JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D12107, doi:10.1029/2008JD011580, 2009

Reconstruction of near-global annual precipitation using correlations with sea surface temperature and sea level pressure

Thomas M. Smith,[1,2] Philip A. Arkin,[2] and Mathew R. P. Sapiano[2]

Received 4 December 2008; revised 8 April 2009; accepted 13 April 2009; published 18 June 2009.

[1] An indirect precipitation analysis method is described which allows analysis of large spatial-scale and multidecadal variations over land and oceans beginning 1900. The method uses covariance between precipitation and analyses of sea level pressure (SLP) and sea surface temperature (SST). Both SLP and SST analyses can be produced using in situ data for the 20th century. Here a canonical correlation analysis is developed to specify annual precipitation anomalies from annual anomalies of SLP and SST on a 5° spatial grid. Covariance relationships are computed using 26 years of satellite-based precipitation data beginning 1979 and are used to analyze annual average precipitation anomalies for the full period. This indirect analysis indicates global variations consistent with the satellite-based analysis for the recent period. Cross-validation testing shows most skill in the tropics where variations are largest, with decreasing skill at higher latitudes, and large-scale averages have much more skill than at individual locations. For the full period over oceans the analysis indicates increasing precipitation with increasing temperature over the 20th century. That oceanic change is correlated with the change from climate models, but the analysis change is more than twice as strong as the change indicated by the models. Over land the analysis is consistent with gauge observations over the 20th century, which are independent observations before 1979. This study shows that indirect precipitation analyses can show many climate-scale variations that cannot be resolved in studies based on direct analysis of precipitation data.

Citation: Smith, T. M., P. A. Arkin, and M. R. P. Sapiano (2009), Reconstruction of near-global annual precipitation using correlations with sea surface temperature and sea level pressure, *J. Geophys. Res.*, *114*, D12107, doi:10.1029/2008JD011580.

1. Historical Data and Analyses

[2] Global mean surface temperature has increased substantially during the past 100 years [e.g., *Trenberth et al.*, 2007], and projections based on current climate models [*Randall et al.*, 2007] point to accelerating increases in the future. Global changes in precipitation have not been measured with the same confidence as those in temperature, but recent observational studies based on the past 20–30 years [*Wentz et al.*, 2007; R. F. Adler, personal communication, 2008] have found increases that correlate well with the increases in atmospheric moisture expected from the changes in temperature. However, theoretical considerations [*Held and Soden*, 2006] suggest that the increases in precipitation should be less than that found in recent observations. Regional and global changes in precipitation are expected to be of enormous importance to society, and confidence in the projections of climate models can only be established by verifying the skill of such models in simulating observed changes over the past century. Here we focus on large-scale precipitation variability since 1900, and explore how well the precipitation can be analyzed using the available data. We also discuss some of the issues that must be confronted to verify that our analyses are adequate to validate climate-model simulations.

[3] Historical data from the period before the time when comprehensive satellite data became available are typically not spatially or temporally complete. However, many surface measurements of temperature and precipitation are available over the 20th century, especially over land [*Trenberth et al.*, 2007]. Precipitation information over the oceans is extremely sparse in the presatellite period, before 1979. Here we discuss a way to use the available data to analyze large-scale and annual oceanic and land precipitation, and present results of the analysis.

[4] Some oceanic historical climate analyses have been produced, including sea surface temperature (SST) [*Kaplan et al.*, 1998; *Rayner et al.*, 2003; *Smith et al.*, 2008a] and sea level pressure (SLP) [*Allan and Ansell*, 2006]. Those data sets use the relatively sparse in situ data together with spatial covariance statistics based on the more recent data that include satellite measurements. We call such a historical climate analysis a reconstruction. Typically, reconstructions can only resolve large-scale variations, with spatial scales of hundreds of kilometers and temporal scales interannual or longer. Both monthly SST and SLP can be reconstructed because they have large scales and because oceanic measurements of both are available throughout the 20th century [*Woodruff et al.*, 1998].

[1] SCSB, STAR, NESDIS, NOAA, College Park, Maryland, USA.
[2] CICS, ESSIC, University of Maryland, College Park, Maryland, USA.

0148-0227/09/2008JD011580$09.00

[5] Historical precipitation reconstructions have been computed using similar approaches, using the network of available gauge data [*Xie et al.*, 2001; *Efthymiadis et al.*, 2005; *Smith et al.*, 2008b]. Typically, empirical orthogonal functions (EOFs) of a modern satellite-based analysis are used to define spatial covariance modes. Gauge data are used to find appropriate weights for the set of EOF modes in historical periods. We will refer to this as the reconstruction from gauges (RG) method. The RG method evaluated here uses gauge data from the Global Historical Climate Network (GHCN) [*Vose et al.*, 1998] beginning in 1900. Since historical reconstructions are produced from gauge data, this is a direct reconstruction method. In this study we do not use historical gauge data for reconstructions, but rather we use relationships between precipitation and several better defined historical analyses to indirectly reconstruct the precipitation.

[6] Reconstructing oceanic precipitation using a direct method is more difficult than SST or SLP. Both SST and SLP have relatively large spatial-temporal scales, so the available historical observations of these variables from ships may be used to reconstruct climate-scale variations. With precipitation times scales are shorter for individual events so intermittent ship sampling is much less useful for reconstruction. In addition, there are many fewer quantitative oceanic precipitation observations, compounding the problem. There are intermittent historical ship observations of qualitative variables such as present weather and cloudiness, which have been used to evaluate some oceanic precipitation characteristics [*Petty*, 1995]. At any place there may be one or a few historical observations. By their nature these variables change rapidly over time, and they are noisier and more difficult to use than SST or SLP. Rain gauges at fixed locations average many observations each month and thus are much less noisy than the marine observations, but gauge data are available only from continental and island locations and they can only be used to reconstruct variations on the largest scales.

