Dear Editor,

We thank you for your constructive criticism and comments. We acknowledge that comparing PET with the other drought indicators is not giving an additional value to the manuscript.

According that we've revised the paper by excluding all the comparison with PET. In order to do that we've changed the text accordingly (highlighted with track changes) and figures 5, 7 and 8 were modified.

Best regards,

Gustavo Naumann.

Comparison of drought indicators derived from multiple datasets over Africa

Gustavo Naumann[1], Emanuel Dutra[2], Paulo Barbosa[1], Florian Pappenberger[2], Fredrik Wetterhall[2] and Jürgen Vogt[1].

[1]{European Commission, Joint Research Centre, Ispra, Italy}

[2]{European Centre for Medium Range Weather Forecasts, Reading, United Kingdom}

Correspondence to: G. Naumann (gustavo.naumann@jrc.ec.europa.eu)

Abstract

Drought monitoring is a key component to mitigate impacts of droughts. Lack of reliable and up-to-date precipitation datasets is a common challenge across the Globe. This study investigates different datasets and drought indicators on their capability to improve drought monitoring in Africa. The study was performed for four river basins located in different climatic regions (the Oum er-Rbia in Morocco, the Blue Nile in Eastern Africa, the Upper Niger in Western Africa, and the Limpopo in South-Eastern Africa) as well as the Greater Horn of Africa.

The five precipitation datasets compared are the *ECMWF ERA – Interim reanalysis*, the *Tropical Rainfall Measuring Mission* satellite monthly rainfall product 3B-43, the *Global Precipitation Climatology Centre* gridded precipitation dataset, the *Global Precipitation Climatology Project* Global Monthly Merged Precipitation Analyses, and the *Climate Prediction Center Merged Analysis of Precipitation*. The set of drought indicators used includes the Standardized Precipitation Index, the Standardized Precipitation-Evaporation Index, and Soil Moisture Anomalies. ~~Potential Evapotranspiration was also compared with the different drought indicators since it's a key process in the water cycle that can influence drought conditions.~~

A comparison of the annual cycle and monthly precipitation time series shows a good agreement in the timing of the rainy seasons. The main differences between the datasets are in the ability to represent the magnitude of the wet seasons and extremes. Moreover, for the areas affected by drought, all the drought indicators agree on the time of drought onset and

recovery although there is disagreement on the extent of the affected area. In regions with limited rain gauge data the estimation of the different drought indicators is characterised by a higher uncertainty. Further comparison suggests that the main source of differences in the computation of the drought indicators is the uncertainty in the precipitation datasets rather than the estimation of the distribution parameters of the drought indicators.

1 Introduction

Assessment of drought impacts requires understanding of regional historical droughts as well as the behaviours on human activities during their occurrences. Traditional methods for drought assessment are mainly based on water supply indices derived from precipitation time-series alone (Heim 2002). A sparse distribution of rain gauges and short or incomplete historical rainfall records may, however, lead to significant errors in the estimation of water supply indices derived from precipitation time-series.

As a consequence of drought, many countries in Africa have seen recurrent famines that affected millions of people (Rojas et al., 2011). Since precipitation is fundamental for rain-fed crops in these drought-prone regions, improvements in drought monitoring and early warning will improve our capacity to detect, anticipate, and mitigate famine (Wilhite et al, 2000, Rowland et al., 2005). However, the lack of reliable and up-to-date climatological data in many regions of Africa hinders the development of effective real-time drought monitoring and early warning systems.

Recently, several rain gauge and remote sensing based estimations of precipitation became available, which exhibit discrepancies and limitations in representing rainfall at local and regional scale. This has been highlighted for daily and monthly precipitation datasets by Dinku et al (2007; 2008) and Hirpa et al (2010). The authors studied a relatively dense station network over the Ethiopian highlands and found that at a monthly time scale and a spatial resolution of 2.5º CMAP and TRMM 3B-43 performed very well with a bias of less than 10% and a root mean square error of about 25%. Thiemig et al. (2012; 2013) found that the Rainfall Estimation Algorithm and TRMM 3B-42 showed a high potential in reproducing the interannual variability, the spatial and quantitative distribution and the timing of rainfall events.

Liebmann et al., 2012, studied the spatial variations in the annual cycle comparing GPCP with TRMM and gauge-based Famine Early Warning System datasets. They found that GPCP estimates are generally higher than TRMM in the wettest parts of Africa, but the timing of the annual cycle and onset dates are consistent. Dutra et al., 2013a, found significant differences (mainly in the equatorial area) in the quality of the precipitation between the ERA-Interim, GPCP and the Climate Anomaly Monitoring System – Outgoing Longwave Radiation Precipitation Index (CAMS-OPI) datasets for different river basins in Africa. From these studies it is evident that the question on which dataset best represents African precipitation is still not sufficiently answered.

The difficulty in establishing a "ground truth" of precipitation in Africa also affects the uncertainty in the calculation of derivatives of precipitation, like drought indicators, since the relationship between the quality of a precipitation product and any drought indicator is nonlinear. This means that errors in the precipitation can be amplified or dampened when a drought index is computed. Previous works have reviewed and compared several drought indicators (Heim 2002; Anderson et al 2011, Shukla et al 2011; Vicente-Serrano et al 2012). However, an agreement between different indicators is not necessarily observed as the capability to detect droughts changes between indicator, system and region.

The main goal of this study was to identify the main sources of uncertainty in the computation of the drought indicators. Furthermore, an assessment was done on the ability of the different datasets and drought indicators (SPI, SPEI~~, PET~~ and SMA) to represent the spatio-temporal features of droughts in different climate regimes across Africa.

2 Data and Methods

2.1 Study area

The analysis was performed at continental level over Africa with particular focus on the areas falling in four river basins (Oum er-Rbia, Limpopo, Niger, and Eastern Nile) as well as the Greater Horn of Africa (GHA). The regions were defined as the land areas inside each bounding box (see Figure 1). The area and geographical extent of the study areas are provided in Table 1. The regional study areas selected cover a range of climates and socio-economic systems in Africa.

2.2 Precipitation Data

The five precipitation datasets used were the ECMWF ERA-INTERIM (ERA-I) Reanalysis (approximately 0.7°x0.7°, bilinear interpolation to 0.5°x0.5°), Tropical Rainfall Measuring Mission (TRMM) satellite monthly rainfall product 3B-43 (0.25°x0.25°), the Global Precipitation Climatology Centre (GPCC) gridded precipitation dataset V.5 (0.5°x0.5°), the Global Precipitation Climatology Project (GPCP) Global Monthly Merged Precipitation Analyses (2.5°x2.5°) and the CPC Merged Analysis of Precipitation (CMAP, 2.5°x2.5°) (Table 2).

This work uses the TRMM Multisatellite Precipitation Analysis estimation computed at monthly intervals as TRMM 3B-43 dataset for the period 1998-2010 (Huffman et al., 2007). This product combines the estimates generated by the TRMM and other satellite products (3B-42) with the Climate Anomaly Monitoring System gridded rain gauge data and/or the GPCC global rain gauge data at 0.25°x0.25° resolution. The GPCC full reanalysis version 5 (Rudolf et al., 1994) was used for 1979 to 2010. 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.

The ECMWF ERA-I reanalysis, the latest global atmospheric reanalysis produced by ECMWF extends from 1 January 1979 to the present date. See Dee et al. (2011) for detailed descriptions of the atmospheric model used in ERA-I, the data assimilation system, the observations used, and various performance aspects. The ERA-I configuration has a spectral T255 horizontal resolution (about 0.7°x0.7° in the grid-point space) with 60 model vertical levels. For the present application, the monthly precipitation means were spatially interpolated (bilinear) to a regular 0.5°x0.5° grid. Three-hourly ERA-I precipitation estimates are produced by 12 h model integrations starting at 00UTC and 12UTC. These short-range forecasts are therefore mainly constrained by the analysis of upper-air observations of temperature and humidity, from satellites and in situ instruments.

The Global Precipitation Climatology Project (GPCP, Huffman et al., 2009) combines the precipitation information available from several sources such as the Special Sensor Microwave/Imager (SSM/I) data from the United States Defence Meteorological Satellite

Program satellites, infrared precipitation estimates computed primarily from geostationary satellites, low-Earth orbit estimates including the Atmospheric Infrared Sounder Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (TOVS), and Outgoing Longwave Radiation Precipitation Index data from the NOAA series satellites. The gauge data included are assembled and analyzed by the Global Precipitation Climatology Centre (GPCC). The latest version of GPCP v2.2 that was used is available since January 1979 to December 2010 in a regular 2.5°x2.5° grid.

The CPC Merged Analysis of Precipitation ("CMAP") is a technique which produces pentad and monthly analyses of global precipitation in which observations from rain gauges are merged with precipitation estimates from several satellite-based sensors (infrared and microwave). The analysis are on a 2.5 x 2.5 degree latitude/longitude grid and extend back to 1979. For further information refer to Xie and Arkin, (1997).

2.3 Drought indicators

The set of hydro-meteorological indicators analysed included the Standardized Precipitation Index (SPI), Standardized Precipitation-Evaporation Index (SPEI), and Soil Moisture Anomalies (SMA). ~~A discussion on the possible relation of Potential Evapotranspiration (PET) with the different indicators is also presented. Even if Evaporation~~PET ~~is not a drought indicator itself, it is one of the key processes in the water cycle and usually gives a more complete accounting of drought variability and~~that ~~can help to understand differences in drought indicators. However, the estimation of continental evapotranspiration fluxes is complex and can be a source of uncertainties when assessing drought conditions (Trambauer et al., 2014).~~ The SPI was computed with all the datasets (ERA-I, TRMM, and GPCP) since it only uses precipitation data. The SPEI was computed with precipitation and potential evapotranspiration from ERA-I, as well as with precipitation from GPCP and potential evapotranspiration from ERA-I. SMA and PET were directly obtained from the ERA-I reanalysis. The individual drought episodes from the time series of all indicators were determined by considering different thresholds of the standardized indicators. The duration of each dry event was determined as the number of consecutive months with negative values of the drought indices ~~(positive for PET)~~ over the period 1998-2010. The monthly drought fractional area was computed for different thresholds but is only shown for the values below the -1.0 threshold.

