

Monitoring Drought Conditions and Their Uncertainties in Africa Using TRMM Data

G. NAUMANN, P. BARBOSA, H. CARRAO, A. SINGLETON, AND J. VOGT

Institute for Environment and Sustainability, Joint Research Centre, Ispra, Italy

(Manuscript received 3 May 2012, in final form 6 July 2012)

ABSTRACT

The main objective of this study is to evaluate the uncertainties due to sample size associated with the estimation of the standardized precipitation index (SPI) and their impact on the level of confidence in drought monitoring in Africa using high-spatial-resolution data from short time series. To do this, two different rainfall datasets, each available on a monthly basis, were analyzed over four river basins in Africa—Oum er-Rbia, Limpopo, Niger, and eastern Nile—as well as at the continental level. The two precipitation datasets used were the Tropical Rainfall Measuring Mission (TRMM) satellite monthly rainfall product 3B43 and the Global Precipitation Climatology Centre full-reanalysis gridded precipitation dataset. A nonparametric resampling bootstrap approach was used to compute the confidence bands associated with the SPI estimation, which are essential for making a qualified assessment of drought events. The comparative analysis of different datasets suggests that for reliable drought monitoring over Africa it is feasible to use short time series of remote sensing precipitation data, such as those from TRMM, that have a higher spatial resolution than other gridded precipitation data. The proposed approach for drought monitoring has the potential to be used in support of decision making at both continental and subcontinental scales over Africa or over other regions that have a sparse distribution of rainfall measurement instruments.

1. Introduction

Methods for drought assessment are based mainly on water supply indices derived from precipitation time series alone. Over Africa, the main limitation for developing effective real-time drought monitoring and early warning systems is the lack of reliable and up-to-date precipitation data in many regions of the continent. A sparse distribution of rain gauges and short or incomplete historical rainfall records pose further problems. This lack of information can lead to significant errors in the estimation of statistical parameters for deriving water supply indices from the precipitation time series.

When the availability of in situ data is scarce, it is necessary to determine whether to use longer, but spatially sparse, time series or to use shorter time series with higher spatial resolution. Prior studies suggest that

within-station substitution with a moderate length history (about 10 yr) performs better than spatial interpolation of long time series for representing the spatial-temporal variability of large-scale climatological conditions, such as time-averaged precipitation (Willmott et al. 1996). In this way, Rhee and Carbone (2011) studied the effect of drought estimation with limited precipitation data across different climatic regions in the United States. They show that the standardized precipitation index (SPI) values that are based on short-term records generally produced smaller cross-validation mean absolute errors values than the spatially interpolated SPI values when the lengths of records were equal to or longer than 10 yr for all SPI time scales. This relation was high even when the lengths of records were only 5 yr. These results using short-term records show that including as many stations with moderate lengths of records (at least 10 yr) as possible can improve the representation of spatial-temporal variability of drought. The authors also perform an analysis using 5.5 yr of the Tropical Rainfall Measuring Mission (TRMM) 3B43 record. They stated that the TRMM data could not outperform the spatially interpolated daily precipitation in most regions. In some regions (such as mountainous regions or those without in situ measurements), however, the use

Corresponding author address: Gustavo Naumann, Climate Risk Management Unit, Institute for Environment and Sustainability, Joint Research Centre - European Commission, Via E. Fermi, 2749 - TP 280, I-21027 Ispra (VA), Italy.
E-mail: gustavo.naumann@jrc.ec.europa.eu

of remote sensing–derived data (even with 5.5-yr record length) could outperform spatial interpolation of long-term gauge data.

The TRMM satellite has proved to be useful for precipitation monitoring in regions, such as areas of central Africa, for which station data are difficult to obtain or in which there is poor station coverage (Jenkins 2000). In fact, TRMM precipitation products have been extensively validated at ground sites around the world, some of these in Africa. Nicholson et al. (2003) show that TRMM estimations are in excellent agreement with gauge data over West Africa on monthly to seasonal time scales, with a root-mean-square (RMS) error of around 1 mm day^{-1} at monthly resolution.