[7] Indirect precipitation reconstructions may be produced using the covariance between precipitation and other variables. Here we evaluate a canonical correlation analysis (CCA) [e.g., *Barnett and Preisendorfer*, 1987], used to indirectly reconstruct precipitation from combined SST and SLP predictors. These predictor fields exploit the physical relationships between them and precipitation on interannual or longer time scales. Low SLP is associated with convergence and more precipitation [e.g., *Sapiano et al.*, 2006]. In the tropics high SST is associated with enhanced atmospheric convection [*Graham and Barnett*, 1987]. The CCA is used to evaluate the climate-scale associations of these two with precipitation. As with the RG method, statistical relationships are developed using satellite-derived precipitation data from the recent period. Those relationships are then used with historical SLP and SST data to define historical precipitation. The training period for the CCA may incorporate gauges as part of the modern analysis. However, no gauge data are used in the historical reconstruction. We will refer to this as the CCA method.

2. Satellite Base Analyses

[8] Statistical reconstructions require high-quality base data from which statistics are computed. For oceanic analyses only satellites provide enough spatial coverage for computing these statistics, so satellite-based analyses are used. Several analyses are considered for computing analysis statistics. The optimum interpolation (OI) of *Sapiano et al.* [2008] combines only passive microwave and ERA-40 forecast precipitation to form a homogeneous analysis. However, because of limits in the availability of the homogeneous input data, the OI is only available for ten full years, 1992–2001. This is not long enough to confidently define statistics for multidecadal variations.

[9] There are several merged analyses beginning 1979, when infrared-based satellite estimates of precipitation with near global coverage first became available [*Arkin and Ardanuy*, 1989]. One complete global analysis is the Global Precipitation Climatology Project (GPCP) analysis [*Huffman et al.*, 1997; *Adler et al.*, 2003]. It is more complex than the OI since it uses a changing mix of inputs, including gauge observations, over different regions and at different times. The GPCP is a high-quality analysis intended for climate studies, and we use it here to develop base statistics and for some validation of the CCA. This analysis incorporates many observations and thus should be more accurate than model precipitation estimates. Reanalyses, which do incorporate many observations, are not used for base data because they have known problems with convective precipitation [*Sapiano et al.*, 2008]. A potential problem with GPCP is its use of multiple satellites. There are adjustments between satellites to minimize biases, and the data do not have obvious biases associated with changing satellite inputs (e.g., see the time series of Figures 1 and 2 below). We intend to continue work on producing more homogeneous analyses using satellite data, which could strengthen the results presented here based on GPCP. However, because of the GPCP's relative stability, it is unlikely that using a more homogeneous base analysis will change the major results of this study.

[10] Another analysis considered is the Climate Prediction Center Merged Analysis of Precipitation (CMAP) [*Xie and Arkin*, 1997]. This combines different satellite estimates to form a precipitation map, which is then adjusted using a fit to gauge data, including island gauges. Over the tropical oceans only a few gauge data are used to make large-scale adjustments. These large-scale adjustments are useful for producing month-to-month variations consistent with the limited gauge data, and this helps to makes CMAP useful for analysis of El Niño/Southern Oscillation (ENSO) variations. However, the adjustment can adversely affect the multidecadal signal. For example, in CMAP the multidecadal signal has the opposite sign as the GPCP multidecadal signal [*Yin et al.*, 2004]. Thus, we use GPCP here. Each of these satellite-based analyses covers roughly 25% of the reconstruction period, which begins 1900.

3. Canonical Correlation Analysis

[11] Canonical correlation analysis (CCA) is used to identify and quantify associations between two sets of variables. It does so by finding the maximum correlation between linear combinations of the two sets of variables. The maximization criterion in the basis training period leads to an eigenvalue problem that generates a basis for the reconstruction. We use the CCA method in the EOF spectral

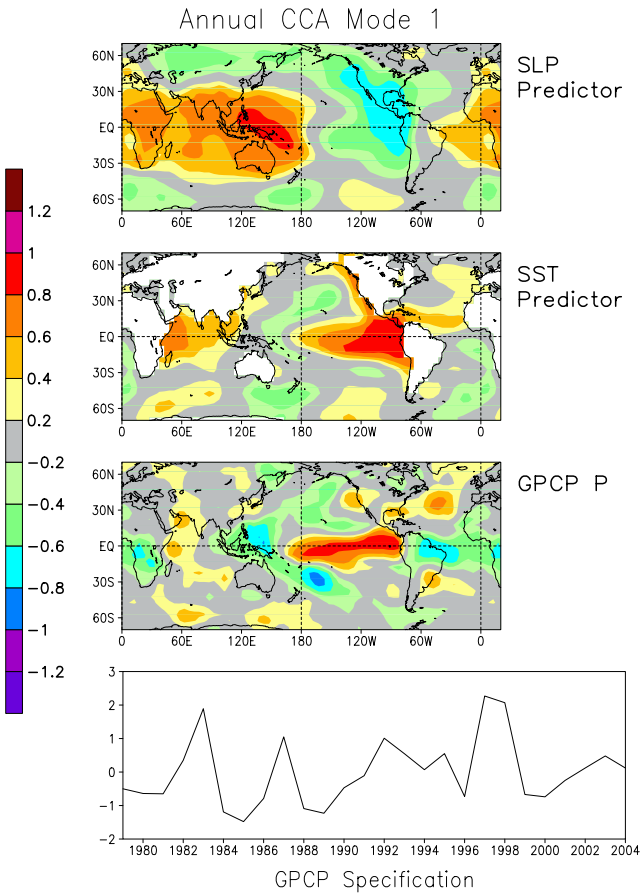

Figure 1. First canonical correlation analysis (CCA) mode computed using sea level pressure (SLP) and sea surface temperature (SST) as predictors for the Global Precipitation Climatology Project (GPCP) precipitation for the 1979–2004 training period.

space developed by *Barnett and Preisendorfer* [1987]. The EOF decomposition helps to smooth the analysis and filter out noise in the training period data. An analysis is found by projecting predictor data onto the predictor CCA modes, to compute weights for the predictand CCA modes. A weighted sum of the predictand modes gives the analysis. Details, including equations fully describing the CCA, are given in the appendix of *Barnett and Preisendorfer* [1987].

[12] In our CCA, anomalies of annual average SLP and SST are the predictors used to reconstruct simultaneous annual average precipitation anomalies. The direct result of the CCA is a normalized specification of precipitation. We remove the normalization by multiplying the analysis by the training period standard deviation, yielding a dimensional precipitation anomaly.

