A Precipitation Climatology and Dataset Intercomparison for the Western United States

KRISTEN J. GUIRGUIS AND RONI AVISSAR

Department of Civil and Environmental Engineering, Duke University, Durham, North Carolina

(Manuscript received 19 September 2006, in final form 13 June 2008)

ABSTRACT

This paper presents the results of a regionalization study of the precipitation climate of the western United States using principal component analysis. Past eigen-based regionalization studies have relied on rain gauge networks, which is restrictive because rain gauge coverage is sparse, especially over complex terrain that exists in the western United States. Here, the use of alternate data products is examined by conducting a comparative regionalization using nine precipitation datasets used in hydrometeorological research. Five unique precipitation climates are identified within the western United States, which have centers and boundaries that are physically reasonable and that highlight the relationship between the precipitation climatology and local topography. Using the congruence coefficient as the measure of similarity between principal component solutions, the method is found to be generally stable across datasets. The exception is the National Centers for Environmental Prediction–Department of Energy (NCEP–DOE) Reanalysis 2, which frequently demonstrates only borderline agreement with the other datasets. The loading pattern differences among datasets are shown to be primarily a result of data differences in the representation of (i) precipitation over the Rocky Mountains, (ii) the eastward wet-to-dry precipitation gradient that occurs during the cold season, (iii) the magnitude and spatial extent of the North American monsoon signal, and (iv) precipitation in the desert southwest during spring and summer. Sensitivity tests were conducted to determine whether the spatial resolution and temporal domain of the input data would dramatically affect the solution, and these results show the methodology to be stable to differences in spatial/temporal data features. The results suggest that alternate data products can be used in regionalization studies, which has applications for rain gauge installation and planning, climate research, and numerical modeling experiments.

1. Introduction

The western United States is a complex landscape consisting of coastal zones, mountains, basins, and plateaus. Its precipitation climatology is diverse as a result of interactions between orography and atmospheric dynamical processes occurring on multiple temporal and spatial scales. Competing climate controls and local differences in terrain and proximity to coastal moisture sources are such that unique precipitation climates are found within relatively short spatial distances (Mock 1996). This can be challenging for climate modeling because localities in close proximity may respond differently not only to real climate features but also to model errors. For example, model weaknesses associated with convection parameterizations will likely cause more error for the parts of the domain experiencing convective thunderstorm activity, whereas the failure to adequately resolve topography will be a greater problem for regions exhibiting complex terrain. As such, the systematic analysis of simulation results for each unique precipitation climate in the western United States is likely to lend insight into model strengths and weaknesses as well as into the study of local weather or climate characteristics or land-atmosphere feedback.

A useful approach for regionalizing a spatial domain into subdomains based on local precipitation climatology is principal component analysis (PCA). Many studies have used PCA and related eigen techniques to identify patterns of precipitation covariability and isolated regions that are spatially cohesive with respect to their precipitation regimes (e.g., Beaudoin and

Corresponding author address: Roni Avissar, Department of Civil and Environmental Engineering, Edmund T. Pratt School of Engineering, Duke University, P.O. Box 90287, Durham, NC 27708-0287.
E-mail: avissar@duke.edu

DOI: 10.1175/2008JHM832.1

Rousselle 1982; Richman and Lamb 1985; Ehrendorfer 1987; Richman and Lamb 1987; White et al. 1991; Carter and Elsner 1997; Comrie and Glenn 1998; Hawkins et al. 2002; Dinpashoh et al. 2004; Gutzler 2004; Diem 2006; Diem and Brown 2006; Gochis et al. 2006). The regions identified by PCA have been used for many purposes such as identifying persistent patterns of thunderstorm activity (Easterling 1990), developing a statistical rainfall forecasting model (Carter and Elsner 1997), linking synoptic-scale circulation patterns to precipitation anomalies (Diem 2006), developing an index of North American monsoon (NAM) variability (Gutzler 2004), identifying snow–NAM teleconnections (Ellis and Hawkins 2001; Hawkins et al. 2002), elucidating atmosphere–rainfall relationships (Diem and Brown 2006), and many others.

However, past PCA-based regionalization studies have relied on rain gauge networks, which limit the utility of the method because many parts of the world are covered sparsely by, or are void of, these networks. Additionally, for the western United States with its complex terrain, rain gauge coverage tends to be clustered in the low elevations, whereas coverage in orographic zones is sparse. Given that numerous datasets are available for the western United States, including satellite data and reanalysis, it is not clear a priori whether rain gauge data would yield the most representative results in a regionalization study. Additionally, the ability to extend a study domain beyond land areas using global products would be useful for numerical modeling studies because these areas often include an oceanic region. However, all precipitation products contain errors and many are available at a coarse spatial resolution on the order of 2.5° × 2.5°, and it is not clear whether this coarse resolution would affect the regionalization results. Therefore, this study investigates the utility of alternate data products for regionalization studies by conducting a comparative regionalization of the precipitation climate of the western United States using several observation-based precipitation datasets.

There are several sources of precipitation data currently available for the western United States, including measurements from rain gauges, ground-based radar, Earth-orbiting satellites, and reanalysis products. In general, rain gauge data are considered to be the most accurate where they are available, but coverage over mountain regions is sparse and rain gauge data are additionally known to suffer low bias as a result of wind and evaporation (e.g., Legates and Willmott 1990). Ground Doppler radar estimates are available at a relatively high spatial/temporal resolution over much of the central and eastern United States, but mountainous terrain interferes with low-level radar scans, making Doppler precipitation estimates in the western United States problematic (Serafin and Wilson 2000; Maddox et al. 2002). Error in radar estimates also occurs from brightband effects and algorithm misinterpretation of ground clutter (e.g., Klazura et al. 1999). Satellite measurements offer the advantage of increased spatial coverage, including over mountains. Geosynchronous infrared (geo-IR) platforms provide near-complete spatial coverage equatorward of 40°. However, IR-based precipitation estimates are based on cloud-top brightness temperatures, which have only a limited empirical relationship with surface precipitation (Arkin and Meisner 1987; Arkin and Xie 1994). Microwave estimates from polar-orbiting satellites are more physically based but have a poor temporal sampling rate of one or two observations per day (Adler et al. 1993) for a given location. Additionally, microwave-scattering algorithms do not function over snow- or ice-covered surfaces (Xie and Arkin 1997), which limits their use for some northern and high-elevation areas. Merged datasets attempt to retain the desirable spatial/temporal features of the geo-IR datasets while drawing data from other sources for calibration purposes or to fill in data gaps to improve the accuracy of the overall combined dataset (e.g., Xie and Arkin 1997; Adler et al. 2003). The accuracy of these merged products, however, is affected by the shortcomings of the raw input data and by the merging methodology (e.g., Gruber et al. 2000; Adler et al. 2001; Chen et al. 2002; Yin et al. 2004). Global and regional reanalysis products assimilate observations of atmospheric and ocean data and use a numerical model to generate precipitation fields. However, the accuracy of these fields is heavily influenced by the performance of the model and its subgrid-scale parameterizations (Kistler et al. 2001), including those for cumulus convection, which do not perform optimally in complex terrain.

For this precipitation regionalization and climatology study, we consider the bulk of the observation-based precipitation data products that are available for moderately long-term (~15 yr) climatology studies of the western United States. The datasets included in the study are the (i) Global Precipitation Climatology Center (GPCC) monitoring product; (ii) Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset, version 2; (iii) Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP); (iv) CPC retrospective United States and Mexico daily precipitation analysis (USMex); (v) Precipitation-elevation Regressions on Independent Slopes Model (PRISM); (vi) National Centers for Environmental Prediction–Department of Energy (NCEP–DOE) Reanalysis 2

TABLE 1. Main characteristics of the datasets included in this study.

Dataset	Spatial resolution	Temporal domain	Data source
GPCC[a]	2.5° × 2.5°	From 1986	Rain gauge
GPCP[b]	2.5° × 2.5°	From 1979	Rain gauge, satellite
CMAP[c]	2.5° × 2.5°	From 1979	Rain gauge, satellite
NCEP2[d]	210 km × 210 km	From 1979	Reanalysis
GMFD[e]	1° × 1°	1948–2000	Reanalysis, rain gauge, satellite
USMex[f]	1° × 1°	From 1948	Rain gauge
NARR[g]	32 km × 32 km	From 1979	Reanalysis
VIC[h]	1/8° × 1/8°	1950–2000	Rain gauge with orographic adjustment
PRISM[i]	4 km × 4 km	From 1890	Rain gauge with orographic adjustment

[a] Global Precipitation Climatology Center monitoring product
[b] Global Precipitation Climatology Project Combined Precipitation Dataset, version 2
[c] Climate Prediction Center Merged Analysis of Precipitation
[d] National Centers for Environmental Prediction–Department of Energy Reanalysis 2
[e] Global Meteorological Forcing Dataset for land surface modeling
[f] Climate Prediction Center retrospective United States and Mexico daily precipitation analysis
[g] North American Regional Reanalysis
[h] Variable Infiltration Capacity Retrospective Land Surface Dataset
[i] Parameter-elevation Regressions on Independent Slopes Model

(NCEP-2); (vii) North American Regional Reanalysis (NARR); (viii) Variable Infiltration Capacity (VIC) Retrospective Land Surface Dataset; and (ix) Global Meteorological Forcing Dataset (GMFD) for land surface modeling. First, we conduct a series of PCAs using each of the nine datasets rescaled to a common 2.5° × 2.5° grid and quantitatively compare their loading patterns. Next, we regionalize the domain based on the relative strength of gridcell loadings for each dataset, and then we compare the results for the different datasets and physically interpret any differences in the initial precipitation data. Finally, we test the stability of the method to differences in spatial/temporal data features. Section 2 describes the datasets and section 3 discusses rescaling. Section 4 describes the methodology of the PCA and regionalization. Section 5 contains a discussion of the results. Conclusions are given in section 6.

