JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D04110, doi:10.1029/2007JD009132, 2008

Click Here for Full Article

Assessing objective techniques for gauge-based analyses of global daily precipitation

Mingyue Chen,[1] Wei Shi,[1] Pingping Xie,[1] Viviane B. S. Silva,[1,2] Vernon E. Kousky,[1] R. Wayne Higgins,[1] and John E. Janowiak[1]

Received 6 July 2007; revised 6 October 2007; accepted 21 November 2007; published 29 February 2008.

[1] Three objective techniques used to obtain gauge-based daily precipitation analyses over global land areas are assessed. The objective techniques include the inverse-distance weighting algorithms of Cressman (1959) and Shepard (1968), and the optimal interpolation (OI) method of Gandin (1965). Intercomparisons and cross-validation tests are conducted to examine their performance over various parts of the globe where station network densities are different. The gauge data used in the examinations are quality controlled daily precipitation reports from roughly 16,000 stations over the global land areas that have been collected by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC). Data sources include daily summary files from the Global Telecommunication System (GTS), and the CPC unified daily gauge data sets over the contiguous United States (CONUS), Mexico, and South America. All three objective techniques are capable of generating useful daily precipitation analyses with biases of generally less than 1% over most parts of the global land areas. The OI method consistently performs the best among the three techniques for almost all situations (regions, seasons, and network densities). The Shepard scheme compares reasonably well with the OI, while the Cressman method tends to generate smooth precipitation fields with wider raining areas relative to the station observations. The quality of the gauge-based analyses degrades as the network of station observations becomes sparser, although the OI technique exhibits relatively stable performance statistics over regions covered by fewer gauges.

Citation: Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, *113*, D04110, doi:10.1029/2007JD009132.

1. Introduction

[2] Despite the rapid progress achieved in the last two decades in estimating precipitation from radar and satellite observations and in simulating precipitation through numerical models, gauge observations continue to play a critical role in documenting the characteristics of precipitation over global land areas [*Huffman et al.*, 1997; *Xie and Arkin*, 1997; *Adler et al.*, 2003]. Gauge observations have the longest recording period which makes them the most suitable sources from which the long-term mean and variability of precipitation on various timescales can be defined. In addition, gauge observations are the only source of precipitation that is obtained via direct measurement. Radar estimates, satellite estimates, and model predictions are indirect in nature and hence must be calibrated or verified using the gauge observations [e.g., *Xie and Arkin*, 1995; *Ebert and Manton*, 1998; *Adler et al.*, 2001; *McCollum et al.*, 2002].

[3] Several sets of precipitation climatologies have been constructed over global land areas by interpolating gauge-observed monthly climate normals [*Legates and Willmott*, 1990; *Hulme*, 1991; *New et al.*, 1999]. Furthermore, gauge-based analyses of monthly and daily precipitation have been constructed over global and regional domains by several research groups around the world [e.g., *Bradley et al.*, 1987; *Willmott and Matsuura*, 1995; *Rudolf*, 1993; *Schneider*, 1993; *Xie et al.*, 1996, 2007; *Dai et al.*, 1997; *Higgins et al.*, 2000; *New et al.*, 2000; *Shi et al.*, 2001; *Chen et al.*, 2002; *Maurer et al.*, 2002]. These gauge-based analyses have been used in a wide range of applications, including weather and climate monitoring, climate diagnostics, verification of numerical models and satellite products, and hydrological studies [e.g., *Morrissey et al.*, 1995; *Dai and Wigley*, 2000; *Sorooshian et al.*, 2000; *Higgins and Shi*, 2000, 2001; *Roads et al.*, 2001; *Higgins et al.*, 2004; *Xue et al.*, 2005; *Yatagai et al.*, 2005, *Ebert et al.*, 2007; *Xie et al.*, 2007].

[4] At the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) gauge observations have long been utilized for climate monitoring,

[1]Climate Prediction Center, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA.
[2]Also at RS Information Systems, Inc. (RSIS), McLean, Virginia, USA.

0148-0227/08/2007JD009132$09.00

climate analysis and climate forecast verification activities (A. Kumar, An overview of operational activities at Climate Prediction Center (CPC) to be submitted to *Weather and Forecasting*, 2007). The Climate Anomaly Monitoring System (CAMS), which is a real-time surface climate database of monthly precipitation and temperature observations, was established in 1984 to aid in the analysis of climate anomalies. In essence, CAMS is an archive of station reports of monthly mean precipitation and surface air temperatures collected from the daily Global Telecommunication System (GTS) data and the monthly CLIMAT reports [*Ropelewski et al.*, 1985]. Monthly precipitation reports from ~6000 stations are available for each month and are inserted in CAMS. Analyzed fields of monthly precipitation are created over the global land areas by interpolating the CAMS gauge observations [*Xie et al.*, 1996]. These gauge-based analyses are then combined with estimates derived from satellite observations to generate improved near-global precipitation analyses [the CPC Merged Analysis of Precipitation (CMAP); *Xie and Arkin*, 1996, 1997] on a 2.5° latitude/longitude grid. A streamlined version of CMAP, called CAMS-OPI [*Janowiak and Xie*, 1999], is also produced on a real-time basis by merging the CAMS gauge-based analysis and the OLR-based Precipitation Index [OPI; *Xie and Arkin*, 1998] satellite estimates.

[5] In addition to the observations that are available from the GTS and the CLIMAT reports described above, CPC also collects daily precipitation reports from other sources. These sources include the National Weather Service (NWS) River Forecast Centers (RFC), the Hydrologic Automated Data System (HADS), and national collections from meteorological agencies in Mexico and countries in South America (mainly Brazil), East Asia, and Africa. Together with the GTS data, these additional daily gauge observations produce relatively dense gauge networks over the regions. For instance, over the contiguous United States (CONUS), the total number of daily precipitation reports from the combined daily gauge data sets collected at NOAA/CPC is about 8000 compared to about 500 from the GTS alone. Analyses of daily precipitation are created on a 0.25° latitude/longitude grid over the CONUS-Mexico domain [*Higgins et al.*, 2000], on a 1.0° latitude/longitude grid over South America [*Shi et al.*, 2001; *Silva et al.*, 2007], and on a 0.5° latitude/longitude grid over East Asia [*Xie et al.*, 2007] by interpolating the GTS and regional station data.

[6] However, the precipitation variations in the global gauge-based and merged analyses do not always match with those in the regional gauge-based analyses primarily due to differences in the gauge station data and interpolation algorithms used in defining the gauge analyses [*Xie et al.*, 2003]. This inconsistency may cause problems in applications where joint use of the global and regional analyses is required.

[7] To address this problem, a project was launched recently at CPC to construct unified gauge and gauge-satellite merged analyses of daily precipitation over both global and regional domains. The data set is derived from a set of quality controlled input gauge and satellite using a robust objective analysis technique. The entire project consists of four major components, i.e., (a) construction of a unified data set of gauge observations and satellite estimates of global and regional precipitation; (b) quality control of these daily gauge reports; (c) creation of gauge-based analyses; and (d) production of the gauge-satellite merged analyses of precipitation. As a first step of this massive project, station daily precipitation reports from GTS over the global land areas and national collections from US, Mexico, and South American countries have been combined to form a preliminary version data set of the unified gauge observations. Quality control procedures have been developed and applied to this unified global gauge data set to flag suspicious reports [*Chen et al.*, 2008].

[8] The purpose of this paper is to select a common objective technique for producing gauge-based analyses of daily precipitation on global and regional domains over land. This is done by assessing the performance of three interpolation algorithms currently used at CPC to generate its operational precipitation products. The three objective methods that were tested include the inverse-distance weighting interpolation algorithms of *Cressman* [1959] and *Shepard* [1968], and the optimal interpolation (OI) technique of *Gandin* [1965]. The Cressman method is used at CPC to create regional analyses over the CONUS-Mexico and the South America domains [*Higgins et al.*, 2000; *Shi et al.*, 2001; *Silva et al.*, 2007]. The technique developed by *Shepard* [1968] is used to produce GTS gauge-based global analyses [*Xie et al.*, 1996], and the OI technique is implemented to construct East Asia daily precipitation maps [*Xie et al.*, 2007]. Although several other objective techniques are also used by scientists of different institutions for gauge interpolation [e.g., *Barnes*, 1964; *Dai et al.*, 1997; *New et al.*, 2000], the three algorithms to be examined here are among those most widely used and present stable performances [*Creutin and Obled*, 1982; *Bussières and Hogg*, 1989; *Chen et al.*, 2002].

