RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2017.43.312

ScienceAsia 43 (2017): **312–318**

Improved GP algorithm for the analysis of sleep stages based on grey model

Yabing Li, Songyun Xie*, Jin Zhao, Chang Liu, Xinzhou Xie

NPU-TUP Joint Laboratory for Neural Informatics, Northwestern Polytechnical University, Xi'an, China, 710129

*Corresponding author, e-mail: syxie@nwpu.edu.cn

Received 6 Jun 2017
Accepted 7 Sep 2017

ABSTRACT: Correlation dimension analysis of EEG signals is widely used to access sleep stages. However, the standard Grassberger-Procaccia (GP) algorithm used to calculate the correlation dimension is very time consuming. To overcome this problem, an algorithm that combines the grey model and GP algorithm (GM-GP) is proposed. The results show that the correlation dimensions computed from GP and GM-GP are highly correlated, and the significance between the CDs in different stages of GM-GP is similar to GP. Furthermore, the computation time of the proposed method is at most 5% of that of the GP. The proposed algorithm is suitable for the real-time monitoring of sleep stages, which can provide a deeper understanding of brain function.

KEYWORDS: correlation dimension, time consuming, sleep analysis, GM-GP

INTRODUCTION

In addition to medical prevention, sleep analysis can also contribute to psychophysiological monitoring[1,2]. In general, sleep consists of six stages including wake (stage 0), light sleep (stages 1 and 2), delta or deep sleep (stages 3 and 4), and rapid eye movement (REM)[3,4]. Previous studies have also demonstrated that electroencephalograph (EEG) signals vary with sleep progresses, thus EEG has been widely used for assessing sleep stages[5,6]. Hassan et al extracted many features such as tunable-Q factor wavelet transform[7], spectral features, empirical mode decomposition[8], Gaussian parameters and statistical features[9] from EEG signals to classify sleep states. Liang et al employed multi-scale entropy and auto-regressive model parameters to classify sleep stages[10], considering that the characteristics of EEG signals are somewhat chaotic, not only the traditional features, the correlation dimension (CD) derived from nonlinear dynamical analysis are also applied to investigate the dynamic characteristics of EEG. As a measure of dimensionality of the space occupied by a set of random points, CD has been widely employed to analyse EEG signals and has also been proven to be effective when applied to analysing the dynamics of the brain function underlying behaviour[11,12]. It has been reported that the CD of human EEG changes as the sleep stages change, particularly in the wake and REM stages, the cortex is more active in these two stages than in other stages[13,14]. By analysing the relationship between sleep stages and CD, Ehlers et al indicated that the CDs of the wake and REM stages are significantly higher than those of the other stages[15,16]. Rajendra et al studied the six different types of sleep signals using the CD, and the results denoted that nonlinear parameters can be used to quantify the cortical function at different sleep stages[17]. Although CD can reveal information of the activity source, it requires a considerable amount of computation time, for example 30 s for a typical case ($N = 6000$, Fs $= 200$ Hz, duration of session $= 30$ s). Hence it is necessary to find more efficient algorithm for CD.

The grey model (GM), which was proposed by Deng, is primarily used for solving complicated mathematical problems and determining the implicit message of an uncertain or incomplete system[18]. The GM has evolved in many fields, such as forecasting, image denoising, which is achieved by reducing the amounts of data. Xie et al processed images (denoising, an edge detection method) using the grey prediction model, and the results indicated that it is feasible and effective on image denoising[19]. Hsu et al proposed a novel image compression using the GM on a dynamic window, which does not require an extra decoder and only utilizes the modelling parameters to reconstruct the image by reversing the operation of GM(1,1)[20].

In this paper, an algorithm that combines the GM and the Grassberger-Procaccia (GP) is proposed to improve the computational efficiency. The EEG signal is first compressed using the GM, then the CD of the compressed data is calculated using the GP algorithm. The performance of the proposed algorithm is evaluated by comparison with the GP algorithm, and the results demonstrate that GM-GP has the ability to assess sleep stages with few computations, and can be further used to realize real-time monitoring for the entire sleep sessions.

