World Scientific
www.worldscientific.com

HHT ANALYSIS OF THE GLOBAL AVERAGE MONTHLY PRECIPITATION DATA

SAMUEL S. P. SHEN* and DAVID NEW[†]

Department of Mathematics and Statistics
San Diego State University, San Diego, CA 92182, USA
*shen@math.sdsu.edu
†davidandrewnew@gmail.com

THOMAS M. SMITH

CICS/ESSIC, University of Maryland and NOAA/STAR
National Environmental Satellite, Data and Information Service (NESDIS)
5825 University Research Court, Suite 4001
College Park, MD 20740, USA
Tom.Smith@noaa.gov

PHILLIP A. ARKIN

CICS/ESSIC, University of Maryland
5825 University Research Court
Suite 4001, College Park, MD 20740, USA
parkin@essic.umd.edu

This paper uses the Hilbert–Huang transform (HHT) method to make time–frequency diagnostic analyses of four monthly time series of the global precipitation: MERG (1900–2008), REOF (1900–2008), GPCP (1979–2009), and CMAP (1979–2009). All these data are the global land and ocean average of precipitation anomalies with respect to the mean of the entire data period. The MERG and REOF are spectral reconstructions based on historical data. The GPCP and CMAP are based on station gauge data and satellite remote sensing data. We have made the following analysis of the four datasets: (a) extract intrinsic mode functions (IMF) by HHT empirical model decomposition (EMD) sifting, (b) calculate the mean frequency and energy of each IMF, (c) calculate the Fourier spectra to compare with the IMF spectral properties, (d) calculate the Hilbert spectra and display the time–frequency variation of the precipitation time series, and (e) calculate the basic statistics of the four datasets, including mean, standard deviation, skewness, kurtosis and inter-correlation among the datasets. Our analysis results indicate the following: (i) IMFs may contain physical signals of MJO (Madden–Julian oscillation), monsoon, annual cycle, and ENSO (El Nino southern oscillation), (ii) Hilbert spectra appears to be an effective tool to display the time-frequency change of a precipitation time series and can help identify critical characteristics for improving data aggregation method and climate models, (iii) among the four datasets, MERG is the smoothest data

and has the smallest variance and hence the smallest IMF energies, while the CMAP has the largest, followed by GPCP and REOF, and (iv) the nonlinear and nonstationary annual cycle is the IMF3 for all the four datasets, which is modulated by ENSO signals.

Keywords: Hilbert–Huang transform; global precipitation; intrinsic mode function.

1. Introduction

A reliable estimate of the monthly or annual global total precipitation over the Earth from observations is important for assessing the influence of climate change on the global hydrologic cycle. Currently, there exist several observational datasets and over 20 model simulations of global precipitation. Satellite-based datasets began in 1979 while reconstructed and model data began in 1900 or earlier. These datasets are useful for understanding climate change and for the improvement of climate models. Although the observed data share many common characteristics and confirm some large-scale climate dynamics, these datasets have significant differences. The most noticeable difference is that the model simulations consistently yields an annual global average precipitation around 3.1 mm/day, which is about 15% higher than the observed or reconstructed data of 2.7 mm/day [Arkin *et al.* (2007); Smith *et al.* (2008)]. Even among the observed datasets, differences are apparent. Some have a positive linear trend while others negative. The purpose of this paper is to introduce the Hilbert–Huang Transform (HHT) method into the analysis of global precipitation and to provide new perspectives of spectral properties beyond the traditional Fourier analysis. The HHT analysis will help identify differences in the four observed datasets of the monthly global average precipitation. These datasets are (i) CMAP [CPC Merged Analysis of Precipitation, Xie and Arkin (1996), (1997), (1998); Janowiak and Xie (1999)], (ii) GPCP [Global Precipitation Climatology Project, Adler *et al.* (2003); Ebert *et al.* (1996); Huffman *et al.* (1997)], (iii) MERG [merged analysis of the monthly global precipitation, Smith *et al.* (2010)] and REOF [Reconstruction by Empirical Orthogonal Functions, Smith *et al.* (2011); Efthymiadis *et al.* (2005)]. The REOF data used here is the REOF version 2 data REOFv2 [Smith *et al.* (2011)], an improvement of the original REOF [Smith *et al.* (2008)]. These global data include areas of both land and ocean as well as ice. There are other datasets that include land only [Dai *et al.* (1997)]. Our global monthly data are displayed in Fig. 1 with CMAP and GPCP starting from 1979 and MERG and REOF starting from 1900. Five aspects of the HHT time frequency analyses have been considered: Intrinsic mode functions (IMF) derived from the procedures of empirical mode decomposition (EMD), IMF energy analysis, IMF mean period analysis, Hilbert spectral analysis, and Fourier analysis. These diagnostic analyses indicate clear differences of spectral characteristics of CMAP, GPCP, MERG and REOF data, which may suggest a need for revisiting the methods of aggregating the data from stations, satellite remote sensing, and other instruments. The results may also help improve precipitation simulations in general circulation climate models.
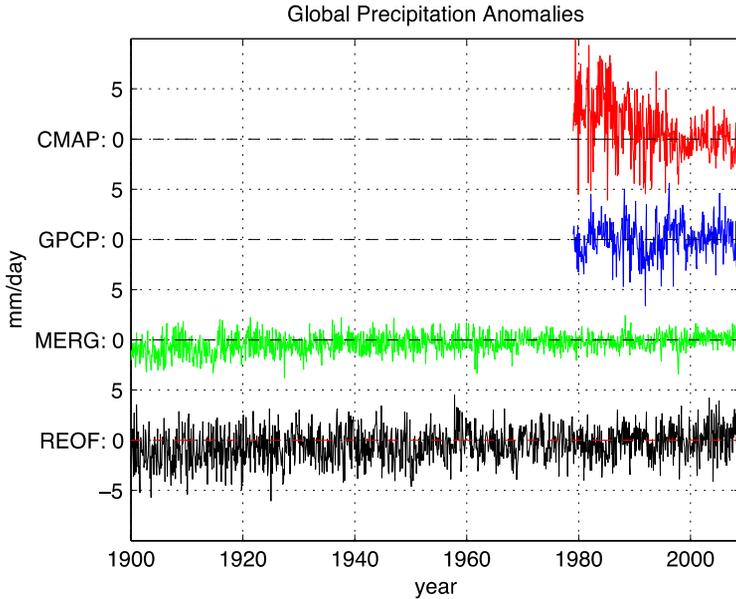
Fig. 1. Four monthly global land and ocean average precipitation datasets: CMAP (1979–2008), GPCP (1979–2008), MERG (1900–2008), and REOFv2 (1900–2008).

Data and method are described in Sec. 2. Results are in Sec. 3. Conclusions and discussion are in Sec. 4.

2. Data and Method

2.1. *Data*

The four observed global average precipitation datasets CMAP, GPCP, MERG and REOF are depicted in Fig. 1. The main features of these four datasets are briefly described below.

CMAP is a monthly $2.5° \times 2.5°$ latitude–longitude gridded global precipitation dataset from January 1979 to current. The global average CMAP data used here is the area-weighted spatial average of this gridded data from 1979 to 2008. CMAP is a blended product of the data from several satellites and the ground rain gauges. Thus, CMAP is an observed dataset processed with specific blending algorithms. CMAP satellite-based product is adjusted monthly over low and mid-latitude oceans according to the average of rain gauge observations at atolls and low islands in the tropical Pacific. Since precipitation over the atolls and low islands is heavier than the adjacent oceanic satellite-derived estimates, the adjustment led to an increase in oceanic precipitation. Although those atolls and islands are in the central and western part of the tropical Pacific, this adjustment was applied to the global tropics and subtropics. More satellite sensors joined the dataset in the 1990s [Simpson *et al.* (1996); Wentz and Spencer (1998)]. An interpolation-type blending

produces a weighted average of all the satellite sensors and the ground gauges. The blending also smoothens the data being interpolated.

