Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations

MINGYUE CHEN,* PINGPING XIE, AND JOHN E. JANOWIAK

Climate Prediction Center, National Centers for Environmental Prediction, National Weather Service, National Oceanic and Atmospheric Administration, Camp Springs, Maryland

PHILLIP A. ARKIN

Office of Global Programs, Office of Oceanic and Atmospheric Research, National Oceanic and Atmospheric Administration, Silver Spring, Maryland

(Manuscript received 12 July 2001, in final form 14 November 2001)

ABSTRACT

This paper describes the initial work toward the production of monthly global (land and ocean) analyses of precipitation for an extended period from 1948 to the present. Called the precipitation reconstruction (PREC), the global analyses are defined by interpolation of gauge observations over land (PREC/L) and by EOF reconstruction of historical observations over ocean (PREC/O). This paper documents the creation of the land component of the analyses (PREC/L) on a 2.5° latitude/longitude grid for 1948–2000. These analyses are derived from gauge observations from over 17 000 stations collected in the Global Historical Climatology Network (GHCN), version 2, and the Climate Anomaly Monitoring System (CAMS) datasets. To determine the most suitable objective analysis procedure for gridding, the analyses generated by four published objective analysis techniques [those of Cressman, Barnes, and Shepard, and the optimal interpolation (OI) method of Gandin] were compared. The evaluation demonstrated two crucial points: 1) better results are obtained when interpolating anomalies rather than the precipitation totals, and 2) the OI analysis procedure provided the most accurate and stable analyses among the four algorithms that were tested. Based on these results, the OI technique was used to create monthly gridded analyses of precipitation over the global land areas for the 53-yr period from 1948 to 2000. In addition, some diagnostic investigations of the seasonal and interannual variability of large-scale precipitation over the global land areas are presented. The mean distribution and annual cycle of precipitation observed in the PREC/L showed good agreement with those in several published gauge-based datasets, and the anomaly patterns associated with ENSO resemble those found in previous studies. The gauge-based dataset (PREC/L) will be updated on a quasi-real-time basis and is available online (ftp.ncep.noaa.gov/pub/precip/50-yr).

1. Introduction

In recent years, several algorithms have been developed to construct analyses of global monthly precipitation by merging multiple individual data sources. One such algorithm developed by a group at the National Aeronautics and Space Administration Goddard Space Flight Center combines gauge observations with estimates derived from IR, outgoing longwave radiation (OLR), Special Sensor Microwave Imager (SSM/I), and Television IR Observational Satellite (TIROS) Operational Vertical Sounder (TOVS) observations (Adler et al. 1993, 1994; Huffman et al. 1995). It has been applied successfully to construct the global monthly precipitation analyses for the Global Precipitation Climatology Project (GPCP) for the period from 1979 to the present (Huffman et al. 1997; Adler et al. 2001, manuscript submitted to *Bull. Amer. Meteor. Soc.*, hereinafter ADLR). Another merging algorithm is that of Xie and Arkin (1996), which takes the gauge observations, satellite estimates derived from IR, OLR, Microwave Sounding Unit, and SSM/I and the precipitation distributions from the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis as inputs. Using this algorithm, a global monthly precipitation dataset, called the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997), has been created for the same period as for the GPCP product. Both the GPCP and the CMAP monthly precipitation datasets have been applied widely in climate analysis (Curtis and Adler 2000), numerical model verifications

* Additional affiliation: RS Information Systems, Inc. (RSIS), McLean, VA.

Corresponding author address: Dr. Mingyue Chen, Climate Prediction Center, NOAA/NWS/NCEP, 5200 Auth Road, No. 800, Camp Springs, MD 20746.
E-mail: mingyue.chen@noaa.gov

(Stephenson et al. 1998; Janowiak et al. 1998; Dai et al. 2001), hydrological research (Trenberth and Guillemott 1998), and other investigations (Yang et al. 1999).

Both of these merging procedures rely heavily on the estimates derived from satellite observations, which are not available for the period before the 1970s. Many applications, especially those related to climate variability research, however, require precipitation analyses for a longer period. Although the GPCP and CMAP datasets cover a 22-yr period that includes 4–5 ENSO events, it is highly desirable to have a monthly dataset for an extended period so as to facilitate improved understanding of long-term climate variability and to determine how well numerical models can represent and predict it (Kumar et al. 1996; Kumar and Hoerling 1997; Rowell 1998).

The only source of precipitation measurements available for the period before the 1970s is that of the rain gauge observations. Although gauge observations suffer from various problems, such as poor spatial sampling over oceanic and unpopulated land areas, temporal inhomogeneity in historical records, and uncertainty in undercatchment due to wind and evaporation effects (Sevruk 1982, 1989; Hulme 1995), their extended record makes them a valuable source for precipitation ''climatologies'' and time series.

One of the most widely used climatologies of monthly precipitation is that of Jaeger (1976), in which analyzed values on a 4° latitude × 5° longitude global grid are defined by subjective estimation based on several regional precipitation maps mainly representative of the period from 1931 to 1960. Another well-known precipitation climatology is that of Legates and Willmott (1990), which is constructed on a 0.5° latitude/longitude grid. Over land areas, this monthly climatology is defined by objectively interpolating gauge observations at over 24 000 stations; over the oceans it is based on empirical estimates derived from ship observations of present weather (Dorman and Bourke 1979, 1981). Recently, New et al. (1999) published a climatology of monthly terrestrial precipitation representative of the mean status over a 30-yr period from 1961 to 1990. Monthly precipitation values on a 0.5° latitude/longitude grid are calculated by interpolating corresponding climatological normals at over 19 000 gauge stations included in the dataset of Eischeid et al. (1991) and several other sources (e.g., reports from national meteorological agencies) collected by the Climate Research Unit (CRU) of the University of East Anglia (UEA) in the United Kingdom.

Many efforts have been focused on the construction of the time series of monthly precipitation from gauge observations. Assuming that accumulated station observations of precipitation follow the gamma distribution, Bradley et al. (1987) converted the anomaly of precipitation into percentiles for 1477 Northern Hemisphere stations collected by Bradley et al. (1985) and interpolated them onto grid points at an interval of 400 km for a period beginning from the middle of the nineteenth century. Diaz et al. (1989) extended this analysis to the entire global land area by including gauge observations at 656 Southern Hemisphere stations taken from archives maintained by the CRU/UEA and NCAR.

Dai et al. (1997) described a unique effort in the construction of a gauge-based analysis of monthly precipitation over global land areas. First, EOF analysis is performed on the original time series of gauge observations from approximately 5300 stations collected in Eischeid et al. (1991) to remove temporal inhomogeneity caused by instrumental changes throughout the record period. A monthly analysis of global land precipitation is then defined on a 2.5° latitude/longitude grid for the period from 1900 to 1988. This analysis was combined with that of Xie and Arkin (1997) to document the global patterns of ENSO-induced precipitation anomalies (Dai and Wigley 2000).

A series of studies has been reported by a group in the CRU/UEA to create datasets of large-scale precipitation specifically suited for evaluation of general circulation models (GCMs). Hulme (1991, 1992) constructed a monthly analysis of precipitation on a 5° latitude/longitude grid over the global land areas for 1950–80 by interpolating quality-controlled gauge observations at approximately 2500 stations, based on the aforementioned archive of Eischeid et al. (1991) and enhanced by other CRU collections, and applied it to evaluate outputs from several GCMs. Hulme (1994) recreated the gauge-based precipitation analysis on a 2.5° latitude × 3.75° longitude grid for the period from 1900 to 1992 using gauge observations at over 4900 stations collected in their updated database. Doherty et al. (1999) combined this gauge-based analysis with oceanic precipitation estimates derived from the OLR data observed by the polar-orbiting National Oceanic and Atmospheric Administration (NOAA) satellites (Gruber and Krueger 1984) to create an analysis of monthly precipitation over the global tropical areas for the period from 1974 to 1994. Most recent, the CRU group, as reported by New et al. (2000), created monthly analyses of precipitation and several other climatic variables for the period from 1901 to 1996. The precipitation analysis is defined on a fine spatial resolution of 0.5° latitude/longitude over global land areas using gauge observations at up to 13 500 stations collected in the CRU archives, which have been updated for over 20 years.

As a key component of the GPCP (Arkin and Xie 1994), which was established to provide large-scale precipitation datasets for many World Climate Research Programme activities, the Global Precipitation Climatology Centre (GPCC) in Germany has developed a sophisticated system of data selection, quality control, and objective interpolation for constructing gauge-based analyses over the global land areas (Rudolf 1993; Schneider 1993). As of December of 2000, they have finished the monthly analyses on 2.5° and 1.0° latitude/

longitude grids for their quasi-real-time monitoring product for the period from January of 1986 to September of 2000 using gauge observations of almost 7000 stations obtained from the Global Telecommunication System (GTS). Applying the same algorithm as the GPCC, Xie et al. (1996) created a gauge-based analysis of global land precipitation on a 2.5° latitude/longitude grid for the period from 1971 to 1995 using gauge observations at over 6000 stations obtained from the version-1 dataset of the Global Historical Climatology Network (GHCN; Vose et al. 1992) of the NOAA National Climatic Data Center (NCDC) and the Climate Anomaly Monitoring System (CAMS; Ropelewski et al. 1985) of NOAA/CPC. The gauge-based analyses of Xie et al. (1996) for 1979–1985 and of the GPCC for 1986–present are used to construct the merged analysis of monthly precipitation in both the GPCP and the CMAP (ADLR; Xie and Arkin 1997).

All of the gauge-based precipitation analyses described above cover primarily only the land areas of the globe. Over the vast oceanic areas, direct quantitative observations of precipitation exist neither over a broad region nor for an extended period before the 1970s, for which no satellite observations are available, although there are several published efforts to construct a climatology of oceanic precipitation and its frequency using ship-observed present weather reports (Jaeger 1976; Dorman and Bourke 1979, 1981; Petty 1995; Dai 2001a,b).

Motivated by the increasing demand for a high-quality dataset of observation-based large-scale precipitation over both the land and oceanic areas of the globe for the period before the 1970s, we have launched a project to construct an analysis of monthly precipitation for an extended period from 1948 to the present. Although several gauge-based monthly analyses already exist (e.g., Dai et al. 1997; New et al. 2000), the recent release of the GHCN version-2 (GHCN2) dataset, which includes gauge observations of monthly precipitation for over 23 000 stations (Peterson and Vose 1997), makes it possible to create an analysis with improved quality over land. This extended analysis of monthly precipitation over the global land areas is defined by interpolating gauge observations collected in the GHCN2 and CAMS datasets. Over the global oceanic areas, the analysis is defined by projecting historical records of coastal, island, and ship observations onto spatial variation patterns of large-scale precipitation observed by satellites for later years using a technique applied successfully by Smith et al. (1996) to reconstruct sea surface temperature fields. The extended analysis of global monthly precipitation is targeted to begin in 1948 in consideration of the discontinuity and unavailability of meteorological observations over many stations during and before World War II. The spatial resolution of the dataset is 2.5° latitude/longitude, the same as those in the GPCP and CMAP merged analyses of monthly precipitation.