MATERIALS AND METHODS

Materials

The EEG data for analysis were obtained from the CAP sleep-EDF Database from the MIT-BIH Database Distribution, which is a collection from the Sleep Disorders Centre of the Ospedale Maggiore of Parma, Italy. The waveforms include at least 3 EEG channels (F3 or F4, C3 or C4 and O1 or O2, referred to as A1 or A2), EOG, EMG of muscle, and EKG. The 8 healthy subjects included in the study did not present any neurological disorders and were free of drugs that affect the central nervous system. Furthermore, the sleep stages (W = wake, S1–S4 = sleep stages, R = REM) were coded by expert neurologists, which is in accordance with the R&K rules[21]. In this work, the sleep stages were analysed based on P4–O2 and the maximum available sample size was 120. And each epoch lasted 30 s.

For all subjects, the raw EEG data were band-pass filtered between 1 Hz and 40 Hz. After that, the bad trials, which contained obvious artefacts (eye blink, eye movement, and EMG), were rejected based on EEGLAB for further analysis.

GP algorithm

CD describes the dimensionality of the underlying process in relation to the geometrical reconstruction in phase space. In this study, we estimated CD using an algorithm based on the GP algorithm[22]. The value of CD can be computed by the slope of the curve $\ln C(r)$ versus $\ln r$ over a scale-invariant interval,

$$\mathrm{CD} = \lim_{r\to\infty} \frac{\ln C(r)}{\ln r},$$

where the correlation integral $C(r)$ is

$$C(r) = \frac{2}{N^2}\left(\sum_{i,j} H(r - |y_i - y_j|)\right),$$

when y_i and y_j are the points of the trajectory in the phase space, N is the number of points, r is the radial distance, and H is the Heaviside function.

Note that the choices of an appropriate time lag τ and embedding dimension m are important. For the time lag τ, the first minimal value of the autocorrelation function is utilized. For the choice of embedding, the method is based on the idea that in the passage from dimension d to $d+1$, up to the slope of $\ln C(r)$ versus $\ln r$ is invariant[23].

The GM-GP algorithm

The CD achieved by the GP algorithm can be used to quantify the cortical function at different sleep stages, but it requires a considerable amount of time for the computation, making this algorithm less attractive. To overcome the shortcomings of the GP algorithm, GM-GP is proposed, which is described as follows.

The GM(1,1) can be expressed as

$$x^{(0)}(k) + az^{(1)}(k) = b,$$

where $z^{(1)}(k)$ is defined as the mean of $x^{(1)}(k)$, when $x^{(1)}(k)$ is the first-order accumulated generate operation (AGO) sequence of the original sequence $x^{(0)}(k)$ with n samples, marked 1-AGO, which can be expressed as

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$$
$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)),$$

where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, for $k = 1, 2, \ldots, n$. For the estimate for the parameter of GM(1,1), it can be expressed as

$$P = \begin{bmatrix} a \\ b \end{bmatrix} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}Y,$$

where

$$B = \begin{bmatrix} -z(2) & 1 \\ -z(3) & 1 \\ \vdots & \vdots \\ -z(n) & 1 \end{bmatrix}, \; Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}.$$

There are two propositions of GM(1,1) for describing the considered system, the development coefficient a and grey input b. The development coefficient a reflects the developmental tide of the GM(1,1) and b represents the total external excitation[18]. These parameters can reflect the trend of the original data to some degree, and thus the grey model can be used to compress data, which is shown

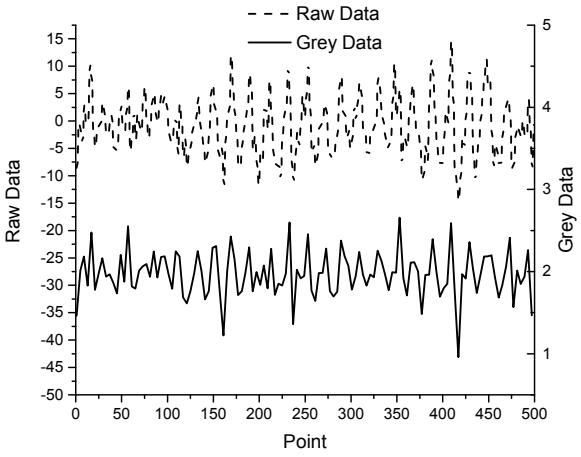

Fig. 1 Comparison of raw data with grey data (4 samples); the solid line represents the compressed data using grey model, while the short dashed line means raw data of CD for EEG signals.

