Automatic sleep stage identification: difficulties and possible solutions

Sukhorukova, N[1*] Stranieri, A[1] Ofoghi, B[1] Vamplew, P[1] Saleem, M[1] Ma, L[1] Ugon, A[2,3] Ugon, J[1] Muecke, N[1] Amiel, H[2] Philippe, C[2] Bani-Mustafa, A[1] Huda, S[1] Bertoli, M[1] Lévy, P[2] Ganascia, J-G[3]

1 Centre for Informatics and Applied Optimisation, University of Ballarat, Australia
2 Tenon Hospital, Paris, France
3 Laboratoire d'Informatique de Paris 6, France

*Corresponding author: n.sukhorukova@ballarat.edu.au

Abstract

The diagnosis of many sleep disorders is a labor intensive task that involves the specialised interpretation of numerous signals including brain wave, breath and heart rate captured in overnight polysomnogram sessions. The automation of diagnoses is challenging for data mining algorithms because the data sets are extremely large and noisy, the signals are complex and specialist's analyses vary. This work reports on the adaptation of approaches from four fields; neural networks, mathematical optimisation, financial forecasting and frequency domain analysis to the problem of automatically determining a patient's stage of sleep. Results, though preliminary, are promising and indicate that combined approaches may prove more fruitful than the reliance on a single approach.

Keywords: Sleep stage identification, data mining.

1 Introduction

Sleep Stage Identification (SSI) is the first step in the process of modern sleep disorder diagnostics. Currently, the identification of stages 1, 2, 3, REM and Awake is performed manually using rules drafted for medical practitioners based on the frequency and amplitude of waves recorded during polysomnogram sleep sessions (PSG). A polysomnogram sleep session (PSG) includes measures of eye movement (EOG), brain wave fluctuations (EEG), heart rhythm (ECG), muscle activity (EMG), respiratory effort and other biophysiological characteristics while a patient is asleep.

SSI is a time consuming, manual process that requires a great deal of skill and expertise in scanning PSG graphs and applying SSI rules. Recent advances in computing performance has made computer-based automatic scoring of sleep stages very attractive.

However, sleep practitioners report that existing automated techniques are not accurate enough to be routinely used (Robert, Guilpin et al. 1998).

PSG scoring experts apply rules based on the visual appearance of frequencies and amplitudes of waves on screen rather than using quantitative data describing frequencies and amplitudes. A survey of existing automatic tools for SSI by Bashashati, Fatourechi et al. (2007) and Rajeev and Gotman (2002) reveals that the majority of approaches apply signal processing (SP) methods (Bashashati, Fatourechi et al. 2007), Artifical Neural Network (ANN) methods (Robert, Guilpin et al. 1998) or Wavelet Transformations (Virkkalaa, Hasan et al. 2007). Approaches based on Financial Forecasting, Mathematical Optimisation and Hidden Markov Models have been deployed with other time series data and could conceivably lead to accurate classifications of SSI.

Much of of the challenge in automated SSI is due to the translation of open textured standards to mathematical models (Rajeev and Gotman 2002) and the dimension of the problem. Raw data for one patient for 10 hours results in a single file more than 300 MB large with over 3,600,000 observations. Further, over 65% of records are sleep stage 2 and less than 5% for sleep Stage 1 and 3. This adds to the complexity of the challenge. Further, PSG data contains a great deal of noise. With practice, experts are able to ignore noise to an extent that is challenging for automated scoring tools. In addition, SSI can be performed differently by two sleep practitioners with an 80% level of agreement.

Rules for SSI originally were standardized by Rechtschaffen and Kales (1968). Since then, the rules have been updated numerous times. The most recent version is reported by Iber, Ancoli-Israel et al. (2007). A comprehensive explanation on why this update was necessary can be found in Schulz (2008).

In SSI doctors rely on the visual presentation of waves. Recorded waves are signals and therefore it is natural to analyze them with existing SP techniques, especially given the theoretical advances in this field in recent decades. The drawback of this approach is that in many cases SP completely ignores manual scoring characteristics, which are not described in general scoring rules, but are often taken into account by medical doctors.

Since doctors are not experts in SP they learn the shapes of the waves through their visual characteristics rather than wave characteristics used in SP.

