An Intercomparison of the Spatiotemporal Variability of Satellite- and Ground-Based Cloud Datasets Using Spectral Analysis Techniques

JING LI

NASA Goddard Institute for Space Studies, and Department of Applied Physics and Applied Math, Columbia University, New York, New York

BARBARA E. CARLSON

NASA Goddard Institute for Space Studies, New York, New York

WILLIAM B. ROSSOW

Cooperative Remote Sensing Science and Technology Center, City College of New York, New York, New York

ANDREW A. LACIS

NASA Goddard Institute for Space Studies, New York, New York

YUANCHONG ZHANG

NASA Goddard Institute for Space Studies, and Department of Applied Physics and Applied Math, Columbia University, New York, New York

(Manuscript received 31 July 2014, in final form 26 January 2015)

ABSTRACT

Because of the importance of clouds in modulating Earth's energy budget, it is critical to understand their variability in space and time for climate and modeling studies. This study examines the consistency of the spatiotemporal variability of cloud amount (CA) and cloud-top pressure (CTP) represented by five 7-yr satellite datasets from the Global Energy and Water Cycle Experiment (GEWEX) cloud assessment project, and total cloud fraction observation from the Extended Edited Cloud Reports Archive (EECRA). Two spectral analysis techniques, namely combined maximum covariance analysis (CMCA) and combined principal component analysis (CPCA), are used to extract the dominant modes of variability from the combined datasets, and the resulting spatial patterns are compared in parallel. The results indicate that the datasets achieve overall excellent agreement on both seasonal and interannual scales of variability, with the correlations between the spatial patterns mostly above 0.6 and often above 0.8. For seasonal variability, the largest differences are found in the Northern Hemisphere high latitudes and near the South African coast for CA and in the Sahel region for CTP, where some differences in the phase and strength of the seasonal cycle are found. On interannual scales, global cloud variability is mostly associated with major climate modes, including El Niño–Southern Oscillation (ENSO), the Pacific decadal oscillation (PDO), and the Indian Ocean dipole mode (IODM), and the datasets also agree reasonably well. The good agreement across the datasets supports the conclusion that they are describing cloud variations with these climate modes.

1. Introduction

Clouds cover about 70% of the earth's surface and play an important role in the climate system by strongly altering the incoming solar and outgoing thermal radiation. Knowledge of cloud properties such as cloud coverage, cloud-top pressure, and microphysical properties is thus essential to accurately estimate the energy budget. Currently, most information on cloud properties is derived by satellite remote sensing techniques. Various sensors have been developed to detect clouds and to facilitate the retrieval of cloud properties from the

Corresponding author address: Jing Li, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025.
E-mail: jl2862@columbia.edu

DOI: 10.1175/JCLI-D-14-00537.1

spaceborne measurements. These sensors have different spectral designs and viewing geometries, and therefore different retrieval algorithms have been developed for the different satellite sensors according to their capabilities. These include multispectral imagers such as the Advanced Very High Resolution Radiometer (AVHRR; Foster and Heidinger 2013) and Moderate Resolution Spectroradiometer (MODIS; Platnick et al. 2003), hyperspectral infrared sounders such as the Atmospheric Infrared Sounder (AIRS; Aumann et al. 2003), and lidar instruments such as *Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations* (*CALIPSO*; Winker et al. 2009). Cloud observations can also be made from the surface by visual observations at weather stations, which also provide long-term cloud amount and type information worldwide.

However, cloud climatologies derived from different satellite sensors and the surface can exhibit considerable differences and systematic biases due to sources of uncertainties such as surface brightness, pixel resolution, cloud inhomogeneity for satellites (Li et al. 2004), and observing procedure changes for surface observation (Eastman and Warren 2013). Therefore, it is important to compare and evaluate the agreement and disagreement between different cloud observational datasets in order to make the best use of the data in climate and modeling studies, as well as to improve current retrieval algorithms. Many comparison studies have been made previously focusing on different datasets and different variables. The cloud assessment project initiated by the GEWEX Radiation Panel provides a very comprehensive, coordinated intercomparison of global cloud climatologies from more than 10 publicly available datasets (Stubenrauch et al. 2013), by evaluating the global mean and latitudinal distribution as well as the seasonality of several key cloud parameters. Other examples include the comparison of zonal mean high cloud amount between the International Satellite Cloud Climatology Project (ISCCP) and the Stratospheric Aerosol and Gas Experiment (SAGE) (Liao et al. 1995), comparison of ISCCP cloud parameters to the High-Resolution Infrared Sounder (HIRS) (Stubenrauch et al. 1999), comparison of cloud-top heights between MODIS and MISR using collocated measurements at specific sites (Naud et al. 2002), a more thorough evaluation of MODIS, MISR, and ISCCP cloud-top height and optical depth products by Marchand et al. (2010), comparison between cloud height retrievals by AIRS, MODIS, and *CALIPSO* for collocated scenes (Weisz et al. 2007), and the evaluation of AIRS climatology with *CALIPSO* and *CloudSat* (Stubenrauch et al. 2010). In these studies, the comparisons were generally focused on the analysis of mean spatial maps, zonally averaged time series, or collocated scenes. While these types of approaches are useful to assess data accuracy, they are unsuited for the evaluation of the spatial and temporal variability of cloud properties. In this study, we focus on the intercomparison of the spatiotemporal variability in different archived datasets.

Spectral analysis techniques have proven to be effective in identifying and isolating dominant variability both spatially and temporally through the decomposition of the data covariance matrix. These methods can also be used in the comparison of multiple datasets of the same physical parameter; by comparing the spatial and temporal modes extracted from different datasets, the agreements and disagreements in spatiotemporal variability can be examined in parallel. In earlier studies, Li et al. (2013, 2014a,b,c) applied four different spectral decomposition techniques to the analysis and comparison of aerosol datasets retrieved by multiple satellite sensors and a ground-based sun photometer network. These studies indicated that these spectral methods have the following advantages: 1) they reduce data dimensionality and limit the comparison to only the first few dominant modes of variability; 2) the coherency or lack thereof in both the spatial and temporal dimensions can be simultaneously examined and compared; 3) the methods are capable of extracting the common modes of variability from different datasets, which enables parallel comparison to better identify agreements and disagreements; and 4) the leading modes can often be explained by or related to physical phenomena, such as regular seasonal cycle, climate oscillations, or events.

With respect to cloud observations, Rossow et al. (1993) used empirical orthogonal function analysis to intercompare the space and time variability of global cloud distribution in ISCCP and three other datasets. Their study mainly focused on the examination of the seasonal cycle. In this study, we aim to extend the Rossow et al. (1993) analysis by examining interannual variability as well, and by incorporating more and newer satellite- and ground-based datasets. We focus on basic cloud parameters of cloud amount (CA) and cloud-top pressure (CTP), which provides cloud height information, and both seasonal and interannual variability are examined and compared. Section 2 introduces the five satellite datasets and one surface dataset used in the study. Section 3 describes the two spectral analysis methods used for space–time comparison, namely combined principal component analysis (CPCA) and combined maximum covariance analysis (CMCA). The results are presented in section 4 and the conclusions are presented in section 5.

2. Datasets

The primary source of satellite datasets is the GEWEX Cloud Assessment database (Stubenrauch et al. 2013).

The GEWEX Cloud Assessment project was initiated in 2005 by the GEWEX Radiation Panel to compare different global cloud data products from multispectral imagers, infrared sounders, and active lidars with the International Satellite Cloud Climatology Project (Rossow and Schiffer 1999). The data provided by GEWEX Cloud Assessment database are in gridded, monthly mean format and cover a wide range of satellite sensors with different instrumental characteristics and retrieval capabilities. In this study, we focus the comparison on two basic variables: cloud amount and cloud top pressure, from five satellite datasets, namely the ISCCP, the Advanced Very High Resolution Radiometer data processed by the Pathfinder Atmospheres Extended program (PATMOS-x), the Moderate Resolution Imaging Spectroradiometer from both the *Aqua* and *Terra* platforms, and the Atmospheric Infrared Sounder. Only daytime data are considered in this study.

The ISCCP project (Rossow and Schiffer 1999) uses a combination of multispectral imager observations onboard weather satellites and global retrievals of the atmospheric thermodynamic state to derive cloud properties. The ISCCP record spans 30 yr, is the only product that resolves diurnal variations, and has emerged as a baseline description of cloudiness in the earth's atmosphere (Rossow and Schiffer 1991, 1999). The only commonly available spectral bands are visible (VIS) and infrared (IR). ISCCP cloud detection is based on space and time variability of the IR and VIS radiances. IR radiances are used to derive cloud-top temperature, corrected for IR transparency using VIS radiances during daytime, and the CTP product is obtained using atmospheric temperature, water vapor profile, and ozone abundance from the TOVS product (Rossow and Schiffer 1991). For the purpose of spatiotemporal comparison, we prefer a long data record, and the ISCCP data provided by GEWEX Cloud database terminate in 2007. We also note that the version of ISCCP in the GEWEX Cloud database is not the official version but a special version prepared especially to facilitate comparisons and to illustrate certain aspects of the cloud retrievals. The GEWEX ISCCP data version only includes observations at four local times: 0300, 0900, 1500, and 2100. It also does not include the day–night algorithm correction implemented in the officially released ISCCP D2 data. Moreover, it was interpolated to $1^\circ \times 1^\circ$ resolution from the original 280-km equal area product. Because of the differences between the GEWEX version and the officially released ISCCP product, the ISCCP data used here are from the official ISCCP version D2 with all observation times from 0000 to 2100. The ISCCP CA and CTP data have been sorted to include only daytime observations at each location. Specifically, the reported UTC time in ISCCP D2 data is converted to local time for each grid box (local time = UTC + longitude × 12/180); then only data collected between 0600 and 1800 local time are selected as daytime observations.

The AVHRR instruments are multispectral imagers with five spectral channels onboard a series of National Oceanic and Atmospheric Administration (NOAA) satellites. Cloud detection by the NOAA PATMOS-x processing system is based on Bayesian classifiers derived from *CALIPSO* (Heidinger et al. 2012) and the retrieval is based on an optimal estimation approach (Heidinger and Pavolonis 2009). Cloud-top temperature is first retrieved using the two IR channels, and the CTP information is derived using cloud-top temperature and interpolating within the NCEP numerical weather prediction (NWP) temperature profile (Heidinger and Pavolonis 2009). We average observations made at two daytime overpasses: 0730 and 1330 into one monthly mean dataset. Temporal biases in diurnal variability due to orbital drift have been noted in the PATMOS-x cloud climatology by Foster and Heidinger (2013). However these biases are not significant due to the relatively short record used here and the less drifting in the newer satellites.

MODIS is also a multispectral imager with 36 channels. It was launched onboard the EOS-*Terra* satellite in December 1999 and later on EOS-*Aqua* in May 2002. Here we use Collection 5 cloud products from both platforms retrieved by the MODIS science team. Cloud amount is determined using spectral testing (Ackerman et al. 1998; Frey et al. 2008; Menzel et al. 2008) and CTP is retrieved using CO_2 slicing (Menzel et al. 2008). The equatorial overpass time is 1030 for *Terra* and 1330 for *Aqua*. The two MODIS data are combined in this study to avoid over contribution from the MODIS sensor and algorithm in the analysis.

The AIRS instrument is an IR sounder launched also onboard the EOS-*Aqua* platform in 2002. It has three IR spectral bands with high spectral resolution. The AIRS-LMD cloud retrieval product in the GEWEX archive uses spectral emissivity coherency for cloud detection and a weighted χ^2 method to determine the CTP (Stubenrauch et al. 2010; Guignard et al. 2012). Only one daytime measurement at 1330 is available for AIRS.

Except for the ISCCP data, all other satellite datasets obtained from the GEWEX Cloud Assessment database are gridded onto the same $1^\circ \times 1^\circ$ spatial resolution. The time period, January 2003 to December 2009 used here, corresponds to the longest period of overlapping data.

In addition to satellite observations, we also incorporate one ground-based cloud dataset in order to yield a more complete comparison. The surface data are provided by the Extended Edited Cloud Report Archive (EECRA; Hahn and Warren 1999; Eastman and Warren 2013). The

EECRA integrates surface-based cloud observations from worldwide weather stations and covers most of the land area. Ship observations are also available for some ocean regions (Hahn et al. 2009). Here, we use monthly mean daytime EECRA cloud amount data from both land and ocean. The land data are provided for each individual station while the ocean data are gridded into $10° \times 10°$ grids. The temporal coverage of land and ocean data are also slightly different. For land, we use data from January 2003 to December 2009, while for ocean the data are only updated to December 2008. Moreover, as a complete time series is essential for constructing the temporal variance matrix used in the spectral analysis, we select 1376 land stations and 157 ocean grid boxes that have no missing monthly data points for the 2003–09 and 2003–08 periods, respectively. Overall, the spatial coverage of the stations is excellent for Europe, Asia, and the Northern Hemisphere oceans. The distribution of the stations is relatively sparse for Africa and South America. The coverage for North America is also low for the study period due to the switching to automatic weather stations, whose reports are not compatible with human observations (Dai et al. 2006).