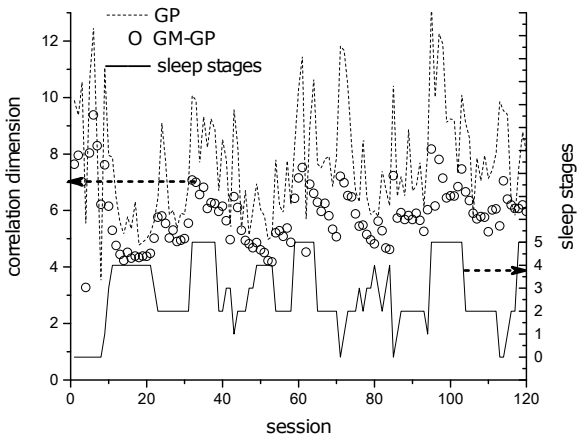

Fig. 3 Comparison of CD using different algorithms at session. The left and right vertical axes represent the CD and sleep stages, respectively. For the left vertical axis, the dash corresponds to the proposed algorithm; the hollow circles correspond to the GP algorithm and for the right vertical axis, the solid line represents the sleep stages.

in Fig. 1. It is clear that the grey data have the same trend as the raw data.

For the proposed algorithm, the original data are first compressed using GM(1,1) to obtain new data. Then the new data are used to calculate CD as input for the GP algorithm. By compressing the data, the amount of computation is decreased. The proposed algorithm is shown in Fig. 2.

All the processing steps were performed using MATLAB (Version R2014a, MathWorks), and some results were obtained using SPSS (IBM SPSS Statistics 22).

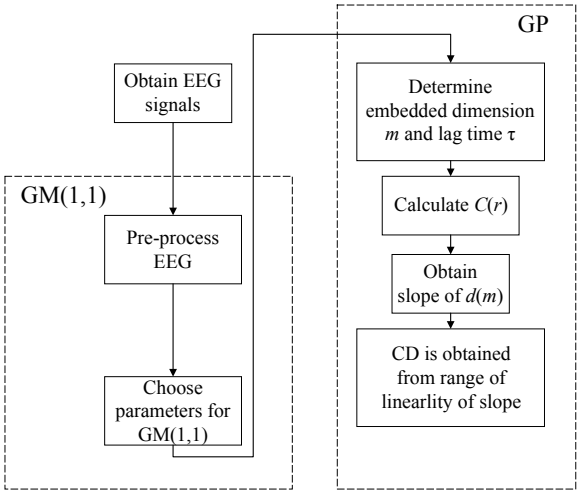

Fig. 2 The flow of GM-GP algorithm.

RESULTS AND DISCUSSION

The time varied CDs of one subject computed by GM-GP and GP are shown in Fig. 3. Considering the time consuming, 120 sessions were selected, which were acquired uniformly over the entire stages. From the results, although there are some differences between the results obtained using GM-GP and GP, the trends as the sleep stages change are the same. The CDs of sleep stage 0 (wake) and REM are larger than those of the other stages, which indicate that the brain is active in these stages.

The correlations between CDs computed using GP and CDs computed using GM-GP with different GM parameters (sample points) are shown in Fig. 4. For each subject, 20 samples for each stage were selected to perform statistical analysis. From the results, the CDs computed using GM-GP with 4 sample points are highly correlated with those computed using the GP algorithm. The performances of different sample points were also given in Fig. 5. It is clear that there was a higher correlation with GP method using grey-4 compared to others. From the result, the proposed method was more efficient, which was particularly in accuracy and stability.