[9] Assessment of the performance of gauge interpolation algorithms has been the topic of several published studies. *Creutin and Obled* [1982] examined several well-known schemes in deriving analyzed fields of event total precipitation over regions of intense and highly spatially varying rainfall and recommended the optimal interpolation (OI) of *Gandin* [1965]. *Bussières and Hogg* [1989] compared the performance of four algorithms [*Barnes*, 1964; *Cressman*, 1959; *Shepard*, 1968; and OI of *Gandin*, 1965] in defining daily precipitation analysis over Canada and concluded that the OI does the best job while the technique of *Shepard* [1968] performs almost as well. *Legates* [1987] evaluated several objective procedures for defining a monthly climatology and found that a spherical adaptation of *Shepard*'s [1968] method [*Willmott et al.*, 1985] was the best for interpolating over 24,000 gauge observations of long-term mean precipitation into 0.5° lat/lon grid over global land areas. On the basis of these results, *Shepard* [1968] algorithm was chosen by the Global Precipitation Climatology Centre (GPCC) to construct monthly precipitation analyses from about 7000 quality controlled gauge observations over global land areas [*Rudolf et al.*, 1994]. *Chen et al.* [2002] performed an inter-comparison of four interpolation algorithms and confirmed that the OI technique is the best for construction of gauge-based analyses of monthly and pentad precipitation over the global land areas.

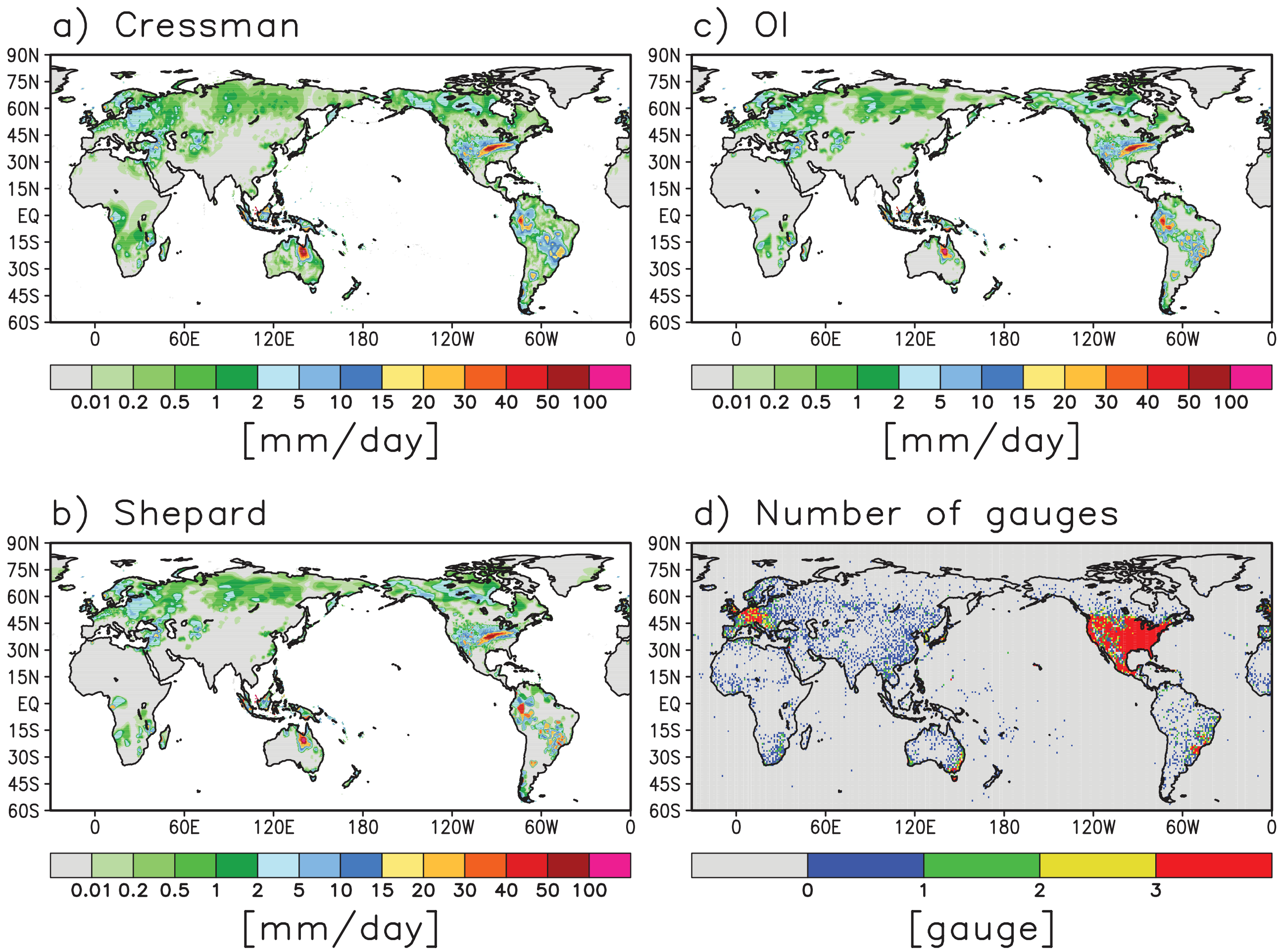

Figure 1. Distribution of precipitation (mm day^{-1}) for 5 January 2005, defined by interpolating station observations through the algorithms of (a) *Cressman* [1959], (b) *Shepard* [1968], and (c) the optimal interpolation (OI) of *Gandin* [1965], together with (d) number of gauge reports available in a 1.0°lat/lon grid box.

[10] Performance of an objective analysis technique differs in generating analyzed fields of precipitation with different spatial structures sampled by different observation networks. The objective technique to be selected here needs to be capable of defining daily precipitation analyses with reliable quality over the global land regions for all seasons and from gauge data from networks of highly variable station densities. None of the above mentioned studies has thoroughly addressed these issues for applications of global daily precipitation. In particular, the sensitivity of the gauge analysis quality to the gauge network density is largely unknown, despite the critical importance of that characteristic for reliable gauge interpolation over the global land areas where mean station-to-station distance may vary from ~30 km over CONUS to ~500 km over tropical Africa. In this paper, we will describe a comprehensive assessment of the performance for the three objective algorithms for interpolating daily precipitation over the global land areas. Section 2 describes the three objective techniques to be examined and the gauge data to be interpolated; section 3 presents results of an inter-comparison of the precipitation analyses generated by the different algorithms, cross-validation tests and the gauge network density impact experiments, while a summary of the results is given in section 4.

2. Algorithms and Gauge Data

2.1. The Objective Analysis Techniques to be Examined

[11] The three objective analysis techniques to be examined in this study are the inverse-distance weighting methods of *Cressman* [1959] and *Shepard* [1968], and the Optimal Interpolation (OI) algorithm of *Gandin* [1965]. They were selected because of their operational applications at CPC and for their wide utilization for analyzing observation fields by many institutions around the world.

[12] In the method of *Cressman* [1959], a "first-guess" field of interpolated values at the target grid points is first defined. The first-guess field is then corrected by the weighted mean of the differences between the observations and the interpolated values at gauge locations within a predetermined search distance from the target grid point. This process is repeated four times, with decreasing search distance. The search distances for the four reiterations are adjusted for gauge networks of different density and for