GPCP is also a global observed precipitation dataset and has several temporal and spatial resolutions. The one used here is the area-weighted average of the $2.5° \times 2.5°$ monthly global precipitation data from January 1979 to current. GPCP also uses both satellite data and the ground gauge data but with different aggregation algorithms. Instead of using the weighted average of all the satellite products as CMAP does, GPCP uses the best estimate from a single satellite.

The MERG reconstruction dataset merges a monthly regional reconstruction with an annual global reconstruction [Smith *et al.* (2010)]. This global monthly data has a 5° resolution and covers from January 1900 to current. The monthly reconstruction used for MERG fits the available land and island gauge data to a set of spatial modes in order to produce reconstructions separately in the northern extra tropics, the tropics, and the southern extra tropics. Those three regional analyses are then combined. The annual ocean-area reconstruction used for MERG is based on canonical correlation analysis (CCA), which predicts historical annual precipitation anomalies from the annual sea-surface temperature (SST) and sea-level pressure (SLP) anomaly data. The CCA was judged to be more reliable for resolving multi-decadal variations over oceans, and therefore the ocean-area multi-decadal component of the combined regional analysis was replaced by the CCA component. Since the CCA prediction has a smoothing effect, the variance of the MERG data is smaller than the variances of CMAP and GPCP.

The REOF is another reconstructed dataset also at 5° resolution from January 1900 to current. The reconstruction based entirely on empirical orthogonal function (EOF) modes was computed as an alternative to the MERG reconstruction. The selected EOF modes are anchored on the historical gauge data using a least square approach. Computationally, a set of 20 global-annual EOFs is computed. Annual-average gauge data is analyzed by fitting to this set of modes, upon screening to remove modes not adequately sampled. Next a set of 40 monthly increment modes is computed to analyze the increment between the monthly and annual averages. Monthly gauge increments are analyzed using this set of modes. The sum of the annual and monthly increment components forms the REOF gridded data. Compared to the MERG data, the REOF data has less smoothing, and thus has larger variance (see Table 1).

The linear trend and the first four statistical moments of the datasets are shown in Table 1. For the 1979–2008 period, this table indicates the following. (a) CMAP has the largest positive mean (0.87 mm/day) mostly attributed to the 1980s large positive anomalies, and REOF has the smallest negative mean (−0.10 mm/day). (b) CMAP has the largest standard deviation (2.81 mm/day), and MERG has the smallest standard deviation (0.74 mm/day) and hence is the temporally smoothest time series. (c) CMAP is skewed right, but the other three skewed left. (d) The kurtosis for each dataset is positive and hence the probability distribution is leptokurtic, i.e. more peaked than the normal distribution. (e) The REOF yields the strongest

Table 1. Statistical moments and linear trends of the CMAP, GPCP, MERG, and REOFv2 data.

	Mean (mm/day)	Standard deviation (mm/day)	Skewness (dimensionless)	Kurtosis (dimensionless)	Trend ((mm/day)/decade)
			1979–2008		
MERG	−0.08	0.74	−0.30	1.17	0.1624
REOFv2	−0.10	1.47	−0.06	−0.01	0.3193
GPCP	−0.01	1.73	−0.10	0.69	0.0958
CMAP	0.87	2.81	0.43	0.42	−1.5153
			1900–2008		
MERG	−0.36	0.94	−0.21	0.16	0.0895
REOF	−0.49	1.62	−0.10	−0.11	0.1126

positive trend of 0.3193 (mm/day)/decade, while CMAP yields the strongest negative trend of −1.5153 (mm/day)/decade. The negative trend might be attributed to the large positive anomalies in the early 1980s.

For the period of 1900–2008 and for MERG and REOF data, the means (−0.36 mm/day for MERG and −0.49 mm/day for REOF) are smaller than the 1979–2008 period (−0.08 mm/day for MERG and −0.10 for ROEF). The standard deviations are about 20% larger than the 1979–2008 period. The skewness and kurtosis maintain the same sign, although the kurtosis of MERG in 1900–2008 is much smaller. The trends of 1900–2008 (0.0895 (mm/day)/decade for MERG and 0.1126 (mm/day)/decade for REOF) are much weaker than those of 1979–2008 (0.1624 (mm/day)/decade, and 0.3193 (mm/day)/decade).

The inter-correlations of the four datasets in the period of 1979–2008 are shown in Table 2 and have very small values. These low correlations may be due to strong noise and weak signal in each of the datasets.

2.2. *Method*

The HHT analysis method used here follows that of Huang *et al.* [1998], a time domain analysis using EMD, and a frequency domain analysis using Hilbert spectra [Huang and Shen (2005); Shen *et al.* (2005)]. The time–frequency analysis can identify instantaneous frequency and its associated amplitude of a time series. The HT (Hilbert transform) is an effective tool for calculating phase and amplitude, but

Table 2. Inter-correlations of the CMAP, GPCP, MERG, and REOFv2 data, and energy of mode R.

	MERG	REOFv2	GPCP	GPCP	Mode R energy (mm/day)^2
MERG	1				0.1961
REOFv2	0.2144	1			0.3423
GPCP	0.2870	0.2355	1	1	0.0124
CMAP	0.0539	−0.0051	0.2127	0.2127	3.0355

it often cannot be directly applied to a time series. For example, when the HT was applied to the analysis of the nonlinear vibrations [Feldman (1997)], the instantaneous frequency could not be calculated from the original time series. One problem is the non-differentiable phase function, which leads to nonphysical instantaneous frequency as a temporal derivative of the phase. HHT attempts to overcome this difficulty by sifting the original time series by the EMD procedure and leads to the intrinsic mode functions (IMF). The original time series is decomposed into a finite number of IMFs, each of which are either quasi-periodic and quasi-symmetric, or quasi-monotone and have a properly defined temporal derivative of the phase function. Thus, the IMFs allow the depiction of high-resolution Hilbert-spectra for nonlinear and nonstationary data. Mathematically, the phase functions and their numerical derivatives are well defined for these IMFs, which satisfy the following two conditions:

(i) In the temporal domain of the data, the number of extrema and the number of zero-crossings must either be equal or differ at most by one, and

(ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

With these requirements, each IMF oscillates in a narrow frequency band, a reflection of quasi-periodicity and weak nonlinearity. Of course, the nonconstant frequency means nonstationarity. When an IMF $c(t)$ is found, its HT, $H\{c(t)\}$, can be found. Then

$$c(t) + iH\{c(t)\} = a[t]\exp[i\theta(t)] \tag{1}$$

is well defined for a given t. This equation yields the instantaneous frequency $\omega(t) = d\theta/dt$ and its associated amplitude $a(t)$. The triplet (t, ω, a) forms the Hilbert spectral power [Shen *et al.* (2005)].