2.3.1 Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (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 timescale (e.g. 3, 6, 12 months). It is computed by fitting a parametric Cumulative Distribution Function (CDF) to a homogenized precipitation time-series and applying an equi-probability transformation to the standard normal variable. This gives the SPI in units of number of standard deviations from the median.

Typically, the gamma distribution is 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 advantage of being bounded on the left at zero and positively skewed (Thom 1958; Wilks 2002). Moreover, Husak et al. (2007) and Naumann et al. (2012) have shown that the gamma distribution adequately models precipitation time-series in most of the locations over Africa. In this study we use the Maximum-Likelihood Estimation (MLE) method to estimate the parameters of the gamma distribution.

A persistent negative anomaly of precipitation is the primary driver of drought, resulting in a successive shortage of water for different natural and human needs. Since SPI values are given in units of standard deviation from the standardised mean, negative values correspond to drier periods than normal and positive values correspond to wetter periods than normal. The magnitude of the departure from the median is a probabilistic measure of the severity of a wet or dry event.

~~2.3.2~~ Standardized Precipitation-Evaporation Index (SPEI) ~~and Potential Evapotranspiration (PET)~~

The Standardized Precipitation Evapotranspiration Index (SPEI, Vicente-Serrano et al., 2010) is based on precipitation and temperature data, and it has the advantage of combining different time dimensions (like the SPI) with the capacity to include the effects of temperature variability on drought. The calculation combines a climatic water balance, the accumulation of a water deficit/surplus at different time scales, and an adjustment to a log-logistic probability distribution. SPEI is similar to SPI, but it includes the temperature impact via the potential evapotranspiration (PET) that is calculated following Thornthwaite (1948). In the current work, we used ERA-I 2-meter temperature to derive PET, and the multiscalar index is

calculated as P-PET over the different time-scales and normalized (like the SPI) using the log-logistic probability distribution.

~~2.3.3~~2.3.2 Soil Moisture Anomalies (SMA)

Soil moisture anomalies were derived from ERA-I simulations by removing the mean annual cycle. Further standardization could be achieved by fitting the soil moisture distribution to a probability distribution (similar to SPI or, SPEI) such as the Beta distribution (Sheffield et al., 2004) or just a simple z-score (Dutra et al., 2008). In the current work we compare the SMA z-score following the considerations depicted in Dutra et al., 2008. By normalizing the soil moisture with the z-score, a classification scheme is obtained that is similar and comparable to that of McKee et al. (1993) and Vicente Serrano et al. (2012).

2.4 Evaluation metrics

The precipitation datasets and drought indicators were assessed using different scores available in the hydroGOF R-Package (Zambrano-Bigarini, 2013): Spearman's correlation coefficient (r), Mean Absolute Difference (MAD), Percent Bias (PBIAS) between two products and the Index of Agreement (d). Details of the Evaluation scores are listed in the appendix.

A direct quantitative assessment at continental level is difficult due to the lack of an actual validation dataset that represents the ground truth with adequately high spatial or temporal resolution. The performance metrics (mean absolute difference, relative bias and index of agreement) were used to diagnose the relative reliability of each indicator over different drought properties. This analysis does not assume that a single dataset or indicator is better than the other but highlights their temporal and spatial coherency.

3 Results and discussion

3.1 Comparison of global precipitation datasets

The datasets analysed are based on in-situ data (GPCC), remote sensing estimations (TRMM, GPCP) and a global circulation model (ERA-I). The datasets are not completely independent. For example, TRMM and GPCP are mainly based on remote sensing data and GPCP uses

GPCC over land). Figure 2 shows the mean annual precipitation for the ERA-I, GPCC, GPCP, CMAP and TRMM datasets over Africa. There is an overall agreement between the datasets with respect to the mean as well as the general spatial patterns of annual precipitation. These datasets agree on the north-south gradient from the Sahara desert areas in the North to the tropical savannahs in the Sahel (an area centered at approximately 10°N spanning from the Atlantic Ocean in the west to the Red Sea in the east). The datasets also agree in the areas of maximum precipitation over the African rainforests related to the location of the Inter-tropical Convergence Zone (ITCZ), as well as in the drier climate of the south-western part of Africa. The main differences are observed in the tropical area and over un-gauged areas. In transition regions from the Sahel to the Sahara TRMM estimations can exceed GPCC more than twofold while TRMM is substantially lower than the other estimations along the southwestern coast of West Africa (Liebmann et al., 2012). There is also a tendency of higher precipitation in the tropical rainforest in GPCP (Liebmann et al., 2012) and ERA-I (Dutra et al., 2013a, b) compared with the other datasets. ERA-I overestimates the rainfall in the central African region which is likely to be associated with a substantial warm bias in the model due to an underestimation of aerosol optical depth in the region (Dee et al., 2011).

For all the datasets and regions analysed the mean annual cycle of precipitation shows good agreement with respect to the onset and end of the rainy season. This is true even for the GHA region which is characterized by two rainy seasons (Figure 3). However, with respect to intensity the results are more heterogeneous. Although in the Limpopo and Oum er-Rbia basins there is a good agreement between the datasets, for the basins located in the tropical band the discrepancies are higher with an overestimation of ERA-I in the Eastern Nile Basin and GHA and an underestimation in the Niger basin.

Apparently the density of rain gauges plays a role in determining the agreement between datasets. The best gauged regions (Oum er-Rbia and Limpopo; Table 1) are those with the lowest dispersion in terms of annual cycle. These two regions (Oum er-Rbia and Limpopo) are located outside the tropical region, and their precipitation variability is mainly controlled by large-scale synoptic weather systems, while in the tropical region small-scale convective events play an important role. In these regions, model uncertainties (for example land-atmosphere coupling), uncertainties in satellite retrievals as well as poor gauge cover contribute to the large spread in the mean annual cycles computed with the different datasets.

The monthly datasets show a reasonable agreement over all regions in terms of the correlation coefficients which are usually greater than 0.8 (Table 3). The CMAP dataset deviates with values below 0.7 in some regions. Oum er-Rbia and Limpopo areas show the best agreement between datasets with MAE values below 10 mm/month. The bias in those two regions is below 20 % in all the cases except when TRMM and CMAP are compared (30%).

The biggest differences were observed for ERA-I in the Blue Nile and GHA regions. In these regions the overestimation of monthly precipitation reached 40 mm/month and the bias can reach 90% in the Blue Nile and around 50% in the GHA. These discrepancies are mainly due to the problems of representation of the mean annual cycle of precipitation by atmospherics models and the lack of in-situ observations (Mariotti et al., 2011, Dutra et al., 2013b).

3.2 Comparison of drought indicators

The monthly patterns of drought over Africa for January 2000, 2003, 2006 and 2009 show that dry areas (indicators with negative values) are generally depicted in more than one indicator, but their consistency varies with the drought characteristics, as well as the spatial and temporal coverage (Figure 4). Although there is in general a good spatial correspondence between all the indicators over the study period, there are also areas where there is no agreement between some indicators, such as in Central Africa between SPI and SPEI.

Figure 5 shows the index of agreement (d) between all the drought indicators computed with ERA-I. Overall, the index of agreement shows that there is a good correspondence between indicators in all regions with mean d values greater than 0.5 for almost all the comparisons. ~~PET seems to be uncoupled with the other indicators with~~ showing low values of d. ~~Maybe using different PET products could improve the index of agreement, since~~ This may reflect ~~the high disparities between different PET products that still is observed in the continethave been reported for Africa (Trambauer et al., 2014).~~ However, the effect on the computations of the SPEI is not major, since the agreement of this indicator with the others is still high. Only the inner Niger Delta is characterized by a weaker agreement, where d is often below 0.5.

Figure 6 shows the evolution of drought areas in 2000, 2003, 2006 and 2009 characterized by the number of indicators below a certain threshold. In almost all areas there is a good agreement, with usually more than 3 indicators reporting drought conditions per grid cell. However, there are some areas with only one indicator below the defined threshold, mostly

over Central Africa. There is scope to take advantage of these discrepancies and agreements and propose the construction of a composite indicator (Svoboda et al., 2002; Sepulcre-Canto et al., 2012; Hao and AghaKouchak, 2013). The development of a single composite drought indicator could improve the detection of the onset of a drought and help to monitor its evolution more efficiently, at the same time providing information on the uncertainty in the data. This will allow decision makers and stakeholders to better handle uncertainties in early warning systems.

The individual drought episodes were computed from the time series of all indicators considering as dry periods all values of standardized indicators below zero. The duration of each dry event was determined as the number of consecutive months with negative values ~~(positive for PET)~~ for the period 1998-2010. The average duration of dry episodes lasted between 2 to 6 months for all indicators, with the largest differences in duration for different indicators being found in the Niger basin and in the GHA (Figure 7). Overall, dry periods measured with SPEI tend to be 1 or 2 months more persistent if compared with the other estimations~~, while PET is the variable with less memory~~.

Figure 8 shows the monthly fractional area under standardised values below -1.0. For the areas that are under drought, all the datasets agree with the time of onset and recovery but there are sometimes disagreements on the area affected and this disagreement tends to be dependent on the threshold selected. In general there is a better agreement if the areas covered by any standardised indicator below -1.0 are considered. In this analysis the Niger basin and Greater Horn of Africa present more discrepancies reaching a difference of more than 50% between SPI and SPEI estimations during the 2009/2010 and 2005/2006 periods respectively. The soil moisture anomalies tend to define less generalised droughts as it is hard to reach half the region under dry conditions. However, even if the magnitude of the area is smaller with respect to the other indicators, the soil moisture shows a good correspondence except for the period 2000/2002 in the inner Niger delta.

In order to define how the selected threshold could affect the agreement between datasets a correlation analysis was performed between different thresholds of SPI and the areas affected by droughts in each region. Here the results of the different SPI estimations are presented, however similar results were found for the other indicators (not shown). For almost all regions (except for Oum er-Rbia where this relationship is almost constant) the correlation

between the different SPI's is higher for thresholds closer to zero (Figure 9). To consider a higher threshold (i.e. less negative) to define areas affected by drought (e.g. -0.8 or -1), therefore, will reduce the disagreement between indicators. However it puts a limit to the detection of the significance and severity of a drought. These results highlight that the main differences between the indicators appear in the extreme events.