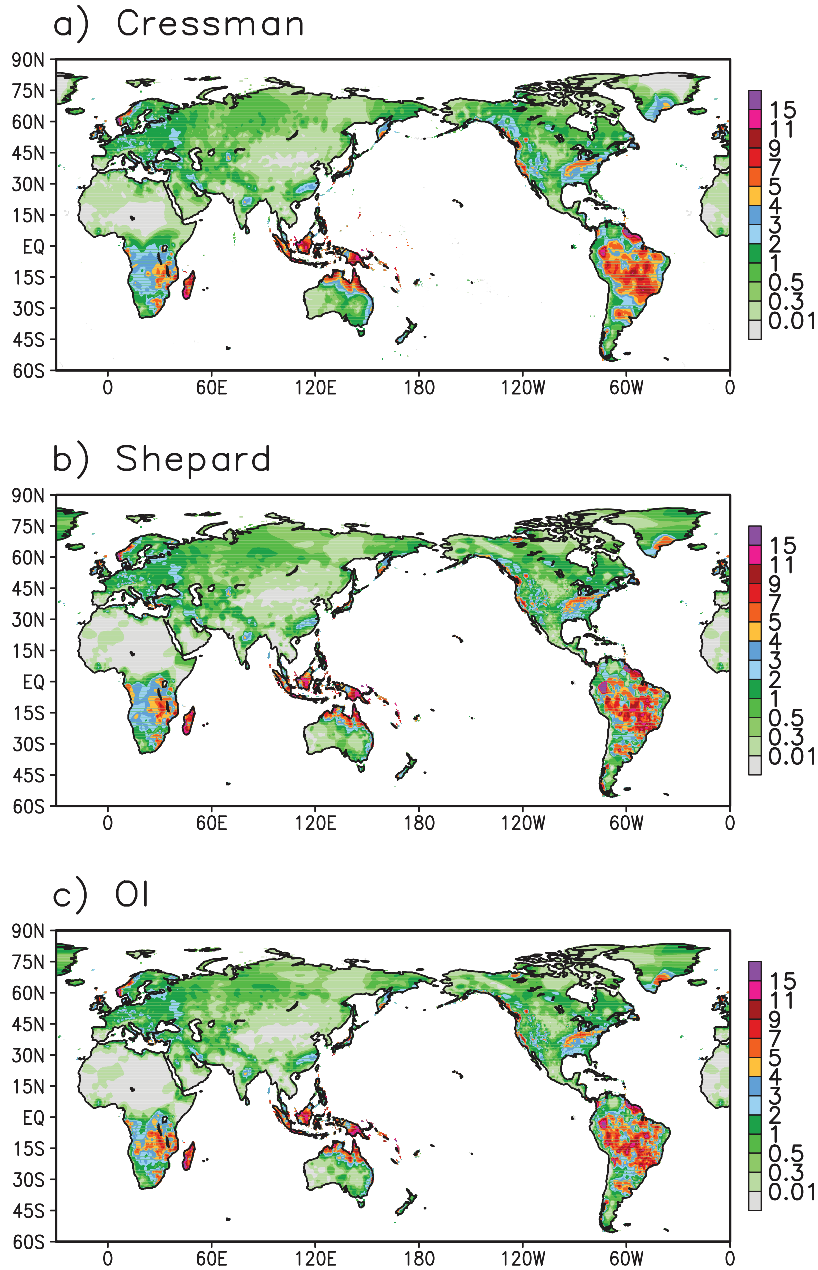

Figure 2. Mean precipitation (mm day^{-1}) for January 2005, defined by interpolation of station reports using the algorithm of (a) *Cressman* [1959], (b) *Shepard* [1968], and (c) OI of *Gandin* [1965].

precipitation fields of different spatial structures. In our implementation [*Higgins et al.*, 2000], the ratio among the four search distances is fixed to 9: 1.9: 1.3: 0.7, while the absolute magnitude of them is tuned for each individual continent and for each season to optimize the interpolation performance. For instance, the search distance for the last reiteration is ~25 km and ~120 km, respectively, for interpolation of June–July August daily precipitation over CONUS and Africa. The weighting function is defined as $(D_m^2 - D_s^2)/(D_m^2 + D_s^2)$, where D_m and D_s are the search distance and the station-gauge distance, respectively. The form of the function yields a slowly decreasing weight with the increasing distance, ensuring robust large-scale distributions with small-scale features smoothed [*Bussières and Hogg*, 1989]. The first guess field is zero rainfall in our implementation.

[13] In the algorithm of *Shepard* [1968], the interpolated value of precipitation at a target grid point is computed as a weighted mean of observed values at nearby gauge stations within a search distance. The search 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-grid point distance, creating a sharp gradient in the analyzed precipitation fields. In addition, directional correction is implemented to account for any uneven distribution of gauges in different directions from the target. The version of the *Shepard* [1968] algorithm that is tested in this study is the

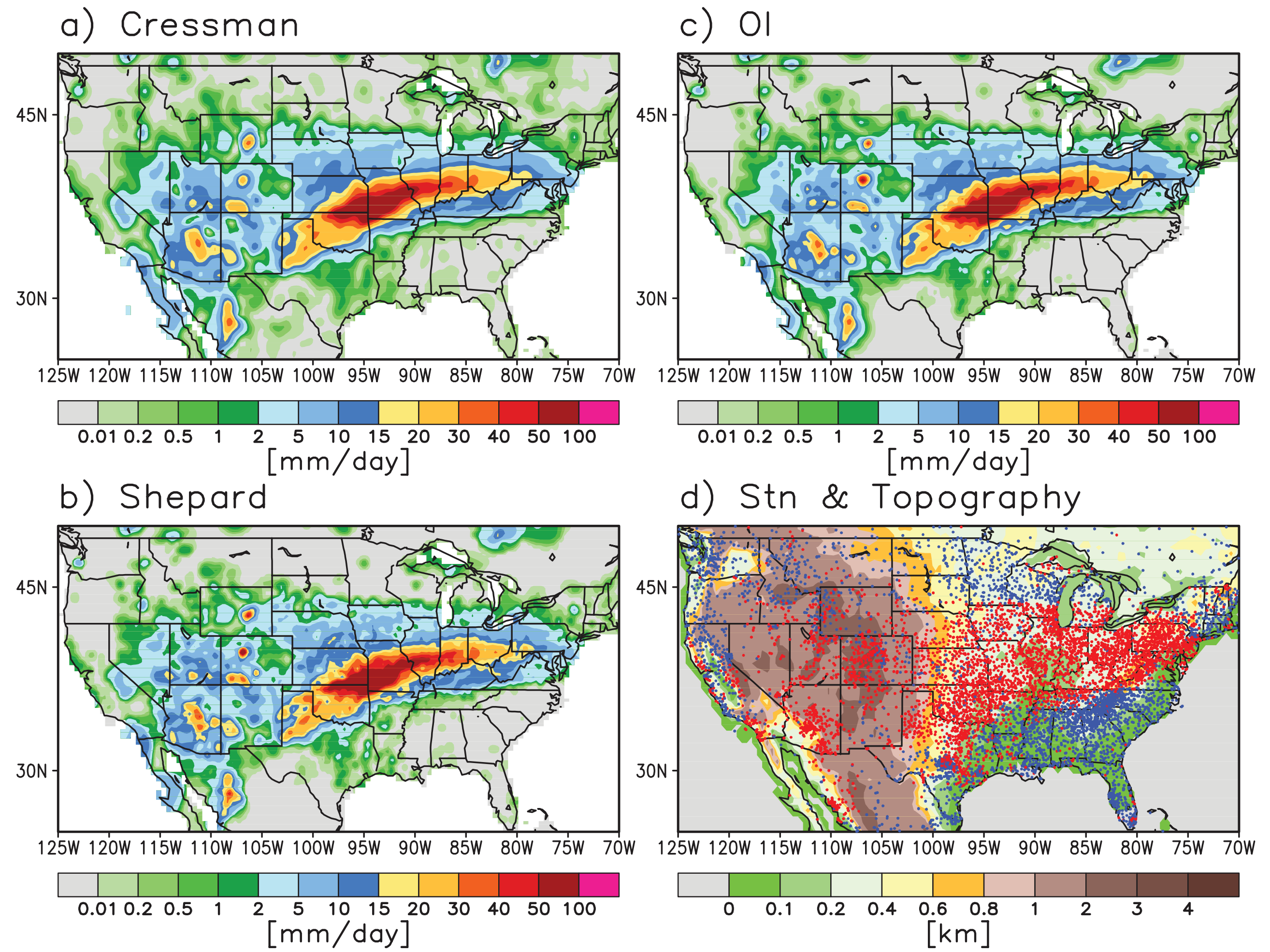

Figure 3. Daily precipitation distribution (mm day^{-1}) for 5 January 2005 over Contigious United States based on interpolation of gauge observations through the algorithm of (a) *Cressman* [1959], (b) *Shepard* [1968], and (c) OI of *Gandin* [1965]. Locations of gauges reporting rain and no-rain events are plotted in red and blue, respectively, in panel d, together with the background color representing elevations.

so-called Spheremap implementation developed by *Legates and Willmott* [1990], in which distances and angles are calculated on a spherical coordinate system. The *Shepard* [1968] algorithm is used by the Global Precipitation Climatology Centre [GPCC; *Rudolf*, 1993; *Schneider*, 1993] of Germany and the CPC [*Xie et al.*, 1996] to construct their gauge-based analyses of global precipitation.

[14] The OI technique of *Gandin* [1965] defines the analyzed value at a grid point by modifying a first-guess field with the weighted mean of the differences between the observed and the first-guess values at station locations within a search distance. While in the Cressman and Shepard methods, the weighting coefficient is a function only of gauge-grid point distance, in the OI technique it is determined from the variance and co-variance structure of the target precipitation fields. The implementation of the OI algorithm tested in this inter-comparison was developed by *Xie et al.* [2007] to construct daily precipitation analyses over East Asia. The creation of the daily precipitation analysis is conducted in three steps. First, analyzed fields of daily precipitation climatology are defined from historical gauge observations collected at CPC. Gridded fields of the ratio between the daily precipitation and daily climatology is then computed by interpolating the corresponding values at the gauge locations through the OI technique. Daily precipitation analysis is finally defined by multiplying the fields of the daily climatology and the daily ratio. By interpolating the ratio of total rainfall to the climatology, instead of the total rainfall itself, the OI is capable of better representing the spatial distribution of precipitation, especially over regions with substantial orographic effects [*Xie et al.*, 2007].

[15] In this work, we evaluate the performance of three widely used objective techniques, aiming to select the best one for our operational applications. From that point of view, no efforts have been made to change the contents of the operational packages. In creating the analyses using the methods of *Cressman* [1959] and *Shepard* [1968], total precipitation is interpolated, although interpolating anomalies or ratios to the climatology generally yield better results [*Chen et al.*, 2002; *Xie et al.*, 2007]. Our primary principal here is to make our comparisons fair to all of the ‘‘operational packages’ involved.

