Advances in Adaptive Data Analysis
Vol. 1, No. 4 (2009) 667–679

SEARCH FOR CLIMATE TRENDS IN SATELLITE DATA

ALEXANDER RUZMAIKIN* and JOAN FEYNMAN†

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California, 91109, USA
**alexander.ruzmaikin@jpl.nasa.gov*
†joan.feynman@jpl.nasa.gov

Unambiguous determination of trends is the central theme of climate change studies. A promising way to validate and quantify a warming trend of few degrees on a time scale of a century, predicted by climate models, is the use of measurements provided by satellites. However, climate variables, such as temperature and concentration of water vapor in the Earth's atmosphere, are noisy and subject to seasonal and interannual variabilities. The lifetime of any single satellite is relatively short (< 5–10 years) and producing long-term data records by using successive satellite introduces inter-calibration problems. Standard methods of trend determination, such as the least square fit of a linear trend, require sufficiently long time series and thus are not effective for the analysis of satellite data. Here, we applied the Empirical Mode Decomposition (EMD) and find it to be more efficient in the search for climate trends in the relatively short time series provided by satellites. We give examples of climate time series analyses using the EMD and discuss the problems we encounter in calculation and interpretation of trends extracted from the data limited in time extent.

Keywords: Empirical Mode Decomposition; trends; climate data from satellites.

1. Introduction

Climate models have predicted a linear trend of a few degrees per century caused by global warming due to anthropogenic forcing.[1] This translates into a small temperature trend of about 0.1–0.2 K per decade. The trend is superimposed on numerous much stronger natural climate variabilities, among them: the seasonal variations, interannual variations due to El Nińo, the Quasi-Biennial Oscillation (QBO), volcanic, and solar irradiance variability. Validation of the anthropogenic trend and separating it from the natural variabilities is a challenging task. One of the puzzling results of the trend determination is a substantial disagreement (about 0.1 K per decade) between the trend estimated at the Earth's surface (as recorded by thermometers placed at many locations and used in the IPCC reports) and a trend in the low troposphere found from satellite measurements.[2]

Satellite measurements of temperature, water vapor and other climate variables provide an excellent spatial coverage of the Earth's surface and the multiple layers of its atmosphere. The measurements are typically taken twice per day for a satellite in sunsynchronous polar orbits. However, the climate variables are noisy, and as mentioned above, are subjected to seasonal and inter-annual variability. In principle, the trend could be separated from noise and natural variabilities by using sufficiently long time series. Standard linear trend determinations, such as the least square fit to a line[2,3] require very long time series (20–30 years) to obtain the desirable accuracy. But satellite lifetimes are short (< 5–10 years) and long-term data records combining successive satellites have inter-calibration problems impeding the integrity and accuracy of the resulting time series.

Here we apply the Empirical Mode Decomposition (EMD)[4] and the concept of an adaptive trend[5] to selected satellite data to investigate the role of the time series length and noisiness of the data in determination of trends.

2. Data and Approach

For this investigation, we used the measurements of the solar UV irradiance obtained by the Solar Radiance Climate Explorer (SORCE) satellite[6] and the measurements of the Earth's atmospheric temperature obtained by the Atmospheric Infrared Spectrometer (AIRS) on board of the Aqua satellite.[7] The SORCE solar UV data at 280 nm (corresponding to Mg II spectral line) are complemented by the earlier measurements to form a time series of about 10,000 daily data points (27 years). The AIRS' 6 years of measurements generated 2190 daily data points. To better understand and evaluate the quality of analyses of these satellite data we generated an artificial data set of variable length, which simulates a controlled trend and basic natural variabilities.

We apply the EMD method,[4] which is ideally suited for analysis of non-stationary and nonlinear climate processes. The data noise is a special issue. Numerical experiments carried out by Flandrin *et al.*[9] and Wu and Huang[10] showed that the EMD applied to noise acts as dyadic filter expanding the data into normally distributed intrinsic mode functions. Wu and Huang[10] have further shown that the product of variances of these functions and their mean periods T is constant. Since each intrinsic mode function is normally distributed its variance obeys the χ^2 distribution with the degrees of freedom equaled to the number of data points times the variance. By knowing the variance distribution function one can determine the confidence intervals relative to the variance–mean period line, using the procedure similar to the determination of confidence intervals for the Gaussian distribution. As derived by Wu and Huang,[10] the confidence intervals for each variance (the confidence gate) are determined by