[13] Initially monthly and annual average reconstructions were tested using CCAs, and we found that most CCA skill was associated with interannual and longer variations. Thus, the CCAs discussed here use annual average data. The annual average training data have also been slightly smoothed spatially as described below, in order to focus on large-scale interannual and longer-period variations. We begin with monthly and 5° spatial data. Spatial smoothing for the CCA consists of averaging the 5° spatial data over moving 15° regions, along with annual averaging over calendar years. The spatial smoothing is done to further concentrate the CCA modes on large-scale variations. Reconstructions are computed for the near-global region, 75°S–75°N. Poleward of this area the satellite-based data needed to form statistics may not be reliable.

[14] The GPCP precipitation data used for training the CCA begin in 1979, which is therefore the first year of the CCA training period. We use the SLP historical analysis of *Allan and Ansell* [2006], which ends with 2004. Thus we use as the 26-year training period 1979–2004. There are updates of the historical SLP analysis for several years beyond 2004, but the updates are computed differently and have greater variance than the historical values. In the future we may obtain an improved SLP analysis for recent years, which would allow the CCA training period to be extended several years. The SST historical analysis used for training and analysis with the CCA is the oceanic component of *Smith et al.* [2008a].

[15] As discussed above, the predictor and predictand fields are both filtered using a set of EOF modes, and those filtered fields are used for the CCA. The analysis is itself produced by a weighted sum of CCA modes. It is necessary to set the number of modes for filtering. The maximum number of CCA modes that can be used is the minimum of the number of filtering modes for the two fields, predictor and predictand. Here for simplicity we use the same number

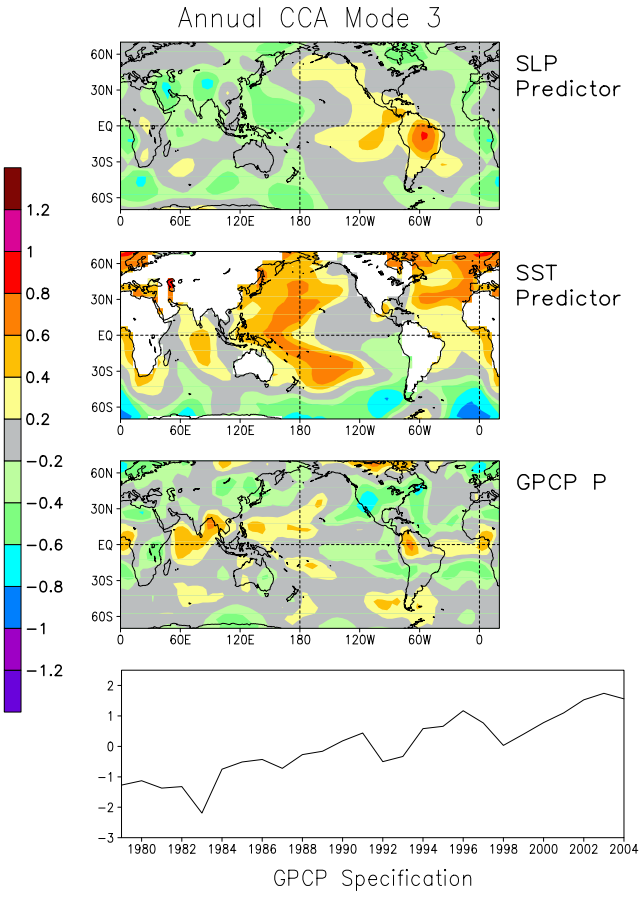

Figure 2. Third CCA mode computed using SLP and SST as predictors for the GPCP precipitation for the 1979–2004 training period.

of predictor and predictand modes and tested the analysis to determine a suitable number of modes to use for analysis. That simplifying assumption is adequate for the present study, which is intended to show the usefulness of this approach. We are in the process of developing an improved base period analysis for the satellite period, similar to the analysis of *Sapiano et al.* [2008] but for a longer period. That improved analysis will be used to recompute the CCA reconstruction, using additional tuning to ensure that we use the optimal number of modes for analysis. As discussed in the summary below, the CCA with improved data and tuning will be combined with other reconstructions to produce an improved merged 20th-century reconstruction.

[16] We tested using up to 10 modes for the CCA. However, the version of CCA used here has a cut off for the number of modes used. If a CCA mode explains less than 1% of the amount of variance explained by the first CCA mode, then that and all higher-CCA modes are not used in the analysis. This truncation of the maximum number of modes is used to minimize noise in the analysis. Testing showed that for our analyses, the code consistently used fewer than 10 modes. We tested using monthly, seasonal, and annual average data. In this analysis 7 modes passed this test for inclusion, which is the number of modes used here for both EOF filtering of the data and for the CCA analysis.

[17] Examination of the 7 CCA modes shows that the first 6 account for most of the variance explained by the CCA, with a large drop off in variance for mode 7. The mode 6 variance is 56% of the mode 1 variance while the mode 7 variance is 16% of the mode 1 variance. Thus, it is possible that additional tuning would eliminate mode 7. Because that mode accounts for little variance its contribution to the CCA is small, so more detailed tuning would produce slight if any changes to the reconstruction.

[18] The first two modes indicate variations associated with ENSO variations. For example, the first mode (Figure 1) shows typical ENSO SLP and SST variations, and the associated precipitation variations. The time series indicates positive peaks associated with warm episodes in 1983, 1987, and 1997–1998. Cool episodes are less marked in the time series of this mode, although there are troughs associated with cool conditions in 1984–1985, 1988–1989, and 1999–2000.

[19] Most multidecadal variance in the training period is associated with the third CCA mode (Figure 2). Some multidecadal variations can be mixed in with the other CCA modes, but examination of the full set of modes showed its concentration in mode 3. Mode 3 indicates decreasing SLP over most of the Earth, with increases over central Africa, South America, and the tropical and north Pacific region near the coast of the Americas. The SST associated with this mode indicates warming except in the Southern Ocean and southeast Pacific, where cooling is indicated. Precipitation associated with this mode is increasing in the tropical Indian and west tropical Pacific oceans, consistent with the GPCP-based analysis of *Smith et al.* [2006].