The main problem with ANN approaches is that the dimension of the problem challenges most learning algorithms. Several simplifications have been used including using fewer variables in order to overcome the dimensionality problem. However, simplified models are not accurate enough to meet the needs of sleep disorder specialists. Problems associated with the use of ANN and SP approaches suggest the need to explore combinations of ANN and SP with other approaches.

In this study, approaches based on financial forecasting, mathematical optimisation, frequency domain analysis and neural networks have been adapted for SSI. The approaches have been applied to data supplied and classified from overnight sleep records from 100 patients from the Tenon Hospital sleep research group in Paris. Each approach is described and results presented in the sections below, before providing a cross-approach analysis and concluding remarks.

2 Neural network approach

ANN is a network composed of artificial nodes that process input activation for transmission to connected nodes. Input vectors to the ANN are treated as a temporal sequence whose analysis requires consideration of a set of prior input vectors. (Waibel, Sawai, et al. 1989) used Time-Delay Neural Networks (TDNNs) for speech recognition. The delay-based methodology of TDNNs, which reduces the high dimensionality of the input data to the network, is very important in the SSI, due to the length of input sequences.

A TDNN is a type of dynamic ANN where the output of the network at time t_i is not only dependent on the input p_i at this time, but also on a range of previous inputs p_{i-1}, p_{i-2},…, p_{i-n} corresponding to t_{i-1}, t_{i-2},…, t_{i-n} where n is the delay length that is to be considered by the network. The main benefit obtained when using TDNN's is that there is no need for the network to contain many input nodes to deal with the whole set of delayed input vectors. The sequential data (original signal information) is presented to the network over time and the network is trained to deal with desired steps of delay.

A focused TDNN (delay only at the input layer) was configured with 1 input layer, 3 hidden layers, and 1 output layer using MATLAB's ANN package. Six input layer nodes represent PSG variables, EEG Curve 1, 2 & 3, EOG Curve 1 & 2 and EMG. Each hidden layer included 6 nodes, and the output layer comprised 6 nodes corresponding to sleep stage classes, Awake, Stage 1, 2, 3 and REM. The input signals were first normalized to the range of [-1,+1] and then converted to time sequences. The delay length of the network was set to 1 second (equal to 100 input vectors).

We have implemented a focused TDNN (delay only at the input layer) with an input layer, 3 hidden layers, and an output layer using MATLAB's ANN package. The training procedure is carried out with 500 epochs. The classification accuracy obtained as a baseline for comparison with other approaches is 76.15% of correctly classified records. A total number of records=33,407 were used to train the network.

3 Financial forecasting

A forecasting approach applied to financial market predictions by Bertoli and Stranieri (2004) was adapted in this study to predict sleep stages using data on six PSG variables. Like the TDNN, the approach is based on the intuition that a classification at a point in a series depends on classifications on previous sequences. The approach combines subsequence conditional probabilities to perform a classification in a way that is scalable to large data sets. The adapted forecasting algorithm was applied to data collected from the same patient in an overnight sleep session which included over 3.5 million records on six real-valued PSG signals. The size of the data makes this dataset challenging for any algorithm.

The real valued raw data was first converted to five point interval data labelled BI (big increase), SI (small increase), N (no change), SD (small decrease) and BD (big decrease). Threshold values for the intervals derived from percentiles. The algorithm was applied to discover all unique sequences shorter than 7. The intuition being that a sequence such as BI, BI, BI, SI, N, SD and SI could be discovered on each variable that could discriminate one sleep stage from another.

The confusion matrix (CM) represented in Table 1 depicts classifications made by the Tenon Hospital sleep experts against classifications predicted using the FF approach. This CM illustrates the forecasting approach has some promise given the large and noisy data set, however the prevalence of Stage 2 classifications in the training set led to relatively high mis-classifications. It was also found that the thresholds used in the transformation of real values to interval data for the classification labels (BI, SI, N, SD, BD) had significant impact so further work is required to identify optimal mappings. Unexpectedly, the experiments also found that the length of the pattern used to make the prediction did not need to be particularly long and that a pattern length of 6 or more did not result in any improvement to the predictions but did have a detrimental effect on the processing time.

