

New Rainfall Datasets for the Congo Basin and Surrounding Regions

S. E. NICHOLSON AND D. KLOTTER

Department of Meteorology, Florida State University, Tallahassee, Florida

A. K. DEZFULI

*Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, and
Science Systems and Applications, Inc., Lanham, Maryland*

L. ZHOU

University at Albany, State University of New York, Albany, New York

(Manuscript received 25 January 2018, in final form 30 April 2018)

ABSTRACT

This paper describes three new rainfall datasets that have been developed for equatorial Africa. The development relies on acquisition of recent gauge data from the relevant countries and statistical methods to fill in gaps in coverage. Two of the three datasets are gridded with spatial resolutions of 2.5° and 5.0°, and the third is regionally aggregated and based purely on gauge data. The work is based on a total of 1826 gauge records in the analysis sector, of which only several hundred operate in recent years. The gridded datasets were produced and validated by using a period of dense coverage (1947–72) to “calibrate” a spatial reconstruction method, which is then utilized to grid data for the remaining years. The period 1973–2010 served as a validation period. The validation was carried out by comparing the gridded values with values obtained by simple averaging of station data in grid boxes with an adequate number of stations. This exercise clearly showed that the statistical reconstruction approach based on principal components produced far superior results than those from the more commonly used kriging. The gridded datasets cover each month of the year, six seasons, and annual rainfall, and they commence in 1921 and extend through 2014. In contrast, the gauge-only regional dataset covers varied time periods, depending on the geographical region in question. Records for several regions cover nearly all of the twentieth century and most extend to 2014.

1. Introduction

The Congo basin (Fig. 1) is the location of the world’s second largest area of rain forest, as well as the world’s most intense thunderstorms. The region is currently of concern because of evidence of a significant multidecadal drying trend, one of the most significant worldwide (Malhi and Wright 2004; Asefi-Najafabady and Saatchi 2013; Zhou et al. 2014; Harris et al. 2017). Zhou et al. (2014) utilized several satellite indicators of vegetation greenness and water content during the April–June (AMJ) season to examine trends in forest vigor and their relationship to climatic variables. They demonstrated a decline in vegetation greenness and terrestrial water storage (TWS) for the period 2000–12 and in water content of aboveground plant

matter for the period 1988–2010 over most of the still-intact forest region. That study attributed this forest greenness and moisture decline to the long-term drying trend.

An understanding of this decline requires examination of both anthropogenic and meteorological factors. Unfortunately, equatorial Africa has long represented a major gap in our understanding of the tropical climate system (Todd and Washington 2004; Washington et al. 2013). While research of the last decade has improved the situation, the meteorology of the Congo basin is somewhat of an enigma. Part of the problem is the scarcity of rainfall data, especially after the early 1990s (Alsdorf et al. 2016). The meteorological services of both Angola and the Democratic Republic of the Congo, countries that include most of the Congo rain forest, essentially ceased to function for decades and have been slow to rebuild their services.

Corresponding author: S. E. Nicholson, snicholson@fsu.edu

DOI: 10.1175/JHM-D-18-0015.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

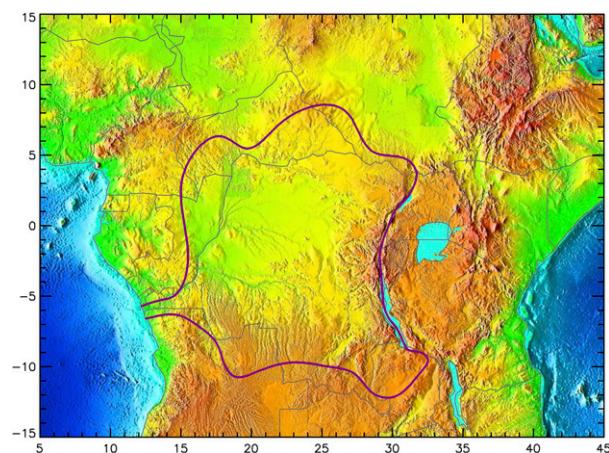


FIG. 1. Schematic of topography in equatorial Africa, with the red line showing the region of the Congo basin.

Because of the scarcity and difficulty in accessing gauge data, most studies of the region depend on satellite precipitation estimates. For example, Zhou et al. (2014) used TRMM (Huffman et al. 2007) and GPCP (Adler et al. 2003), as well as the GPCC gauge dataset (Schneider et al. 2015). However, satellites estimates generally do not perform well over the Congo basin (McCollum et al. 2000; Awange et al. 2016). Moreover, GPCC, which is used often used as the benchmark in validation, has very little station coverage over the Congo basin for the last two decades.

The current study seeks to expand and improve coverage in recent years by acquiring new gauge data for the various countries of equatorial Africa and by utilizing statistical methods for filling in gaps in coverage. The result is three datasets: two are gridded at spatial resolutions of 2.5° and 5.0° and the third is regionally aggregated. The regional dataset has a coarser resolution, but is based on a larger number of stations than the grid boxes at either resolution. Hence, it can help to verify the accuracy of the gridded datasets.

Trend analyses of rainfall are also carried out and compared with the findings of Zhou et al. (2014) for the period 1985–2012. Calculations are performed on both the regional and gridded datasets. In most cases, only the AMJ time period is considered, as that is the season of significant long-term drought and of greatest greenness and moisture decline.

2. Data

a. Station network

The first author created a monthly African precipitation dataset that included some 1400 stations and extended from 1901 to 1998, although few stations covered the entire

99-yr span. The dataset, hereafter termed NIC131, is described in numerous publications (e.g., Nicholson 1986; Nicholson and Kim 1997; Nicholson et al. 2000). It was assembled from a large number of sources, including several writing campaigns to African meteorological services and personal visits to these services in some countries.

Updates were recently carried out for the continent as a whole (Nicholson et al. 2018), so that the current version of NIC131 now includes some 3000 stations. In the region of interest in the current study, there are now 1826 stations (Fig. 2). The number of stations in countries in and/or bordering on the Congo basin are as follows: Gabon, 28; Cameroon, 41; Central African Republic (CAR), 54; South Sudan, 20; Rwanda, 61; Burundi, 19; Tanzania, 123; Republic of the Congo, 36; Democratic Republic of the Congo (DRC; formerly Zaire), 148; Chad, 109; Uganda, 70; Nigeria, 42; Zambia, 92; and Angola, 99.

Figure 3 shows the number of stations available in each year in the original and in the current archive for the Congo basin study region in Fig. 2. The largest number of stations is available in the period of roughly 1940–2000, but a few stations operated even prior to 1900 (Nicholson et al. 2012a). A decline in station numbers is clearly apparent since the 1960s, when roughly 1200 stations are available in each year. By 2014 the number has decreased to less than 200.

Figure 2 shows the station distribution in four time periods (1947–72, 1973–84, 1985–98, 1999–2014). The decline of the station network is clearly apparent. The gauge dataset is most extensive during the period 1947–73. Within that period adequate stations exist so that there is generally at least one station in each $2.5^\circ \times 2.5^\circ$ grid box (Fig. 4). Notably, hundreds more stations existed in the DRC and Angola at that time, but the records have not been digitized and incorporated into the archive used here. This time period is used to produce spatial patterns [principal components (PCs)] that allow gaps in earlier and later years to be filled. That procedure is described in section 3c. From here on, for the sake of simplicity, 1947–73 is referred to as the calibration period (see section 3a).

The paucity of station records in the Congo basin itself in recent years is illustrated in Fig. 5, which shows the total number of station months available in 1° grid boxes during the period 1995–2012. This period is examined because NIC131 includes extensive DRC data through 1994, but little after that time. The absence of records in most of the DRC and Angola is clearly apparent. Figure 5 also shows the post-1994 station density in GPCC, which includes the original NIC131 dataset. The current version of NIC131 now includes recent updates from the DRC and numerous other countries in equatorial Africa.

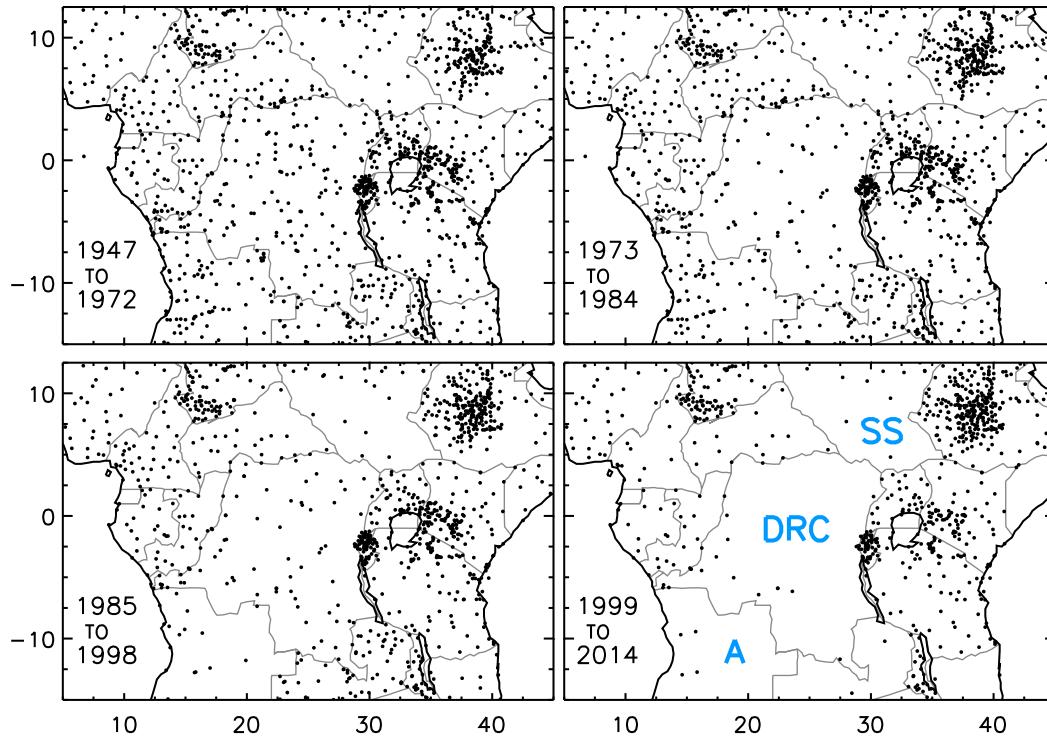


FIG. 2. Station network around Congo basin in four time periods: 1947–72, 1973–84, 1985–98, 1999–2014. In the map for 1999–2014 the letters SS, DRC, and A respectively identify the countries of South Sudan, the Democratic Republic of the Congo, and Angola.