While satellite sensors are able to detect clouds during both day and night, and night observations are also attempted by EECRA, the data volume and accuracy are reduced during night for EECRA due to illumination requirements. As a result, to avoid possible effects of diurnal cloud changes, we use daytime only measurements for all satellites and EECRA. The EECRA daytime measurements are defined as those made between 0600 and 1800 local time (Eastman and Warren 2013).

3. Methods

Spectral decomposition techniques are very effective and efficient in reducing data dimensionality and in identifying the major modes of variability in space and time. Previously we applied and developed several spectral methods to intercompare different aerosol observations (Li et al. 2013, 2014a,b,c). In this study, we use combined principal component (CPCA) analysis for CTP comparison and combined maximum covariance analysis (CMCA) for CA comparison. Below we briefly describe these two techniques, while their detailed mathematical description can be found in Li et al. (2014b) and Li et al. (2014c), respectively.

The CPCA is a modification of the traditional principal component analysis (PCA) by decomposing the covariance matrix of the different data fields combined. In this way, the leading modes will maximize the variance explained by the sum of the elements in the combined field (Bretherton et al. 1992). Specifically, a combined data matrix is constructed by

$$\mathbf{D} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \tag{1}$$

where $\mathbf{X}_1$ to $\mathbf{X}_n$ are different datasets (e.g., different satellite observations of CTP), whose columns are the number of spatial locations and rows are observations at each location. Then PCA decomposition is performed on the covariance matrix of this joint field to obtain the orthogonal temporal modes (time series). The spatial modes are found by projecting the temporal modes back to each individual data matrix. As a result, each CPCA mode comprises an individual spatial pattern for each dataset and a shared time series that represents how these spatial patterns evolve over time. Because usually the first few leading modes explain the bulk of the variance in all data fields, and the spatial modes of different datasets correspond to the single time series, the comparison is effectively reduced to the comparison of the spatial patterns from only the first few dominant modes. An important note on CPCA is that by combining different datasets as shown in Eq. (1), we are assuming equal weights for the fields being combined. In other words, different fields should in theory have the same order of magnitude in spatial and temporal variability. In this study, this prerequisite is satisfied, as the different datasets are measurements of the same physical parameter (CTP). Although ISCCP has different spatial mapping than the other datasets, this problem is reconciled by balancing the information density using the area of each grid as the weight; that is, before combining the datasets, the value of each grid box is weighted by the area of that grid box. Nonetheless, caution must be taken when using CPCA for datasets of different parameters.

While CPCA successfully achieves parallel comparison across the satellite datasets, it is not suitable for scattered ground-based observations. This is because unlike regularly gridded data, the spatial representativeness (or the weight of each station) is difficult to determine, and blindly combining satellite and ground observations may result in serious biases in the decomposition results. Therefore, for the CA comparison, for which data are available from both ground-based and satellite measurements, we use the CMCA method developed by Li et al. (2014c). The CMCA is an extension of CPCA and another spectral method—maximum covariance analysis (MCA, also known as singular value decomposition analysis)—and accommodates both multisensor satellite

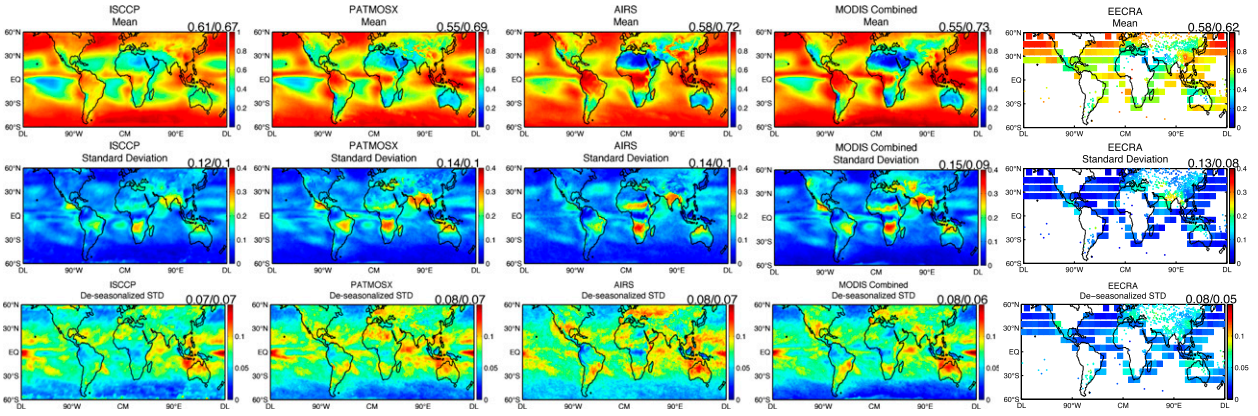

FIG. 1. (top) Multiyear averaged global mean CA (color scale runs from 0 to 1 in 0.2 increments), (middle) CA standard deviation fields (color scale runs from 0 to 0.4 in 0.1 increments), and (bottom) deseasonalized CA standard deviation fields (color scale runs from 0 to 0.1 in 0.05 increments) of (left to right) the five CA datasets: ISCCP, PATMOS-x, AIRS, MODIS Combined, and EECRA. The numbers on the top-right corner of each panel indicate land/ocean averages of the spatial map. All datasets agree well in the spatial patterns and global means.

data and ground-based observation. The MCA method is used to find coupled modes between two datasets through the decomposition of their cross-covariance matrix (Bretherton et al. 1992; Li et al. 2014a). It has no strict assumption and can be basically applied to any two data fields. The leading modes have the property of maximizing the covariance in the two data fields. And because the covariance is the product of the correlation of the two data fields and the variance of each individual data field, the dominant modes represent the variability that both have the highest correlation and account for the most variance of each individual dataset. Therefore, agreements as well as differences between datasets in terms of spatiotemporal variability can be found by examining the correlation between the temporal modes and comparing the spatial patterns. In CMCA, MCA is performed between the combined satellite field constructed by Eq. (1) and the ground observation field, so that the dominant modes of variability of all different datasets can be extracted and compared in parallel. Note that although the EECRA ocean data are gridded, the distribution of the grid boxes is sparse when compared with the resolution of the satellite datasets. We thus also use CMCA for the analysis of ocean CA data instead of CPCA.

The construction of the data covariance/cross-covariance matrix requires the removal of the temporal mean at each spatial location. Because global cloud variability exhibits distinct seasonal cycles, the remaining data matrix will be dominated by seasonal variability, which allows the comparison of different datasets on annual or seasonal time scale. Moreover, we also construct a deseasonalized dataset by removing the multiyear averaged seasonal cycle from each grid box to further examine interannual variability. Finally, because of the absence of daytime measurements during polar nights, the spatial domain of this analysis is restricted to $\pm 60°$ latitude.

4. Results

We present the analysis and comparison for CA followed by CTP. For each variable, we first briefly compare the time series mean (which is removed in the subsequent spectral analysis), standard deviation (mainly representing seasonal variability), and deseasonalized standard deviation (representing interannual variability) fields. We then present and discuss CMCA results for CA, separately for land and ocean due to different EECRA data formats, and CPCA results for CTP.

a. CA

CA, or cloud fraction, is the fraction of pixels that contain clouds with respect to the total number of valid pixels. It is the first-order indicator of global cloud distribution. Figure 1 shows the time series mean, standard deviation, and deseasonalized standard deviation fields for the five different datasets. The spatial patterns of all three rows in Fig. 1 agree very well. The two numbers on the upper right corner of each map indicate global mean values for land/ocean, weighted by the area of each grid box for the satellite datasets. Globally, on average CA is high in the intertropical convergence zone (ITCZ) and midlatitude storm tracks, and low for the arid regions of the subtropics. The mean CAs for the different datasets are very close, ranging from 0.55 to 0.61 over land, and from 0.62 to 0.73 over ocean, which is consistent with Stubenrauch et al. (2013) that global mean CA over ocean is approximately 0.1–0.15 higher than that over

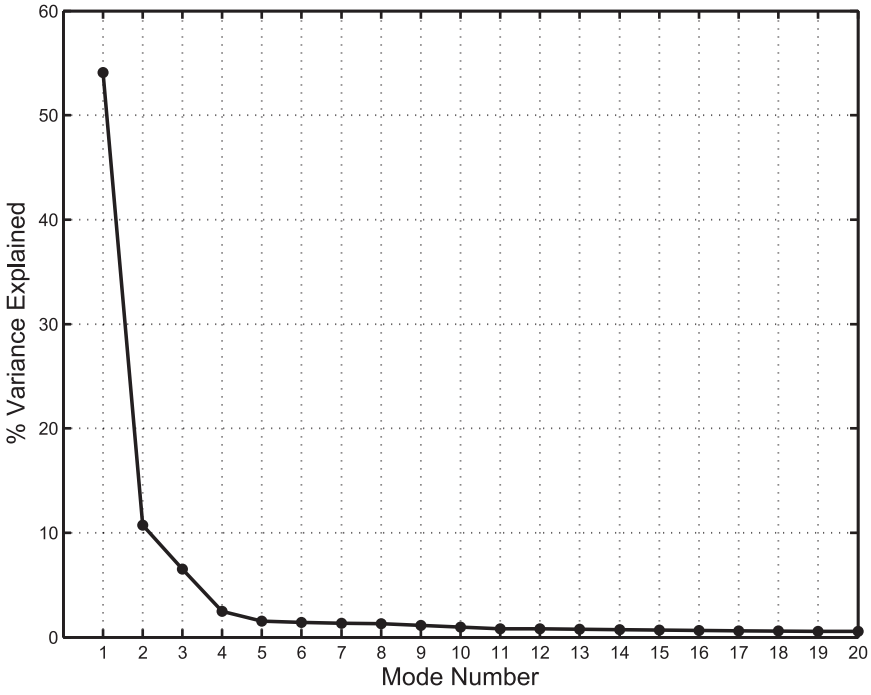

FIG. 2. Variances explained by the first 20 CMCA modes of CA full dataset with seasonal cycle left in over land.

land. The EECRA ocean average (0.62) is the lowest because the data are not complete for many high CA regimes including the tropics and Southern Hemisphere midlatitudes. For the standard deviation fields, the highest signals are found for several land regions including South Africa, South America, India, and northern Australia, indicating the strongest seasonality over these regions, while the seasonal variability is generally lower over the oceans. The global mean standard deviations for the individual datasets are also reasonably close, with land values from 0.12 to 0.15 and ocean values 0.08 to 0.10. The deseasonalized standard deviation fields also agree quite well across the datasets. The highest variability is found in the Niño-4 region of the east Pacific and the South Pacific convergence zone (SPCZ), while the midlatitude oceans and tropical lands have the lowest variability. Globally averaged interannual CA variability is around 0.08 over land and 0.07 over ocean.

It is also worth noting that some differences still exist in Fig. 1. For example, ISCCP has lower CA over much of South America. ISCCP and AIRS have lower seasonal variability over Europe, where ISCCP also has lower interannual variability.

1) CMCA ANALYSIS OVER LAND

This section presents the CMCA decomposition results for land data and compares the modes of variability for the different datasets. The analysis is first performed on the full dataset with the seasonal cycles left in, in order to examine seasonal variability. To determine the number of significant modes, we first examine the variance explained by the first 20 modes, which is shown in Fig. 2. Mode 1 is clearly dominant, explaining nearly 60% of the total variance. Modes 2 and 3 also stand out by accounting for significantly higher variance than the following modes, while the variance falls below 1% for mode 4. We thus consider the first three modes to be the leading modes.