$$\ln(\mathrm{Var}) = \ln T \pm k(2/N)^{-1/2} \exp(\ln T/2), \tag{1}$$

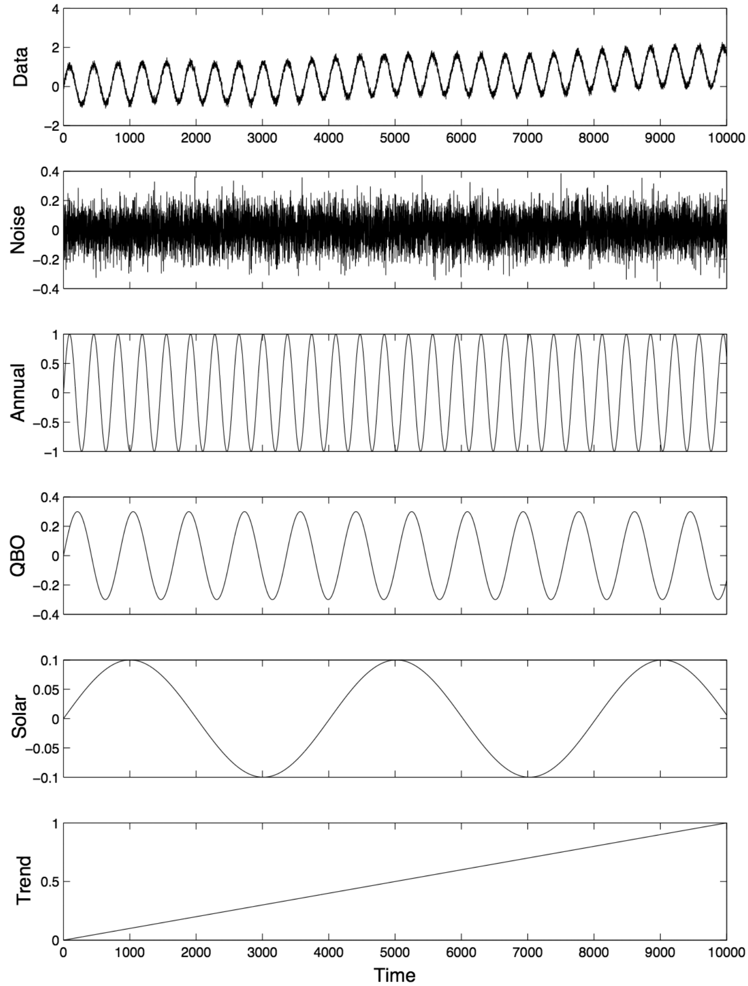

Fig. 1. The artificial "daily" data set (top panel) composed of the five components shown in separate subsequent panels: the white noise, the annual oscillation, the Quasi-Biennial (2.3 year period) Oscillation, the solar 11-year variability, and a linear trend $x = ct, c = 10^{-4}$.

where N is the number of data points and $k = 0.675$ for 99% significance level. Note that the width of the confidence gate is proportional to $N^{-1/2}$, i.e. it is narrower for larger N. Thus the variance–period dependence and the confidence gate allow an effective separation of the oscillating parts of signals from the noise. However, it still remains uncertain how to separate contributions of the signal and noise to a trend, which formally corresponds to $T \to \infty$, since the confidence gates diverge in this limit. Here we define "the trend" as the residual (= data minus the sum of oscillating EMD modes). And we assume that the trend is statistically significant if the last oscillating mode is statistically significant.