[20] Decreasing precipitation associated with the third CCA mode is indicated over the east tropical Pacific, east Africa, in many midlatitudes regions in both hemispheres. The decrease in east African precipitation has been linked to warming in the Indian Ocean, which affects the large-scale atmospheric circulation over the region [*Funk et al.*, 2008]. Over the Northern Hemisphere, models have indicated drought conditions associated with cool east Pacific and warm west Pacific SSTs, which combined with warm Indian SSTs force drought conditions [*Hoerling and Kumar*, 2003]. These SST patterns are all evident in CCA mode 3. In addition, *Seidel et al.* [2008] show evidence that midlatitude dry zones are shifting poleward associated with global warming, consistent with the midlatitude drying trend for many regions in CCA mode 3. Over the U.S. area, dust bowl drought conditions have been shown to be associated with cool east Pacific and warm Atlantic SSTs [*Schubert et al.*, 2004], which again is evident in this mode. In the CCA analysis conditions associated with this mode can be modified by the other six CCA modes, but this is clearly the dominant mode associated with multidecadal precipitation variations in the base period.

[21] A weakness of the CCA reconstruction is that only seven modes are used, and variations that do not fit this limited set of modes cannot be resolved. Thus, we can only hope to resolve large-scale variations, and we cannot specify on small scales with this method. In addition, the modes themselves are only as good as the base data used to compute them and any residual biases in the GPCP will contaminate the modes and the analysis. However, where we can validate the CCA it appears to reliably represent large-scale variations, as discussed below. In the future we hope to further validate and improve this analysis using new satellite-based analyses.

4. Results

[22] Results of the CCA are discussed to illustrate average variations and the analysis skill. These results show that many of the large-scale historical variations can be resolved using this indirect reconstruction.

4.1. Near-Global Averages

[23] First the near global (75°S–75°N) average variations are considered. This is the full CCA analysis region, and averages are used to show how well the analysis does on the largest scales. Averages over land areas are computed with subsampling to use only data in locations where there are GHCN gauge values. The GHCN samples most regions, and using all available land sampling would yield similar comparisons, but we use common sampling to avoid sampling differences. Variations are shown for the CCA, for the GHCN gauges, and for the GPCP training data (Figure 3).

[24] For 1979–2004 all three are dependent. Gauge data are used in GPCP, although the GPCP gauges are slightly different, and the GPCP is used to train the CCA for that period. In that period all three show similar large-scale variations when averaged over common-sampling land regions. For any two of the three, the correlation between them is about 0.7 over 1979–2004. Before 1979 the CCA and GHCN are independent, since the CCA does not use gauges as a predictor. In the independent period, averages over common-sampling land regions for the CCA and GHCN are almost as strongly correlated (correlation = 0.6). These correlations include the effects of interannual and multidecadal variations. The GPCP shows larger var-

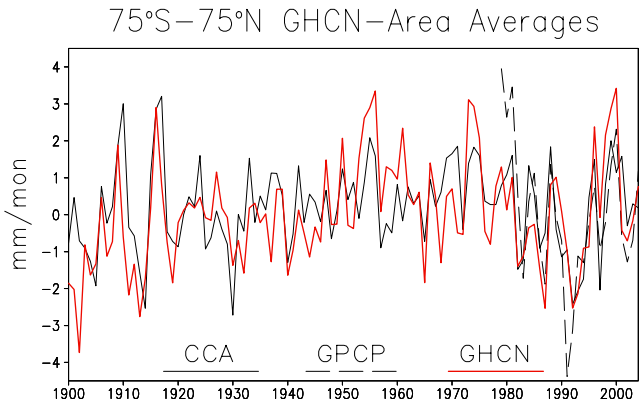

Figure 3. Averages over Global Historical Climate Network (GHCN) gauge locations, 75°S–75°N, from CCA, GPCP, and GHCN.

iations than either the CCA or GHCN in several years. These variations are due to the inclusion of satellite data in GPCP, and it is not clear which data set, GHCN or GPCP, is more reliable in those years.

[25] Similar averages are computed over ocean areas (Figure 4). These ocean areas exclude 5° grid squares that contain any land to avoid contamination by GPCP gauge and land satellite values. Again the CCA and GPCP show similar variations (correlation = 0.8). Over the entire period the CCA ocean average is closely related to the average SST anomaly (see Table 1). This shows a close relationship between interannual to multidecadal global precipitation and global SST, consistent with climate model studies [e.g., *Held and Soden*, 2006; *Randall et al.*, 2007].

[26] Combined land and ocean CCA averages are similar to the ocean averages discussed above. For the recent period the GPCP global average shows an increase of roughly 2% per °C global average temperature [*Adler et al.*, 2008]. Here we measure this percent change using a linear fit of average precipitation to average land and ocean global temperature, obtained from NCDC. The precipitation is computed as a percent of the 1992–2001 GPCP mean. Using the GPCP data this yields a change of 1.8% per °C over 1979–2004, while for the same period the CCA change is 3.1% per °C. The slightly stronger change for CCA may be because the CCA does not resolve the large-amplitude anomalies shown in Figure 3, which may reduce the overall tendency in the relatively short GPCP time series. Over the entire 1900–2004 period the CCA change is 2.4% per °C.

[27] Comparisons to climate model precipitation are done using an ensemble average from models used in the IPCC Assessment Report 4 (AR4) [*Randall et al.*, 2007]. The near-global averages from the climate models are compared to similar averages from the CCA, from SST, and from GHCN gauges (Table 1). The AR4 climate model uses greenhouse gas forcing, but it is not phase linked to ENSO episodes in the observations. Thus, for these comparisons the regional and annual averages are smoothed using an 11-year binomial filter, to minimize interannual variations while maintaining multidecadal variations. The ocean area averages exclude 5° grid squares that contain any land while the land area averages exclude grid squares that are not sampled by the GHCN. Note that correlations against SST are always correlations with the global ocean area temperature, including several averages over land regions. Similarly, correlations against GHCN are always correlations against land area precipitation.

[28] Over oceans averages of both CCA and AR4 are strongly correlated to SST averages. This shows that multidecadal SST variations control much of the global average multidecadal precipitation variations. Over land the CCA and GHCN correlation is high, but the AR4 average land precipitation is weakly negatively correlated with CCA and GHCN. This indicates problems with the multidecadal AR4 estimates over land. The AR4 climate models may have trouble simulating multidecadal land precipitation variations, where changing land surface properties can complicate simulations.