The objective of this paper is to report the construction of the monthly precipitation analysis over the global land areas by interpolating gauge observations in the GHCN2 and CAMS datasets. Section 2 describes the selection of the interpolation algorithms based on an intercomparison of several published methods, section 3 reports on the construction of the gauge-based analysis for the 53-yr period from 1948 to 2000, section 4 shows comparisons with several published datasets of gauge-based precipitation and applications of the 53-yr analysis in some simple diagnostics of annual and interannual variations of large-scale precipitation, and a summary is given in section 5.

2. Interpolation algorithm

The determination of the methodology for constructing a gauge-based analysis of precipitation involves two major issues: 1) design of interpolation strategy (directly interpolating the total precipitation or defining it by adding the interpolated anomaly to a climatology) and 2) selection of an interpolation algorithm to define the analyzed values at grid points from irregularly distributed station observations. The GPCC (Rudolf 1993; Schneider 1993) and Xie et al. (1996) defined their grid fields by directly interpolating the total values of the monthly precipitation observations; Bradley et al. (1987), Diaz et al. (1989), Hulme (1992, 1994), Dai et al. (1997), and New et al. (2000) chose the anomaly of precipitation as their interpolation quantity. It has not been shown, however, which approach yields better representation of precipitation, although it is argued that the anomaly scheme is better because some of the spatial variation of total precipitation is associated with local topography and is thus steady (Dai et al. 1997).

Many different interpolation algorithms have been applied to define these gauge-based precipitation analyses. Bradley et al. (1987) and Diaz et al. (1989) applied an approach similar to that of Barnes (1964), in which value at a grid point is calculated by taking the mean of observations at four nearby stations weighted by a negative exponential function of the square of the distance between the station and grid point. Hulme (1992, 1994) employed the algorithm of Thiessen's polygon, in which the analyzed value for a grid box is given as the mean of the surrounding gauge observations weighted by areas closest to each of the gauges within the grid box. In constructing their climatology of monthly precipitation, New et al. (1999) used a thin-plate spline-fitting technique that takes the precipitation as a function of latitude, longitude, and elevation. Later, however, when creating the time series of monthly precipitation anomaly, New et al. (2000) implemented a distance-weighting interpolation algorithm very similar to that of Shepard (1968).

Creutin and Obled (1982) examined several well-known schemes, including the nearest neighbor method, the arithmetic mean, the spline-plate fitting, the optimal

interpolation (OI), the kriging, and the EOF interpolation methods and recommended the OI of Gandin (1965) for its excellent results and relatively easy implementation. Bussieres and Hogg (1989) compared the performance of four algorithms (Cressman 1959; Barnes 1964; Shepard 1968; Gandin 1965) and concluded that OI does the best job and the method of Shepard (1968) performs almost as well. Both of these studies investigated the analyses of total precipitation of mesoscale with spatial scales of ~100 km and timescales of ~24 h. The suitability of interpolation schemes for total precipitation with much larger scales is reported by Legates (1987), who, after evaluating several procedures, found that a spherical adaptation of the Shepard method (Willmott et al. 1985) is the best for their interpolation of over 24 000 gauge observations of long-term (generally over 10 yr) mean precipitation. It is not clear, however, how well various algorithms perform in creating fields of total precipitation for temporal scales between daily and annual or how the accuracy of the analysis is affected by defining it through directly interpolating the total values or by adding the interpolated anomaly to the climatology.

To fill in this gap and to design the best strategy to construct the gauge-based analysis of global monthly precipitation, an intercomparison of several interpolation algorithms was conducted for both the total value and the anomaly of precipitation for daily, pentad, and monthly timescales.

a. Candidate interpolation algorithms

Four published interpolation algorithms are selected for their popularity and feasibility in implementation. These algorithms are the Cressman (1959), the Barnes (1964), the Shepard (1968), and the OI of Gandin (1965).

In the method of Cressman (1959), a first-guess field of interpolated values at the target grid points is defined. This first-guess field is then corrected by the weighted mean of the differences between the observed and interpolated values over gauge locations within a searching distance. This process is repeated four times, with decreasing search distance. In our implementation, the first guess is defined as the arithmetic mean of the gauge observations, and the searching distances for the four iterations are 320, 240, 150, and 120 km, respectively.

The algorithm of Barnes (1964) is also an interpolation technique based on iterative correction using distance weighting. It differs from the Cressman method in that the searching distance is fixed and the weighting coefficient is a negative exponential function of the gauge–gridpoint distance. After several trials, we set the searching distance for the Barnes as 300 km, the same as that used by Dai et al. (1997), who derived it based on the spatial correlation distance of gauge-observed precipitation.

In the algorithm of Shepard (1968), the interpolated value at a target grid point is defined as a weighted sum of observed values at nearby gauge stations within a searching distance. The searching distance is variable depending on the gauge network density, so that 4–10 gauges are included in the calculation. The weighting function is inversely proportional to the gauge–gridpoint distance, yielding a sharper decrease than those in the Cressman and Barnes. In addition, directional correction is implemented to account for any uneven distribution of gauges in different directions.

The OI technique of Gandin (1965) defines the analysis value at a grid point by modifying a first-guess field with the weighted mean of the differences between the observed values and the first-guess values at station locations within a searching distance. Whereas in the Cressman, Barnes, and Shepard methods the weighting coefficient is a function only of gauge–gridpoint distance, in the OI technique it is determined from the variance and covariance relationship of the target precipitation field. In this study, the long-term mean of precipitation (see following sections for details) is used as the first guess, and the searching distance is set to 300 km as in Dai et al. (1997).

b. Intercomparison strategy

To assess which of the four objective analysis methods produces the most accurate precipitation estimates, an evaluation strategy was devised. Cross-validation tests were conducted for each of the four selected interpolation algorithms and for two sets of gauge observation data, one for daily precipitation using a dense network over the United States and the other for monthly precipitation over the entire global land area. For each dataset, gauge reports at 10%, randomly selected, of the stations were withdrawn and those at the remaining 90% of the stations were used to define the analysis values at locations of the withdrawn stations by interpolation. This process was conducted 10 times so that each gauge was withdrawn once. The analysis values produced by interpolation using the four selected algorithms were then compared with the corresponding observations to examine their performance. Note that in this cross-validation practice, the analysis values at the 10% of withdrawn gauge locations are calculated by directly interpolating observations at the remaining 90% of gauges rather than by interpolating analysis values at nearby regular grid points.

For the three interpolation algorithms based on distance weighting (the Cressman, Barnes, and Shepard), two different approaches were undertaken to define the analysis values of precipitation. In the first approach, the analysis value of total precipitation is defined by directly interpolating the observations of total precipitation at surrounding gauge stations; in the other approach, values of precipitation anomaly are first defined by interpolating the observed anomaly, and the analyzed values of total precipitation are then defined by adding

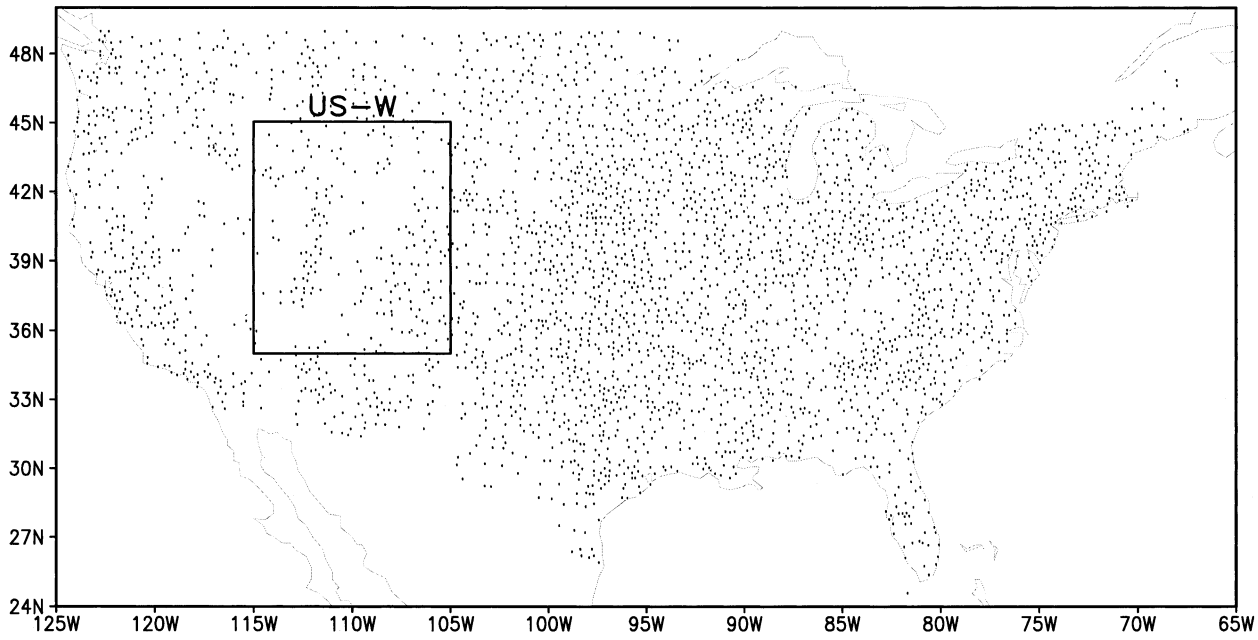

FIG. 1. Distribution of the gauge stations included in the NCDC ''Surface Land Daily Cooperative Summary of the Day'' dataset of daily precipitation with 80% or higher reporting rate during 1979–93. The labeled box indicates the subregion of the western U.S. mountainous area (see text).

the interpolated anomalies to the long-term mean. In the following discussions, we will refer to them as the direct and indirect approaches, respectively, for convenience. As described in section 2a, the OI technique implicitly analyzes anomalies, and thus only the indirect approach is applied for this technique.

Three statistical parameters were calculated to assess the agreement between the interpolated and the observed precipitation. They are the correlation, the bias, and the root-mean-square (rms) error. The correlation is an index of how well the spatial and temporal variation patterns match. The bias, defined as differences between the mean values of interpolated precipitation and those of the withdrawn station observations, provides a quantitative measure of how well the magnitudes agree. A positive bias indicates that interpolated values are larger than withdrawn observations. The rms error, defined as the root-mean-square differences between the analyzed values and the observations, helps us to understand how much the two sets of values differ in quantity. Both the bias and the rms error are normalized by the mean values of the observations. Because the frequency distribution of precipitation is highly skewed toward 0 (no rain), especially for short accumulations (e.g., daily), the normalized bias and rms error show very large values for cases with small mean precipitation (e.g., over dry regions).

c. Intercomparison over the United States

The first set of intercomparisons of the four interpolation algorithms was conducted using the gauge observations of daily precipitation over the United States collected in the ''Surface Land Daily Cooperative Summary of the Day'' by NCDC (Hughes et al. 1992). The original dataset includes gauge observations of daily precipitation at over 3800 stations for the period from 1979 to 1993. Only data for 3115 stations with 80% or higher reporting rates were used in this study. As shown in Fig. 1, these gauges are distributed relatively uniformly over the conterminous portion of the United States, with an average station-to-station distance of about 60 km.