3. Results

3.1. *MERG data diagnostics*

The MERG dataset (1900–2008) yields 8 IMFs and a nonlinear trend, denoted as mode R (Fig. 2). The largest variations are in the first two modes. The main variations of the data are in IMF1, 2, 3 and R. Mode R is a nonlinear trend and also shows a clear nonstationarity. The periods of IMF1, 2 and 3 are approximately 1/4, 1/2 and 1.0 year (Fig. 3(a)). Thus, IMF1 may be contributed mainly by the tropical ocean precipitation associated with Madden–Julian Oscillation (MJO) whose period is around 2–3 months. IMF2 may be attributed to the tropical monsoon precipitation which typically peaks twice a year due to a seasonal wind reverse. This semi-annual cycle might be the consequence of the strong boreal summer and winter monsoon precipitation in the near tropical areas over the ocean. The two rainy seasons in Malaysia belong to this type of monsoon precipitation.
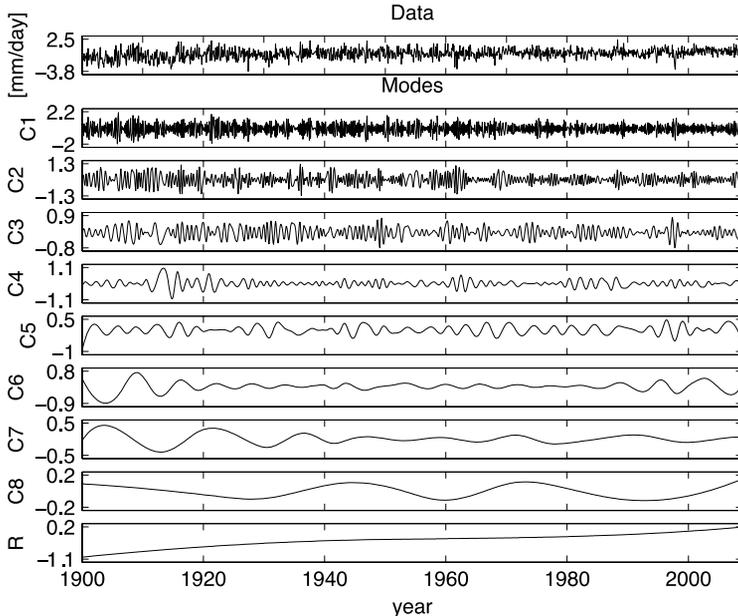
Fig. 2. IMFs and the nonlinear trend of MERG (1900–2008) data. The top panel shows the global average monthly precipitation anomaly data. The last panel is the nonlinear trend of the data. The middle panels are the IMFs of the data.

IMF3 is the annual cycle. Although the data are the anomalies and the global precipitation does not have a strong annual cycle, the year-after-year repeat pattern is still in the data. This annual oscillation is much weaker compared to the quarter-year or semi-year oscillations. The IMF energies and periods in Fig. 3 show the relative energy magnitude of the IMFs and their corresponding periods. Here the energy of an IMF is defined as

$$E_n = \frac{1}{T} \sum_{t=1}^{T} h_n^2(t) \tag{2}$$

which is the variance if the IMF has a zero mean.

The IMF energy decays roughly exponentially with base 2 from mode 1 to mode 5, suddenly increases with mode 6 and decreases exponentially again to mode 8, and finally has a large surge at mode R to 0.2 $[\text{mm/day}]^2$ due to the large positive trend (Fig. 3(b) and Table 3). Figure 3 does not show this energy since the trend mode does not have a period. The trend energies are shown in the last column of Table 3. The energies of the first three modes are approximately $\frac{1}{2}, \frac{1}{7}, \frac{1}{15}$, respectively. The mean period of the IMFs increases exponentially in base 2 from mode 1 to mode 7. Period doubling is a property of white noise in EMD analysis [Wu and Huang (2004)] and is often presented in the EEMD, perhaps due to the addition of white noise to the data although the white noise's magnitude may be small. For all the eight modes, the period doubling is almost strictly followed: $1/4, 1/2, 1, 2$, etc.
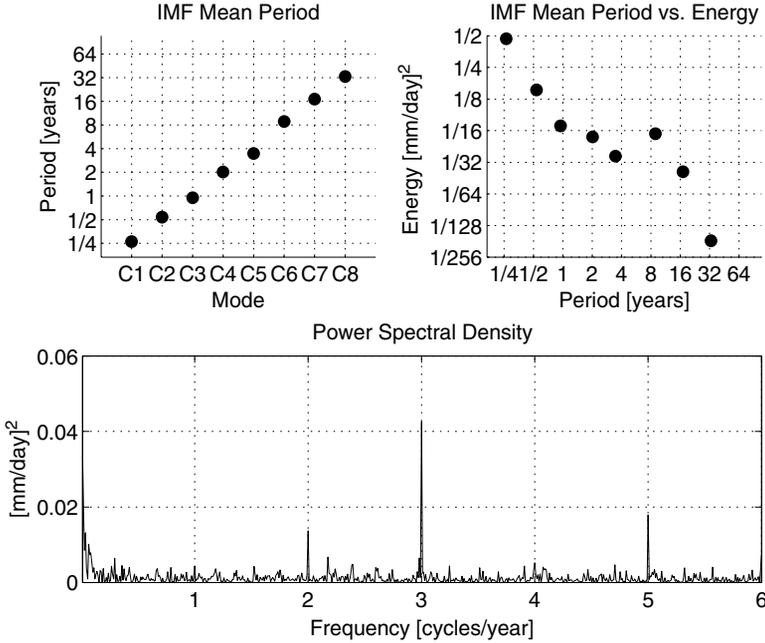
Fig. 3.   Mean period and energy of IMFs, and Fourier spectra of the MERG data. The top two panels use the logarithmic scale.

(see Fig. 3(a)). IMFs 4–6 have periods of 2–8 years and may correspond to El Nino Southern Oscillations (ENSO). These three IMFs' energy is small compared with the first mode and R mode, and hence IMFs 4–6 may be regarded as a minor modulation of the high frequency global precipitation's variation. IMF3 is regarded as the residual of the seasonal cycle that was removed when anomalies were computed, and thus has a small energy as well. Further, like the global average temperature, the global average precipitation has a small annual cycle. When considering precipitation over a region with a strong annual cycle, the annual cycle IMF will have large energy. The last two modes' periods are greater than 16, indicating very slow oscillations as depicted by the IMFs in Fig. 2.

The trend of the MERG precipitation data is positive, with an increase about 1.0 [mm/day] per century. This trend is not linear and is not a regression result. The positive direction of the trend is consistent with that of GPCP, AR4 models, and REOF analysis.

Fourier spectra of the MERG data are shown in Fig. 3(c), which has a frequency range [0, 6]. The maximum frequency limit 6 is due to our use of monthly data (12 data points in a year) and a necessity of a maximum and a minimum per cycle ($12/2 = 6$ cycles per year). Figure 3(c) indicates a dominant power peak at frequency 3 (3 cycles per year) and a major peak at frequencies 5 (Fig. 3(c)). These two peaks may be related to the first IMF's energy in Fig. 3 and may be attributable to MJO precipitation. The distribution of one IMF's power into several Fourier

spectral power may be due to the nature of Fourier spectra's linear dispersion, while nonlinear data analysis using HHT can retain energy in a single mode.

Another major Fourier spectral peak is at frequency 2, i.e. the semi-annual cycle, corresponding to the second IMF whose mean frequency is approximately 2. This semi-annual signal may be attributable to the monsoon precipitation which has a strong semi-annual cycle in most locations around the globe.