The spatial patterns and time series for the first three CMCA modes are presented in Fig. 3. The black curves in the bottom row of Fig. 3 show the time series (PCs) associated with the spatial patterns for each satellite dataset shown in the first five rows, while the red curves indicate the temporal evolution of the EECRA spatial modes shown in the sixth row. The PCs of the satellite and EECRA data are highly correlated for all three modes, with correlations (indicated by the R value) well above 0.9, indicating coherent variability. PCs 1 and 2 exhibit strong summer–winter and spring–fall seasonal cycles, respectively, while PC 3 displays a semiannual variability (with summer–winter and spring–fall peaks). The seasonality in modes 1 and 2 agrees well with the earlier results (Rossow et al. 1993). The spatial maps associated with the individual satellite datasets also appear to agree well. Although the EECRA data miss the signal in several regions due to insufficient coverage,

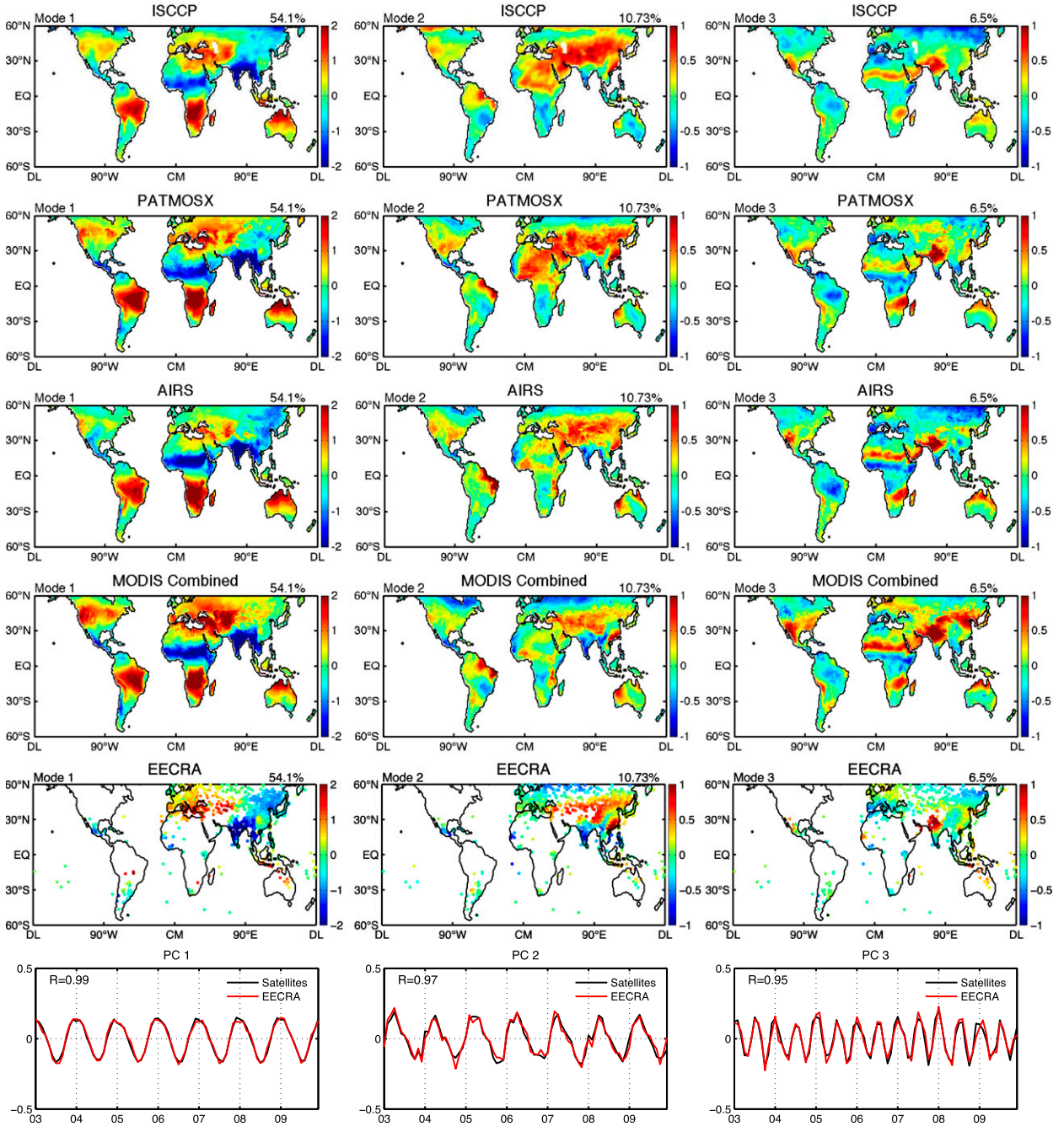

FIG. 3. (left to right) The first three CMCA modes for the full CA dataset over land. The top five rows show the spatial patterns of each dataset—(top to bottom) ISCCP, PATMOS-x, AIRS, MODIS Combined, and EECRA—with one column for each mode, and the bottom row shows the PC time series. The numbers in the top-right corner of each spatial mode indicate the percentage of variance explained by that mode. The *R* value on the upper left of each PC panel indicates the correlation coefficient between the satellite and EECRA PC. For the spatial modes, red colors indicate positive anomalies, which vary in phase with the corresponding PC; while the blue colors are negative anomalies that vary out of phase with the PC. The PC correlations are quite high and the spatial patterns all agree very well, which suggests that different datasets are highly consistent in representing the major seasonal variability of CA.

where the data are complete they capture the dominant variability shown in the satellite spatial maps. To examine the agreement between the spatial patterns more quantitatively, we calculated the correlation coefficients between the spatial maps for each mode and list the results in Table 1. All correlations in Table 1 are calculated with respect to the MODIS spatial modes, which temporarily serve as the reference. All data except for

TABLE 1. Land full dataset for CA, showing the correlation coefficients between the spatial patterns of modes 1, 2, and 3 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOS-x	AIRS	EECRA
Mode 1	0.85	0.91	0.83	0.85
Mode 2	0.61	0.77	0.81	0.92
Mode 3	0.73	0.79	0.75	0.71

ISCCP have been downscaled from 1° × 1° to 2.5° × 2.5°. The downscaling retains all the variability of the original patterns (figures not shown). From Table 1, we can see that the correlation coefficients are reasonably high. All correlations are above 0.6, with many cases above 0.8.

The high correlations in both the spatial patterns and temporal variability are very encouraging. Nonetheless, we note some differences. Most of the disagreements appear to come from high latitudes in the Northern Hemisphere. For example, in mode 1, PATMOS-x and MODIS show strong positive signals over Europe and North America, while ISCCP and AIRS only have weak or even negative signals over these two regions. The EECRA signal for Europe better agrees with PATMOS-x and MODIS. Also, for the East Asia region (including Siberia), ISCCP, AIRS, and EECRA exhibit extensive negative anomalies, PATMOS-x has only a weak negative projection over this region, but MODIS indicates a positive anomaly. Since mode 1 is associated with a regular summer–winter seasonal cycle, and in winter these latitudes often have extensive snow/ice cover, these differences suggest possible issues with cloud detection above the snow or ice surface. Therefore, we further compare the seasonal cycles for NH high latitudes to identify the source of the disagreements. Because EECRA data do not cover most of North America, we separated the NH high latitudes into two subregions: between 40° and 60°N in Europe and Asia and between 40° to 60°N in North America. Figure 4 shows the multiyear averaged CA seasonal cycles averaged within these two subregions. From Figs. 4a and 4b, it is clearly seen that both MODIS and PATMOS-x data exhibit a strong seasonal cycle for the two NH high-latitude regions, with minimum CA in July and August and maximum CA in December and January. As this seasonality is out of phase with PC 1, which has summer maximum and winter minimum, NH high latitudes appear as negative signals in mode 1 for these two datasets. However, the seasonality for the other three datasets, ISCCP, AIRS, and EECRA, is much weaker. The phase of the seasonal cycles also disagrees with that of MODIS and PATMOS-x. For ISCCP, AIRS, and EECRA, CA is slightly higher during the spring and fall months from April to June and from October to November, respectively, but lower for

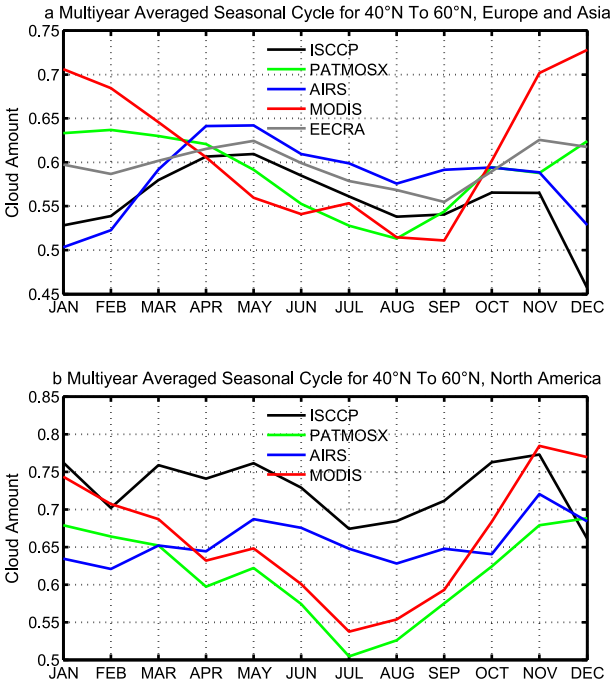

FIG. 4. Multiyear averaged seasonal cycle for the NH high latitudes for the five datasets: (a) Europe and (b) Asia and North America. These regions appear to have the largest discrepancy among the CMCA mode 1. From this figure, we are able to see that the disagreements are due to the different phase and strength of the seasonal cycle for PATMOS-x and MODIS from the other three datasets.

the summer months of July and August. This seasonality is also different from PC 1, which results in a weak projection of PC 1 on the spatial maps of these three datasets. Several factors may account for the discrepancies in satellite retrievals, including bright surfaces that often appear in high latitudes during the winter months. The discrepancy in the CA seasonal cycle for NH high latitudes suggests that large uncertainties still exist in cloud detection for this region, and improvements are likely needed in the satellite algorithms over bright surfaces to yield a consistent result. Although the two MODIS data are combined here for better diurnal resolution, MODIS-*Aqua* and -*Terra* can also have differences due to factors such as cloud diurnal cycle, sensor calibration, and sensitivity, which will be investigated in the future.

Next, we examine interannual variability through CMCA analysis of the deseasonalized dataset. The variance explained by the first 20 CMCA modes is shown in Fig. 5. The variance associated with the deseasonalized modes is in general smaller than those of the full dataset, because seasonal variability accounts for the bulk of the variance. Based on the sharp drop of variance from mode 4 to mode 5 shown by Fig. 5, we consider the first four modes to be significant. Figure 6 presents the first four modes and the corresponding time

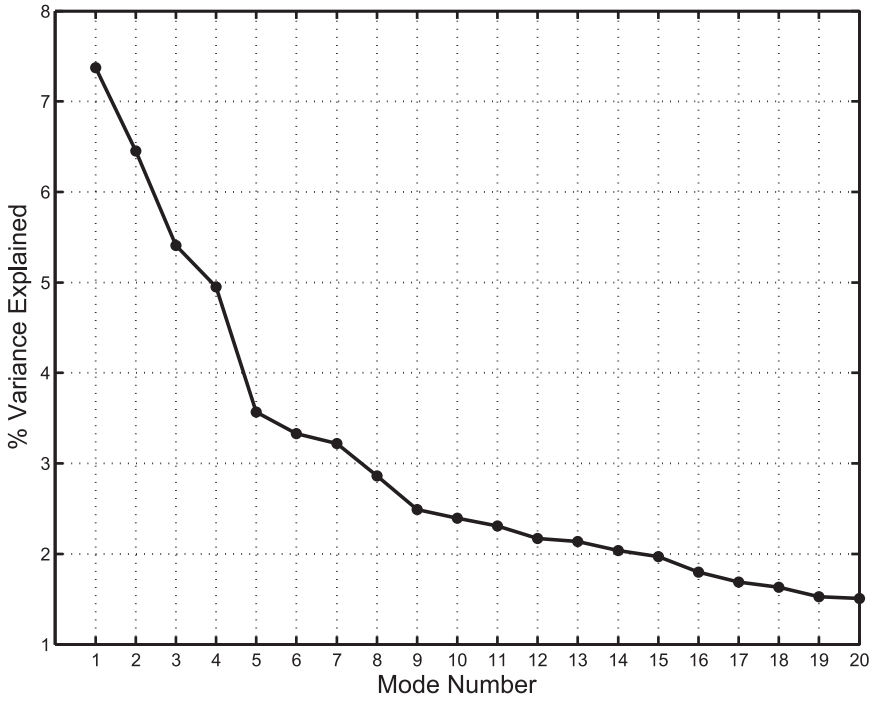

FIG. 5. Variances explained by the first 20 CMCA modes of deseasonalized CA dataset over land.

series. The PC time series are much noisier and reflect perturbations of the CA apart from the regular seasonal cycle. The PCs of satellite and EECRA are also highly correlated, with correlations well above 0.8. Moreover, although the deseasonalized spatial patterns appear noisier than those of the full dataset (Fig. 3), the different datasets agree remarkably well. In fact, the correlations between the deseasonalized spatial modes are higher than those for the full data. Table 2 lists the correlation coefficients with respect to MODIS, which show that all correlations are close to or above 0.8. This result is again very encouraging in that while different datasets may differ on seasonal scales for certain regions, they are highly consistent on interannual time scales. It is also interesting to note that modes 3 and 4 in Fig. 6 capture an ENSO pattern. We correlate these two time series with the multivariate ENSO index (MEI; Wolter and Timlin 2011) provided by the NOAA Earth System Research Laboratory, and find that the correlations are 0.38 and 0.52, respectively, for PC 3 and PC 4. Also their spatial patterns exhibit positive anomalies over several regions, especially high latitudes in Asia and Europe, suggestive of ENSO teleconnection from tropics to high latitudes.