		Predicted				
Actual		A	S 1	S 2	S 3	REM
	A	**70,000 45%**	35,000 **23%**	44,000 **28%**	1,000 **1%**	5,000 **3%**
	S1	9 **4%**	**195 81%**	28 **12%**	3 **1%**	5 **2%**
	S2	317,000 **2%**	1,700,000 **9%**	**13,000,000 60%**	1,300,000 **7%**	2,400,000 **13%**
	S3	0	1 **2%**	1 **2%**	**45 96%**	0
	REM	23 **4%**	22 **4%**	103 **20%**	2 0%	**374 71%**

Table 1. Confusion matrix for financial forecasting approach

4 Non-smooth optimisation

The adaptation of non-smooth optimisation to SSI is based on minimising the deviation between the actual PSG curve and modelled wave patterns. This approach extracts wave characteristics (similar to SP), but these characteristics are more flexible than "standard" SP characteristics and are targeting wave shape descriptions. These characteristics can be used for explicit description of wave shape patterns (similar to ANN), but the dimension of the problem is considerably lower.

The EEG curves are taken to be the sum of two sine curves. The first curve (lower frequency) represents a general trend which is passing through the whole observation sub-period. The second one (higher frequency) is the actual behaviour of the curve along the general trend. The amplitude of each curve is modelled as a piece-wise linear function. This approach allows more precise curve patterns then in the case of classical sine curves where the amplitude is scalar. Additionally, it allows for abrupt changes in the wave patterns with the piecewise linear function (non-smoothness). In our experiments we use non-smooth optimisation techniques from the GANSO library (Ganso 2006).

All the experiments have been performed on an EEG curve with the horizontal axis corresponding to time. First, the higher frequency sine curve was obtained (Figure 1). This curve is the first approximation of EEG data. The accuracy of approximation is improved by taking into account the general trend of the curve. In Figure 2, the general trend is plotted against the data which represents the difference between the original data and the first trend. Finally, Figure 3 represents the final pattern which follows the original EEG data quite well. In these experiments the subinterval corresponds to 5 seconds of sleep, therefore for each epoch we construct 6 patterns. The dimension of this problem is 12. The dimension of an ANN problem would be 500. This suggests the use of the output of the optimisation problem as an input for ANN could lead to good results.