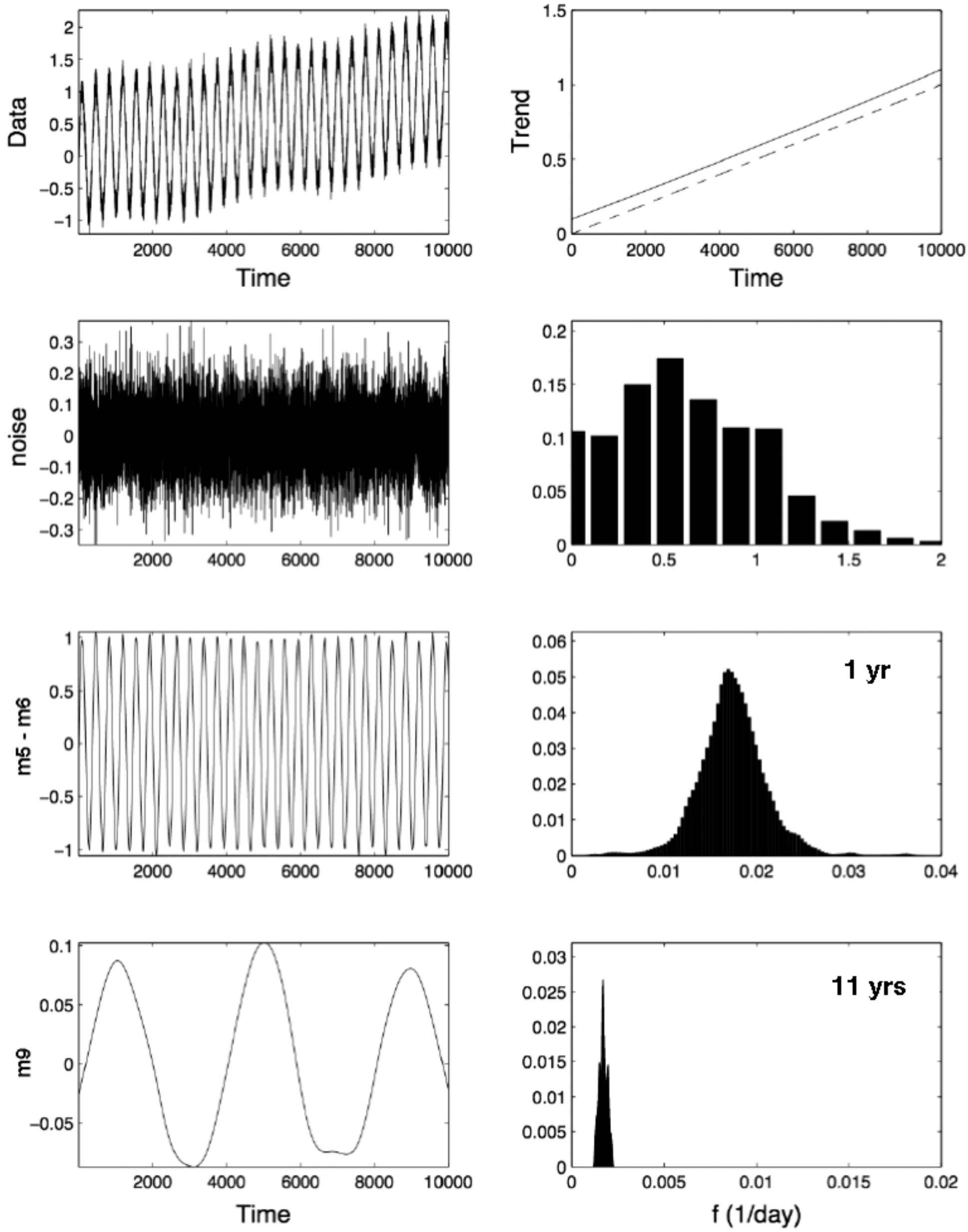

Fig. 2. The combined EMD modes (left panels), the residual (trend) (upper right panel), and histograms of their instantaneous frequencies (on corresponding right panels) for the data shown in Fig. 1. The histograms are calculated using the Matlab code (ksdensity.m) based on the Gaussian kernel weighting of bins. The numbers on the three right panels show the mean periods determined as the mean values of the histograms. The calculated (solid) and input (dash) trends are shown in the upper right panel.

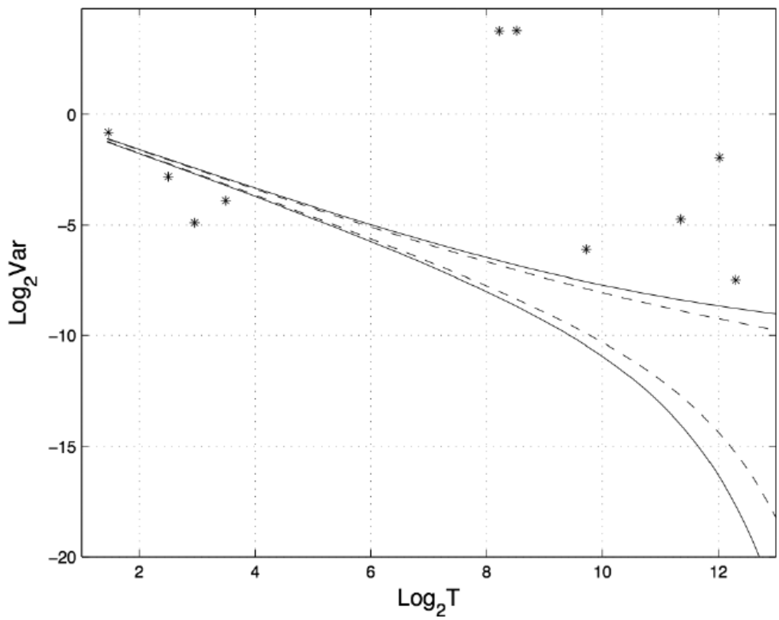

Fig. 3. Evaluation of statistical significance of the modes shown in Fig. 2 using the Log(Var)–Log T diagram. The solid and dashed lines outline the 99% and 95% of significance, respectively. The variances (energies) of the modes are shown by asterisks.

In our analysis, we used the EMD code written by Zhaohua Wu and now freely available.[8] In this code the number of EMD modes is set equal to the number of modes for a white noise time series of the same length as the signal length. The number of modes depends on the number of the data points[10] as $\log_2 N$. The major advantage of using this code is its capability of performing the ensemble EMD (EEMD).[4] An EEMD mode at a certain scale is a mean of EMD modes at the same scale each of which is obtained from the data plus a white noise realization. The EEMD allows eliminating a mixing of the EMD modes, i.e. having two or more EMD modes with the same or very close periodicity. The mixing is caused by the presence of time intervals in the data at which specific, usually low frequency variability dominates. A shortcoming of the EEMD code is that it typically generates a larger number of modes than a directly applied EMD implemented in other codes. As a result some modes should be combined if one is interested in their physical interpretation.

Since we are dealing with discrete time series sampling can affect the EMD introducing a nonzero contribution to the residual.[11] To ensure that there is no loss of extrema (interpolation between which is critical for the EMD algorithm) we follow these author's advice[11] that "*the sampling period be at most one half of the minimum distance between extrema in the signal, and the sampling period should be longer than any interval where the signal is constant*". However, the role of sampling has not been fully understood yet.