Adeyewa and Nakamura (2003) conducted a 36-month climatological assessment of the TRMM precipitation radar, TRMM 3B43, and Global Precipitation Climatology Project satellite products over the major climatic regions in Africa. The study shows that 3B43 closely matches rain gauge data, suggesting that the goal of the algorithm was largely achieved. Dinku et al. (2007) studied a relatively dense station network over the Ethiopian highlands to evaluate the performance of different satellite products over complex topography in the tropics. The authors found that, for those evaluated at a monthly time scale, Climate Prediction Center (CPC) Merged Analysis of Precipitation and TRMM 3B43 performed very well, with a bias of less than 10% and an RMS of about 25%. The TRMM Multisatellite Precipitation Analysis (TMPA) estimation provides reasonable performance at monthly scales but has lower skill in correctly specifying moderate- and light-event amounts at short time intervals, in common with other finescale estimators (Huffman et al. 2007).

As a step toward the implementation of a cost-effective drought early-warning system in many undergaged regions around Africa, this paper describes and tests an approach of unbiased SPI estimation using TRMM 3B43 data over Africa. To quantify bias and uncertainties, a nonparametric bootstrap approach was performed.

2. Data and methods

a. Precipitation datasets

This work uses the TMPA estimation (computed at monthly intervals as the TRMM 3B43 dataset for the period 1998–2010) that combines the estimates generated by TRMM and other satellite products (3B42) with the Climate Anomaly Monitoring System gridded rain gauge data produced by the National Oceanic and Atmospheric Administration CPC and/or the global rain gauge product produced by the Global Precipitation

TABLE 1. Definition of African regions by latitude and longitude, basins totally or partially included, and TRMM ($0.25^\circ \times 0.25^\circ$) and GPCC total number of grid points ($1^\circ \times 1^\circ$). For GPCC, the percentage of stations per grid and the percentage of pixels without stations are shown in parentheses.

Area	Lat and lon boundaries	Basin	TRMM grid cells	GPCC grid cells
A	31°–35°N, 10°W–0°	Oum er-Rbia	640	36 (0.52, 0.65)
B	6°–18°N, 10°W–0°	Niger	1920	120 (0.23, .70)
C	7°–17°N, 30°–40°E	Blue Nile	1600	100 (0.23, 0.75)
D	26°–20°S, 25°–34°E	Limpopo	1224	54 (0.56, 0.44)

Climatology Centre (GPCC). The output is rainfall for $0.25^\circ \times 0.25^\circ$ grid boxes for each month.

The reference dataset that was used is version 5 of the GPCC full reanalysis (Rudolf et al. 1994) for the period 1951–2009. This dataset is based on quality-controlled precipitation observations from a large number of stations (up to 43 000 globally) with irregular coverage in time. These datasets were evaluated over four river basins in Africa—Oum-er-Rbia (A), Niger (B), eastern Nile (C), and Limpopo (D)—as well as at the continental scale. A short description of these regions is shown in Table 1 and Fig. 1.

The TRMM product and the reference data are not completely independent, however, although TRMM is mainly based on remote sensing data. Figure 1 shows the mean and standard deviation of annual precipitation for the TRMM and GPCC datasets over Africa. An overall agreement is shown between datasets with respect to average and interannual variability as well as the mean spatial patterns of annual precipitation. These datasets agree on the north–south gradient from the desert areas to the tropical savannas and rain forests related to a precipitation maximum due to the location of the inter-tropical convergence zone (ITCZ). Although this agreement exists, the record length of the two datasets is different.

b. Standardized precipitation index

The SPI was developed by McKee et al. (1993, 1995) to provide a spatially and temporally invariant measure of the precipitation deficit (or surplus) for any accumulation time scale. It is computed by fitting a parametric cumulative distribution function (CDF) to a homogenized precipitation time series and applying an equiprobability transformation to the standard normal variable. This gives the SPI in units of number of standard deviations from the median.