Also, the bias between estimations indicates an acceptable departure between estimations from normal conditions until values near -0.5 (Figure 10). Below this threshold the bias increases exponentially surpassing quickly a bias of 100% around SPI values of -1. For Niger and GHA regions there is only a reasonable agreement between ERA-I and GPCC estimations.

Generally in the Oum er-Rbia and Limpopo basins, both extra-tropical regions, the agreement is high, possibly due to the greater number of in-situ observations and the importance of large- scale synoptic weather systems in these areas.

For the basins located between the tropics a greater disagreement is observed due to different factors. The main common factor is the remarkable absence of observations to calibrate and test the datasets. These deficiencies are also more evident in complex mountainous areas such as the Eastern Nile basin. Furthermore, droughts in equatorial regions are mainly driven by the absence of convective events during the rainy season. These mesoscale dimension events are hard to be reproduced by models and even difficult to monitor in areas with scarce in-situ rain gauges.

For drier regions, such as the inner Niger delta and the GHA, the estimation of the distribution parameters needed for the computation of the standardized indicators can be biased (or lower bounded) by the large amount of zero or near null precipitation observations. As depicted in Wu et al. (2007), the estimation of the gamma probability density function and the limited sample size in dry areas reduce the confidence of the SPI values. In these cases, the SPI may never attain the necessary threshold and hence failing to detect some drought occurrences (e.g. SPI always above -1 in Niger and GHA). The discrepancies between indicators for lower thresholds over regions with limited rain gauge data is characterised by the uncertainties of extreme values. This suggests that the main sources of error are the uncertainties in the precipitation datasets that are propagated in the estimation of the distribution parameters of the drought indicators.

The above discussion underlines the fact that drought monitoring and assessment is a difficult task, not only due to the nature of the phenomenon, but also due to the limitations inherent in the availability of long-term and high quality datasets for extended regions. The meteorological datasets as well as the indicators and models used must be selected carefully and their limitations need to be taken into account. As a consequence no definite conclusion can be drawn for the use of a single dataset or indicator. Depending on the region to be studied, different combinations may have to be chosen.

Our results further underline the value of maintaining an operational monitoring network at country, continental or even global level since indirect observations have their intrinsic uncertainties linked to the availability and reliability of 'ground truth' for their calibration. Without proper calibration, model (ERA-I) or algorithm (TRMM) inherent errors can propagate resulting in large drought indicator uncertainties, bringing no added value with respect to using standard climatology.

The development of a combined indicator based on a probabilistic approach (e.g., Dutra et al., 2013c) could be useful as a monitoring product at continental level. However, at local scale the kind of indicator and the source of data must be chosen carefully taking into account their limitations.

4 Conclusions

This study evaluated the capabilities of different drought indicators, ~~(~~including SPI, SPEI~~,~~ and SMA~~, as well as PET estimations and SMA),~~ in detecting the timing and extension of drought across Africa, using five different precipitation datasets (TRMM, ERA-Interim, GPCC, GPCP and CMAP). The analysis was performed on a Pan-African scale and on a regional scale focused on four river basins and on the Greater Horn of Africa.

A comparison of the annual cycle and monthly precipitation time series shows a good agreement in the timing of the peaks, including the Greater Horn of Africa where there are two rainy seasons. The main differences are observed in the ability to represent the magnitude of the wet seasons.

The monthly mean precipitation datasets shows good agreement over all regions with the only exception of the CMAP dataset that shows a lower agreement. In the Oum er-Rbia and

Limpopo basins there is a good agreement between the datasets with mean absolute differences below 10 mm/month. The bias in those two regions is below 20 %. The worst performance of ERA-I was observed in the Blue Nile basin, overestimating the monthly precipitation up to 40 mm/month with a bias of up to 92% with respect to the other datasets. Also in the GHA region the bias is around 50% with an overestimation of up to 17 mm/month.

The comparative analysis between TRMM, ERA-I, GPCP and GPCC datasets suggests that it is feasible to use TRMM time series with high spatial resolution for reliable drought monitoring over parts of Africa. It is possible to take advantage of this dataset mainly at regional level due to its high spatial resolution. However, higher discrepancies in SPI estimations are shown in mountainous areas and areas with a sparse in situ station density. On the other hand, drought monitoring at continental level with ERA-I performs better outside the areas influenced by the ITCZ.

The comparison between drought indicators suggests that the main discrepancies are due to the uncertainties in the datasets (driven by a lack of ground information, uncertainties in the estimation algorithms or the parameterization of the convection) rather than to the estimation of the distribution parameters. This is why the SPI estimations for the Oum er-Rbia and Limpopo regions exhibit a better agreement between estimations. While for the other regions the discrepancies between datasets are in many cases acceptable, greater discrepancies are observed for the inner Niger Delta when comparing ERA-I estimations with the other datasets.

Regarding the areas that are under drought, all the indicators agree with the time of onset and recovery but there are sometimes disagreements with respect to the area affected, and the level of disagreement tends to be dependent on the threshold selected.

It is proposed to integrate different indicators and accumulation periods in the form of a multivariate combined indicator in order to take advantage of their different drought properties. The probabilistic nature of such an approach would be very helpful for decision makers and for the combined analysis of multiple risks.

Acknowledgements

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. Gustavo Naumann thanks Mauricio Zambrano-Bigarini for the important discussions about different metrics to compare indicators and to make available the hydroGOF R-package.

Appendix A

The Spearman correlation represents the Pearson correlation coefficient computed using the ranks of the data. Conceptually, the Pearson correlation coefficient is applied to the ranks of the data rather than to the data values themselves. The Spearman coefficient is a more robust and resistant alternative to the Pearson product-moment correlation coefficient (Wilks, 2002). Computation of the Spearman rank correlation can be described as:

$$r = 1 - \frac{6 \sum R_i^2}{n(n^2 - 1)} \quad (1)$$

where R_i is the difference in ranks between the ith pair of data values. In cases of ties, where a particular data value appears more than once, all of these equal values are assigned their average rank before computing the Ri's.

The Mean Absolute Difference (MAD) measures the average magnitude of the differences in a set of different estimations of a certain indicator. It measures accuracy for continuous variables without considering the direction of the error. Also, this quantity is usually used to measure how close are two datasets or indicators as shown in equation 2

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |X1_i - X2_i| \quad (2)$$

where X1 and X are the values of precipitation or drought indicator of dataset 1 and n represents the number of pairs The percent bias (PBIAS) measures the average tendency of the values of a certain dataset to be larger or smaller than a reference one.

$$PBIAS = 100 \frac{\sum_{i=1}^{n} (X1_{i-} - X2_i)}{\sum_{i=1}^{n} X1_i} \quad (3)$$

The optimal value of PBIAS is 0, with low-magnitude values indicating accurate representation of drought indicators. Positive values indicate an overestimation bias, whereas negative values indicate an underestimation bias. It must be taken into account that this metric depends on which dataset is considered to represent the observations.

The Index of Agreement (d) developed by Willmott (1981) as a standardized measure of the degree of model prediction error varies between 0 and 1. A value of 1 indicates a perfect match, and 0 indicates no agreement at all (Willmott, 1981). The index of agreement can

detect additive and proportional differences in the observed and simulated means and variances; however, it is overly sensitive to extreme values due to the squared differences (Legates and McCabe, 1999).

$$d = 1 - \frac{\sum(X1 - X2)^2}{\sum(|X2 - \overline{X1}| + |X1 - \overline{X1}|)} \quad (4)$$

References

Anderson, M. C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J. R., and Kustas, W. P.: Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. Journal of Climate, 24(8), 2025-2044, 2011.

Dee, D., and Coauthors: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart. J. Roy. Meteor. Soc., 137, 553–597, 2011.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., and Ropelewski, C. F.: Validation of satellite rainfall products over East Africa's complex topography. International Journal of Remote Sensing, 28(7), 1503-1526, 2007.

Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S. J., & Ropelewski, C. F.: Validation of high-resolution satellite rainfall products over complex terrain. International Journal of Remote Sensing, 29(14), 4097-4110, 2008.

Dutra, E., Viterbo, P., and Miranda, P. M. A.: ERA-40 reanalysis hydrological applications in the characterization of regional drought, Geophys. Res. Lett., 35, L19402, doi:19410.11029/12008GL035381, 2008.

Dutra, E., Giuseppe, F. D., Wetterhall, F., & Pappenberger, F.: Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index. Hydrology and Earth System Sciences, 17(6), 2359-2373, 2013a.

Dutra, E., Magnusson, L., Wetterhall, F., Cloke, H. L., Balsamo, G., Boussetta, S., & Pappenberger, F.: The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. International Journal of Climatology, 33, 7, 1720–1729, 2013b.

Dutra, E., Wetterhall, F., Di Giuseppe, F., Naumann, G., Barbosa, P., Vogt J., Pozzi W., and Pappenberger F.: Global meteorological drought: Part I - probabilistic monitoring, to be submitted to HESS, 2013c.

Hao, Z., & AghaKouchak, A.: Multivariate Standardized Drought Index: A parametric multi-index model. Advances in Water Resources, 57, 12-18, 2013.

Heim Jr, R. R.: A review of twentieth-century drought indices used in the United States. Bulletin of the American Meteorological Society, 83(8), 1149-1165, 2002.

Hirpa, F. A., Gebremichael, M., and Hopson, T.: Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. Journal of Applied Meteorology and Climatology, 49(5), 1044-1051, 2010.

Huffman and Coauthors: The TRMM Multi-satellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined sensor precipitation estimates at fine scales. J. Hydrometeor., 8, 38–55, 2007.

Huffman, G.J, R.F. Adler, D.T. Bolvin, and Gu G.: Improving the Global Precipitation Record: GPCP Version 2.1. Geophys. Res. Lett., 36,L17808, doi:10.1029/2009GL040000, 2009.

Husak, G. J., Michaelsen, J., and Funk, C.: Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. International Journal of Climatology, 27(7), 935-944, 2007.

Legates, D. R., and McCabe Jr., G. J.: Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation, Water Resour. Res., 35(1), 233–241, 1999.

Liebmann, B., Bladé, I., Kiladis, G. N., Carvalho, L. M., B. Senay, G., Allured, D., Leroux, S., and Funk, C.: Seasonality of African precipitation from 1996 to 2009. Journal of Climate, 25(12), 4304-4322, 2012.