ANOVA was conducted to detect the statistical significance between either the two independent variables or their interaction. The results are shown in Table 1 and Table 2. As shown in Table 1, the CDs computed from GP in different sleep stages are statistically significant. Similar results were obtained

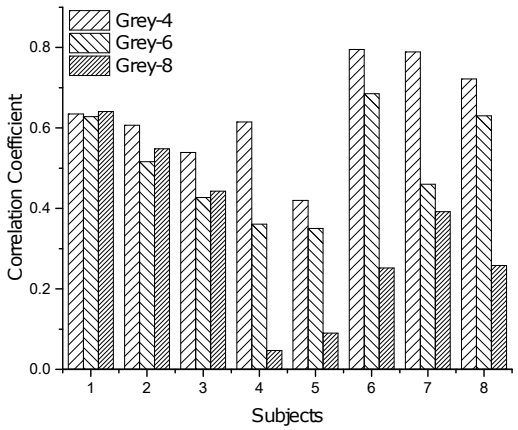

Fig. 4 Correlation between the results obtained by GP and the proposed algorithm. Grey-4, grey-6 and grey-8 represent 4, 6, and 8 sample points for parameter of GM(1,1), respectively.

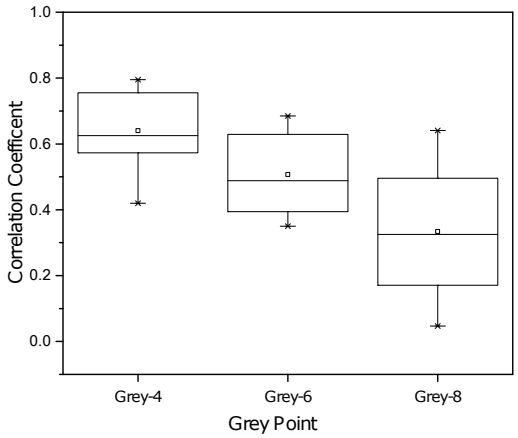

Fig. 5 Performance of recognition for three different samples in eight different participants. Each box represents the 25–75th percentiles, and central line is the median value, the tiny vertical lines extend to the most extreme data not considering as outliers, which are plotted individually. Statistical significant differences are marked ($p < 0.05$).

Table 1 Analysis of significance test of 8 subjects for different sleep stages with GP algorithm.

Subject		Wake	Sleep1	Sleep2	Sleep3	Sleep4	REM
S1	Wake		0.018	0.128	0	0.003	0.055
	Sleep1	0.018		0.091	0	0	0.001
	Sleep2	0.128	0.091		0	0.026	0.007
	Sleep3	0	0	0		0.006	0
	Sleep4	0.003	0	0.026	0.006		0
	REM	0.055	0.001	0.007	0	0	
S2	Wake		0.001	0.006	0	0.001	0
	Sleep1	0.001		0.026	0.181	0.018	0.107
	Sleep2	0.006	0.026		0.257	0.185	0.079
	Sleep3	0	0.181	0.257		0.156	0.006
	Sleep4	0.001	0.018	0.185	0.156		0.022
	REM	0	0.107	0.079	0.006	0.022	
S3	Wake		0.028	0.005	0.105	0.582	0.336
	Sleep1	0.028		0.709	0.455	0.086	0.002
	Sleep2	0.005	0.709		0.218	0.022	0
	Sleep3	0.105	0.455	0.218		0.282	0.011
	Sleep4	0.582	0.086	0.022	0.282		0.132
	REM	0.336	0.002	0	0.011	0.011	
S4	Wake		0	0.014	0	0	0.016
	Sleep1	0		0.202	0.286	0.002	0.179
	Sleep2	0.014	0.202		0.02	0	0.947
	Sleep3	0	0.286	0.02		0.043	0.017
	Sleep4	0	0.002	0	0.043		0
	REM	0.016	0.179	0.947	0.017	0	
S5	Wake		0.001	0.168	0.385	0.345	0.005
	Sleep1	0.001		0	0	0	0
	Sleep2	0.168	0		0.608	0.661	0.14
	Sleep3	0.385	0	0.608		0.941	0.048
	Sleep4	0.345	0	0.661	0.941		0.056
	REM	0.005	0	0.14	0.048	0.056	
S6	Wake		0.138	0	0	0	0.104
	Sleep1	0.138		0.005	0	0	0.709
	Sleep2	0	0.005		0.042	0	0
	Sleep3	0	0	0.042		0.013	0
	Sleep4	0	0	0	0.013		0
	REM	0.104	0.709	0	0	0	
S7	Wake		0.001	0	0	0.008	0
	Sleep1	0.001		0	0.022	0.461	0.079
	Sleep2	0	0		0.001	0	0
	Sleep3	0	0.022	0.001		0.073	0.181
	Sleep4	0.008	0.461	0	0.073		0.154
	REM	0	0.079	0	0.181	0.154	
S8	Wake		0	0	0.002	0.019	0
	Sleep1	0		0.003	0	0	0.013
	Sleep2	0	0.003		0	0	0.034
	Sleep3	0.002	0	0		0.069	0
	Sleep4	0.019	0	0	0.069		0
	REM	0	0.013	0.034	0	0	

