Sleep and Wakefulness State Detection in Nocturnal Actigraphy Based on Movement Information

Alexandre Domingues*, Teresa Paiva, and J. Miguel Sanches, *Senior Member, IEEE*

***Abstract*—Wrist *actigraphy* (ACT) is a low-cost and well-established technique for long-term monitoring of human activity. It has a special relevance in sleep studies, where its noninvasive nature makes it a valuable tool for behavioral characterization and for the detection and diagnosis of some sleep disorders. The traditional *sleep/wakefulness* state estimation algorithms from the nocturnal ACT data are unbalanced from a sensitivity and specificity points of view since they tend to overestimate *sleep* state, with severe consequences from a diagnosis point of view. They usually maximize the overall accuracy that does not take into account the highly unbalanced state distribution. In this paper, a method is proposed to appropriately deal with this unbalanced problem, achieving similar sensitivity and specificity scores in the state estimation process. The proposed method combines two *linear discriminant classifiers*, trained with two different criteria involving movement detection to generate a first state estimate. This result is then refined by a *Hidden Markov Model*-based algorithm. The global accuracy, the sensitivity, and the specificity of the method are 77.8%, 75.6%, and 81.6%, respectively, performing better than the tested algorithms. If the performance is assessed only for movement periods, this improvement is even higher.**

***Index Terms*—Actigraphy (ACT), hidden Markov model (HMM), linear discriminant classifier (LDC), movement detection, sleep/wake estimation.**

I. Introduction

WRIST actigraphy (ACT) has received great attention since the publication, by the *American Sleep Disorders Association*, of the guidelines for its application in the clinical environment [1]. Its relevance in the scope of sleep disorders is well documented in the reviews [2], [3], where it is shown that the number of publications including ACT is rapidly increasing despite its performance still being far from the polysomnography (PSG), the golden standard for the sleep disorder diagnosis [4].

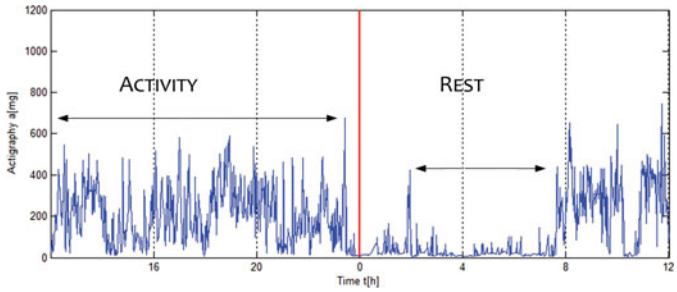

Fig. 1. ACT data acquired over a 24-h period. Rest and activity periods are clearly distinguishable.

The highly portable and noninvasive nature of ACT sensors makes them ideal for long-term monitoring applications. They are a valuable tool to gather behavioral information and to obtain estimates of some sleep parameters, such as sleep efficiency and fragmentation [5], and for the characterization of the circadian cycle [6], [7].

Fig. 1 shows a typical segment of ACT data over a 24-h period, the rest and activity periods along the circadian cycle are clearly visible.

Certain disorders, such as circadian phase shifts, are accurately detected from the ACT data [8], but sleep staging and accurate *sleep/wakefulness* (*S/W*) state discrimination are still open issues and active fields of research.

The different levels of agreement between PSG and ACT reported in the literature have raised some issues regarding the validity of ACT for *S/W* estimation [9]–[12] and the metrics used to evaluate the suggested algorithms [13].

The validation of the ACT prediction rates is typically made from the hypnogram obtained from the PSG data. Although this information is accurate, it is also unbalanced from a state distribution point of view. In fact, in a healthy subject hypnogram, at least 85% of the epochs correspond to *Sleep* state [14]–[16].

Thus, the high accuracies and sensitivities reported in *S/W* estimation using the nocturnal ACT data, often mask the low specificity associated to the poor wake detection ability, as reported in [9]. Table I shows some of the most relevant results obtained in *S/W* state estimation in adults.[1] It illustrates how diverse is the performance of the methods and mainly, how different are the sensitivity and specificity in most of them.

In the assessment of the performance of new algorithms and comparison with existing methods, common datasets and

Manuscript received February 11, 2013; revised June 5, 2013 and August 13, 2013; accepted August 20, 2013. Date of publication September 4, 2013; date of current version January 16, 2014. This work was supported by project the FCT (ISR/IST plurianual funding) through the PIDDAC Program funds and FCT project Detection of Brain Micro-states in Fibromyalgia (PTDC/SAU-BEB/104948/2008). *Asterisk indicates corresponding author.*

*A. Domingues is with Institute for Systems and Robotics/Bioengineering Department, Instituto Superior Técnico, Technical University of Lisbon, 1349–055 Lisbon, Portugal (e-mail: adomingues@gmail.com).

T. Paiva is with the Centro de Electroencefalografia e Neurologia Clinica, Faculdade de Medicina da Universidade de Lisboa, 1649-004 Lisbon, Portugal (e-mail: teresapaiva0@gmail.com).

J. M. Sanches is with Institute for Systems and Robotics/Bioengineering Department, Instituto Superior Técnico, Technical University of Lisbon, 1349–055 Lisbon, Portugal (e-mail: jmrs@ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2280538

[1]The methods from Sadeh *et al*. and Hedner *et al*. were implemented to compare the performance of the proposed algorithm.

TABLE I
PERFORMANCES REPORTED IN THE LITERATURE FOR *S/W* ESTIMATION USING THE ACT DATA IN ADULT POPULATIONS

Author	Sens (%)	Spec (%)	Acc (%)	G-mean (%)
Cole et al. [6]	95	**64**	88	**78**
Sadeh et al. [7]	96	**75**	92	**84**[1]
Kushida et al. [17]	90	**45**	80	**64**
Ancoli-Israel et al. [18]	−−	−−	93	−−
Hedner et al. [19]	89	**69**	86	**78**[1]
Sivertsen et al. [11]	95	**36**	83	**59**
Paquet et al. [12]	95	**54**	90	**72**

figures of merit must be used. In [14], for instance, the authors implemented and compared the algorithms described in [6] and [7] obtaining results significantly different than the originally reported.

In order to overcome the intrinsic limitations of ACT, new approaches are being explored. In [20], cardiorespiratory signals are combined with the ACT data yielding promising results, and in [21], the authors show that it is possible to accurately characterize the human activity from the accelerometer data in a nonsleep scope. In [13], the authors employ artificial neural networks and decision trees to score the infant ACT data obtaining relevant results and stressing the importance of using common metrics. In [22] and [23], the authors hypothesized that movements during *wakefulness* and *sleep* states are intrinsically different, not only in terms of magnitude but from a spectral and statistical distribution perspective. While movements during the *sleep* state are typically random and without purpose, i.e., *purposeless*, movements during *wakefulness* state are more coherent and correlated, usually with a defined purpose.

The method proposed in this paper for *S/W* estimation, is based on a *movement detector* (MD) designed to discriminate and detect movement and quietness events from the nocturnal ACT data.