These records have not yet been incorporated into GPCC, which was used by Zhou et al. (2014), so that the current NIC131 dataset is more complete in recent years.

The decline in the number of records in the NIC131 archive since approximately 1980 is a result both of the decline of the networks in many countries and the increased difficulty and cost for obtaining records. The post-2010 decline reflects the time lag in making data available outside of the meteorological services, as well as the timing of the data acquisitions from the various countries.

The long-term decline in the number of operative stations since the peak in the 1960s and 1970s coincides with a major administrative change, in that most African countries gained their independence from colonial powers around 1960 (Nicholson 1994). Unfortunately, this was followed in most cases by an economic decline, which severely impacted the meteorological services. The resulting impact was threefold. The observing stations declined in number, and in many countries the regular publication of meteorological data ceased, seriously impeding data acquisition by researchers. At the same time, most meteorological services started charging very high sums for providing data. Particularly hard-hit by economic and social factors were the Democratic Republic

of the Congo, which became independent in 1960, and Angola, which became independent in 1975. In both cases, the meteorological services all but ceased to function for one or more decades. Fortunately, they are currently being built back up, but with far fewer reporting stations than were operating prior to independence.

Despite the changes of station network and paucity of data in some time periods, there are nevertheless a large

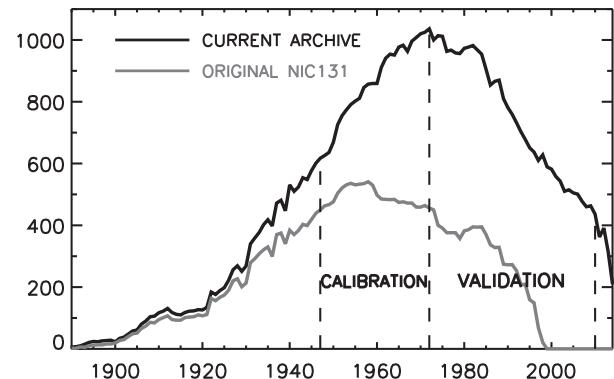


FIG. 3. Stations available by year in the original NIC131 archive vs the current version of the archive. Vertical lines delineate the calibration and validation periods.

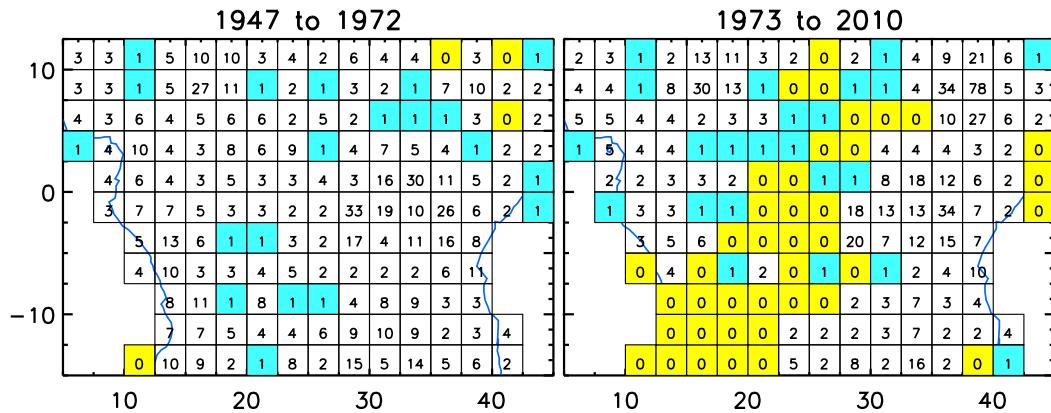


FIG. 4. Number of stations in each grid box available in at least 50% of the years. (left) The period 1947–72 and (right) the period 1973–2010. Grid boxes with no stations are shaded in yellow; those with only one station are shaded in blue.

number of very long station records (Fig. 6). Nearly half the station records are 50 years or longer and some 200 stations have 70 or more years of record.

b. Quality control

The quality control procedure is threefold. First, extreme values are statistically identified and outliers are examined and assessed (via comparison with nearby stations, with previously documented extremes, and with extreme events). Second, each individual record is

subjectively scanned to identify any obvious problems and extreme values. The latter are further examined via comparison with values at nearby stations, etc. The visual scanning helps to pick up patterns that might not be objectively identified (e.g., runs of identical values). Third, each station record is correlated with the multistation time series for the geographic region in which it lies (see section 3d). If the station–region correlation is very low, the station is further scrutinized. Low correlations may indicate a problem, such as a discontinuity. If no specific

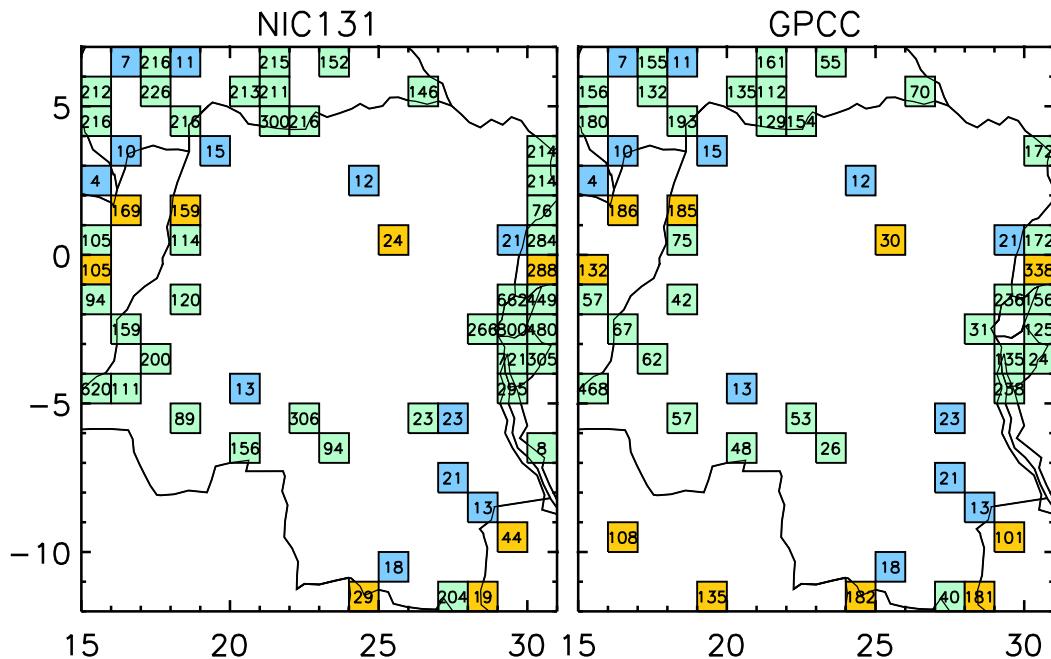


FIG. 5. Total number of months with available rainfall records during the period 1995–2012. (left) Tabulations for the NIC131 dataset and (right) tabulations for GPCC. Grid boxes in orange/green are those with a greater number of records in GPCC/NIC131. Blue indicates the same number of stations in both datasets.

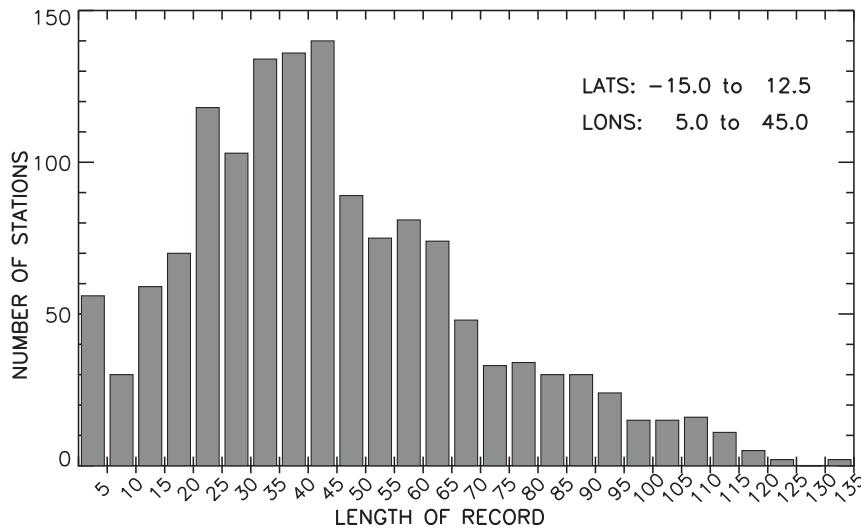


FIG. 6. Histogram of the number of years of record for stations in the analysis sector.

problem is identified (i.e., no outliers or inhomogeneities are evident), the station is noted and retained in the dataset, but excluded from regional averages (see section 3d).

3. Methodology

a. Overview

Because of the large spatial variability of tropical rainfall (Nicholson 2011) an individual station record does not provide reliable information on areally averaged rainfall. To reduce the random error in individual station records, multistation averages should be used in rainfall analysis. In Sahelian Africa, for example, Lebel and Le Barbé (1997) find that roughly five rainfall stations are needed to reduce the error in seasonal totals to 3% for a $1^\circ \times 1^\circ$ region; the error associated with a single station is roughly 5%. The paucity of stations in recent years makes even that criterion difficult to meet in many parts of the analysis region. Thus, it is difficult to establish with certainty rainfall trends over much of the Congo basin.

Statistical approaches can help to solve this problem. Accordingly, the ultimate goal of this project is to create a reliable, gridded rainfall dataset for the Congo basin and surrounding countries. However, because of the relatively sparse network in some regions, datasets are created on three spatial scales. Originally, an attempt was made to produce data at the $2^\circ \times 2^\circ$ spatial scale. However, by expanding to 2.5° , the number of stations per grid, and hence reliability, went up significantly. This resolution is compatible with many other relevant datasets, such as GPCC gauge data.