To examine the effect of temporal averaging on the accuracy of interpolation, pentad and monthly precipitation values were calculated for each station by accumulating daily reports. Long-term means of daily, pentad, and monthly precipitation were then defined for each of the 3115 gauge stations using the corresponding data for the 15-yr period. These long-term means were then used as the first guess in the OI-based analysis and to define the total precipitation from the interpolated anomaly in the indirect approach using the three distance-weighting algorithms.

Although including data for longer periods yields more stable statistics, cross-validations for the interpolation of daily, pentad, and monthly precipitation were conducted for only 62 days (January and July of 1989), 37 pentads (pentads 1–37 of 1989), and 24 months (January 1989–December 1990), respectively, to reduce the computation amount to a manageable level. Because the spatial variability of precipitation varies from area to area, statistics were calculated for the entire

TABLE 1. Cross-validation results for daily, pentad, and monthly precipitation over the United States.

Region	Time	Cressman		Barnes		Shepard		
		Direct	Indirect	Direct	Indirect	Direct	Indirect	OI indirect
				Correlation				
U.S.	Daily	0.773	0.865	0.782	0.860	0.750	0.842	0.848
	Pentad	0.875	0.904	0.877	0.898	0.858	0.892	0.902
	Monthly	0.902	0.924	0.902	0.925	0.890	0.921	0.929
US-W	Daily	0.564	0.766	0.600	0.784	0.550	0.723	0.738
	Pentad	0.736	0.797	0.752	0.797	0.725	0.787	0.802
	Monthly	0.780	0.863	0.793	0.874	0.764	0.864	0.883
				Relative rms error (%)				
U.S.	Daily	204.5	160.2	199.8	162.7	215.6	173.9	170.8
	Pentad	83.8	73.6	82.8	75.7	89.3	78.4	74.7
	Monthly	37.9	33.4	37.7	33.3	40.1	34.2	32.4
US-W	Daily	312.6	230.2	294.0	223.7	318.3	255.1	245.5
	Pentad	132.4	117.8	127.3	116.0	134.6	120.7	115.7
	Monthly	60.3	48.2	58.5	46.2	62.2	47.9	44.5
				Relative bias (%)				
U.S.	Daily	1.4	2.8	1.4	3.9	0.2	3.0	0.5
	Pentad	0.5	0.5	0.4	0.8	0.5	0.8	0.5
	Monthly	0.1	0.3	0.1	0.3	0.1	0.2	0.2
US-W	Daily	5.2	7.9	5.4	12.3	0.7	7.4	−0.5
	Pentad	2.8	5.8	3.3	5.7	1.0	3.4	1.3
	Monthly	0.7	−0.4	1.7	0.1	0.1	0.4	0.4

United States and a subregion over the western U.S. mountainous areas (US-W; 35°–45°N, 105°–115°W). Table 1 presents the validation results for daily, pentad, and monthly precipitation.

As expected, the accuracy of the interpolated values improves with increasing temporal scale. For the OI-based interpolation of precipitation over the entire United States, the correlation, the relative rms error, and the relative bias are 0.848, 171%, and 0.5%, respectively, for daily accumulation; 0.902, 75%, and 0.5% for pentad; and 0.929, 32%, and 0.2% for monthly scale. The agreement between the interpolated values and the independent observations for the western U.S. region, however, is much worse than that for the entire United States. The OI-based interpolation of daily precipitation shows a correlation of 0.738 and a relative rms error of 246% for this region.

When the total value of precipitation is defined by directly interpolating observations from nearby stations (i.e., the direct approach), the three distance-weighting algorithms, the Cressman, Barnes, and Shepard, yield very similar statistics, showing reasonable agreements of the interpolated values with those from the withdrawn gauge observations. The correlations between the interpolated monthly precipitation and the gauge observations are 0.902, 0.902, and 0.890, respectively, for the Cressman, Barnes, and Shepard algorithms.

When the total value of precipitation is defined as the summation of the long-term mean and the anomaly interpolated from nearby stations (i.e., the indirect approach), the agreements in pattern for the analyses produced by the three distance-weighting algorithms, as measured by the correlation and the relative rms error, exhibit consistent improvements when compared with those based on the direct approach. The correlation coefficients for the monthly precipitation based on the indirect approach over the entire United States are 0.924, 0.925, and 0.921, respectively, as compared with 0.902, 0.902, and 0.890 for the analyses based on the direct approaches. Particularly significant is the improvement over the western mountainous areas, for which the total precipitation receives substantial influences from topography. The correlation coefficients for monthly precipitation are 0.863, 0.874, and 0.864 for the analyses based on the indirect approach by the Cressman, Barnes, and Shepard algorithms, respectively, as compared with 0.780, 0.793, and 0.764 through the direct approach.

The agreement in magnitude, however, degrades for the indirect approach based on the three distance-weighting techniques, especially for applications to precipitation fields with large variability (shorter timescale and/or significant topographical effects). The relative biases for indirectly defined analyses of daily precipitation over the western mountainous areas are 7.9%, 12.3%, and 7.4% for the Cressman, Barnes, and Shepard methods, as compared with 5.2%, 5.4%, and 0.7% based on the direct approach. Positive biases in the indirect approach may be caused mainly by the limitation that the total precipitation has to be set to 0 when the summation of the climatology and the anomaly is negative.

The point values of precipitation interpolated by the OI algorithm, which by definition is an indirect approach, show equivalently good pattern agreements (correlation and rms error) when compared with those produced by the three distance-weighting algorithms through the indirect approach. However, the OI-based

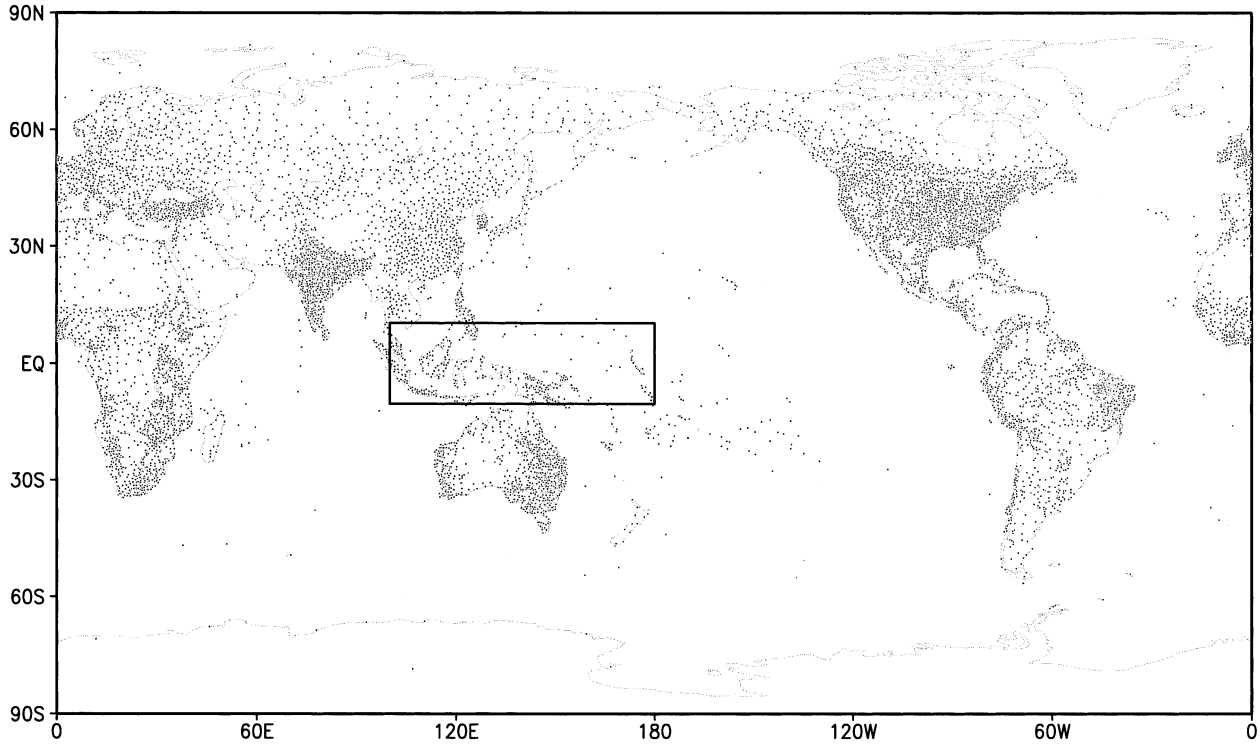

FIG. 2. Distribution of the gauge stations with 10-yr-or-longer recording periods from 1951 to 1990 collected in the GHCN2 and CAMS datasets. The labeled box indicates the subregion of the Indonesian islands (see text).

interpolation yields smaller biases, in general, than those based on the distance-weighting techniques, especially for shorter timescales. The correlation coefficient and relative bias for the interpolation of daily precipitation over the entire United States are (0.848, 0.5%) for the OI, and they are (0.865, 2.8%), (0.860, 3.9%), and (0.842, 3.0%) for the Cressman, Barnes, and Shepard algorithms through the indirect approach. In considering agreements in both the pattern and magnitude, the OI-based algorithm presents superior performance in interpolating precipitation for all temporal accumulation scales and over most regions in the United States.

d. Intercomparison over the global land areas

To understand further the performance of the four candidate algorithms in defining precipitation fields with various spatial variability, a similar intercomparison was conducted for land areas over the globe using station observations of monthly precipitation. A combination of two individual datasets was used: the version-2 dataset of the GHCN, compiled by NCDC, and the CAMS, maintained by the NOAA CPC.

As discussed in detail in the following section, the combined dataset of GHCN2 and CAMS is a collection of monthly precipitation reports from over 23 000 stations for the period from 1948 to 2000. The long-term mean of monthly precipitation was first calculated for each of the 17 493 stations with records greater than 10 years during the period from 1951 to 1990. This long-term mean was then used in the OI-based analysis as the first guess and in the indirect approach using the three distance-weighting algorithms to define total precipitation from the interpolated anomalies. Figure 2 shows the distribution of gauges for which the long-term mean of monthly precipitation is defined. Western Europe, India, East Asia, the east and west coastal areas of Australia, the United States, and some coastal areas in Africa and South America are covered by relatively dense gauge networks; precipitation over Amazonia and tropical Africa is observed by sparsely distributed gauges with notable gaps.