2. Data

The nine datasets included in this study and their spatial/temporal characteristics are shown in Table 1. A brief description is provided below.

a. Global Precipitation Climatology Center monitoring product

The GPCC monitoring product provides monthly precipitation data from 1986 for global land areas based on rain gauge data from approximately 7500 stations (Fuchs et al. 2007). Its source data are surface synoptic weather reports (SYNOP) and monthly climate bulletins (CLIMAT), which are quality controlled by automated and manual processes (Rudolf et al. 1994). The SYNOP provide information on precipitation totals for time intervals ranging from 1 to 24 h. Incomplete monthly time series are not uncommon and can result in local monthly precipitation estimates that are based on less than optimal temporal information (Rudolf et al. 1994; Rudolf and Schneider 2005). Following the quality-control (QC) processes, a precipitation value is assigned to a location based on source reliability information. The data are interpolated to a 0.5° × 0.5° grid and spatially averaged over 2.5° × 2.5° (Rudolf et al. 1994; Rudolf and Schneider 2005).

b. Global Precipitation Climatology Project Combined Precipitation Dataset, version 2

The GPCP (Adler et al. 2003) is a global, merged satellite-gauge data product that is available monthly from 1979 at a 2.5° × 2.5° spatial resolution. The combination method uses a multistep process in which lesser-biased data sources are used to calibrate higher-biased sources to reduce the overall bias of the end product. For latitudes spanning 40° north–south, precipitation estimates from the IR-based Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI; Arkin and Meisner 1987) are compared with Special Sensor Microwave Imager (SSM/I) data when both are simultaneously available. The GPI–SSM/I bias is calculated and used to calibrate the GPI precipitation estimates everywhere. Poleward of 40°, the satellite component is a combination of SSM/I and Television and Infrared Observation Satellite (TIROS)

Operational Vertical Sounder (TOVS) data. The multisatellite data are adjusted to agree with large-scale (5×5 grid cells each at 2.5°) averages of rain gauge data to reduce satellite bias. The corrected satellite data are then merged with rain gauge data using inverse error weighting to incorporate gauge-measured local variability (Huffman et al. 1997). Beginning in 1986, the gauge component of the GPCP is the GPCC monitoring product (section 2a), which is corrected for systematic rain gauge error using Legates and Willmott (1990). In the pre-SSM/I period (before July 1987), satellite estimates are from the outgoing longwave radiation precipitation index (Xie and Arkin 1998).

c. *Climate Prediction Center Merged Analysis of Precipitation*

The CMAP (Xie and Arkin 1996) is a monthly product that is available from 1979 at a $2.5° \times 2.5°$ resolution. It combines precipitation estimates from satellite IR, SSM/I, and Microwave Sounding Unit (MSU) with rain gauge data in two steps. The first step is to reduce random error, and the second step is to reduce bias. Over land, the first step uses inverse (random) error variance weighting to combine the satellite products, which estimates the random error from large-scale averages of rain gauge data. The second step merges the combined satellite data with rain gauge data to reduce overall bias. Using the blending algorithm of Reynolds (1988), the combined product is assumed equal to rain gauge data for those grid areas having adequate coverage, whereas other values are calculated by solving a form of the Poisson equation. The result is that the satellite product determines the precipitation pattern, and the rain gauge data constrains the amplitude. From 1986, the rain gauge component is the GPCC monitoring product (section 2a).

d. *CPC retrospective U.S. and Mexico daily precipitation analysis*

The USMex is a gridded rain gauge product available at a $1° \times 1°$ resolution from 1948. The U.S. data are based on the CPC Unified Raingauge Data (URD; Higgins et al. 2000), which are composed of (i) the daily CPC Cooperative (COOP) dataset (~7000 rain gauges from 1992), (ii) the National Climatic Data Center (NCDC) COOP dataset (~8000 rain gauges from 1948), and (iii) the hourly precipitation dataset as described in Higgins et al. (1996; ~2500 rain gauges from 1948). Accounting for overlap between networks, the URD typically represents 13 000–15 000 daily rain gauge sites; however, prior to 1992 the number of rain gauges approaches 8000. The Mexican data are based on approximately 200 rain gauges prior to 1990 and 600 rain gauges thereafter. The data are quality controlled for duplicate and extreme values and bias corrected for spurious zero measurements against collocated ground radar data. The gridding of the data uses a modified Cressman scheme (Cressman 1959). Although there is likely some overlap between this USMex rain gauge data product and the GPCC monitoring product described above, the GPCC data contains only a fraction of the information contained in the USMex dataset (i.e., 7500 rain gauges globally for the GPCC data versus 13 000–15 000 in the United States for the USMex data). These datasets also differ in their interpolation methods, and interpolation alone has been shown to affect both correlation and bias with respect to actual point measurements (Chen et al. 2002).

e. *Precipitation-elevation Regressions on Independent Slopes Model*

PRISM is an analytical model that uses point source precipitation observations together with a digital elevation model (DEM) to provide gridded precipitation data products for the United States (Daly et al. 1994). The development of PRISM was motivated by the need for precipitation data over elevated terrain where rain gauge data are sparse. It attempts to resolve some important problems associated with other orographic interpolation methods, which rely heavily on neighboring observations to approximate data-void localities even though neighboring sites may exhibit important orographic differences. The PRISM methodology is i) to use a DEM to estimate the elevation of each available gauge station, and ii) to group stations according to their orographic commonality (or "facet") with respect to orographic features such as elevation, face orientation, slope, and proximity to coasts. Precipitation at a given DEM cell is estimated using a regression of precipitation versus elevation, where the stations sharing the same facet as the grid cell of interest are used to determine the parameters of the regression. The primary point source estimates used as input to PRISM are from the NCDC Historical Climate Network (~1200 stations), CPC COOP (~8000 stations), and Natural Resources Conservation Service snowpack telemetry (SNOTEL) data (~730 sites). PRISM data are available from 1890 as monthly or annual averages with a spatial resolution of 32×32 km^2.

f. *NCEP–DOE Reanalysis 2*

NCEP-2 (Kanamitsu et al. 2002) is a global data product created from a spectral data assimilation system that merges observations from many sources.

These include rawinsondes for upper air variables, satellites for vertical temperature soundings and cloud drift winds, aircraft for wind and temperature, ocean reports of surface variables, synoptic weather reports over land, and pentad CMAP data as a precipitation correction to improve soil moisture fields (Kalnay et al. 1996; Kanamitsu et al. 2002). Following extensive QC, the data are assimilated using a T62 (~210 km) global spectral model, which has 28 vertical levels and includes parameterizations of the major physical processes such as convection, large-scale precipitation, boundary layer physics, and so on. The assimilation process generates additional variables that are derived from observed fields. These derived variables (class C) include clouds, surface fluxes, and precipitation. Their reliability is heavily dependent on the performance of the model and its parameterizations (Kistler et al. 2001). The NCEP-2 data are available in 6-h increments from 1979.

g. North American Regional Reanalysis

NARR (Mesinger et al. 2006) is an atmospheric and hydrology dataset covering North America. NARR uses the regional Eta Model (forced at its lateral boundaries by NCEP-2) to assimilate observations in a similar manner used for global reanalysis products. In addition to a higher spatial (32×32 km^2 with 45 levels) and temporal (3 hourly) resolution as compared to global reanalysis, NARR uses additional and improved input data and contains a better representation of land surface hydrology and land–atmosphere interactions. A particular improvement is the assimilation of observed precipitation. NARR uses precipitation observations to correct atmospheric moisture and energy fields, which subsequently leads to improvements in the model-derived precipitation fields. NARR data are available from 1979.

h. Variable Infiltration Capacity Retrospective Land Surface Dataset

The VIC land surface dataset is based on 50-yr (1950–2000) simulations with the VIC land surface hydrology model. The simulations were designed to provide a high-resolution (3 hourly and $1/8^\circ \times 1/8^\circ$) dataset of land surface states and fluxes for the continental United States and for parts of Canada and Mexico (Maurer et al. 2002). The model includes a soil–vegetation–atmosphere transfer scheme and is forced by soil and land use data and observations of precipitation, air temperature, wind, humidity, and radiation. Model-derived variables include snow water and multiple-layer soil moisture tendencies, surface temperature, and latent and sensible heat fluxes. The VIC dataset also provides the model-forcing data including precipitation. The precipitation data are based on daily totals from the CPC COOP stations, with an approximate density of one gauge per 700 km^2 (Maurer et al. 2002), which is gridded to a $1/8^\circ \times 1/8^\circ$ resolution (Shepard 1984; Widmann and Bretherton 2000). Variation as a result of orography is incorporated into the data using an adjustment factor for each month and grid cell, which is calculated as the ratio of monthly mean precipitation from PRISM (section 2e) to that of the COOP data.

i. Global Meteorological Forcing Dataset for Land Surface Modeling

GMFD (Sheffield et al. 2006) is a high-resolution (3 hourly and $1^\circ \times 1^\circ$) dataset spanning 1948–2000, developed by Princeton University's Land Surface Hydrology Research Group to provide near-surface meteorology variables on space–time scales necessary for forcing land surface and hydrology models. The forcing data includes precipitation fields, which are developed using NCEP global reanalysis as the primary input and uses other data to correct known reanalysis errors and to disaggregate the corrected precipitation fields to the desired 3-hourly, 1° resolution. The precipitation observations are the Climatic Research Unit (CRU) monthly climate variables (which include rain day statistics), the GPCP daily precipitation product, and the Tropical Rainfall Measuring Mission (TRMM) 3-hourly precipitation. The corrections to the NCEP data are for (i) rain day frequency statistics (Sheffield et al. 2004), which involves resampling NCEP daily precipitation to match the rain day statistics of CRU, GPCP, and TRMM observations; and (ii) rain gauge undercatch, which uses monthly adjustment ratios from Adam and Lettenmaier (2003). Spatial downscaling uses probabilistic relationships between precipitation intensity observed on a daily reanalysis grid, with observed gridcell fractional coverage as determined from higher-resolution precipitation observations. Temporal downscaling uses probability density functions derived from TRMM.