The gamma distribution is typically the parametric CDF chosen to represent the precipitation time series (e.g., McKee et al. 1993, 1995; Lloyd-Hughes and Saunders 2002; Husak et al. 2007), since it has the

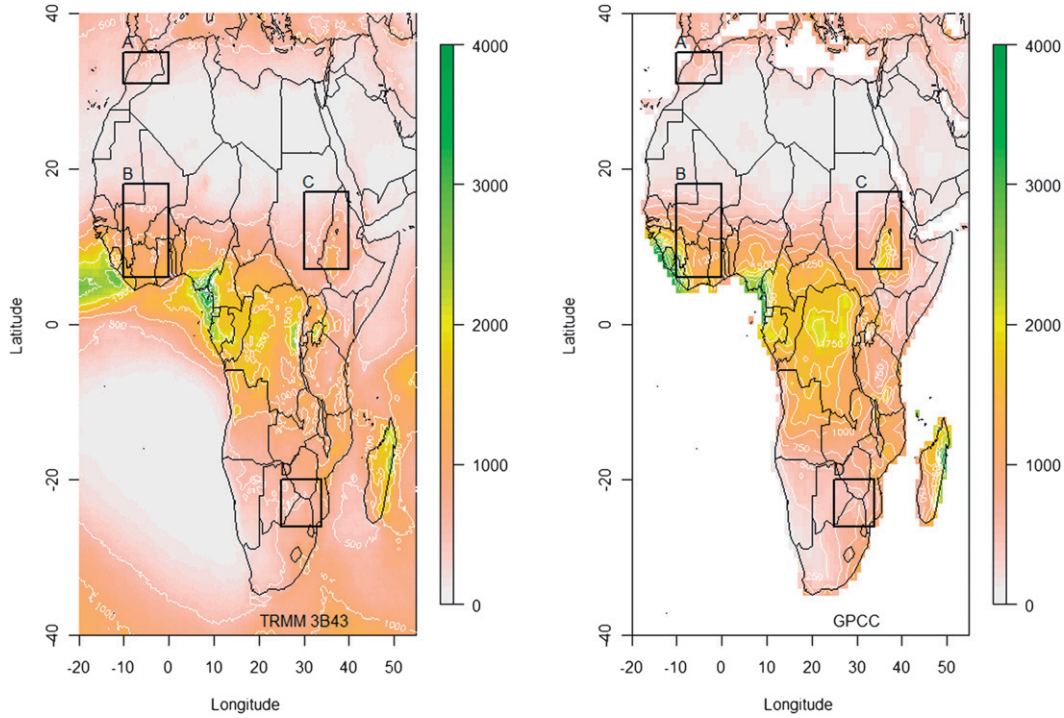


FIG. 1. Annual mean precipitation (mm) for the (left) TRMM 3B43 and (right) GPCC datasets. The labeled rectangles show the African regions used in this analysis: Oum-er-Rbia (A), Inner Niger Delta (B), Blue Nile (C), and Limpopo basin (D).

advantage of being bounded on the left at zero and positively skewed (Thom 1958; Wilks 2002). Moreover, Husak et al. (2007) have shown that the gamma distribution adequately models precipitation time series in roughly 98% of locations over Africa.

In this study we use the maximum-likelihood estimation method to estimate the parameters of the gamma distribution. To show the effect of period length on the estimation of the gamma distribution, the shape and scale parameters for each month using 15 (1994–2009), 30 (1979–2009), and 60 (1949–2009) yr of GPCC data were calculated for the Oum-er-Rbia basin (Fig. 2).

c. Nonparametric bootstrap method

Here we use a nonparametric bootstrap method to estimate the parameters and confidence intervals of the gamma distribution. This is done for aggregated precipitation for the 3-month time scale over the 13-yr time series of the TRMM 3B43 dataset from 1998 to 2010.

Consider that a random sample of observations, $X = \{X_1, X_2, \dots, X_n\}$, is used to obtain a sample estimate θ_s of a parameter of interest θ that can be the shape or scale parameter that defines the gamma distribution for X . The purpose of bootstrap simulation is to estimate uncertainty (bias and variance) associated with the sample estimate θ_s . According to Efron and Tibshirani (1993),

a random sample size of size n is drawn with replacement from the original sample.