The number of gauge stations reporting precipitation varies with time, with a maximum of about 17 000 around 1970 (see Fig. 3 and discussion in the next section for details). To examine the performance of the algorithms in constructing precipitation fields observed by gauge networks of various density, the gauge data for January of 1970 were selected as the target month of the global intercomparison. In total, there are 16 682 gauge stations reporting monthly precipitation for that month.

Validation statistics were calculated for the entire global land area and for a subregion over the Indonesian islands (IND.ISL; 10°S–10°N, 100°E–180°). Table 2 shows the correlation, the relative rms error, and the relative bias of the interpolated point values of monthly precipitation against the withdrawn independent gauge

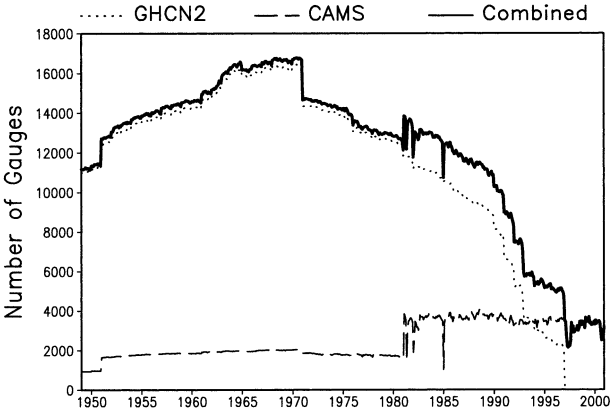

FIG. 3. Time series of numbers of gauges reporting monthly precipitation in GHCN2 (dotted line), CAMS (dashed line), and their combination (solid line).

observations for the entire global land area and the Indonesian islands.

As found in the regional intercomparison over the United States, interpolation of anomalies yields better pattern agreement than direct derivation from station observations of total precipitation for the three distance-weighting algorithms. Over the entire global land area, correlation coefficients for the indirect approach are 0.877, 0.857, and 0.895 for the Cressman, Barnes, and Shepard algorithms, and those for the direct approach are 0.760, 0.801, and 0.831, respectively (Table 2, top). Significant improvements in pattern agreement are observed over the Indonesian Maritime Continent region, where the topographic effects on the spatial distribution of total precipitation are substantial. The correlation coefficients are as low as 0.4–0.5 when directly interpolating total precipitation. These statistics go up to over 0.7 for the indirect approaches.

The indirect approach, defining total precipitation by adding the interpolated anomaly to the climatology, however, tends to produce analyses of monthly precipitation with larger bias than those derived through the direct approach, especially for those using the Barnes and Shepard algorithms. The bias of interpolated values over the entire global land area is −0.3% for the direct approach using the Barnes algorithm; it is 1.7% for the indirect approach (Table 2, bottom).

The OI-based interpolation of monthly precipitation yields good agreements in both the pattern and magnitude. Both the pattern agreement, measured by the correlation and the relative rms error, and the magnitude agreement, measured by bias, are equally as good as or marginally superior to those for the three distance-weighting algorithms with the indirect approach over the entire global land area and most of the subregions.

The same intercomparison was repeated for the same month using gauge reports from 50%, 25%, and 10% of all available observations so as to examine the impact of varying gauge network density on the accuracy of the interpolated values of monthly precipitation. Table 3 shows the results for the entire global land area. The interpolation accuracy (correlation and relative rms error) degrades with the decreasing gauge network density for all algorithms and for both the direct and indirect approaches of interpolation. The indirect interpolation approach for the three distance-weighting algorithms and the OI scheme, however, appears more successful in interpolating station observations with less density. For the Barnes algorithm with the direct interpolation approach, the correlation decreases from 0.801 based on all available gauges to 0.582 based on 10% gauges, and it changes from 0.857 to 0.750 if the indirect approach is applied (Table 3, top).

Although the regional and global intercomparisons of the interpolation algorithms were conducted using limited data and although fluctuation and uncertainty exist in the statistics, three conclusions can be drawn from the results presented above:

1) When distance-weighting algorithms, such as those of Cressman (1959), Barnes (1964), and Shepard (1968), are applied, the issue that most influences the interpolation accuracy is not the interpolation algorithm (e.g., weighting function, searching distance) but the interpolation strategy (direct or indirect approaches). Once the strategy is determined, the accuracy of the interpolated fields is relatively insensitive to the algorithm used.

TABLE 2. Cross-validation results for monthly precipitation over the global land areas.

	Cressman		Barnes		Shepard		
Regions	Direct	Indirect	Direct	Indirect	Direct	Indirect	OI Indirect
			Correlation				
Globe	0.760	0.877	0.801	0.857	0.831	0.895	0.896
IND.ISL	0.402	0.749	0.445	0.722	0.544	0.761	0.772
			Relative rms error (%)				
Globe	92.1	65.9	80.7	71.6	75.9	60.6	60.5
IND.ISL	67.3	45.5	59.3	49.3	58.2	44.5	42.8
			Relative bias (%)				
Globe	3.5	2.3	−0.3	1.7	0.4	0.9	0.5
IND.ISL	0.5	0.7	−1.2	3.3	1.0	1.2	0.1

TABLE 3. Cross-validation results for gauge networks of reduced density.

Density of gauges	Cressman		Barnes		Shepard		
	Direct	Indirect	Direct	Indirect	Direct	Indirect	OI indirect
				Correlation			
Full	0.760	0.877	0.801	0.857	0.831	0.895	0.896
50%	0.732	0.874	0.711	0.851	0.810	0.891	0.900
25%	0.692	0.847	0.683	0.835	0.774	0.873	0.880
10%	0.635	0.741	0.582	0.750	0.716	0.832	0.811
				Relative rms error (%)			
Full	92.1	65.9	80.7	71.6	75.9	60.6	60.5
50%	94.5	66.4	101.1	72.6	80.1	61.6	59.3
25%	103.5	74.2	105.4	76.9	87.8	67.8	66.1
10%	105.6	96.0	111.8	90.4	94.6	75.7	81.4
				Relative bias (%)			
Full	3.5	2.3	−0.3	0.9	0.4	0.9	0.5
50%	−0.3	0.9	0.8	0.9	0.4	0.9	0.3
25%	2.6	1.4	1.0	0.4	−0.5	0.4	0.0
10%	−0.4	1.4	−11.3	0.0	−2.0	0.0	−0.6

2) For the distance-weighting algorithms, the indirect approach, which defines the total precipitation by adding the interpolated anomaly to the climatology, yields substantial improvements in pattern agreement but degradation to some extent in magnitude agreement when compared with the direct approach, which creates precipitation analysis by directly interpolating observations of total precipitation from nearby stations.
3) In most cases, the OI algorithm of Gandin (1965) exhibits superior performance in constructing precipitation analysis when compared with the distance-weighting algorithms.

3. Construction of global analysis from 1948

Based on the results for the intercomparison of the four algorithms, it was decided to construct the analysis of monthly precipitation for the period from 1948 to 2000 by interpolating gauge observations using the OI algorithm.

Two sets of gauge observations of monthly precipitation were used jointly to construct the analysis. One is GHCN2 dataset, (Peterson and Vose 1997) and the other is the CAMS dataset (Ropelewski et al. 1985). These two datasets were described in section 2d.

The GHCN2 dataset contains quality controlled monthly precipitation data at over 23 000 stations over global land areas for an extensive period from the late 1690s to 1996. Not every station, however, has a precipitation observation for every month. As shown in Fig. 3, the number of GHCN2 gauge stations (dotted line) reporting precipitation increases from about 11 000 in 1948 to about 17 000 in 1970 and then declines down to about 3000 in 1996. In addition, the spatial distribution (Fig. 4, top) of the gauges is uneven in the dataset. Western Europe; India; eastern China; coastal areas of Australia, Africa, and South America; and the United States are covered by relatively dense networks, but there are few gauges over central and northern Africa, central Asia, and the Amazon basin.

The monthly gauge observations in CAMS are based on two data sources, one obtained from the historical records collected and edited by NCAR for the period before 1981 and the other accumulated on a real-time basis from the GTS for the period after 1981 (Ropelewski et al. 1985). Although only an automatic screen check and gross quality control are conducted for monthly precipitation data, the CAMS relies heavily on ''CLIMAT'' station reports, which are of high quality. The number of gauge stations reporting precipitation (Fig. 3, dashed line) is roughly 1900 and 3500 for the two periods, respectively. In comparison with the GHCN2 dataset, the CAMS gauge network (Fig. 4, middle) is denser over Europe and central and East Asia but is much sparser over the rest of the global land areas.

Although adding more gauges does not necessarily lead to improvement in the quality of the resulting precipitation analysis, the extremely poor coverage of the gauge observations in both datasets in some of the land areas and the unavailability of the GHCN2 gauge observations for recent years required combination of the two gauge datasets (Fig. 4 bottom) so as to construct the global precipitation analysis of the highest possible quality. This is done by comparing the station identification and location and the time series of monthly precipitation. As a result, 1976 stations among the total of 5652 stations in the CAMS dataset were determined to be independent and were added to the GHCN2 dataset. Figure 3 shows the time series of the number of gauges available in the combined gauge dataset (solid line) from 1948 to 2000. Inclusion of the CAMS gauge data greatly improves the gauge coverage for the period after 1980 and makes it possible to construct the analysis for recent years.

As described in section 2a, the OI technique of Gan-

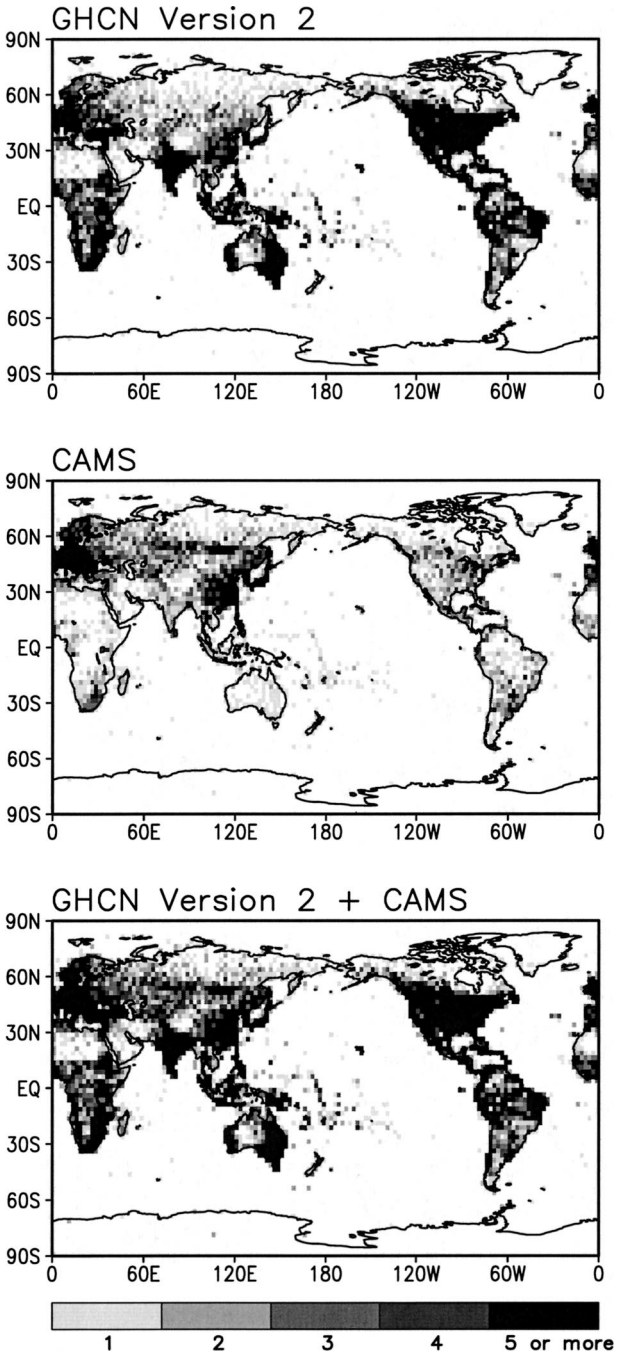

FIG. 4. Distribution of numbers of gauges in 2.5° latitude/longitude grid boxes available from (top) GHCN2, (middle) CAMS, and (bottom) their combination.