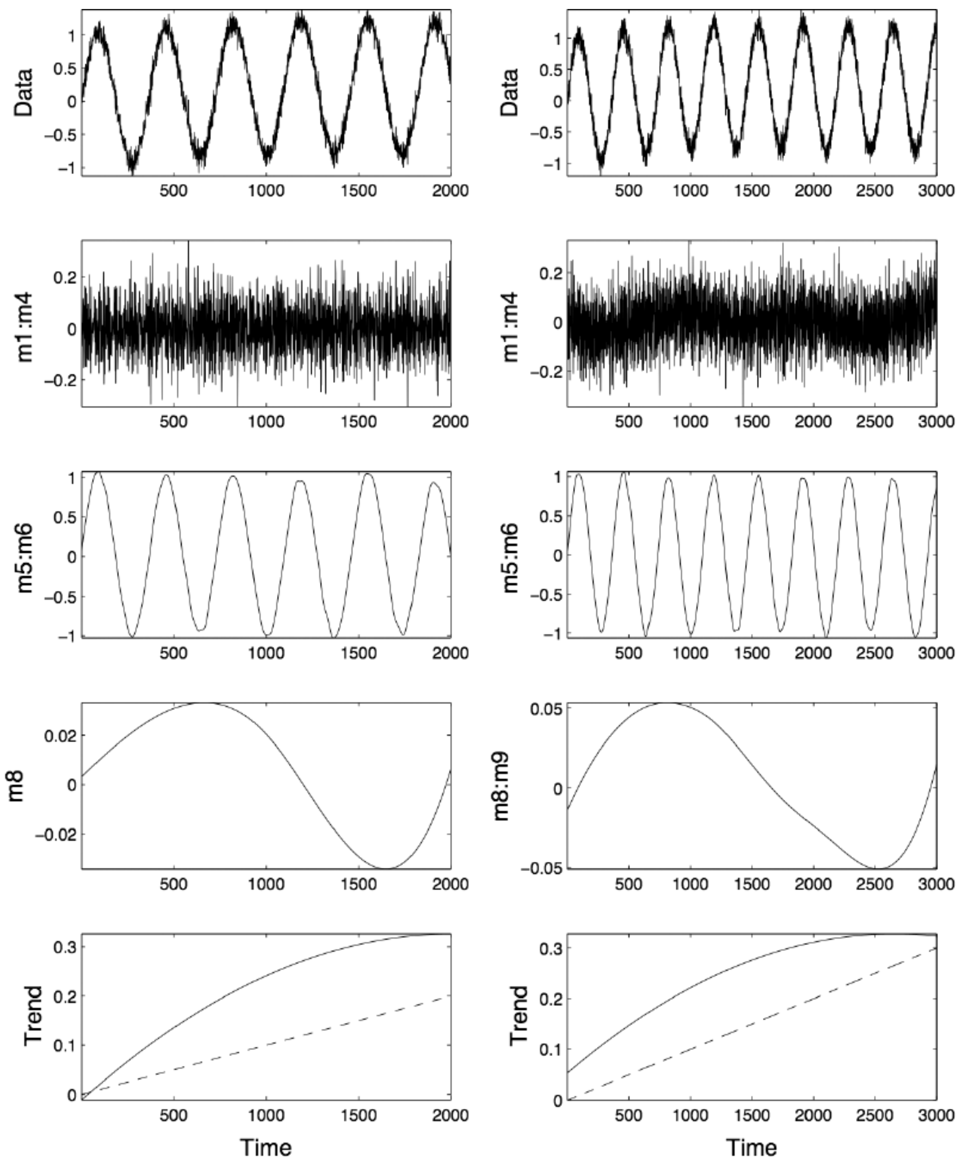

Fig. 4. The combined EMD modes obtained for shorter lengths of data shown in Fig. 1: 2000 points (left panels), 3000 points (right panels). Modes from 1 to 4 reproduce the noise. Modes 5 and 6 display the annual cycle.

3. Analysis of Artificial Data

We start with the analysis of an artificial data set generated to simulate the expected oscillations and trend (Fig. 1).

The application of the EMD algorithm to these data yields 13 modes for $N =$ 10,000. The modes are averages of a 300 member ensemble runs, but the results

Table 1. Slopes and their 1σ errors for the trends determined by the two methods. The slopes and errors are expressed in the units of 10^{-4} corresponding to the data trend $x = t/10,000$.

Method	10^4	5×10^3	3×10^3	2×10^3
LS fit	0.99 ± 0.02	0.80 ± 0.07	-0.08 ± 0.15	1.05 ± 0.3
EMD residual fit	1 ± 0.00	0.99 ± 0.001	1.05 ± 0.001	0.96 ± 0.004

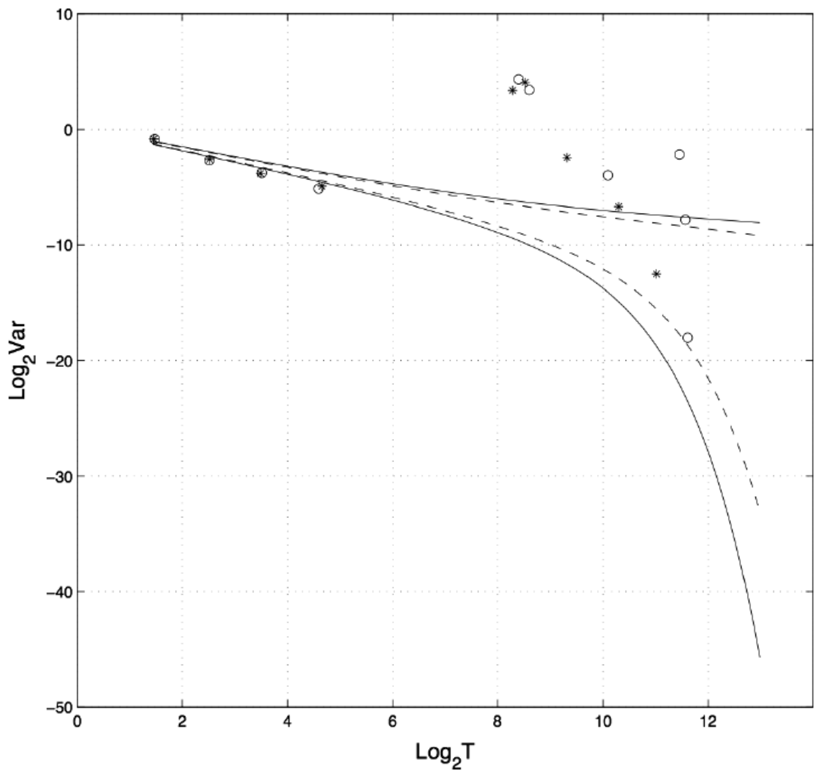

Fig. 5. The significance of the EMD modes obtained for the shorter test time series. The circles (stars) display the variances of the 3000 (2000) point data series.

only weakly depend on the ensemble number. The modes and distributions of their instantaneous frequencies corresponding the input noise, annual, solar oscillations and the trend are shown in Fig. 2. We see that the trend is well reproduced, as well as the annual and solar oscillations. However, the latter come with small amplitude modulations.