Clearly, the greater the number of stations in the spatial aggregation, the lower the error in the rainfall estimate. For this reason, a 5.0° gridded dataset and a somewhat larger-scale regional dataset were also created. The approach for gridding rainfall estimates is described in sections 3b and 3c, and that for the regional estimates is described in section 3d. The gridding is based on kriging for the period 1947–72 (see section 3b) and on spatial reconstruction (see section 3c) for the periods 1921–46 and 1973–2014. The latter approach is based on PCs that quantify the spatial teleconnections within the region. The rationale for the two different approaches is that 1) it is well known that kriging does not perform well in cases of low station density and 2) spatial reconstruction requires calibration, that is, calculation of PCs, on a temporally invariant dataset. The approach taken is to use 1947–72 as a calibration period for the spatial reconstruction method that is applied in earlier and later years. Although the overall station number is very high into the 1980s (Fig. 3), in the Congo basin itself the station number and spatial distribution is optimal during the period 1947–72. During those years the station network in the analysis sector is relatively complete, so that kriging was required in few cases.

The period 1973–2010 serves as a validation period, the final year being chosen based on station density. Initially, kriging and spatial reconstruction were both used to produce a gridded dataset for that period, and the results validated. However, as anticipated, the spatial reconstruction approach was far superior (see section 3d), so that the final gridded dataset was produced using spatial reconstruction, which covers the period 1921–2014. It is limited to those years because the station network is very

sparse in earlier and later years. However, at the regional level more early data are available, so that for most regions the dataset commences in the earliest years of the twentieth century.

Data are produced at monthly, seasonal, and annual time scales. Six seasonal aggregations are produced: January–March (JFM), AMJ, July–September (JAS), October–December (OND), March–May (MAM), and October–November (ON). The latter two seasons best capture the rainy seasons in East Africa. It should be noted that the seasonal and annual values are not simply the sum of monthly values. In both of those cases, the procedure commences at station level. This approach to temporal aggregation serves to reduce the error in the estimates.

The task of creating a gridded dataset for the Congo basin is made difficult by the large rainfall gradients within the region. While most of the area exhibits a similar seasonal cycle, with two peak seasons occurring in the boreal spring and boreal autumn, the annual totals are exceedingly diverse. In the northwest, in the mountains of Cameroon, mean annual rainfall in most areas exceeds 3000 mm. At one station, Debundscha, mean annual rainfall exceeds 10 000 mm. Just to the south and also in the northeast lie deserts with mean annual rainfall on the order of 100–300 mm. However, this problem is offset by the fact that, despite the disparate climatology, large areas are relatively homogeneous with respect to interannual variability (Nicholson 1986). That fact is a key to the creation of the datasets described in this article.

b. Gridded estimates for the “calibration” period 1947–72

Figure 4 shows the total number of stations in each 2.5° grid box of the analysis sector during the period 1947–72. The number varies between 0 and 33, but most boxes contain between 3 and 10 stations. Station density is particularly high in Chad, Rwanda, Uganda, and parts of Kenya and Ethiopia. It is very low over the DRC, South Sudan, and the CAR (see Fig. 2 for location).

Ideally, spatial averages at 2.5° should have five or more stations. Xie and Arkin (1995), for example, calculated that when five or more gauges are available at this resolution, the error in the simple station average is generally less than 10%. Clearly, because of low station density, simple averaging would not be feasible over all of the Congo basin. For that reason, two approaches to filling in missing data were considered, the commonly used kriging technique (Isaaks and Srivastava 1989) and “natural neighbor” (Watson 1992). Both interpret a grid box’s value based on a weighted sum of values at local, neighboring points. In most cases, the results were nearly

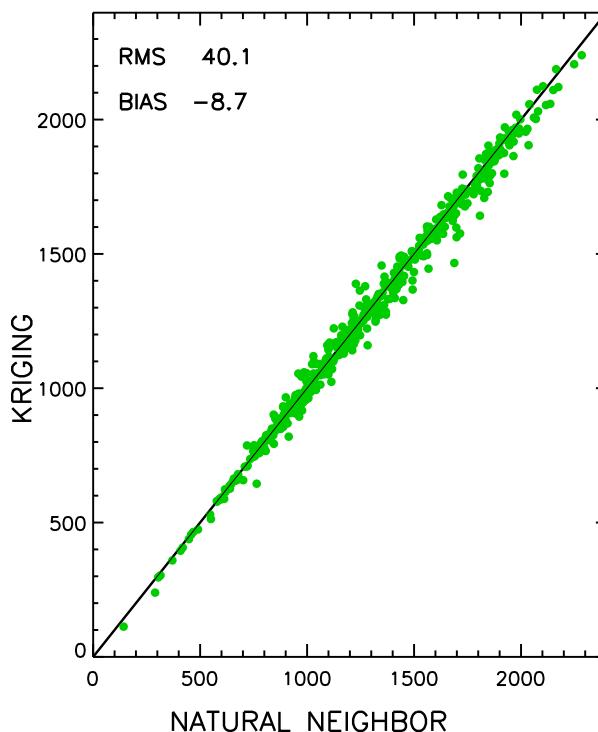


FIG. 7. Kriged values vs natural neighbor estimates (mm) for boxes deemed missing during the 1947–72 period.

identical (Fig. 7). However, the natural neighbor approach tended to underestimate rainfall at very wet stations (see section 3d). For that reason, kriging was selected to produce the gridded dataset used to calibrate the spatial reconstruction.

The final gridded dataset for the 1947–72 period was created by using simple station averages for grid boxes with “adequate” data and filling in the remaining grid boxes via kriging. The task, then, was to decide how to define “adequate,” that is, enough stations to ensure a reasonable spatial average. This was done, in conjunction with the validation described in section 3d, by way of a procedure that is somewhat akin to cross validation. All grid points with a station number exceeding some threshold were denoted as “missing” and the value was then computed via kriging. The average and the kriged values were then compared. For 2.5° grid boxes, thresholds of two, three, and four stations were tested. For 5.0° grid boxes, thresholds of five, six, and seven stations were tested. The results are described in section 3d.

Using the AMJ season as an example, Fig. 8 shows for each grid box the percent of years in which two or more stations were available during the period 1947–72. This figure was 100% in most of the analysis sector. However, within the area of Congo rain forest, one of the grid boxes had at no time two or more stations available

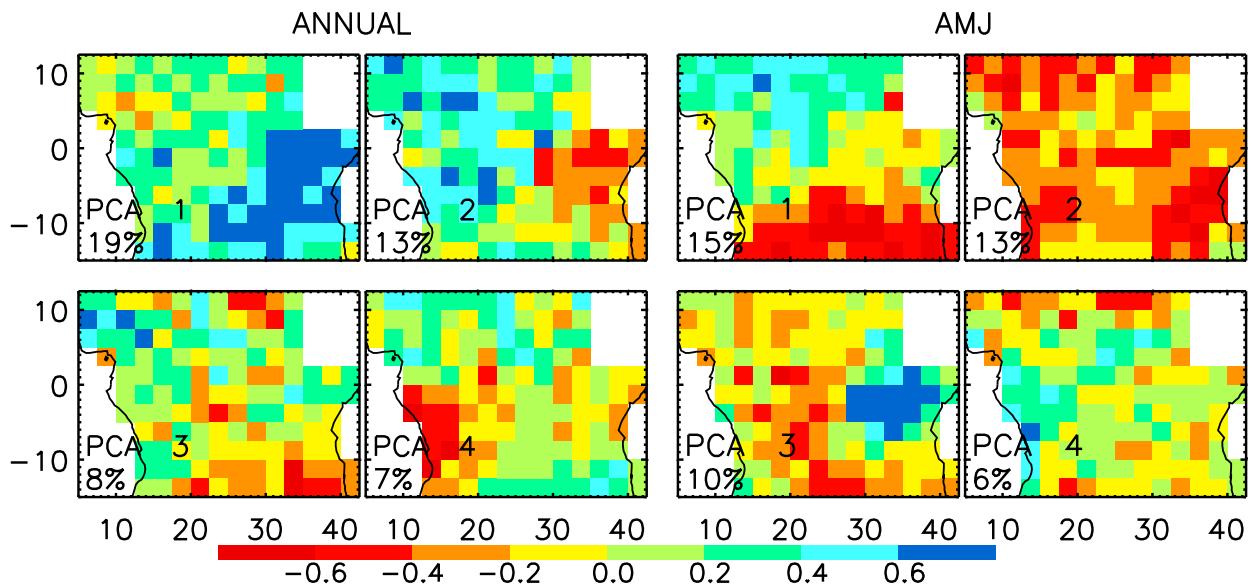


FIG. 9. The first four annual and AMJ PCs with percent variance explained. Colors indicate dimensionless PC loadings.

spatial loadings prevailing throughout most of the equatorial region. The few negative loadings are small and generally in the far northwest. The second PC is dominated by negative loadings in the east and positive loadings in the west. PC3 shows mainly a north–south opposition, with positive loadings concentrated north of the equator and negative loadings concentrated south of the equator. PC4 has a more diffuse pattern, but is dominated by strongly negative loadings in the western sector south of the equator.

For the AMJ season, PC1 shows mainly a north–south opposition. PC2 shows a relatively uniform pattern of negative loadings throughout the analysis sector. PC3 is dominated by positive loadings over eastern equatorial Africa, where AMJ is the main rainy season. Elsewhere loadings are predominantly negative. PC4 is much more diffuse, with the strongest loadings over western equatorial Africa.

d. Validation of the dataset

Validation is carried out by comparing the results of kriging and/or spatial reconstruction with the values obtained by simple averaging of the station data in grid boxes with an adequate number of stations. The first task is to define “adequate.” This is assessed for the calibration period 1947–72, variously testing definitions of two, three, or four stations for 2.5° grid boxes and five, six, or seven stations for 5° grid boxes. This step serves two purposes. One is to select a threshold for defining missing grid boxes that needed to be kriged during the calibration period or needed to be estimated via spatial reconstruction in other years. The second is to choose a

station threshold in selecting grid boxes for the validation of estimates during the 1973–2010 period.

In testing these various thresholds for “adequate,” each grid box and year with an adequate number of stations is treated as missing. Its value is then filled in, with kriging in one case and with natural neighbor in another. Then the so-derived value is compared with that based on simple averaging of the data for the actual stations available in that grid box and year. This procedure is repeated for each grid box with an adequate number of stations and the results plotted in a scatter diagram.