The Hilbert spectra help depict the time-frequency variations of a time series. A given time, t, corresponds to a frequency, $\omega[t]$, and an amplitude, $a[t]$. Thus, on the (t, ω) plane, each point corresponds to an amplitude that is a function of both time, t, and frequency, ω, but the time and frequency are not independent; rather, they are related by a function $\omega[t]$. The triplet $(t, \omega[t], a[t])$ determines a point in three-dimensional space (t, ω, a) [Shen *et al.* (2005)]. For a given time and a given IMF mode we can find a frequency by the Hilbert transform $\omega[t]$, hence a point on the (t, ω) time–frequency plane, and find the corresponding amplitude, $a[t]$. One can find this $a[t]$ for all IMFs except the last one which is a nonlinear trend. For any other grid point on the (t, ω) plane which does not correspond to an IMF, the amplitude is assigned to be zero. These amplitudes form the discrete Hilbert-spectra. For practical calculation, we define a grid over the (t, ω) plane. The time resolution is monthly, and hence the t-axis has 1308 (= 109 year times 12 months) points for the MERG and REOF data. For monthly data, the maximum frequency in the units of cycles/year is $12/2 = 6$, which is both the range of the instantaneous frequency ω and the range of Fourier power spectra. This range is divided into $1308/2$ grid points. Thus, the (t, ω) domain is January 1900 to December 2008 and 0 to 6 cycles/year. This rectangular domain has $1308 \times (1308/2)$ uniformly divided grid points. If the grid point (t, ω) corresponds to an IMF, then $a(t)$ is possibly nonzero. Otherwise it is zero. Since most points do not correspond to IMFs, the (t, ω, a) values are dominated by zeros and are difficult to analyze visually. To better visualize the spectra, we apply a Gaussian filter to the triplet (t, ω, a) to depict the amplitude variations on the time–frequency domain (see Fig. 4). Modes 1 to 8 are displayed in the Hilbert spectra. Each mode follows a trace of bright dots. The dots of Mode 1 scatter around over the top half of the plot. Mode 2 traces the dots around a frequency of 2 cycles/year. Mode 3 is the annual cycle and has a clear trace around the frequency of 1 cycle/year. The dots of Modes 4, 5 and 6 stack together because of low frequency oscillations. Because they stack together, the discrete Hilbert spectra of the low frequency modes are almost continuous over the time–frequency plane after the Gaussian smoothing. The Hilbert spectra shown in Fig. 4 indicate the following: (a) The frequency domain of the first two modes has a large range from 2 to 6 cycles/year, (b) the annual cycle mode IMFs has a range of 0.3–1.5 cycles/year, (c) the IMF4, . . . , 8 have a narrow frequency range 0–1 cycles/year since they represent very slow oscillations, (d) the amplitude of IMF1, 2, and 3 varies little according to time, and hence the nonstationarity of the first three modes is very weak, (e) the variation of the amplitude of IMF4, . . . , 8 is apparent and manifests both nonstationarity and nonlinearity, and (f) the strongest
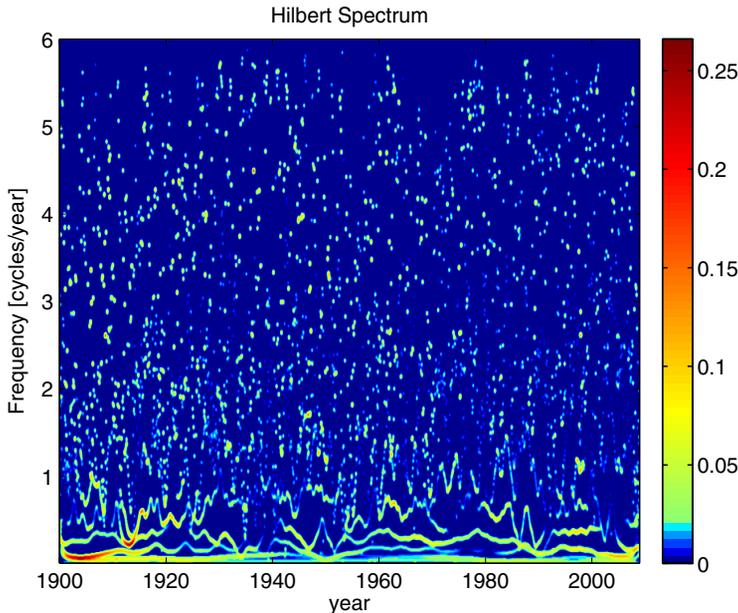
Fig. 4. Hilbert spectra of the MERG data (1900–2008). The vertical frequency axis has units [cycle/year].

nonlinearity is reflected in the nonlinear trend mode R which is not shown in the Hilbert spectra.

3.2. *REOF* (1900–2008) *data diagnostics*

REOF reconstructed data were generated from global EOFs by a linear regression. Both REOF and MERG use EOFs based on GPCP, and much of the variance in each reconstruction is from fitting gauge-analysis anomalies to a set of EOF modes. Thus, the characteristics of the REOF data and the MERG data have many similarities. REOF data also have eight IMFs and a positive nonlinear trend (see Fig. 5). Figure 6 indicate that the mean periods between REOF and MERG data are very similar: The mean period of the first IMF is also about 1/4 year, those of IMF2 and 3 are 1/2 and 1, respectively. So the IMF3 is the annual cycle. The IMF2 is the semi-annual cycle. IMFs 4–8 have little variation and contain little energy. The IMF energies decay exponentially from mode 1 to mode 4, and from mode 6 to mode 8 at a faster rate. However, there are differences between the REOF and MERG modes. The annual cycle IMF3 of the REOF has a much stronger modulation, which is shown in both IMF3 time series of Fig. 5 and the Hilbert spectra of Fig. 7. In addition, REOF has a stronger nonlinear trend, with an increase of about 1.6 mm/day from 1900 to 2008. This is consistent with Sec. 2's linear trend results which are 0.1126 (mm/day)/decade for ROEF and 0.0895 (mm/day)/ decade for MERG. The REOF IMFs have stronger energies. The total energy of all
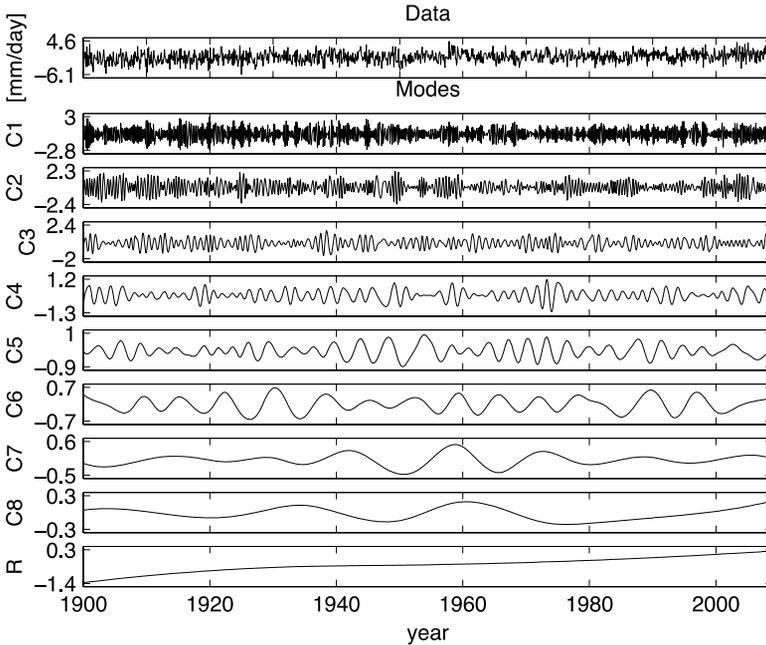
Fig. 5.   IMFs and trend of the REOFv2 data (1900–2008).

the 8 IMFs and mode R is about 3.0 $(mm/day)^2$ while that of MERG is only about 1.0 $(mm/day)^2$. This means that ROEF is less smoothed than MERG and retains more variations of global precipitation. In particular, the ROEF trend contains an energy of 0.34 $(mm/day)^2$ and thus shows a strong nonlinear upward trend that might be correlated with a SST warm trend in the same period.