2) CMCA ANALYSIS OVER OCEAN

In this section, we present and discuss CMCA results for ocean data. Similar to land, we first examine seasonal variability. The first three CMCA modes of full (seasonal cycle included) ocean data capture the bulk of the variance (~60%, Fig. 7). Figure 8 shows the spatial patterns and time series of modes 1, 2, and 3 for the ocean data. Likewise, PCs 1 and 2 represent winter–summer and spring–fall seasonality respectively, while PC 3 exhibits a semiannual seasonal cycle. High correlations are also found between the PCs of satellite and EECRA data over ocean. Qualitative comparison between the five spatial patterns of each mode indicates excellent agreement. The correlation coefficients between the spatial patterns are calculated against MODIS and the values are given in Table 3. The correlations are even higher for ocean results than for land except for AIRS. For ISCCP and PATMOS-x, the correlations of all three modes with MODIS are close to or above 0.9. AIRS has comparatively lower correlations, especially for mode 2. Going back to Fig. 8, we find that this low correlation is mainly attributable to two regions, 1) the north Mexican coast and 2) the subtropical Atlantic close to the South African coast (regions indicated by the black rectangles on the mode 2 spatial maps). Removing these two regions, the correlation between AIRS and *Aqua* MODIS mode 2 increases from 0.51 to 0.76. For region 1, only the AIRS spatial map shows positive anomalies, whereas weak negative anomalies are found in the other datasets. For region 2, all of the ISCCP, PATMOS-x, MODIS, and EECRA spatial maps exhibit a strong positive anomaly, while the AIRS mode has negative anomalies with a narrower spatial extent.

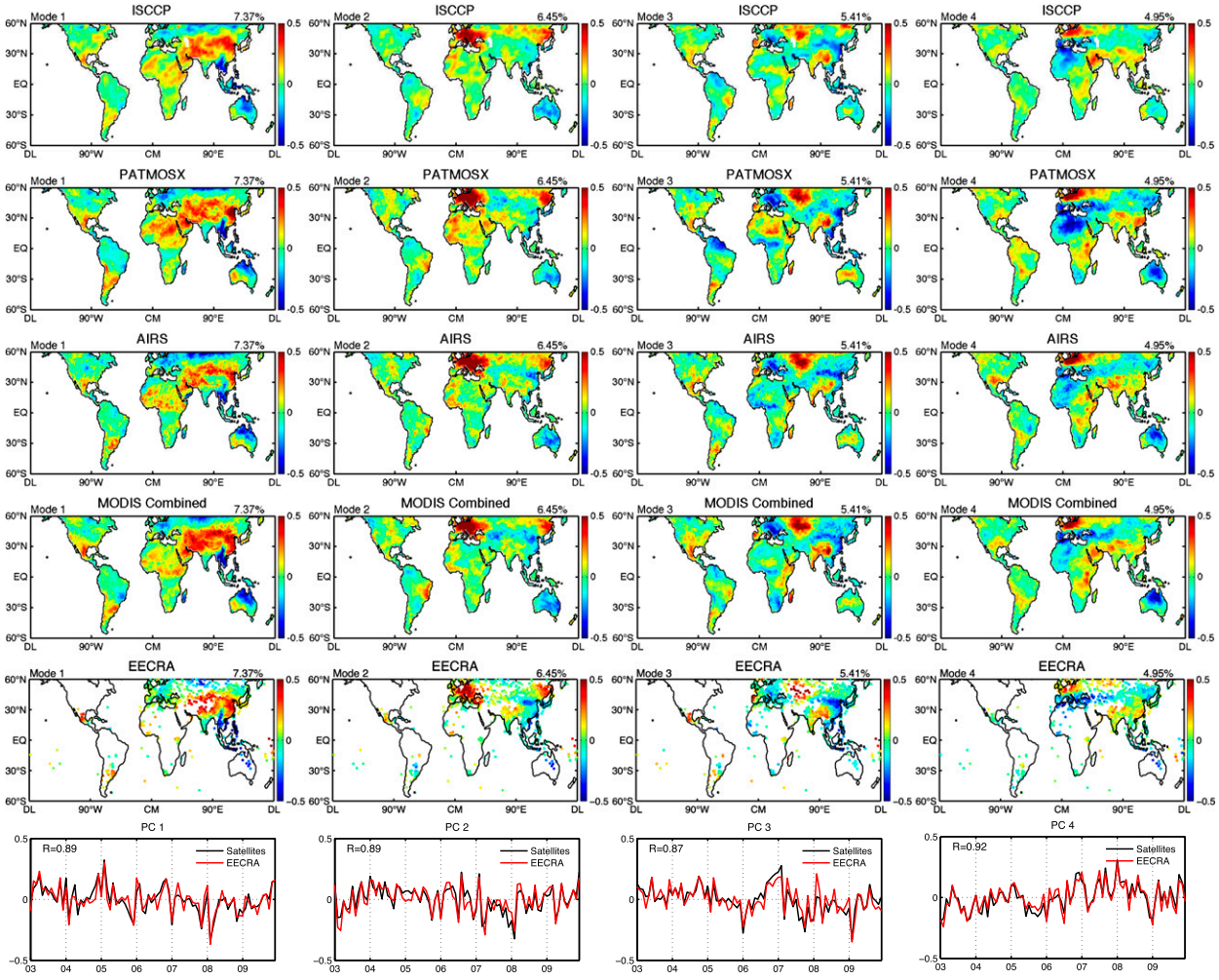

FIG. 6. As in Fig. 3, but for the first four CMCA modes for deseasonalized CA data land. The deseasonalized modes all agree well in both spatial and temporal variability, indicating that the different datasets are also highly consistent in characterizing interannual variability of CA.

Since these two regions are primarily dominated by low-level marine stratus clouds, this suggests problems in the seasonality of low clouds represented by the AIRS data. We therefore further compare the mean seasonal cycle averaged over the two black rectangles in order to identify the cause of the disagreements. The results are shown in Fig. 9. From Fig. 9a, it is clearly seen that the AIRS seasonal cycle is completely out of phase with the other datasets. It displays a June and July minimum, whereas all of the other datasets reach a maximum during these two months. For region 2 (Fig. 9b), the mean seasonal cycles of all datasets except AIRS are very consistent, with the highest CA during the fall months (September–November). For AIRS, interestingly, the first half of the seasonal cycle (i.e., from January to August) agrees well with the other datasets. However, AIRS CA drops during the second half of the year from August to December, without reaching a maximum value in the fall months. This different behavior of AIRS data leads to a phase shift of its seasonal cycle with respect to the other datasets as well as PC 2, which then results in a negative projection of PC 2 on AIRS map for the South African coast region. AIRS is known to have issues with low-level cloud detection

TABLE 2. Deseasonalized land dataset for CA, showing the correlation coefficients between the spatial patterns of modes 1 to 4 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOS-x	AIRS	EECRA
Mode 1	0.87	0.88	0.83	0.90
Mode 2	0.85	0.89	0.85	0.91
Mode 3	0.76	0.81	0.82	0.82
Mode 4	0.80	0.80	0.84	0.89

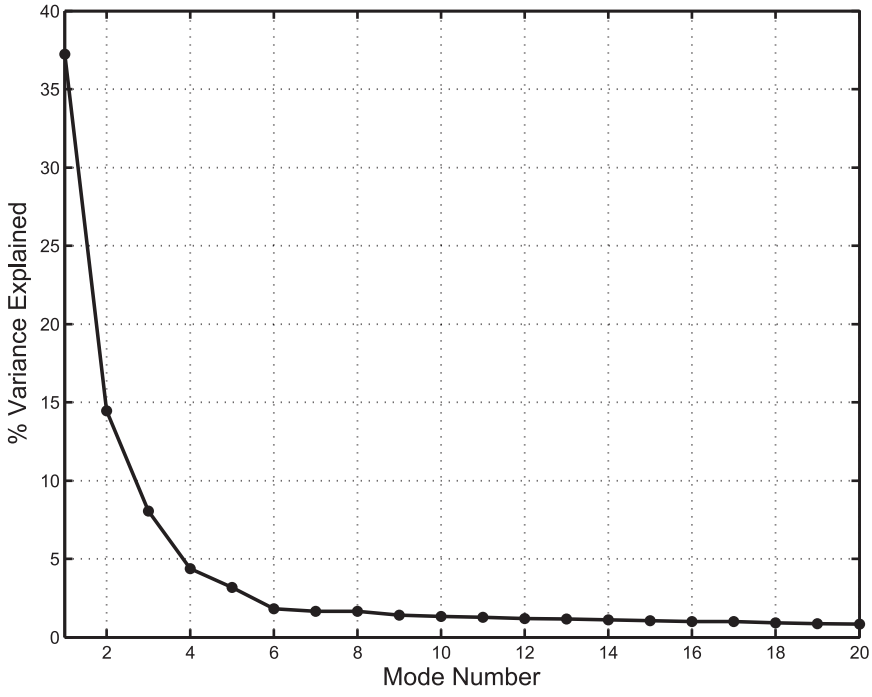

FIG. 7. Variances explained by the first 20 CMCA modes of CA full dataset with seasonal cycle left in over ocean.

and the results here confirm this drawback with IR-only measurements.

The interannual variability for ocean CA is also analyzed and compared. Figure 10 shows the variance explained by the first 20 CMCA modes. Based on the large decrease of the variance from mode 4 to mode 5 and the small differences in the variance explained by the higher-order modes, we select the first four modes for further examination, shown in Fig. 11. The spatial patterns for all modes (Fig. 11) also agree well. And the correlations between the time series are well above 0.9. Mode 1 of all datasets exhibit an ENSO-like pattern, with positive anomalies over the central Pacific and negative anomalies over the west Pacific warm pool. PC 1 also appears to reflect ENSO variability, and the peak in late 2006 and drop in early 2008 could correspond to the El Niño and La Niña events of these two years, respectively. To confirm this, we again correlate the PC 1 time series with MEI. It is found that MEI is significantly correlated with PC 1 of both satellite and EECRA data, with correlation coefficients of 0.81 and 0.69, respectively. The latter is lower mainly because EECRA data do not cover the central Pacific Niño-4 region and thus do not reflect the positive anomalies. No significant correlation is found between MEI and the PC of the other modes. As mode 1 explains the most variance of the deseasonalized data, its correlation with ENSO suggests that ENSO is the primary factor affecting the interannual variability of global cloud amount over oceans. Similar ENSO-related cloud pattern changes were also reported by Park and Leovy (2004) and Zhu et al. (2007). The correlations between the spatial patterns of modes 1 to 4 are also remarkably high, as indicated in Table 4. Mode 4 of PATMOS-x has a slightly higher variability overall compared to the other five datasets. This phenomenon is consistent with the overall higher variability found in the PATMOS-x deseasonalized standard deviation field (Fig. 1). Nonetheless, the distribution of the signals of PATMOS-x mode 4 is still in agreement with the other datasets.

To briefly summarize the results for CA intercomparison, by using the CMCA technique, the major spatial and temporal variability on both seasonal and interannual time scale are successfully extracted. The five different datasets agree remarkably well in the spatiotemporal variability for both land and ocean. The correlations of the spatial patterns are mostly above 0.7 and often reach 0.9. For land data, differences mainly lie in the NH high latitudes, where MODIS and PATMOS-x have a strong summer–winter seasonal cycle. However, the seasonality in the other three datasets is comparably weaker, exhibiting a phase difference. For ocean, AIRS seasonal cycle is out of phase from the other four datasets over the subtropical Atlantic region near South African coast. The interannual variability is highly coherent across the datasets for both land and ocean.

