band of interest was located in a region of the AF spectrum corresponding to anomalous respiration. The statistical significance test and the diagnostic performance assessment of the features confirmed the usefulness of the information contained in AF and RRV. Results from the FSLR procedure suggested that data extracted from them can complement each other. Moreover, the *AF-RRV* model improved the diagnostic performance of all the single features. The best results obtained from this study improved the results from those studies which involved thermistor and are comparable to those involving NP. Therefore, the proposed methodology could be useful to help in SAHS diagnosis.

Acknowledgments

This work has been partially supported by the project VA111A11-2 from Consejería de Educación de la Junta de Castilla y León, by the Proyectos Cero on Ageing from Fundación General CSIC, by Consejería de Educación de la Junta de Castilla y León (Orden EDU/1204/2010) and by the European Social Found.

References

Abásolo D, Hornero R, Gómez C, García M and López M 2006 Analysis of background activity in Alzehimer's disease patients with Lempel–Ziv complexity and central tendency measure *Med. Eng. Phys.* **28** 315–22

Álvarez D, Hornero R, Abásolo D, del Campo F and Zamarrón C 2006 Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection *Physiol. Meas.* **27** 399–412

Álvarez D, Hornero R, Marcos J V and del Campo F 2010 Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis *IEEE Trans. Biomed. Eng.* **57** 2816–24

BaHammam A 2004 Comparison of nasal prong pressure and thermistor measurements for detecting respiratory events during sleep *Respiration* **71** 385–90

Carlile J and Carlile N 2008 Repeat study of 149 patients suspected of having sleep apnea but with an AHI<5 *Sleep* **31** A153

Casolo G, Balli E, Fazi A, Gori C, Freni A and Gensini G 1991 Twenty-four-hour spectral analysis of heart rate variability in congestive heart failure secondary to coronary artery disease *Am. J. Cardiol.* **67** 1154–8

de Chazal P, Heneghan H, Sheridan E, Reilly R, Nolan P and O'Malley 2003 Automated processing of single-lead electrocardiogram for the detection of obstructive sleep apnoea *IEEE Trans. Biomed. Eng.* **50** 686–96

Chen H, Lowe A A, Bai Y, Hamilton P, Fleetham J A and Almeida F R 2009 Evaluation of a portable recording device (ApneaLink^{TM}) for case selection of obstructive sleep apnea *Sleep Breath* **13** 213–9

Cohen M E, Hudson D L and Deedwania P C 1996 Applying continuous chaotic modelling to cardiac signal analysis *IEEE Eng. Med. Biol. Mag.* **15** 97–102

Cysarz D, Zerm R, Bettermann H, Frühwirth M, Moser M and Kröz M 2008 Comparison of respiratory rates derived from heart rate variability, ECG amplitude, and nasal/oral airflow *Ann. Biomed. Eng.* **36** 2085–94

De Almeida F R, Ayas N T, Ueda H, Hamilton P, Ryan F C and Lowe A A 2006 Nasal pressure recordings to detect obstructive sleep apnea *Sleep Breath* **10** 62–69

Erman M K, Stewart D, Einhorn D, Gordon N and Casal E 2007 Validation of the ApneaLink^{TM} for the screening of sleep apnea: a novel and simple single-channel recording device *J. Clin. Sleep Med.* **3** 387–92

Farré R, Montserrat J M, Rotger M, Ballester E and Navajas D 1998 Accuracy of thermistors and thermocouples as flow-measuring devices for detecting hypopnoeas *Eur. Respir. J.* **11** 179–82

Flemons W W, Douglas N J, Kuna S T, Rodenstein D O and Wheatley J 2004 Access to diagnosis and treatment of patients with suspected sleep apnea *Am. J. Respir. Crit. Care Med.* **169** 668–72

Flemons W W, Littner M R, Rowley J A, Gay P, Anderson W M, Hudgel D W, McEvoy D and Loube D I 2003 Home diagnosis of sleep apnea: a systematic review of the literature *Chest* **124** 1543–79

Gergely V, Pallos H, Mashima K, Miyazaki S, Tanaka T, Okawa M and Yamada N 2009 Evaluation of the usefulness of the SleepStrip for screening obstructive sleep apnea–hypopnea syndrome in Japan *Sleep Biol. Rhythms* **7** 43–51

Grover S S and Pittman S D 2008 Automated detection of sleep disordered breathing using a nasal pressure monitoring device *Sleep Breath* **12** 339–45

Han J, Shin H B, Jeong D U and Park K S 2008 Detection of apnoeic events from single channel nasal airflow using 2nd derivative method *Comput. Methods Programs Biomed.* **98** 199–207

Hornero R, Aboy M, McNames J and Goldstein B 2005 Interpretation of approximate entropy. Case studies in the analysis of intracranial pressure during elevations in traumatic brain injury *IEEE Trans. Biomed. Eng.* **53** 1671–80

Hornero R, Alonso A, Jimeno N, Jimeno A and López M 1999 Nonlinear analysis of time series generated by schizophrenic patients *IEEE Eng. Med. Biol. Mag.* **3** 84–90

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

Hosmer D W and Lemeshow S 1999 *Applied Logistic Regression* (New York: Wiley)

