

AUTOMATED SLEEP STAGING TECHNIQUE BASED ON THE EMPIRICAL MODE DECOMPOSITION ALGORITHM: A PRELIMINARY STUDY

JING DONG* and DONGDONG LIU†

*Department of Biomedical Engineering
Peking University, Beijing, 100871, China
*corrinedong@gmail.com
†pqu2@163.com*

CHENG ZHANG‡, JING MA§ and GUANGFA WANG¶

*Department of Pulmonary Medicine, Peking University First Hospital
Beijing, 100871, China
‡zhangcheng033@163.com
§majjmail@163.com
¶wangguangfa@hotmail.com*

DAN GUO

*Sleep Disorders Center, Mei Tan General Hospital
Beijing, 100028, China
public_scenery@163.com*

YANHUI LIU

*DynaDx Corporation
Mountain View, CA 94041, USA
yanhui.liu@gmail.com*

HUA ZHONG

*Biomed-X Research Center, Academy of Advanced Interdisciplinary
Peking University,
Beijing, 100871, China
snoopyzhong@126.com*

JUE ZHANG||

*Department of Biomedical Engineering, Peking University
Beijing, 100871, China
zhangjue@pku.edu.cn*

CHUNG-KANG PENG^{||}

*Margret & H.A. Rey Institute of Nonlinear Dynamics in Physiology and Medicine
Division of Interdisciplinary Medicine and Biotechnology
Beth Israel Deaconess Medical Center
Harvard Medical School, Boston, MA 02215, USA
cpeng@bidmc.harvard.edu*

JING FANG

*Biomed-X Research Center, Academy of Advanced
Interdisciplinary, Peking University
Beijing, 100871, China
jfang@pku.edu.cn*

An automatic sleep staging method is proposed to score wakefulness and three nonrapid eye movement (NREM) stages S1, S2 and S3, based on the Empirical Mode Decomposition (EMD) algorithm. Patients with sleep disorders were tested using this method. Good agreements between manual and automatic staging results were achieved in terms of their Cohen's Kappa value.

Keywords: Automatic sleep staging; Empirical Mode Decomposition; sleep disordered breathing subjects.

1. Introduction

Human sleep was described as a succession of five recurring stages by Rechtschaffen and Kale (R&K) in 1968: the rapid eye movement (REM) stage and four nonrapid eye movement (NREM) stages, S1, S2, S3, S4 [Rechtschaffen and Kales (1968)]. It has not been changed until 2007 the American Academy of Sleep Medicine updated part of it. S3 and S4 were grouped into one stage [Norman *et al.* (2000)]. R&K rules have been the world-wide standard to describe sleep structure for clinic. All the stage scoring work has to be done manually. It is mainly based on three signals: electroencephalograms (EEG), electrooculograms (EOG), and electromyogram (EMG). The R&K rules defined some characteristic waves according to the amplitude, frequency and shape of the EEG signal. For example, alpha waves with a frequency of 8 Hz or more but less than 13 Hz, delta waves with a frequency of less than 4 Hz and with amplitudes greater than $75 \mu\text{V}$, the K-complex waveform consisting of a brief negative high-voltage peak, followed by a slower positive complex and a final negative peak, and sleep spindles, a burst of 12–16 Hz waves that occur for 0.5–1.5 s. Sleep stages are scored mainly based on these waves. For example, if the delta waves occupy more than 20% in 30 s, this epoch would be scored as Stage 3 sleep. It is known that such manual work of sleep staging is very complex and onerous. Although performed by trained personnel, the reliability and coherence of manual sleep scoring are usually unsatisfactory because of subjective judgment and human error [Norman *et al.* (2000)]. A number of automatic sleep staging methods have been studied in the last decades, including period analysis, spectral analysis, multiple discriminant analysis, pattern recognition technique, wave detection,