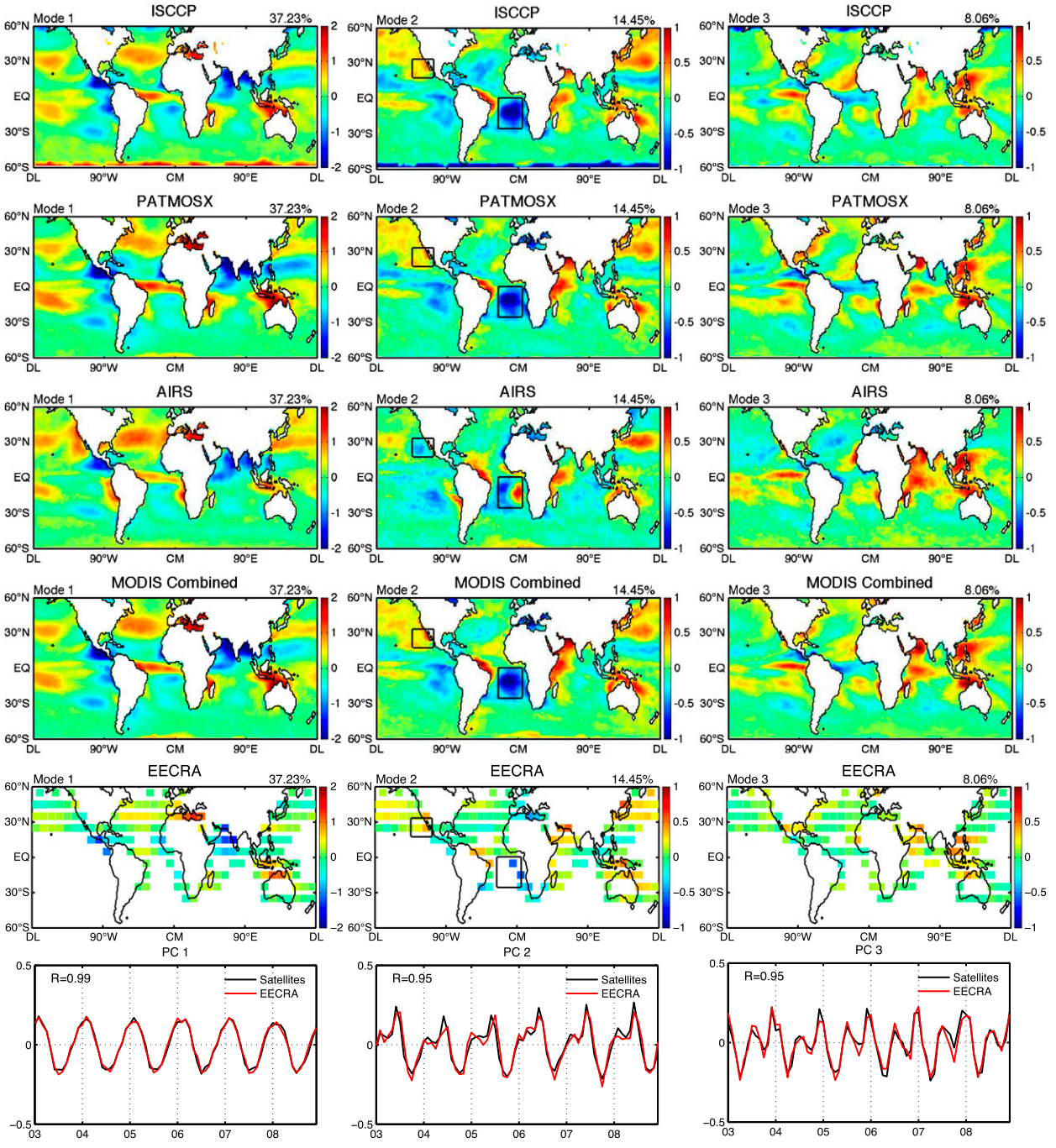

FIG. 8. As in Fig. 3, but over ocean. PCs 1 and 2 exhibit a strong seasonal cycle while PC 3 represents semiannual variability. The correlations for the PCs are very high. The spatial modes are also highly coherent. Two major regions with disagreements are the northern Mexican coast and South African coast, which are marked by the black rectangles on mode 2 spatial maps.

b. CTP

CTP indicates cloud height, and as clouds at different altitudes have different radiative effects, accurate information of CTP is important in studying the earth's energy budget. CTP is also needed in the retrieval of many other atmospheric and surface properties from the space. Figure 12 shows the averaged global CTP distribution, CTP standard deviation, and deseasonalized standard deviation fields for the four satellite datasets. CTP is not available from surface retrievals. The global CTP distribution appears consistent across the four datasets,

TABLE 3. Ocean full dataset for CA, showing the correlation coefficients between the spatial patterns of modes 1, 2, and 3 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOS-x	AIRS	EECRA
Mode 1	0.83	0.96	0.75	0.93
Mode 2	0.68	0.94	0.51	0.87
Mode 3	0.76	0.93	0.75	0.80

especially between AIRS and MODIS. CTP distribution shows well-known cloud characteristics, including low CTP (high clouds) in the ITCZ, west Pacific, and some desert regions and high CTP (low clouds) in the central subtropical gyres and southern Indian Ocean. Overall CTP is lower over land than over ocean. The regions with the highest seasonality include the subtropical regions in both hemispheres, inferred from the standard deviation fields (middle row of Fig. 12), while the interannual variability is largest over tropical oceans, especially the west Pacific, subtropical oceans (especially the Indian Ocean), the SPCZ, and the Gulf of Mexico, and some land regions including India, northern Australia, and the eastern United States, as shown by the deseasonalized standard deviation fields (bottom row of Fig. 12). The agreement between the spatial patterns and global mean values for the satellite datasets is less good for CTP than for CA. CTP retrieval by passive sensors usually involves external inputs such as atmospheric temperature profiles, and differences in these ancillary datasets may result in uncertainties in the retrieved CTP (e.g., Jin and Rossow 1997; Stubenrauch et al. 1999, 2013; Wang et al. 2000). In addition, different sensors may have different sensitivities to the detection of high, thin cirrus clouds (King et al. 2013; Rossow and Schiffer 1991; Stubenrauch et al. 2013). The ISCCP data appear to have the most differences in both mean and standard deviation fields compared to the other datasets. The ISCCP annual mean CTP appears lower over the oceans but higher over the Sahara desert compared to all the other datasets. It also has lower variability for most tropical and subtropical regions, in particular the tropical Pacific and Atlantic, South America, and South Africa. Biases in the ISCCP retrieval toward midlevels and its lower sensitivity to thin cirrus are likely to affect its CTP variability.

The CPCA method is applied to the combined multisensor CTP fields in order to extract and compare the spatiotemporal variability. As mentioned in section 2, each CPCA mode will have one shared time series and a spatial pattern for each individual dataset. The analysis is first performed on the full dataset to emphasize seasonal variability. Figure 13 shows the variances for the first 20 CPCA modes, according to which we determine

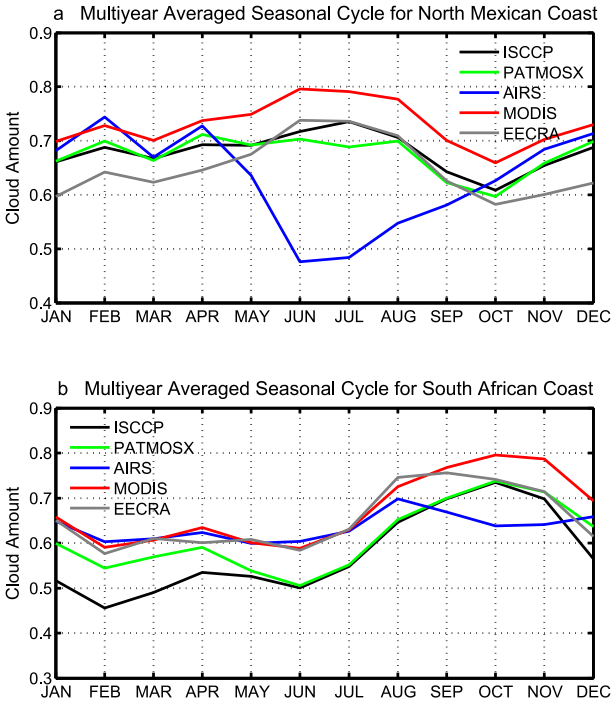

FIG. 9. As in Fig. 4, for the (a) northern Mexican and (b) South African coastal regions. We can see that while the AIRS seasonal cycle is out of phase for region A and has a phase shift for region B (due to the CA drop from August to December).

the first two modes, accounting for >50% of the total variance, to be dominant. Unlike results for CA, mode 3 does not exhibit semiannual cycles and is actually associated with the Pacific decadal oscillation (PDO). However, it is almost identical to mode 1 of deseasonalized analysis and thus is not shown here but will be discussed further in the deseasonalized section. The spatial patterns and PC time series of modes 1 and 2 are shown in Fig. 14. Both modes are clearly associated with distinct seasonal cycles. Mode 1 represents hemispheric differences in the CTP seasonality for the subtropical regions. Specifically, the Northern Hemisphere exhibits the lowest CTP or the highest cloud height during the local summer [June–August (JJA)], when convection is also at maximum, whereas for the Southern Hemisphere the seasonal cycle is reversed. Mode 2 exhibits a (boreal) spring–fall seasonal cycle, with negative anomalies found in the tropics and East Asia and positive anomalies in the midlatitudes. A weak positive trend is also observed in the PC of mode 2, which we find is mainly associated with decreasing CTP over the tropical and subtropical Pacific and tropical Atlantic and increasing CTP over the tropical central Pacific (figure not shown), and is confirmed by Tan et al. (2015). A cross comparison of the spatial patterns for each mode indicates good agreement between the different CTP datasets. Similar to the CA analysis,

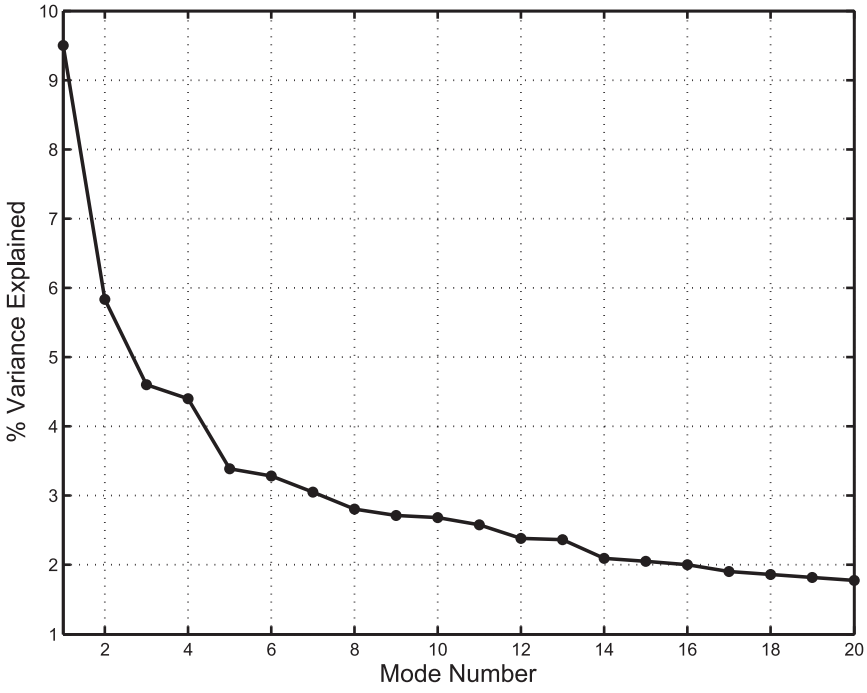

FIG. 10. Variances explained by the first 20 CMCA modes for deseasonalized CA data over ocean.

correlation coefficients against MODIS *Aqua* spatial modes are computed and tabulated in Table 5. We are able to see that all correlations are reasonably high. Only ISCCP mode 2 has relatively lower correlation, which is likely associated with the lack of positive signal over the tropical and subtropical Pacific. This result is consistent with the overall low variability in ISCCP over this region found in the standard deviation fields (Fig. 12). In addition, we note that the Sahel region also has large disagreements in both modes 1 and 2. For mode 1, this region appears with strong positive anomalies in PATMOS-x and AIRS maps but no detectable signals show up for ISCCP and MODIS. For mode 2, only the AIRS pattern has negative anomalies for the Sahel while neutral or weakly positive signals are found in the other datasets. As a result, we further investigate this problem by comparing the averaged CTP seasonal cycle for the Sahel (10°–15°N, 15°W–35°E) in Fig. 15. Figure 15 clearly shows that the seasonal cycles for PATMOS-x and AIRS are almost reversed with respect to ISCCP and MODIS. The overall magnitude of variability is also much stronger for PATMOS-x and AIRS datasets. This difference explains the strong projection of PC 1 on PATMOS-x and AIRS spatial maps over the Sahel, while the projection is weak or even reversed for the other three datasets. Also the agreement between AIRS and PATMOS-x implies that these differences are associated with cirrus detection, and that this region may be subject to larger cirrus retrieval uncertainties. In addition to the Sahel, which has the strongest contrast across the datasets, a few other regions also disagree. These include East Asia (including Siberia) in mode 1, where only MODIS shows strong positive anomalies, and the western United States, where a weak negative anomaly is found in the ISCCP.