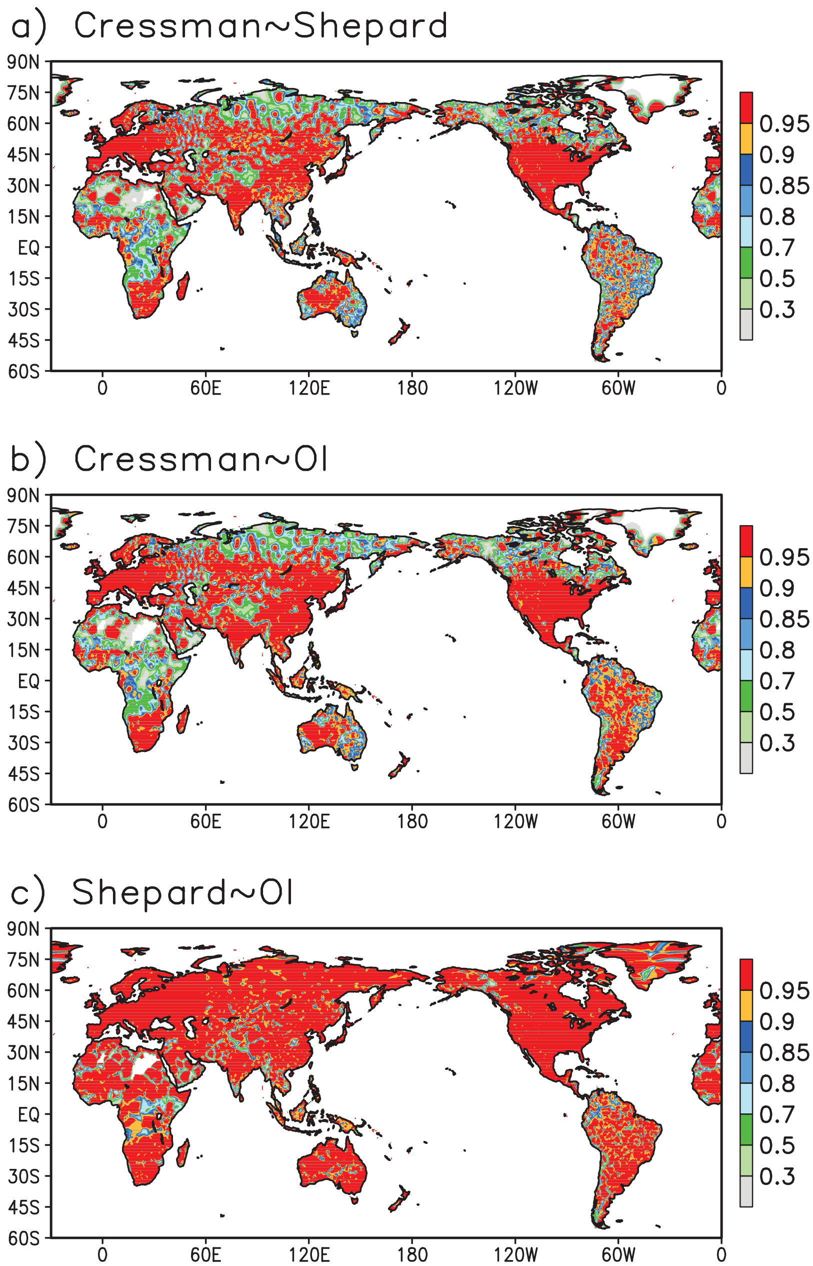

Figure 4. Serial correlation among the gauge-based daily precipitation analysis derived using the Cressman, Shepard, and OI algorithms. The correlation is computed for each 0.5° lat/lon grid for January, April, July, and October of 2005.

2.2. Gauge Data Used in the Assessment

[16] Gauge observations of daily precipitation from four individual data sets are combined and used in examining the performance of the objective interpolation algorithms. These include the Global Telecommunication System (GTS) daily summaries files archived at CPC for a period from 1977 to the present; the CPC unified data set of daily station precipitation over CONUS [*Higgins et al.*, 2000] starting from 1948; a data set of daily precipitation over Mexico provided to CPC by the National Meteorological Service (SMN) of Mexico [*Higgins et al.*, 2000] dating back to 1948, and a collection of daily precipitation reports over South America [*Shi et al.*, 2001; *Silva et al.*, 2007] beginning from 1979. These individual data sets are selected for their availability in real-time operations at CPC. While the GTS gauge network covers most of the global land areas, the station density in those data is less than desirable over many important regions of the world. Individual data sets collected at CPC from various sources (meteorological, hydrological agencies and other organizations), meanwhile, provide dense gauge networks over several of the regions that have sparse gauges in the GTS data.

[17] In this study, gauge observations from the individual regional data sets over CONUS, Mexico, and South America are combined with those from the GTS over the rest of the world to create an enhanced data set of gauge observations of daily precipitation over the entire global land areas.

Table 1. Summary of Cross-Validation Tests for the Gauge-Based Analyses of Daily Precipitation Over the Global and Regional Domains for the Entire Periods of January, April, July, and October of 2005[a]

	Cressman		Shepard		OI	
	Correlation	Bias, %	Correlation	Bias, %	Correlation	Bias, %
Global	0.706	0.251	0.709	−0.085	0.735	-0.349
U.S.	0.793	0.754	0.784	−0.118	0.811	−0.467
S. America	0.653	0.244	0.723	0.084	0.724	0.369
E. Asia	0.592	0.572	0.567	−0.533	0.596	−0.332
Indo. Islands	0.290	−1.772	0.253	−0.677	0.313	−0.831
Australia	0.553	1.381	0.588	−0.579	0.592	−0.207
Africa	0.364	3.316	0.354	1.259	0.377	−0.778

[a]Correlation and bias are computed through comparisons with the withdrawn independent gauge observations.

Comprehensive quality control procedures are applied to the raw station data to remove suspicious reports with zero and extremely large values. This is done through comparisons with climatological statistics at the target station, concurrent observations from nearby stations, and corresponding radar images, satellite estimates and numerical model forecasts of daily precipitation [*Higgins et al.*, 2000; *Shi et al.*, 2001; hereinafter referred to as *Chen et al.*, 2008].

[18] Figure 1d presents an example of the gauge distributions for 5 January 2005. In total, daily precipitation reports from ~16,000 stations are available from the combined global station data set, composed of ~8500, ~1100, ~1100 stations over CONUS, Mexico, and South America, respectively, and ~5500 from the GTS data set over the rest of the world. The gauge network is quite dense over most of the CONUS, east coast of Brazil, Western Europe, South Africa and the coastal regions of China and Australia, while precipitation is poorly sampled over most of the African continent, western China, central Australia and the Amazon.

3. Examination of Results

3.1. Inter-Comparisons of the Daily Analyses Derived by the Three Algorithms

[19] Analyzed fields of daily precipitation are created on a 0.5°lat/lon grid over the global land areas for 2005 through interpolation of the combined global station reports using the three objective analysis techniques. Results are then compared to each other and with independent observations to examine their performance. Figure 1 shows an example of the global daily precipitation analyses (grid box average) generated by the three algorithms, together with the number of gauge reports available in each 1.0°lat/lon grid box. As expected, close agreement is observed in the large-scale global precipitation patterns generated by the three algorithms. Major precipitation areas (with similar magnitudes) over southern Africa, Europe, Siberia, Australia, northern Canada, US, and South America are well depicted in all three analyses. Differences, however, are observed in the small scale features and in the extension of raining areas. The Cressman technique (Figure 1a) tends to generate precipitation fields with smoother spatial distribution and larger precipitation areas than the Shepard (Figure 1b) and the OI (Figure 1c) methods, especially over regions of sparse gauge networks (e.g., southern Africa, South America and central Australia). These differences, however, tend to decrease as daily fields are accumulated to form precipitation distributions for extended periods. As shown in Figure 2, only minor differences are visible among the global distribution of monthly precipitation generated by the three objective analysis algorithms examined here, though the Cressman still presents slightly larger areas of precipitation.

[20] Figure 3 provides a close look at the daily precipitation analyses for 5 January 2005, over the CONUS where daily reports from ~8000 stations are available over this area of $\sim 8 \times 10^6$ km^2. The precipitation distribution for this day is governed primarily by a passing frontal system extending from Arizona to the northeast corner of the country. Precipitation is well organized into a band over central and eastern CONUS, while it spreads out and is more or less scattered over western mountainous areas of the nation. All of the three algorithms produced a precipitation distribution with a maximum of 50–100 mm/day over the corner of Kansas, Missouri, Okalahoma, and Arkansas. It is clear that the precipitation generated by the Cressman technique is smoother and presents larger raining areas than those of the OI and Shepard, especially over the western mountainous areas with precipitation of scattered distributions is observed by a relatively sparse gauge network.