Using the k th bootstrap sample of b bootstrap simulations and denoting by

$$X(k) = \{X_1^*, X_2^*, \dots, X_N^*\}, \quad k = 1, 2, \dots, b, \tag{1}$$

a new bootstrap estimate θ_k^* of θ_s can be obtained. The set of $\theta^* = \{\theta_1^*, \theta_2^*, \dots, \theta_b^*\}$ constitutes the sampling distribution of θ_s . The bootstrap estimate of the bias is given as

$$\text{bias} = (\theta_m^* - \theta_s), \tag{2}$$

where θ_m^* represents the average of all bootstrap estimates θ^* . This leads to the bias-corrected estimator of the parameter θ :

$$\theta = \theta_s - \text{bias} = 2\theta_s - \theta_m^*. \tag{3}$$

The variance S of θ_s is estimated as

$$S(\theta_s) = \frac{1}{b-1} \sum_{k=1}^b (\theta_k^* - \theta_m^*)^2. \tag{4}$$

This is a statistical resampling technique that does not require preselected distribution or models to fit the data,

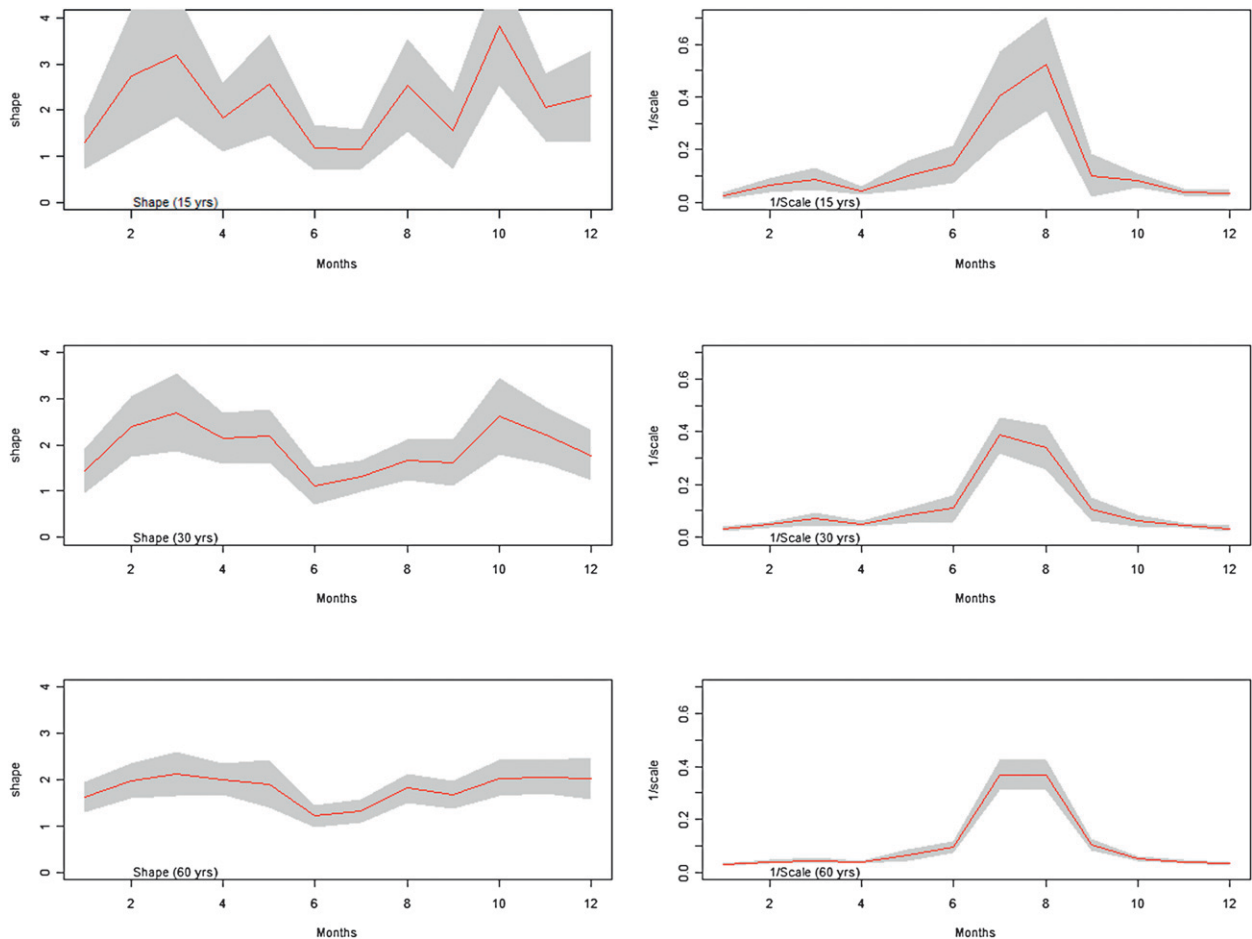


FIG. 2. Annual cycle of shape and scale gamma parameters of GPCP precipitation at region A (Oum er-Rbia) for different record lengths (red) and its uncertainties (light gray).