3. Data rescaling

The precipitation datasets used in this study have spatial resolutions ranging from $1/8^\circ$ to 2.5°. To compare these data directly, the data were rescaled prior to the analysis to a common $2.5^\circ \times 2.5^\circ$ grid, extending from the Pacific coast to slightly east of the Rocky Mountains and from northern Mexico to slightly north of the U.S.–Canadian border (27.5°–50.0°N and 130.0°–102.5°W). For datasets having a horizontal resolution

smaller than 2.5°, simple box averaging was employed to rescale the data. For those datasets having an original horizontal grid size of 1° or smaller, upscaling was done iteratively to prevent the propagation of undefined values that occur over the Pacific Ocean for rain gauge data as well as to prevent the spatial propagation of small-scale, intense precipitation events to the coarser scale. Bilinear interpolation was then employed to adjust the 2.5° data such that all datasets have collocated gridcell centers. Because PCA has been shown to be sensitive to domain shape (e.g., Richman and Lamb 1985; Richman 1986), we opted to modify the PRISM data at its southern boundary to account for its lack of data over northern Mexico by filling five data-void grid cells centered at 28.75 °N with values of their nearest northern neighbor. The temporal resolutions of the nine datasets vary from 3 hourly to monthly. However, monthly averages of temporally high-resolution data are often available. Monthly products were used for all datasets except the USMex, for which we constructed monthly mean values from daily data. The period of study is January 1986–July 2000, which is the time of maximum overlap between datasets. Units of measure have been converted to mm month^{-1} for all datasets.

4. Methods

S-mode, rotated PCA operating on the correlation matrix is used to disaggregate the domain of the western United States into regions that are unique and spatially cohesive with respect to their precipitation climate. Prior to the analysis, monthly precipitation values are weighted by $(\cos\varphi)^{1/2}$, where φ is latitude, to correct for latitudinal differences in grid spacing (e.g., Wilks 2006). The correlation matrix is used as the dispersion matrix because there are large differences in precipitation variance across the domain and using the correlation matrix (rather than the covariance matrix) prevents the high-variance grid cells from disproportionately influencing the results (Jolliffe 2002; Wilks 2006). The correlation coefficient can be sensitive to data nonnormality, and monthly precipitation data are skewed. However, normality is not required for PCA to be valid (Wilks 2006). In preliminary analyses, we tested the sensitivity of the methodology to transformations of the input data (which brought the data closer to normal) and found no notable differences in the resulting principal component (PC) loading patterns.

The appropriate number of PCs to rotate is selected using the scree test (Cattell 1966), the log eigenvalue (LEV) diagram, and the North et al. (1982) eigenvalue separation test. The scree test and LEV diagram use a plot of the eigenvalue spectrum (eigenvalues in decreasing order) and the log eigenvalue spectrum, respectively, and the appropriate truncation point is observed as a discontinuity in slope. For the scree plot, this appears as a transition point between the steeply sloped part of the spectrum to the left and the gently sloped portion to the right (Wilks 2006), which marks the point where the cumulative variance explained with each added PC is approaching zero. For the LEV diagram, the log eigenvalues that form an approximately straight line indicate the exponential decay of PCs dominated by uncorrelated noise (e.g., Wilks 2006), so these and subsequent PCs would not be retained. The eigenvalue separation test is applied to avoid separating closely spaced PCs. Neighboring PCs whose eigenvalues are close in size are associated with large sampling errors whereby different samples may produce different linear combinations of the closely spaced PCs. As such, the patterns produced by closely spaced PCs ("effectively degenerate multiplets") are not independent but rather random mixtures of the true population PCs (North et al. 1982), and these should not be separated in a rotated PCA (North et al. 1982; Jolliffe 2002). Neighboring multiplets are identified as those having overlapping standard error bars, where standard error is estimated by $\delta = \lambda(2/N)^{1/2}$, and where λ is the eigenvalue and N is the sample size (North et al. 1982). Because monthly precipitation data are autocorrelated, the standard error approximation is modified in this study to use an effective sample size. Specifically, we use $N^* = N(1 - r^2)/(1 + r^2)$, which is appropriate for variances (Bretherton et al. 1999), where r is the domain-averaged autocorrelation coefficient. The autocorrelation ranges from 0.34 to 0.39 for the nine datasets.

The PCs are rotated using the direct oblimin oblique rotation method. Before rotating, each PC is weighted by the square root of their corresponding eigenvalue, which makes interpretation more intuitively meaningful because the weighted PC elements are then a measure of the correlation between the original data and the PC time series (e.g., Richman and Lamb 1987; Wilks 2006). The rotated PCA is used because rotation has been shown to improve the interpretability of the solution and can eliminate some problems caused by the orthogonality constraint associated with unrotated PCs such as Buell patterns, which are artificial and misleading patterns resulting from domain shape rather than any physical relationship between variables (Buell 1979) and larger sampling errors as discussed above (Richman 1986; Richman and Lamb 1987; Jolliffe 2002; Wilks 2006). Additionally, rotated PCs (RPCs) show more similarity to the original data than unrotated solutions (Richman 1986; Wilks 2006) and, for regional-

TABLE 2. Interpretation of congruence coefficient representing the degree of likeness between rotated loading vectors.

≥0.98	Excellent	
≥0.92	Good	<0.98
≥0.82	Borderline	<0.92
≥0.68	Poor	<0.82
	Terrible	<0.68

ization studies in particular, rotated PC solutions have been shown to be more stable to changes in domain size and shape (Richman and Lamb 1985; Richman 1986; White et al. 1991). The oblique rotation rather than an orthogonal rotation is chosen because oblique methods have been shown to be more stable with respect to changes in spatial and temporal domain and also superior in achieving simple structure (White et al. 1991).

The rotated PCs are used to regionalize the domain of the western United States according to the maximum loading method (Comrie and Glenn 1998). This approach assigns each grid cell to the RPC onto which it loads most strongly, and the grid cells assigned to the same RPC are then grouped together into distinct "regions," which results in the regionalization of the domain.

The degree of similarity between rotated PC solutions obtained from different datasets is assessed by quantitatively comparing their respective loading matrices using the congruence coefficient (Harman 1976):

$$g = \frac{\sum_{j=1}^{n} (b_{ja} b_{jb})}{\left[\left(\sum_{j=1}^{n} b_{ja}^2 \right) \left(\sum_{j=1}^{n} b_{jb}^2 \right) \right]^{1/2}}, \quad (1)$$

where $\mathbf{b}_a$ is a loading from rotated loading vector **A** from one solution, $\mathbf{b}_b$ is a loading from rotated loading vector **B** from another solution, and n is the number of variables in each eigenvector, which for S-mode PCA corresponds to spatial positions. A value of +1 for the congruence coefficient represents perfect agreement between solutions, and a value of −1 represents perfect inverse agreement and a value of 0 indicates no agreement. The congruence coefficient is preferred to the correlation coefficient for measuring pattern similarity because it preserves the mean (whereas the correlation coefficient measures deviations from the mean), and the mean is an important feature of a PC loading vector (Richman 1986). Following Richman (1986) and the references therein, the guidelines shown in Table 2 are adopted as an indicator of the degree of likeness between solutions, which are based on the Monte Carlo studies of the distribution of the congruence coefficient by Korth and Tucker (1975). From this table, the congruence coefficient is biased toward higher values as compared to the correlation coefficient and, according to Richman (1986), any coefficient smaller than 0.7 represents a match that is not any better than would be expected by "randomly spinning" the PC axes prior to matching a set of PC loading vectors.

Two additional comparative RPC analyses are performed following the methodology just described to test the sensitivity of the method to differences in spatial resolution and temporal domain. The USMex and GMFD data are used for these sensitivity tests because they have long records and a desirable original 1° × 1° spatial resolution. The 1° scale is small enough to provide a meaningful comparison with the 2.5° data, while not being so small as to become computationally difficult to process. The RPC solutions obtained using 1° versus 2.5° data are compared using USMex and GMFD data. Next, using 1° data from USMex and GMFD, the affect of temporal sample size is tested by comparing results obtained using data from January 1986 to July 2000 versus the longer time series of January 1950–December 2000. In each case, the loading patterns are compared using the congruence coefficient as the indicator of similarity.