We subsequently examine interannual variability through CPCA analysis of the deseasonalized datasets. Figure 16 plots the variance explained for the first 20 modes. There is a sharp drop from mode 3 to mode 4, and higher-order modes beyond mode 4 all explain very small fractions of the variance. We thus choose to examine the first three modes, and their spatial patterns and time series are presented in Fig. 17. Ocean cloud variability has been found sensitive to climate oscillations associated with sea surface temperature anomalies, such as ENSO (e.g., Park and Leovy 2004; Zhu et al. 2007; Zelinka and Hartmann 2011). Because globally, most cloud interannual variability comes from the oceans (bottom row of Fig. 12), we find that modes 1, 2, and 3 of deseasonalized CTP data (Fig. 17) respectively capture the signature of climate oscillations of the PDO, ENSO, and the Indian Ocean dipole mode (IODM). This conclusion is drawn by correlating the PC time series with the corresponding climate indices, and by comparing the spatial patterns with the associated SST anomalies. The PDO index (Zhang et al. 1997; Mantua et al. 1997) is

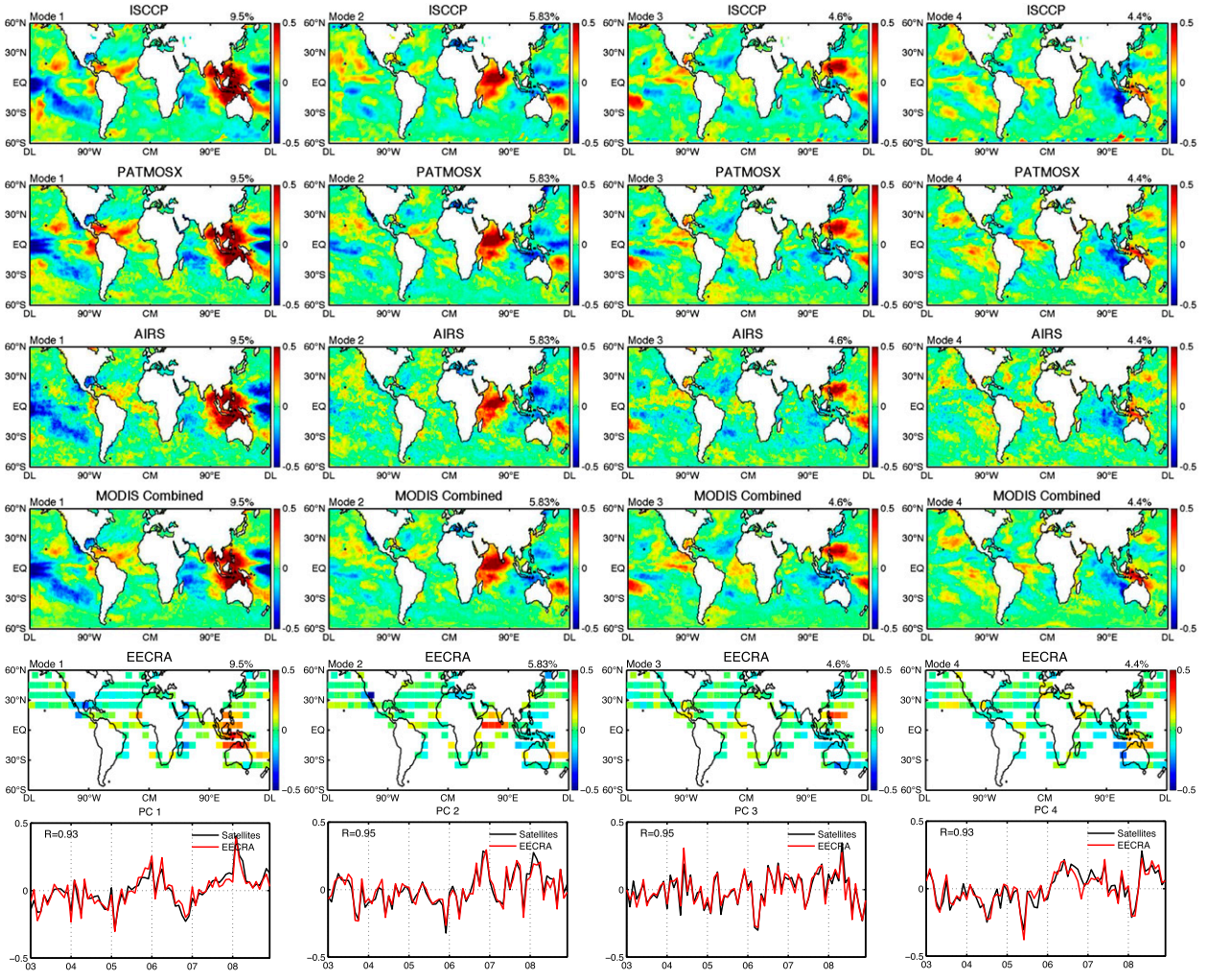

FIG. 11. As in Fig. 6, but for ocean data. Mode 1 represents the variability associated with ENSO. The correlations between the satellite and EECRA PCs with the multivariate ENSO index (MEI) are 0.81 and 0.69, respectively. Again, excellent agreement is found for all spatial patterns and the satellite and EECRA PCs are highly correlated.

provided by the University of Washington's Joint Institute for the Study of the Atmosphere and Ocean (JISAO), and the IODM index [or dipole mode index (DMI); Saji et al. 1999] is provided by the Japan Agency for Marine-Earth Science and Technology, which is derived from the HadISST dataset. The PDO index and MEI and DMI time series for the study period are superimposed on PCs 1, 2, and 3, respectively. The PDO index is scaled by 0.1 and MEI and DMI are scaled by 0.2 to yield a comparable order of magnitude to the PC time series. The PCs and the climate indices vary coherently, although the PCs usually appear noisier than the climate indices. This is reasonable as the decomposition is performed on global datasets, while the signals of these climate modes are concentrated over a particular region, such as tropical Pacific or the Indian Ocean. The correlation between PCs 1, 2, and 3 with the PDO index, MEI, and DMI are 0.64, 0.68, and 0.40, respectively. Their significance levels are 0.98, 0.99, and 0.92, respectively, after correcting for lag-1 autocorrelation. In addition to the temporal consistency, the spatial patterns also resemble the SST anomalies for the PDO, ENSO, and IODM. Mode 1 exhibits positive anomalies for the east Pacific and negative anomalies for the central Pacific. A

TABLE 4. Deseasonalized ocean dataset for CA, showing the correlation coefficients between the spatial patterns of modes 1 to 4 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOS-x	AIRS	EECRA
Mode 1	0.93	0.94	0.87	0.84
Mode 2	0.89	0.94	0.81	0.84
Mode 3	0.86	0.93	0.77	0.84
Mode 4	0.77	0.88	0.74	0.75

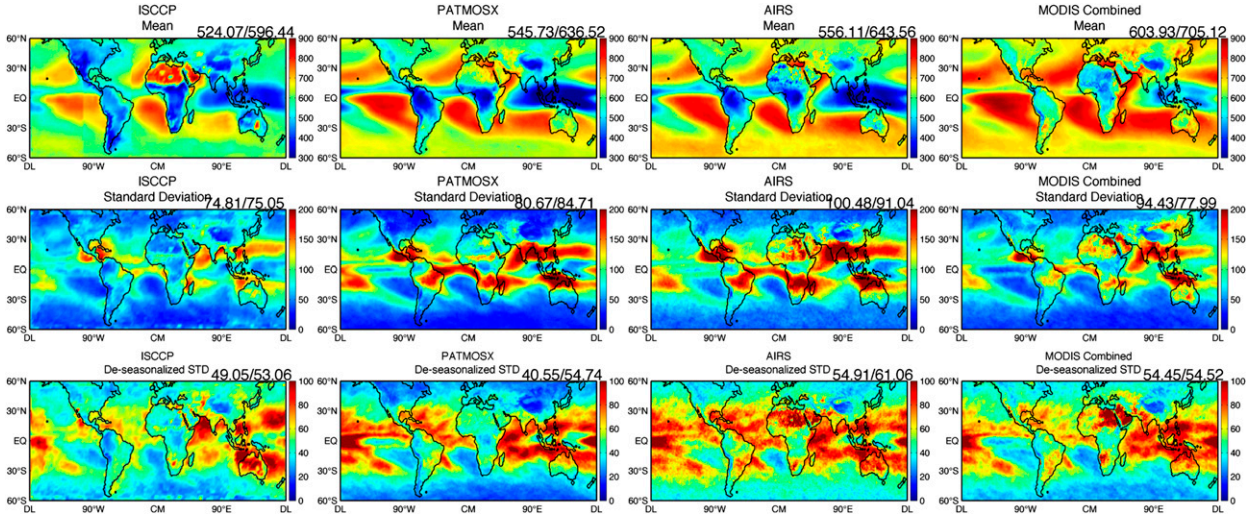

FIG. 12. (top) Multiyear averaged global mean CTP (color scale runs from 100 to 900 in 100 increments), (middle) CTP standard deviation fields (color scale runs from 0 to 200 in 50 increments), and (bottom) deseasonalized CTP standard deviation fields (color scale runs from 0 to 100 in 20 increments) of (left to right) the four CA datasets: ISCCP, PATMOS-x, AIRS, and MODIS Combined. The numbers on the top-right corner of each panel indicate land/ocean averages of the spatial map. Note that the color scale for the seasonal variability (middle row) is twice as large as that for the interannual variability (bottom row). All datasets agree well in the spatial patterns and global means.

strong positive anomaly also appears over the west Pacific warm pool, which may be due to the ''mode leak'' caused by overlapping spatial patterns of ENSO and PDO; that is, the ENSO and PDO types of variability are not strictly orthogonal and therefore CPCA analysis does not completely isolate them into two separate modes. The pattern of mode 2 agrees well with the SST anomaly induced by ENSO events. And the dipole feature for Indian Ocean in mode 3 resembles the SST pattern for the IODM positive phase (Saji et al. 1999).

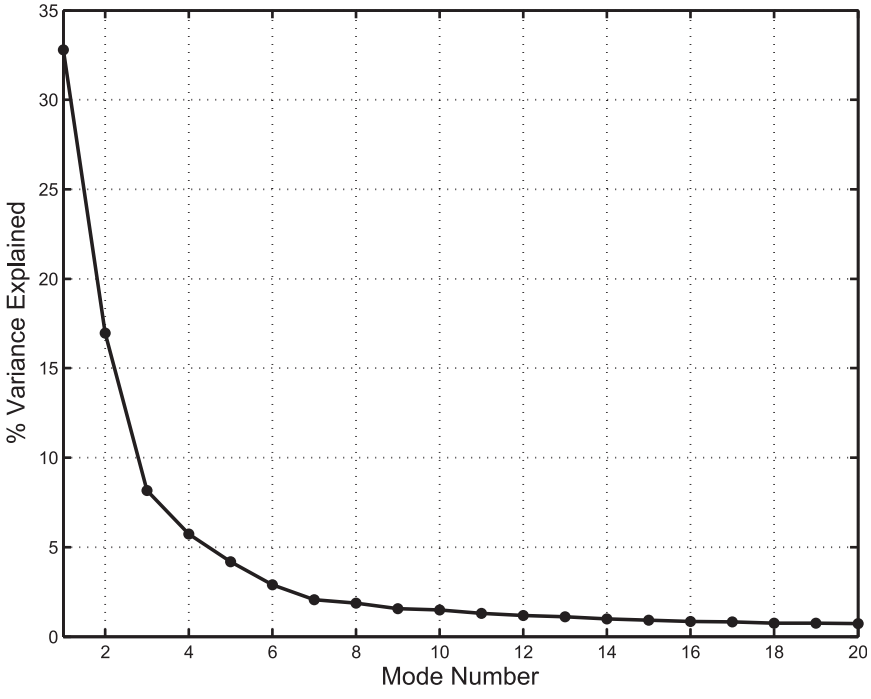

FIG. 13. Variances explained by the first 20 CPCA modes for full CTP dataset.