because the simulations have the same distributional properties of the original dataset. On the other hand, this behavior may produce spurious fine structure in the original data (Silverman 1986).

3. Results

In Fig. 2 it is shown that the confidence interval decreases as the record length of data increases, meaning that the stability of the coefficients that can be estimated with reasonable accuracy increases as the sample size increases. Stability is usually understood to mean that the fitted distribution (i.e., parameters) is also applicable to independent future data so that the parameters would be substantially unchanged if based on a different sample of the same kind of data. Furthermore, the error in the estimation of the scale parameter is greater for dry months (June–August). This result suggests that the error in the estimation is not constant during the year

and also that it changes for each climate region, being greater for dry climates.

This lack of stability for the shorter time series could produce greater uncertainties when the drought indices are calculated. Figure 3 shows the empirical distribution of the 3-month TRMM averaged precipitation over the Limpopo basin for February. This distribution was fitted using kernel, gamma, and unbiased gamma distributions using the bootstrap technique [Eq. (3)]. The unbiased estimation fits best (i.e., lower Kolmogorov–Smirnov distance: 0.15 for unbiased, 0.18 for kernel, and 0.23 for uncorrected estimation) when compared with the other approaches. Figure 3 also shows the distribution estimation and the family of distributions associated with the bootstrap resampling. It is shown that the members could vary widely, but the mode is in general well represented by the majority of members.

At the pan-African level, Fig. 4 shows the spatial distribution of the unbiased shape and scale parameters using 3-monthly averaged TRMM precipitation data.

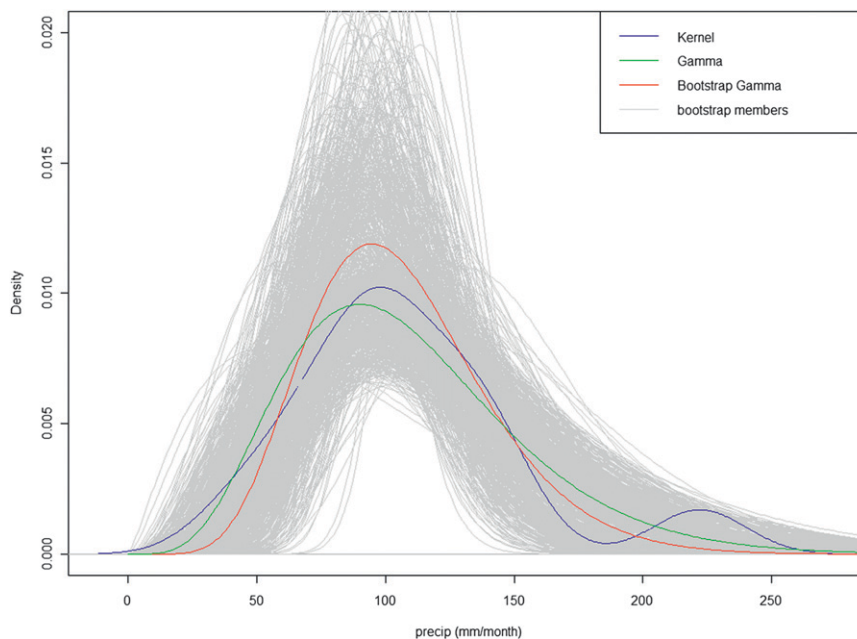


FIG. 3. Gamma and kernel probability distribution functions of TRMM precipitation at Limpopo basin. Blue: kernel distribution, green: gamma distribution, red: bootstrap gamma distribution, and gray: gamma distribution of each bootstrap member.