din (1965) defines the analysis value at a target grid point by modifying a first-guess value at the grid point with weighted mean of the differences between the observed values and the first-guess values at station locations within a search distance. Long-term means of monthly precipitation were used as the first guess in defining our gauge-based analysis. First, long-term means of monthly precipitation were calculated at gauge locations from the station observations collected in the combined dataset described above. Long-term means of monthly precipitation at target grid points were then defined by interpolating those over gauge locations, using the algorithm of Shepard (1968). Raising the minimum years of observations required in calculating long-term means for each station improves the quantitative accuracy and stability at station locations, but it reduces the total number of qualified stations and therefore degrades the spatial representativeness of the resulting first-guess fields. After examining the effects of temporal and spatial sampling on the accuracy of interpolated fields of long-term means, Willmott et al. (1996) and Hulme and New (1997) both concluded that including additional stations yields more improvements than the degradation caused by reducing the number of years that define the climatology. To determine the best threshold for defining the long-term means, we tried several options ranging from 5 to 40 yr and found that a base period of 40 yr (1951–90) with at least 10 years of observations results in analyses of long-term means of monthly precipitation with balanced spatial representativeness and quantitative accuracy. Of over 23 000 stations available from the combined gauge dataset, 17 493 satisfy the requirements (Fig. 2). Note that reasonable gauge coverages are available over the Amazon basin and tropical Africa when using these criteria.

An analysis of monthly precipitation over global land areas was then constructed by interpolating the monthly gauge observations in the combined dataset by using the OI algorithm with the monthly long-term mean precipitation as the first guess. Following GPCC and Xie et al. (1996), the gauge-based analysis was first defined at grid points of 0.5° latitude/longitude intervals over the global land areas. Areal mean precipitation over each 0.5° latitude/longitude grid box was then defined as the arithmetic mean of the point values at its four corners. The areal mean precipitation over each 2.5° latitude/longitude grid box was calculated as the mean of the 25 0.5° grid boxes weighted by their areas. This gauge-based analysis of monthly precipitation was defined over global land areas for the 53-yr period from 1948 to 2000. Figure 5 is an example of the monthly precipitation analysis (top) and the number of gauges (bottom) in each 2.5° latitude/longitude grid box for January of 1970. Precipitation zones associated with the major convection centers are observed over the Southern Hemisphere portions of Africa, South America, and the Indonesian maritime continent. Additional rain areas appear over the eastern coastal areas of the United States, over western Europe, and over the northern and eastern coasts of Australia.

Several shortcomings exist in the gauge-based analysis of monthly precipitation because of problems and imperfections in the input station observations and in the interpolation algorithms. No modification is applied in this work to correct the bias in the station observa-

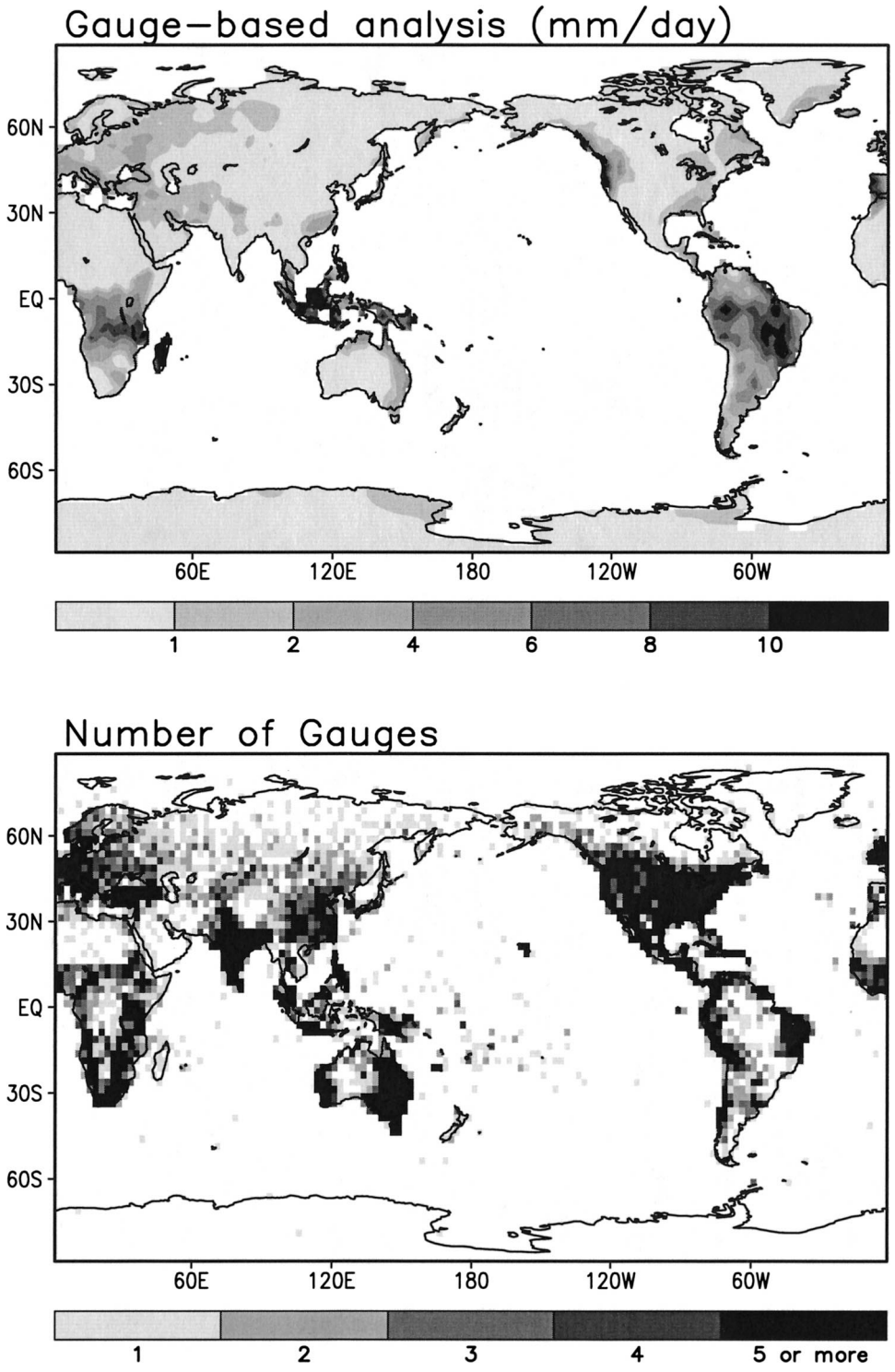

FIG. 5. (top) Gauge-based analysis of global monthly precipitation (mm day^{-1}) for Jan 1970 and (bottom) the number of gauges in each 2.5° latitude/longitude grid box used to define the analysis.

tions caused by the gauge undercatchments or to remove the discontinuity introduced from the changes in instruments. Although the gauge undercatchment is a formidable error source, especially during winter and early spring over mid- and high-latitude land areas, it is very difficult, if not impossible, to correct it month by month without coincident information about the precipitation type and the wind speed. Accurate removal of discon-

tinuities also requires additional information about the time of the instrument changes, which is not generally available. The analysis ideally should be defined using station reports at the same stations throughout the target period to ensure temporal homogeneity. Only a small portion of the stations have full coverage for the 53-yr period, and monthly precipitation data from all reporting stations were used to create the analysis. Because of these factors, the resulting gauge-based analysis of monthly precipitation may result in underestimates of monthly precipitation over areas and for seasons with substantial snowfall, and its time series may also be contaminated with artificial long-term variability and trend. Another potential source of error of the gauge-based analysis lies in the handling of the orographic effect on precipitation. Although the OI algorithm is able to reproduce the precipitation distribution over mountainous areas better by adding interpolated anomalies to a monthly climatology based on a dense network of historical observations (see section 2), systematic error may occur over areas for which historical station observations are not available and nearby stations are unrepresentative. This may cause serious problems in some applications that require accurate estimates of heavy rainfall over a region with mountainous topography (Nijssen et al. 2001).

As mentioned in the introduction, our final goal is to create an analysis of monthly precipitation for an extended period from 1948 to the present by interpolating gauge observations over land and by EOF reconstruction of historical records over ocean. The gauge-based analysis of monthly precipitation described above is the land portion of this extended analysis. For convenience, we will label the extended analysis over the entire globe as PREC (precipitation reconstruction) and the gauge-based analysis over land and the reconstruction over ocean as PREC/L and PREC/O, respectively.

4. Validation and application of the gauge analysis

The new 53-yr monthly analysis of global land precipitation (PREC/L) was compared with four published gauge-based precipitation analyses and was used to examine the seasonal and interannual variations of large-scale precipitation over the land areas of the globe.

The four datasets of gauge-based monthly precipitation to be compared here are those of New et al. (2000), Dai et al. (1997), the GPCC gauge-based analysis (Rudolf 1993), and Xie et al. (1996). The monthly analysis of New et al. (2000) was originally constructed on a 0.5° latitude/longitude grid over global land areas for the period from 1901 to 1996. However, only the precipitation data averaged on 5.0° latitude/longitude grid boxes were available free of charge and were used in this comparison. The dataset of Dai et al. (1997) contains a time series of monthly precipitation anomalies over the global land areas. Total precipitation was defined by adding these anomalies to the monthly climatology of Shea (1986). All three of the remaining gauge-based analyses were created on a 2.5° latitude/longitude grid over global land areas and covered time periods from 1900 to 1988 in Dai et al. (1997), from 1986 to the present in the GPCC, and from 1971 to 1996 in Xie et al. (1996). Because of the different temporal coverages in the various datasets, two sets of comparisons were conducted: one among our 53-yr analysis (PREC/L), New et al. (2000), and Dai et al. (1997) for the period from 1948 to 1988, and the other among the PREC/L, GPCC, and Xie et al. (1996) for the period from 1986 to 1995. Only the 2.5° latitude/longitude grid boxes with 50% or more land coverage were included in the comparison.