All modes are statistically significant although the significance level falls with the increase of the mode period (Fig. 3).

Shortening the data length generates fewer EMD modes (Fig. 4) but, as shown in Table 1, leads to a degradation of the quality of extraction of the trend and amplitudes of oscillating components. The last oscillating mode and thus the trend

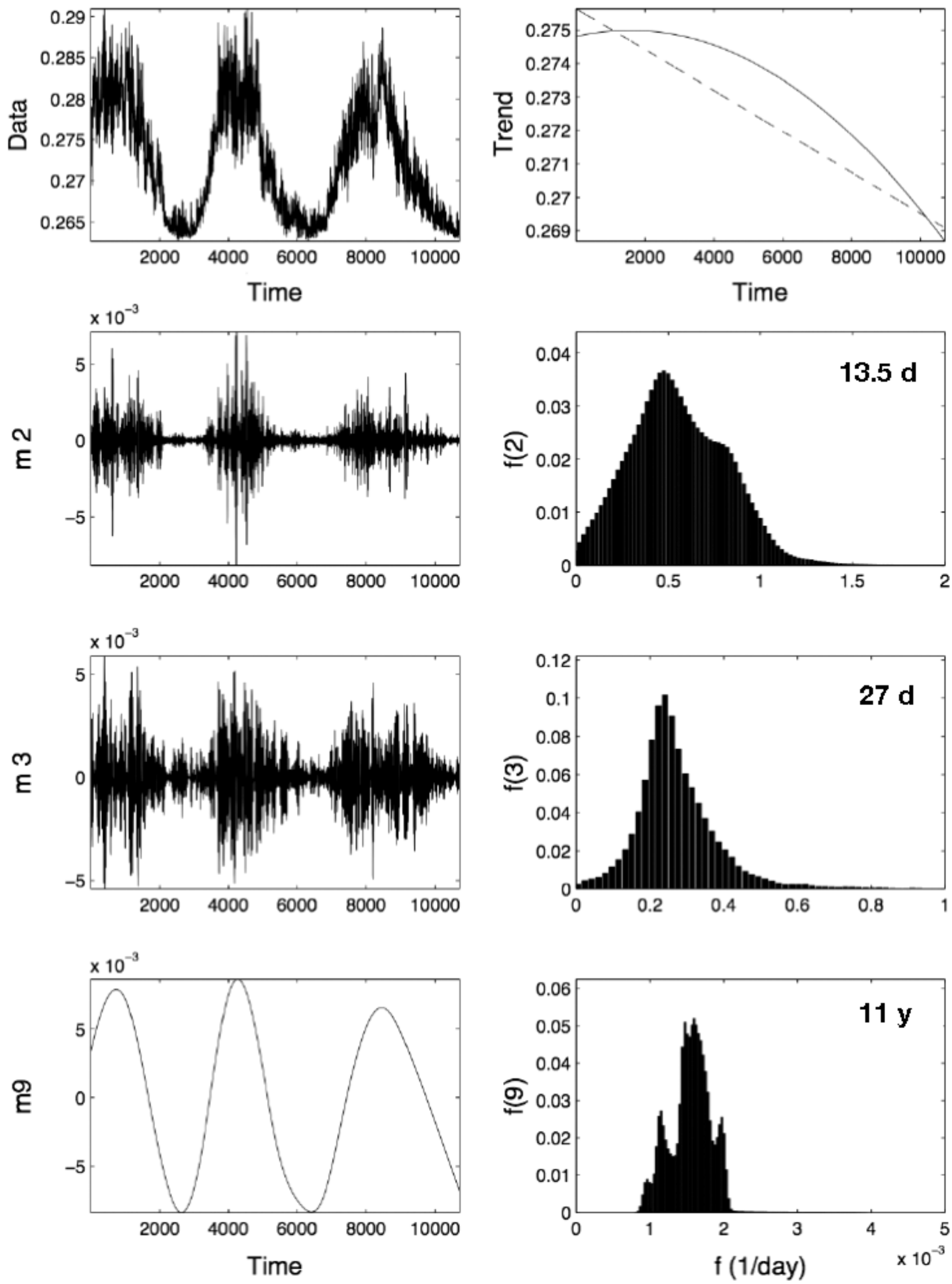

Fig. 6. The main EMD modes (left panels) and histograms of their instantaneous frequencies (right panels) of the solar UV irradiance in 1980–2008. Modes 2 and 3 arise due to the sources of the UV radiation on the Sun rotating with the rate and half rate of solar rotation.