Figure 10 shows the results for annual data for the calibration period 1947–72 at 2.5° spatial resolution. When missing annual values are supplied via kriging, the root-mean-square error (RMSE) using the simple average of the stations in the grid box as a reference is 157, 161, and 172 mm, respectively for two-, three-, and four-station thresholds. The bias in these cases is -20 , -26 , and -32 mm, respectively. Error statistics are similar for estimates obtained via the natural neighbor approach, but underestimation at high rainfall totals is clearly apparent. In both cases bias and RMSE are slightly lower with two stations. Given the similar results for the various thresholds tested, a decision was made to define adequate as a grid box containing two stations, based on the following considerations. Random error of gauge averages is lower with higher station numbers. However, the larger the number of grids with values from station data only, the lower the error potentially introduced by the statistical approach. Also, the lower threshold increases the number of grid boxes that can be used in the validation.

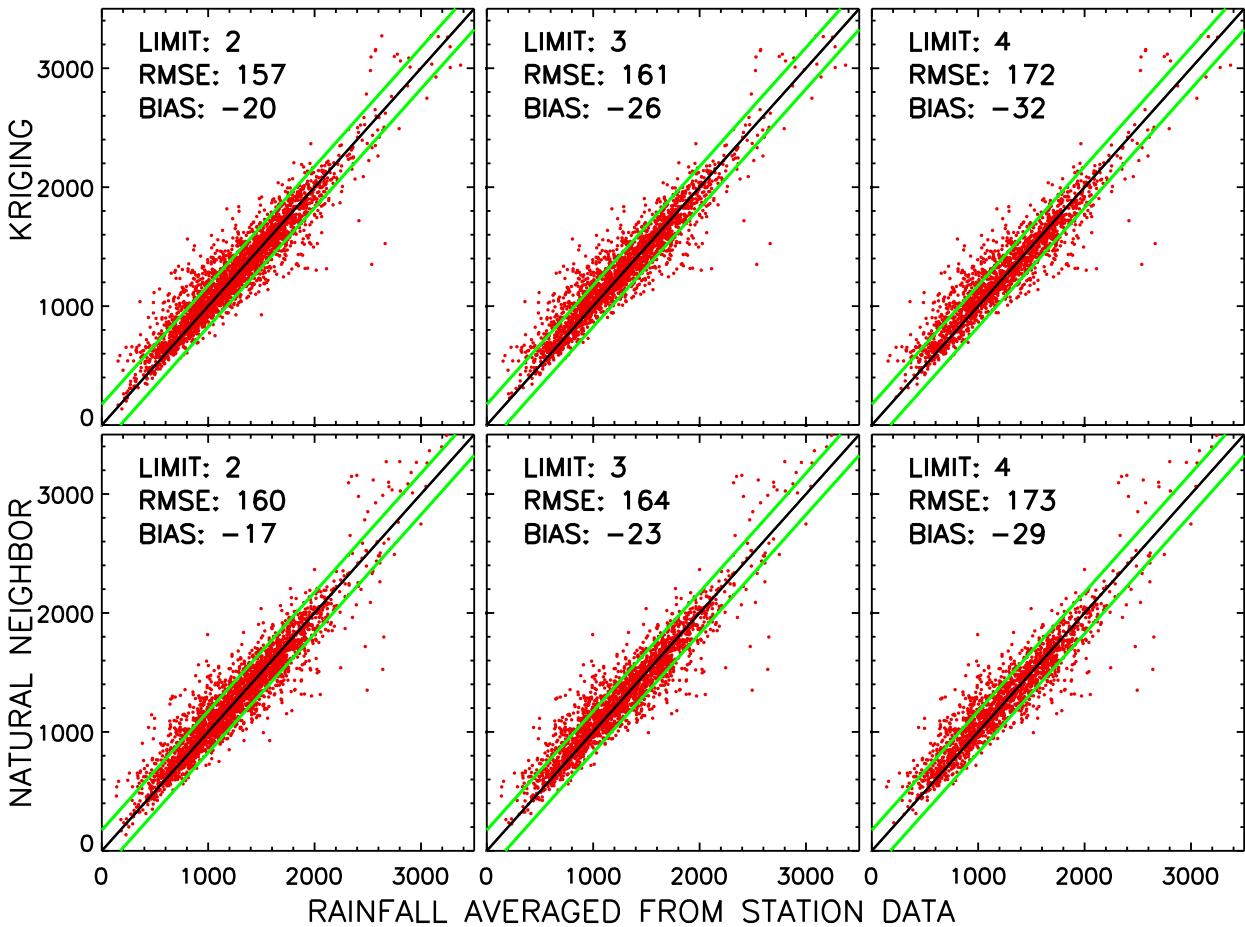


FIG. 10. Validation of annual estimates (mm) via (top) kriging and (bottom) natural neighbor for the period 1947–72 at 2.5° spatial resolution. The three graphs differ in terms of the threshold for including a grid box in the validation, using thresholds of two, three, and four stations in a grid box for inclusion.

A similar test was conducted for 5° data. In that case, station thresholds of five, six, and seven were used in selecting grid boxes to be used for validation. As with the 2.5° analysis, there is little difference in the results. However, results with the five-station threshold had slightly lower bias and RMSE, so that threshold was chosen defining missing boxes that need to be kriged during the 1947–72 period, for choosing missing grid boxes for which spatial reconstruction needed to be applied in other years, and for validating the spatially reconstructed data during the period 1973–2010.

The kriged values of seasonal data were also validated for the period 1947–72, but using only the two-station and five-station threshold in selecting grid boxes for the validation. The results are summarized in Table 1. RMSE is on the order of 70–80 mm for seasonal data and 209 mm for annual data at 5° resolution and slightly higher at 2.5° resolution. Notably, kriging was required for relatively

few grid boxes (see Fig. 7). Of the 132 grid boxes in the analysis sector, only 7 were kriged in all years. Only 27 were kriged in more than 30% of the years.

For the period 1973–2010, the validation is carried out in a fashion similar to that for 1947–72. However, in this case, the validation is used both to estimate error and to compare rainfall estimates based on kriging with those

TABLE 1. Validation of kriged estimates of rainfall for the period 1947–72 at 2.5° and 5.0° spatial resolution. Bias is in millimeters, RMSE values are in millimeters per year or per season.

	Bias		RMSE	
	5°	2.5°	5°	2.5°
Year	-37.8	2.0	209	215
JFM	-3.7	3.7	68	80
AMJ	-16.5	-3.8	74	86
JAS	-10.4	-3.0	81	94
OND	-9.3	4.5	81	90

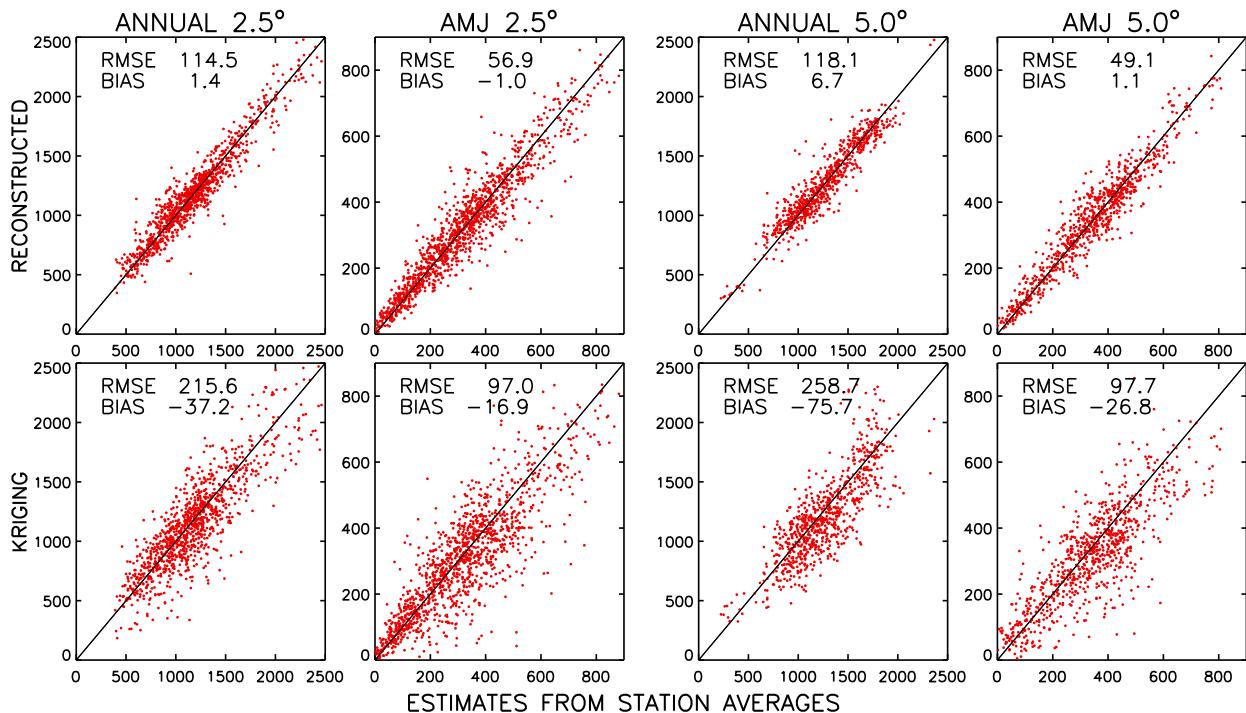


FIG. 11. Validation of annual and AMJ estimates (mm) via (top) spatial reconstruction and (bottom) kriging for the period 1973–2010 at 2.5° and 5.0° spatial resolution.

based on spatial reconstruction. Figure 11 shows the results for annual and AMJ data at spatial resolution of both 2.5° and 5.0°. Additional results for monthly and seasonal data are summarized in Table 2. The superior results with spatial reconstruction are clearly apparent in Fig. 11 and Tables 2 and 3 for both annual and AMJ data. For spatial reconstruction the RMSE for seasonal data is similar at both spatial scales and ranges from 42 to 60 mm. RMSE for monthly data is roughly half that. Annual RMSE is 118 mm for 5° data and 114 mm for 2.5° data. The RMSE for kriging is roughly twice that for spatial reconstruction. Moreover, the bias is minimal in the case of spatial reconstruction but as large as -75.7 mm for annual estimates via kriging.