[21] Figure 4 presents the serial correlation between daily precipitation analyses generated by different interpolation algorithms. Correlations higher than 0.95 (red color) are observed among the daily precipitation values generated by the three algorithms over most of the global land areas. In general, analyses produced by the OI and the Shepard techniques exhibit close agreement, while poor correlation (<0.5) appears between the Cressman analysis and other two precipitation data sets over areas with sparse gauge networks (e.g., tropical Africa, Tibet, and northern Russia). No systematic variations are observed in the correlation maps calculated for different seasons (not shown). While other factors may be involved as well, differences between the daily analyses derived by the Cressman algorithm and those by the other two techniques are caused largely by the differing weighting functions. With weighting coefficients decrease slowly with distance, the Cressman algorithm relies more on observations at distant stations over regions of poor gauge networks.

3.2. Cross-Validation Tests

[22] To quantify the performance of the three objective analysis techniques in interpolating daily precipitation, cross-validation tests are conducted over the global land areas. To this end, 10% of the stations were randomly selected from the entire databases of N (~16,000) stations over the global land areas. This is done by arranging the stations in order (1st, 2nd, ... Nth), generating a series of N random numbers ranging from 0–10 using a Fortran utility and picking out stations with corresponding random number between 0–1. Daily precipitation reports for these stations are withdrawn and the gauge observations for the remaining 90% of stations are used to define the analyzed values at the locations of the withdrawn stations. This process is repeated 10 times so that each station is withdrawn once. Visual inspection of the 10 sets of withdrawn stations shows

Table 2. Summary of Cross-Validation Tests for the Gauge-Based Analyses of Daily Precipitation Performed for January, July, April, July, and October of 2005 over (a) Northern, and (b) Southern Hemispheres

	Cressman		Shepard		OI	
	Correlation	Bias, %	Correlation	Bias, %	Correlation	Bias, %
(a)						
January	0.798	0.724	0.793	−0.228	0.806	−0.534
April	0.734	0.906	0.766	−0.044	0.783	−0.491
July	0.672	0.762	0.658	0.080	0.690	−0.580
October	0.743	0.285	0.740	−0.349	0.771	−0.238
(b)						
January	0.481	−2.268	0.514	−1.048	0.606	1.151
April	0.466	−6.658	0.486	1.390	0.620	0.390
July	0.466	−6.398	0.474	−0.992	0.590	−0.168
October	0.445	0.473	0.440	−1.497	0.661	−0.887

relatively homogeneous distributions. The analyzed values at the station locations are then compared to the corresponding withdrawn station observations to assess the quantitative accuracy of the interpolated daily precipitation fields.

[23] Table 1 shows the comparison statistics of the cross-validation tests for the three interpolation algorithms over the global land areas as well as six individual regions around the world. The correlations and biases are calculated from the analyzed daily precipitation values and the original daily observations station by station for the global land areas and sub-regions. All of the three objective techniques were able to generate analyzed fields of daily precipitation with reasonable spatial distribution and close magnitude agreement. Bias is less than 1% relative to the mean gauge-observed precipitation over most of the global land areas. The correlation is higher than 0.5 for all global regions, except for Indonesia and Africa where heavy tropical rainfall with large spatial variations is sampled by sparse gauge networks. The best performance is observed over the United States where daily precipitation is monitored by a national network of ~8000 gauges with an average area of ~1000 km^2 per gauge. Analyzed fields derived by the OI shows consistently better correlation with the withdrawn independent gauge observations, compared to those based on the Cressman and Shepard techniques. The correlations computed over the entire global land areas reaches 0.735 for the OI-based analysis, suggesting good overall performance in analyzing daily precipitation.

[24] Tables 2a and 2b present summaries of the cross-validation tests for the gauge-based analyses of daily precipitation validation for January, April, July, and October of 2005, over the northern and southern hemispheres, respectively. The OI exhibits the best correlation for all of the four individual months and over both of the hemispheres. In particular, over the southern hemisphere where gauge networks are relatively poor, the OI-based precipitation analysis shows substantially improved correlation (0.6 or higher) than that based on the Cressman and the Shepard (0.4–0.5). Biases of the OI-based analyses are comparable or a little bit smaller than those of the analyses created by the other two techniques.

[25] One important statistic of precipitation fields is the probability density function (PDF) of precipitation intensity. In general, interpolating point observations yields analyzed fields with reduced occurrences for both high and low (no-rain) precipitation amounts compared to those of the orig-

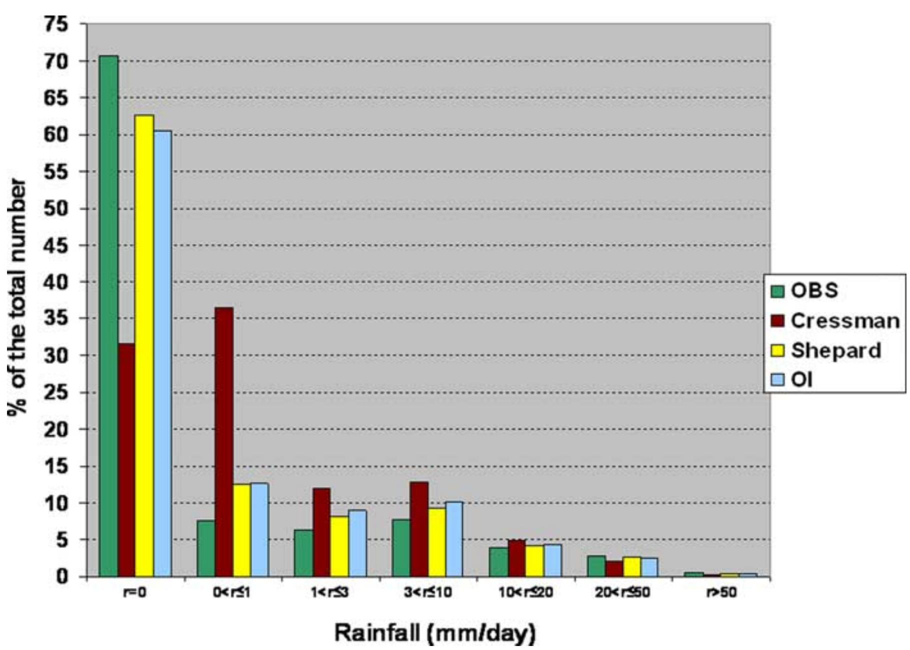

Figure 5. Probability density function (PDF,%) of daily precipitation amount (mm day^{-1}) defined by the gauge station reports (green), analyses based on the algorithms of Cressman (red), Shepard (yellow), and OI (blue). PDFs are computed for precipitation amounts observed/interpolated at all stations over the entire global land areas and for the entire cross-validation tests for January, April, July, and October of 2005.

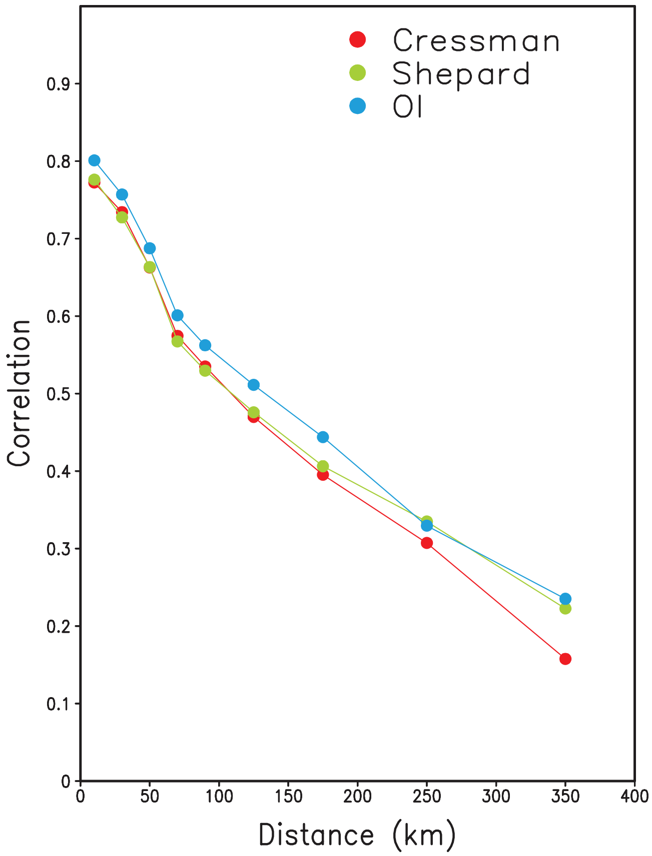

Figure 6. Relationship between the station daily precipitation analysis correlation at a withdrawn station and the distance from that station to the closest gauge with daily reports. Correlation between the station reports and the analyzed values is first computed for each withdrawn station for the entire cross-validation period of January, April, July and October of 2005. Mean correlation values are then defined for nine bins according to the distance to the closest station. Results for gauge-based analyses using the Cressman, Shepard, and OI algorithms are plotted in red, green and blue, respectively.