Lloyd-Hughes, B., and Saunders, M. A.: A drought climatology for Europe. Int. J. Climatol., 22, 1571–1592, 2002.

Mariotti, L., Coppola, E., Sylla, M. B., Giorgi, F., and Piani, C.: Regional climate model simulation of projected 21st century climate change over an all-Africa domain: Comparison analysis of nested and driving model results. J. Geophys. Res., 116, D15111, doi:10.1029/2010JD015068, 2011.

McKee T.B., Doesken N.J., and Kleist J.: The Relationship of Drought Frequency and Duration to Time Scales. Proc. 8th Conf. on Appl. Clim., 17-22 Jan. 1993, Anaheim, CA, 179-184, 1993.

McKee, T. B., Doesken, N. J., and Kleist, J.: Drought monitoring with multiple time scales. Proc. Ninth Conf. on Applied Climatology, Dallas, TX, Amer. Meteor. Soc. 233-236, 1995.

Naumann, G., Barbosa, P., Carrao, H., Singleton, A., and Vogt, J.: Monitoring Drought Conditions and Their Uncertainties in Africa Using TRMM Data. Journal of Applied Meteorology and Climatology, 51(10), 1867-1874, 2012.

Rojas, O., Vrieling, A., and Rembold, F.: Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sensing of Environment 115.2, 343-352, 2011.

Rowland, J., Verdin, J., Adoum, A., and Senay, G.: Drought monitoring techniques for famine early warning systems in Africa, in: Monitoring and Predicting Agricultural Drought: A Global Study, edited by: Boken, V. K., Cracknell, A. P., and Heathcote, R. L., Oxford University Press, New York, 252–265, 2005.

Rudolf, B.,W. Rueth, and Schneider, U.: Terrestrial precipitation analysis: Operational method and required density of point measurements. Global Precipitation and Climate Change, M. Desbois and F. Desahmond, Eds., Springler-Verlag, 173–186, 1994.

Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H., and Vogt, J.: Development of a Combined Drought Indicator to detect agricultural drought in Europe. Natural Hazards and Earth System Science, 12(11), 3519-3531, 2013.

Sheffield, J., Goteti, G., Wen, F. H., and Wood, E. F.: A simulated soil moisture based drought analysis for the United States, J. Geophys. Res., 109(D24), 2004.

Shukla, S., Steinemann, A. C., & Lettenmaier, D. P.: Drought monitoring for Washington State: Indicators and applications. Journal of Hydrometeorology, 12(1), 66-83, 2011.

Svoboda, M. D., and Coauthors: The drought monitor. Bull. Amer. Meteor. Soc., 83, 1181-1189, 2002.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and De Roo, A.: Validation of Satellite-Based Precipitation Products over Sparsely Gauged African River Basins. Journal of Hydrometeorology, 13(6), 1760-1783, 2012.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., and De Roo, A.: Hydrological Evaluation of Satellite-Based Rainfall Estimates over the Volta and Baro-Akobo Basin. Journal of Hydrology, 499, 324-338, 2013.

Thom, H. C.: A note on the gamma distribution. Monthly Weather Review, 86(4), 117-122, 1958.

Thornthwaite, C. W.: An approach toward a rational classification of climate. Geogr. Rev., 38, 55–94, 1948.

~~Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H., and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrol. Earth Syst. Sci., 18, 193-212, doi:10.5194/hess-18-193-2014, 2014.~~

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. Journal of Climate, 23(7), 1696-1718, 2010.

Vicente-Serrano, S. M., Beguería, S., Gimeno, L., Eklundh, L., Giuliani, G., Weston, D., El Kenawy, A., López-Moreno, J., Nieto, R., Ayenew T., Konte, D., Ardö, J., and Pegram, G. G.: Challenges for drought mitigation in Africa: The potential use of geospatial data and drought information systems. Applied Geography, 34, 471-486, 2012.

Wilhite, D. A., and Svoboda, M., D.: "Drought early warning systems in the context of drought preparedness and mitigation." Early warning systems for drought preparedness and drought management, 1-21, 2000.

Wilks D.S.: Statistical Methods in the Atmospheric Sciences. Elsevier Academic Press Publications, 467 pp., 2002.

Willmott, C. J.: On the validation of models. Physical Geography, 2, 184–194, 1981.

Wu, H., Svoboda, M. D., Hayes, M. J., Wilhite, D. A., and Wen, F.: Appropriate application of the standardized precipitation index in arid locations and dry seasons. International Journal of Climatology, 27(1), 65-79, 2007.

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

Zambrano-Bigiarini M.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-7. http://CRAN.R-project.org/package=hydroGOF, 2013.

Table 1. Geographical extent of the African regions and number of grid cells analysed for each dataset. For GPCC, the percentage of stations per grid and the percentage of pixels without stations are respectively shown between brackets.

Region	**Area (10^6xKm^2)**	**Longitude-Latitude**	**GPCC Grid cells**
A - Oum er-Rbia	0.49	[10ºW-0ºE]X[31ºN-35ºN]	36 (52, 65)
B - Niger	1.48	[10ºW-0ºE]X[6ºN-18ºN]	120 (23, 70)
C - Eastern Nile	1.23	[30ºE-40ºE]X[7ºN-17ºN]	100 (23, 75)
D -Limpopo	0.94	[25ºE-34ºE]X[26ºS-20ºS]	54 (56, 44)
E -GHA	2.22	[40ºE-52ºE]X[2ºS-12ºN]	180 (15, 85)

Table 2. Description of global datasets available in near-real time that could be used for monitoring precipitation conditions at continental level.

Datasets	*resolution*	*period*	*Source*	*Update*
ERA INTERIM	0.5ºx0.5º	1979-present	ECMWF Reanalysis	½ month
TRMM 3B-43 v.6	0.25ºx0.25	1998-present	Remote Sensing Estimate (combination 3B-42, CAMS and/or GPCC)	1 or 2 months
GPCC v.5	0.5ºx0.5º	1901-2010	In-situ data	1 month
GPCP v.2.2	2.5ºx2.5º	1979-2010	Remote Sensing Estimate(merged from microwave, infrared and sounder data and precipitation gauge analyses (GPCC).	irregular
CMAP	2.5ºx2.5º	1979-2009	Remote Sensing Estimate (GPI, OPI,S SM/I scattering, SSM/I emission and MSU + NCEP/NCAR Reanalysis)	irregular

Table 3. Correlation coefficient (r), Mean absolute difference (MAD) and percent bias (%) between the different precipitation datasets averaged over each region for the common period 1998-2010. All correlations are significant at 99%.

		TRMM			*GPCC*			*GPCP*			*CMAP*			*ERA-I*		
		r	**MAD**	**BIAS**	**r**	**MAD**	**BIAS**	**r**	**MAD**	**BIAS**	**r**	**MAD**	**BIAS**	**r**	**MAD**	**BIAS**
OER	**TRMM**	-	-	-	0.99	2.5	2.7	0.99	2.9	6.7	0.74	7.8	42.8	0.95	7.3	26.3
	GPCC	0.99	2.5	-2.6	-	-	-	0.99	2.5	4.2	0.94	4.7	23.1	0.95	6.7	24.4
	GPCP	0.99	2.9	-6.2	0.99	2.5	-4	-	-	-	0.73	6.5	33.9	0.95	5.7	18.4
	CMAP	0.74	7.8	-30	0.94	4.6	-18.7	0.73	6.5	-25.3	-	-	-	0.68	7.0	-11.6
	ERA-I	0.95	7.3	-20.8	0.95	6.6	-19.6	0.95	5.7	-15.5	0.68	7.0	13.1	-	-	-
NIG	**TRMM**	-	-	-	0.99	5.8	-1.9	0.98	13.6	-14.5	0.8	13.9	7.2	0.94	23.2	8
	GPCC	0.99	5.8	1.9	-	-	-	0.99	11.6	-14.1	0.97	6.9	-1	0.95	22.2	8.3
	GPCP	0.98	13.6	17	0.99	11.5	16.4	-	-	-	0.82	16.7	25.4	0.95	25.8	26.4
	CMAP	0.8	13.8	-6.7	0.97	6.9	1	0.82	16.8	-20.3	-	-	-	0.78	25.8	0.7
	ERA-I	0.94	23.1	-7.4	0.95	22.2	-7.7	0.95	25.8	-20.9	0.78	25.8	-0.7	-	-	-
ENL	**TRMM**	-	-	-	0.94	17.6	-23.7	0.93	17.4	-22.4	0.82	15.3	-0.6	0.93	43.9	-48.1
	GPCC	0.94	17.6	31	-	-	-	1	2.7	1.9	0.97	12.1	22.5	0.97	29.9	-32.3
	GPCP	0.93	17.4	28.9	1	2.66	-1.9	-	-	-	0.85	14.3	28.2	0.97	30.1	-33.1
	CMAP	0.82	15.3	0.6	0.97	12.1	-18.4	0.85	14.3	-22	-	-	-	0.86	43.4	-47.8
	ERA-I	0.93	43.9	92.8	0.97	29.9	47.6	0.97	30.1	49.5	0.86	43.4	91.7	-	-	-
LIM	**TRMM**	-	-	-	0.98	7.03	8.9	0.97	8.4	6.7	0.76	12.6	20.6	0.96	10.4	9
	GPCC	0.98	7.0	-8.2	-	-	-	0.99	5.1	-3.3	0.91	8.3	1.8	0.98	8.1	-1.5
	GPCP	0.97	8.4	-6.3	0.99	5.1	3.4	-	-	-	0.79	9.9	13	0.97	8.8	2.1
	CMAP	0.76	12.6	-17	0.91	8.3	-1.8	0.79	9.9	-11.5	-	-	-	0.79	12.8	-9.6
	ERA-I	0.96	10.4	-8.2	0.98	8.1	1.5	0.97	8.8	-2.1	0.79	12.8	10.6	-	-	-
GHA	**TRMM**	-	-	-	0.82	9.8	-4.2	0.88	6.6	1.7	0.72	9.2	11.2	0.84	17.8	-34
	GPCC	0.82	9.8	4.4	-	-	-	0.9	8.2	7.1	0.84	9.4	8.4	0.83	17.1	-30.9
	GPCP	0.88	6.6	-1.7	0.9	8.2	-6.6	-	-	-	0.7	9.6	9.3	0.92	16.4	-35.1
	CMAP	0.72	9.2	-10.1	0.84	9.4	-7.8	0.7	9.6	-8.5	-	-	-	0.61	22.7	-40.6
	ERA-I	0.84	17.8	51.5	0.83	17.1	44.7	0.92	16.4	54.1	0.61	22.7	68.4	-	-	-