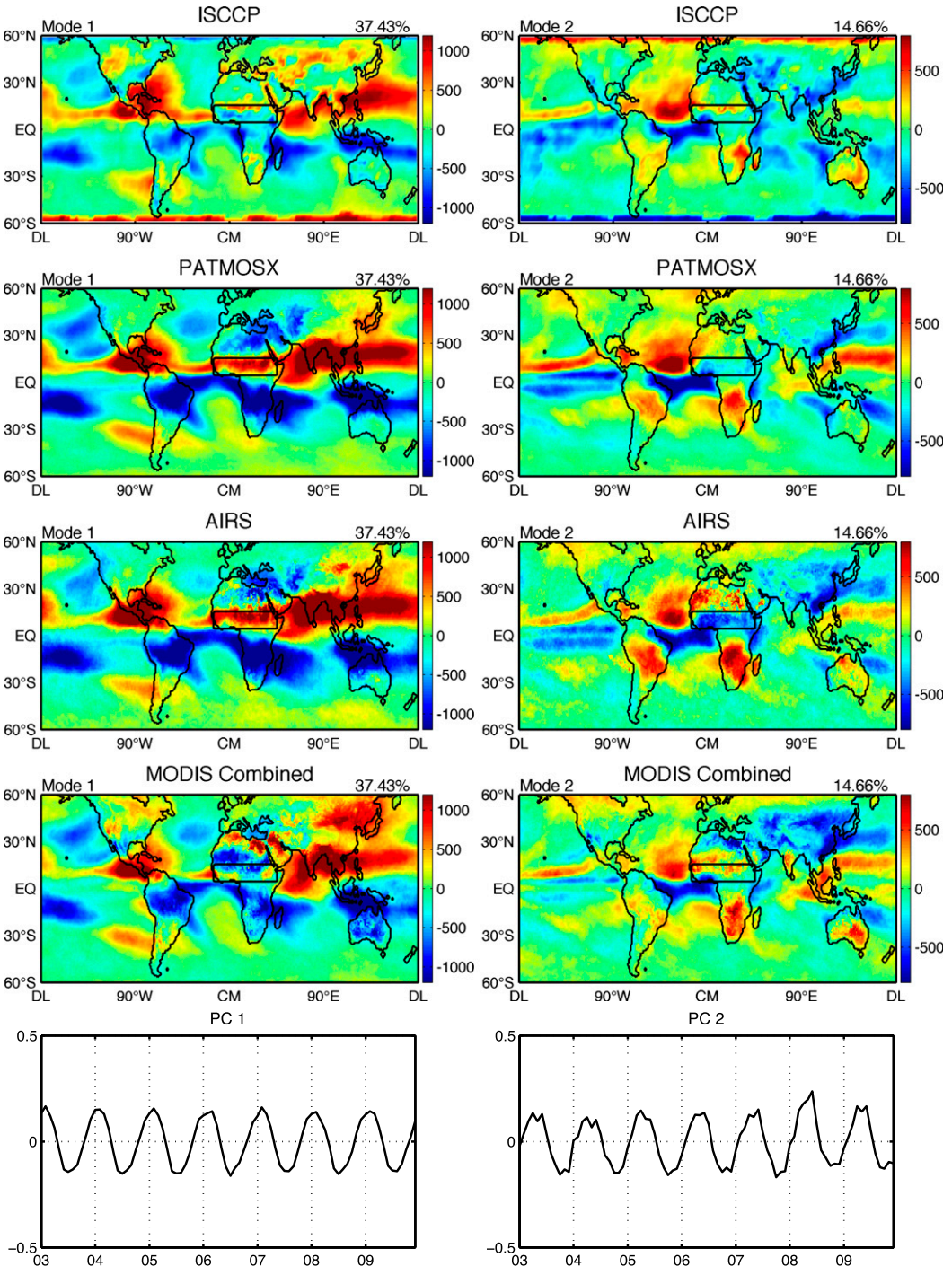

FIG. 14. As in Fig. 3, but for the two leading CPCA modes of the full CTP dataset; and without EECRA. Both modes exhibit distinct seasonality and the spatial modes for the four datasets agree very well. The major difference comes from the Sahel region (marked by the black rectangle on the spatial maps). In mode 1, this region exhibit positive anomaly for PATMOS-x and AIRS but weakly negative signals for the other data. For mode 2, only AIRS data indicate a negative anomaly over this region, while no significant signal is found for the other datasets.

TABLE 5. Full dataset for CTP, showing the correlation coefficients between the spatial patterns of modes 1 and 2 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOSX	AIRS
Mode 1	0.69	0.84	0.80
Mode 2	0.55	0.76	0.75

Nonetheless, the current data record is still relatively short compared to the time scale of decadal climate variability such as the PDO, and a longer time series is likely required to further investigate the relationship between the CTP and PDO.

Table 6 lists the correlation coefficients of the spatial patterns with those for MODIS *Aqua*. Most of the correlations are above 0.7, indicating good agreement. ISCCP modes 1 and 2 agree less well with the other datasets. Major regions of discrepancy include the Indian Ocean and west Pacific. These regions are dominated by deep convection that has strong diurnal cycles, so that diurnal sampling may explain part of the differences in perceived interannual variations. The good agreement across different datasets for all three modes further strengths the association of CTP variability with these major climate modes.

5. Conclusions

This study presents an intercomparison of the spatiotemporal variability of five publicly available cloud observational datasets. These datasets include retrievals from multispectral imagers, IR sounders, and surface-based human observation. Knowledge of the spatiotemporal cloud variability is critical in understanding the role of clouds in the climate system, and our intercomparison study provides an assessment of how this is represented by different datasets.

Two spectral decomposition techniques, namely CMCA and CPCA, are used to extract the major modes of variability from all datasets combined. These methods are particularly useful in space–time comparison as they allow the simultaneous examination of both spatial and

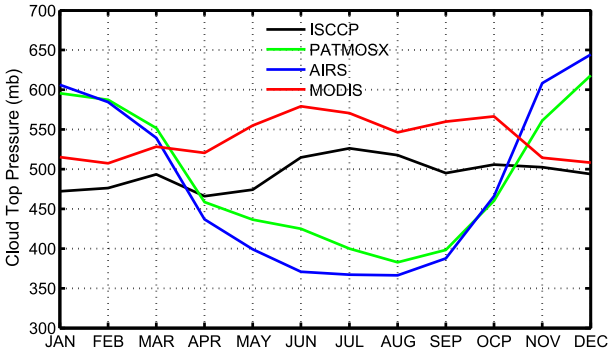

FIG. 15. Multiyear averaged seasonal cycle for the Sahel (marked by the black rectangle in Fig. 14) for the five datasets minus EECRA. The seasonal cycles for PATMOS-x and AIRS are clearly out of phase with ISCCP and MODIS.

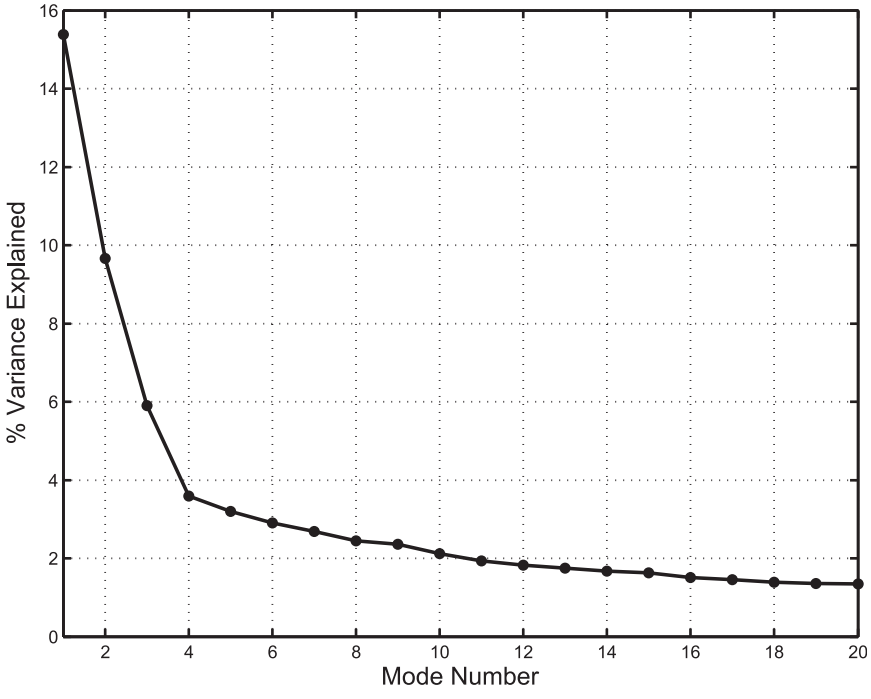

FIG. 16. Variances explained by the first 20 CPCA modes for deseasonalized CTP dataset.

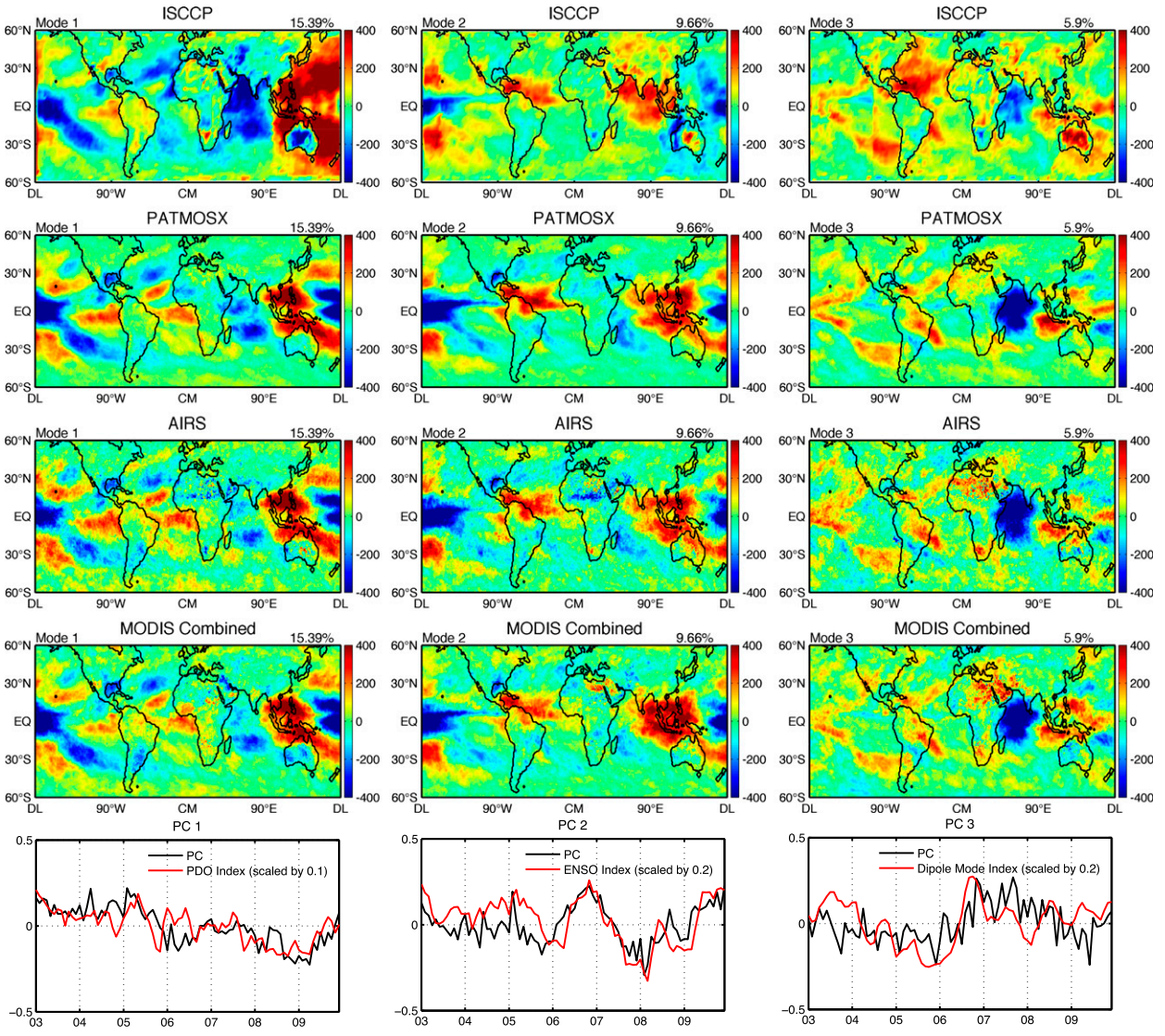

FIG. 17. As in Fig. 3, but for the first three modes of deseasonalized CTP dataset; and without EECRA. Modes 1, 2, and 3 are found to be correlated with the PDO, ENSO, and the IODM, respectively. The PDO index and MEI and DMI time series are plotted on top of the PC time series, and the correlations between PCs 1–3 with the PDO index, MEI, and DMI are 0.64, 0.68, and 0.40, respectively. The spatial patterns of these three modes also resemble the SST anomalies associated with these three types of climate oscillations. The spatial patterns of PATMOS-x, AIRS, and MODIS data are quite consistent. ISCCP captures the major spatial patterns, but with some differences over the Indian Ocean and west Pacific.

temporal variability and, furthermore, associate the comparison with physical processes such as seasonal cycles and signatures of climate modes.