[29] Over all regions the CCA and AR4 estimates are correlated, but the CCA change with temperature is much stronger than the AR4 change. As noted above, over land and oceans combined the overall CCA change is 2.4% per °C, but the comparable AR4 change is 0.9% per °C, or roughly a third as strong. More work is needed to resolve causes for this difference.

4.2. Validation Testing

[30] Most skill in the CCA comes from specification of warm and cool ENSO episodes. In years with an episode the CCA closely resembles GPCP, while in non-ENSO years the resemblance is not as good. To quantify and better understand the overall skill, and differences between land and ocean area skill, cross-validation testing is done. The cross-validation CCA is computed by analyzing each year in the training period using a set of modes computed excluding data from that year. The gives a roughly independent set of modes for the analysis of each year. This is done for the full 26-year period, and results are compared to the GPCP data using temporal correlation at each location. Because this is a correlation of individual locations and not a correlation of the global average, values are lower than in the comparisons discussed above.

[31] Zonal averages of the cross-validation correlation skill show that skill is highest in the tropics and least at high latitudes (Figure 5). These spatial averages of correlations at individual locations are much lower than correla-

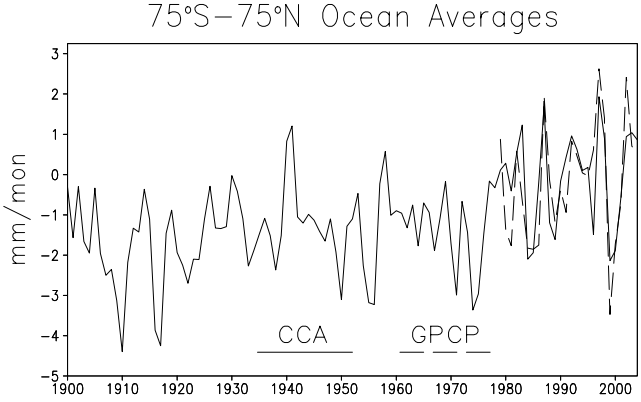

Figure 4. Averages over ocean locations, 75°S–75°N, from CCA and GPCP.

Table 1. Correlations of Annual and Near-Global (75°S–75°N) Averages of Combinations of GHCN, SST, CCA, and AR4 Precipitation Estimates for the 20th Century[a]

	GHCN Land	Ocean	All Regions
GHCN, CCA	0.64	−0.11	0.44
GHCN, AR4	−0.35	0.44	0.36
SST, CCA	−0.08	0.81	0.91
SST, AR4	−0.32	0.85	0.84
CCA, AR4	−0.24	0.68	0.79

[a]GHCN, Global Historical Climate Network; SST, sea surface temperature; CCA, canonical correlation analysis; AR4, Intergovernmental Panel on Climate Change Assessment Report 4. Averages are computed over GHCN land regions, over ocean regions, and over all regions, as indicated. All time series are smoothed with an 11-year binomial filter.

tions of the spatial averages. The land cross-validation correlation skill is lower than over oceans (Table 2), suggesting that we can trust oceanic variations at least as much as the land variations.

[32] Most of the cross-validation skill is concentrated in the low latitudes (Figure 5), where precipitation variations are strongest. The low latitudes have more ocean than land, compared to the Northern Hemisphere high latitudes, which helps to explain the larger ocean global average in Table 2. Poleward of roughly 45° the CCA is less reliable. This is not surprising since the satellite-based GPCP is also most reliable at low latitudes where convective precipitation is dominant. Except where there are gauges to anchor it, such as at some Northern Hemisphere locations, the high-latitude GPCP may be noisier and thus covariations with the predictor data at high latitudes may be lower. Future satellite-based analyses may improve this high-latitude precipitation, allowing for more reliable reconstructions in those regions.

[33] Since the ocean area CCA is highly correlated with the SST, a test CCA was produced using only SST as a predictor. The global average of the correlation skill of this analysis is compared to the average correlation of the CCA using both SST and SLP predictors. For these comparisons correlation skill against GPCP is computed without cross-validation, giving higher correlations (Table 3). This comparison shows that much of the correlation skill is derived from the SST, but the SLP noticeably improves the correlation.

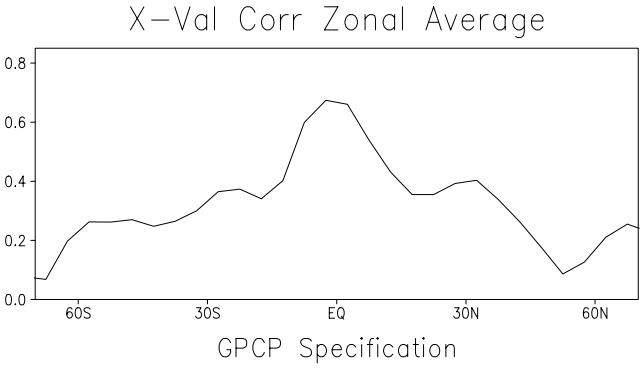

Figure 5. Zonal average cross-validation correlation skill for the CCA.

Table 2. Global Average CCA Correlation Cross-Validation Skill Averaged Over Land, Sea, and All Areas

	Area		
	Land	Sea	All
Correlation	0.27	0.37	0.36

[34] If we square the correlations in Table 3 and compare variance explained by the CCAs, we see that the SST CCA explains roughly half the variance of the SST and SLP CCA over oceanic regions. If SST and SLP were linearly dependent then adding SLP would add no additional skill to the CCA, so the SST and SLP must be contributing independent information to the CCA. Over land the SST CCA explains even more of the SST and SLP variance, suggesting that most of the skill over land is derived from teleconnections from oceanic regions.

4.3. Climate Variations

[35] Some regional climate tendencies in the CCA are discussed here. First the tropics are considered (Figure 6). Averages are computed over all regions and over ocean regions only. In the tropics both show an increasing tendency over the analysis period. However, the ocean area averages have much more interannual variability because of ENSO variations.