5. Results

a. Principal component analysis and regionalization

Figure 1 gives the scree graphs, LEV diagrams, and eigenvalue separation tests for each dataset. Based on our interpretation of Figs. 1a–c, five PCs are retained and rotated for each dataset. The slope discontinuity observed on the scree and LEV graphs occurs at the fifth or sixth eigenvalue; however, the choice of retaining five versus six PCs is subjective, and the level of ambiguity varies by dataset. The eigenvalue separation test shown in Fig. 1c is a plot of $\lambda_k - \delta_k$ and $\lambda_{k+1} + \delta_{k+1}$, where λ_k is the kth eigenvalue, λ_{k+1} is its neighboring eigenvalue, and δ_k and δ_{k+1} are their corresponding standard error estimates. The intersection of these two lines is an indication that neighboring eigenvalues are not adequately separated. In Fig. 1c, effective degeneracy occurs at the sixth PC for nearly all datasets and, therefore, five PCs are retained. The first five unrotated PCs collectively account for between 76% and 85% (depending on dataset employed) of the variance contained in the original data (Table 3).

The loading patterns corresponding to the five RPCs for each dataset are shown in Fig. 2 and are named and ordered according to their geographic centers: Pacific Northwest, West Coast, Southwest, Northern Plains,

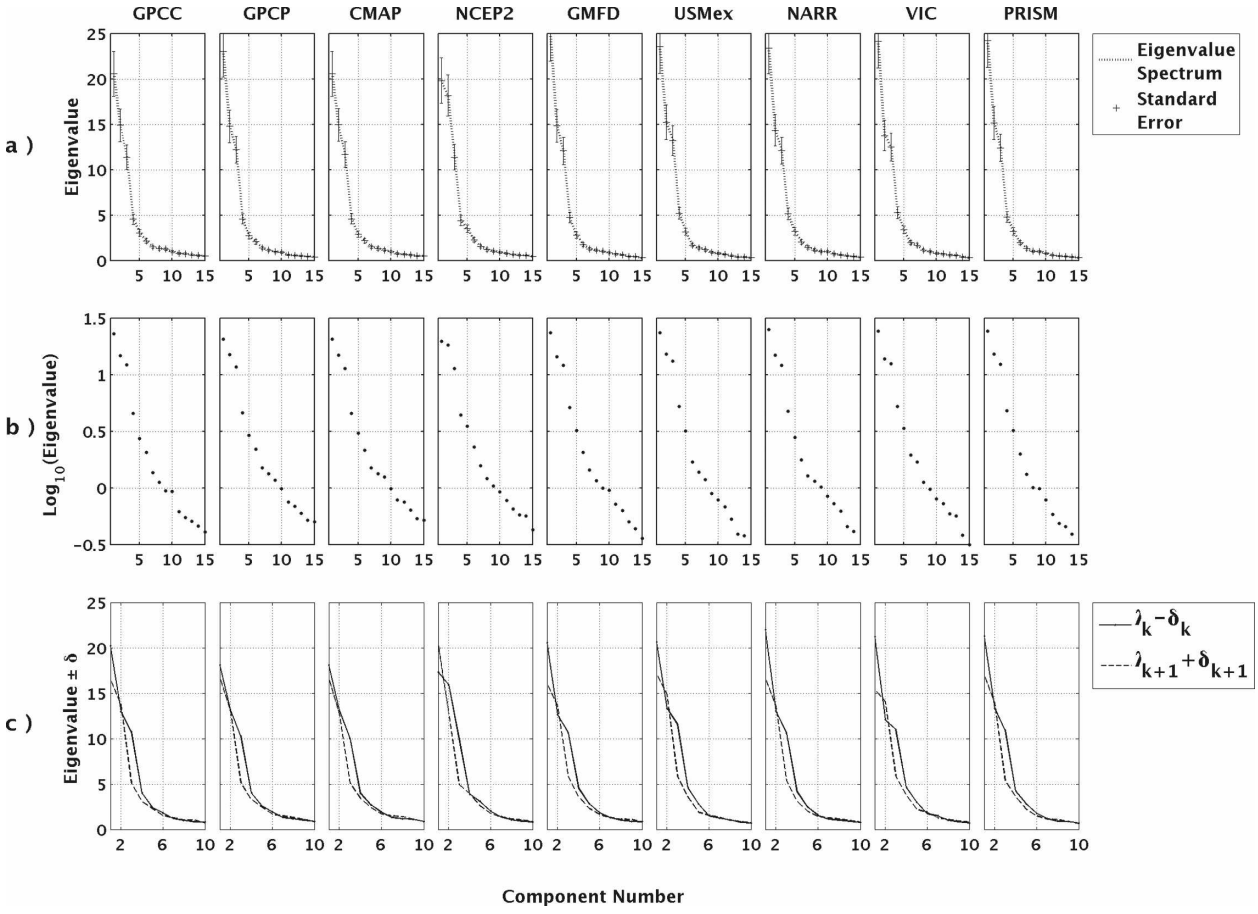

FIG. 1. (a) Scree plot, (b) LEV diagram, and (c) eigenvalue separation test for the nine precipitation datasets.

and Colorado Plateau. The rotated loading patterns given in Fig. 2 show the important influence of topography on the precipitation climate of the western United States. All datasets indicate a spatial pattern of variation centered over the Pacific Northwest, which is bounded to the east/southeast by the Rocky Mountains. There is also a California-centered pattern bounded to the east by the Rocky Mountains. The Southwest loading pattern is centered over northern Mexico and extends into Arizona and New Mexico, and is bounded to the north by the Colorado Plateau. The Northern Plains loading pattern is centered over the northern Great Plains at the state of Montana, and the Rocky Mountains bind the south/southwestward extent of the precipitation variability pattern. The fifth RPC for all datasets is centered over the high-elevation areas of the central Rocky Mountains near the Colorado Plateau.

The results of the principal component–based regionalization using the maximum loading method are shown in Fig. 3. Figure 4 gives the monthly climatology for each of the five regions shown in Fig. 3 according to the nine datasets. The Pacific Northwest region extends from the Pacific coast to the Rocky Mountains, and its southern boundary is near the northern limit of the Sacramento Valley and the southern limit of the Cascades. This region has a cold-season precipitation regime with a maximum in November–January (Fig. 4) that is associated with cyclonic storms emanating from the northern Pacific and dry summers. The West Coast

TABLE 3. Percent variance explained by the first five principal components.

	PC component				
Dataset	1	2	3	4	5
GPCC	28.95	20.99	16.03	6.4	4.30
GPCP	32.42	20.80	17.2	6.44	3.85
CMAP	28.94	21.05	16.45	6.45	4.10
NCEP2	27.92	25.60	16.02	6.20	4.93
GMFD	35.32	20.88	16.99	6.66	3.93
USMex	33.16	21.45	18.62	7.36	4.47
NARR	32.94	20.19	17.04	7.25	4.51
VIC	34.02	19.37	17.62	7.42	4.74
PRISM	34.08	21.35	17.48	6.75	4.52

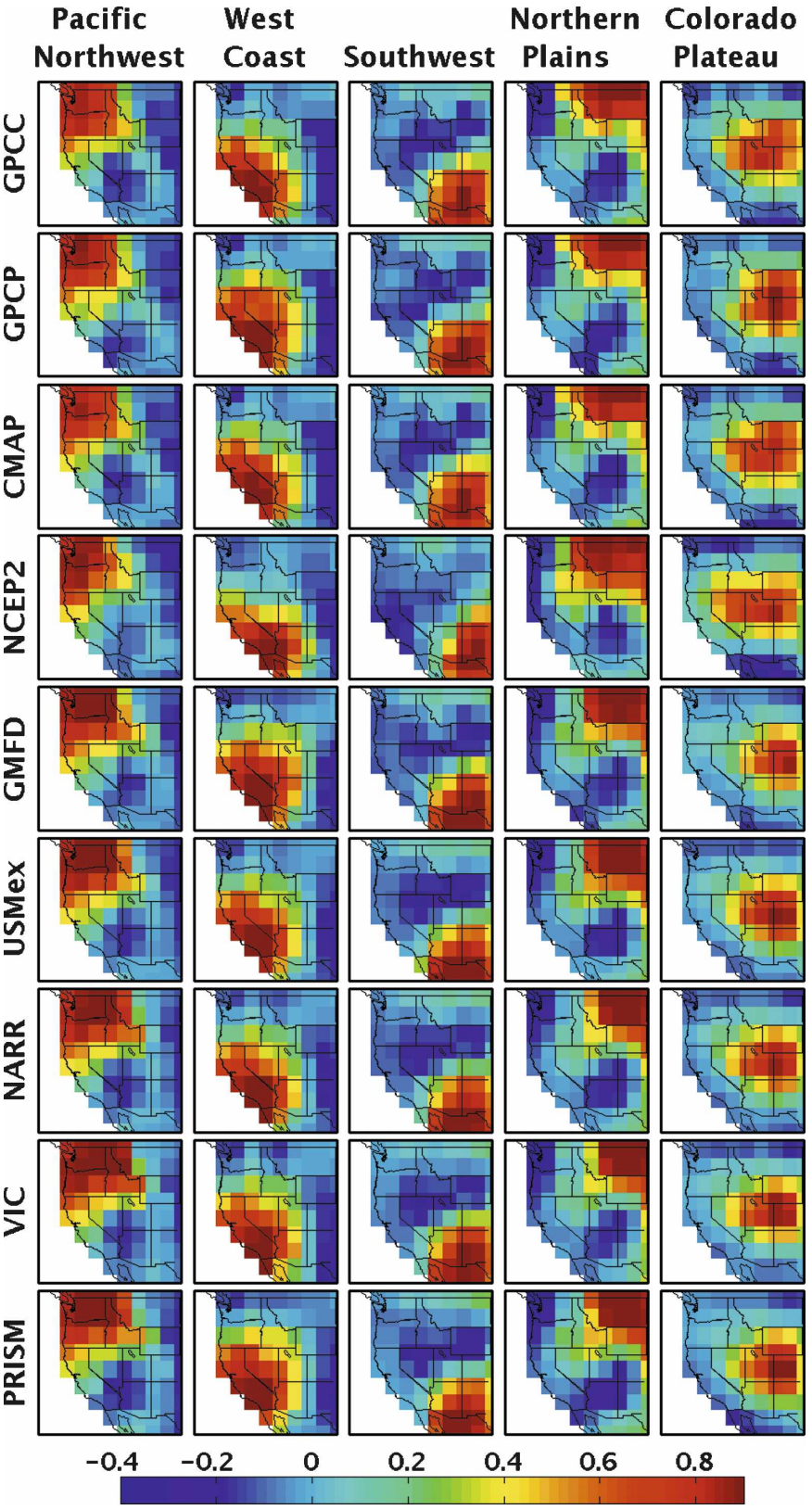

FIG. 2. Rotated loading patterns corresponding to the five retained PCs and for each dataset.