Table 4. Spearman correlation coefficient (r), mean absolute difference (MAD) between the different SPI-3 estimations averaged over each region for the common period 1998-2010

		TRMM		*GPCC*		*GPCP*		*ERA-I*	
		r	**MAD**	**r**	**MAD**	**r**	**MAD**	**r**	**MAD**
Oum er-Rbia	**TRMM**	-	-	0.89	0.28	0.81	0.38	0.84	0.37
	GPCC	0.89	0.28	-	-	0.81	0.35	0.81	0.34
	GPCP	0.81	0.38	0.81	0.35	-	-	0.74	0.5
	ERA-I	0.84	0.37	0.81	0.34	0.74	0.5	-	-
Niger	**TRMM**	-	-	0.85	0.26	0.79	0.38	0.71	0.5
	GPCC	0.85	0.26	-	-	0.91	0.29	0.72	0.46
	GPCP	0.79	0.38	0.91	0.29	-	-	0.67	0.65
	ERA-I	0.71	0.5	0.72	0.46	0.67	0.65	-	-
Blue Nile	**TRMM**	-	-	0.54	0.54	0.53	0.55	0.6	0.5
	GPCC	0.54	0.54	-	-	0.92	0.27	0.57	0.41
	GPCP	0.53	0.55	0.92	0.27	-	-	0.67	0.46
	ERA-I	0.6	0.5	0.57	0.41	0.67	0.46	-	-
Limpopo	**TRMM**	-	-	0.91	0.28	0.84	0.39	0.8	0.46
	GPCC	0.91	0.28	-	-	0.92	0.27	0.91	0.33
	GPCP	0.84	0.39	0.92	0.27	-	-	0.88	0.35
	ERA-I	0.8	0.46	0.91	0.33	0.88	0.35	-	-
GHA	**TRMM**	-	-	0.58	0.4	0.65	0.44	0.61	0.44
	GPCC	0.58	0.4	-	-	0.86	0.29	0.58	0.42
	GPCP	0.65	0.44	0.86	0.29	-	-	0.68	0.45
	ERA-I	0.61	0.44	0.58	0.42	0.68	0.45	-	-

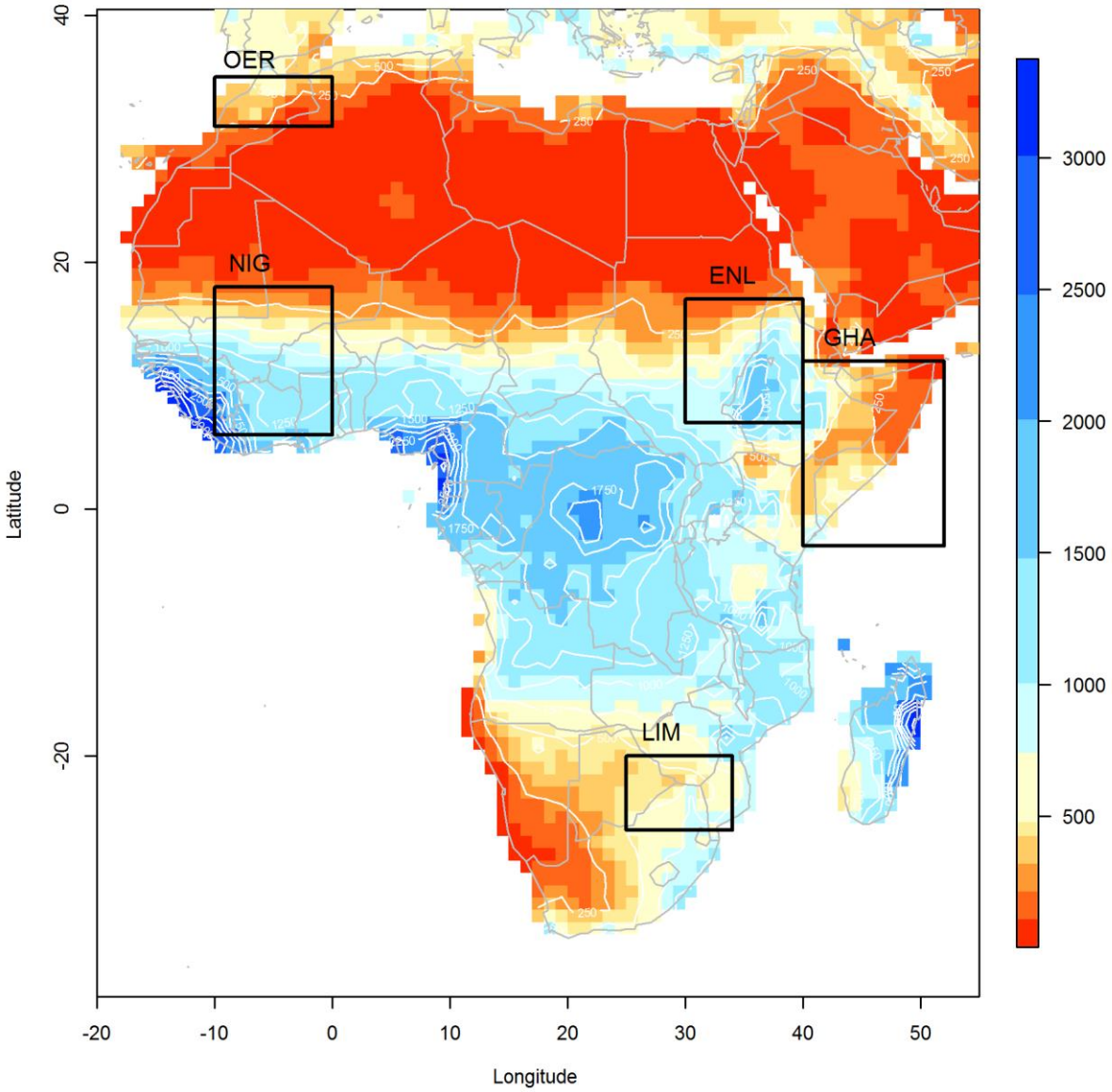


Figure 1. Annual mean precipitation from the GPCC dataset and African regions used in this analysis as defined in Table 1. (OER: Oum er-Rbia; NIG: Inner Niger Delta; ENL: Eastern Nile, LIM: Limpopo basin and GHA: Greater Horn of Africa.

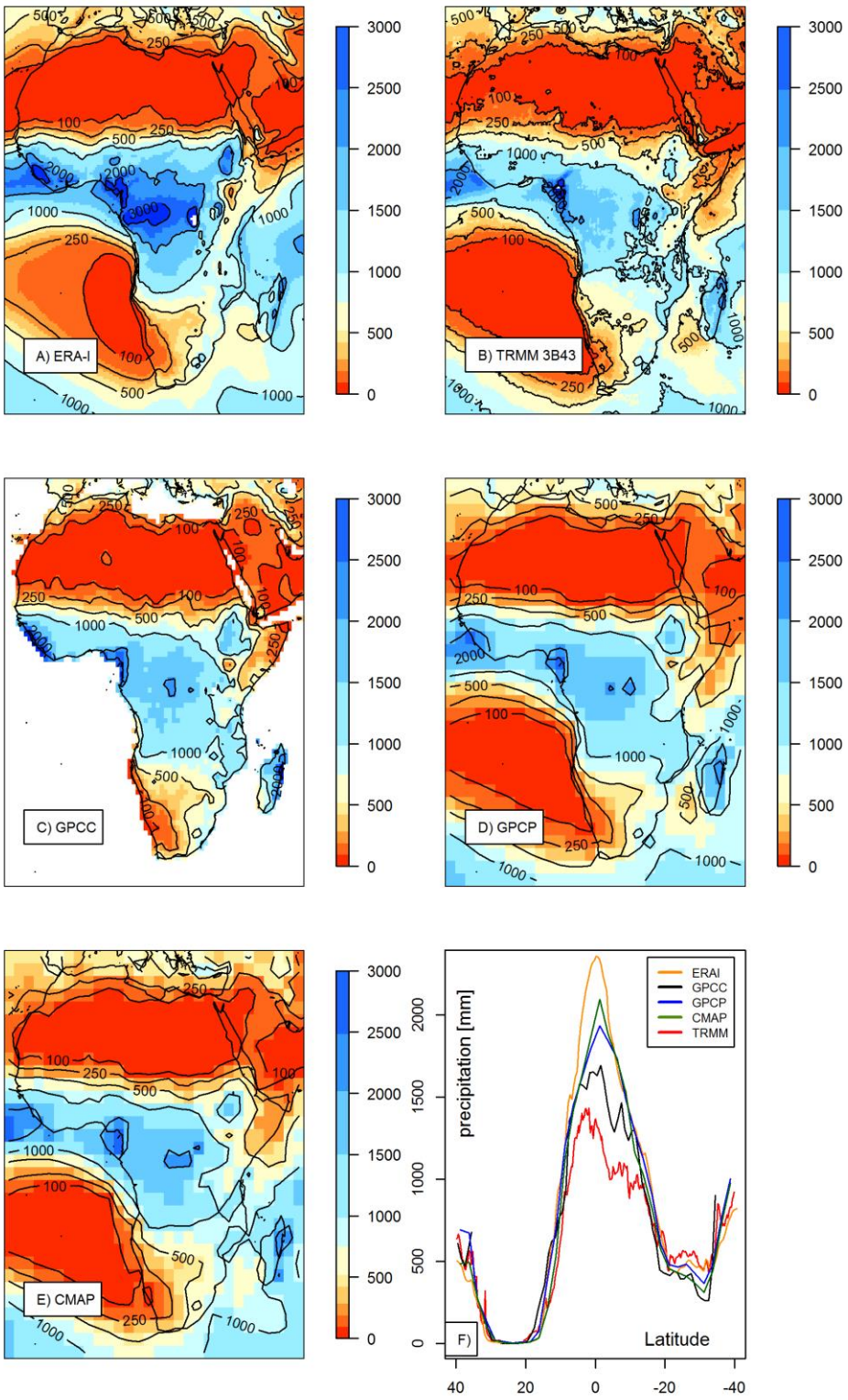


Figure 2. A-E) Mean annual precipitation (mm/year) from different datasets for the common period 1998-2010, F) longitudinal cross section at 25°E of mean annual precipitation.