The classification/testing procedure is performed in three stages. After the feature extraction, an estimation of the S/W state is obtained for each epoch with a linear discriminant classifier (LDC) [24], with parameters θ^* and θ^*_M, during the quietness and movement epochs, respectively, as represented in Fig. 3. The previous result is then refined with a *Hidden Markov Model* (HMM) that incorporates statistical information computed from the training data.

The features used in the LDC's are the coefficients of a Rayleigh mixture distribution, to describe the first-order statistics of the ACT data, the residues of an *Autoregressive* (AR) model fitted to the data, describing its high-order statistics and finally the signal energy that takes into account the intensity of the signal, the most important feature used in the traditional approaches for ACT data processing.

ACT-based *S/W* estimation algorithms typically combine several features extracted from the intensity and frequency counts of the recorded signal. In these algorithms, periods of strong activity are normally scored as *wakefulness* and long periods quietness as *sleep*. This strategy leads to acceptable and relevant accuracies but also to the well-documented limitations of ACT such as the poor ability to detect wakening episodes during quietness periods, very typical in insomnia, and the generalized tendency to overestimate *sleep* [8], [13]. Since sleeping is the natural state during the night, when estimating *S/W* states, for sleep disorders diagnosis purposes, it is generally more important to accurately estimate *wakefulness* than *sleep*. The traditional state estimation algorithms usually maximize the overall accuracy that does not take into account this unbalanced state distribution, which leads to poor wakefulness state detection rate.

The proposed method is designed to achieve a similar performance during movement and quietness periods and it is tuned, using the *Geometric mean* (G-mean) [25] as the optimization criteria, for a balance between *sleep* and *wakefulness* detection ability. Thus, minimizing the tendency of ACT to underestimate the *wakefulness* periods [8], [13].

The presented algorithm is optimized for proper S/W estimation in the scope of sleep disorders diagnosis. Other applications may have different requirements not fulfilled by this method. For example, the detection of rest and activity periods for ambulatory blood pressure monitoring, such as the work described in [26], requires an algorithm less sensitive to microwakening episodes.

II. METHODS

This section describes the method used to estimate the *S/W* states from the nocturnal ACT data. These data are acquired with an actigraph sensor located at the nondominant wrist of the subject. The sensor, which is basically a 3-D accelerometer, provides the magnitude of the acceleration vector.

A. Algorithm

The complete state estimation method, displayed in Fig. 3, is composed by a preprocessing step, feature extraction, training, state estimation, and a final classification refinement.

Two preprocessing operations are performed on the data: 1) magnitude normalization and dc component removal, and 2) movement segmentation.

The magnitude normalization and dc component removal is required to minimize the interpatient and interdevice variability. This procedure, performed in a sliding window basis, is done according to

$$\tilde{a}(n) = \frac{a(n) - \boldsymbol{\mu}(\boldsymbol{n})}{\boldsymbol{\sigma}(\boldsymbol{n})} \tag{1}$$

where $a(n)$ is the ACT sample, $\boldsymbol{\mu}(\boldsymbol{n})$ and $\boldsymbol{\sigma}(\boldsymbol{n})$ are the mean and standard deviation of the data within the 5 min window centered at the nth sample, respectively, and $\tilde{a}(n)$ is the normalized ACT sample.

In a second preprocessing step, movement events are identified on the normalized data with the MD displayed in Fig. 4.

This detector is composed of 1) a noncausal low-pass stretching filter and 2) a threshold-based binarization block. The smoothing filter computes the movement envelope, its width is controlled by the parameter p, defined as $p = 5$ for a sampling frequency of 1Hz.

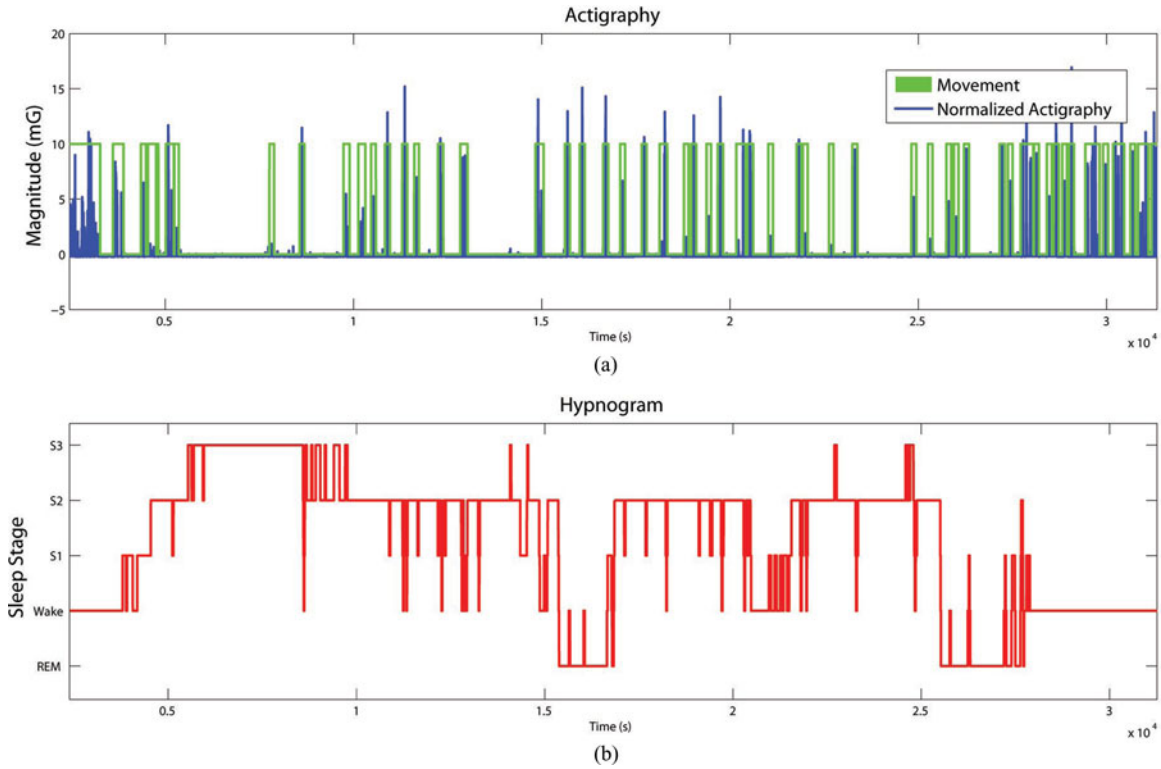

Fig. 2. (a) ACT data and detected movements (top) and (b) Hypnogram (bottom). The binary movement function was rescaled for visualization purposes.

Let $S(t) = |\frac{dF(t)}{dt}|$ denote the sensibility of the MD and $F(t)$ the percentage of movement detected as a function of the threshold, t. The optimal value, t^*, is obtained for each dataset computing S and finding $t^* = \arg\min_x \frac{d^2 S(t)}{dt^2}$. The selected threshold corresponds to the point where the sensibility is more stable.