Figure 6 shows the spatial distributions of annual mean precipitation averaged for the 41-yr period from 1948–88 as obtained from the PREC/L, New et al. (2000), and Dai et al. (1997) (left column) and for the 10-yr period from 1986 to 1995 from the PREC/L, GPCC, and Xie et al. (1996) (right column). Several blank areas are observed in the analyses of New et al. (2000) and Dai et al. (1997) due to their assignments of the analysis over no-gauge grid boxes as missing. Otherwise, the annual averages for the 41-yr period obtained from the three datasets, all of which were defined through adding interpolated anomalies to the climatology (the indirect approach), show very similar spatial distribution patterns, characterized by rainbands associated with the intertropical convergence zone, the South Pacific convergence zone, and midlatitude storm tracks.

The spatial distribution patterns of annual means for the 10-yr period are also very similar among the PREC/L, GPCC, and Xie et al. (1996). Noticeable differences in magnitude, however, are observed over India, the Amazon, and several other tropical land areas, for which our new 53-yr analysis (PREC/L) presents heavier rainfall than those in the GPCC and Xie et al. (1996). A detailed examination (results not shown here) revealed that most of these differences are caused by biases in the analyses of the GPCC and Xie et al. (1996) within regimes of precipitation with large spatial variations observed by sparse gauge networks. The direct approach, like those of the GPCC and Xie et al. (1996), defines total precipitation by interpolating observations from surrounding stations, resulting in biases for cases where there are systematic differences between the precipitation values at the target grid point and its surrounding stations. The indirect approach, like the OI used in the PREC/L, is able to reproduce better the magnitude by defining the long-term means using observations from dense gauge networks for a historical period and then modifying them by anomalies from stations with reports for the target month.

Table 4 shows the precipitation averaged over the two hemispheres and the whole globe and for the seasonal periods of December–January–February (DJF) and June–July–August (JJA), as well as the entire year. Only

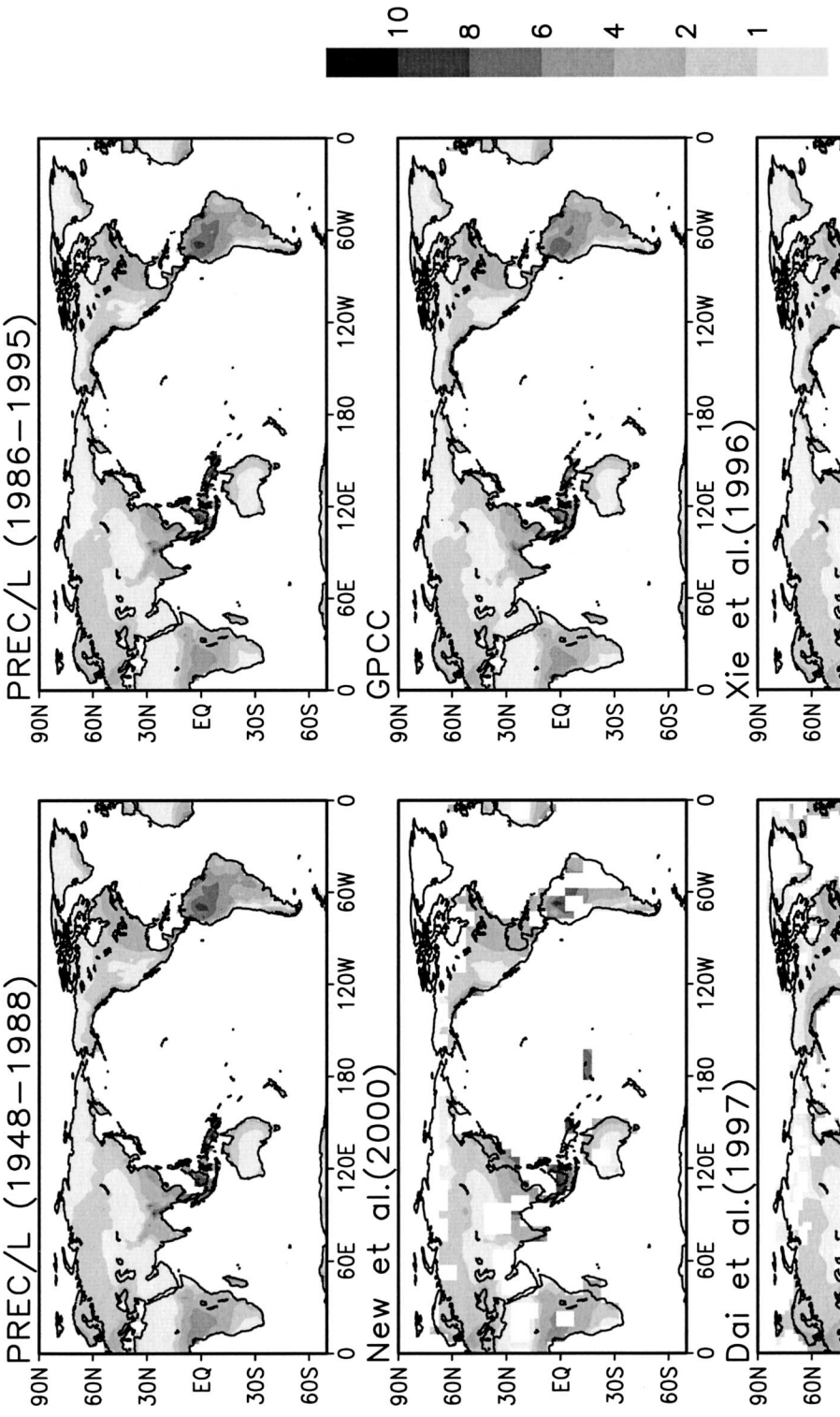

FIG. 6. Mean annual precipitation (mm day^{-1}) defined for (left) a 41-yr period (1948–88) from (top) our gauge-based analysis (PREC/L), (middle) New et al. (2000), and (bottom) Dai et al. (1997); and (right) for a 10-yr period (1986–95) from (top) our new gauge-based analysis (PREC/L), (middle) the GPCC, and (bottom) Xie et al. (1996).

TABLE 4. Mean precipitation (mm day^{-1}) over 2.5° latitude/longitude grid boxes with 50% or more land.

Sources	DJF			JJA			Annual		
	Globe	NH	SH	Globe	NH	SH	Globe	NH	SH
				(1948–88)					
PREC/L	1.97	1.03	4.34	2.59	3.15	1.16	2.18	1.96	2.73
New et al.(2000)	2.01	1.05	4.41	2.61	3.18	1.19	2.21	1.99	2.76
Dai et al.(1997)	1.99	1.05	4.37	2.55	3.06	1.25	2.18	1.94	2.78
				(1986–95)					
PREC/L	2.06	0.99	4.70	2.46	2.93	1.32	2.18	1.84	3.00
GPCC	1.97	0.93	4.51	2.29	2.73	1.21	2.06	1.74	2.84
Xie et al.(1996)	2.01	0.92	4.70	2.25	2.68	1.19	2.05	1.70	2.89

the grid boxes with analyzed values of precipitation available from all datasets are included in the calculation of the long-term mean (Table 4) and annual cycle (Fig. 7). In general, the long-term averages obtained from our PREC/L analysis agree very well with those of New et al. (2000) and Dai et al. (1997). The annual mean precipitation over the global land areas is 2.18 mm day^{-1} while those over the Northern and Southern Hemispheres are 1.96 and 2.73 mm day^{-1}, respectively. The PREC/L, however, exhibits larger values of average precipitation than those defined from the gauge-based analyses of GPCC and Xie et al. (1996).

Figure 7 shows the annual cycle of precipitation averaged over the land areas of the two hemispheres and

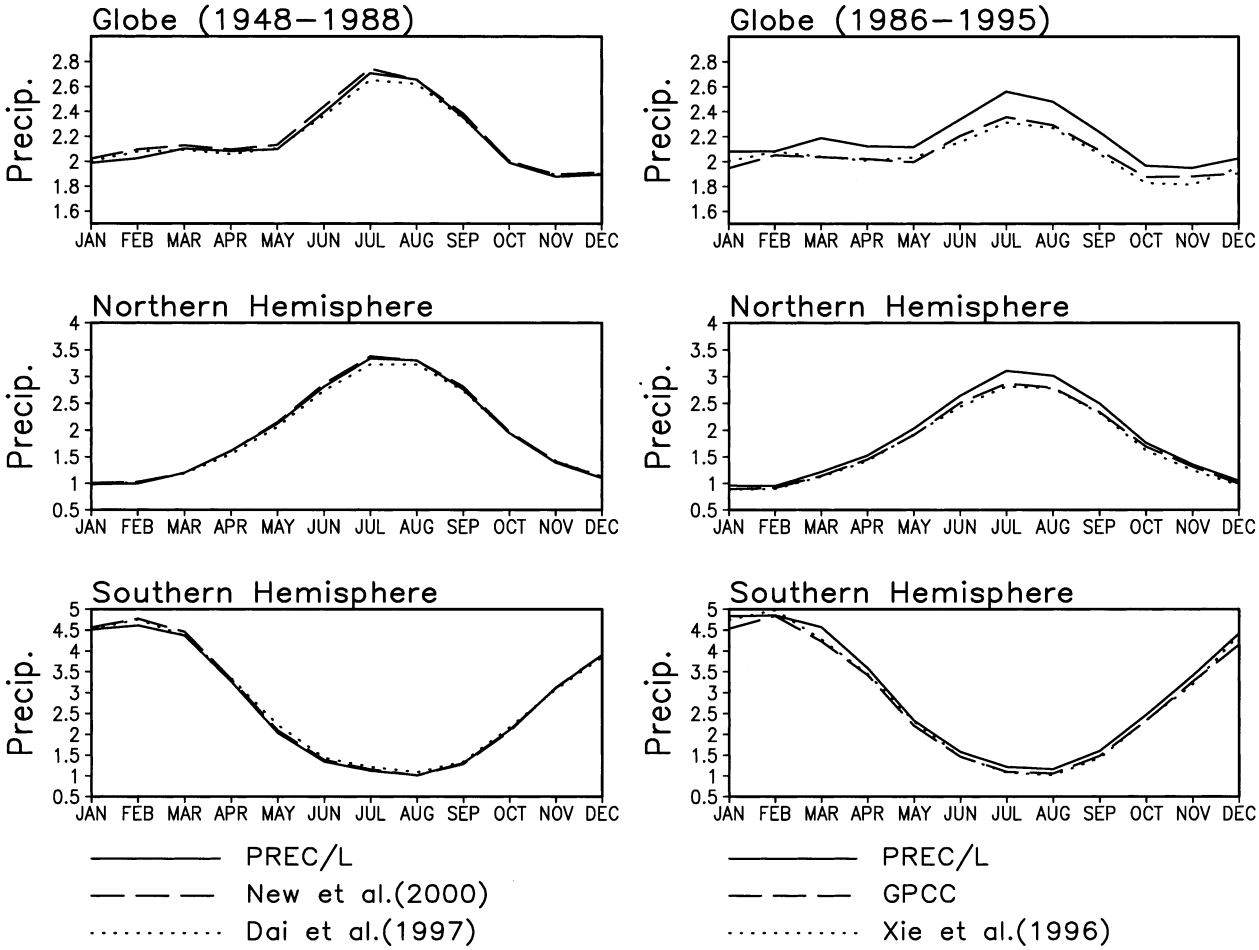

FIG. 7. Annual cycles of mean precipitation (mm day^{-1}) over 2.5° latitude/longitude grid boxes with at least 50% land coverage over (top) the entire globe, and the (middle) Northern and (bottom) Southern Hemispheres as defined by several datasets. (left) The results for the 41-yr period from 1948 to 1988. (right) Those for the 10-yr period from 1986 to 1995.