Figure 7 demonstrates that the REOF annual cycle has a frequency band from 0.6 to 1.3, which is narrower than that of MERG. The strong modulation of the annual cycle has about 10 peaks from 1900–2008 and may be related to PDO (Pacific Decadal Oscillation).

The Fourier spectra of the REOF data have very different properties from those of the MERG data. The REOF data show a strong peak at frequency 2, a second strong peak at frequency 1, and a third peak at frequency near 0. The first peak may be related to IMF2 (semi-annual cycle), the second peak to IMF3 (annual cycle), and the third peak to mode R (the trend).

3.3. *Monthly GPCP global average precipitation data* (1979–2009)

Compared to the MERG and REOF data, the GPCP data are smoother and have some apparent low-frequency modulation. The GPCP data have 7 IMF modes and a positive nonlinear trend, i.e. mode R, shown in Fig. 8. The first IMF has some intermittency, a mixture between different frequencies. The variance of this mode
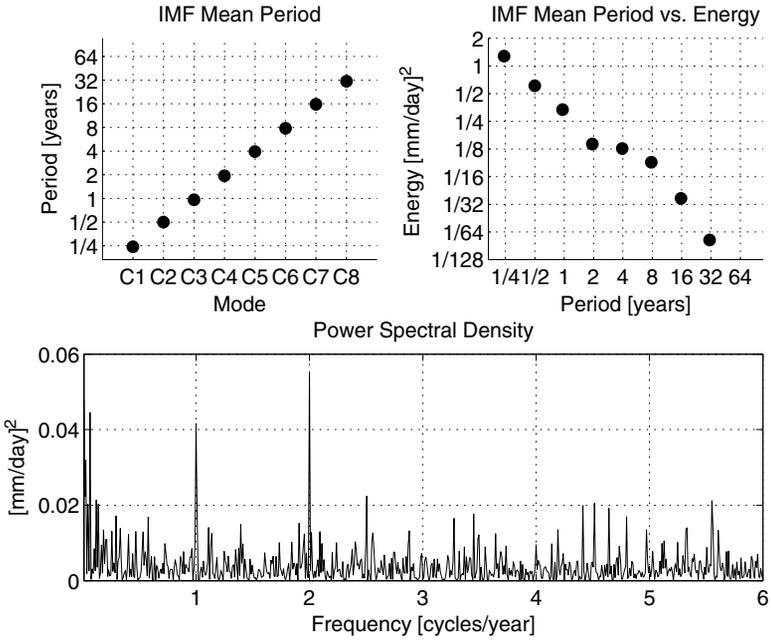
Fig. 6.   Mean period and energy of the REOFv2 IMFs and the Fourier spectra of the REOF data.
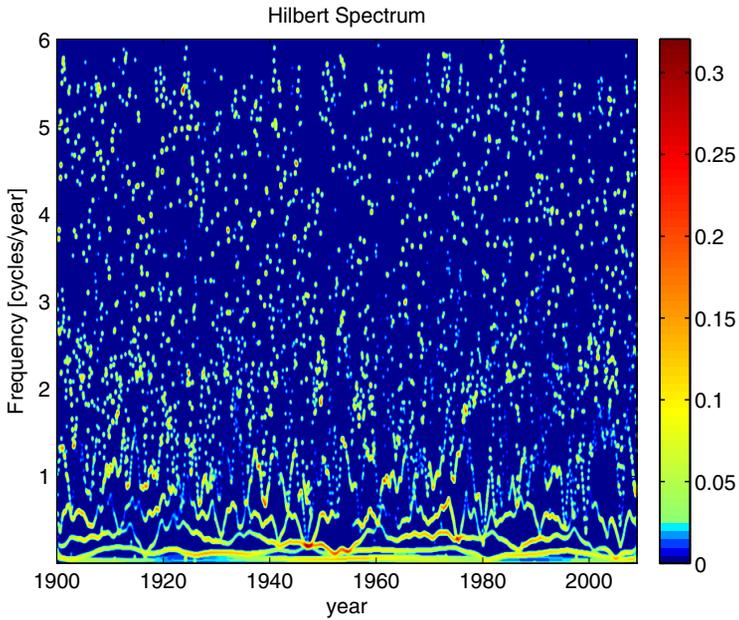

Fig. 7.   Hilbert spectra of the REOFv2 data (1900–2008).
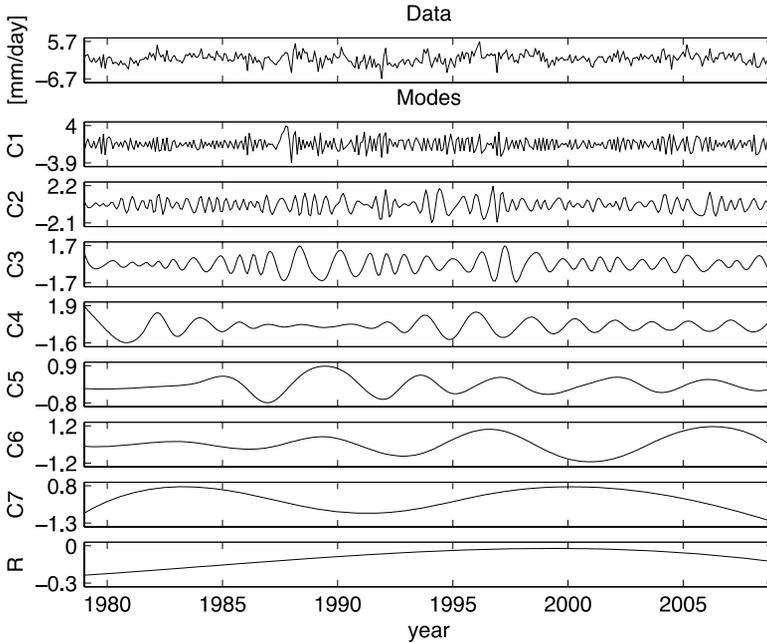
Fig. 8.   IMFs of the monthly GPCP global average precipitation data from 1979–2009.

was obviously larger in the 1990s than in the 1980s, and became smaller in the last decade. IMF3 and 4 have stronger variations in the late 1980s and the 1990s. The third mode is the annual cycle, which is nonstationary. The slow modulation at a period of 2–12 years is in IMF4–7. Whether these modulations are related to ENSO, PDO, or some other defined climate mode requires further investigation. On the other hand, these modulations may have been caused by the methods of data processing and changes of satellite sampling, and hence these slow modulations may have no clear physical implications. If so, the data aggregation algorithms may need to be re-visited.