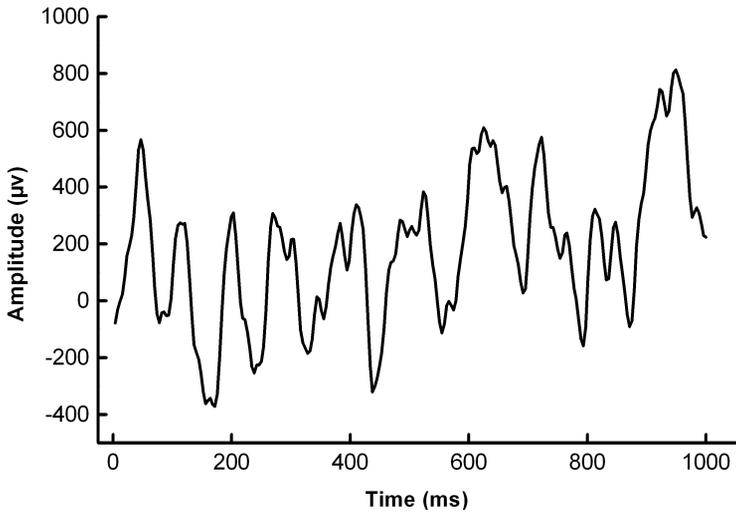
fuzzy clustering, and neural network [Loomis *et al.* (1937); Johnson *et al.* (1969); Larsen and Walter (1970); Martin *et al.* (1972); Kuwahara *et al.* (1988); Stanus *et al.* (1987); Gath and Bar-On (1980); Friedman and Jones (1984); Schaltenbrand *et al.* (1996); Ray *et al.* (1986); Berthomier *et al.* (2007)]. However, most of the proposed methods have only a limited range of viability, and are seldom applied to patients with sleep disorders. This paper proposes an automatic sleep staging method based on empirical mode decomposition (EMD) [Huang *et al.* (1998)], to stage the wakefulness and three NREM sleep states (S1, S2 and S3).

2. Methods

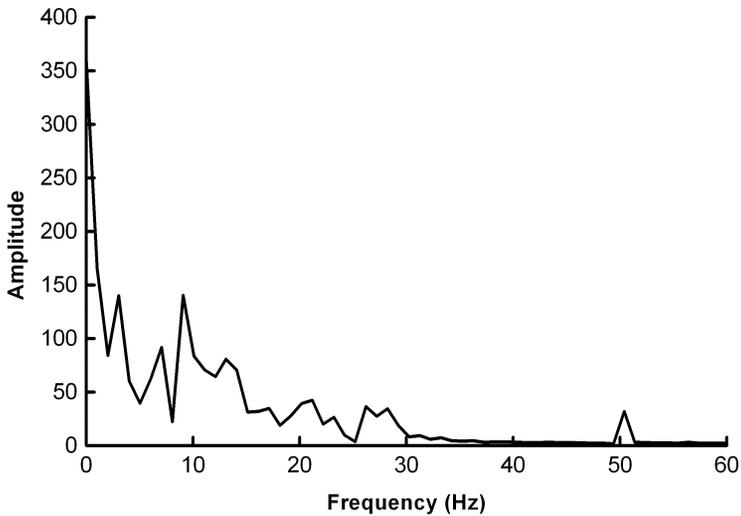
The sleep EEG of 15 patients (mean age 42 ± 14 years, range 21–66 years, 4 females and 11 males) with different severity of obstructive sleep apnea–hypopnea syndrome (OSAHS) were analyzed retrospectively. The whole night polysomnographic (Siesta 2, Compumedics Ltd, Australia) data were obtained from the Sleep Laboratory of Department of Pulmonary Medicine of Peking University First Hospital. The data of each subject, including EEG, ECG, EOG, SpO₂, airflow, thorax efforts, abdominal efforts, snoring sound, leg movement and body position, were annotated with respect to sleep stages and apnea. In this study, only single channel EEG signals were analyzed from the C3 derivation with a sampling rate of 128 Hz or 256 Hz. The sleep stages were visually scored for each 30 s epoch by experienced clinically trained personnel, following the 2007 AASM (American Academy of Sleep Medicine) sleep scoring manual.