Fig. 5 illustrates the determination of t^* for one dataset. The sensibility curve is evaluated for the range $t = [0.5\ 1]$, this range was chosen by direct observation of the data. The average optimal threshold, computed from all datasets, is $\hat{t}^* = 0.68 \pm 0.22$. The output of the detector is a binary function $\tau(n) \in \{m, q\}$, where m corresponds to movement and q to quietness.

Fig. 2 displays an example of the preprocessed data. Fig. 2(a) shows the normalized ACT signal and the movement indicator, and Fig. 2(b) the corresponding hypnogram segment. The hypnogram discriminates five different states, namely *wakefulness, Rem* sleep, and three *non-Rem* sleep states. All epochs marked as *Rem* and *non-Rem* were translated into a single *sleep* label.

The next three sections describe 1) the features, 2) the main classification stage (MCS), and 3) the refinement algorithm, which compose the proposed method, here, called *Movement-based State Detection* (MSD).

1) Features: In this paper, an extended set of features is used, one related to ACT intensity and two describing first and higher order statistics, used to discriminate the intrinsic characteristics of the ACT data. After preprocessing, each ACT time course is divided in contiguous epochs of $T = 30$ s.

Let $\mathbf{w}_j$ represent an L-dimensional window, 210 s long, centered on the jth epoch, where $j \in \{1, \ldots, M\}$ with M the total number of epochs. Features are extracted from each window, $\mathbf{w}_j$, as follows.

1) Energy, E_j—The energy of the epoch is $E_j = \sum_k h(k) w_j^2(k)$, where $\mathbf{h} = \{h(k)\}$ is a Hanning window.
2) Residue, r_j—Residue of the eight-order AR model [27] estimated for each $\mathbf{w}_j$ based on an L dimension window, centered on it, as proposed by the authors in [23]. By using a forward search feature selection algorithm [24], it was concluded that the residue is more discriminative than the coefficients of the AR model in the estimation of the *S/W* state. Fig. 6(b) shows the normalized histograms of the residues obtained for *sleep* and *wakefulness* movements. The residues roughly follow a Gaussian distribution with different means and standard deviations for the two states.
3) Coefficients $\alpha_j(k)$—Coefficients of a three component *Rayleigh Mixture Model* (RMM) [22], fitted to each $\mathbf{w}_j$. It was found that only the coefficients, $\alpha_j(k)$ are discriminative for *S/W* state estimation and not the parameters of the Rayleigh components of the mixture. The number of components in the mixture, L, was selected by fitting several mixture distributions, with different values of L, to the complete set of data and measuring the *goodness-of-fit*. The used optimality criterion was the *Kullback–Leibler* (KL) divergence.
   Table II shows that the value of the KL divergence decreases until $L = 3$ and then remains approximately constant for higher values of L. The selected value was thus $L = 3$.

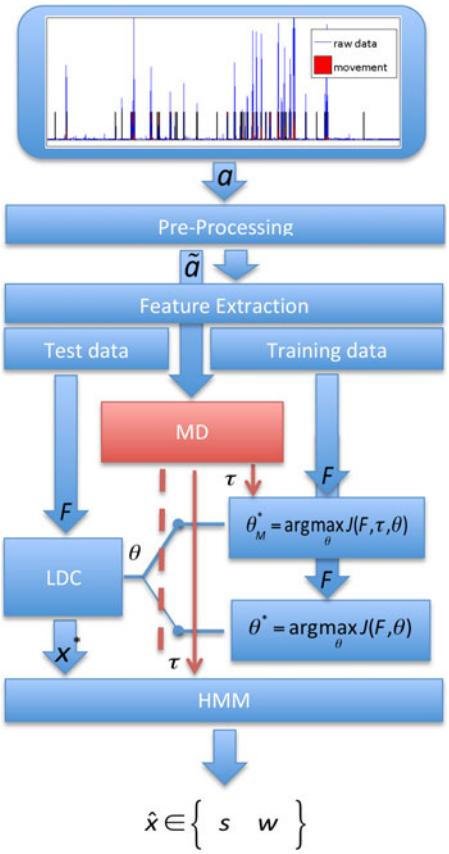

Fig. 3. Fluxogram of the proposed (MSD) algorithm. After preprocessing and feature (F) extraction, an initial estimation (x^*) is made using an LDC with parameters θ^* and $\theta_M{}^*$. The final estimate, $\hat{x}$, is obtained with an HMM-based regularization algorithm.

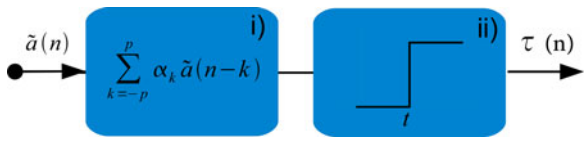

Fig. 4. Structure of the MD.

The physical reasoning to use an RMM is related to the model adopted for the actigraph sensor, as described in [28]. If the acceleration along each axis is described by a zero mean Gaussian distribution, the acceleration magnitude follows a Rayleigh and a Maxwell distribution in 2-D and 3-D, respectively. Fig. 6(a) shows the normalized histograms of Sleep and Wakefulness movement data and the two Rayleigh distributions fitted to the data.

2) Main Classification Stage: The MCS performs the initial estimation of the *S/W* state, denoted as $x^* \in \{s^*, w^*\}$. The used cost function, the G-mean metric, is given by

$$J = \sqrt{\text{sens} * \text{spec}} \tag{2}$$

where sens and spec[2] are the *sensitivity*, the ability of the method to correctly detect *sleep*, and *specificity*, the ability of the method to correctly detect *wakefulness*, respectively. This criterion is adopted for simultaneous maximization of the *sensitivity* and *specificity*, which is not guaranteed when only the overall accuracy (acc) is considered.

[2]$\text{sens} = \frac{\text{TP}}{\text{TP+FN}}$, $\text{spec} = \frac{\text{TN}}{\text{TN+FP}}$, $\text{acc} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$, TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives, respectively.

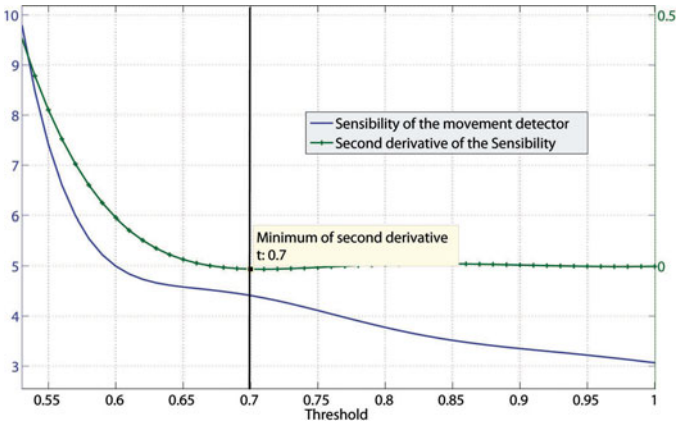

Fig. 5. Sensibility of the MD and the respective second derivative for a given dataset. The minimum of the second derivative is selected as the optimal value for this dataset, corresponding to $t^* = 0.7$.