Level of statistical significance: $p < 0.05$.

using the GM-GP algorithm (shown in Table 2). This demonstrates that the proposed algorithm is effective for distinguishing the sleep stages.

The time consumptions of both algorithms are shown in Table 3. One can observe that the GP algorithm requires considerably more time and points. Compared with the GP algorithm, the computation time of the GM-GP algorithm decreased to below 5% for each subject except S1. In summary, the proposed algorithm is more efficient.

Fig. 6 displays the receiver operating characteristic from our analysis, which indicates that the area in the upper left provides the better performance to discriminate different classifications. And the area under the curve (AUC) is shown in Fig. 7. Compared with other traditional spectral features—spectral centroid (SC), spectral flux (SF), spectral roll-off (SRO), spectral irregularity (SI), and spectral irregularity square (SIsquare), which are used

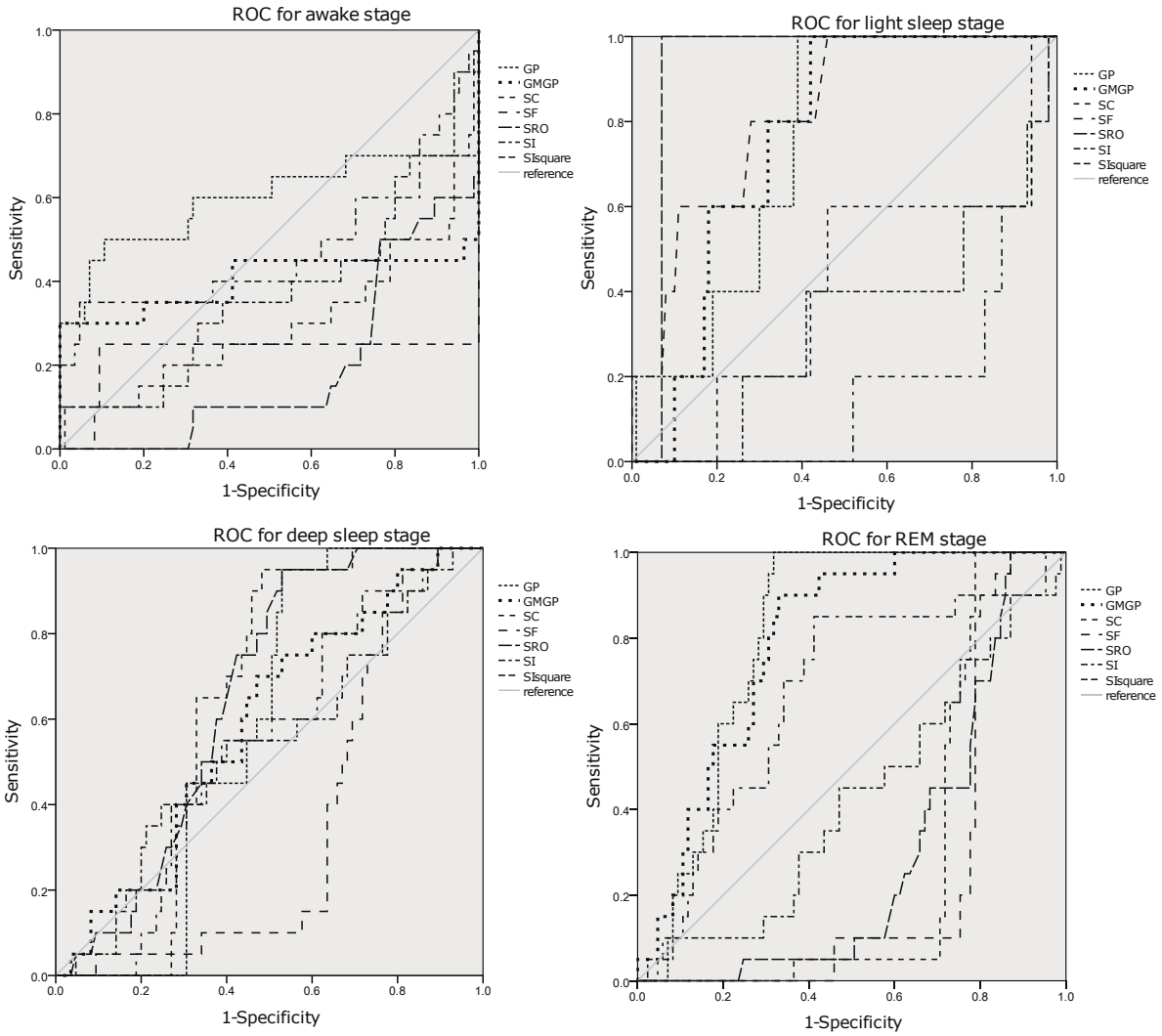

Fig. 6 ROC with different features (SC, SF, SRO, SI, and SIsquare represent spectral centroid, spectral flux, spectral roll-off, spectral irregularity, and spectral irregularity square, respectively) of sleep stages. Figures represent ROC for awake versus others, light sleep versus others, deep sleep versus others, and REM versus others, respectively.

in some literature[24,25]—we conclude that the AUC of GP and GM-GP is significantly better than others. In stages of awake and REM, the AUC of GP and GM-GP is dramatically larger than other features, except for SIsquare of awake stage. The performance in deep stage is similar for all these features without SF. And the AUC of GP and GM-GP is smaller significantly than SF and SRO, while larger overweight the features of the rest.

Limitations

Several limitations have to be taken into account when utilizing the results of these analyses. Firstly, the trend of GM-GP is similar to GP algorithm, while the values of CD for GP are not consistent with the CD for GM-GP method. Secondly, this algorithm requires signals that have chaotic property. Thirdly, there is part individual difference among different subjects, especially for different sample points of grey method.

CONCLUSIONS

In this paper, we presented a modified algorithm for analysing EEGs for different sleep stages. The proposed algorithm combined the grey model and the GP algorithm. The results demonstrated that the proposed algorithm is satisfactory with significant performance according to ANOVA, and thus it is effective for distinguishing sleep stages. Furthermore, it is considerably faster than the standard

Table 2 Analysis of significance test of 8 subjects for different sleep stages with the proposed algorithm.