[36] The tropical averages reflect the increasing global precipitation with temperature, discussed above. To see if the subtropical dry zones are changing, similar averaging is done for the regions 35°S–25°S (Figure 7) and 25°N–35°N (Figure 8). In the southern subtropical zone there are downward tendencies over the entire record, with the tendency increasing after 1970. In the northern subtropical zone there is little tendency until after 1980 when a steep decrease begins. These decreases are consistent with those of *Seidel et al.* [2008], who find evidence for widening of the subtropical subsidence zones with global warming. Differences in the southern and northern zones indicate that in the CCA the zones in the two hemispheres are not completely symmetric. The analysis of the southern indicates steadier drying while the northern zone change is more abrupt. Hemispheric differences may be due to the hemispheric land-sea differences and the influence of different climate modes affecting the hemispheres. For example, Northern Hemisphere CCA rainfall oscillation apparent in Figure 8 may be influenced by the Atlantic multidecadal oscillation [*Enfield et al.*, 2001].

Table 3. Global and Tropical (25°S–25°N) Average CCA Correlation Skill, Averaged for Analysis Using SST and SLP Predictors and for Analysis Using Only SST Predictors[a]

	Predictors	
	SST and SLP	SST
Land global	0.60	0.48
Sea global	0.66	0.47
All global	0.64	0.47
Land tropics	0.60	0.52
Sea tropics	0.73	0.51
All tropics	0.70	0.51

[a]SLP, sea level pressure.

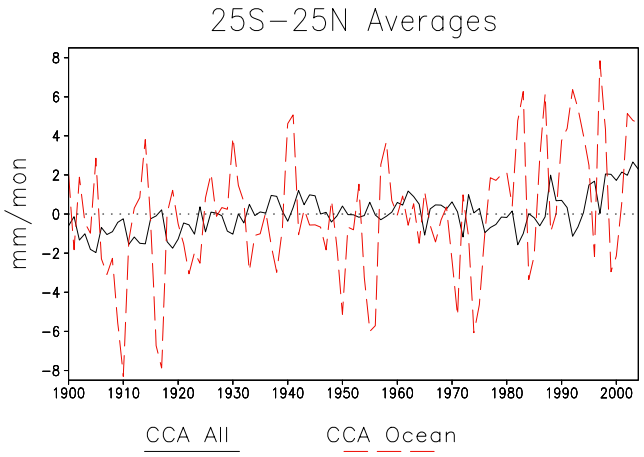

Figure 6. CCA precipitation averaged between 25°S and 25°N and all longitudes for all areas and for only ocean areas. The 1900–2004 mean is removed.

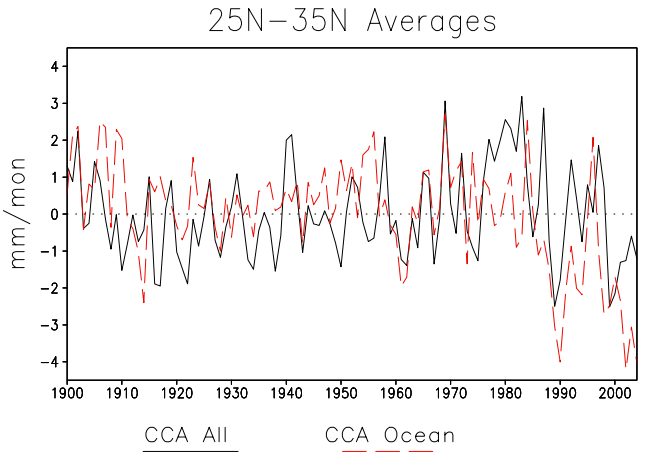

Figure 8. CCA precipitation averaged between 25°N and 35°N and all longitudes for all areas and for only ocean areas. The 1900–2004 mean is removed.

[37] The northern extratropics (Figure 9) have a pattern similar to the northern subtropics. Both show a slight decrease in precipitation from 1900 until the late 1930s, when it begins increasing until the 1980s, followed by a decreasing tendency. In the subtropics (Figure 8) the decrease since 1980 is more pronounced than the early century decrease. In the extratropics (Figure 9) the early century and recent decreases are comparable. In the southern extra tropics (50°S–35°S, not shown) the CCA indicates increasing precipitation correlated with the global SST. That may be because there is almost no land in that region, making the response more closely linked to the SST.

[38] In these regions, the comparable average GPCP anomalies typically correlate well with the CCA anomalies. Highest correlation is over the tropical oceans (0.8) and lowest correlation is over the extratropical Northern Hemisphere oceans (0.5). In other regions the correlations are 0.6 to 0.7. Both GPCP and CCA indicate major interannual variations in the overlap region. However, the GPCP average anomalies tend to be more variable than the CCA anomalies, which is expected since the CCA filters data using seven modes while the GPCP is based on satellite data.

[39] For all of these regions the CCA variance in 1979–2004 is slightly higher than the variance in 1953–1978, over both oceans and land and ocean areas combined. Since the CCA is linear and uses SLP and SST anomalies as predictors, this must be caused be increased variance of those predictors. Compared to earlier periods, those reconstructed predictors may have slightly more variance in the more recent period when there are more in oceanic situ data, so sampling changes could be contributing to the increased variance. However, oceanic in situ sampling generally good after 1950 [e.g., *Smith et al.*, 1996], and variance changes are most likely associated with stronger ENSO episodes in the more recent period [*Xue et al.*, 2003].

[40] At the time of the early century extratropical decrease over northern mid latitudes, there were serious droughts that contributed to dust bowl conditions in the U.S. The CCA suggests that the northern extratropical precipitation is again headed toward dryer conditions, which may be intensified by climate change. A dust bowl area is

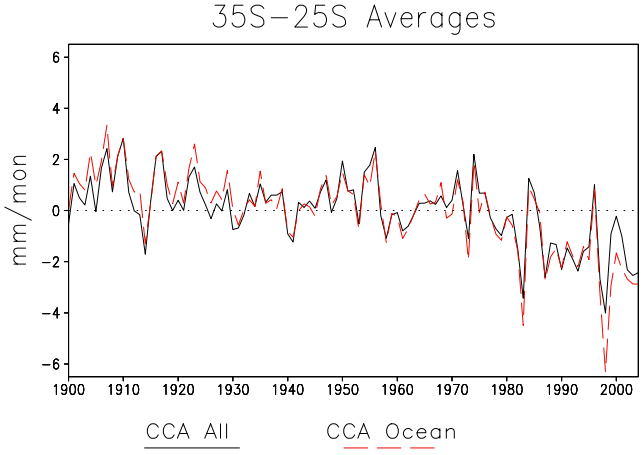

Figure 7. CCA precipitation averaged between 35°S and 25°S and all longitudes for all areas and for only ocean areas. The 1900–2004 mean is removed.