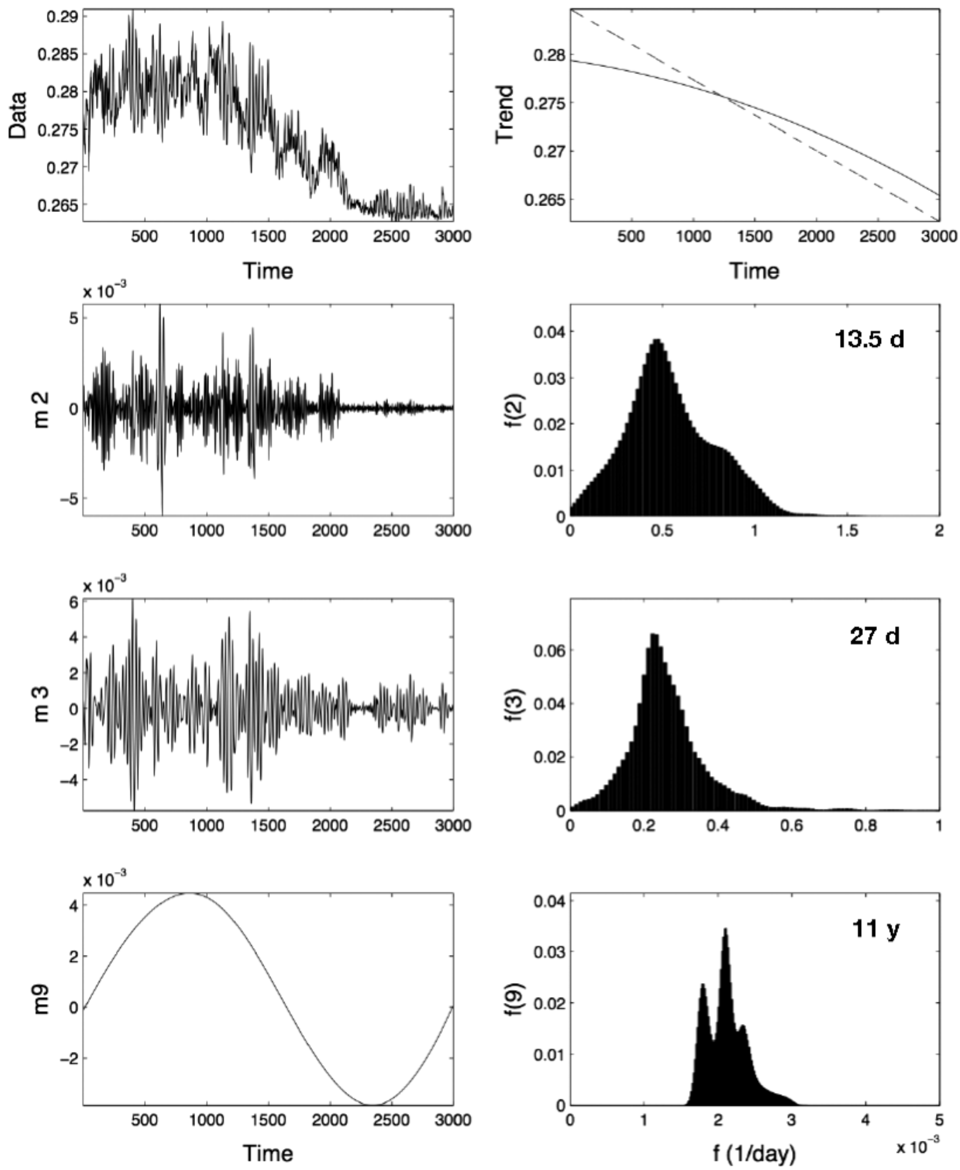

Fig. 7. Same as Fig. 6 but for the irradiance measurements in a shorter time interval, in 2002–2008.

for the 3000 point time series is statistically significant at 95% level. But the corresponding mode and trend for the 2000 point time series are significant only at the level below 90% (Fig. 5).

It is instructive to compare the EMD residuals (trends) with the trends determined by application of the least-square (LS) method directly to the data. The results of the comparison of the trend slopes obtained directly from the data and

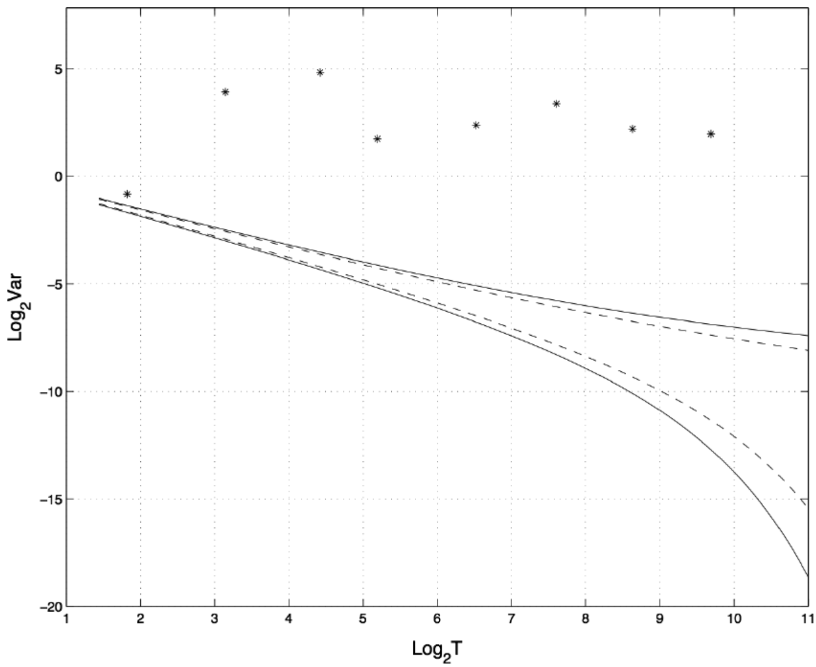

Fig. 8. The diagram of statistical significance of the modes shown in Fig. 6. The modes of Fig. 7 are also significant.

those obtained via fitting the slope to the EMD residuals are shown in Table 1 demonstrating the superiority of the EMD. Although the residuals for less than $N/2 = 5000$ data points become nonlinear the EMD still preserves the sign and a reasonable value of the slope.