At this point a dataset is produced for monthly data, seasonal, and annual data for each of the grid boxes in the analysis sector (Fig. 2) and for each year from 1921 to 2014. To summarize the approach, gridding is via simple spatial averaging when two or more (five or more) stations are available at 2.5° (5.0°) resolution. In other cases, rainfall estimates are based on kriging during the period 1947–72 and on spatial reconstruction in other years. Although in theory this results in a discontinuity, it is considered to be minor because there is so little missing data during the period 1947–72. For that reason, it was deemed unnecessary to apply spatial reconstruction to that period.

The validation exercise suggests that results are excellent, and the dataset was visually scanned to further ensure the quality of results. At this point, it became evident that for a handful of grid boxes near the edges of the analysis sector the results were questionable, that is, temporal discontinuities were clearly apparent. Those grid boxes, 5 along the Atlantic coast and 18 in northeast Kenya and Ethiopia, were removed from the dataset.

e. Regional dataset

The nature of the African precipitation records and the characteristics of the rainfall regime in most areas of the continent pose special problems in the analysis of the interannual variability of rainfall. The African station records are often fragmentary, sometimes discontinuous, and generally cover varied periods of record. As for climatic considerations, the spatial variability of mean rainfall is high and individual station records include a large stochastic component that reflects the randomness of the convective process in the tropics.

These problems can be minimized by using spatial averages instead of individual stations. Thus, for many past studies with the NIC131 dataset (e.g., Nicholson 1986; Nicholson and Kim 1997; Nicholson et al. 2012b) the gauge data were combined into 90 regions spread across the continent. Each region was determined to be

TABLE 2. Validation of estimates of rainfall for the period 1973–2010 via spatial reconstruction at 2.5° and 5.0° spatial resolution. Bias is in millimeters, RMSE values are in millimeters per year or per season.

	Bias		RMSE	
	5°	2.5°	5°	2.5°
Year	6.7	1.4	118	114
JFM	1.9	−0.8	50	46
AMJ	1.1	−1.0	49	57
JAS	0.1	−0.5	48	49
OND	1.5	−0.3	55	52
MAM	1.7	0.9	56	60
JJA	−1.0	−2.9	47	46
ON	1.1	0.0	47	42
Jan	0.2	−0.5	24	23
Feb	0.8	0.0	23	22
Mar	0.8	0.1	28	29
Apr	1.4	−1.2	32	35
May	0.7	−0.2	28	28
Jun	−0.7	−0.6	21	20
Jul	−0.2	−0.6	23	24
Aug	0.7	0.1	25	26
Sep	0.1	0.9	28	24
Oct	0.9	0.3	28	27
Nov	0.1	−0.5	27	27
Dec	0.4	−0.5	27	27

relatively homogeneous with respect to interannual variability. The regional averages are aggregates of station values expressed as a standardized departure from the station’s long-term mean, that is, the departure from the mean divided by the standard deviation. This standardization procedure minimizes the effect of dry stations where departures are large compared with the mean. It allows for a simple adjustment for temporal changes in number of stations in the region.

Correspondingly, the regional values are also expressed in units of standard departures. The use of standardized departures allows for the use of stations with diverse means and variances. This is particularly important when the station network varies over time and station records may contain extensive gaps. For that reason, this formulation is conventionally used in studies of interannual variability of African rainfall (Ali and Lebel 2009). It is possible to convert the standardized departures to units of rainfall, but for statistical analyses the standardized departure series are preferable and the regional dataset presented here utilizes this approach.

A four-step procedure is utilized to transform annual and monthly totals for individual stations into regional rainfall departure series (Nicholson 1986): 1) departure series are calculated for individual stations, 2) a regional average is derived, 3) the homogeneity of each rainfall region is tested via linear correlation and via an *F* test that compares

TABLE 3. Validation of estimates of rainfall for the period 1973–2010 via kriging at 2.5° and 5.0° spatial resolution. Bias is in millimeters, RMSE values are in millimeters per year or per season.

	Bias		RMSE	
	5°	2.5°	5°	2.5°
Year	−75.7	−37.2	259	216
JFM	−9.0	0.7	74	70
AMJ	−26.8	−16.9	98	97
JAS	−19.8	−13.8	121	110
OND	−17.5	−8.0	85	85

temporal and spatial variability within the region, and 4) regional series are corrected for inhomogeneities resulting from changes in station network from year to year. Use of the transformed regionally averaged series reduces two problems inherent in the analysis of rainfall in tropical areas: the highly diverse means and variances and the randomness of the convective process reflected in individual station totals.

The regions were identified in earlier studies and tested for homogeneity (e.g., Nicholson 1986). Stations were initially assigned to a region based on geography and rainfall seasonality. As indicated, the calculation for each region commences with individual station records. The value for each year *j* and station *i* is expressed as a standardized departure from the station mean \bar{r}_i :

$$X_{ij} = (r_{ij} - \bar{r}_i) / \sigma_i. \tag{1}$$

In the above X_{ij} is the standardized departure, and σ_i is the standard deviation over the period for which the mean is calculated. Such calculations can be done for annual, seasonal, or monthly time series.

The regional value R_j is an arithmetic average of the standardized departures from all *i* stations within the sector that are available for the season or year in question. For example, in the case of annual values,

$$R_j = I_j^{-1} \sum_{i=1}^I X_{ij}. \tag{2}$$

The final step is an adjustment for cases in which only a small number of stations comprise the average. Adjustment factors were determined, as in Nicholson (1986), by computing the ratio of the standard deviation of the time series for the full complement of stations versus the standard deviation when the time series consists of *i* number of stations. The analysis suggested that once five stations were available, no adjustment was needed as the variance did not change significantly with station number.

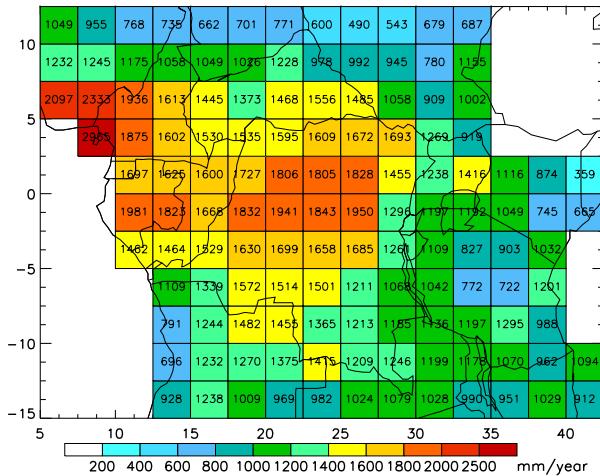


FIG. 12. Map of mean annual rainfall (mm) for the period 1947–2010.

The updated dataset includes many new stations and in some cases additional data for stations already in the original dataset. To determine the appropriate region for new and enhanced stations, the rainfall time series for each station is correlated with the time series for several regions near its geographical location. Each station is assigned to the region with which it was best correlated, if a correlation above roughly 0.4 was found. The cutoff for inclusion was subjectively determined, as it depended on the record length and on the spatial homogeneity of the rainfall regime in the region. In cases with a correlation lower than 0.4, the station is retained in the NIC131 dataset but is excluded from regional averages.

4. Results

The purpose of this study is both to produce a reliable gridded dataset for equatorial Africa and to compare the results based on this dataset with those given in Zhou et al. (2014), based on GPCP and TRMM. The dataset is illustrated with examples of annual and AMJ data (section 4a) and with rainfall anomaly patterns in select individual years (section 4b). The comparison with the results of Zhou et al. (2014) is presented in section 4c.

a. Climatological means

Figure 12 shows the gridded climatological mean of annual rainfall for the period 1947–2010. The figure shows high spatial gradients in the western and eastern portions of the region, but relatively weak spatial gradients within the Congo basin itself. In western Cameroon and southeastern Nigeria annual rainfall

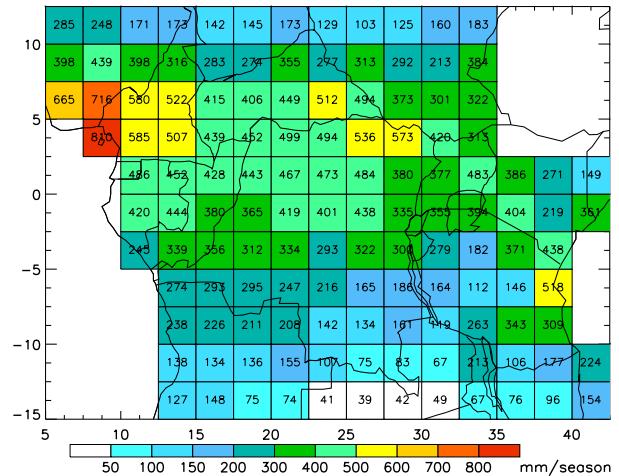


FIG. 13. Map of mean AMJ rainfall (mm per season) for the period 1947–2010.

is on the order of 3000 mm, but rapidly diminishes southward to some 700 mm in the coastal region of Angola. In the east annual rainfall is on the order of 700–1000 mm, and as low as 359 mm in the driest regions. In contrast, throughout the Democratic Republic of the Congo, which comprises most of the Congo basin, annual rainfall is on the order of 1600–1800 mm and peaks at 1950 mm in the center of the basin.