The MCS is composed of the following three steps.

1) Classification of the test data with LDC(θ^*). The parameters θ^* were obtained through the maximization of the cost function (2), taking into account the whole training dataset.
2) Classification of the test data with LDC(θ^*_M). LDC(θ^*_M) is trained using the whole data but the cost function (2) is optimized taking into account only the movement data ($\tau(n) = m$).
3) Combination of the previous results, where quietness epochs are scored from LDC(θ^*) and movement epochs are scored using LDC(θ^*_M).

3) HMM Regularization Algorithm: This final procedure refines the results obtained in the MCS leading to the final estimation, $\hat{x} \in \{\hat{s}, \hat{w}\}$.

An HMM was chosen for this task since it models processes which have a temporal relation between states, which is the case in the sleep/wake cycle.

Two hidden states are considered, $x \in \{s, w\}$, where s and w refer to *sleep* and *wakefulness* states, respectively.

Let us consider $x^* \in \{s^*, w^*\}$, the output, of the MCS, $\tau \in \{m, q\}$ the output of the MD, and $t \in \mathbb{N}$ the time, in seconds, since the last movement (quietness periods) or since the patient started to move (movement periods).

The observation model takes into account the following information, extracted in the training step:

1) The accuracy rate of the MCS, given by $P(x^*|x)$.
2) The conditional distribution of the activity given the state, expressed as $P(\tau|x)$ and shown in Table III.
3) The duration of the quietness and movement periods during *sleep* and *wakefulness*. Expressed as

$$P(t|x,\tau) = \begin{cases} \mathcal{N}(\sigma_x), & \text{if } \tau = m \\ \mathcal{E}(\lambda_x), & \text{if } \tau = q \end{cases} \tag{3}$$

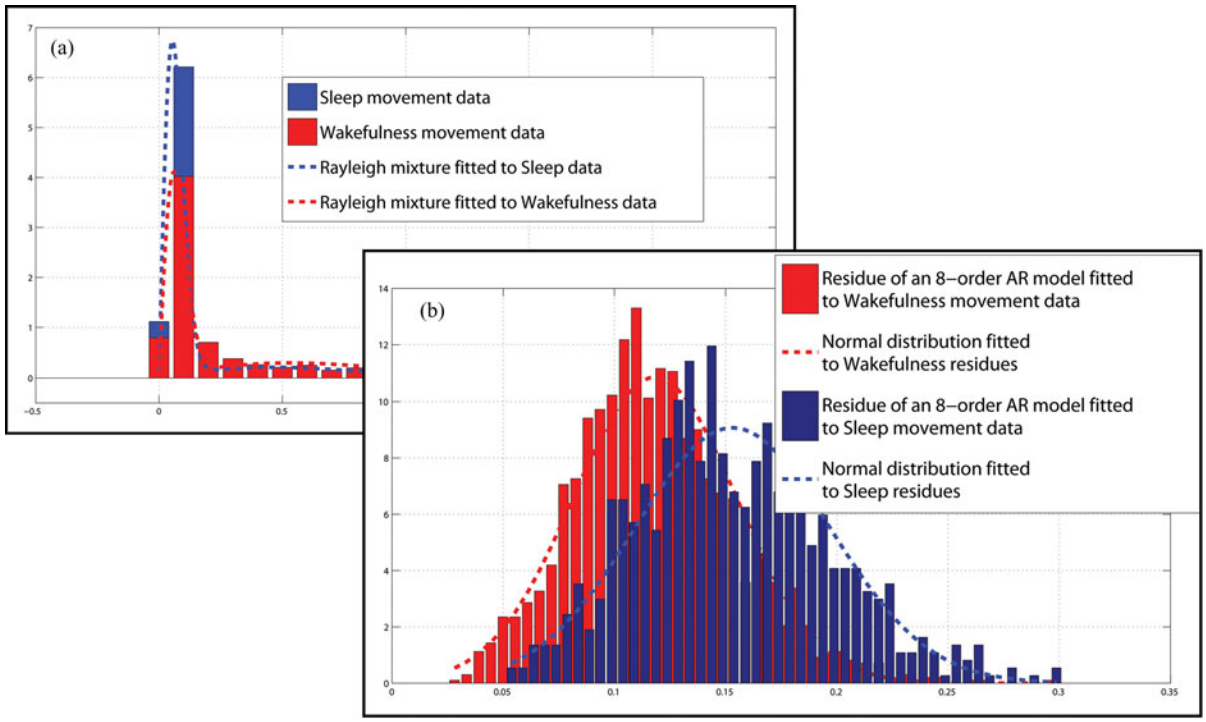

Fig. 6. Normalized histograms of (a) Sleep (blue) and Wakefulness (red) movement data and the respective Rayleigh distributions and (b) Sleep (blue) and Wakefulness (red) residues, obtained using a order eight AR model fitted to the data.

TABLE II
KL DIVERGENCE USING DIFFERENT NUMBER OF COMPONENTS FOR THE RAYLEIGH MIXTURE DISTRIBUTION

L	1	2	3	4	5
$\overline{KL}$	0.230	0.042	**0.036**	0.036	0.035

TABLE III
CONDITIONAL PROBABILITIES OBTAINED FROM THE RELATIVE FREQUENCY OBSERVED IN THE REAL HYPNOGRAMS

$P(x\|\tau)$			
$P(w\|m)$	0.58 ± 0.19	$P(s\|m)$	0.42 ± 0.18
$P(w\|q)$	0.15 ± 0.1	$P(s\|q)$	0.85 ± 0.09
$P(\tau\|x)$			
$P(m\|w)$	0.18 ± 0.07	$P(q\|w)$	0.82 ± 0.07
$P(m\|s)$	0.06 ± 0.03	$P(q\|s)$	0.94 ± 0.04

$x \in \{s, w\}$ denote *sleep* and *wakefulness* states and $\tau \in \{m, q\}$ denote *movement* and *quietness*.

where $\mathcal{N}(\sigma_x)$ is a Gaussian probability distribution with zero mean and $\mathcal{E}(\lambda_x)$ an exponential probability distribution.

$\mathcal{E}(\lambda_x)$ is an exponential probability distribution, $P(t|\lambda_{s,w})$, giving the probability of a *quietness* period of length t being observed during *sleep* and *wakefulness*. This distribution arises naturally if movement events are assumed to be a stochastic Poisson process. The parameters λ_s and λ_w are computed as

$$\widehat{\lambda}_x = \frac{1}{\bar{t}_x} \tag{4}$$

where $\bar{t}_x$ is the mean duration of all *quietness* intervals for *sleep* ($x = s$) and *wakefulness* ($x = w$) states. $\mathcal{N}(\sigma_x)$ is a zero mean Gaussian probability distribution, $P(t|\sigma_{s,w})$, giving the probability of a movement of length t being observed during *sleep* and *wakefulness* states. The parameters σ_s and σ_w are the standard deviation of the duration of all movements recorded during *sleep* and *wakefulness* states, respectively.