the entire globe. The annual cycles averaged for the 41-yr period from 1948 to 1988 are very close for the three sets of gauge-based analyses examined here. In particular, our PREC/L and the New et al. (2000) show almost exactly the same curve over the Northern Hemisphere. Over the Northern Hemisphere, the annual cycle of precipitation peaks in July–August; over the Southern Hemisphere, the peak is found in January–February. The annual cycles defined from the Dai et al. (1997) dataset show somewhat smaller peak values than those from the other two during boreal summer. The annual cycles for the 10-yr period are very close for the GPCC and the Xie et al. (1996), reflecting the fact that the two of them were created by the same algorithm and from similar gauge networks. These annual cycles, however, are consistently smaller than that from our PREC/L, especially for months with heavier precipitation. As discussed before, this difference can likely be attributed to the biases in the GPCC and the Xie et al. (1996) datasets.

One of the potential applications of this 53-yr gauge-based monthly analysis is the examination of interannual variability of large-scale precipitation over the land areas of the globe. Here, as an example of such applications, the interannual variability associated with the El Niño–Southern Oscillation (ENSO) phenomenon is described. Although there are many ways to define an ENSO episode, we adopted a simple and objective approach in which a cold/warm ENSO episode is declared for a season if the SST anomaly over the Niño-3 area (5°N–5°S, 150°–90°W) is ≤ -0.5°C or ≥ 0.5°C. This classification results in 45 cold- and 52 warm-event seasons, more than 10 events for each season, and 99 normal seasons for the period from 1950 to 1998. These classification results are in general agreement with those of Ropelewski and Halpert (1989), which are based on the Southern Oscillation index. Particularly noticeable is that, with 10 or more warm or cold events observed for each season during the period, it becomes possible to quantify the precipitation variability observed among different ENSO episodes with increased confidence using this new gauge-based analysis.

To this end, mean precipitation was composited for the warm and cold events as determined above. The departure of these composite fields from overall mean depicts shifts in precipitation that are associated with the ENSO cycle. In Fig. 8 (top), we show these departures for the DJF season, after having normalized them by the standard deviation of seasonal means over the 53-yr period. As expected, the pattern of anomalous precipitation is similar to the observations of Ropelewski and Halpert (1987, 1989), Janowiak (1988), Janowiak and Arkin (1991), Dai et al. (1997), and Dai and Wigley (2000). By normalizing the anomalies, however, other high-latitude signals appear that are diminished when analyzing nonnormalized data. For example, wetter-than-normal conditions tend to be observed in Mexico and the southern parts of the United States, during warm episodes, and normal to anomalously dry conditions tend to prevail there during cold events.

Because the dataset extends over 53 years, an ample number of ENSO events exist in the record so that we can assess variability in precipitation among the events with reasonable confidence. To examine this variability, we plot the standard deviation of the normalized anomalies among the cold- and warm-event DJF seasons (Fig. 8, bottom). Note that we plot these statistics only in regions for which the seasonal mean precipitation exceeds 0.3 mm day^{-1}; that is, there is no shading where the mean precipitation is less than 0.3 mm day^{-1}. In general, there is more variability in precipitation during cold episodes than during warm episodes. To be specific, this is observed over much of Australia, southern Africa, the eastern United States, and in northern South America. In contrast, more variability is observed among warm episodes in southeastern China, parts of Indonesia, and Mexico than among cold events.

5. Summary and future work

Analyses of monthly precipitation have been constructed on a 2.5° latitude/longitude grid over global land areas for the period from 1948 to 2000 by interpolating gauge observations at over 17 000 stations collected in two individual datasets: the Global Historical Climatology Network, version 2, of NOAA/NCDC and the Climate Anomaly Monitoring System of NOAA/CPC.

First, an intercomparison was conducted among four published algorithms [Cressman (1959); Barnes (1964), Shepard (1968), and the optimal interpolation method of Gandin (1965)] to examine their performance in interpolating gauge observations of daily, pentad, and monthly precipitation. The results showed that all of these algorithms perform better in interpolating anomaly of precipitation than the total values and that the OI scheme yields superior and stable statistics in most cases when compared with the other three algorithms. The correlation between the analysis values and the withdrawn independent station observations is about 0.8 and the bias is almost 0 for interpolation of monthly precipitation using the OI algorithm.

Based on these results, the OI algorithm was selected to construct the analysis of monthly precipitation over global land areas for a 53-yr period from 1948 to 2000. First, quality control was conducted for the station observations of monthly precipitation collected in the GHCN2 and CAMS datasets to remove erroneous and questionable values and to remove redundancy. Climatologies of monthly precipitation were then defined for stations with records of 10 years or longer during the 40-yr period from 1951 to 1990 and were used to construct analyzed fields of monthly precipitation climatology over the global land areas. Last, analyses of monthly precipitation were created by interpolating the

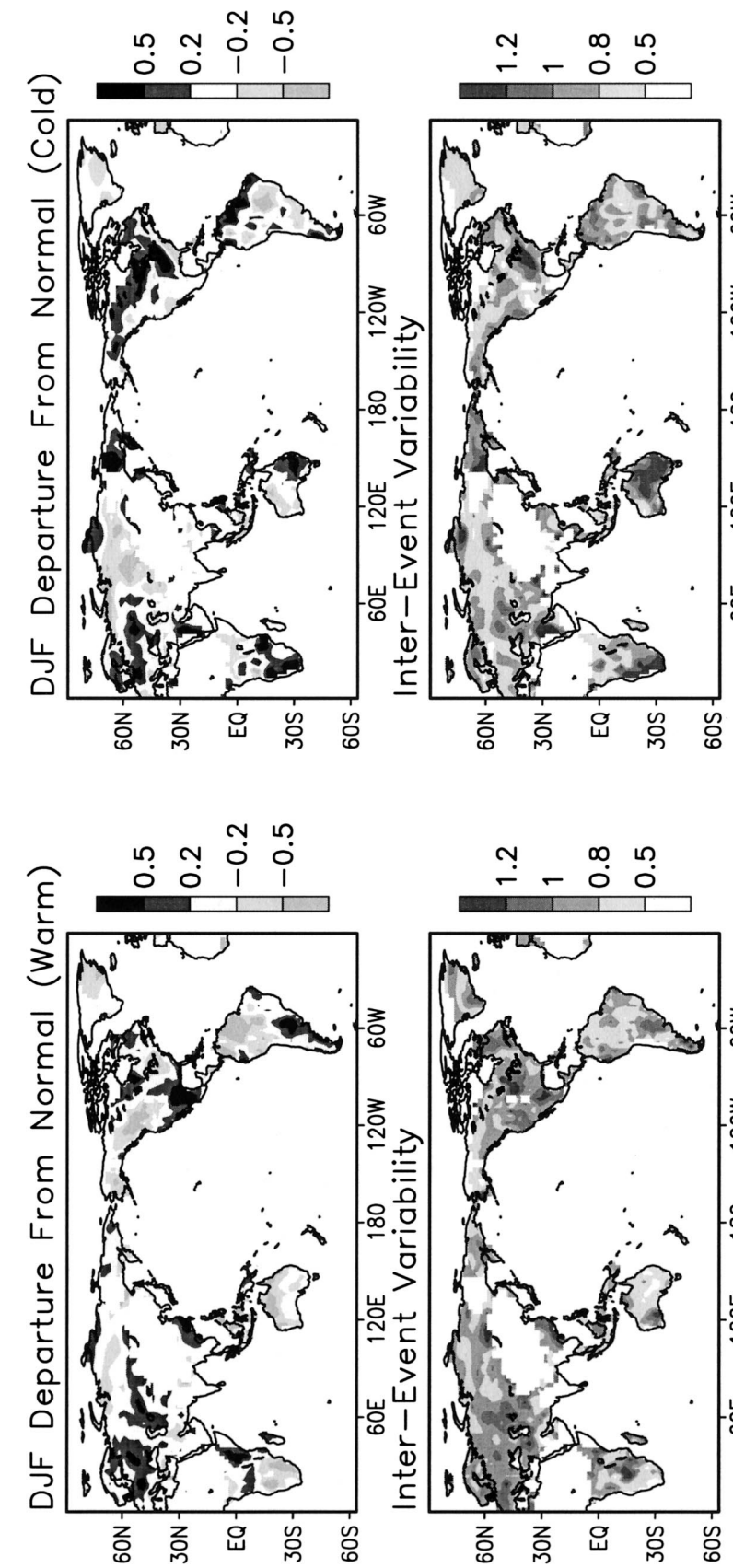

FIG. 8. (top) The normalized anomalous precipitation for DJF for (left) warm-event and (right) cold-event ENSO composites from the 53-yr PREC/L analysis. (bottom) The precipitation variability among (left) warm and (right) cold events. This ''interevent'' variability is defined as the standard deviation of the normalized anomalies for each year in the composites in the top panel.

quality-controlled monthly station data via the OI algorithm using the climatology as the first guess.

The 53-yr gauge-based analysis of monthly precipitation over land (PREC/L) was compared with four published gauge-based datasets: New et al. (2000), Dai et al. (1997), the GPCC, and Xie et al. (1996). Close agreements in both the phase and the magnitude of the mean annual cycles were observed among the PREC/L, Dai et al. (1997), and New et al. (2000), in which total precipitation was defined by adding interpolated anomalies to the climatology. The PREC/L, however, exhibits larger precipitation amounts when compared with those of GPCC and Xie et al. (1996), in which analysis values were calculated by direct interpolation of precipitation observations over surrounding gauge stations. PREC/L was used to examine the interannual variations in large-scale precipitation and showed ENSO-related anomaly patterns similar to those reported by previous studies.