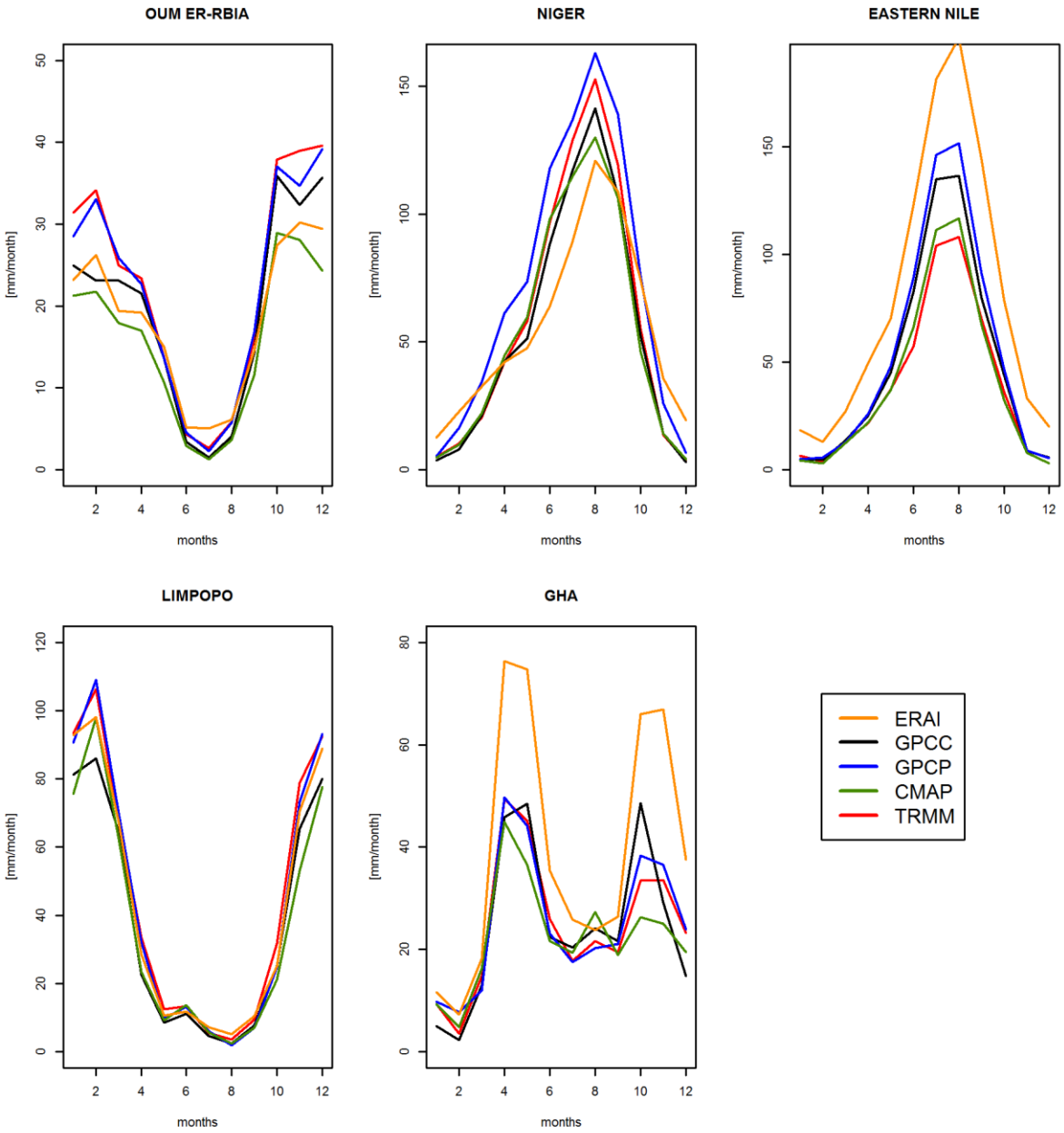
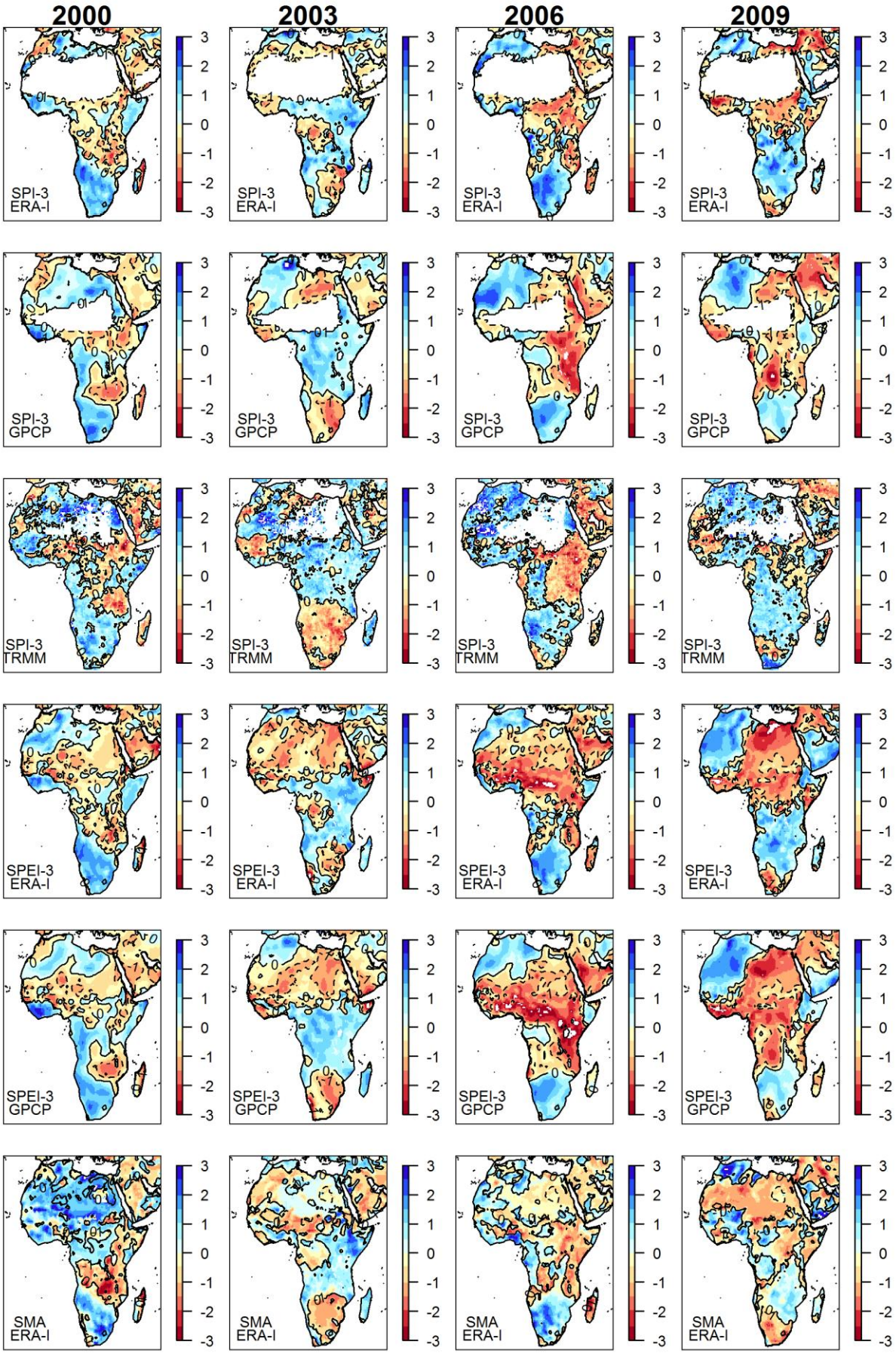


Figure 3. Mean annual cycle of precipitation from the different datasets averaged over the five regions defined in Figure 1 (OER: Oum er-Rbia, NIG: Inner Niger Delta, NIL: Eastern Nile, LIM: Limpopo basin and GHA: Greater Horn of Africa) for the common period 1998-2010.


2000
2003
2006
2009
SPI-3 ERA-I
SPI-3 GPCP
SPI-3 TRMM
SPEI-3 ERA-I
SPEI-3 GPCP
SMA ERA-I
3
2
1
0
-1
-2
-3

1

Figure 4. Monthly standardized anomalies in SPI-3 (ERA-I, GPCP, TRMM), SPEI (ERA-I and GPCP) and Soil Moisture (SMA) for January 2000, 2003, 2006 and 2009. Solid lines indicates the zero contour. White areas represent regions where it was not possible to compute the gamma parameters for SPI due the large amount of zeros.