4. EMD of Solar Irradiance

Now we apply the EMD to the solar irradiance data, first using the full available set 10,000 data points and a set abbreviated to 3000 data points. The modes presented here and in the next section are determined as the means of an ensemble of 300 modes, each of which is obtained from the data plus an independent realization of the white noise. (The choice of a different ensemble number in the range 100–500 does not affect the results.) For comparison with the results of analysis of the artificial data we note that in this case in addition to the basic 11-year solar cycle there are also modes properly reflecting the full and half solar rotation rate (27-day and 13.5-day oscillations). Figures 6 and 7 show the main modes and histograms of their instantaneous frequencies. In both cases all modes are highly statistically significant (Fig. 8). We also recover a trend, which can be interpreted as a declining phase of the 88-year (Gleissberg) cycle or a transition to the next Global Solar Minimum.

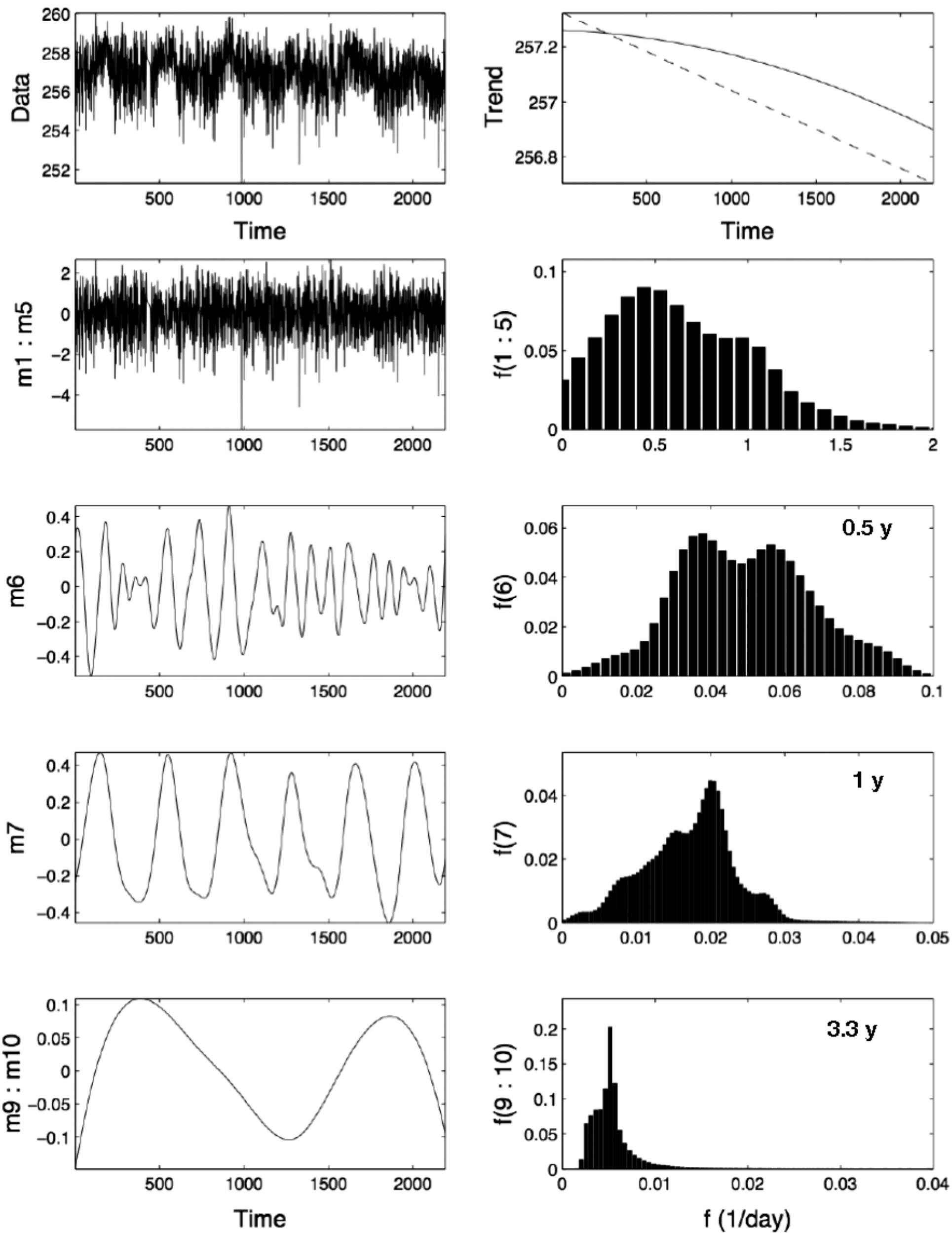

Fig. 9. The main EMD modes (left panels) and histograms of their instantaneous frequencies (right panels) of the Earth's infrared radiance over tropic ocean in 2002–2008. Modes 7 and 8 represent the annual variability. Modes 9 and 10 can be interpreted as caused by the ENSO (El Nińo – Southern Oscillation) variability.