Figure 9 shows mean periods and energies of the GPCP IMFs, as well as the GPCP Fourier spectra. The GPCP IMF energies do not decay exponentially according to the mode number as the MERG and REOF IMF energies. Instead, the energies decay sharply from IMF1 to mode 2. The energies of IMF2, 3, 4, 6 and 7 are nearly flat and constrained within 1/4 and 1/2 $[\text{mm/day}]^2$. The slow-decaying behavior of the energy indicates the strong low-frequency variations in the range of 2–12 years, which are not well represented in REOF and MERG data perhaps because of the use of EOFs that do not retain the low frequency oscillations well. The EOF and CCA reconstruction in MERG and REOF data depend on the base data for the EOFs. The MERG and REOF reconstructions used the same base dataset GPCP that may not have sufficiently long history to effectively reflect this low-frequency variation.
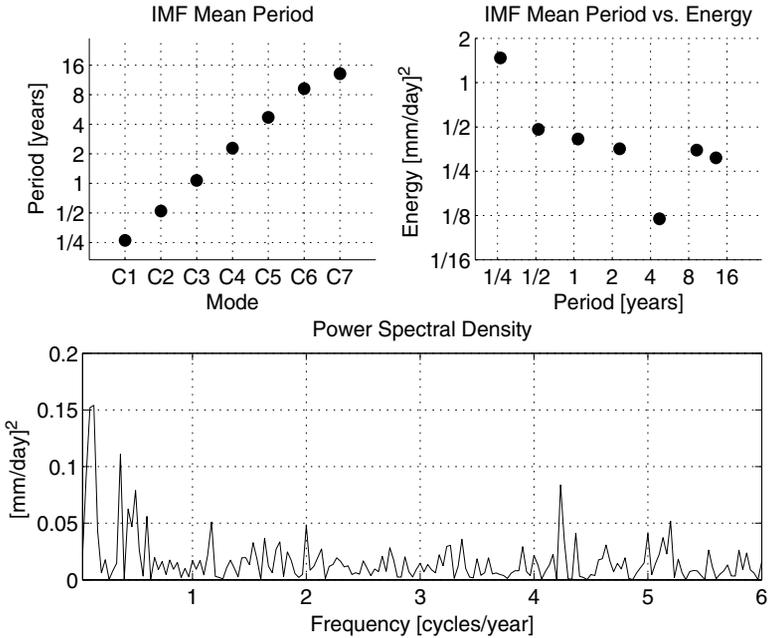
Fig. 9.   Mean period and energy of the GPCP IMFs and the Fourier spectra of the GPCP data.

The mean period for IMF1 is about 1/4 year and increases via an approximate period-doubling relation to 15 years for IMF7. This is similar to that of MERG and REOF data. Because the GPCP IMF energies do not decay exponentially, the energies are thus not from white noise, and hence the low-frequency variations may be physically meaningful assuming that they are not generated from data aggregation algorithms [Wu and Huang (2004)]. The total energy is about 3.0 $(mm/day)^2$, which is about the same as that of REOF and more than MERG. Nonlinearity must play a very important role here in the low-frequency variation since the main peak of the Fourier spectra (see Fig. 9(c)) is near zero frequency. This strong low-frequency modulation is different from that of the MERG and REOF data. An explanation for this is that the MERG and REOF data were a result of linear regression and can focus energy on some specific modes, while the GPCP are from a fusion of satellite retrieval and in situ gauges, which can more effectively reflect the intrinsic nonlinearity of the Earth climate processes. The above implies that each IMF has strong intermittency, and the global precipitation does not possess a persistent linear and stationary cycle at a fixed frequency.

The strong nonlinearity and nonstationarity are also clearly depicted in the Hilbert spectra of the GPCP data (Fig. 10). The Hilbert spectra for the four IMFs clearly indicate strong energy in the 1990s with variation frequencies between 0 and 2 cycles/year. Whether these strong IMF energies in the 1990s are due to physical processes or to changes in observational methods or to the data processing methods
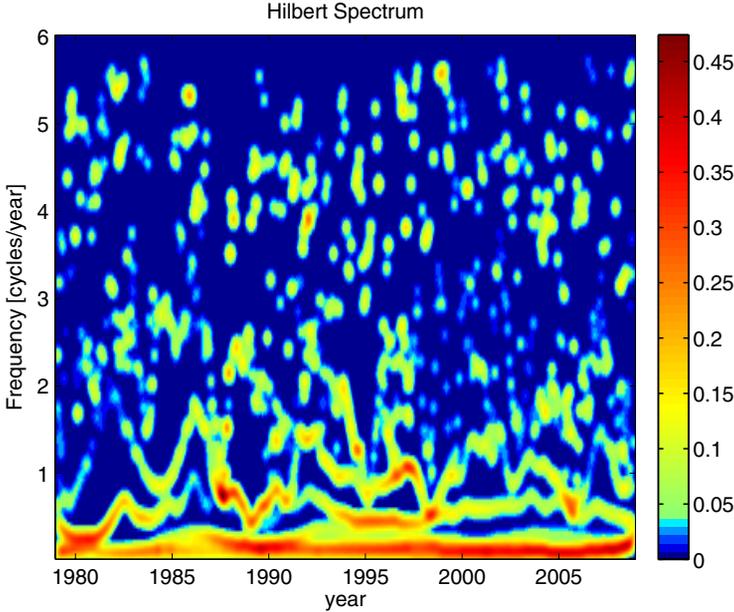
Fig. 10.   Hilbert spectra of the monthly GPCP global precipitation data (1979–2009).

is to be investigated. Using conventional Fourier analysis, one cannot detect the interdecadal signal with 30 years of data. An advantage of EMD method is that EMD can extract a low frequency signal with a short section of data [Huang *et al.* (1998); Huang and Shen (2005)]. However, this is only an empirical conclusion and has not been mathematically proven yet. Nonetheless, it cannot be ruled out that the slow variation is due to instrument change rather than a real physical phenomenon.

3.4. *Monthly CMAP global average precipitation data* (1979–2009)

CMAP is a commonly used global precipitation dataset that is anchored by over 20000 ground stations whose data are blended with satellite data. The CMAP IMFs in Fig. 11 also show some low-frequency modulation, as shown in the GPCP IMFs, but with different modulation properties. The first four modes all show an apparent amplitude decrease around 1995. The amplitudes in the last decade are comparable to those of GPCP in the same period. This means that the CMAP anomaly amplitudes in the 1980s are much bigger than those of GPCP in the same period. This can even be clearly seen from the data themselves. CMAP data oscillate between -6 and $10\,\mathrm{mm/day}$ in the 1980s, while the GPCP data oscillate only between -5 and $4\,\mathrm{mm/day}$. This leads to a question as to which dataset overestimates or underestimates the global precipitation in the 1980s.

A dramatic feature of the CMAP IMFs is that CMAP has a decreasing nonlinear trend. This is in contrast to those increasing trends of MERG, REOF, and GPCP,
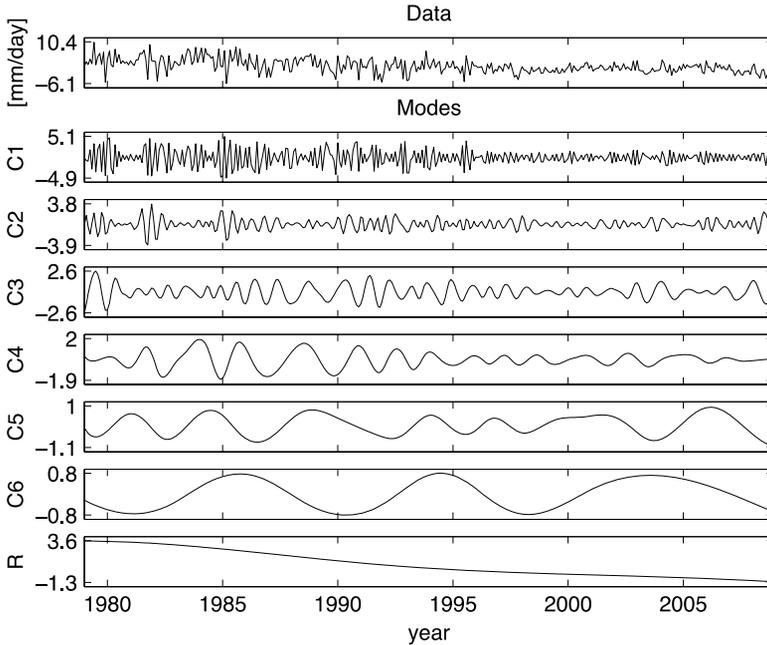
Fig. 11.   IMFs and trend of the CMAP data (1979–2009).

but is consistent with CMAP's negative linear trend in Table 1. Land observations and climate models all support an increasing trend or a positive linear trend [Held and Soden (2006); IPCC (2007); Wentz *et al.* (2007)]. The CMAP's decreasing trend might have been caused by large variance at the beginning of the data period and the adjustment based on the gauges on atolls and low islands, as discussed in Sec. 2, and may not be a physical reality. These asymmetric and large positive values of CMAP in the 1980s have led to this decreasing trend. The consistent negative global precipitation anomalies in the last decade enhanced the decreasing trend. The consistently smaller variation after the mid-1990s might be due to the over-smoothing in the method of blending multiple sources of data, particularly more satellite data. It is to be determined whether these two apparent asymmetric distributions of the global precipitation in the 1980s and the last decade are a physical manifestation of the oceanic precipitation or nonphysical biases caused by the data aggregation methods.