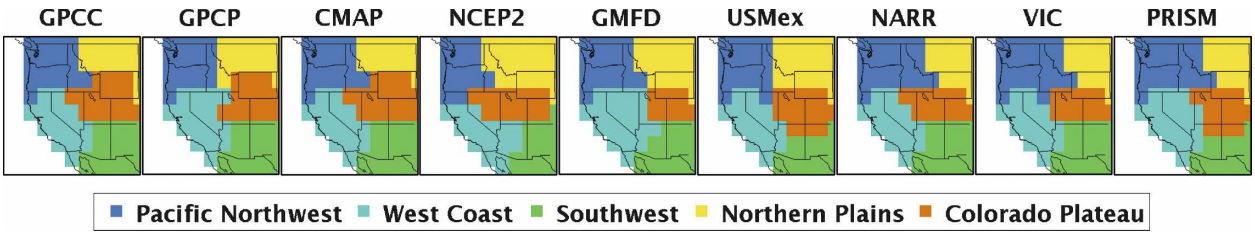

FIG. 3. The five precipitation regions obtained from each dataset using the maximum loading method.

region extends from the Oregon–California border southward to include the entire state of California and extends eastward to include all or part (depending on the dataset) of Nevada. The eastern boundary of the West Coast region is in the vicinity of where the Great Basin meets the Rocky Mountains. This region also has a winter precipitation regime similar to the Pacific Northwest but with its maximum occurring January–March. This seasonal shift in the cold-season maximum is attributed to a more southern concentration of cyclonic storm activity and the southward progression of the jet stream (Trewartha 1981; Mock 1996). The Southwest region is bound to the west at approximately the Arizona–California border in the vicinity of the Mojave Desert, which marks the transition between the Sonora Desert and the higher elevations of the Great Basin. Its northern boundary is near the Mogollon Rim, which marks the southern edge of the Colorado Plateau in Arizona. This Southwest region experiences a summer precipitation climate, with a peak from July–September marking the arrival and duration of the North American monsoon. The Northern Plains region is bound to the west by central or eastern Idaho (depending on the dataset) near where the Rocky Mountains intersect the northern Great Plains. This region includes most or all of the state of Montana; however, the horizontal extent region varies among datasets with respect to the proportion of Idaho and Wyoming contained. The Northern Plains region experiences an early summer (May–July) precipitation regime arising from the summer land–sea temperature gradient, which allows for advection of Gulf of Mexico moisture into the deep continental interior (Trewartha 1981). The Colorado Plateau region spans the high-elevation areas of Colorado and Utah (and Wyoming according to the CPCC, GPCP, and CMAP data). This region exhibits a bimodal precipitation regime that includes an early summer precipitation maximum similar to the precipitation climate of the Northern Plains and another maximum in late summer during the North American monsoon. For the Colorado Plateau, the positioning of the region itself (Fig. 3) explains some of the differences observed among the datasets with respect to the monthly climatology. For example, the summer peak centered in August is not observed according to the GPCP, CMAP, or GPCC data. This is probably because the Colorado Plateau region extends farther north according to these three datasets, thus the region is out of the range of primary North American monsoon influence.

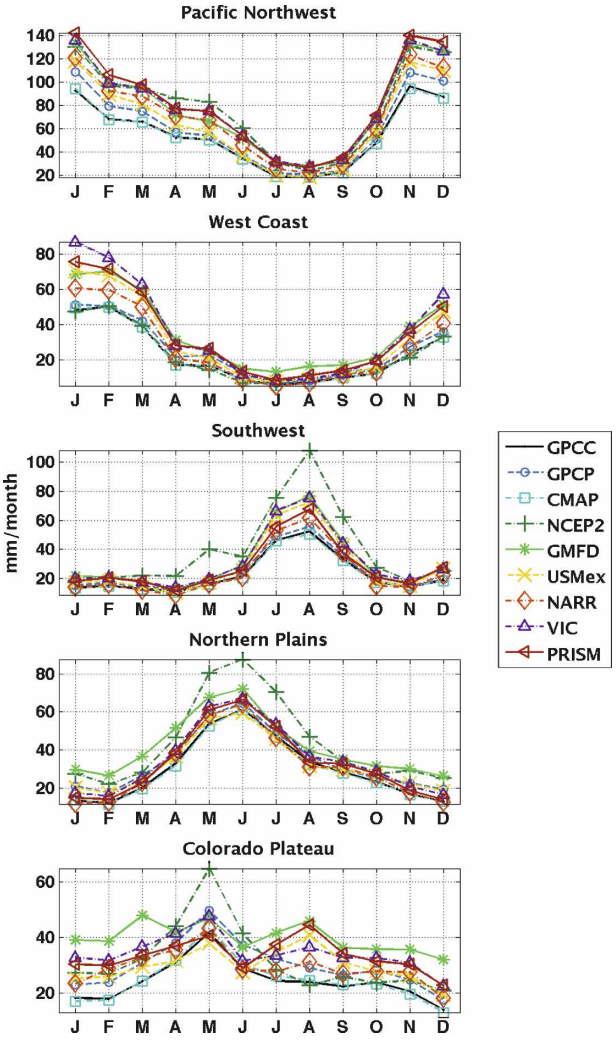

FIG. 4. Monthly climatology (January–December) for the five regions.

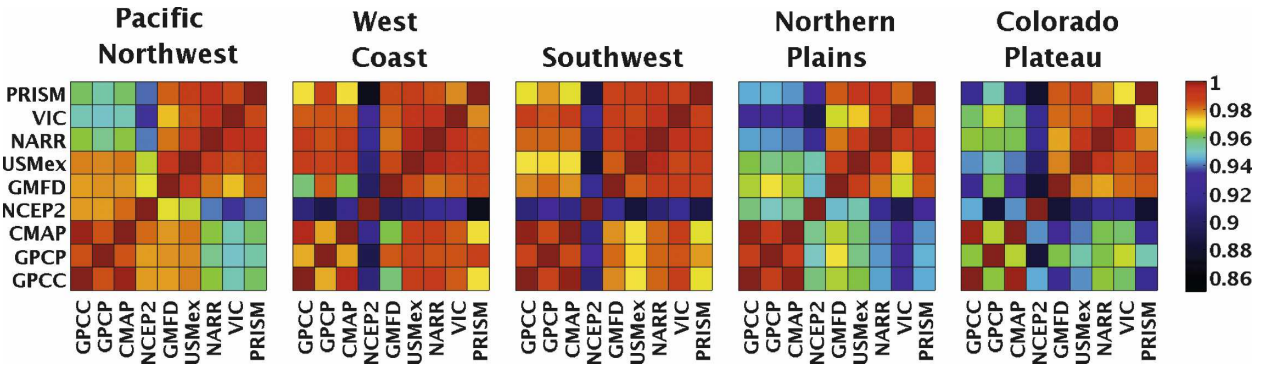

FIG. 5. Congruence coefficient for each data pair.

b. Dataset intercomparison

Figure 5 gives the congruence coefficients for all data pairs and for each of the five rotated loading patterns shown in Fig. 2. For the Pacific Northwest loading pattern, there is a high level of similarity among data with congruence coefficients above 0.93 (the good-to-excellent range) for all data pairs. For the West Coast and Southwest, the similarity among data is very high in general, with congruence coefficients at or above 0.97 for most data pairs. The exception is NCEP-2, which generally exhibits only "borderline" agreement with the other data (g = 0.86–0.93). For the Northern Plains loading pattern, congruence is generally in the good-to-excellent range ($g > 0.92$) except for the comparisons between NCEP-2 with NARR and VIC, which show only borderline agreement (g = 0.89–0.92). For the Colorado Plateau, there is good-to-excellent congruence ($g > 0.92$) among datasets with the exception of NCEP-2, which exhibits borderline congruence (g = 0.88–0.92) with all but two datasets (GPCC and CMAP, g = 0.94). The clustering of high-congruence coefficients in the southwest and northeast quadrants of each panel in Fig. 5 highlights two data groups that exhibit very similar loading patterns. These are the GPCP, CMAP, and GPCC datasets (Group A) and the NARR, USMex, GMFD, VIC, and PRISM datasets (Group B). These datasets show a higher level of congruence within groups (g = 0.97–1.0, g = 0.98–1.0, g = 0.98–1.0, g = 0.97–1.0, g = 0.96–1.0 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau, respectively) and lesser congruence between groups (g = 0.95–0.98, g = 0.96–0.99, g = 0.97–0.99, g = 0.93–0.97, g = 0.92–0.97 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau, respectively).