The work reported here is an integral part of our efforts to construct an analysis of monthly precipitation over the entire globe for an extended period from 1948 to the present. Known as precipitation reconstruction, this monthly analysis of precipitation consists of gauge-based analysis over land and the reconstructed fields over ocean. The 53-yr gauge-based analysis (PREC/L) described in this paper is the land portion of the analysis over the entire globe (PREC). Work is under way to define monthly analysis of precipitation over oceanic areas (PREC/O) for the same period as PREC/L. This oceanic precipitation analysis will be derived by EOF reconstruction that combines the spatial variability distribution patterns defined from satellite observations for later years with temporal variation information taken from historical records of gauge and ship observations. This technique was developed by Smith et al. (1996), who successfully applied it to construct sea surface temperature analyses for the presatellite era.

Acknowledgments. We thank S. Curtis, R. Bruno, A. Dai, A. Gruber, M. Halpert, G. J. Huffman, M. Hulme, M. New, D. J. Seo, T. Smith, E. Yarosh, and X.-B. Zhang for their invaluable discussions and comments throughout this study. We are also indebted to the editor and three anonymous reviewers, whose comments lead to improvements to the original manuscript. The 53-yr gauge-based analysis of global land precipitation was available through anonymous FTP at ftp.ncep.noaa.gov/pub/precip/50-yr. In addition, an analysis on 0.5° latitude/longitude is also available on request.

REFERENCES

Adler, R. F., A. J. Negri, P. R. Keehn, and I. M. Hakkarinen, 1993: Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.,* **32,** 335–356.

——, G. J. Huffman, and P. R. Keehn, 1994: Global rain estimates from microwave adjusted geosynchronous IR data. *Remote Sens. Rev.,* **11,** 125–152.

Arkin, P. A., and P. Xie, 1994: The Global Precipitation Climatology Project: First Algorithm Intercomparison Project. *Bull. Amer. Meteor. Soc.,* **75,** 401–419.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.,* **3,** 396–409.

Bradley, R. S., P. M. Kelly, P. D. Jones, C. M. Goodess, and H. F. Diaz, 1985: A climatic data bank or Northern Hemisphere land areas, 1851–1980. U.S. Department of Energy Tech. Rep. 017.

——, H. F. Diaz, J. K. Eischeid, P. D. Jones, P. M. Kelly, and C. M. Goodess, 1987: Precipitation fluctuations over Northern Hemisphere land areas since the mid-19th century. *Science,* **237,** 171–175.

Bussieres, N., and W. Hogg, 1989: The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. *Atmos.–Ocean,* **27,** 521–541.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374.

Creutin, J. D., and C. Obled, 1982: Objective analysis and mapping techniques for rainfall fields: An objective comparison. *Water Resour. Res.,* **18,** 413–431.

Curtis, S., and R. F. Adler, 2000: ENSO indices based on patterns of satellite-derived precipitation. *J. Climate,* **13,** 2786–2793.

Dai, A., 2001a: Global precipitation and thunderstorm frequencies. Part I: Seasonal and interannual variations. *J. Climate,* **14,** 1092–1111.

——, 2001b: Global precipitation and thunderstorm frequencies. Part II: Diurnal variations. *J. Climate,* **14,** 1112–1128.

——, and T. M. L. Wigley, 2000: Global patterns of ENSO induced precipitation. *Geophys. Res. Lett.,* **27,** 1283–1286.

——, I. Y. Fung, and A. D. Del Genio, 1997: Surface observed global land precipitation variations during 1900–88. *J. Climate,* **10,** 2943–2962.

——, T. M. L. Wigley, B. A. Boville, J. T. Kiehl, and L. E. Buja, 2001: Climates of the twentieth and twenty-first centuries simulated by the NCAR climate system model. *J. Climate,* **14,** 485–519.

Diaz, H. F., R. S. Bradley, and J. K. Eischeid, 1989: Precipitation fluctuations over global land areas since the late 1800s. *J. Geophys. Res.,* **94,** 1195–1210.

Doherty, R. M., M. Hulme, and C. G. Jones, 1999: A gridded reconstruction of land and ocean precipitation for the extended tropics from 1974 to 1994. *Int. J. Climatol.,* **19,** 119–142.

Dorman, C. E., and R. H. Bourke, 1979: Precipitation over the Pacific Ocean, 30°S to 60°N. *Mon. Wea. Rev.,* **107,** 896–910.

——, and ——, 1981: Precipitation over the Atlantic Ocean, 30°S–70°N. *Mon. Wea. Rev.,* **109,** 554–563.

Eischeid, J. K., H. F. Diaz, R. S. Bradley, and P. D. Jones, 1991: A comprehensive precipitation dataset for global land areas. U.S. Department of Energy Tech. Rep. DOE/ER-6901T-H1, 81 pp. [Available from National Technical Information Service, U.S. Dept. of Commerce, Springfield, VA 22161.]

Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields.* Israel Program for Scientific Translations, 242 pp.

Gruber, A., and A. F. Krueger, 1984: The status of the NOAA outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.,* **65,** 958–962.

Huffman, G. J., R. F. Adler, B. R. Rudolf, U. Schneider, and P. R. Keehn, 1995: Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *J. Climate,* **8,** 1284–1295.

——, and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.,* **78,** 5–20.

Hughes, P. Y., E. H. Mason, T. R. Karl, and W. A. Brower, 1992: United States historical climatology network daily temperature and precipitation data. Tech. Rep. ORNL/CDIAC-50, NDP-042, Carbon Dioxide Information Analysis Center, ORNL, Oak Ridge, TN, 40 pp.

Hulme, M., 1991: An intercomparison of model and observed global precipitation climatologies. *Geophys. Res. Lett.,* **18,** 1715–1718.

——, 1992: A 1951-80 global land precipitation climatology for the evaluation of general circulation models. *Climate Dyn.,* **7,** 57–72.
——, 1994: Validation of large-scale precipitation fields in general circulation models. *Global Precipitations and Climate Change,* M. Desbois and F. Desalmand, Eds., Springer-Verlag, 387–405.
——, 1995: Estimating global changes in precipitation. *Weather,* **50,** 34–42.
——, and M. G. New, 1997: Dependence of large-scale precipitation climatologies on temporal and spatial sampling. *J. Climate,* **10,** 1099–1113.
Jaeger, L., 1976: Monatskarten des Niederschlags für die Ganze Erde (Monthly precipitation maps for the whole earth). Berichte des Deutscher Wetterdienstes, 33 pp.
Janowiak, J. E., 1988: An investigation of interannual rainfall variability in Africa. *J. Climate,* **1,** 240–255.
——, and P. A. Arkin, 1991: Rainfall variation in the Tropics during 1986–1989, as estimated from observations of cloud-top temperature. *J. Geophys. Res.,* **96,** 3359–3373.
——, A. Gruber, C. R. Kondragunta, R. E. Livezey, and G. J. Huffman, 1998: A comparison of NCEP–NCAR reanalysis precipitation and the GPCP rain gauge–satellite combined dataset with observational error considerations. *J. Climate,* **11,** 2960–2979.
Kumar, A., and M. P. Hoerling, 1997: Interpretation and implications of the observed inter–El Niño variability. *J. Climate,* **10,** 83–91.
——, ——, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate,* **9,** 115–129.
Legates, D. R., 1987: A climatology of global precipitation. *Publ. Climatol.,* **40,** 85 pp.
——, and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge corrected, global precipitation. *Int. J. Climatol.,* **10,** 111–127.
New, M. G., M. Hulme, and P. D. Jones, 1999: Representing twentieth century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate,* **12,** 829–856.
——, ——, and ——, 2000: Representing twentieth century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate,* **13,** 2217–2238.
Nijssen, B., G. M. O'Donnell, and D. P. Lettenmaier, 2001: Predicting the discharge of global rivers. *J. Climate,* **14,** 3307–3323.
Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.,* **78,** 2837–2849.
Petty, G. W., 1995: Frequencies and characteristics of global oceanic precipitation from shipboard present-weather reports. *Bull. Amer. Meteor. Soc.,* **76,** 1593–1616.
Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.
——, and ——, 1989: Precipitation patterns associated with the high index phase of the Southern Oscillation. *J. Climate,* **2,** 268–264.
——, J. E. Janowiak, and M. S. Halpert, 1985: The analysis and display of real time surface climate data. *Mon. Wea. Rev.,* **113,** 1101–1106.
Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Climate,* **11,** 109–120.
Rudolf, B., 1993: Management and analysis of precipitation data on a routine basis. *Int. Symp. on Precipitation and Evaporation,* Bratislava, Slovakia, WMO, 69–76.
Schneider, U., 1993: The GPCC quality-control system for gauge-measured precipitation data. *Proc. Analysis Methods of Precipitation on a Global Scale: Report of a GEWEX Workshop,* Koblenz, Germany, GPCC, WCRP-81, WMO/TD-588, A5–A9.
Sevruk, B., 1982: Methods of correction for systematic error in point precipitation measurement for operational use. Operational Hydrology Rep. 21, WMO Rep. 589, 91 pp.
——, 1989: Reliability of precipitation measurements. *Proc. Int. Workshop on Precipitation Measurements,* St. Moritz, Switzerland, WMO/IAHS/ETH, 13–19.
Shea, D. J., 1986: Climatological atlas: 1950–1979. NCAR Tech. Note TN-269+STR, 35 pp. [Available from National Center for Atmospheric Research, Boulder, CO 80307.]
Shepard, D., 1968: A two dimensional interpolation function for regularly spaced data. *Proc. 23d National Conf. of the Association for Computing Machinery,* Princeton, NJ, ACM, 517–524.
Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperature using empirical orthogonal functions. *J. Climate,* **9,** 1403–1420.
Stephenson, D. B., F. Chauvin, and J.-F. Royer, 1998: Simulation of the Asian summer monsoon and its dependence on model horizontal resolution. *J. Meteor. Soc. Japan,* **76,** 237–265.
Trenberth, K. E., and C. J. Guillemott, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP reanalysis. *Climate Dyn.,* **14,** 213–231.
Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. K. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea-level pressure, and station pressure data. Tech. Rep. ORNL/CDIAC-53, Carbon Dioxide Information Analysis Center Oak Ridge National Laboratory, Oak Ridge, TN, 26 pp.
Willmott, C. J., C. M. Rowe, and W. D. Philpot, 1985: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Amer. Cartogr.,* **12,** 5–16.
——, S. M. Robeson, and M. J. Janis, 1996: Comparison of approaches for estimating time-averaged precipitation using data from the USA. *Int. J. Climatol.,* **16,** 1103–1115.
Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate,* **9,** 840–858.
——, and ——, 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.
——, B. Rudolf, U. Schneider, and P. A. Arkin, 1996: Gauge-based monthly analysis of global land precipitation from 1971–1994. *J. Geophys. Res.,* **101** (D14), 19 023–19 034.
Yang, S., K.-M. Lau, and P. S. Schopf, 1999: Sensitivity of the tropical Pacific Ocean to precipitation-induced freshwater flux. *Climate Dyn.,* **15,** 737–750.