Iber C, Ancoli-Israel S, Chesson A and Quan S F 2007 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Westchester, IL: American Academy of Sleep Medicine)

Korten J B and Haddad G G 1989 Respiratory waveform pattern recognition using digital techniques *Comput. Biol. Med.* **19** 207–17

Lempel A and Ziv J 1976 On the complexity of finite sequences *IEEE Trans. Inf. Theory* **24** 530–6

Levendowski D, Steward D, Woodson B T, Olmstead R, Popovic D and Westbrook P 2009 The impact of obstructive sleep apnea variability measured in-lab versus in-home on sample size calculations *Int. Arch. Med.* **2** 1–8

Lindberg E, Berne C, Elmasry A, Hedner J and Janson C 2006 CPAP treatment of a population-based sample—What are the benefits and the treatment compliance? *Sleep Med.* **7** 553–60

Lindberg E, Carter N, Gislason T and Janson C 2001 Role of snoring and daytime sleepiness in occupational accidents *Am. J. Respir. Crit. Care Med.* **164** 2031–5

López-Jiménez F, Kuniyoshi F H S, Gami A and Somers V K 2008 Obstructive sleep apnea: implications for cardiac and vascular disease *Chest* **133** 793–804

Marshall N S, Barnes M, Travier N, Campbell A J, Pierce R J, McEvoy R D, Neill A M and Gander P H 2006 Continuous positive airway pressure reduces daytime sleepiness in mild to moderate sleep apnoea: a meta-analysis *Thorax* **61** 430–4

Morillo D S, Rojas J L, Crespo L F, León A and Gross N 2009 Poincaré analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hipopnea syndrome *Physiol. Meas.* **30** 405–20

Nakano H, Tanigawa T, Furukawa T and Nishina S 2007 Automatic detection of sleep-disordered breathing from single-channel airflow record *Eur. Respir. J.* **29** 728–36

Patil S P, Schneider H, Schwartz A R and Smith P L 2007 Adult obstructive sleep apnea: pathophysiology and diagnosis *Chest* **132** 325–37

Penzel T, McNames J, de Chazal P, Raymond B, Murray A and Moody G 2002 Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings *Med. Biol. Eng. Comput.* **40** 402–7

Pincus S M 1991 Approximate entropy as a measure of system complexity *Proc. Natl Acad. Sci.* **88** 2297–301

Pincus S M 2001 Assessing serial irregularity and its implications for health *Ann. New York Acad. Sci.* **954** 245–67

Poyares D, Guilleminault C, Rosa A, Ohayon M and Koester U 2002 Arousal, EEG spectral power and pulse transit time in UARS and mild OSAS subjects *Clin. Neurophysiol.* **113** 1598–606

Poza J, Hornero R, Abásolo D, Fernández A and García M 2007 Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease *Med. Eng. Phys.* **29** 1073–83

Roche F, Gaspoz J M, Court-Fortune I, Minini P, Pichot V, Duverney D, Costes F, Lacour J R and Barthélémy J C 1999 Screening of obstructive sleep apnea syndrome by heart rate variability analysis *Circulation* **100** 1411–5

Rofail L M, Wong K K H, Unger G, Marks G B and Grunstein R R 2010 The role of single-channel nasal airflow pressure transducer in the diagnosis of OSA in the sleep laboratory *J. Clin. Sleep Med.* **6** 349–56

Sassani A, Findley L J, Kryger M, Goldlust E, George C and Davidson T M 2004 Reducing motor-vehicle collisions, cost, and fatalities by treating obstructive sleep apnea syndrome *Sleep* **27** 453–8

Shochat T, Hadas N, Kerkhofs M, Herchuelz A, Penzel T, Peter J H and Lavie P 2002 The SleepStrip^{TM}: an apnoea screener for the early detection of sleep apnoea syndrome *Eur. Respir. J.* **19** 121–6

Tonelli A C, Martinez D, Vasconcelos L F T, Cadaval S, Lenz M C, Costa S, Gus M, Abreu-Silva O, Beltrami L and Fuchs F D 2009 Diagnostic of obstructive sleep apnea syndrome and its outcomes with home portable monitoring *Chest* **135** 330–6

Welch P D 1967 The use fast Fourier transform of the estimation of power spectra: a method based on time averaging over short, modified periodograms 1967 *IEEE Trans. Audio Electroacoust.* **15** 70–3

Wong K K H, Jankelson D, Reid A, Unger G, Dungan G, Hedner J A and Grunstein R R 2008 Diagnostic test evaluation of a nasal flow monitor for obstructive sleep apnea detection in sleep apnea research *Behav. Res. Methods* **40** 360–6

Wootters W K 1981 Statistical distance and Hilbert space *Phys. Rev.* D **23** 357–62

Young T, Peppard P E and Gottlieb D J 2002 Epidemiology of obstructive sleep apnea *Am. J. Respir. Crit. Care* **165** 1217–39

Zhang X S, Roy R J and Jensen E W 2001 EEG complexity as a measure of depth anesthesia for patients *IEEE Trans. Biomed. Eng.* **48** 1424–33

Zweig M H and Campbell G 1993 Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine *Clin. Chem.* **39** 561–77