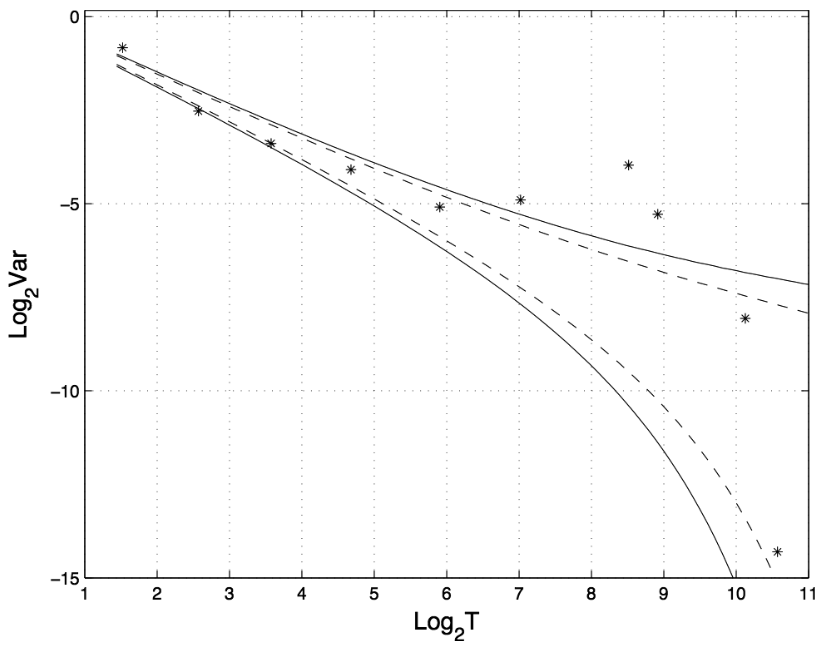

Fig. 10. The significance diagram for the modes shown in Fig. 9. The last mode and thus the trend are statistically significant at 95% level.

5. EMD of Earth's Infrared Irradiance

Next, we apply the EMD to the Earth's infrared irradiance data. In this case, we have only 2192 data points obtained from the Airs measurement. The data refer to the height of 400 hPa (about 5 km) in the atmosphere over the tropical ocean (0–30° N) in the time period 2002–2008. In this case, the basic oscillation is the annual cycle. Figure 9 shows the EMD modes and histograms of their instantaneous frequencies. The significance diagram in Fig. 10 indicates that modes 6–8 are statistically significant at 99% level. The last mode and thus the trend is significant at 95% level. Based on the significance diagram we can combine the first nonsignificant modes into a noise and a fewer oscillating modes (Fig. 9).

6. Conclusions

By consistent separation of noise and natural variability the EMD allows the identification of trends in time series of a limited length. In our tests, we found that the EMD is superior to the direct trending (de-trending) the data by standard least-square techniques. Our numerical experiments show that a qualitatively good identification of a mode could be achieved for time series that include one or two periods of oscillation. This sends an encouraging message to the space community on the possibility of extracting long-term oscillations and trends being limited by the finite time of the measurements.

Acknowledgments

We are grateful to Zhaohua Wu for providing the EMD code and numerous helpful discussions. We thank Norden Huang and the participants of the HHT conference in Guangzhou 2008 for the attention and useful questions related to our work. We acknowledge the use of the SORCE data, and we thank the Airs Project Scientist Hartmut Aumann for the help and discussion of the Airs data. The work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. *IPCC Climate Change 2007 — The Physical Basis* (Cambridge University Press, New York, 2007).
2. B. D. Santer,T. M. L. Wigley, D. J. Gaffen, L. Bengtsson, C. Doutriaux, J. S. Boyle, M. Esch, J. J. Hnilo, P. D. Jones, G. A. Meehl, E. Roeckner, K. E. Taylor and M. F. Wehner, Intrepreting differential temperature trends at the surface and in the lower troposphere, *Science* **21** (2008) 841–846.
3. S. S. Leroy, J. G. Anderson and G. Ohring, Climate signal detection times and constraints on climate benchmark accuracy requirements, *J. Climate* **287** (2000) 1227–1232.
4. N. E. Huang and Z. Wu, A review on Hilbert–Huang transform: Method and its applications to geophysical studies, *Adv. Adapt. Data Anal.* **1** (2009) 1–23.
5. Z. Wu, N. E. Huang, S. R. Long and C.-K. Peng, On the trend, detrending, and the variability of nonlinear and non-stationary time series, *Proc. Natl. Acad. Sci. USA* **104** (2007) 14889–14894.
6. G. Rottman, The SORCE Mission, *Solar Phys.* **230** (2005) 7–25, http://lasp.colorado.edu/sorce/index.htm.
7. M. T. Chahine, T. S. Pagano, H. H. Aumann, R. Atlas, C. Barnet, J. Blaisdell, L. Chen, M. Divakarla, E. J. Fetzer, M. Goldberg, C. Gautier, S. Granger, S. Hannon, F. W. Irion, R. Kakar, E. Kalnay, B. H. Lambrigtsen, S. Y. Lee, J. Le Marshall, W. W. McMillan, L. McMillin, E. T. Olsen, H. Revercomb, P. Rosenkranz, W. L. Smith, D. Staelin, L. L. Strow, J. Susskind, D. Tobin, W. Wolf and L. Zhou, The Atmospheric Infrared Sounder (AIRS): Improving weather forecasting and providing new data on greenhouse gases, *BAMS* **87** (2006) 911–926.
8. http://rcada.ncu.edu.tw/research1.htm
9. P. Flandrin, G. Rilling and P. Goncalves, Empirical mode decomposition as a filter bank, *IEEE Signal Process. Lett.* **11** (2004) 112–114.
10. Z. Wu and N. E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method, *Proc. Roy. Soc. London A* **460** (2004) 1597–1611.
11. G. Rilling and P. Flandrin, Sampling effects on the empirical mode decomposition *Adv. Adapt. Data Anal.* **1** (2009) 43–59.