The results reveal good agreement for both the CA and CTP, and on both seasonal and interannual time scales. The spatial modes of different datasets are highly correlated, with the majority of the correlations above 0.7 and often reaching 0.9, which is a strong support that all datasets capture the dominant cloud variations.

For the CA, the differences mainly lie in NH high latitudes and marine stratus cloud regions of northern Mexican coast and West African coast. For NH high latitudes, PATMOS-x and MODIS have a strong summer–winter seasonal cycle, while the seasonality of ISCCP, AIRS, and EECRA is relatively weaker and out of phase. For the marine stratus regions, off the northern Mexican coast the AIRS data are out of phase with all of the other datasets. AIRS also appears to have a phase shift in the seasonal cycle due to the decrease in cloud fraction from August to December off the West African coast. With respect to the CTP, major seasonal cycle differences are found over the Sahel region, where the PATMOS-x and

TABLE 6. Deseasonalized dataset for CTP, showing the correlation coefficients between the spatial patterns of modes 1, 2, and 3 of different datasets with those of MODIS.

Dataset	ISCCP	PATMOSX	AIRS
Mode 1	0.49	0.93	0.86
Mode 2	0.53	0.92	0.85
Mode 3	0.51	0.87	0.77

AIRS seasonal cycle is out of phase with ISCCP and MODIS. The interannual variability for ISCCP data might have artifacts for Indian Ocean and west Pacific. For the study period, ISCCP data over this region are directly from the Multifunctional Transport Satellite (MTSAT) of Japan Meteorological Agency, which has some anomalies in its lowest visible radiances and results, to some extent, in the detection of extra thin cirrus. On interannual scales, global CA variability is primarily governed by ENSO variability in the tropical Pacific, while the dominant variability in the CTP is associated with the PDO, ENSO, and the IODM.

The agreements reached in the CA and CTP parameters are encouraging, and increase our confidence in using these datasets for climate and modeling studies. In addition, the problems and inconsistencies identified through the comparison provide information for further improvements of the datasets or algorithms, such as the planned new ISCCP version, in order to produce a consistent cloud climatology. Future work will concentrate on other derived cloud variables, especially cloud optical depth and particle size. We will also further examine the variability of high, middle, and low cloud amounts, as these are equally important parameters in determining Earth's energy budget. Moreover, while the comparison in this study focuses on observational datasets, the spectral analysis techniques can also be useful tools in the intercomparison of model results, as well as between model and observations. These will also be interesting topics for future study.

Acknowledgments. We thank the GEWEX Cloud Assessment work group for providing the satellite datasets used in this study. The GEWEX Cloud Assessment datasets were obtained from the ClimServ Data Center of IPSL/CNRS, at http://climserv.ipsl.polytechnique.fr/gewexca/. The ISCCP D2 data were downloaded from NASA Langley Research Center Atmospheric Science Data Center (ASDC) at http://eosweb.larc.nasa.gov. The EECRA surface data are provided by the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (http://rda.ucar.edu/datasets/ds292.2/). The data were accessed on 10 September 2013. The MEI ENSO index is provided by the NOAA Earth System Research Laboratory, available from http://www.esrl.noaa.gov/psd/enso/mei/. The PDO index is downloaded from http://jisao.washington.edu/pdo/PDO.latest, and the DMI index is downloaded from http://www.jamstec.go.jp/. This study is funded by NASA Climate Grant 509496.02.08.04.24. Jing Li also acknowledges Hal Maring and the NASA Radiation Science program for providing funding for this investigation.

REFERENCES

Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, 1998: Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.,* **103,** 32 141–32 157, doi:10.1029/1998JD200032.

Aumann, H. H., and Coauthors, 2003: AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. *IEEE Trans. Geosci. Remote Sens.,* **41,** 253–264, doi:10.1109/TGRS.2002.808356.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560, doi:10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2.

Dai, A., T. R. Karl, B. Sun, and K. E. Trenberth, 2006: Recent trends in cloudiness over the United States: A tale of monitoring inadequacies. *Bull. Amer. Meteor. Soc.,* **87,** 597–606, doi:10.1175/BAMS-87-5-597.

Eastman, R., and S. G. Warren, 2013: A 39-yr survey of cloud changes from land stations worldwide 1971–2009: Long-term trends, relation to aerosols, and expansion of the tropical belt. *J. Climate,* **26,** 1286–1303, doi:10.1175/JCLI-D-12-00280.1.

Foster, M. J., and A. Heidinger, 2013: PATMOS-x: Results from a diurnally corrected 30-yr satellite cloud climatology. *J. Climate,* **26,** 414–425, doi:10.1175/JCLI-D-11-00666.1.

Frey, R. A., S. A. Ackerman, Y. Liu, K. I. Strabala, H. Zhang, J. R. Key, and X. Wang, 2008: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5. *J. Atmos. Oceanic Technol.,* **25,** 1057–1072, doi:10.1175/2008JTECHA1052.1.

Guignard, A., C. J. Stubenrauch, A. J. Baran, and R. Armante, 2012: Bulk microphysical properties of semi-transparent cirrus from AIRS: A six year global climatology and statistical analysis in synergy with geometrical profiling data from CloudSat-CALIPSO. *Atmos. Chem. Phys.,* **12,** 503–525, doi:10.5194/acp-12-503-2012.

Hahn, C. J., and S. G. Warren, 1999: Extended edited synoptic cloud reports from ships and land stations over the globe, 1952–1996. Numerical Data Package NDP-026C, 79 pp., accessed February 2014, doi:10.3334/CDIAC/cli.ndp026c.

——, ——, and J. London, 2009: Extended edited synoptic cloud reports from ships and land stations over the globe, 1952–1996 (updated to 2009). Numerical Data Package NDP-026C, 76 pp., accessed February 2014, doi:10.3334/CDIAC/cli.ndp026c.

Heidinger, A. K., and M. J. Pavolonis, 2009: Gazing at cirrus clouds for 25 years through a split window. Part I: Methodology. *J Appl. Meteor. Climatol.,* **48,** 1100–1116, doi:10.1175/2008JAMC1882.1.

——, A. T. Evan, M. J. Foster, and A. Walther, 2012: A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *J. Appl. Meteor. Climatol.,* **51,** 1129–1144, doi:10.1175/JAMC-D-11-02.1.

Jin, Y., and W. B. Rossow, 1997: Detection of cirrus overlapping low-level clouds. *J. Geophys. Res.,* **102,** 1727–1737, doi:10.1029/96JD02996.

King, M. D., S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, 2013: Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. *IEEE Trans. Geosci. Remote Sens.,* **51,** 3826–3852, doi:10.1109/TGRS.2012.2227333.

Li, J., B. E. Carlson, and A. A. Lacis, 2013: Application of spectral analysis techniques in the inter-comparison of aerosol data. Part I: An EOF approach to analyze the spatial-temporal variability of aerosol optical depth using multiple remote sensing data sets. *J. Geophys. Res. Atmos.,* **118,** 8640–8648, doi:10.1002/jgrd.50686.

——, ——, and ——, 2014a: Application of spectral analysis techniques in the inter-comparison of aerosol data. Part II: Using maximum covariance analysis to effectively compare spatiotemporal variability of satellite and AERONET measured aerosol optical depth. *J. Geophys. Res. Atmos.,* **119,** 153–166, doi:10.1002/2013JD020537.

——, ——, and ——, 2014b: Application of spectral analysis techniques in the inter-comparison of aerosol data. Part III: Using combined PCA to compare spatiotemporal variability of MODIS, MISR, and OMI aerosol optical depth. *J. Geophys. Res. Atmos.,* **119,** 4017–4042, doi:10.1002/2013JD020538.

——, ——, and ——, 2014c: Application of spectral analysis techniques to the inter-comparison of aerosol data. Part IV: Combined maximum covariance analysis to bridge the gap between multi-sensor satellite retrievals and ground-based measurements. *Atmos. Meas. Tech. Discuss.,* **7,** 3503–3547, doi:10.5194/amtd-7-3503-2014.

Li, Z., M. C. Cribb, F. L. Chang, and A. P. Trishchenko, 2004: Validation of MODIS-retrieved cloud fractions using whole sky imager measurements at the three arm sites. *Proc. 14th ARM Science Team Meeting,* Albuquerque, NM, ARM, 1–6.

Liao, X., W. B. Rossow, and D. Rind, 1995: Comparison between SAGE II and ISCCP high-level clouds: 1. Global and zonal mean cloud amounts. *J. Geophys. Res.,* **100,** 1121–1135, doi:10.1029/94JD02429.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.,* **78,** 1069–1079, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Marchand, R., T. Ackerman, M. Smyth, and W. B. Rossow, 2010: A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS. *J. Geophys. Res.,* **115,** D16206, doi:10.1029/2009JD013422.

Menzel, W. P., and Coauthors, 2008: MODIS global cloud-top pressure and amount estimation: Algorithm description and results. *J. Appl. Meteor. Climatol.,* **47,** 1175–1198, doi:10.1175/2007JAMC1705.1.

Naud, C., J.-P. Muller, and E. E. Clothiaux, 2002: Comparison of cloud top heights derived from MISR stereo and MODIS CO_2-slicing. *Geophys. Res. Lett.,* **29,** doi:10.1029/2002GL015460.

Park, S., and C. B. Leovy, 2004: Marine low-cloud anomalies associated with ENSO. *J. Climate,* **17,** 3448–3469, doi:10.1175/1520-0442(2004)017<3448:MLAAWE>2.0.CO;2.

Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey, 2003: The MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.,* **41,** 459–473, doi:10.1109/TGRS.2002.808301.

Rossow, W. B., and R. A. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.,* **72,** 2–20, doi:10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2.

——, and ——, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.,* **80,** 2261–2287, doi:10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.

——, A. W. Walker, and L. C. Garder, 1993: Comparison of ISCCP and other cloud amounts. *J. Climate,* **6,** 2394–2418, doi:10.1175/1520-0442(1993)006<2394:COIAOC>2.0.CO;2.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature,* **401,** 360–363.

Stubenrauch, C. J., W. B. Rossow, F. Chéruy, A. Chédin, and N. A. Scott, 1999: Clouds as seen by satellite sounders (3I) and imagers (ISCCP). Part I: Evaluation of cloud parameters. *J. Climate,* **12,** 2189–2213, doi:10.1175/1520-0442(1999)012<2189:CASBSS>2.0.CO;2.

——, S. Cros, A. Guignard, and N. Lamquin, 2010: A 6-year global cloud climatology from the Atmospheric Infrared Sounder AIRS and a statistical analysis in synergy with CALIPSO and CloudSat. *Atmos. Chem. Phys.,* **10,** 7197–7214, doi:10.5194/acp-10-7197-2010.

——, and Coauthors, 2013: Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel. *Bull. Amer. Meteor. Soc.,* **94,** 1031–1049, doi:10.1175/BAMS-D-12-00117.1.

Tan, J., C. Jakob, W. B. Rossow, and G. Tselioudis, 2015: Increases in tropical rainfall driven by changes in frequency of organized deep convection. *Nature,* **519,** 451–454, doi:10.1038/nature14339.

Wang, J., W. B. Rossow, and Y.-C. Zhang, 2000: Cloud vertical structure and its variations from a 20-yr global rawinsonde dataset. *J. Climate,* **13,** 3041–3056, doi:10.1175/1520-0442(2000)013<3041:CVSAIV>2.0.CO;2.

Weisz, E., J. Li, W. P. Menzel, A. K. Heidinger, B. H. Kahn, and C.-Y. Liu, 2007: Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals. *Geophys. Res. Lett.,* **34,** L17811, doi:10.1029/2007GL030676.

Winker, D. M., M. A. Vaughan, A. Omar, Y. Hu, K. A. Powell, Z. Liu, W. H. Hunt, and S. A. Young, 2009: Overview of the CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Oceanic Technol.,* **26,** 2310–2323, doi:10.1175/2009JTECHA1281.1.

Wolter, K., and M. S. Timlin, 2011: El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.,* **31,** 1074–1087, doi:10.1002/joc.2336.

Zelinka, M. D., and D. L. Hartmann, 2011: The observed sensitivity of high clouds to mean surface temperature anomalies in the tropics. *J. Geophys. Res.,* **116,** D23103, doi:10.1029/2011JD016459.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate,* **10,** 1004–1020, doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zhu, P., J. J. Hack, J. T. Kiehl, and C. S. Bretherton, 2007: Climate sensitivity of tropical and subtropical marine low cloud amount to ENSO and global warming due to doubled CO_2. *J. Geophys. Res.,* **112,** D17108, doi:10.1029/2006JD008174.