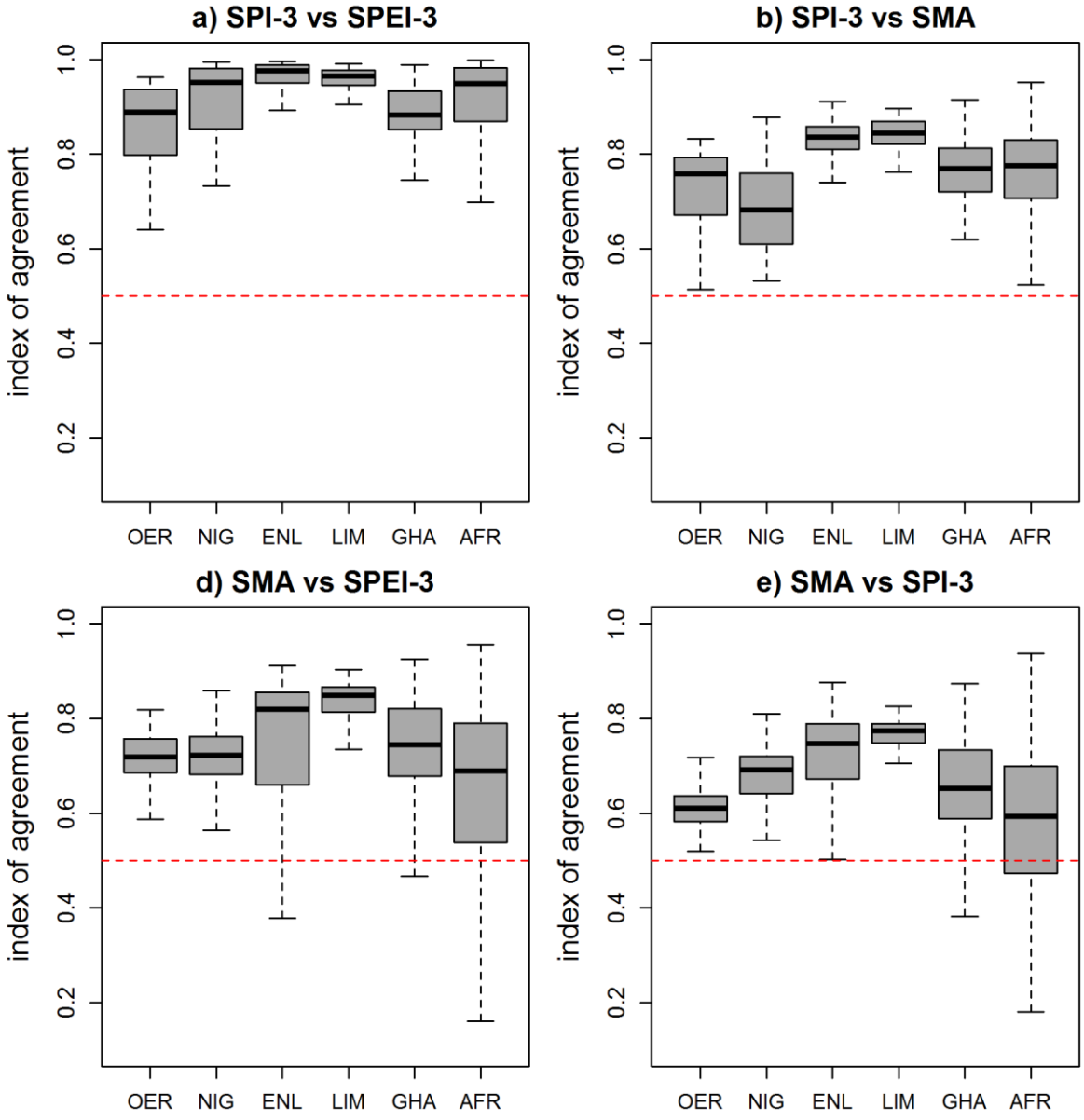


Figure 5. Index of agreement (d) between SPI, SPEI~~,~~ and SMA ~~and PET~~ computed using ERA-I for the five case studies and the whole continent. (OER: Oum er-Rbia, NIG: Inner Niger Delta, NIL: Eastern Nile, LIM: Limpopo basin and GHA: Greater Horn of Africa).

Dashed lines extend from 5th to 95th percentile of estimations, boxes extend from 25th to 75th percentile and middle horizontal lines within each box indicate the mean for each region.

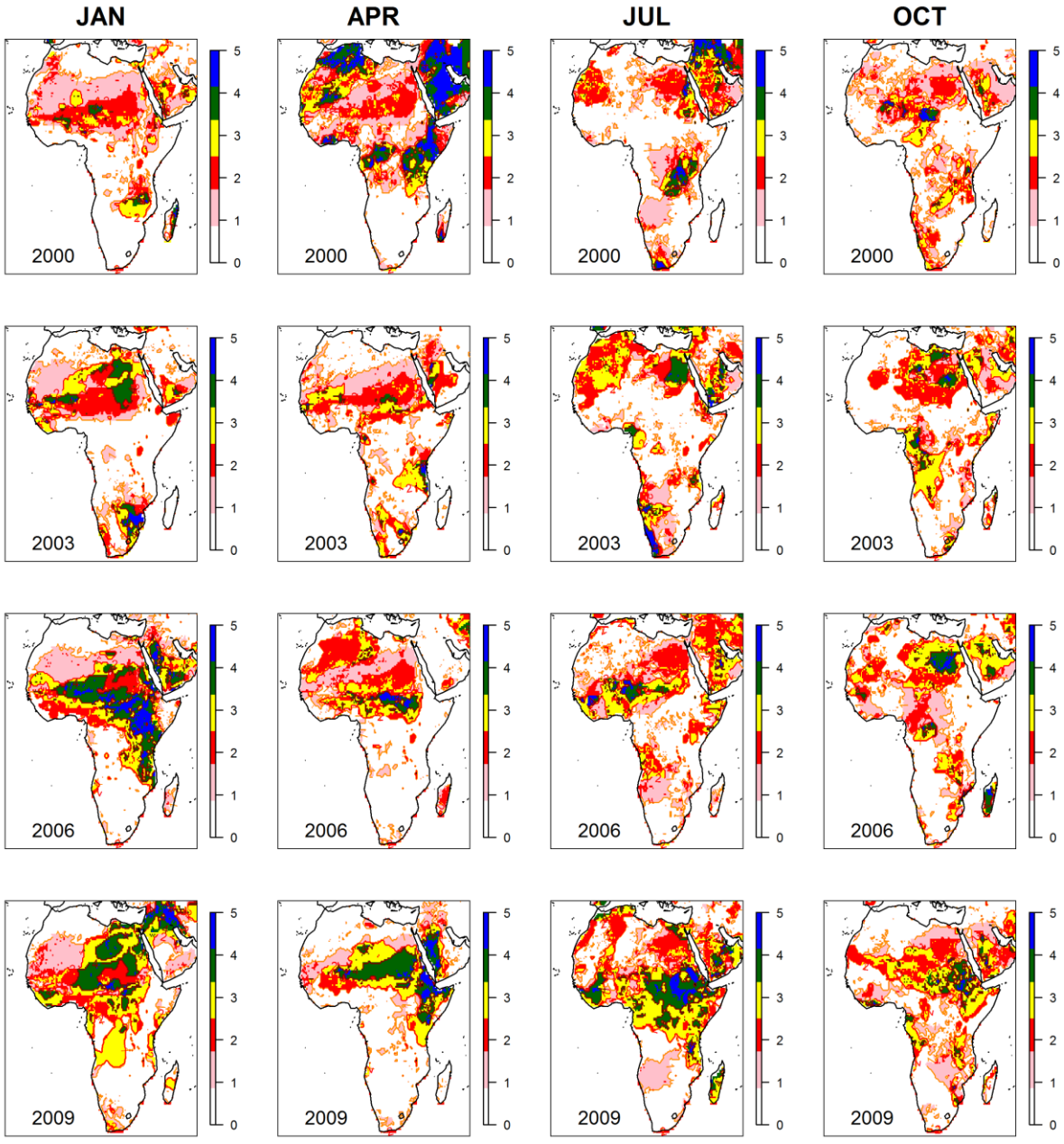


Figure 6. Month by month evolution of droughts in 2000, 2003, 2006 and 2009 according to grid cells with SPI-3/SPEI-3 computed using ERA-I GPCP, and TRMM below -1.0. Values are ranged between 0 (no dataset with SPI-3/SPEI-3 below the threshold) and 5 (all datasets below threshold).

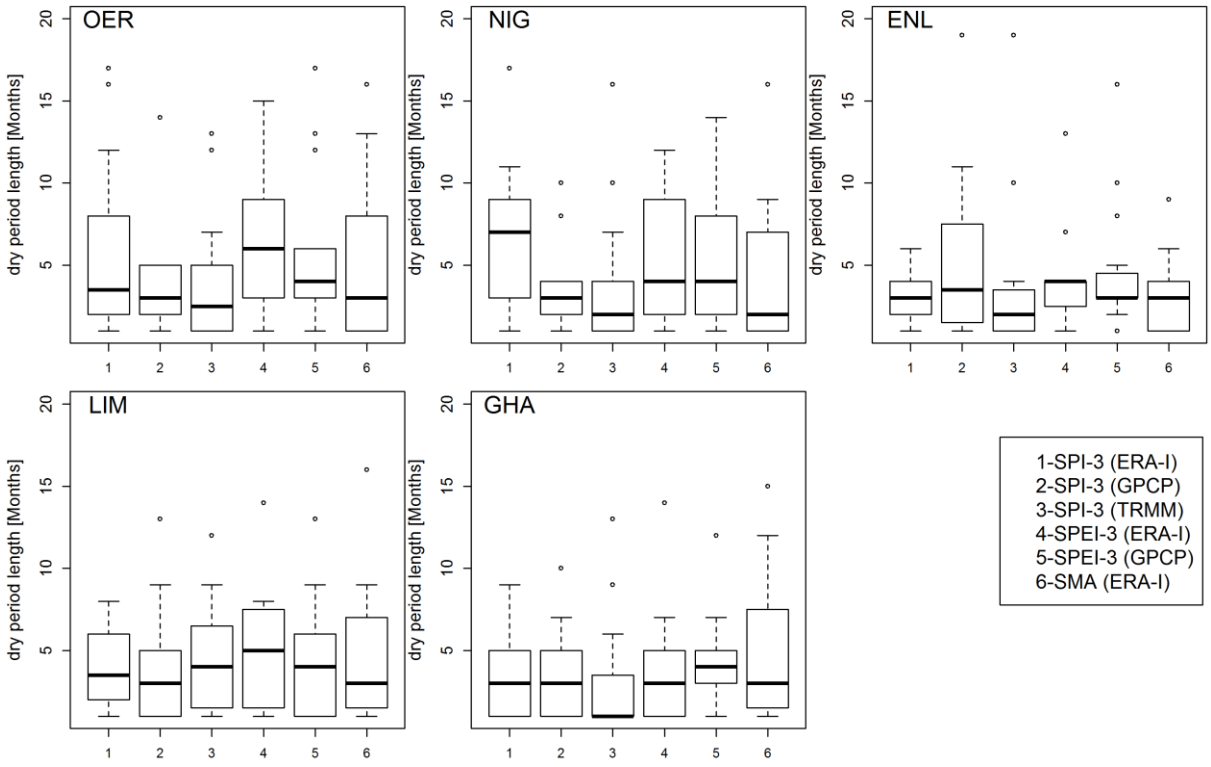


Figure 7. Duration of dry periods for the standardized indicators below zero in the common period 1998-2010. (OER: Oum er-Rbia, NIG: Inner Niger Delta, NIL: Eastern Nile, LIM: Limpopo basin and GHA: Great Horn of Africa). Dashed lines extend from 5th to 95th percentile of estimations, boxes extend from 25th to 75th percentile and middle horizontal lines within each box indicate the mean for each region.