Figure 1 Actual behaviour and main frequency

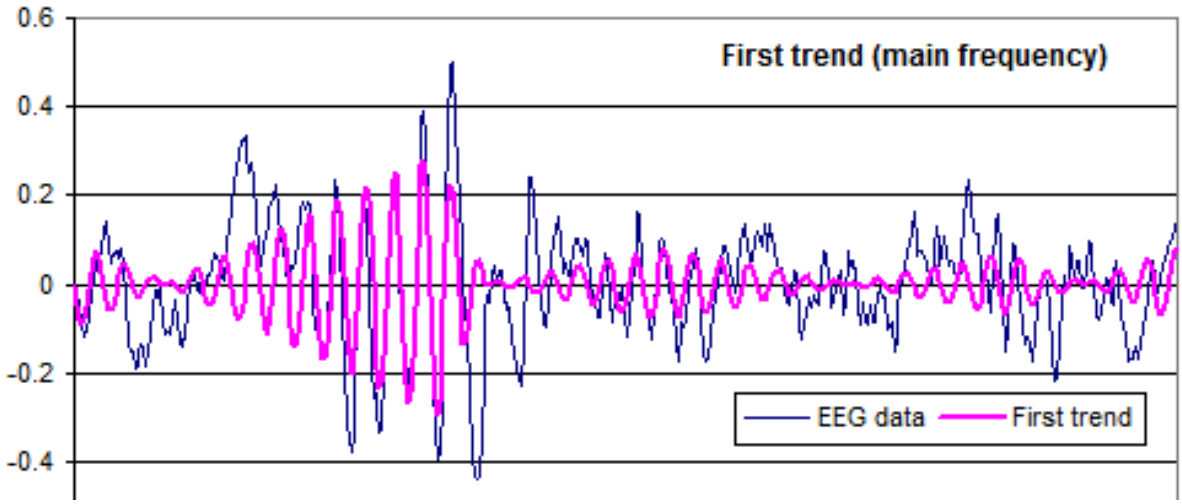


Figure2 General trend

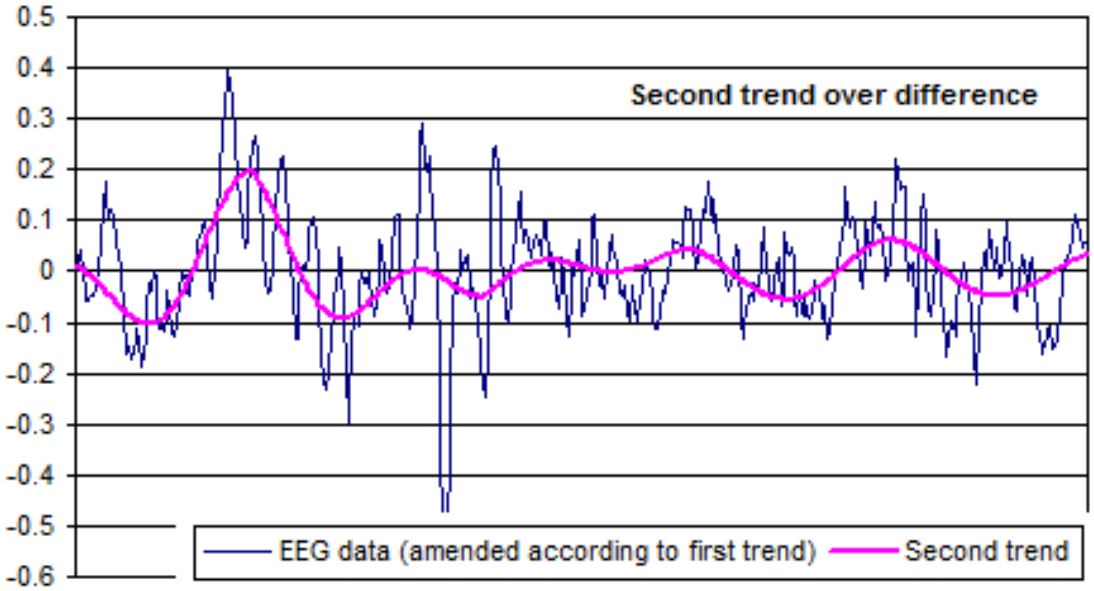
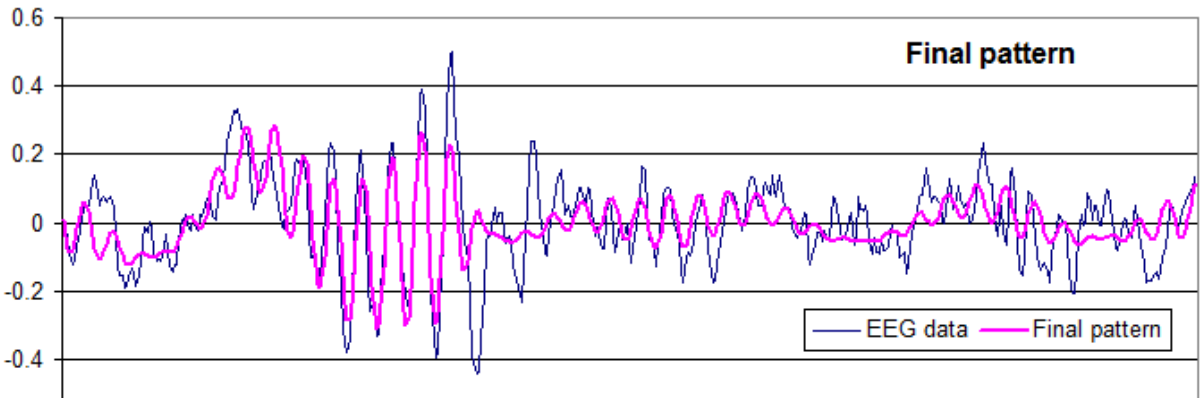


Figure 3 Final pattern (sum of the first and the second trend)



5 Frequency domain analysis

The frequency domain analysis (FDA) approach is based on the basic concept of windowing the signal in the time domain and then taking it into the frequency domain, also called Short-Time Fourier Transform (STFT). The resultant signal is mapped into a two dimensional function of time and frequency. As signals of EEG, EOG, and EMG are not stationary, hence, such techniques give limited precision over this conversion. EEG power spectra has been used in the literature for detecting behavioural microsleeps and estimating the alertness (Jung, Makeig et al. 1997), (Peiris, Jones et al. 2006).

EEG, EMG, and EOG signals are taken into the frequency domain with a window of 30 seconds. The frequency components can be divided into four bands: δ (< 4 Hz), θ (4 – 7 Hz), α (8 – 13 Hz) and β (> 13 Hz) (Carney, Berry et al. 2005). Once these frequency components are separated, the power spectral density is plotted for the window. The power spectral density can be calculated as $\Phi(\omega) = \frac{(F(\omega)F^*(\omega))}{2\pi}$, where $F^*(\omega)$ is the complex conjugate of the frequency matrix.