Traditional Fourier spectral analysis is not applied to sleep staging in our analysis. Fourier transform requires stationary signals and decompose them into cosine or sine waves spanning the whole data length with constant amplitude. However, EEGs are typically nonstationary signals with time varying frequency and amplitude. As a result, applying Fourier transform on nonstationary signals will induce spurious harmonic components that cause erroneous energy spreading. For example, according to R&K rules, alpha waves lies within the range of 8–13 Hz. Clinically, the alpha wave is identified by manually counting the number of peaks within 1 second (Fig. 1(a)). However, it is hard to pick out the center frequency of this alpha wave from its Fourier power spectrum (Fig. 1(b)), which makes automatic staging a difficult task.

The EMD algorithm has been developed to analyze nonstationary signals. It directly extract the energy associated with various intrinsic time scales by decomposing the signal into a group of intrinsic mode functions (IMFs) that often contain more relevant physical meanings. By applying the Hilbert transform on these IMFs, one can extract instantaneous frequencies and amplitudes that are variable in time. For example, the alpha wave in Fig. 1(a) was processed using the EMD (Fig. 2(a)). By choosing the third mode and applying the AR model [Percival and Walden (1993)] of order 16 on its power spectrum, it is seen that the peak power was around 10 Hz (Fig. 2(b)).



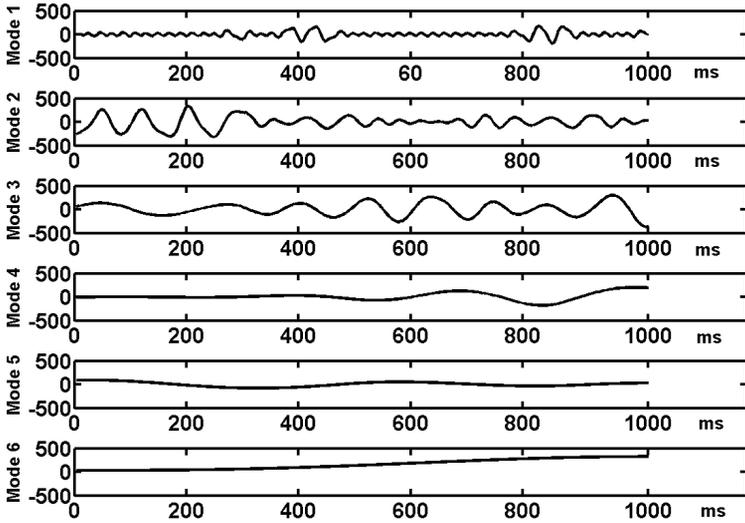
(a)



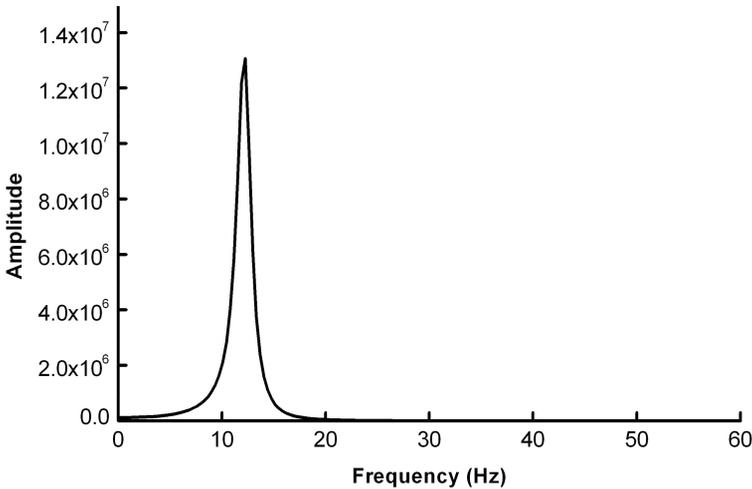
(b)

Fig. 1. (a) Typical EEG signal with alpha wave oscillation. (b) Fourier spectrogram of the EEG shown in (a). It is difficult to recognize the center frequency of the alpha wave.