The AMJ seasonal mean for 1947–2010 (Fig. 13) shows a similar spatial pattern, except that this season becomes very dry in the northern and southern extremes of the analysis sector. Rainfall is concentrated in the latitudinal span from 7.5°N to 2.5°S, with seasonal means on the order of 300–600 mm in most areas. This season is much wetter in the Atlantic coastal sectors north of the equator (500–800 mm) and much drier in the Atlantic coastal sector south of the equator (100–400 mm). AMJ rainfall to the north and east of the Congo basin is on the order of 100–400 mm in most locations.

b. Select individual years

To illustrate the year-to-year variability in this region, Figs. 14 and 15 present rainfall anomalies in select years. The annual anomalies, as well as the anomalies in the MAM and ON seasons, are presented. These seasons are considered because in most cases the largest anomalies are in eastern equatorial regions, where this is the most appropriate seasonal aggregation. Included are three extremely wet years (1961, 1997, 2006) and several extremely dry years (1949, 1983, 2005). To compare locations with very different rainfall climatologies, rainfall is represented by a standardized departure (departure

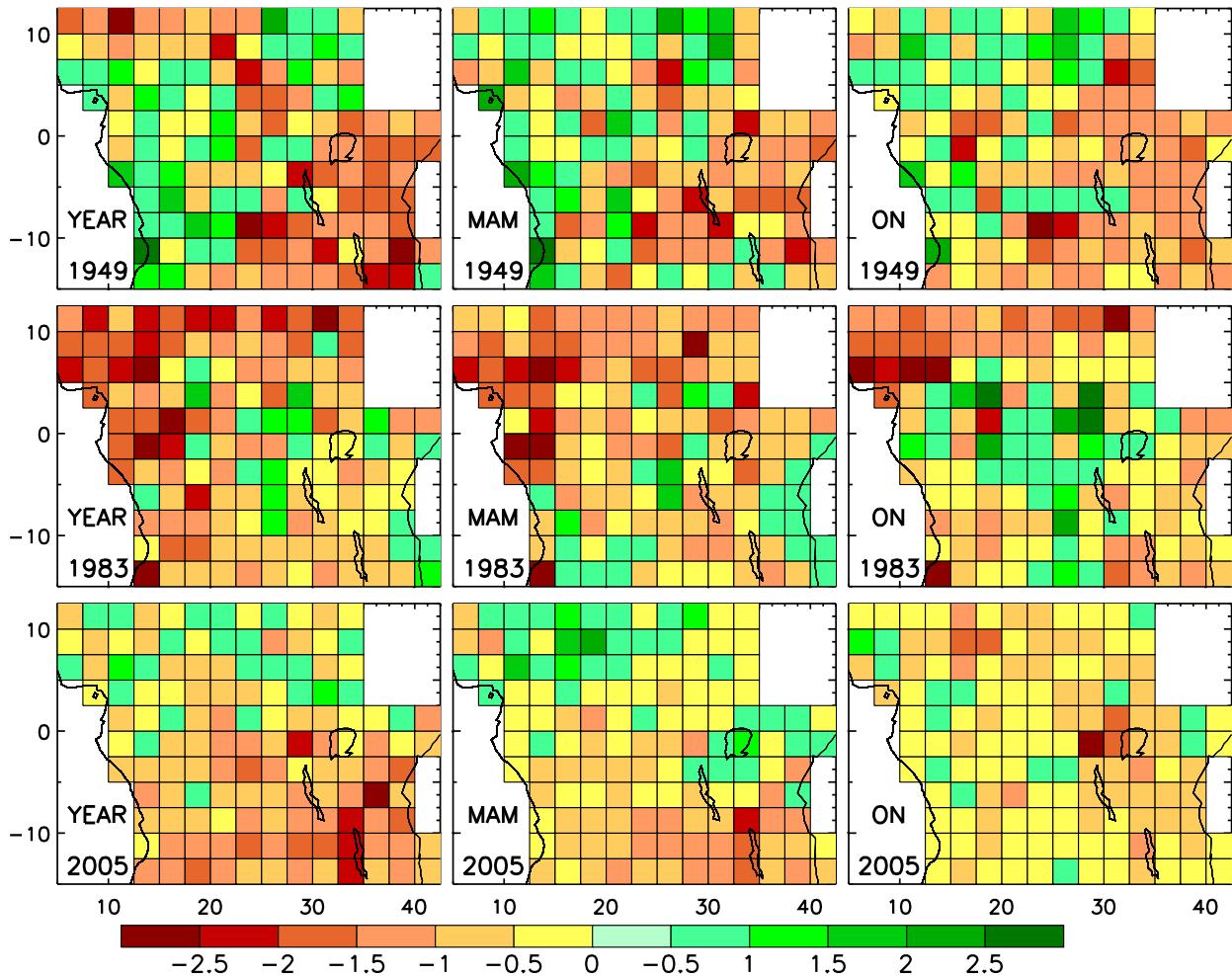


FIG. 14. Standardized (see section 3e) annual, MAM, and ON rainfall anomalies in three dry years.

from the 1947–2010 mean, divided by the standard departure for that same period).

Two of the three dry years (1983 and 2005) are characterized by below-average rainfall almost ubiquitously within the analysis sector (Fig. 14). In both cases abnormally dry conditions are evident not only in the annual rainfall but in both the MAM and ON seasons as well. In 1949 the driest region is eastern equatorial Africa between roughly the equator and 15°S, but rainfall is above normal in many of the grid boxes, especially to the west and north. As in 1983 and 2005, the anomaly patterns are similar in both seasons. Note that 1983 was a year of dry conditions throughout nearly the entire continent. In both 1949 and 1983, rainfall in many areas is more than 1.5 standard deviations below normal. In 2005 the anomalies are more uniform across the region in all seasons and generally weaker than in 1949 and 1983.

The spatial anomaly patterns are quite different in the three wet years (Fig. 15). The two main differences are

that the anomalies are not very spatially uniform and that the anomalies tend to be of opposite sign in the two seasons in eastern areas of the analysis sector. The year 1961 was phenomenally wet throughout the equatorial region, with extraordinary rainfall occurring in the driest sectors of eastern Africa during the ON season (Nicholson 2011). There have been several case studies of that year (e.g., Flohn 1983), which saw rapid rises of the East African lakes. In the regions with extreme wet conditions during ON, the MAM season was anomalously dry. The wet conditions of 1997 are much more moderate for the year as a whole, but extreme in eastern equatorial Africa. The MAM season shows a highly variable spatial pattern with both negative and positive anomalies being weak. In 2006 the very wet conditions are confined to eastern equatorial Africa, with the annual anomalies looking very much like those of ON. However, in both seasons anomalies are relatively

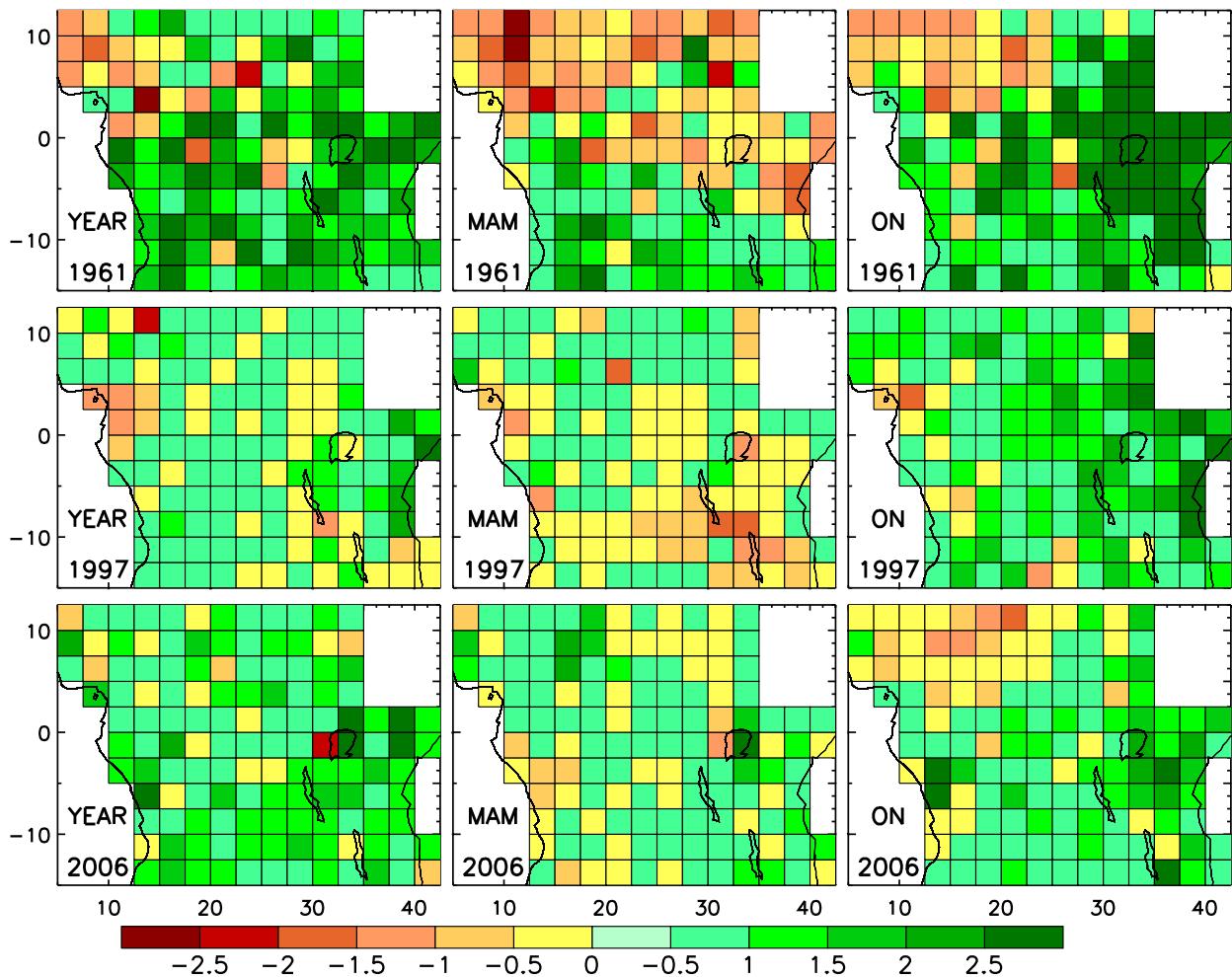


FIG. 15. Standardized (see section 3e) annual, MAM, and ON rainfall anomalies in three wet years.

weak compared to the other five years illustrated in Fig. 14 and 15.

c. Interannual variability of AMJ rainfall

To compare results with those of Zhou et al. (2014), rainfall trends for the AMJ season are calculated over the period 1985–2012. The results of trend analysis in that study are shown in Fig. 16. Those based on the spatially reconstructed dataset produced here are shown in Fig. 17.

Notably, the results of the two analyses are not strictly comparable because of different spatial resolutions (2.0° vs 2.5°) and because Zhou's analysis is based on GPCC, which has few gauge stations during that time period. However, similarities are evident in trends over the period 1985–2012. In both cases, there are strong negative trends in eastern and southern portions of the Congo basin sector studied by Zhou et al. (2014).