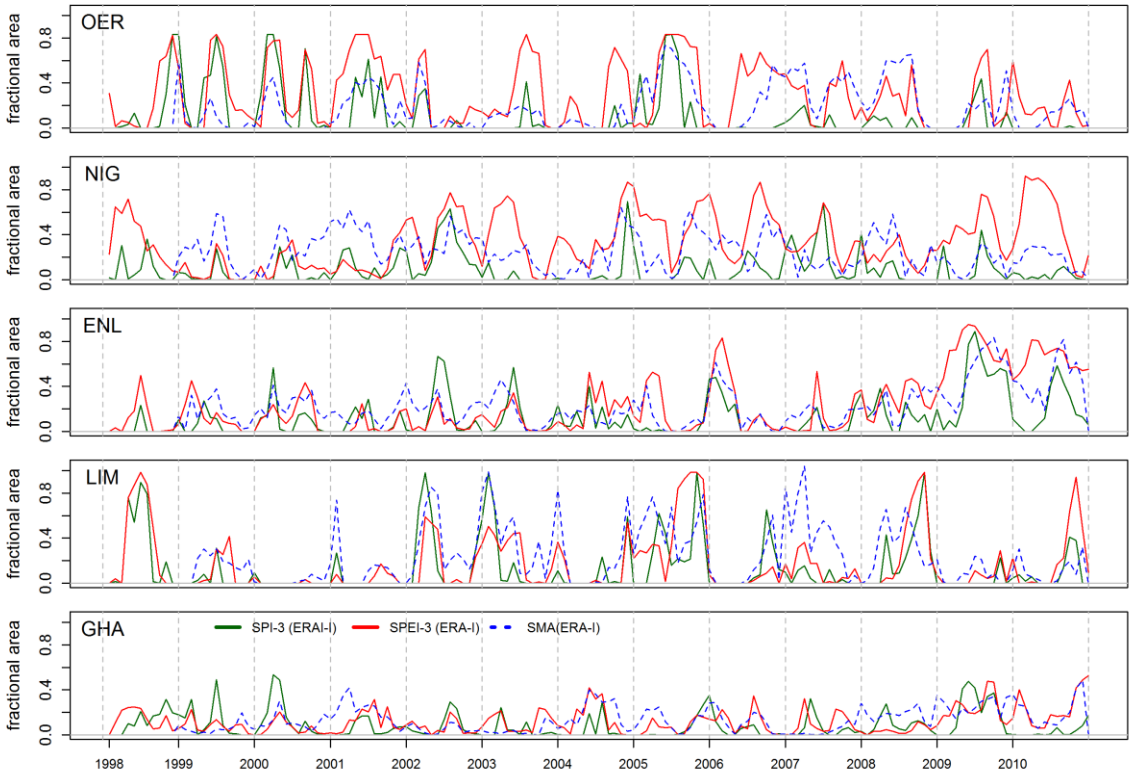


Figure 8. Fractional area of each region under SPI, SPEI and SMA ~~and PET z-scores~~ below -1.0 for the period 1998-2010. (OER: Oum er-Rbia, NIG: Inner Niger Delta, NIL: Eastern Nile, LIM: Limpopo basin and GHA: Greater Horn of Africa).

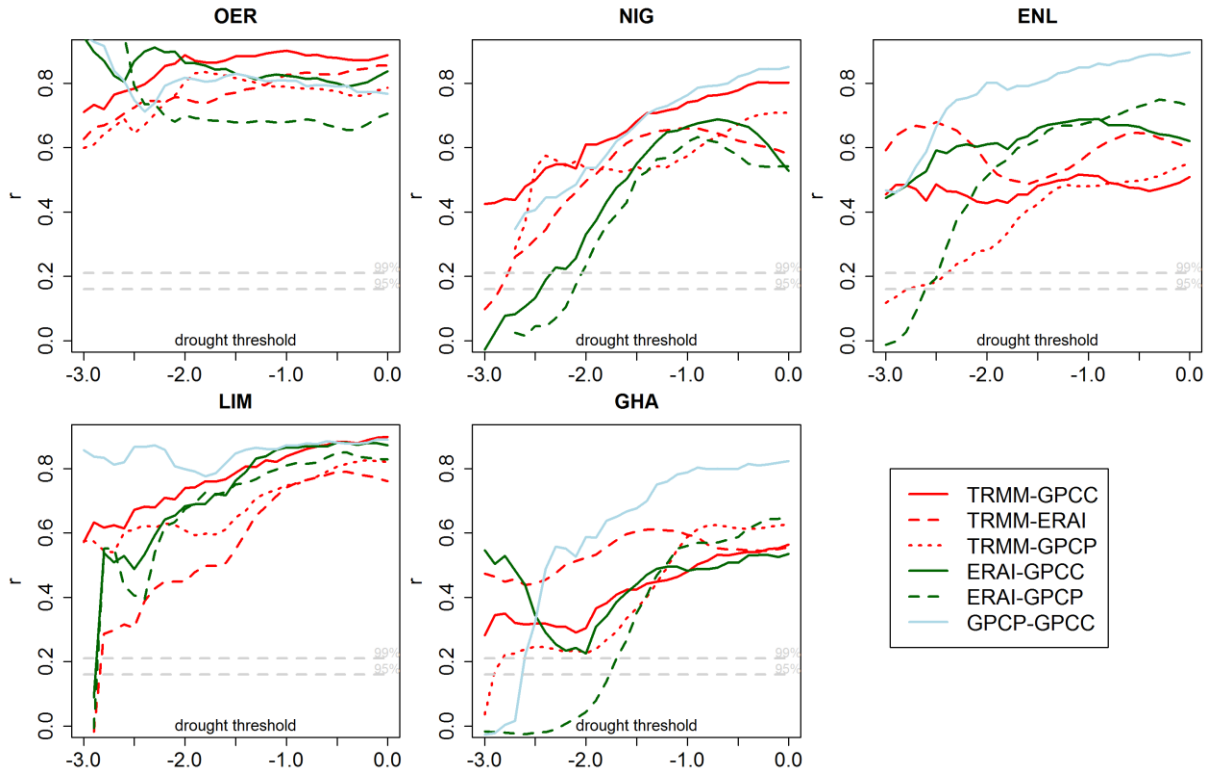


Figure 9. Correlation coefficient of fractional areas under drought between different datasets and thresholds. The horizontal axis represents the SPI threshold below which areas are considered to be under drought.

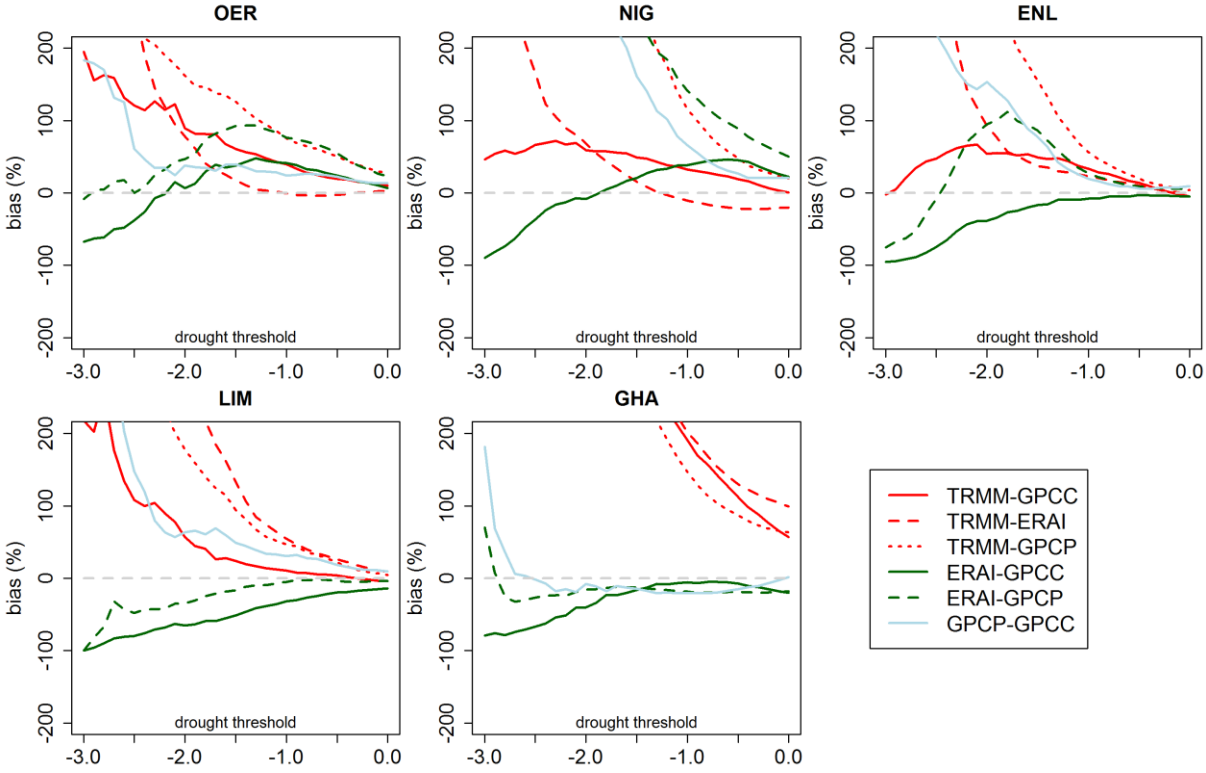


Figure 10. Relative bias between the estimation of fractional areas under drought for different datasets and thresholds. The horizontal axis represents the SPI threshold below which areas are considered to be under drought.