Fig. 7(a) shows the histograms of the length of the quietness periods during *sleep* and *wakefulness* states. As expected, the mean value is larger during sleep and the periods tend to be longer. Fig. 7(b) shows the histogram of the length of the recorded movements, as expected, during wakefulness, movement duration is typically longer and has a higher standard deviation.

The probability of any observation $y = \{x^*, \tau, t\}$ given the state $x \in \{s, w\}$ is expressed as

$$\begin{aligned} P(x^*, \tau, t|x) &= P(x^*|x)P(t, \tau|x) \\ &= P(x^*|x)P(\tau|x)P(t|x, \tau). \end{aligned} \tag{5}$$

The transition matrix is computed from the training data as

$$P = \begin{bmatrix} \frac{N(\mathrm{SS})}{N(\mathrm{SS})+N(\mathrm{SW})} & \frac{N(\mathrm{SW})}{N(\mathrm{SS})+N(\mathrm{SW})} \\ \frac{N(\mathrm{WS})}{N(\mathrm{WS})+N(\mathrm{WW})} & \frac{N(\mathrm{WW})}{N(\mathrm{WS})+N(\mathrm{WW})} \end{bmatrix} \tag{6}$$

where $N(.)$ is a counting operator for ss, sw, ws, and ww, corresponding to *sleep–sleep*, *sleep–wakefulness*, *wakefulness–sleep*, and *wakefulness–wakefulness* transitions, respectively.

The hidden state, $x(t)$, is estimated along the time from the observations, $y(t)$, and the model parameters. The initial probabilities are set to 0.0 and 1.0 for *Sleep* and *Wakefulness* states,

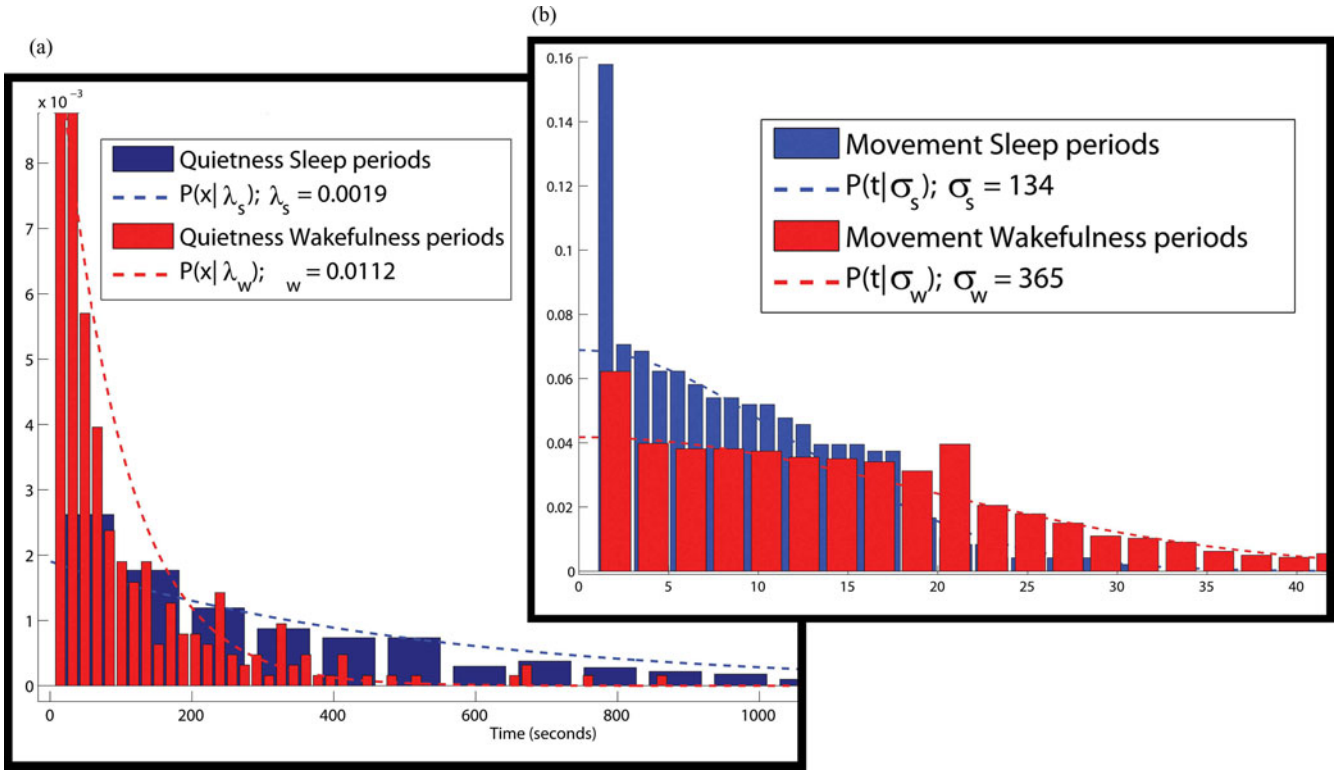

Fig. 7. Normalized histograms of (a) the length of *quietness* periods and (b) duration of the recorded movements, during *sleep* (blue) and *wakefulness* (red) states.

respectively, because all the patients were awake in the beginning of the exam, and the optimal solution, the most probable state sequence, is computed using the *Viterbi Algorithm* [29].

B. Comparative Methods

In this section, two of the methods for *S/W* discrimination listed in Table I, are described, as well as the adaptations required for their use with the existing datasets. The two methods, here called Sadeh's [7] and Hedner's [19], were selected by their reported performance, particularly its $Sens$ and $Spec$ balance.

1) Sadeh's Algorithm: In [7], Sadeh *et al.* propose a scoring algorithm, using 60 s epochs, that linearly combines several features in the following discriminative function:

$$PS = \boldsymbol{\alpha}\boldsymbol{\theta}^T \tag{7}$$

where $\boldsymbol{\alpha}$ is a vector of adjustable parameters defined in [7] as

$$\boldsymbol{\alpha} = [7.601; -0.065; -1.08; -0.056; -0.703]\,. \tag{8}$$

$\boldsymbol{\theta}$ is a vector of features extracted from the data expressed as

$$\boldsymbol{\theta} = [1; \mu; Nat; \sigma; LogAct] \tag{9}$$

where μ is the mean number of activity counts on a 11 min window centered in the current epoch, Nat is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts in a window of 11 min, σ is the standard deviation of the activity on the last 6 min, and $LogAct$ is the natural logarithm of the number of activity counts during the scored epoch plus 1.

A given epoch is scored as *sleep* if $PS >= 0$ and *wakefulness* otherwise.

Since the initial algorithm was developed for a different Actigraph device and database, the five parameters from the discriminative function were optimized for the current data. The optimal parameters were found maximizing the cost function given by (2) leading to

$$\boldsymbol{\alpha}^* = [4.097; -0.528; -0.51; -0.259; -1.65]\,. \tag{10}$$

2) Hedner's Algorithm: In [19], Hedner *et al.* present an *S/W* state estimation algorithm with focus on sleep apnea patients. The algorithm is divided in four distinct steps with the last one aiming at the detection of periodic movements, typical from the apnea patients. Here, the last step is discarded and the algorithm is as follows.