Based on the EMD algorithm, the automatic sleep staging process started with pre-processing the whole night polysomnographic EEG data through 20 iterations with 7 modes. The frequencies of 2–5 modes were calculated and the mode with the main frequency 0–20 Hz was selected (Fig. 3). Secondly, the mean instantaneous frequency (per second) was calculated by performing Hilbert transform on selected mode. Thirty mean instantaneous frequencies were acquired for each epoch. During



(a)



(b)

Fig. 2. (a) IMFs of the EEG signal shown in Fig. 1(a). The alpha wave oscillations are decomposed into mode 2 and mode 3 (see the discussion section of the text for details). (b) Spectrogram of the 3rd mode of the alpha wave using the EMD analysis. It is clear to see that the center frequency was around 10 Hz.

measurements, EEG signals would be interfered by leg movement of the patients, resulting in data fluctuation or overflow. Therefore, to eliminate the influence of bad data, the average value of the 20 frequencies around the middle was calculated and used as the characteristic frequency of this epoch. Finally, clustering of the data was done by applying the K-means algorithm [Hartigan (1975)], which grouped

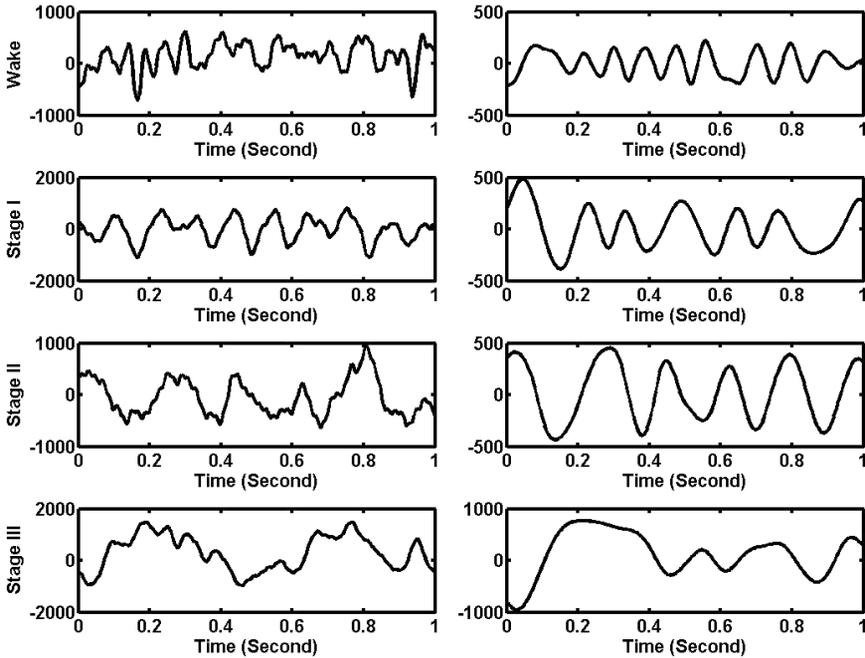


Fig. 3. Typical EEG signals of different stages (left panels) and the selected modes of different stages (right panels). The graph shows the frequency in a descending order from Wake to Stage 3 in both the left and right figures. The cycles per second of selected modes are visually apparent.

the characteristic frequencies into four clusters with their center frequencies in a descending order: wakefulness, S1, S2 and S3.

3. Results

Prior to applying the clustering algorithm, characteristic frequencies were linearly normalized within the range of 0–3. Meanwhile, each manually staged state was assigned with a value: wakefulness is 0, S1 is 1, S2 is 2 and S3 is 3. Plotting together the linearly normalized frequencies and manual staging results for a patient, it is seen that similar trends were yielded (Fig. 4).