inal station observations. While caution is needed when interpreting the differences, examination of the PDF functions, or histograms, of precipitation intensity provides us with a qualitative sense of how well the relative intensity of precipitation events are reproduced in the analyzed fields. The histograms of daily precipitation at all stations over global land areas is largely dominated by no-rain events which has a frequency of occurrence of ~70%, while the probability of daily precipitation greater than 50 mm is 0.6% during the four-month period in 2005 (Figure 5, green bars). The OI and Shepard techniques capture this feature quite well. The frequencies for no-rain (rain greater than 50 mm day^{-1}) are 60.6% (0.4%), and 62.6% (0.5%), respectively, for the analyses generated by the OI (blue bars) and the Shepard (yellow bars) algorithms. The Cressman method (red bars) significantly under-estimates the frequency of no-rain while over-represents regions with light rainfall. The frequency of no-rain days is only 32%, less than half of that of the station observations and the analyses derived by the OI and Shepard techniques. Meanwhile, the frequency of light rain ($R < 1$ mm/day) is ~3 times as much as those in the observations and the analyses based on the OI and Shepard techniques.

[26] The performance of the gauge-based analyses is further investigated in relation to the density of gauge networks from which station values are reported. For this purpose, the serial correlation between the daily analyses and the corresponding gauge observations at each withdrawn station is calculated for the four selected months (January, April, July, and October of 2005) of cross-validation tests. Mean correlation values are then computed for nine groups of stations that are determined based on the distance between the target station and the closest stations from which gauge data are available for interpolation. The distance to the closest reporting gauge station is a good index of the gauge network density. Figure 6 presents the relationship between the correlation and gauge network density for the Cressman (red), Shepard (yellow) and OI (blue) techniques. Clearly, the quality of the interpolated field of daily precipitation improves as the gauge network becomes denser. Correlation may reach ~0.8 if precipitation reports are available from a station within 20 km, while it degrades to less than 0.4 if no stations are located within 200 km. Overall, the OI presents better statistics than the other two techniques, especially over regions with low gauge density.

[27] The cross-validation results described above corroborate the superiority of the OI technique in analyzing daily precipitation fields over various regions and for different seasons. Previous examinations [e.g., *Creutin and Obled*, 1982; *Bussières and Hogg*, 1989] have focused on a specific season and/or regions. In addition, our results also demonstrate the strong ability of the OI technique in reproducing the PDF of the precipitation events with high fidelity.

3.3. Impacts of Gauge Network Densities

[28] As described in section 2.2, CONUS is covered by a very dense network of gauges with about 8000 stations reporting precipitation daily. Comparisons of the daily precipitation analyses based on synthetically sparse gauge densities by using selected subsets of the gauges provides an opportunity to examine the impacts of varying gauge density to the quantitative accuracy of the resulting analysis. Therefore to further quantify the impacts of the gauge network density on the accuracy of gauge-based analyses of daily precipitation, cross-validation tests were conducted for analyses over the CONUS region using only 50%, 20%, 5%, 1%, and 0.5% of all available stations.

[29] First, 10% of the daily precipitation reports were withdrawn randomly from the full set of gauge observations and were retained as independent data to verify the quantitative accuracy of analyses derived from subsets of the remaining 90% of gauge reports. In particular, random subsets of gauges that are composed of 50%, 20%, 10%, 5%, 1% and 0.5% of all available gauge reports were constructed. The mean station-to-station distance is ~30 km over the CONUS with all data included, while in a network composed of only 0.5% of all available gauges, the distance increases to ~400 km which is approximately the same as that over tropical Africa in our combined global gauge data set. These sub-sampled gauge data are then used

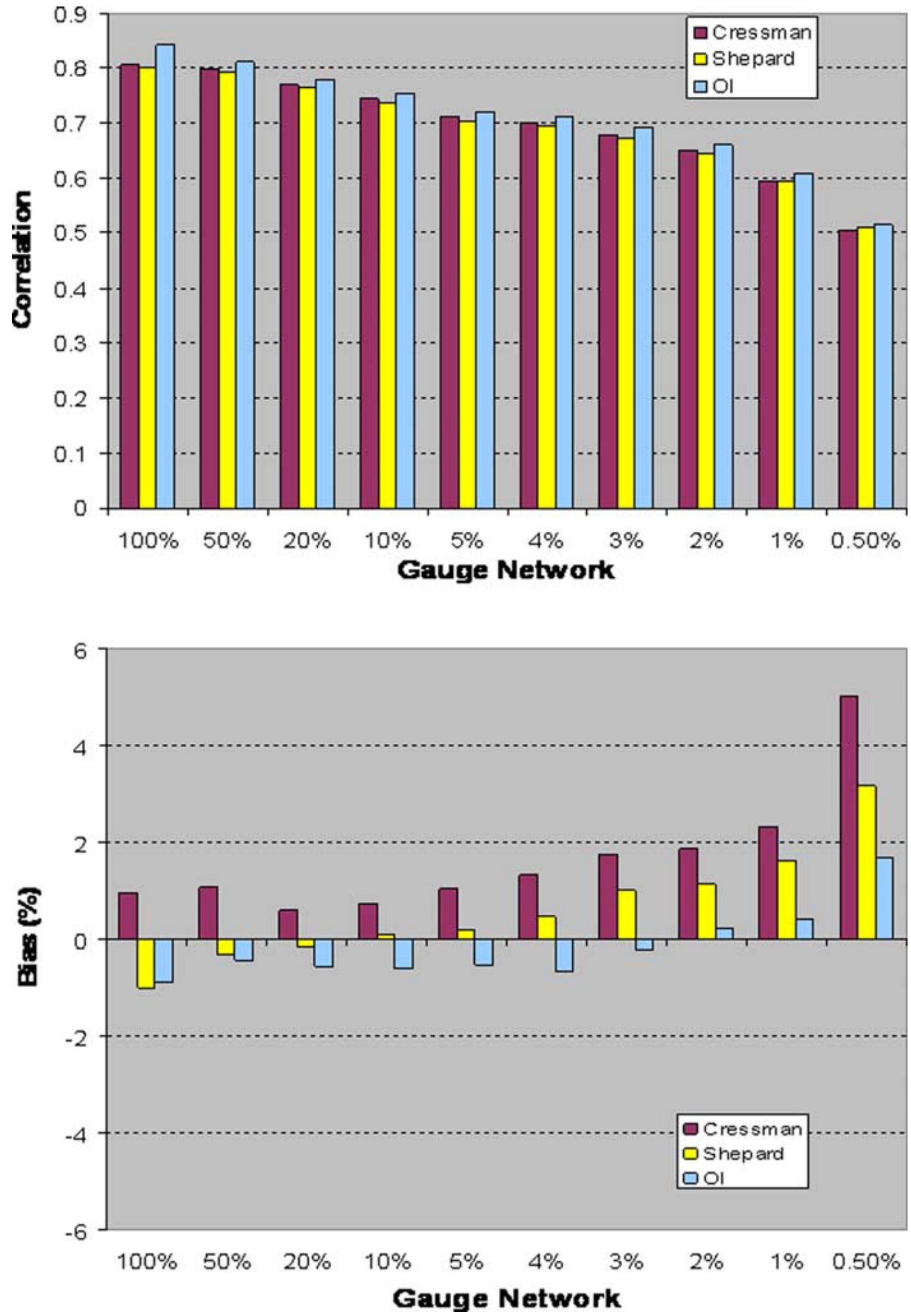

Figure 7. Correlation (top) and bias between the withdrawn independent station observations and analyzed daily precipitation values interpolated from station reports from a subset (0.5% – 100%) of all available stations. Results for gauge-based analyses derived from the Cressman, Shepard and OI algorithms are plotted in red, yellow and blue, respectively.

to define the analyzed fields of daily precipitation and compared against the 10% withheld, independent gauge reports. The inter-comparison statistics for each selected subset cross-validations are calculated station by station.