This estimation shows a regional consistency and is in agreement with the findings of Husak et al. (2007). Large shape values tend to follow the ITCZ through each of the monthly maps. Regions with higher shape values indicate that the probability of events drier than average becomes similar to the probability of events wetter than average, since the distribution is more symmetrical (Wilks 2002). The maxima are located near the coasts of Gabon, the Republic of the Congo, and the Democratic Republic of the Congo in January, whereas in July they are located between 0° and 10°N mainly in western and central Africa. The wet seasons in these regions are driven by the ITCZ, accompanied by large and consistent rainfall in the observations. The areas with a higher scale parameter are mainly observed in the poleward borders of the ITCZ where the rainfall could be more variable. This larger variability is mainly due to fluctuations in the position of the ITCZ across the continent and may give one region heavy rainfall for an extended period of time while causing another to have an abbreviated rainy season.

Figure 5 shows the 3-month SPI estimated with unbiased CDF estimation using a 13-yr baseline for TRMM (including confidence intervals) and a 60-yr baseline for GPCC over the four regions. It is shown that in general TRMM exceeds the peaks when compared with GPCC. This means that the main source of error is due to the estimation of the tails of the distribution, and this is

related to the stability of the time series. The GPCC estimation is within the TRMM estimation confidence intervals, however. This is confirmed by the correlation coefficient r that is shown in Fig. 5, which is statistically significant for all river basins. The lack of agreement is observed only for a few specific periods in the Blue Nile and to a minor extent in the Niger basin. Examples are the years of 2001, 2002, 2004, 2006, and 2008 in the Blue Nile basin, and 2000 and 2007 in the Niger basin. The largest differences are observed for the regions with the most complex orography or a lack of in situ information: the Blue Nile basin and the Niger basin, respectively. In addition, as shown in Table 1, these regions also have the lowest station density per grid, with up to 75% of pixels without any ground observation for the GPCC dataset. This means that more in situ data are needed to improve the GPCC precipitation datasets and the TRMM calibration as well.

4. Conclusions

The comparative analysis between the TRMM and GPCC datasets suggests that for reliable drought monitoring over Africa it is feasible to use TRMM time series that have a higher spatial resolution than other gridded datasets like GPCC. Higher discrepancies in SPI estimations are shown in mountainous areas and areas with sparse in situ station density.