Figure 6 shows the difference between each loading pattern and an ensemble mean, where the ensemble mean was created from the average of the nine datasets for each RPC as

$$E(i, r) = \frac{1}{N}\sum_{n=1}^{N} x_n(i, r), \tag{2}$$

where $E(i, r)$ is the ensemble mean for the rth RPC at position i, and $x_n(i, r)$ is the nth ensemble member for the rth RPC at position i, and N is the number of ensemble members. To relate the differences observed in Fig. 6 with the original precipitation data, we consider the time-varying amplitudes of the five PCs. Specifically, we isolate the months in the January 1986–July 2000 record in which the PC amplitudes are strongest, which we take as those above the 80th percentile. The average precipitation for those months corresponding to the largest amplitudes is shown in Fig. 7 for each dataset.

In Fig. 6 for the Pacific Northwest, loading pattern differences between datasets occur primarily along the Rocky Mountains from Idaho to Colorado. This is probably related to the differences in original spatial resolution among datasets where the higher resolution data (NARR, USMex, GMFD, VIC, and PRISM) are better able to represent the orographic enhancement of precipitation that occurs along the western Rocky Mountains than the coarser data. Some of this high-resolution information is carried over to the coarser scale when the data are regridded to the 2.5° spatial resolution (Guirguis and Avissar 2008). This enhanced precipitation along the Rockies for the Group B datasets is also seen in the raw precipitation data shown in Fig. 7. In Fig. 6 for the West Coast, the NCEP2 data are shown to differ from the ensemble over most of the domain, which is a result of a shift to the southeast of the West Coast loading pattern for the NCEP-2 data as compared to the others (Fig. 2). Considering Fig. 7, this shift appears to be related to the overestimation of precipitation in the vicinity of Arizona, which causes grid cells in the area to load more strongly onto the West Coast RPC than the other datasets. Also for the West Coast loading pattern and considering the other

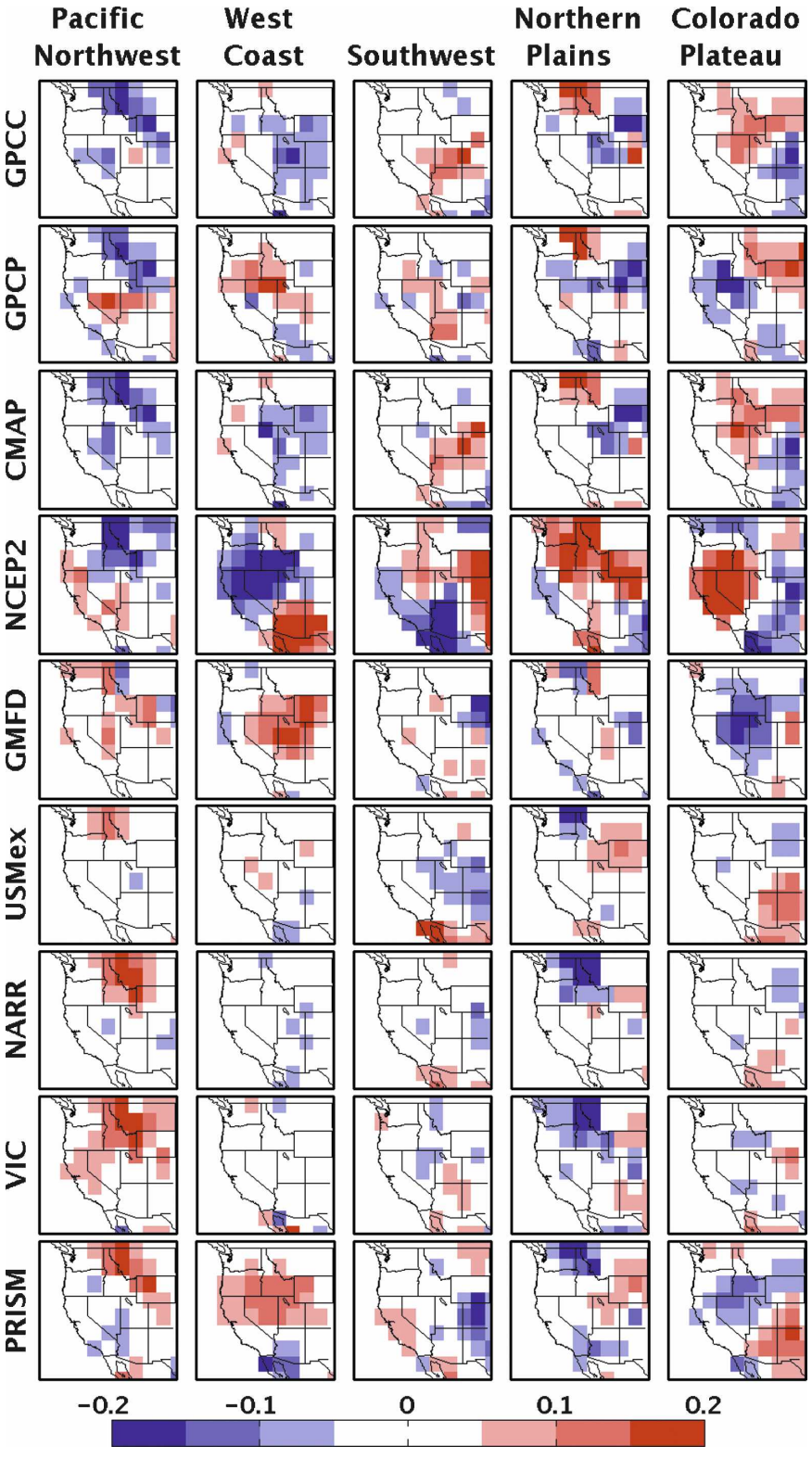

FIG. 6. Difference from ensemble of rotated loading patterns, where the ensemble was created from the average of the nine datasets for each RPC.

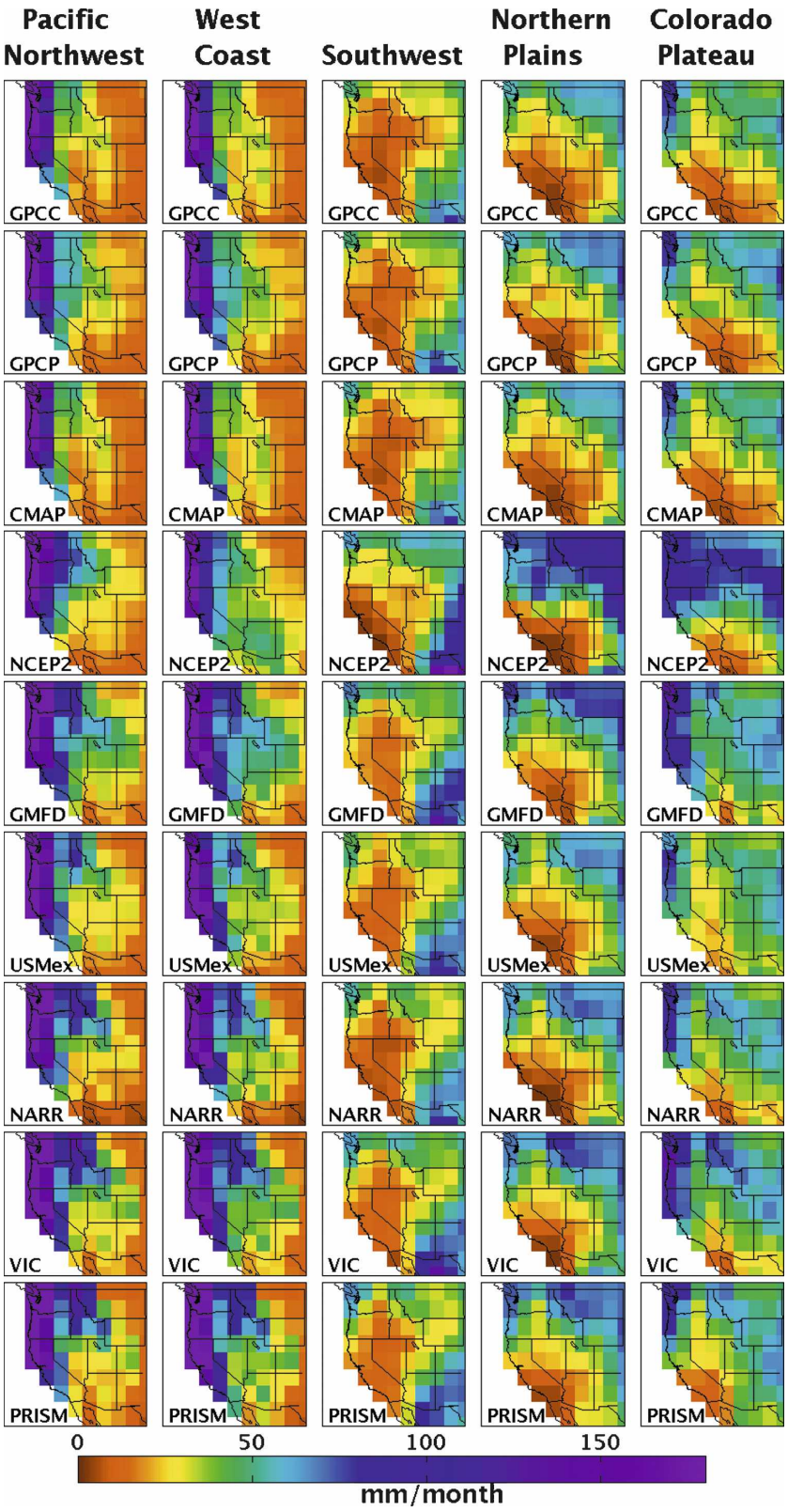

FIG. 7. Mean precipitation for those months in which the PC amplitudes exceed the 80th percentile.