After clustering the frequencies into four states, the automatic and manual staging results were compared (Fig. 5). The two graphs bear similar pattern with the exception during the frequent state-transition periods.

Table 1 gives the detailed comparison between the two methods for all patients: the value in each cell specifies the number of epochs that were assigned to the state in column by manual method and the state in row by automatic.

A more quantitative comparison was done with epoch-by-epoch agreement calculation, which was defined as the percentage of epochs that were assigned to the same state. The agreements of 15 patients ranged from 52% to 70% (mean

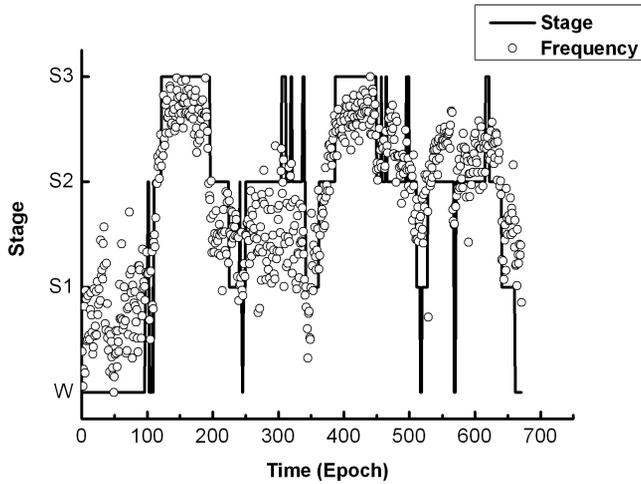


Fig. 4. Normalized frequencies (circle) and manual staging result (line). The trends of normalized frequencies and manual staging are similar.

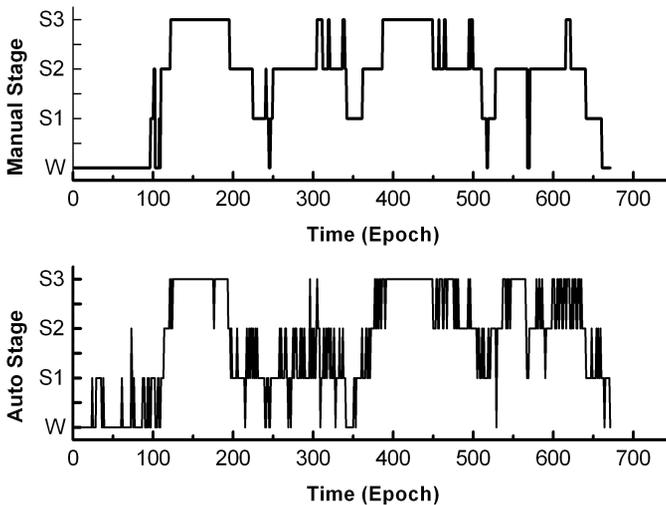


Fig. 5. Manual and automatic staging results of a patient. The two graphs bear similar pattern except for the local discontinuities in automatic staging.

agreement $60\% \pm 5\%$). In this study, Cohen's kappa [Landis and Koch (1977)] was used to assess agreement for comparisons between auto stage and manual stage. Kappa value is a statistical measure of inter-rater agreement for qualitative items. As a rule of thumb, values of Kappa from 0 to 0.20 are considered bad, 0.21 to 0.40 poor, 0.41 to 0.60 reasonable, 0.61 to 0.80 good, and more than 0.80 outstanding. The Kappa value of the above measurements was 0.44, indicating that the agreement between manual and automatic scoring was reasonable.

Table 1. Four-state epoch classification: auto staging versus manual staging.

| | | Auto staging | | | |
|----------------|------|--------------|------|------|------|
| | | Wake | S1 | S2 | S3 |
| Manual staging | Wake | 740 | 676 | 116 | 8 |
| | S1 | 236 | 553 | 167 | 3 |
| | S2 | 235 | 1109 | 2458 | 432 |
| | S3 | 42 | 99 | 581 | 1777 |

Note: It is easy to recognize the disagreements between auto staging and manual staging mainly come from adjacent stages. And there are few mistakes between wake and slow wave sleep (S3).