REMs are generally characterized by a number of features (Pressman 2007), i.e., a low voltage, fast frequency EEG. This is marked by an increase in Φ(β)

and relative decrease in the spectral densities of low frequency components. These characteristics will be exploited to detect REM.

According to (Pressman 2007) it is not essential that all the characteristics described before for REM detection are present simultaneously. The presence of only two features out of three can be accepted as a valid REM stage. Figure 4 illustrates the Short Time Fourier Transform (STFT) plot of EMG and Figure 5 represents the STFT for EOG signals. Figure 6 presents the STFT plot of EEG signal and Figure 7 depicts the manual scoring of sleep stages where Stage 5 is REM. One case is described below to explain the REM detection.

Figure 4 EMG in frequency domain against time and sleep stages

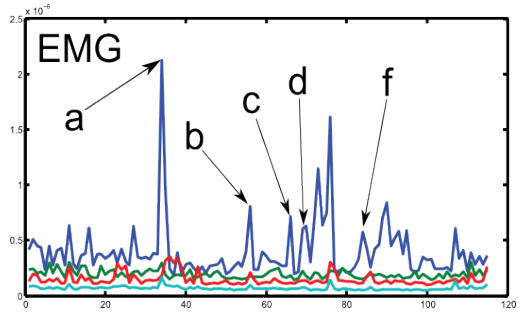


Figure 5 EOG in frequency domain against time and sleep stages

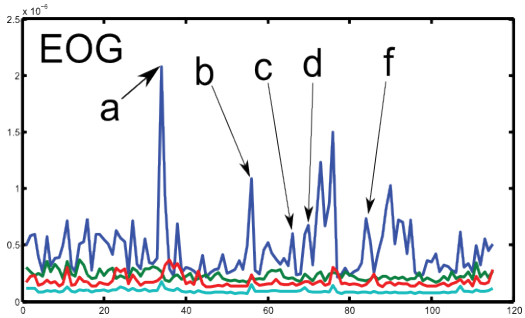


Figure 6 EEG in frequency domain against time and sleep stages

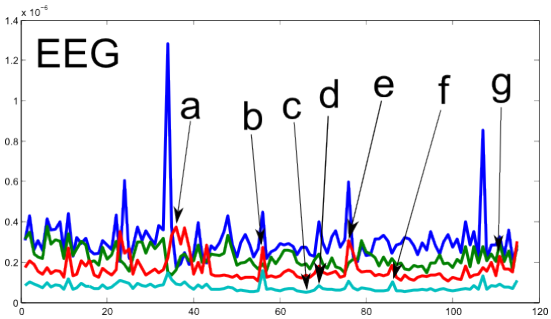
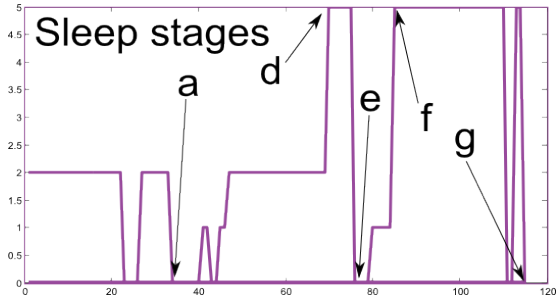


Figure 7 Sleep stages



REM starts with rise in EMG Φ(δ), as Φ(δ) corresponds to very low frequency components. 'a' and 'b' cannot be considered the start of REM stage because EEG Φ(α) also increases sharply with EMG Φ(δ). 'c' is also not the start of REM because there should be a small rise in EEG Φ(β). 'd' is the start of REM as there is an increase in EMG Φ(δ) and a small increase in EEG Φ(β). There is also a rise in all the frequencies of EOG. At 'e', the REM finishes at a sharp increase of EEG Φ(α), marking an awake stage. 'f' marks the start of REM as there is an increase in EMG Φ(δ) accompanied by small increase in EEG Φ(β). 'g' marks the finish of this REM stage as there is an increase in EEG Φ(α). The last REM for a short time duration has not been detected. All these rules, which are inferred from the characteristics of REM stages described by doctors, can be elegantly implemented using a state machine. However, thresholds of different frequency components that would trigger the state change varies from case to case.