1) Determination of the background movement activity of the patient throughout the night, σ.
2) Bandpass filter between 2 and 2.5 Hz, leading to a signal regarded as the energy of the activity.
3) For each 30 s epoch, values of energy below σ are discarded and the remaining energy is integrated using a 5-min Hanning window. Values below a fixed threshold, θ are scored as *sleep* and values above are scored as *wakefulness*.

The two parameters, σ and θ were computed maximizing the cost function given by (2).

III. Results

This section describes the data used in this paper, presents the experimental results obtained with the MSD, and compares them with the two methods described in the previous section. The

performance is assessed with several *Figures of Merit* (FOM). These FOMs are computed in a *leave-one-patient-out* cross validation basis, where each patient dataset is tested after training the classifier with the remaining data. All the classification routines are implemented using *PRTools* [24] for MATLAB.

A. Data

The nocturnal ACT data were acquired with a *Somnowatch* device, from Somnomedics, placed in the nondominant wrist of the subjects, acquiring with a sampling rate of 1 Hz. The core of these devices is a 3-D accelerometer that measures the acceleration along three orthogonal axis with a configurable output format.

Here, the output of the actigraph is the acceleration magnitude. Some authors suggest that this configuration, also known as digital integration, is the most reliable to measure activity levels [8], [30].

The nocturnal ACT data were jointly acquired with the PSG data for validation purposes and the hypnogram obtained from the PSG by trained technicians, is used as a ground truth to identify *sleep* and *wakefulness* states in epochs of 30 s. Twenty-nine adult subjects (age 48 ± 13 years, 13 Males, 16 Females), with no particular prediagnosed sleep disorder, participated in this study.

The *Sleep Efficiency* (SE), computed as the ratio between total *sleep time* and total *bed time* was obtained for each patient. All the values of SE fell within the range 75%–85%. These values are below the typical values found in healthy subjects, usually above 85% [31], which indicated sleep disturbances, although not necessarily pathological.

The normalization step applied to the data reduces the variability observed in the datasets recorded with distinct devices. This step also contributes to the generalization of the described algorithm to the data acquired with different models/brands of actigraph devices.

B. ACT Data Characterization

The nocturnal ACT data are highly unbalanced from a state distribution point of view. This can be confirmed in Table III where experimental conditional distribution means and standard deviation values, computed from the relative frequencies observed in the real hypnograms are displayed.

As expected, during movement periods, the most frequent state is *wakefulness*, $P(w|m) = 0.58$, although closely followed by *sleep*, $P(s|m) = 0.42$. During quietness periods, the gap between the two states is larger, $P(s|q) = 0.85$ against $P(w|q) = 0.15$. This observation suggests high correlations between movement and *wakefulness* state and quietness and *sleep* state, respectively, but in fact, the probability of a patient moving during wakefulness is much smaller than the probability of not moving, as can be seen from $P(m|w)$ and $P(q|w)$. This fact illustrates the main limitation of nocturnal ACT for *S/W* state estimation: although the methods are based on the recorded movements, they only occur during $6 \pm 3\%$ of the time in the whole register.

The information from Table III clarifies why simple empirical classification rules can actually lead to apparently impressive performances. The traditional FOMs, accuracy, sensitivity, and specificity, are not able to cope with the type of unbalanced data present on the nocturnal ACT.

TABLE IV
PERFORMANCE OF TWO NAIVE CLASSIFICATION POLICIES: M1) ALL DATA ARE SCORED AS SLEEP AND M2) MOVEMENT IS SCORED AS *WAKEFULNESS* AND QUIETNESS IS SCORED AS *SLEEP*

	$Sens(\%)$	$Spec(\%)$	$Acc(\%)$	G-mean(%)
M1	100	0	80.4 ± 5.9	0
M2	97.2 ± 2.5	22.9 ± 7	84.8 ± 5.4	47.2

TABLE V
MEAN AND STANDARD DEVIATION OF THE Ssb, $Spec$ AND Acc AND G-MEAN OBTAINED FOR THE DIFFERENT COMPONENTS OF THE MSD

		$Sens$ (%)	$Spec$ (%)	Acc (%)	G-mean (%)
LDC(θ^*)	*i*)	75.9 ± 9.8	74.8 ± 9.6	75.9 ± 8.7	75.3
	ii)	15.8 ± 11.6	96.6 ± 2.9	65.7 ± 12.2	39.1
LDC(θ^*_M)	*ii*)	74.5 ± 16.8	63.4 ± 11.6	66.5 ± 11.3	68.7
MCS	*i*)	72.1 ± 3.2	73.6 ± 11.1	71.3 ± 8.2	72.8
MSD	*i*)	75.6 ± 8.3	81.6 ± 7.5	77.8 ± 8.1	78.5
	ii)	73.8 ± 14.2	73.5 ± 9.5	75.5 ± 9.2	73.7

The performance was assessed considering i) all the data and ii) only movement data.

The classification results displayed in Table IV, obtained with two naive methods, are used to illustrate the previous point.

The method M1 classifies all the data as *sleep* resulting in a surprising global accuracy of 80.4%. This result is relevant because it shows that a good accuracy is not a good indicator of the performance of the method since it has no ability to detect *wakefulness* state. The second method, M2, classifies all the quietness periods as *sleep* and all the movement periods as *wakefulness*. Even by misclassifying all the *sleep* epochs during movement periods, the method is able to achieve a sensitivity of 97.2% and global accuracy of 84.8%. The limitation of both the methods is revealed by the low specificity and G-mean.

C. S/W Classification

Table V summarizes the results obtained with the individual LDC classifiers, $LDC(\theta^*)$ and $LDC(\theta^*_M)$, the results of the intermediate MCS (which is the combination of both LDC's) and the final MSD. Some results are presented for two distinct scenarios; 1) when all the data are considered in the classification and 2) when only movement data are considered ($\tau(n) = m$).

The $LDC(\theta^*)$ classifier achieves a global accuracy of 75.9% with sensitivity, specificity, and G-mean in the same range. The limitation of this classifier arises when only the movement data are considered, with the sensitivity falling to 15.8%. This means that during movement periods the classifier tends to classify all epochs as *wakefulness*.