Likewise, both show positive trends in more western sectors, particularly in the northwest. This pattern of upward AMJ trends in the northwest and downward trends in the south and east prevails over most of the area between 12.5°N and 20°S . This suggests large-scale factors in the trends. The most noticeable contrast in the results of the two analyses is that negative trends prevail over larger areas in the Zhou et al. (2014) analysis.

An examination of the regional datasets can also shed some light on the question of recent rainfall trends in the Congo basin. As indicated earlier, these may be more reliable than the gridded data since they incorporate a larger number of stations than do individual grid boxes and because no statistical manipulation has been performed. The time series for 14 regions in or abutting the Congo basin are shown in Fig. 18. All but four have data through 2012. The location of these regions is shown in Fig. 19.

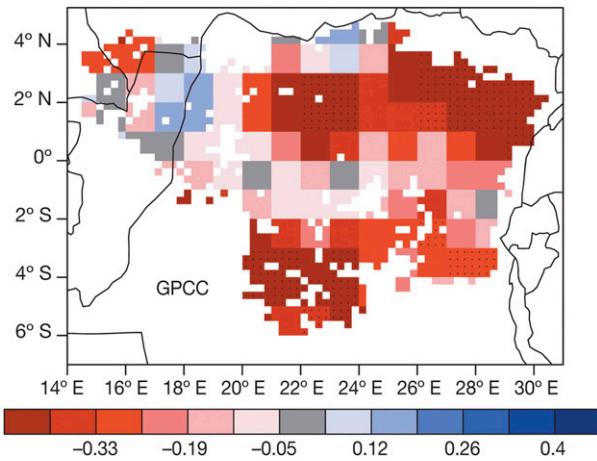


FIG. 16. AMJ rainfall trends ($\text{mm day}^{-1} \text{decade}^{-1}$) 1985–2012 from Zhou et al. (2014), based on GPCC.

Many of the regions show a decline in AMJ rainfall commencing sometime in the 1960s. This is the case for regions 91, 90, 85, 41, 40, 39, 32, 26, and 21. Several, such as regions 21 and 26, show a recovery in very recent years. The time series for regions 30, 34, 35, and 36 suggest a general increase in rainfall since the 1960s.

During the period 1985–2012, the time series suggest negative trends over most of these regions during the AMJ season. A quick scan suggests this for regions 32, 36, 39, 40, 41, 42, 85, and 91. These border the eastern and southern portions of the Congo basin. Positive trends are evident in the regions to the northwest of the Congo basin: 21 and 26. Unfortunately, within the core of the basin, the trends are less clear, partially as a result of inadequate data in some regions. In the most central region, 30, where the data extend only to 2005, there is little evidence of a downward trend. That is true also for the remaining regions: 34, 35, and 90.

The AMJ trends over the period 1985–2012 are quantified in Fig. 20. A pattern of downward trends in the east and south, upward trends in the northwest, and weak trends in the central basin is very clear. This is also consistent with the inherent patterns of interannual variability in the region: typically, an east–west opposition (Nicholson 2014) and local, orographic controls on rainfall in the central basin (Jackson et al. 2009). The trends are significant at the 10% level or better in over half of the regions shown.

5. Summary

This article describes a newly created gridded rainfall dataset for equatorial Africa. It includes rainfall estimates at both 2.5° spatial resolution and 5.0° spatial resolution. It includes monthly, seasonal, and annual

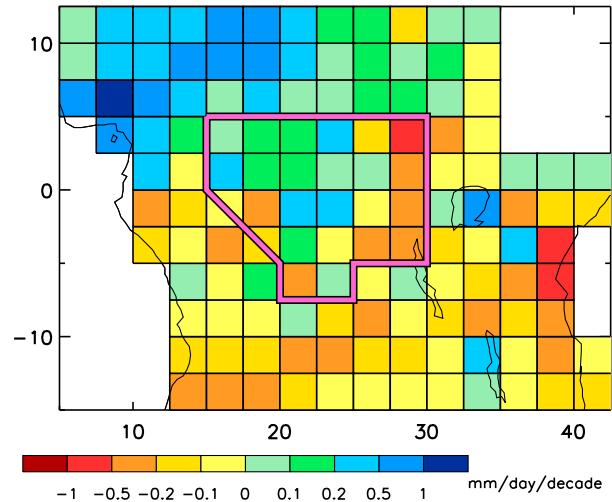


FIG. 17. AMJ rainfall trends 1985–2012 based on the spatially reconstructed dataset. The red line indicates the approximate area of the study by Zhou et al. (2014).

data, with six sets of monthly aggregations comprising the seasons: JFM, AMJ, JAS, OND, MAM, and ON. Two seasonal datasets for boreal spring and two for boreal autumn were produced because various authors utilize different definitions of the rainy season in equatorial Africa. Thus, the availability of both six sets of seasonal estimates facilitates the use of this dataset by the research community.

The dataset extends from 1921 to 2014. During the period calibration period 1947–72, missing values are filled via kriging. During other years, spatial reconstruction is utilized. Spatial reconstruction is not used during the period 1947–72 because it must be applied to spatial arrays without missing data. To do that, kriging would necessarily have been applied to missing 1947–72 data prior to the application of spatial reconstruction. This would compound the error in the estimates. Moreover, the gauge network was dense enough that little kriging was necessary.

The validation exercises suggest that the gridded rainfall estimates based on spatial reconstruction are excellent. At both 2.5° resolution and 5.0° resolution the RMSE is generally 20–30 mm for monthly data, 40–60 mm for seasonal data, and ~ 115 mm for annual data. There is little bias in any of the estimates. Errors for the period 1947–72, in which kriging is used to fill in gaps, are notably larger but kriging is applied to relatively few grid boxes. It should be noted, however, that most of the validation covered areas on the periphery of the basin. Given the similar meteorological circumstances within the basin, it is assumed that the error statistics are roughly valid for that area as well.

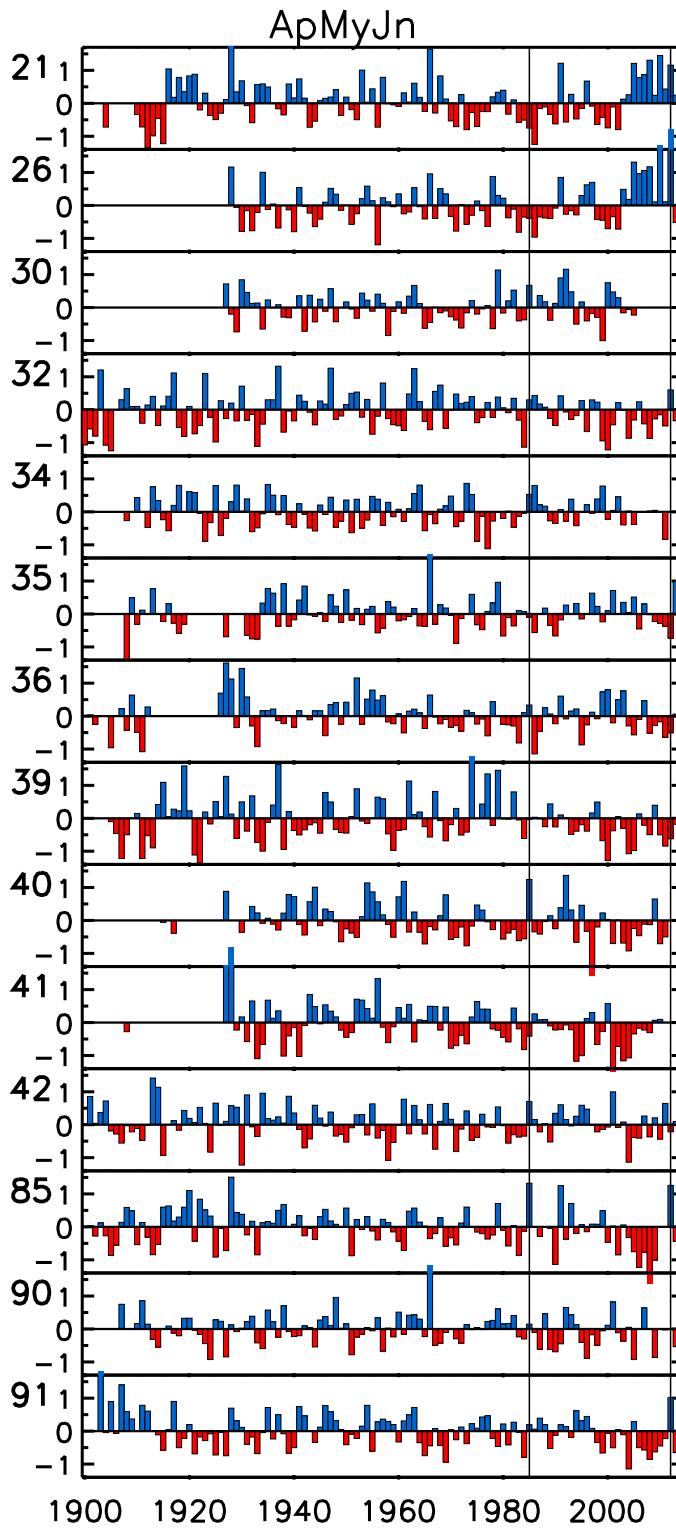


FIG. 18. Regional data series, with rainfall expressed as a regionally averaged standard departure from the long-term mean.

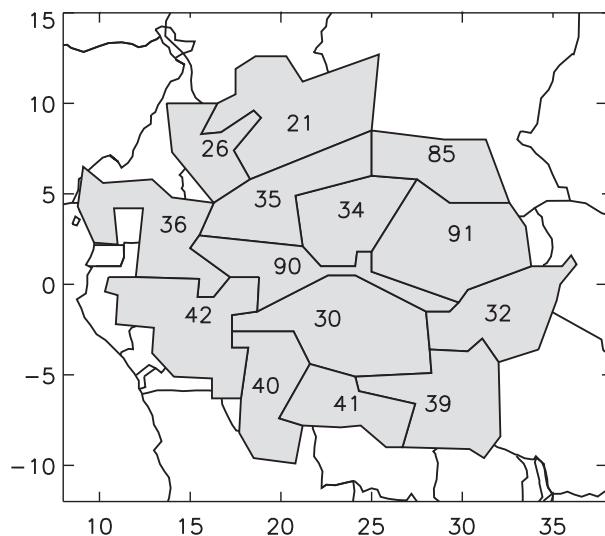


FIG. 19. Location of the regions in Fig. 18.