Subject		Wake	Sleep1	Sleep2	Sleep3	Sleep4	REM
S1	Wake		0.23	0.103	0	0	0.132
	Sleep1	0.23		0.065	0	0	0.023
	Sleep2	0.103	0.065		0	0	0.117
	Sleep3	0	0	0		0.008	0
	Sleep4	0	0	0	0.008		0
	REM	0.132	0.023	0.117	0	0	
S2	Wake		0.015	0	0	0	0
	Sleep1	0.015		0.125	0.106	0.155	0.008
	Sleep2	0	0.125		0.022	0.199	0.831
	Sleep3	0	0.106	0.022		0.022	0.006
	Sleep4	0	0.155	0.199	0.022		0.032
	REM	0	0.008	0.831	0.006	0.032	
S3	Wake		0.002	0	0.001	0.002	0
	Sleep1	0.002		0.711	0.388	0.134	0.488
	Sleep2	0	0.711		0.256	0	0
	Sleep3	0.001	0.388	0.256		0.276	0.002
	Sleep4	0.002	0.134	0	0.276		0
	REM	0	0.488	0	0.002	0	
S4	Wake		0	0.01	0	0	0
	Sleep1	0		0.05	0.1	0	0.068
	Sleep2	0.01	0.05		0	0	0.7
	Sleep3	0	0.1	0		0	0.54
	Sleep4	0	0	0	0		0
	REM	0	0.068	0.7	0.54	0	
S5	Wake		0	0.003	0.71	0.001	0
	Sleep1	0		0.004	0	0	0.315
	Sleep2	0.003	0.004		0.005	0	0
	Sleep3	0.71	0	0.005		0.019	0
	Sleep4	0.001	0	0	0.019		0
	REM	0	0.315	0	0	0	
S6	Wake		0.233	0	0	0	0.011
	Sleep1	0.233		0	0	0	0.76
	Sleep2	0	0		0	0	0
	Sleep3	0	0	0		0	0
	Sleep4	0	0	0	0		0
	REM	0.011	0.76	0	0	0	
S7	Wake		0	0	0	0	0.002
	Sleep1	0		0.007	0.143	0.128	0
	Sleep2	0	0.007		0.12	0.091	0
	Sleep3	0	0.143	0.12		0.052	0
	Sleep4	0	0.128	0.091	0.052		0
	REM	0.002	0	0	0	0	
S8	Wake		0.107	0.063	0.056	0.087	0
	Sleep1	0.107		0.057	0.029	0.006	0.018
	Sleep2	0.063	0.057		0.029	0.048	0.018
	Sleep3	0.056	0.029	0.029		0.14	0.059
	Sleep4	0.087	0.006	0.048	0.14		0.059
	REM	0	0.018	0.018	0.059	0.059	

Level of statistical significance: $p < 0.05$. Proposed algorithm, grey point = 4.

Table 3 Comparison of time consumption between GP and GM-GP.

Sub.	Alg.	Fs (Hz)	$N^{\dagger}$	T (s)	Ratio‡
S1	GP	512	15 360	2605 ± 406	11%
	GM-GP		3840	276 ± 33	
S2	GP	128	3840	307 ± 43	4%
	GM-GP		960	13.3 ± 4.5	
S3	GP	128	3840	296 ± 73	3%
	GM-GP		960	9.4 ± 3.4	
S4	GP	200	6000	610 ± 44	4%
	GM-GP		1500	24.3 ± 8.9	
S5	GP	128	3840	230 ± 27	5%
	GM-GP		960	10.6 ± 4.1	
S6	GP	100	3000	137.2 ± 9.4	2%
	GM-GP		750	2.8 ± 0.0	
S7	GP	200	6000	632 ± 25	5%
	GM-GP		1500	30 ± 10	
S8	GP	100	3000	156.3 ± 6.4	3%
	GM-GP		750	4.4 ± 0.9	

† N represents the points for GP algorithm.
‡ Time ratio between GP and GM-GP.

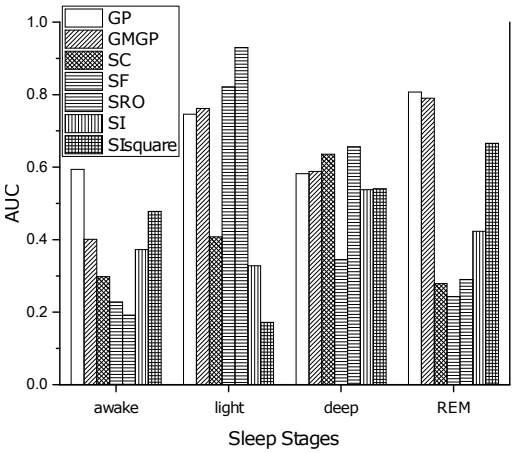

Fig. 7 Areas under the ROC curve with different methods.

GP algorithm. The proposed algorithm is with the ability of real-time monitoring of the sleep stages, which will provide a deeper understanding of the brain function.