IMF2 has a mean frequency around 2 cycles/year. This semi-annual cycle exists in all the four global precipitation datasets, and may be the manifestation of the winter and summer monsoon precipitation. The exact mechanism is still to be validated by climate models and tropical precipitation data.

The IMF mean periods and energies, and the Fourier spectra of the CMAP data are shown in Fig. 12. CMAP's IMF1 has energy 3.5 $[\text{mm/day}]^2$, which is much larger than that of IMF1 from REOF, GPCP, and MERG data. The energies have
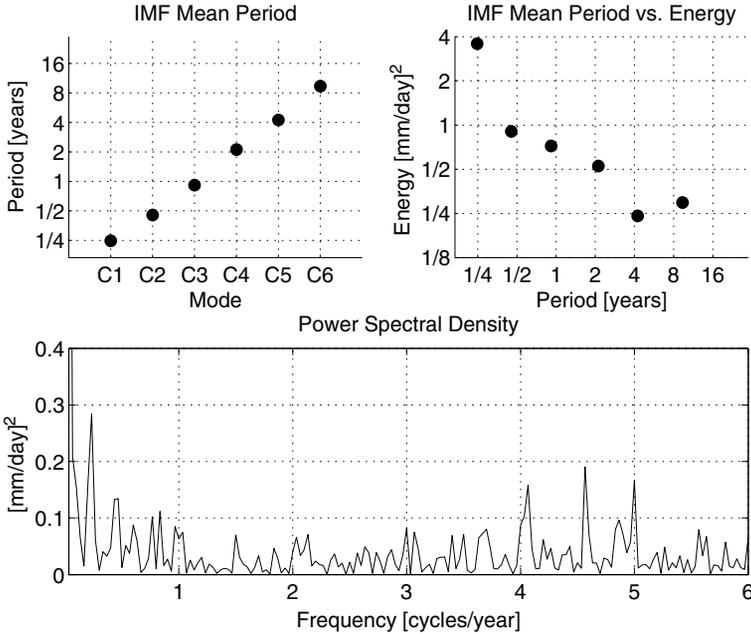
Fig. 12.   Mean period and energy of the CAMP IMFs, and the Fourier spectra of the CMAP data.

a sharp decay from IMF1 to IMF2, and turn to a much slower decay from IMF2 to IMF5, and then change to an increase from IMF5 to IMF6. This behavior is similar to that of GPCP. The total energy from all the six IMFs plus that from mode R is about 9 $[\text{mm/day}]^2$, which is more than twice of that of REOF and GPCP ($3\,[\text{mm/day}]^2$) and nine times of that of MERG ($1.0\,[\text{mm/day}]^2$). The unusually large energy, the decreasing trend, and obvious reduction of the CMAP variance since the mid 1990s are subject to further analysis. GPCP seems more consistent with the results of climate modeling and reconstruction. The similarities between REOF reconstruction and GPCP may be due to the use of GPCP EOF basis in the REOF reconstruction [Smith *et al.* (2010)].

The mean frequencies of the six IMFs are comparable to those of the first 5 IMFs of the other three datasets and follow the rule of period doubling. As discussed earlier, the 3 months, 6 months, 12 months, and other periods may be related to MJO, monsoon, annual cycle, El Nino, and PDO, respectively.

The Fourier spectra of the CMAP data behave similarly to those of GPCP. The major peak is near zero frequency, which corresponds to the nonlinear trend, i.e. mode R. Due to strong nonlinearity inherited from the station data, the pure harmonics do not exist in CMAP, and thus they are not reflected in the IMFs. So, three minor peaks between frequencies 4 and 5 cycles/year may be the result of a strong linear dispersion from IMF5 and 6.
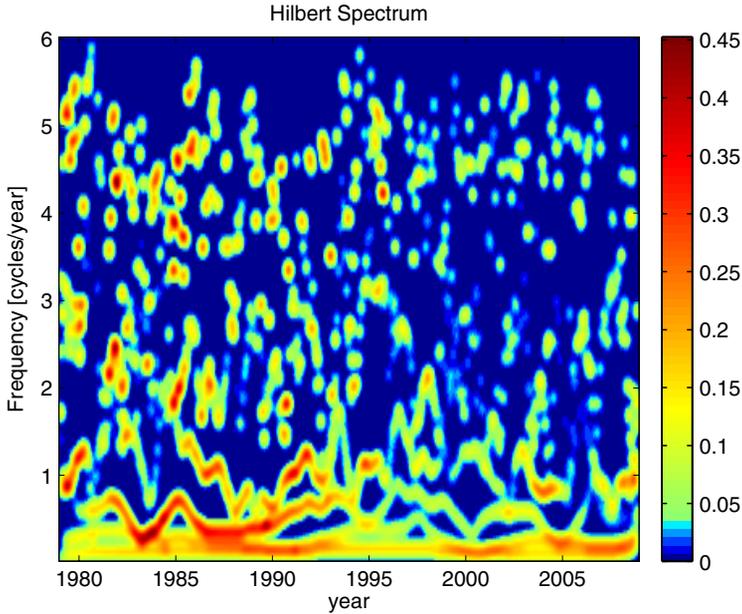
Fig. 13.   Hilbert spectra of the CMAP data (1979–2009).

The nonstationarity is clearly reflected in the Hilbert spectra (Fig. 13). Strong variations with large amplitudes are in the time–frequency domain of the late 1980s and early 1990s in time and 0 to 2 cycles/year and 4 to 6 cycles/year in frequency. The amplitudes have an apparent decrease in the last 15 years. The oscillation frequencies in IMF2 and 3 also decreased in the last 15 years. What has caused this red shift of the two IMFs is still to be found out.

4. Conclusions and Discussion

We have carried out an HHT analysis for the four time series of the monthly global average precipitation: MERG (1900–2008), REOF (1900–2008), GPCP (1979–2008), and CMAP (1979–2008). EMD sifting procedure has been used to decompose the precipitation time series into IMF modes, plus a trend, i.e., mode R, that carries the main signal of nonlinearity and nonstationarity. The mean period and energy of each IMF, Fourier spectra, and Hilbert spectra for each IMF have been calculated and displayed. MERG and REOF reconstructed data yielded 8 IMFs while the GPCP and CMAP observed data yielded 6 IMFs. The IMFs seem providing a way to display some physical signals carried the precipitation data, including MJO, monsoon, annual and ENSO. The Hilbert spectra can clearly display the time–frequency changes in a precipitation time series and may help extract critical information for improving the aggregation method for observed data and for improving climate model parameterizations. However, rigorous statistics inference

has not been made to verify the existence of the weak signals in the global precipitation time series. We conclude that HHT may be a useful tool for global precipitation data diagnostic studies and may provide new perspectives in addition to those of traditional Fourier spectral analysis and other time–frequency analyses.