A very important result is demonstrating how much better the gridded data are when missing data are estimated via spatial reconstruction than via kriging. At 2.5° resolution the RMSE during the period 1973–2010 is generally 70–110 mm for seasonal data and 216 for annual data. At 5.0° resolution the RMSE is on the order of 75–120 mm for seasonal data and 259 mm for annual data. Thus, the error is roughly twice as great as with spatial reconstruction. The bias is also much larger than in the case of spatially reconstructed data, generally on the order of tens of millimeters in seasonal data and roughly 3–4 times larger for annual data.

During the period 1985–2012 AMJ rainfall shows an upward trend in northwestern portions of the Congo rain forest region. This trend is particularly strong in the northwest portion of the overall analysis sector, where it represents a recovery from the well-known drought conditions of the 1980s. Most elsewhere negative trends are apparent, but they are weak in the central Congo basin.

The results are very consistent with those of Zhou et al. (2014). However, we would argue that the trends evident in the dataset produced here are probably more realistic. Two points support this conclusion. One is the greater amount of data in the region, compared with GPCC during recent years. Another is the confirmation of our 1985–2012 trends with the regional data.

The dataset produced here appears to provide an improved product for equatorial Africa. However, it has limitations, particularly with regard to spatial and temporal resolution. The advantage of satellite data is that it can provide greater resolution. Our next step is to validate several satellite or satellite–gauge blended

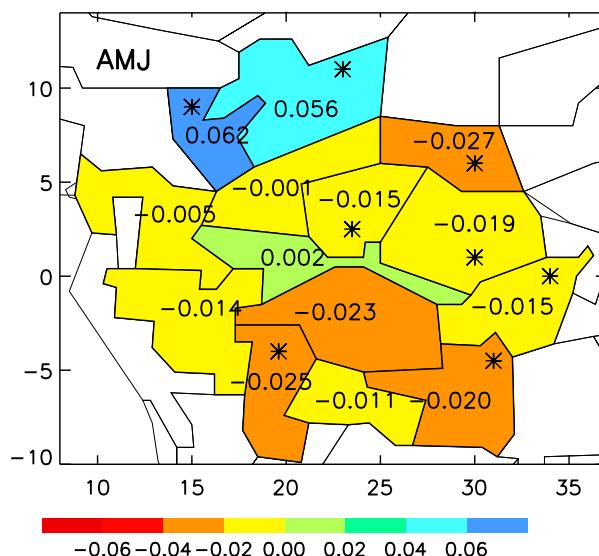


FIG. 20. AMJ rainfall trends 1985–2012 for the regional time series in Fig. 18. An asterisk indicates regions with trends significant at or above the 10% level. Units are standard departures over the analysis period.

products with the gauge data and with the gridded dataset produced here.

Acknowledgments. The authors thank the following individuals, who facilitated the acquisition of data for the project: Gil Mahé, Yolande Munzimi, Paul Dinga, Alphonse Kanga, Nathalie Rouché, and Raphael Tshimanga. The project was also supported by the National Science Foundation (NSF AGS-1535426 and AGS-1535439).

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Ali, A., and T. Lebel, 2009: The Sahelian standardized rainfall index revisited. *Int. J. Climatol.*, **29**, 1705–1714, <https://doi.org/10.1002/joc.1832>.
- Alsdorf, D., and Coauthors, 2016: Opportunities for hydrologic research in the Congo Basin. *Rev. Geophys.*, **54**, 378–409, <https://doi.org/10.1002/2016RG000517>.
- Asefi-Najafabady, S., and S. Saatchi, 2013: Response of African humid tropical forests to recent rainfall anomalies. *Philos. Trans. Roy. Soc. London*, **368B**, 20120306, <https://doi.org/10.1098/rstb.2012.0306>.
- Awange, J. L., V. G. Ferreira, E. Forootan, Khandu, S. A. Andam-Akorful, N. O. Agutu, and X. F. He, 2016: Uncertainties in remotely sensed precipitation data over Africa. *Int. J. Climatol.*, **36**, 303–323, <https://doi.org/10.1002/joc.4346>.
- Flohn, H., 1983: Das Katastrophenregen 1961/62 und die Wasserbilanz des Victoria-See-Gebietes (The catastrophic rains of 1961/62 and the water balance of the Lake Victoria region, in German). *Wiss. Ber. Meteor. Inst. Univ. Karlsruhe*, **4**, 17–34.

- Harris, N. L., and Coauthors, 2017: Using spatial statistics to identify emerging hot spots of forest loss. *Environ. Res. Lett.*, **12**, 024012, <https://doi.org/10.1088/1748-9326/aa5a2f>.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- Isaaks, E. H., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics*. Oxford University Press, 592 pp.
- Jackson, B., S. E. Nicholson, and D. Klotter, 2009: Mesoscale convective systems over western equatorial Africa and their relationship to large-scale circulation. *Mon. Wea. Rev.*, **137**, 1272–1294, <https://doi.org/10.1175/2008MWR2525.1>.
- Lebel, T., and L. LeBarbé, 1997: Rainfall monitoring during HAPEX-Sahel. 2. Point and areal estimation at the event and seasonal scales. *J. Hydrol.*, **188–189**, 97–122, [https://doi.org/10.1016/S0022-1694\(96\)03325-2](https://doi.org/10.1016/S0022-1694(96)03325-2).
- Long, Y., Y. Zhang, and Q. Ma, 2016: A merging framework for rainfall estimation at high spatiotemporal resolution for distributed hydrological modeling in a data-scarce area. *Remote Sens.*, **8**, 599, <https://doi.org/10.3390/rs8070599>.
- Malhi, Y., and J. Wright, 2004: Spatial patterns and recent trends in the climate of tropical rainforest regions. *Philos. Trans. Roy. Soc. London*, **359B**, 311–329, <https://doi.org/10.1098/rstb.2003.1433>.
- McCollum, J. R., A. Gruber, and M. B. Ba, 2000: Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa. *J. Appl. Meteorol.*, **39**, 666–679, <https://doi.org/10.1175/1520-0450-39.5.666>.
- Neukom, R. J., and Coauthors, 2011: Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Climate Dyn.*, **37**, 35–51, <https://doi.org/10.1007/s00382-010-0793-3>.
- Nicholson, S. E., 1986: The spatial coherence of African rainfall anomalies: Interhemispheric teleconnections. *J. Climate Appl. Meteorol.*, **25**, 1365–1381, [https://doi.org/10.1175/1520-0450\(1986\)025<1365:TSCOAR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<1365:TSCOAR>2.0.CO;2).
- , 1994: Recent rainfall fluctuations in Africa and their relationship to past conditions over the continent. *Holocene*, **4**, 121–131, <https://doi.org/10.1177/095968369400400202>.
- , 2011: *Dryland Climatology*. Cambridge University Press, 516 pp.
- , 2014: Spatial teleconnections in African rainfall: A comparison of 19th and 20th century patterns. *Holocene*, **24**, 1840–1848, <https://doi.org/10.1177/0959683614551230>.
- , and J. Y. Kim, 1997: The relationship of the El Niño–Southern Oscillation to African rainfall. *Int. J. Climatol.*, **17**, 117–135, [https://doi.org/10.1002/\(SICI\)1097-0088\(199702\)17:2<117::AID-JOC84>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0088(199702)17:2<117::AID-JOC84>3.0.CO;2-O).
- , B. Some, and B. Kone, 2000: An analysis of recent rainfall conditions in West Africa, including rainy seasons of 1997 El Niño and the 1998 La Niña years. *J. Climate*, **13**, 2628–2640, [https://doi.org/10.1175/1520-0442\(2000\)013<2628:AAORRC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2628:AAORRC>2.0.CO;2).
- , A. K. Dezfuli, and D. Klotter, 2012a: A two-century precipitation dataset for the continent of Africa. *Bull. Amer. Meteor. Soc.*, **93**, 1219–1231, <https://doi.org/10.1175/BAMS-D-11-00212.1>.
- , D. Klotter, and A. K. Dezfuli, 2012b: Spatial reconstruction of semi-quantitative precipitation fields over Africa during the nineteenth century from documentary evidence and gauge data. *Quat. Res.*, **78**, 13–23, <https://doi.org/10.1016/j.yqres.2012.03.012>.
- , C. Funk, and A. H. Fink, 2018: Rainfall over the African continent from the 19th through the 21st century. *Global Planet. Change*, **165**, 114–127, <https://doi.org/10.1016/j.gloplacha.2017.12.014>.
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese, 2015: GPCC full data reanalysis version 7.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. GPCC at DWD, accessed 16 August 2018, https://doi.org/10.5676/DWD_GPCC/FD_M_V7_050.
- Smerdon, J. E., A. Kaplan, D. Chang, and M. N. Evans, 2011: A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J. Climate*, **24**, 1284–1309, <https://doi.org/10.1175/2010JCLI4110.1>.
- Todd, M. C., and R. Washington, 2004: Climate variability in central equatorial Africa: Influence from the Atlantic sector. *Geophys. Res. Lett.*, **31**, L23202, <https://doi.org/10.1029/2004GL020975>.
- Washington, R., R. James, H. Pearce, W. M. Pokam, and W. Moufouma-Okia, 2013: Congo Basin rainfall climatology: Can we believe the climate models? *Philos. Trans. Roy. Soc. London*, **368B**, 20120296, <https://doi.org/10.1098/rstb.2012.0296>.
- Watson, D. F., 1992: *Contouring: A Guide to the Analysis and Display of Spatial Data*. Pergamon Press, 340 pp.
- Xie, P., and P. A. Arkin, 1995: An intercomparison of gauge observations and satellite estimates of monthly precipitation. *J. Appl. Meteorol.*, **34**, 1143–1160, [https://doi.org/10.1175/1520-0450\(1995\)034<1143:AIOGOA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<1143:AIOGOA>2.0.CO;2).
- Zhang, Z., M. Mann, and E. Cook, 2004: Alternative methods of proxy-based climate field reconstruction: Application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. *Holocene*, **14**, 502–516, <https://doi.org/10.1191/0959683604hl727rp>.
- Zhou, L., and Coauthors, 2014: Widespread decline of Congo rainforest greenness in the past decade. *Nature*, **509**, 86–90, <https://doi.org/10.1038/nature13265>.