Acknowledgements: Research reported in this paper is supported by National Natural Science Foundation of China (61273250), scientific and technological plan project of Shaanxi Province (2015GY003) and Joint International Research Centre on Integrated Technique of Brain-Computer for Unmanned System (3102017jc11002).

REFERENCES

1. Pan ST, Kuo CE, Zeng JH, Liang SF (2012) A transition-constrained discrete hidden Markov model for automatic sleep staging. *Biomed Eng Online* **11**, 52.
2. Hassan AR, Haque MA (2017) An expert system for automated identification of obstructive sleep apnea

from single-lead ECG using random under sampling boosting. *Neurocomputing* **235**, 122–30.
3. Kohtoh S, Taguchi Y, Matsumoto N, Wada M, Huang ZL, Urade Y (2008) Algorithm for sleep scoring in experimental animals based on fast Fourier transform power spectrum analysis of the electroencephalogram. *Sleep Biol Rhythm* **6**, 163–71.
4. Zorick T, Smith J (2016) Generalized information equilibrium approaches to EEG sleep stage discrimination. *Comput Math Meth Med* **2016**, 1–8.
5. Hsu YL, Yang YT, Wang JS, Hsu CY (2013) Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing* **104**, 105–14.
6. Peker M (2016) An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms. *Neurocomputing* **207**, 165–77.
7. Hassan AR, Subasi A (2017) A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowl Base Syst* **128**, 115–24.
8. Hassan AR, Bhuiyan MI (2017) Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Comput Meth Programs Biomed* **140**, 201–10.
9. Hassan AR (2016) Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting. *Biomed Signal Process Contr* **29**, 22–30.
10. Liang SF, Kuo CE, Hu YH, Pan YH, Wang YH (2012) Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Trans Instrum Meas* **61**, 1649–57.
11. Stam CJ (2005) Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin Neurophysiol* **116**, 2266–301.
12. Swie YW, Sakamoto K, Shimizu Y (2005) Chaotic analysis of electromyography signal at low back and lower limb muscles during forward bending posture. *Electromyogr Clin Neurophysiol* **45**, 329–42.
13. Botella-Soler V, Valderrama M, Crepon B, Navarro V, Quyen MLV (2012) Large-scale cortical dynamics of sleep slow waves. *PLoS ONE* **7**, e30757.
14. Zappasodi F, Olejarczyk E, Marzetti L, Assenza G, Pizzella V, Tecchio F (2014) Fractal dimension of EEG activity senses neuronal impairment in acute stroke. *PLoS ONE* **9**, e100199.
15. Ehlers CL, Havstad JW, Garfinkel A, Kupfer DJ (1991) Nonlinear analysis of EEG sleep states. *Neuropsychopharmacology* **5**, 167–76.
16. Janjarasjitt S, Scher MS, Loparo KA (2008) Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between sleep state and complexity. *Clin Neurophysiol* **119**, 1812–23.
17. Acharya UR, Faust O, Kannathal N, Chua T, Laxminarayan S (2005) Non-linear analysis of EEG signals at various sleep stages. *Comput Meth Programs Biomed* **80**, 37–45.
18. Liu S, Dang Y, Fang Z, Xie N (2010) *Grey System Theory and Its Application*, 3rd edn, Science Press, Beijing.
19. Xie S, Wang P, Xie Y (2008) New image denoising algorithm based on improved grey prediction model. In: *2008 Congress on Image and Signal Processing*, Sanya, China, pp 367–71.
20. Hsu YT, Yeh J (2000) A novel image compression using grey models on a dynamic window. *Int J Syst Sci* **31**, 1125–41.
21. Terzano MG, Parrino L, Sherieri A, Chervin R, Chokroverty S, Guilleminault C, et al (2001) Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med* **2**, 537–53.
22. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208.
23. Kim HS, Eykholt R, Salas JD (1999) Nonlinear dynamics, delay times, and embedding windows. *Physica D* **127**, 48–60.
24. Hassan AR, Haque MA (2016) Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating. *Biocybern Biomed Eng* **36**, 256–66.
25. Hassan AR, Bashar SK, Bhuiyan MIH (2015) On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram. In: *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, India, pp 2238–43.