datasets, differences are observed in the intermountain region in Fig. 6, which in Fig. 7 appears to be related to the degree of severity of the eastward wet-to-dry precipitation gradient. For example, the GPCC and CMAP show a much stronger precipitation gradient with drier inland conditions, and GMFD shows a weaker gradient with wetter intermountain conditions. For the Southwest, the loading pattern differences appear to result from differences among data in their representation of the North American monsoon. Specifically, in Figs. 6 and 7, the data differ with respect to the strength of the monsoon signal as well as the spatial extent of monsoon influence. In Fig. 6 for the northern plains, the data differences are observed primarily over the Rocky Mountains, which in Fig. 7, are related to the tendency of the Group A datasets to represent drier conditions over Idaho, eastern Washington, Wyoming, Utah, and Colorado as compared to the Group B datasets. Also in Fig. 6 for the Northern Plains, the NCEP-2 data demonstrates much stronger loadings onto the northern plains RPC as compared to the other datasets, which (in Fig. 6) appears to be a result of the comparatively wet conditions observed over the northern half of the domain for the NCEP-2 data. In Fig. 6 for the Colorado Plateau, differences between datasets are observed over much of the domain. These differences are primarily a result of (i) the drier conditions observed over the Southwest region for the GPCC, GPCP, and CMAP data as compared to the other datasets, and (ii) the wetter conditions observed for the intermountain region and along the Rocky Mountains according to PRISM, VIC, NARR, USMex, and GMFD, which is likely related to the original spatial resolution of these datasets. Also for the Colorado Plateau loading pattern, the NCEP-2 data differs greatly from the other datasets, which in Fig. 2 is observed as an elongation to the west of the NCEP-2 Colorado Plateau RPC relative to those of the other datasets. In Fig. 7, this appears to be caused by overestimated precipitation along the northern part of the domain according to NCEP-2.

c. *Space–time sensitivity studies*

Table 4 gives the results of two comparative RPC analyses using data from USMex and GMFD, which were designed to test the sensitivity of the methodology to differences in spatial resolution and temporal domain. The spatial resolution sensitivity study uses Eq. (1), where $\mathbf{b}_a$ is the RPC from the dataset at its original 1° resolution, and $\mathbf{b}_b$ is the loading pattern from the same dataset rescaled to 2.5°. Because the calculation represents a point-to-point comparison, linear interpolation is applied to the coarse-resolution loading vector during the scale sensitivity analysis. The analysis of sensitivity to temporal sample size uses Eq. (1), where $\mathbf{b}_a$ is the loading vector obtained from 1° data over a long record (January 1950–December 200), and $\mathbf{b}_b$ is the loading vector obtained from 1° data from same dataset over a shorter record (January 1986–July 2000).

TABLE 4. Congruence coefficients for (a) precipitation data at 1° vs 2.5° and (b) precipitation data for January 1950–December 2000 vs January 1986–July 2000.

	(a)		(b)	
Rotated PC	USMex	GMFD	USMex	GMFD
Pacific Northwest	0.97	0.97	0.99	1.0
West Coast	0.97	0.97	0.99	1.0
Southwest	0.97	0.96	0.99	1.0
Northern Plains	0.96	0.97	0.99	1.0
Colorado Plateau	0.97	0.96	0.99	1.0

In Table 4, the methodology is shown to be stable to changes in spatial resolution and temporal domain. The rotated loading patterns obtained using a shorter versus longer time domain are shown to be very similar with congruence coefficients exceeding 0.98, representing excellent agreement, for all five RPCs and for both datasets. For the spatial resolution sensitivity analyses, the loading patterns obtained using 1° versus 2.5° data are shown to exhibit good agreement with congruence coefficients in the range of 0.95–0.97. Figures 8a and 8b give the results of the regionalization for the 1° USMex and GMFD data over the shorter and longer records, respectively. A visual comparison of Fig. 8a with Fig. 3 shows the regionalization results to be very similar for the USMex and GMFD data, whether 1° or 2.5° data are used. The regionalization results for the short-term versus the long-term data are shown to be very similar (cf. Figures 8a versus 8b). The minor differences observed are generally a result of the migration of one or two grid cells into a neighboring region. Because grid cells bordering two neighboring regions are often modestly loading onto both corresponding PCs, it is not unexpected to observe this shift from one experiment to the next.

6. Summary and conclusions

This study has identified the precipitation climate regions of the western United States while testing the utility of the methodology, which has been traditionally applied to rain gauge data, to alternate data products. Specifically, we employed rotated PCA with observational precipitation data to identity the dominant patterns of covariability in the western United States and to isolate spatially cohesive regions that experience

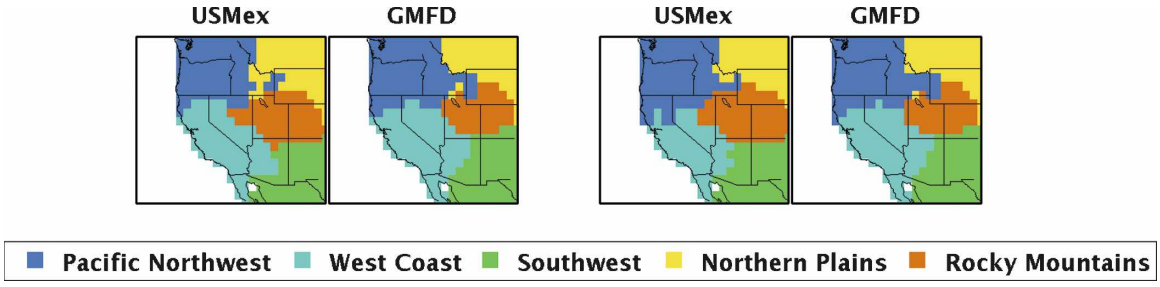

FIG. 8. The five precipitation regions obtained from 1° USMex and GMFD data over (a) the January 1986–July 2000 time domain and (b) the longer period of January 1950–December 2000.

similar temporal variability. This analysis was conducted on nine precipitation datasets including two reanalysis datasets, four rain gauge datasets (two of which use orographic adjustment), two satellite–rain gauge combined datasets, and one multiple-sourced product.

The dominant covariability patterns corresponding to the first five RPCs are centered over Washington, central California, northern Mexico at the border with Arizona and New Mexico, Montana, and Wyoming/Colorado. Using the congruence coefficient as a measure of loading pattern similarity, we compared the patterns obtained from each of the nine precipitation datasets and found the methodology to be reasonably stable to alternate data sources. That is, the congruence coefficient indicated good-to-excellent similarity between most datasets, with the exception of NCEP-2, which frequently demonstrated only borderline agreement with the other datasets. Also from this analysis, loading pattern differences were shown to be related to differences among datasets primarily in their representation of (i) the precipitation over the Rocky Mountains, (ii) the eastward wet-to-dry precipitation gradient that occurs during the cold season, (iii) the magnitude and spatial extent of the North American monsoon signal, and (iv) the precipitation in the desert southwest during spring and summer. The overestimation of precipitation by NCEP-2 was shown to be responsible for the general lack of similarity between NCEP-2 and the other datasets.

To regionalize the domain of the western United States into subregions representing unique precipitation climates, we applied the maximum loading principle, which groups those grid cells exhibiting a similar temporal variability. This process produced five distinct regions whose centers and borders were found to be physically reasonable and which highlight the relationship between the precipitation climatology and the local topography.

Sensitivity experiments were conducted in which we compared the RPC solutions obtained from data having different spatial resolutions and temporal domains but alike otherwise. These results showed the methodology to be stable to spatial/temporal data differences. However, it is worth mentioning that the spatial/temporal characteristics of precipitation data can affect the outcome of the regionalization when the eigenvalue separation test is used for deciding the number of PCs to rotate because standard error calculations are dependent on the sample size (measured in the time domain) and autocorrelation structure (which can be affected by spatial resolution) of the underlying data.

The results of this study have applications for the installation and planning of ground-based instrumentation and/or for climate research. Furthermore, we see interesting uses for numerical modeling experiments. The boundaries obtained from PCA-based regionalization can be used for selecting the extent of a model domain or for the positioning of focused high-resolution grids within a coarser-scale domain. The methodology can also be used to make decisions about grid resolution, where the goal is to ensure that there are an adequate number of grid cells available to represent each distinct region. Additionally, for model evaluation purposes, this methodology provides an analytical way to disaggregate the domain into smaller subdomains to focus on model performance for specific climate or weather features and/or to study local land–atmosphere feedback. Previous precipitation regionalization studies have relied on rain gauge data. For numerical modeling studies, however, satellite data are potentially more useful because climate modeling experiments often include an oceanic region. Additionally, for a domain composed of complex topography as the western United States, it is not clear a priori that rain gauge data are the best choice for a regionalization study. The findings that the methodology shows stability to the choice of precipitation dataset employed therefore extends the utility of the method to include domains where there is uncertainty in observations.