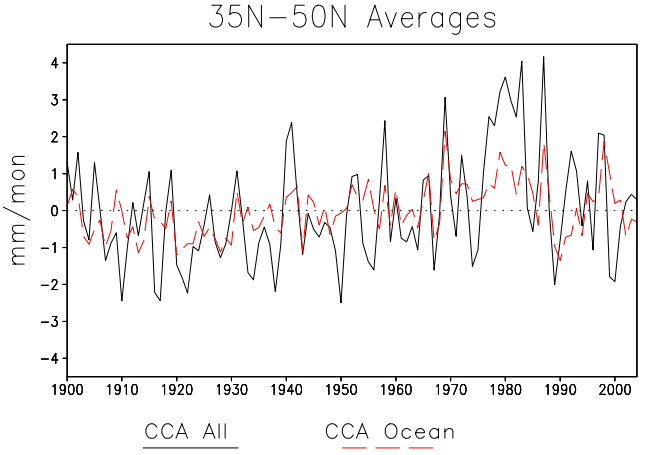

Figure 9. CCA precipitation averaged between 35°N and 50°N and all longitudes for all areas and for only ocean areas. The 1900–2004 mean is removed.

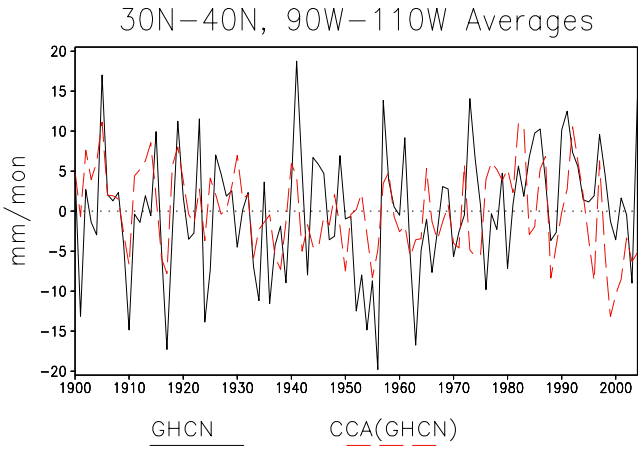

Figure 10. Precipitation averaged between 30°N and 40°N and between 90°W and 110°W from GHCN and the CCA for GHCN sampling areas. The 1900–2004 mean is removed.

approximated by 30°N–40°N and 90°W–110°W. Averages of the GHCN and of the CCA at GHCN sampling regions are shown for this area (Figure 10). There is much more variance in this local region, and the GHCN is more variable than the CCA. However, both averages show decreases in the first part of the century, including fairly persistent dry conditions in the 1930s, followed by wetter conditions until about 1990 when the region began shifting toward dryer conditions.

5. Summary and Conclusions

[41] A method for reconstructing near-global 20th-century precipitation is presented and evaluated. The reconstruction yields multidecadal results consistent with what is expected from climate models for a warming world, but the reconstruction precipitation changes are several times larger than the model changes. When the reconstruction can be compared to independent gauges, the global variations are well correlated. In addition, validation testing suggests that the reconstruction skill over oceans should be at least as good as results over land.

[42] The reconstruction is limited to the analysis of large-scale annual average features that may exhibit covariance with the predictor SST and SLP analyses. In addition, the shape and magnitude of the reconstructed fields are determined by variations in the GPCP training data. Because there are errors in the predictor and GPCP analyses, there will be errors in the reconstruction. Thus, it may not be suitable for giving local details of precipitation variations, and if the GPCP data are biased then the reconstruction based on it will also tend to be biased. In addition, the CCA uses only seven modes and therefore can only resolve large-scale variations.

[43] Reconstructed precipitation in the tropics yields results similar to the global averages, with variations correlated with the global increase in temperature. In the subtropics and northern extratropics there are decreases in precipitation, which are especially pronounced since roughly 1980.

[44] In an earlier study [*Smith et al.*, 2008b] we showed that interannual-scale variations could be reconstructed over most of the earth using a method based on fitting gauge data to sets of spatial modes. However, that method was not able to reliably reconstruct multidecadal signals. This indirect reconstruction of precipitation is able to resolve much of the large-scale multidecadal signal.

[45] In the future we plan to merge a reconstruction based on fitting gauge data to modes with a CCA-based reconstruction, to obtain reconstructed variations for interannual to multidecadal signals beginning 1900. Before merging the analyses we will develop improved satellite era analyses in an attempt to produce a long base analysis that may be more homogeneous than the GPCP multisatellite and in situ analysis. This improved analysis should be of value for climate studies and validating climate models used to simulate future changes in precipitation. We also intend to compare results to other long 20th-century precipitation estimates, such as those based on the method of *Compo et al.* [2006]. These intercomparisons of estimates should help to bracket the range of estimates and better indicate what amount of confidence may be placed in these estimates. The improved analyses will be used to further investigate historical precipitation variations and comparison of those variations to those obtained from models. Such analyses and comparisons should yield an improved understanding of the global hydrologic cycle and its representation in climate models.

[46] **Acknowledgments.** The SLP data were supplied by the U.K. Met. Office Hadley Centre, www.metoffice.gov.uk/hadobs. We thank R. Reynolds, A. Mariotti, and two anonymous reviewers for reviews of earlier drafts and helpful comments. We also thank C.-Y. Chang for assistance analyzing the CCA output. This project is supported in part by the Climate Change Data and Diagnostic Program Element of the NOAA Climate Program Office and the Cooperative Institute for Climate Studies, NOAA grant NA17EC1483. The contents of this paper are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of NOAA or the U. S. Government.

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., G. Gu, J.-J. Wang, G. J. Huffman, S. Curtis, and D. Bolvin (2008), Relationships between global precipitation and surface temperature on interannual and longer timescales (1979–2006), *J. Geophys. Res.*, *113*, D22104, doi:10.1029/2008JD010536.