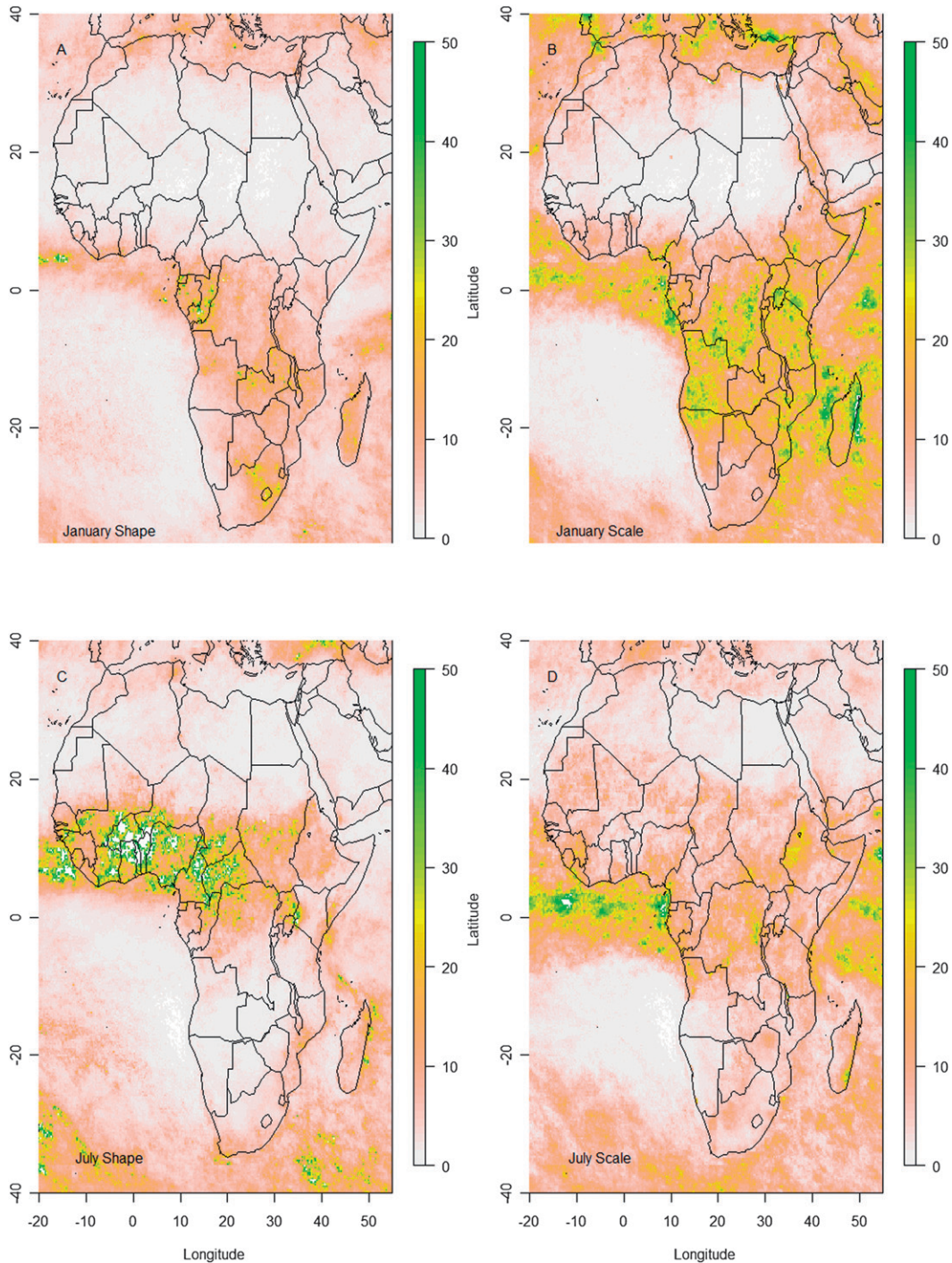


FIG. 4. (left) Shape and (right) scale parameters for (top) January and (bottom) July using unbiased estimation of parameters.

A nonparametric resampling bootstrap approach was used to compute the confidence bands of the sampling uncertainties associated with the SPI estimation. The proposed approach for drought monitoring has the potential to be used in support of decision making at

continental and subcontinental scales over Africa or other regions that have a sparse distribution of rainfall measurement instruments.

This kind of approach could be used to improve the monitoring of rainfall conditions in two ways. The first