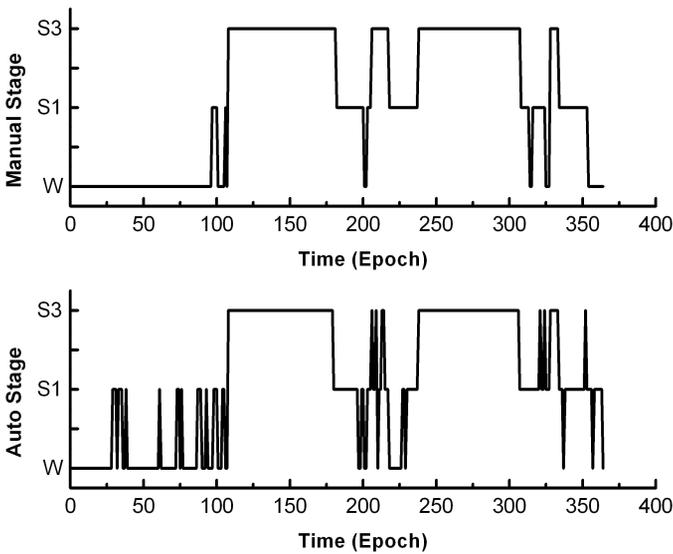


Fig. 6. The staging results of the same patient without S2. It is seen that similar pattern was yielded.

S2 is the transitional stage between active wave sleep and slow wave sleep. Therefore, to get rid of instabilities introduced by S2 data, only wakefulness, S1 and S3 were automatically staged for the same patient (Fig. 6). Higher than 70% mean agreement is achieved and the Kappa value yielded 0.61, which was good.

4. Discussions and Conclusions

Usually, the manual epoch by epoch agreement between experienced scorers is about 73%, with the percentage higher in healthy people (mean 76%, range 65–85%) than in patients with sleep disordered breathing (mean 71%, range 65–78%) [Norman *et al.* (2000)]. The proposed automatic sleep staging method demonstrated that comparable agreements with manual scoring based on the R&K rules

[Iber *et al.* (2007)] can be achieved. Although the macrostructure of the sleep process is more important, it is noteworthy that the major disagreements come from local discontinuities in automatic staging. The discontinuities might due to the excessive sensitivity of the automatic process and also might indicate that the frequent switches associated with some physiologic process. To some extent, R&K rules apply arbitrary EEG criteria on frequency range, amplitude, and brain wave density. As a result, sleep staging does not consistently match other physiologic properties [Brandenberger *et al.* (2005)]. Future sleep analysis might need to integrate different levels of data extraction, based on a set of algorithms to identify the changing features in biologic functions during sleep [Schulz (2008)].

This study represents a preliminary work applying the EMD to automatic sleep staging problem. It is not only because the EMD would be more applicable to EEG signals, but also because the sleep staging rule, i.e. the R&K, itself was born with disadvantages [Penzel and Condrat (2000)]. We started out to automate the traditional staging rule, but something more interesting was found. S2 is the most difficult to automatically quantify in our result. In fact some researchers have reported the duality of S2 with laboratory evidences [Brandenberger *et al.* (2005)]. Overall, our study implies that automatic staging of sleep should not only follow the traditional rules, instead, it shed some light on how to improve those rules. In addition, the algorithm presented in this study might have overlooked some important information by selecting only one IMF to calculate the characteristic frequency. For example, it is clear that in Fig. 2(a) the alpha waves are decomposed into mode 2 in the early segment (0–300 ms), and mode 3 in the latter segment (400–1000 ms). Although the oscillations are mainly in mode 3, it is not an accurate method to calculate the characteristic frequency only using mode 3. Therefore, further studies along this promising direction should integrate other IMFs to get more accurate results.