6 Analysis and discussion

In this project it has been shown that the automated SSI procedure is a complex process which cannot readily be achieved without employing a number of diverse methods. This diversity allows one to overcome the problem of "translating" manual scoring rules into automated algorithms.

In our study we used TDNN which can handle higher dimension data better than other types of ANN. The accuracy of 76% is quite good since 2 manual scorers may also produce different classification results (the level of agreement is around 80%). One possible way to enhance the obtained accuracy is to apply TDNN after dimension reduction using NOM. Another possible way is to detect different sleep stages with different approaches, e.g., to identify REM using FDA and Stage 3 using FF. It was also found that the correct detection of Stage 2 is a challenging task for several methods. This is mainly due to the presence of short lasting events (K-complexes), which are difficult to detect by our methods. One future research direction involves the identification of these events using NOM.

7 Conclusions and further research directions

This project is an attempt to build an automated SSI procedure as a meta-classifier, which involves different methods to solve the problem. We identify strengths of particular methods and distribute the "roles".

In the future we are planning to incorporate these methods in a single procedure. Basing on the research findings of this paper the procedure can be organised as follows:

1. NOM is used as a specific preprocessing tool to convert raw data into a lower dimensional space.
2. Apply ANN methods (or other classification method) to a lower dimensional space of extracted features, obtained on the previous stage.
3. Refine our classification results using FDA and FF for some specific sleep stages (REM and Stage 3 respectively).

7.1 Further research directions

Our future research directions include the meta-classifier building and testing on available data. Also, we are planning to conduct a study on how tis meta-learner would learn from two medical experts scored the same data. As it was mentioned before, the level of agreement between two experts can be as low as 80%.

Another promising method for SSI is the Hidden Markov Model (HMM). HMM has a powerful ability to model signals statistically and represent arbitrarily complex probability density functions of the underlying systems. Previous attempts on sleep stage identification (Flexer A., et.al 2002) using HMM did not have much success. One of the important reasons is that these approaches consider modelling sleep stage sequences using a single HMM with a small number of HMM states. The elaboration of this method is another future research direction.

8 References

Bashashati, A., M. Fatourechi, et al. (2007). 'A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals.' J. Neural Eng.(4): R32-R57.

Bertoli and Stranieri (2004). 'Forecasting on complex datasets with association rules'. 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems. Wellington. Springer. 1170-80.

Carney, P. R., R. B. Berry, et al. (2005). Clinical Sleep Disorders, Lippincott Williams & Wilkins.

Flexer, A., G. Gruber, et al. (2005). "A reliable probabilistic sleep stager based on a single EEG signal." Artificial intelligence in Medicine 33(3): 199-207.

Ganso (2006) http://www.ganso.com.au/

Iber, C., C. Ancoli-Israel, et al. (2007). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Technology and Technical Specifications, Westchester: American Academy of Sleep Medicine.

Jung, T. P., S. Makeig, et al. (1997). "Estimating alertness from the EEG power spectrum." IEEE Transactions on Biomedical Engineering 44(1): 60--69.

Peiris, M. T. R., R. D. Jones, et al. (2006). Detecting Behavioral Microsleeps from EEG Power Spectra. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Pressman, M. (2007). Stages and architecture of normal sleep, UpToDate

Rajeev and Gotman (2002). 'Digital tools in polysomnography.' Journal of clinical neurophysiology 12(2): 136-143.

Rechtschaffenand, A. and A. Kales (1968). A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects. U. G. P. O. US Public Health Service. Washington, DC.

Robert, C., C. Guilpin, et al. (1998). 'Review of neural network applications in sleep research.' Journal of Neuroscience Methods (79): 187-193.

Schulz, H. (2008). 'Rethinking Sleep Analysis.' Journal of Clinical Sleep Medicine 4(4): 99-103.

Virkkalaa, J., J. Hasan, et al. (2007). 'Automatic sleep stage classification using two-channel electro-oculography.' Journal of Neuroscience Methods 166(1): 109-115.

Waibel, A., H. Sawai, et al. (1989) 'Modularity and Scaling in Large Phonemic Neural Networks', IEEE Transactions on Acoustics, Speech and Signal Processing, 37(12): 1888-1898.

Zobel, J. and Dart, P. (2000): Partitioning number sequences into optimal subsequences. *Journal of Research and Practice in Information Technology* **32**(2):121-129.