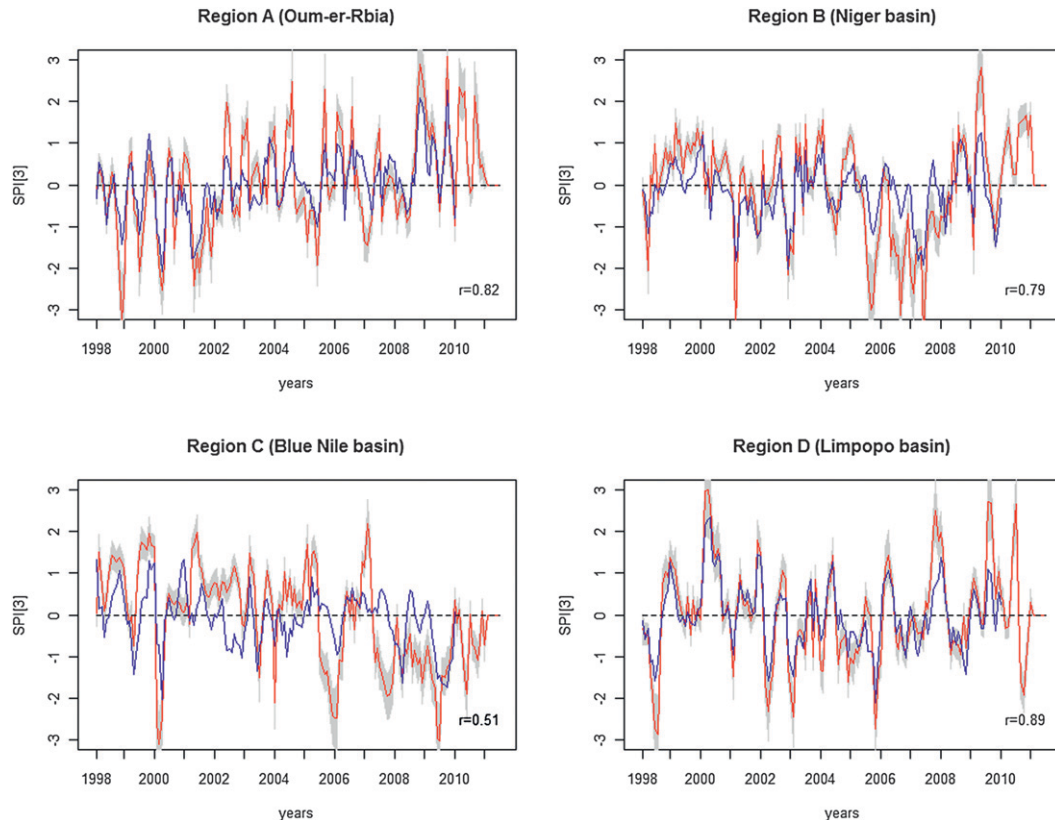


FIG. 5. Three-month SPI time series calculated using 13-yr TRMM (red) and 60-yr GPCP (blue) datasets; also given is the TRMM confidence interval (light gray).

way is to obtain an unbiased estimation of the gamma parameters with short precipitation time series. This could allow the development of a pan-African SPI using near-real-time rainfall estimates. The second is to estimate the confidence bands of SPI. This can prepare decision makers, through the measurement of uncertainties associated with the datasets, to better understand in which situations this tool is more reliable than others. Moreover, this type of approach could enable some forecast applications. For instance, it is possible to use the distribution information for each member of the bootstrap as initial conditions to develop drought scenarios. These types of scenarios could prepare decision makers and local stakeholders to take the appropriate action needed in the cases of high- and low-risk situations.

Acknowledgments. This work was funded by the European Commission Seventh Framework Programme (EU FP7) in the framework of the Improved Drought Early Warning and Forecasting to Strengthen Preparedness and Adaptation to Droughts in Africa (DEWFORA) project under Grant Agreement 265454.

REFERENCES

- Adeyewa, Z. D., and K. Nakamura, 2003: Validation of TRMM radar rainfall data over major climatic regions in Africa. *J. Appl. Meteor.*, **42**, 331–347.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski, 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, **28**, 1503–1526.
- Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- 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.
- Husak, G. J., J. Michaelsen, and C. Funk, 2007: Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int. J. Climatol.*, **27**, 935–944, doi:10.1002/joc.1441.
- Jenkins, G. S., 2000: TRMM satellite estimates of convective processes in central Africa during September, October, November 1998: Implications for elevated Atlantic tropospheric ozone. *Geophys. Res. Lett.*, **27**, 1711–1714.
- Lloyd-Hughes, B., and M. A. Saunders, 2002: A drought climatology for Europe. *Int. J. Climatol.*, **22**, 1571–1592.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. Preprints, *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.

- , —, and —, 1995: Drought monitoring with multiple time scales. *Proc. Ninth Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 233–236.
- Nicholson, S. E., and Coauthors, 2003: Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part II: Validation of TRMM rainfall products. *J. Appl. Meteor.*, **42**, 1355–1367.
- Rhee, J., and G. J. Carbone, 2011: Estimating drought conditions for regions with limited precipitation data. *J. Appl. Meteor. Climatol.*, **50**, 548–559.
- Rudolf, B., W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. *Global Precipitation and Climate Change*, M. Desbois and F. Désalmand, Eds., Springer, 173–186.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 175 pp.
- Thom, H. C., 1958: A note on the gamma distribution. *Mon. Wea. Rev.*, **86**, 117–122.
- Wilks, D. S., 2002: *Statistical Methods in the Atmospheric Sciences*. Elsevier Academic Press, 467 pp.
- Willmott, C. J., S. M. Robeson, and M. J. Janis, 1996: Comparison of approaches for estimating time-averaged precipitation using data from the USA. *Int. J. Climatol.*, **16**, 1103–1115.