This algorithm is the first step to study EEG characteristics of different sleep stages. As a result, there are many aspects of the algorithm that need to be improved in future work. Meanwhile, this result comes from sleep apnea patients. As the sleep architecture is different between normal controls and sleep apnea patients and many arousals that are related to sleep apnea episodes could influence EEG frequency, it should be cautious to translate the present result to healthy subjects.

In summary, the EMD algorithm was applied for automated sleep staging in sleep apnea patients. Good agreements between manual and automatic staging was found during NREM sleep.

Acknowledgments

We would like to thank Prof. Norden E. Huang and Prof. Benjamin Rusak for their helpful suggestions on data analysis. CKP gratefully acknowledges the support from the NIH/NIBIB and NIGMS (U01-EB008577), the DARPA (HR0011-05-1-0057), and the G. Harold and Leila Y. Mathers Charitable Foundation.

References

- Berthomier, C. *et al.* (2007). Automatic analysis of single-channel sleep EEG: Validation in healthy individuals. *Sleep*, **30**: 1587–1595.
- Brandenberger, G. *et al.* (2005). Sleep stage 2: An electroencephalographic, autonomic, and hormonal duality. *Sleep*, **28**: 1535–1540.
- Friedman, L. and Jones, B. E. (1984). Study of sleep–wakefulness states by computer graphics and cluster analysis before and after lesions of the pontine tegmentum in the cat. *Electroencephalogr. Clin. Neurophysiol.*, **57**: 43–56.
- Gath, I. and Bar-On, E. (1980). Computerized methods for scoring of polygraphic sleep recordings. *Comput. Programs Biomed.*, **11**: 217–223.
- Hartigan, J. A. (1975). *Clustering Algorithm*, John Wiley & Sons Inc., New Jersey.
- Huang, N. E. *et al.* (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London Ser., A* **454**: 903–995.
- Iber, C. *et al.* (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, American Academy of Sleep Medicine, Westchester.
- Johnson, L. *et al.* (1969). Spectral analysis of the EEG of dominant and nondominant alpha subjects during waking and sleeping. *Electroencephalogr. Clin. Neurophysiol.*, **26**: 361–370.
- Kuwahara, H. *et al.* (1988). Automatic real-time analysis of human sleep stages by an interval histogram method. *Electroencephalogr. Clin. Neurophysiol.*, **70**: 220–229.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**: 159–174.
- Larsen, L. E. and Walter, D. O. (1970). One automatic methods of sleep staging by EEG spectra. *Electroencephalogr. Clin. Neurophysiol.*, **28**: 459–467.
- Loomis, A. *et al.* (1937). Cerebral states during sleep, as studied by human brain potentials. *J. Exp. Psychol.*, **21**: 127–144.
- Martin, W. B. *et al.* (1972). Pattern recognition of EEG-EOG as a technique for all-night sleep stage scoring. *Electroencephalogr. Clin. Neurophysiol.*, **32**: 417–427.
- Norman, R.G. *et al.* (2000). Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep*, **23**: 901–908.
- Penzel, T. and Condrat, R. (2000). Computer based sleep recording and analysis. *Sleep Med. Rev.*, **4**: 131–148.
- Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications*, Cambridge University Press, New York.
- Ray, S. R. *et al.* (1986). Computer sleep stage scoring – an expert system approach. *Int. J. Biomed. Comput.*, **19**: 43–61.
- Rechtschaffen, A. and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects*, National Institutes of Health (U.S.), Los Angeles.
- Schaltenbrand, N. *et al.* (1996). Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients. *Sleep*, **19**: 26–35.
- Schulz, H. (2008). Rethinking sleep analysis: Comment on the AASM manual for the scoring of sleep and associated events. *J. Clin. Sleep. Med.*, **4**: 99–103.
- Stanus, E. *et al.* (1987). Automated sleep scoring: A comparative reliability study of algorithms. *Electroencephalogr. Clin. Neurophysiol.*, **66**: 448–456.