Some simple statistics of the datasets have also been calculated, including mean, standard deviation, skewness, kurtosis, and linear trend. CMAP has the largest standard deviation due to its adjustment by atolls and low island stations' positive bias. These overestimate biases also cause the right skewness of CMAP data, while the other three data are skewed left. MERG has the smallest standard deviation, indicating that CCA analysis yields a smooth field compared to other methods. The kurtosis of each dataset is positive, indicating a more peaked distribution of global precipitation than a normal distribution. MERG and REOF have a positive linear trend for both periods of 1979–2008 and 1900–2008. The REOF trend is stronger. The GPCP linear trend is positive but weak, while the CMAP linear trend is negative and strong. Again, the strong negative CMAP trend may be caused by the overestimate from the adjustment in the 1980s by the stations on atolls and low islands. The inter-correlations of the four datasets are very small, implying that more effort needs to be made to synchronizing the observations by using climate dynamics when aggregating the observed data together by various interpolation or reconstruction methods [Mariotti and Arkin (2006)].

Although the data studied are the anomaly time series of global precipitation, the annual cycle signal still exists and is the third IMF mode in all the four datasets. CMAP has the largest energy. This is consistent with its largest standard deviation: 2.8 mm/day, four times of that of the MERG data for the same period 1979–2008. All the modes in each dataset approximately follow the rule of period doubling. The energy decays very fast from IMF1 to IMF2. The energy decay slows down from IMF2 in the GPCP and CMAP data. Thus, the annual cycle signal, as the third IMF mode, has some substantial physical signal.

There are many diagnostic and data uncertainty problems to be resolved. For example, it is understood that CMAP has an overestimate of the global precipitation in the 1980s mainly because of the adjustment based on the gauge data on atolls and low islands. It would be useful to study the observed precipitation data from the gauges on these atolls and low island stations and compare the gauge data with the data from satellites, such as TRMM, in the last 10 years. This may quantify the bias introduced to CMAP in the 1980s.

Another problem is the large difference between the trends of MERG and REOF despite the fact that both of them are EOF based reconstructions. MERG is a smoother dataset than ROEF. Then, what is the appropriate and physically meaningful energy for each IMF mode? A new reconstruction for REOF uses CCA for the oceanic precipitation and use EOFs for the global land precipitation. The merge of the two products becomes the new GHCN (Global Historical Climatology Network) based global precipitation dataset [Smith *et al.* (2011); Vose *et al.* (1998)]. We plan

to analyze this dataset with the data from different reanalysis and from different climate models [Compo *et al.* (2006); Onogi *et al.* (2007); Uppala *et al.* (2005)].

The cycles of the first four modes and their corresponding climate dynamic implications may be further tested and explained by considering the signals of the regional average precipitation data. The land and ocean precipitations can be analyzed separately to extract more clear precipitation signals corresponding to MJO, monsoon, annual cycle and ENSO.

Acknowledgments

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin, E. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.*, **4**: 1147–1167.

Arkin, P. A., Cullen, H. and Xie, P. (2007). Oceanic precipitation variability and the North Atlantic Oscillation. *Measuring Precipitation from Space: EURAINSAT and the Future*, eds. Levizzani, Bauer and Turk. Springer, pp. 37–48.

Compo, G. P., Whitaker, J. S. and Sardeshmukh, P. D. (2006). Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**: 175–190.

Dai, A., Fung, I. Y. and Del Genio, A. D. (1997). Surface observed global land precipitation variations during 1900–1988. *J. Climate*, **10**: 2943–2962.

Ebert, E., Manton, M. and Arkin, P. A. (1996). The GPCP intercomparison of rainfall estimates from satellite data for TOGA/COARE. *Bull. Amer. Meteor. Soc.*, **77**: 2875–2887.

Efthymiadis, D., New, M. and Washington, R. (2005). On the reconstruction of seasonal oceanic precipitation in the presatellite era. *J. Geophys. Res.*, **110**: D06103, doi:10.1029/2004JD005339.

Feldman, M. (1997). Non-linear free vibration identification via the Hilbert transform. *J. Sound and Vibration*, **208**: 475–489.

Held, I. M. and Soden, B. J. (2006). Robust response of the hydrological cycle to global warming. *J. Climate*, **19**: 5686–5699.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Q. Zheng, N. C. Yen, Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. Roy. Soc. Lond. A*, **454**: 903–995.

Huang, N. E. and Shen, S. S. P. (eds.), (2005). *Hilbert–Huang Transform and Its Applications.* World Scientific, Singapore, p. 360.

Huffman, G. J., Adler, R. F., Arkin, P. A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B. and Schneider, U. (1997). The global precipitation climatology project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**: 5–20.

IPCC (2007). Summary for Policymakers, In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller. Cambridge University Press, Cambridge, U.K. and New York, NY, USA.

Janowiak, J. E. and Xie, P. (1999). CAMS-OPI: A global satellite-raingauge merged product for real-time precipitation monitoring applications. *J. Climate*, **12**: 3335–3342.

Mariotti, A. and Arkin, P.A. (2006). The North Atlantic oscillation and oceanic precipitation variability. *Clim. Dyn.*, **28**: 35–51, DOI: 10.1007/s00382-006-0170-4.

Onogi, K. and co-authors (2007). The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**: 369–432.

Shen, S. S. P., Shu, T., Huang, N. E., Wu, Z., Karl, T. R. and Easterling, D. R. (2005). HHT analysis of the nonlinear and nonstationary annual cycle of daily surface air temperature data. In the book entitled, *Hilbert–Huang Transform and Its Applications*. eds. N. E. Huang and S. S. P. Shen, World Scientific, pp. 187–210.

Simpson, J., Kummerow, C., Tao, W. K. and Adler, R. F. (1996). On the tropical rainfall measuring mission (TRMM). *Meteorol. Atmos. Phys.*, **60**: 19–36.

Smith, T. M., Sapiano, M. R. P. and Arkin, P. A. (2008). Historical reconstruction of monthly oceanic precipitation (1900–2006), *J. Geophys. Res.*, doi:10.1029/2008JD009851.

Smith, T. M., Arkin, P. A., Sapiano, M. R. P. and Chang, C.-Y. (2010). Merged statistical analyses of historical monthly precipitation anomalies beginning 1900. *J. Climate*, **23**: 5755–5770.

Smith, T. M., Arkin, P. A. and Ren, L. (2011). Analyses of oceanic total cloudiness and some relationships with precipitation and SST beginning 1900. An unpublished manuscript.

Uppala, S. M. and coauthors (2005). The ERA-40 re-analysis, *Quart. J. R. Meteor. Soc.*, **131**: 2961–3012, doi:10.1256/qj.04.176.

Vose, R. S., Peterson, T. C. and Hulme, M. (1998). The global historical climatology network precipitation database: Version 2.0. In *Proceedings of the Ninth Symposium on Global Change Studies*, American Meteorological Society, Boston, Massachusetts.

Wentz, F. J. and Spencer, R. W. (1998). SSM/I rain retrievals within a unified all-weather ocean algorithm. *J. Atmos. Sci.*, **55**: 1613–1627.

Wentz, F. J., Ricciarudlli, L., Hilburn, K. and Mears, C. (2007). How much more rain will global warming bring? *Science*, **317**: 233–235.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. Roy. Soc. Lond. A*, **460**: 1597–1611.

Xie, P. and Arkin, P. A. (1996). Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, **9**: 840–858.

Xie, P. and Arkin, P. A. (1997). Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**: 2539–2558.

Xie, P. and Arkin, P. A. (1998). Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Climate*, **11**: 137–164.