A detailed intercomparison of the nine precipitation

datasets is described in a companion paper (Guirguis and Avissar 2008). In that paper, we provide a multiple dataset analysis of precipitation variability and persistence as well as season- and location-specific assessments of observational precipitation data uncertainty for the western United States and for its five subregions.

Acknowledgments. This research was funded by the National Oceanic and Atmospheric Administration (NOAA) under Grants NA04OAR4310078 and NA050AR4310014. This work was additionally supported by NASA Headquarters under the Earth System Science Fellowship Grant NNG04GQ60H. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or NASA. The authors wish to thank all the data providers for the use of their datasets. The GPCC dataset was provided by the Global Precipitation Climatology Center, Deutscher Wetterdienst (DWD), Germany (available online at http://gpcc.dwd.de). The GPCP, CMAP, and NCEP2 data were provided by the NOAA Climate Diagnostics Center (available online at http://www.cdc.noaa.gov/cdc/data.gpcp.html, http://www.cdc.noaa.gov/cdc/data.cmap.html, and http://www.cdc.noaa.gov/cdc/data.reanalysis2.html, respectively). The USMex dataset was provided by the NOAA Climate Prediction Center (available online at ftp.cpc.ncep.noaa.gov/precip/wd52ws/us-mex). NARR data were obtained from the NOAA National Climate Data Center (available online at http://nomads.ncdc.noaa.gov). The GMFD dataset was provided by the Land Surface Hydrology Research Group at Princeton University (available online at http://hydrology.princeton.edu/data.php). PRISM data were provided by Oregon State University (available online at http://www.prism.oregonstate.edu). The VIC dataset was supplied by the Land Surface Hydrology Research Group at the University of Washington (available online at ftp://ftp.hydro.washington.edu/pub/HYDRO/data/VIC_retrospective/monthly).

REFERENCES

Adam, J. C., and D. P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.,* **108,** 4257, doi:10.1029/2002JD002499.

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

——, C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP–3). *Bull. Amer. Meteor. Soc.,* **82,** 1377–1396.

——, and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Arkin, P. A., and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–84. *Mon. Wea. Rev.,* **115,** 51–74.

——, and P. Xie, 1994: The Global Precipitation Climatology Project: First algorithm intercomparison project. *Bull. Amer. Meteor. Soc.,* **75,** 401–419.

Beaudoin, P., and J. Rousselle, 1982: A study of space variations of precipitation by factor analysis. *J. Hydrol.,* **59,** 123–138.

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladé, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate,* **12,** 1990–2009.

Buell, C. E., 1979: On the physical interpretation of empirical orthogonal functions. Preprints, *Sixth Conf. on Probability and Statistics in the Atmospheric Sciences,* Baniff, AB, Canada, Amer. Meteor. Soc., 112–117.

Carter, M. M., and J. B. Elsner, 1997: A statistical method for forecasting rainfall over Puerto Rico. *Wea. Forecasting,* **12,** 515–525.

Cattell, R. B., 1966: The screen test for the number of factors. *Multivariate Behav. Res.,* **1,** 245–276.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. *J. Hydrometeor.,* **3,** 249–266.

Comrie, A. C., and E. C. Glenn, 1998: Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Climate Res.,* **10,** 201–215.

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

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.,* **33,** 140–158.

Diem, J. E., 2006: Synoptic-scale controls of summer precipitation in the southeastern United States. *J. Climate,* **19,** 613–621.

——, and D. P. Brown, 2006: Tropospheric moisture and monsoonal rainfall over the southwestern United States. *J. Geophys. Res.,* **111,** D16112, doi:10.1029/2005JD006836.

Dinpashoh, Y., A. Fakheri-Fard, M. Moghaddam, S. Jahanbakhsh, and M. Mirnia, 2004: Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *J. Hydrol.,* **297,** 109–123.

Easterling, D. R., 1990: Persistent patterns of thunderstorm activity in the central United States. *J. Climate,* **3,** 1380–1389.

Ehrendorfer, M., 1987: A regionalization of Austria's precipitation climate using principal component analysis. *Int. J. Climatol.,* **7,** 71–89.

Ellis, A. W., and T. W. Hawkins, 2001: An apparent atmospheric teleconnection between snow cover and the North American monsoon. *Geophys. Res. Lett.,* **28,** 2653–2656.

Fuchs, T., U. Schneider, and B. Rudolf, 2007: Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre, Deutscher Wetterdienst, 12 pp.

Gochis, D. J., L. Brito-Castillo, and W. J. Shuttleworth, 2006: Hydroclimatology of the North American Monsoon region in northwest Mexico. *J. Hydrol.,* **316,** 53–70.

Gruber, A., X. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two merged rain gauge–satellite precipitation datasets. *Bull. Amer. Meteor. Soc.,* **81,** 2631–2644.

Guirguis, K. J., and R. Avissar, 2008: An analysis of precipitation

variability, persistence, and observational data uncertainty in the western United States. *J. Hydrometeor.,* **9,** 843–865.

Gutzler, D. S., 2004: An index of interannual precipitation variability in the core of the North American monsoon region. *J. Climate,* **17,** 4473–4480.

Harman, H. H., 1976: *Modern Factor Analysis*. 3rd ed. The University of Chicago Press, 487 pp.

Hawkins, T. W., A. W. Ellis, J. A. Skindlov, and D. Reigle, 2002: Intra-annual analysis of the North American snow cover–monsoon teleconnection: Seasonal forecasting utility. *J. Climate,* **15,** 1743–1753.

Higgins, R. W., J. E. Janowiak, and Y.-P. Yao, 1996: A gridded hourly precipitation data base for the United States (1963–1993). *NCEP/Climate Prediction Center Atlas 1,* National Weather Service, NOAA, U.S. Department of Commerce, 47 pp.

——, W. Shi, E. Yarosh, and R. Joyce, 2000: Improved United States precipitation quality control system and analysis. *NCEP/Climate Prediction Center Atlas 7,* National Weather Service, NOAA, U.S. Department of Commerce, 40 pp.

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

Jolliffe, I. T., 2002. *Principal Component Analysis.* Springer-Verlag, 487 pp.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50–Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Klazura, G. E., J. M. Thomale, D. S. Kelly, and P. Jendrowski, 1999: A comparison of NEXRAD WSR-88D radar estimates of rain accumulation with gauge measurements for high- and low-reflectivity horizontal gradient precipitation events. *J. Atmos. Oceanic Technol.,* **16,** 1842–1850.

Korth, B., and L. R. Tucker, 1975: The distribution of chance congruence coefficients from simulated data. *Psychometrika,* **40,** 361–372.

Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.,* **10,** 111–127.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard, 2002: Weather radar coverage over the contiguous United States. *Wea. Forecasting,* **17,** 927–934.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate,* **15,** 3237–3251.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360.

Mock, C. J., 1996: Climatic controls and spatial variations of precipitation in the western United States. *J. Climate,* **9,** 1111–1125.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* **1,** 75–87.

Richman, M. B., 1986: Rotation of principal components. *Int. J. Climatol.,* **6,** 293–335.

——, and P. J. Lamb, 1985: Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: Some methodological considerations and a regionalization. *J. Climate Appl. Meteor.,* **24,** 1325–1343.

——, and ——, 1987: Pattern analysis of growing season precipitation in southern Canada. *Atmos.–Ocean,* **25,** 137–158.

Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. *Proc. Second Workshop of the International Precipitation Working Group IPWG,* Monterey, CA, EUMETSAT, 231–247.

——, H. Hauschild, W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. *Global Precipitations and Climate Change,* M. Desbois and F. Desalmond, Eds., NATO ASI Series 1, Vol. 26, Springer-Verlag, 173–186.

Serafin, R. J., and J. W. Wilson, 2000: Operational weather radar in the United States: Progress and opportunity. *Bull. Amer. Meteor. Soc.,* **81,** 501–518.

Sheffield, J., A. D. Ziegler, E. F. Wood, and Y. Chen, 2004: Correction of the high-latitude rain day anomaly in the NCEP–NCAR reanalysis for land surface hydrological modeling. *J. Climate,* **17,** 3814–3828.

——, G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate,* **19,** 3088–3111.

Shepard, D. S., 1984: Computer mapping: The SYMAP interpolation algorithm. *Spatial Statistics and Models,* G. L. Gaile and C. J. Willmott, Eds., D. Reidel, 133–145.

Trewartha, G. T., 1981: *The Earth's Problem Climates.* 2nd ed. University of Wisconsin Press, 371 pp.

White, D. R., M. Richman, and B. Yarnal, 1991: Climate regionalization and rotation of principal components. *Int. J. Climatol.,* **11,** 1–25.

Widmann, M., and C. S. Bretherton, 2000: Validation of mesoscale precipitation in the NCEP reanalysis using a new grid-cell dataset for the northwestern United States. *J. Climate,* **13,** 1936–1950.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 648 pp.

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

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

——, and ——, 1998: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Climate,* **11,** 137–164.

Yin, X., A. Gruber, and P. Arkin, 2004: Comparison of the GPCP and CMAP merged gauge–satellite monthly precipitation products for the period 1979–2001. *J. Hydrometeor.,* **5,** 1207–1222.