TABLE VI
MEAN AND STANDARD DEVIATION OF THE Ssb, $Spec$, AND Acc OBTAINED FOR THE THREE CONSIDERED METHODS IN TWO CONDITIONS: I) COMPLETE DATA SET AND II) ONLY MOVEMENT DATA

		$Sens(\%)$	$Spec(\%)$	$Acc(\%)$	G-mean(%)
MSD	i)	$\mathbf{75.6 \pm 8.3}$	$\mathbf{81.6 \pm 7.5}$	$\mathbf{77.8 \pm 8.1}$	**78.5**
	ii)	$\mathbf{73.8 \pm 14.2}$	73.5 ± 9.5	$\mathbf{75.5 \pm 9.2}$	**73.7**
Sadeh	i)	75.1 ± 6.5	61.6 ± 15.2	73.9 ± 3.7	68.0
	ii)	47.3 ± 12.8	75.8 ± 9.2	62.4 ± 12.8	59.9
Hedner	i)	73.6 ± 10.7	68.6 ± 11	74.1 ± 9.3	71.1
	ii)	14.5 ± 14.1	$\mathbf{88.7 \pm 26.1}$	64.2 ± 16.3	35.9

The $LDC(\theta^*_M)$ is only evaluated for the movement segments. It achieves a G-mean of 68.7%, approximately 30% higher than the $LDC(\theta^*)$ during the same periods. On the other hand, the specificity drops to 63.4% due to the *wakefulness* periods misclassified as *sleep*.

The MCS combines the scores from the two LDC's. It achieves a G-mean of 72.8%, when the complete data are considered (the decrease in specificity is due to reason explained before) and the performance during movements is similar to the $LDC(\theta^*_M)$.

Finally, the performance of the MSD clearly reflects the improvement obtained with the HMM. It achieves a G-mean of 78.5% when all the data are considered and 73.7% when limited to movement data.

It is important to stress the balance of $Sens$ and $Spec$ achieved with the proposed method in global terms but especially during movement periods. These results can be very useful when ACT is used together with other physiological data, e.g., *electrocardiography*, whose sensors are typically sensitive to movement artefacts.

In order to assess the generalization capability of the algorithm, the following procedure was performed:

1) Ten datasets were randomly selected from the pool of 29 available datasets.
2) From these ten datasets, five were randomly selected to train the algorithm.
3) The remaining five datasets were used to test the algorithm and the average G-mean was computed.

This procedure was repeated 15 times resulting in an average G-mean of $76.3 \pm 2\%$. This value is only 2% smaller than the G-mean reported in Table V and, together with the low standard deviation, suggests that the reported results should be extensible to other datasets.

The sensibility of the method, to small variations on the MD threshold, was assessed by forcing random variations of $\pm 20\%$ on each dataset threshold. The variation in the G-mean's for 1) All the data and for 2) Movement data (see Table V) was less that 1% in average.

Table VI compares the results obtained with the MSD with the two comparative methods. Using the complete datasets, MSD achieves higher sensitivity, specificity, global accuracy, and G-mean than the considered methods. While the difference in global Acc is relatively small ($\approx 3\%$), the increase in the G-mean is 10.5% and 7.4%. This result clearly illustrates the limitation of using the global accuracy as the only performance metric. When only movement periods are considered the MSD clearly outperforms the comparative methods, which present a bias to classify the movement periods as *wakefulness,* thus, achieving a low $Sens$.

D. Final Remarks

The nocturnal ACT can be roughly clustered in two classes of 1) quietness periods, when only background activity and noise are recorded, and 2) movement periods.

During the periods of quietness the relevant information for *S/W* state estimation lies on the intermovement durations, whose statistics are shown in Fig. 7(a), and on the *a priori* correlation knowledge about activity and state, described in Table III. The ACT signal magnitude itself does not provide useful information during these periods.

On the contrary, during movement periods the magnitude of the ACT signal is relevant for state estimation but the *a priori* information ($P(s|m) = 0.42$ and $P(w|m) = 0.58$) is not so important for estimation process as in the quietness periods ($P(s|q) = 0.85$ and $P(w|q) = 0.15$).

The magnitude characterization in these periods can be performed in a spectral and statistical basis, namely, with first and second order statistics, displayed in Fig. 6.

IV. CONCLUSION

This paper describes a state estimation algorithm (MSD) from the nocturnal ACT data, focused on movement periods obtained from an MD.

The method uses an extended set of features related to the signal magnitude and time events. Two LDC's, trained with the magnitude related features, provide a first state estimation that is refined with a HMM-based algorithm that takes into account time events and *a priori* information about movements and states correlation.

Relevant novelties presented in this paper are related to the optimization strategy in the training process based on the G-mean metrics. The goal is to improve simultaneously the performance in movement and quietness periods as well as the balance between specificity and sensitivity which improves the *wakefulness* detection rate.

A new database of the ACT data was built specifically for this project. With these data, the MSD yields a global accuracy of 77.8%, a sensitivity of 75.6%, and a specificity of 81.6%, revealing a balance in the detection of both the *sleep* and *wakefulness* states, a key issue of this paper. Additionally, under the G-mean metrics the proposed method clearly outperforms the other tested methods. During the movement periods, the method achieves an accuracy of 75.5%, sensitivity of 73.8%, specificity of 73.5%, and G-mean 73.7%.

REFERENCES

[1] M. Thorpy, A. Chesson, S. Derderian, G. Kader, R. Millman, S. Potolicchio, G. Rosen, and P. J. Strollo. (1995). Practice parameters for the use of

actigraphy in the clinical assessment of sleep disorders. American Sleep Disorders Association. *Sleep (Rochester)* [Online]. *18(4)*, pp. 285–287. Available:http://www.ncbi.nlm.nih.gov/pubmed/7618028

[2] A. Sadeh and C. Acebo. (2002, May). The role of actigraphy in sleep medicine. *Sleep Med. Rev.* [Online]. *6(2)*, pp. 113–124. Available:http://linkinghub.elsevier.com/retrieve/pii/S1087079201901820

[3] A. Sadeh. (2011, Aug.). The role and validity of actigraphy in sleep medicine: an update. *Sleep Med. Rev.* [Online]. *15(4)*, pp. 259–67. Available:http://www.ncbi.nlm.nih.gov/pubmed/21237680

[4] D. Leger, S. R. Pandi-perumal, and I. Healthcare, "Review of sleep disorders: Their impact on public health," *Public Health*, vol. 30, no. 7, pp. 92161–92161, 2007.

[5] A. Sadeh, P. J. Hauri, D. F. Kripke, and P. Lavie. (1995, May). The role of actigraphy in the evaluation of sleep disorders. *Sleep* [Online]. *18(4)*, pp. 288–302. Available:http://www.ncbi.nlm.nih.gov/pubmed/7618029

[6] R. J. Cole, D. F. Kripke, W. Gruen, and J. C. Gillin, "Automatic sleep/wake identification from wrist actigraphy," *Sleep*, pp. 461–469, Oct. 1992.

[7] A. Sadeh, K. M. Sharkey, and M. A. Carskadon, "Activity-based sleep-wake identification: An empirical test of methodological issues," *Sleep*, vol. 17, no. 3, pp. 201–207, Apr. 1994.