[30] As expected, the performance of gauge-based analyses improves with the increasing density of gauge data for the three analysis techniques. Correlation coefficients between the gauge-based analyses and the withdrawn station data are about 0.5 when only 0.5% of the gauge data are used in the interpolation, but they are above 0.8 when all data are utilized in the interpolation (Figure 7, top). The daily analyses derived by the OI technique consistently exhibit slightly higher correlation than those produced by the Cressman and Shepard techniques. Both the Cressman and the Shepard techniques show a tendency for generating larger biases for analyses based on gauge networks with relatively sparse density. The OI, meanwhile, shows the best agreement in magnitude with the independent observations for all sets of gauge networks, with a very small bias of less than 1% for networks with 1% or more gauges and ~1.5%

Table 3. Performance Statistics for Analyses Based on a Gauge Network Using 0.5% of all Available Data Over CONUS

	Cressman		Shepard		OI	
	Correlation	Bias, %	Correlation	Bias, %	Correlation	Bias, %
January	0.571	4.330	0.603	2.423	0.577	1.025
April	0.560	2.524	0.551	0.948	0.567	0.092
July	0.376	10.802	0.388	6.807	0.393	4.482
October	0.535	0.984	0.533	1.572	0.550	0.341

for a network of 0.5% of full coverage (Figure 7, bottom). This is partially because in the OI, the interpolation is performed on the ratio between the daily total precipitation and daily climatology instead of the total precipitation itself, thus reducing the chance of spreading precipitation areas into no-rain zones [*Xie et al.*, 2007]. The degradation of the analysis quality caused by the reduced network density is especially large during summer months when small-scale systems make substantial contributions to the daily precipitation. The correlation and bias for daily analyses that are based on the 0.5% gauge network are 0.393 and 4.5%, respectively, compared to 0.577 and 1.0% for winter months when CONUS precipitation is dominated by large-scale synoptic systems (Table 3). The statistics for the analyses of summer months derived from the 0.5% network (or a gauge-to-gauge distance of ×400 km) are comparable with those for analyses over Africa shown in Table 1.

[31] The frequency of occurrence for no-rain events is substantially reduced in the interpolated daily precipitation fields when gauge reports from fewer stations are available (Figure 8). This alias in the PDF is particularly serious in the gauge-based analyses derived by the Cressman algorithm. The percentage of no-rain days is 52%, 42%, and 31%, respectively, in the analyses based on 100%, 10% and 1% of all available gauge reports compared to 72% in the independent observations. The degradation in the fidelity of the PDF is much less in the OI- and Shepard-based analyses which present frequencies of no-rain at 54%, and 53%, respectively, when 1% of the gauge reports are employed.

[32] Overall, results of these impact tests confirm the stable and superior performance of the OI technique in generating daily precipitation analyses from gauge networks of various densities. The Shepard method presents comparison statistics very close to those of the OI, while the Cressman algorithm is capable of analyzing precipitation fields with high pattern correlation but with an erroneously smoothed PDF structure.

4. Summary and Conclusions

[33] A comprehensive assessment has been performed to examine the performance of three published objective analysis techniques in producing daily precipitation analyses by interpolating gauge observations over the global land areas. The three objective techniques include the inverse-distance weighting algorithms of *Cressman* [1959] and *Shepard* [1968], and the optimal interpolation (OI) method of *Gandin* [1965]. The gauge observations used in the examinations are quality controlled daily precipitation reports from ∼16,000 stations over the global land areas collected through combination of four individual data sets available at NOAA Climate Prediction Center (CPC), i.e., the daily summary files from the Global Telecommunication System (GTS), and the CPC unified daily gauge data sets over the contiguous United States (CONUS), Mexico, and South America.

[34] Inter-comparisons and cross-validation tests have been conducted on the analyses of daily precipitation generated by interpolating the station reports by the three objective techniques to examine their performance in characterizing the analyses over various parts of the globe, for all seasons, and from station networks of different densities. Our results show the following:

[35] 1) All of the three objective techniques are capable of generating analyses of daily precipitation with high correlation and close magnitude agreements with independent gauge observations. Biases of the gauge-based analyses are generally less than 1% for analyses produced by all of the three techniques over most parts of the global land areas;

[36] 2) The OI method consistently performs the best among the three techniques, exhibiting the highest correlation and very close probability density function (PDF) of rainfall intensity compared to the independent gauge observations for almost all situations (regions, seasons, and network densities). The Shepard scheme compares well with the OI, while the Cressman tends to generate smooth precipitation fields with broader areas of precipitation relative to the station observations and the analyses based on the OI and the Shepard algorithms;

[37] 3) The quality of the gauge-based analyses degrades as the network of station observations becomes sparser. However, the OI technique exhibits relatively stable performance statistics over regions covered with fewer gauges.

[38] Based on these results, a decision has been made at CPC to use the OI technique to create the unified analyses of daily precipitation over the global land areas. Focusing on applications such as weather/climate monitoring, climate variability studies and model verifications, the analysis is created on a 0.5° lat/lon grid to represent the area-averaged values of daily precipitation over the grid boxes. Work is underway to construct the gauge-based analyses on a real-time basis and for historical periods, and to combine the gauge-based analyses with satellite-based precipitation fields of CPC Morphing Technique [CMORPH, *Joyce et al.*, 2004]. Our final goal is to provide the science community with a suite of high-resolution, high quality gauge-based and gauge-satellite merged analyses of daily precipitation over the global and regional domains.

Figure 8. Probability density function (PDF, %) of daily precipitation amount (mm day^{-1}) defined by the gauge station reports (green), analyses derived by interpolation of station reports from 100% (top), 10% (middle), and 1% of the total available reports, using objective algorithms of Cressman (red), Shepard (yellow), and OI (blue). PDF is computed at the CONUS stations with reports withdrawn from the interpolations and for the entire impact tests period of January, April, July, and October of 2005.

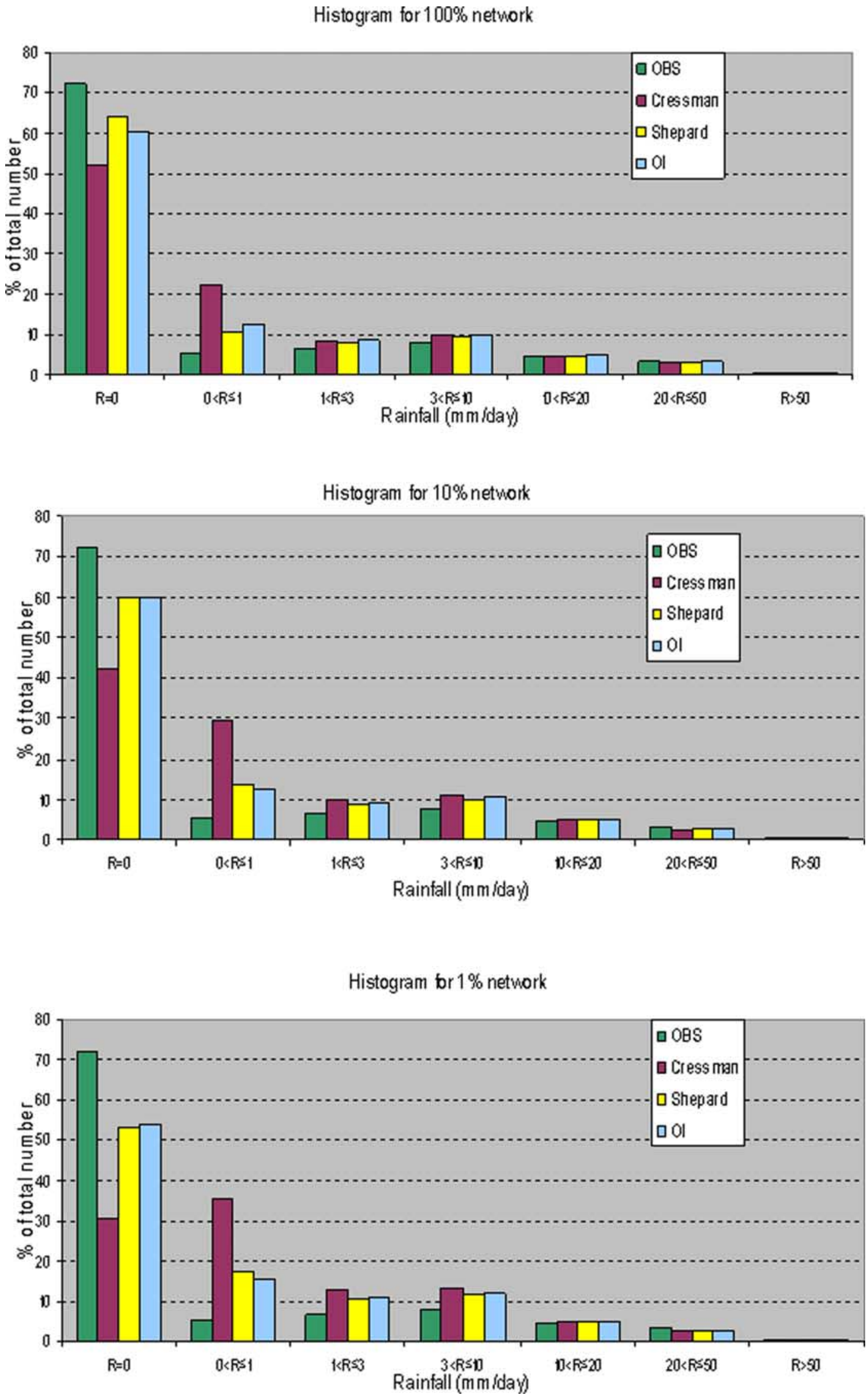
Histogram for 100% network
Histogram for 10% network
Histogram for 1% network
OBS
Cressman
Shepard
OI
% of total number
Rainfall (mm/day)
R=0
0<R≤1
1<R≤3
3<R≤10
10<R≤20
20<R≤50
R>50
0
10
20
30
40
50
60
70
80

Figure 8

[39] **Acknowledgments.** We would like to thank M. Allard, J. Gottschalck, M. Halpert, G. Fulwood, V. Kumar, T. Love, D. Miskus, E. O'Lenic, and W. Thiaw at CPC for their invaluable discussions throughout this project.