Allan, R. J., and T. J. Ansell (2006), A new globally complete monthly historical mean sea level pressure data set (HadSLP2), 1850–2004, *J. Clim.*, *19*, 5816–5842, doi:10.1175/JCLI3937.1.

Arkin, P. A., and P. Ardanuy (1989), Estimating climatic-scale precipitation from space: A review, *J. Clim.*, *2*, 1229–1238, doi:10.1175/1520-0442(1989)002<1229:ECSPFS>2.0.CO;2.

Barnett, T. P., and R. Preisendorfer (1987), Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis, *Mon. Weather Rev.*, *115*, 1825–1850, doi:10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh (2006), Feasibility of a 100-year reanalysis using only surface pressure data, *Bull. Am. Meteorol. Soc.*, *87*, 175–190, doi:10.1175/BAMS-87-2-175.

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

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., *Geophys. Res. Lett.*, *28*, 2077–2080, doi:10.1029/2000GL012745.

Funk, C., M. D. Dettinger, J. C. Michaelsen, J. P. Verdin, M. E. Brown, M. Barlow, and A. Hoell (2008), Warming of the Indian Ocean threatens

eastern and southern African food security but could be mitigated by agricultural development, *Proc. Natl. Acad. Sci. U. S. A.*, *105*, 11,081–11,086, doi:10.1073/pnas.0708196105.

Graham, N. E., and T. P. Barnett (1987), Sea surface temperature, surface wind divergence, and convection over tropical oceans, *Science*, *238*, 657–659, doi:10.1126/science.238.4827.657.

Held, I. M., and B. J. Soden (2006), Robust responses of the hydrological cycle to global warming, *J. Clim.*, *19*, 5686–5699, doi:10.1175/JCLI3990.1.

Hoerling, M., and A. Kumar (2003), The perfect ocean for drought, *Science*, *299*, 691–694, doi:10.1126/science.1079053.

Huffman, G. J., et al. (1997), The Global Precipitation Climatology Project (GPCP) combined data set, *Bull. Am. Meteorol. Soc.*, *78*, 5–20, doi:10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan (1998), Analyses of global sea surface temperature 1850–1991, *J. Geophys. Res.*, *103*, 18,567–18,589, doi:10.1029/97JC01736.

Petty, G. W. (1995), Frequencies and characteristics of global oceanic precipitation from shipboard present-weather reports, *Bull. Am. Meteorol. Soc.*, *76*, 1593–1616, doi:10.1175/1520-0477(1995)076<1593:FACOGO>2.0.CO;2.

Randall, D. A., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis—Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 591–662, Cambridge Univ. Press, New York.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.

Sapiano, M. R. P., D. B. Stephenson, H. J. Grubb, and P. A. Arkin (2006), Diagnosis of variability and trends in a global precipitation dataset using a physically motivated statistical model, *J. Clim.*, *19*, 4154–4166, doi:10.1175/JCLI3849.1.

Sapiano, M. R. P., T. M. Smith, and P. A. Arkin (2008), A new merged analysis of precipitation utilizing satellite and reanalysis data, *J. Geophys. Res.*, *113*, D22103, doi:10.1029/2008JD010310.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2004), On the cause of the 1930s dust bowl, *Science*, *303*, 1855–1859, doi:10.1126/science.1095048.

Seidel, D. J., Q. Fu, W. J. Randel, and T. J. Reichler (2008), Widening of the tropical belt in a changing climate, *Nat. Geosci.*, *1*, 21–24, doi:10.1038/ngeo.2007.38.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes (1996), Reconstruction of historical sea surface temperatures using empirical orthogonal functions, *J. Clim.*, *9*, 1403–1420, doi:10.1175/1520-0442(1996)009<1403:ROHSST>2.0.CO;2.

Smith, T. M., X. Yin, and A. Gruber (2006), Variations in annual global precipitation (1979–2004), based on the Global Precipitation Climatology Project 2.5° analysis, *Geophys. Res. Lett.*, *33*, L06705, doi:10.1029/2005GL025393.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008a), Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*, 2283–2296, doi:10.1175/2007JCLI2100.1.

Smith, T. M., M. R. P. Sapiano, and P. A. Arkin (2008b), Historical reconstruction of monthly oceanic precipitation (1900–2006), *J. Geophys. Res.*, *113*, D17115, doi:10.1029/2008JD009851.

Trenberth, K. E., et al. (2007), Observations: Surface and atmospheric climate change, in *Climate Change 2007: The Physical Science Basis—Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 237–336, Cambridge Univ. Press, New York.

Vose, R. S., T. C. Peterson, and M. Hulme (1998), The Global Historical Climatology Network Precipitation Database: Version 2.0, in *Proceedings of the Ninth Symposium on Global Change Studies*, Am. Meteorol. Soc., Boston, Mass.

Wentz, F. J., L. Ricciarudlli, K. Hilburn, and C. Mears (2007), How much more rain will global warming bring?, *Science*, *317*, 233–235, doi:10.1126/science.1140746.

Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley (1998), COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields, *Phys. Chem. Earth*, *23*, 517–526, doi:10.1016/S0079-1946(98)00064-0.

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, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Xie, P., M. Chen, J. E. Janowiak, P. A. Arkin, and T. M. Smith (2001), Reconstruction of the oceanic precipitation: Preliminary results, *Proc. Annu. Clim. Diagn. Predict. Workshop, 26*. (Available at http://www.cpc.ncep.noaa.gov/products/outreach/publications.shtml)

Xue, Y., T. M. Smith, and R. W. Reynolds (2003), A new SST climatology for the 1971–2000 base period and interdecadal changes of 30-year SST normal, *J. Clim.*, *16*, 1601–1612.

Yin, X., A. Gruber, and P. Arkin (2004), Comparison of the GPCP and CMAP merged gauge–satellite precipitation products for the period 1979–2001, *J. Hydrometeorol.*, *5*, 1207–1222, doi:10.1175/JHM-392.1.

P. A. Arkin and M. R. P. Sapiano, CICS, ESSIC, University of Maryland, College Park, 5825 University Research Court, Suite 4001, College Park, MD 20740, USA.

T. M. Smith, SCSB, STAR, NESDIS, NOAA, 5825 University Research Court, Suite 4001, College Park, MD 20740, USA. (tom.smith@noaa.gov)