[8] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak. (2003, May). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* [Online]. *26(3)*, pp. 342–392. Available:http://www.ncbi.nlm.nih.gov/pubmed/12749557

[9] C. P. Pollak, W. W. Tryon, H. Nagaraja, and R. Dzwonczyk. (2001, Dec.). How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep* [Online]. *24(8)*, pp. 957–965. Available:http://www.ncbi.nlm.nih.gov/pubmed/11766166

[10] C. Acebo and M. K. LeBourgeois. (2006, Mar.). Actigraphy. *Respirat. Care Clin. North Amer.* [Online]. *12(1)*, pp. 23–30, viii. Available:http://www.ncbi.nlm.nih.gov/pubmed/16530645

[11] B. Sivertsen, S. Omvik, O. E. Havik, S. Pallesen, B. Bjorvatn, G. H. Nielsen, S. Straume, and I. H. Nordhus. (2006, Oct.). A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep* [Online]. *29(10)*, pp. 1353–1358. Available:http://www.ncbi.nlm.nih.gov/pubmed/17068990

[12] J. Paquet, A. Kawinska, and J. Carrier, "Wake detection capacity of actigraphy during sleep," *Sleep*, vol. 30, no. 10, pp. 1362–1369, 2007.

[13] J. Tilmanne, J. Urbain, M. V. Kothare, A. V. Wouwer, and S. V. Kothare, "Algorithms for sleep-wake identification using actigraphy: A comparative study and new results," *J. Sleep Res.*, vol. 18, no. 1, pp. 85–98, Mar. 2009.

[14] L. D. Souza, A. A. Benedito-silva, M. Laura, N. Pires, D. Poyares, S. Tufik, and H. M. Calil, "Further validation of actigraphy for sleep studies," *Sleep*, vol. 26, no. 1, pp. 81–85, 2003.

[15] W. W. Tryon. (2004, Feb.). Issues of validity in actigraphic sleep assessment. *Sleep* [Online]. *27(1)*, pp. 158–165. Available:http://www.ncbi.nlm.nih.gov/pubmed/14998254

[16] J. Gale, T. L. Signal, and P. H. Gander. (2005, Aug.). Statistical artifact in the validation of actigraphy. *Sleep* [Online]. *28(8)*, pp. 1017–1018. Available:http://www.ncbi.nlm.nih.gov/pubmed/16218087

[17] C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement. (2001, Sep.). Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med.* [Online]. *2(5)*, pp. 389–396. Available:http://www.ncbi.nlm.nih.gov/pubmed/14592388

[18] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak. (2003, May). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* [Online]. *26(3)*, pp. 342–392. Available:http://www.ncbi.nlm.nih.gov/pubmed/12749557

[19] J. Hedner, G. Pillar, S. D. Pittman, D. Zou, L. Grote, and D. P. White. (2004, Dec.). A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep* [Online]. *27(8)*, pp. 1560–1566. Available:http://www.ncbi.nlm.nih.gov/pubmed/15683148

[20] W. Karlen, C. Mattiussi, and D. Floreano. (2008, Jan.). Improving actigraph sleep/wake classification with cardio-respiratory signals. *Conf. Proc.: Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, [Online]. vol. 2008, pp. 5262–5265. Available:http://www.ncbi.nlm.nih.gov/pubmed/19163904

[21] P. Casale, O. Pujol, and P. Radeva. (2011). "Human activity recognition from accelerometer data using a wearable device," in *Proc. 5th Iberian Conf. Pattern Recognit. Image Anal.*,[Online]. pp. 289–296. Available:http://dl.acm.org/citation.cfm?id=2021341.2021382

[22] A. Domingues, T. Paiva, and J. M. Sanches, "An actigraphy heterogeneous mixture model for sleep assessment," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2012, pp. 2275–2278.

[23] A. Domingues, T. Paiva, and J. M. Sanches, "Statistical characterization of actigraphy data for sleep/wake assessment," in *Proc. 7th Int. Workshop Biosignal Interpret.*, 2012.

[24] PRTools. (2012). "The MATLAB toolbox for pattern recognition," [Online]. Available:http://www.prtools.org/

[25] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, pp. 179–186.

[26] C. Crespo, M. Aboy, J. R. Fernandez, and A. Mojon. (2012, Mar.). Automatic identification of activity rest periods based on actigraphy. *Med. Biol. Eng. Comput.* [Online]. *50(4)*, pp. 329–340. Available:http://www.springerlink.com/index/10.1007/s11517-012-0875-y

[27] R. Takalo, H. Hytti, and H. Ihalainen. (2005, Dec.). Tutorial on univariate autoregressive spectral analysis. *J. Clin. Monitor. Comput.* [Online]. *19(6)*, pp. 401–410. Available:http://www.ncbi.nlm.nih.gov/pubmed/16437291

[28] P. Pires, T. Paiva, and J. Sanches, "Sleep/wakefulness state from actigraphy," in *Pattern Recognition and Image Analysis*, (ser. IbPRIA'09). Berlin, Germany: Springer-Verlag, 2009, pp. 362–369.

[29] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[30] W. Stephen and J. R. Spiro, "Comparing different methodologies used in wrist actigraphy," *Sleep Rev.*, vol. Summer, pp. 1–6, 2001.

[31] R. J. Salin-Pascual, T. A. Roehrs, L. A. Merlotti, F. Zorick, and T. Roth. (1992). Long-term study of the sleep of insomnia patients with sleep state misperception and other insomnia patients. *Amer. J. Psychiatry* [Online]. *149(7)*, pp. 904–908. Available:http://ajp.psychiatryonline.org/cgi/content/abstract/149/7/904

Alexandre Domingues graduated in electronics and telecommunications engineering in 2006, after a period working as a software engineer, he joined a bioinformatics group, where he worked in optimization strategies for metabolic networks. In 2009, he received the Master of Science degree in electrical and computer engineering, and in 2010, he entered the Ph.D. School in Biomedical Engineering at Instituto Superior Técnico, Lisbon, Portugal, working on statistical algorithms for the diagnosis of sleep disorders.

Teresa Paiva graduated from the Medical Faculty of Lisbon, in 1989. She has been a Medical Doctor since 1969, specialized in Neurology and Sleep Medicine, and a Professor of the Medical Faculty of Lisbon. She developed a longstanding clinical work, while implementing national teaching programs, namely the first worldwide Master Degree in Sleep Sciences, and orienting national awareness campaigns for the general public.

J. Miguel Sanches (M'05–SM'11) received the E.E., M.Sc., and Ph.D. degrees from the Lisbon Institute of Technology [Instituto Superior Tcnico (IST)], Technical University of Lisbon, Lisbon, Portugal, in 1991, 1996, and 2003, respectively, and the Habilitation (agregao) degree in biomedical engineering from the University of Lisbon (UL), in 2013.

He is professor in the Bioengineering Department at the Instituto Superior Técnico (IST) from the University of Lisbon in the areas os Signal, Systems and Control theory and he is researcher at the Institute for Systems and Robotics, also in the IST. His research activity is in Biomedical Engineering (BME), mainly in biological and medical image processing and statistical signal processing of physiological and behavioral data.

Dr. Sanches has been a Senior Member of the IEEE Engineering in Medicine and Biology Society since 2011, and is a member of the Bio Imaging and Signal Processing Technical Committee (BISP-TC) of the IEEE Signal Processing Society. He is also the President of the Portuguese Association of Pattern Recognition (APRP), an affiliated of the International Association of Pattern Recognition (IAPR).