References

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, *4*, 1147–1167.
Adler, R. F., C. Kidd, G. Petty, M. Morrissey, and H. M. Goodman (2001), Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3), *Bull. Am. Meteorol. Soc.*, *82*, 1377–1396.
Barnes, S. L. (1964), A technique for maximizing details in numerical weather map analysis, *J. Appl. Meteorol.*, *3*, 396–409.
Bradley, R. S., 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.
Bussières, N., and W. Hogg (1989), The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid, *Atmos. Ocean*, *27*, 521–541.
Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin (2002), Global land precipitation: A 50-year monthly analysis based on gauge observations, *J. Hydrometeorol.*, *3*, 249–266.
Chen M., et al. (2008), Quality control of daily precipitation report at NOAA/CPC, *AMS 12th conferences on IOAS-AOLS*, New Orleans, LA.
Cressman, G. P. (1959), An operational objective analysis system, *Mon. Weather 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.
Dai, A., and T. M. L. Wigley (2000), Global patterns of ENSO induced precipitation, *Geophys. Res. Lett.*, *27*, 1283–1286.
Dai, A., I. Y. Fung, and A. D. del Genio (1997), Surface observed global land precipitation variations during 1900-88, *J. Clim.*, *10*, 2943–2962.
Ebert, E. E., and M. J. Manton (1998), Performance of satellite rainfall estimation algorithms during TOGA COARE, *J. Atmos. Sci.*, *55*, 1537–1557.
Ebert, E. E., J. E. Janowiak, and C. Kidd (2007), Comparison of near-realtime precipitation estimates from satellite observations and numerical models, *Bull. Am. Meteorol. Soc.*, *88*, 47–64.
Gandin, L. S. (1965), *Objective Analysis of meteorological fields.* Israel Program for Scientific Translations, 242 pp.
Higgins, R. W., and W. Shi (2000), Dominant factors responsible for interannual variability of the summer monsoon in the southwestern United States, *J. Clim.*, *13*, 759–776.
Higgins, R. W., and W. Shi (2001), Intercomparison of the principal modes of interannual and intraseasonal variability of the North American Monsoon System, *J. Clim.*, *14*, 403–417.
Higgins, R. W., W. Shi, E. Yarosh and R. Joyce (2000), Improved United States precipitation quality control system and analysis. *NCEP/Climate Prediction Center ATLAS No. 7.*
Higgins, R.W., W. Shi and C. Hain (2004), Relationships between Gulf of California moisture surges and precipitation in the southwestern United States. *J. Climate* (in press).
Huffman, G. J., et al. (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, *78*, 5–20.
Hulme, M. (1991), An intercomparison of model and observed global precipitation climatologies, *Geophys. Res. Lett.*, *18*, 1715–1718.
Janowiak, J. E., and P. Xie (1999), CAMS-OPI: A global satellite-rain gauge merged product for real-time precipitation monitoring applications, *J. Clim.*, *12*, 3335–3342.
Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie (2004), CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, *J. Hydrometeorol.*, *5*, 487–503.
Legates, D. R. (1987), A climatology of global precipitation. Publ. Climatol., *40*, 85 pp.
Legates, D. R., and C. J. Willmott (1990), Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatol.*, *10*, 111–127.
Maurer, E. P., A. W. Wood, J. C. Adam, and D. P. Lettenmaier (2002), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *J. Clim.*, *15*, 3237–3251.
McCollum, J., W. F. Krajewski, R. R. Ferraro, and M. B. Ba (2002), Evaluation of biases of satellite rainfall estimation algorithms over continental United States, *J. Appl. Meteorol.*, *41*, 1065–1080.
Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang (1995), The uncertainty of simple averages using rain gauge networks, *Water Resour. Res.*, *31*, 2011–2017.
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. Clim.*, *12*, 829–856.
New, M. G., M. Hulme, and P. D. Jones (2000), Representing twentieth century space-time climate variability. Part II: Development of a 1901-96 monthly grids of terrestrial climatology, *J. Clim.*, *13*, 2217–2238.
Roads, J. O., S. C. Chen, and F. Fujioka (2001), ECPC's weekly to seasonal global forecasts, *Bull. Am. Meteorol. Soc.*, *82*, 639–658.
Ropelewski, C. F., J. E. Janowiak, and M. S. Halpert (1985), The analysis and display of real time surface climate data, *Mon. Weather Rev.*, *115*, 1101–1106.
Rudolf, B. (1993), Management and analysis of precipitation data on a routine basis, *Int. Symp. On Precipitation and Evaporation*, Bratislava, Slovakia, WMO, 69–76.
Rudolf, B., H. Hauschild, W. Rueth, and U. Schneider (1994), Terrestrial precipitation analysis: Operational method and required density of point measurements, *Global Precipitation and Climate Change*, NATO ASI Seies, *1*(26), 173–186, M. Desbois and F. Desalmand, Eds., Springer Verlag.
Schneider, U. (1993), The GPCC quality-control system for gauge-measured precipitation data, in *Proc. Analysis Methods for Precipitation on a Global Scale: Report of a GEWEX Workshop*, Koblenz, Germany, GPCC, WCRP-81, WMO/TD-588, A5–A9.
Shepard, D. (1968), A two dimensional interpolation function for irregularly spaced data, in *Prod. 23rd National Conf. of the Association for Computing Machinery*, Princeton, NJ, ACM, 517–524.
Shi, W., R. W. Higgins, E. Yarosh, and V. E. Kousky (2001), The annual cycle and variability of precipitation in Brazil, in *NCEP/Climate Prediction Center Atlas*, No.9, National Oceanic and Atmospheric Administration, National Weather Service. [Available from NOAA/Climate Prediction Center, 5200 Auth Road, Camp Springs, MD 20746, USA].
Silva, V. B. S., V. E. Kousky, W. Shi, and R. W. Higgins (2007), An improved gridded historical daily precipitation analysis for Brazil, *J. Hydrometeorol.*, *8*, 847–861.
Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite (2000), Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Meteorol. Soc.*, *81*, 2035–2046.
Willmott, C., and K. Matsuura (1995), Smart interpolation of annually averaged air temperature in the United States, *J. Appl. Meteorol.*, *34*, 2577–2586.
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, *Am. Cartogr.*, *12*, 5–16.
Xie, P., and P. A. Arkin (1995), An intercomparison of gauge observations and satellite estimates of monthly precipitation, *J. Appl., Meteorol.*, *24*, 1143–1160.
Xie, P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*, 840–858.
Xie, P., and P. A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Am. Meteorol. Soc.*, *78*, 2539–2558.
Xie, P., and P. A. Arkin (1998), Global monthly precipitation estimates from satellite-observed outgoing longwave radiation, *J. Clim.*, *11*, 137–164.
Xie, P., B. Rudolf, U. Schneider, and P. A. Arkin (1996), Gauge-based monthly analysis of global land precipitation from 1971 to 1994, *J. Geophy. Res.*, *101*(D14), 19023–19034.
Xie, P., J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis (2003), GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates, *J. Clim.*, *16*, 2197–2214.
Xie, P., M. Chen, J. E. Janowiak, and P. A. Arkin (2005), Orographic enhancements in precipitation: An intercomparison of two gauge-based precipitation climatologies. 20-24 June 2005, Orange County, CA.
Xie, P., A. Yatagai, M. Chen, T. Hayasaka, Y. Fukushima, C. Liu, and S. Yang (2007), A gauge-based analysis of daily precipitation over East Asia, *J. Hydrometeorol.*, *8*, 607–626.
Xue, Y., et al. (2005), Multiscale variability of the river runoff system in China and its long-term link to precipitation and sea surface temperature, *J. Hydrometeorol.*, *6*, 550–570.
Yatagai, A., P. Xie, and A. Kitoh (2005), Utilization of a new gauge-based daily precipitation dataset over monsoon Asia for validation of the daily precipitation climatology simulated by the MRI/JMA 20-km-mesh AGCM, *SOLA*, *1*, 193–196.

M. Chen, J. E. Janowiak, V. E. Kousky, W. Shi, V. B. S. Silva, R. Wayne Higgins, and P. Xie, Climate Prediction Center, NOAA/NCEP, 5200 Auth Road, RM: 605 Camp Springs, MD 20746, USA. (mingyue.chen@noaa.gov)