Published article:

Davis, J, Knippertz, P and Fink, AH (2012) *The predictability of precipitation episodes during the West African dry season.* Quarterly Journal of the Royal Meteorological Society, 139 (673). 1047 - 1058. ISSN 0035-9009

http://dx.doi.org/10.1002/qj.2014

The Predictability of Precipitation Episodes during the West African Dry Season

Jenny Davis and Peter Knippertz*

School of Earth & Environment, University of Leeds, Leeds, UK

Andreas H. Fink

Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany

Keywords: ensemble prediction system, tropical-extratropical interactions, verification, high-impact weather, TRMM, GPCP

*Corresponding author: Dr Peter Knippertz, School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK; E-mail: p.knippertz@leeds.ac.uk

Abstract

Precipitation episodes in tropical West Africa (7–15°N, 10°W–10°E) during the dry season from November to March are rare, but can have significant impacts on human activities reaching from greening of pastures to spoiling harvests and health implications. Previous work has shown a link between these unseasonal rainfalls and extratropical disturbances via a decrease of surface pressure over the Sahara/Sahel and a subsequent inflow of moist air from the Gulf of Guinea. This paper supports the previously stated hypothesis that the extratropical influence leads to a high rainfall predictability through a careful analysis of operational 5-day forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS), which are evaluated using Global Precipitation Climatology Project (GPCP) and Tropical Rainfall Measuring Mission (TRMM) precipitation estimates for the 11 dry seasons 1998/99–2008/09. The long-term regional average of ensemble-mean precipitation lies between the two observational datasets with GPCP being considerably wetter. Temporal correlations between the ensemble mean and observations are 0.8. Standard probabilistic evaluation methods such as reliability and relative operating characteristic (ROC) diagrams indicate remarkably good reliability, sharpness and skill across a range of precipitation thresholds. However, a categorical verification focusing on the most extreme ensemble mean values indicates too many false alarms. Despite the considerable observational uncertainty the results show that the ECMWF EPS is capable of predicting winter rainfall events in tropical West Africa with good accuracy, at least on regional spatial and synoptic time scales, which should encourage West African weather services to capitalize more on the valuable information provided by ensemble prediction systems during the dry season.

1. Introduction

Tropical West Africa is characterized by a monsoon climate with the largest portion of the annual precipitation falling in the boreal summer months (Hastenrath, 1991; Buckle, 1996). The period from around the start of November to the end of March is dominated by the dry and often-dusty northeasterly Harmattan winds and very sporadic rainfall events, which contribute little to the annual total on average. Nevertheless, impacts of these events on the local population can be manifold and include: (A) Harvests such as cotton are often stored to dry outdoors and unexpected rain can cause them to rot (Buckle, 1996; Knippertz and Fink, 2008; 2009). (B) If damp, staple foods such as groundnut and maize can become contaminated by aflatoxins, fungal metabolites that can cause sickness or death in humans and animals (Hell and Mutegi, 2011). (C) Unseasonal rains can significantly improve grazing conditions, e.g. for the herds of kettle nomads, and facilitate agricultural work such as ploughing or building moulds for yam due to enhanced soil moisture. (D) Anomalously moist periods during the dry-season can help to prevent epidemics of meningococcal meningitis, which is widespread in West Africa (Sultan *et al.*, 2005; Thomson *et al.*, 2006). These examples show that reliable predictions of dry-season rainfall events in tropical West Africa on synoptic timescales have the potential to support decision-making processes for a wide range of mitigating actions. Particularly points (A) and (B) above would clearly benefit from the establishment of an early-warning system up to a week ahead.

Given the predominance of summer rainfalls for the annual totals, rather little work has been dedicated to the dynamics and climatology of precipitation during the dry season, mostly in the form of case studies of extreme events (e.g. Knippertz and Martin, 2005; Fall *et al.*, 2007; Knippertz and Fink, 2008; Meier and Knippertz, 2009). Most of these cases occurred over the western parts of tropical and subtropical West Africa, which are occasionally affected directly by very deep upper-level disturbances over the Atlantic Ocean (Fröhlich and Knippertz, 2008). Knippertz and Fink (2008; KF08 hereafter) were among the first to analyze the dynamics of extreme unseasonal rainfall in southern West Africa. The mechanism they proposed is schematically depicted in Figure 1. The presence of a pronounced, positively tilted upper-level trough over northwestern Africa leads to a shift of the subtropical jet (STJ) and a decrease of surface pressure over the Sahara and Sahel. In the particular case KF08 investigated, two low-pressure centres formed, one close to the base of the trough over Algeria associated with unusual precipitation over the northern Sahara and one over tropical West Africa. The latter can be regarded as an intensified and northward shifted wintertime continental heat low. This pressure configuration leads to a significant break in the Harmattan winds and the inflow of moist southerlies from the Gulf of Guinea, which shifts the Intertropical Discontinuity (ITD), the surface boundary between moist maritime and dry continental air, northwards and feeds the unseasonal rainfalls.

A follow-up study by Knippertz and Fink (2009; KF09 hereafter) contains the first-ever statistical analysis of dry-season rainfall events over southern West Africa (7–15°N, 10°W–10°E), based on pentad precipitation from the Global Precipitation Climatology Project (GPCP) and 5-day forecasts made as part of the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40 re-analysis project (Uppala *et al.*, 2005) for the period 1979/80–2001/02. The main conclusions from this work are that (A) the schematic shown in Figure 1 is representative for many of the identified events, (B) the

ECMWF model shows skill in predicting event occurrence on a regional scale up to a week ahead and (C) predictability appears to be enhanced in cases of a clear connection to the extratropics. The latter is typically manifested through a well-defined and persistent upper-level trough or the succession of two troughs, accompanied by an elongated southwest–northeast oriented band of upper- and midlevel clouds stretching from the Tropics to the subtropics along the equatorward side of the STJ. Such bands are often referred to as "Tropical Plumes" (see Knippertz, 2005 and references therein).

This study builds on those by KF08 and KF09, and expands them in the following ways: (A) In order to test the hypothesis of enhanced predictability formulated in KF09, operational ensemble predictions from ECMWF are investigated instead of ERA-40 data. This allows a combination of conventional and probabilistic verification measures to be used. (B) KF09 amongst other authors have demonstrated problems with the old pentad product of the GPCP (Xie *et al.*, 2003) over data-sparse West Africa. The present study uses a new daily version of GPCP (1DD) together with rainfall estimates from the Tropical Rainfall Measuring Mission (TRMM), which includes information from space-born rainfall radar, to assess aspects of observational uncertainty. These improvements in the data quality come at the expense of reduced data availability, which limits this analysis to the 11 dry seasons 1998/99–2008/09. Section 2 provides more information on the datasets used in this study. The results are presented in sections 3 and 4 focussing on evaluation of the ensemble mean and probabilistic analysis, respectively. Section 5 provides a short discussion of the results and gives the main conclusions.

2. Data

Most analyses presented in this paper are based on gridded datasets of estimated and forecast precipitation, respectively, as detailed in the following sections 2.1 and 2.2. Following KF09, these fields are averaged spatially over 7–15°N, 10°W–10°E and accumulated over 5 days to give a regional precipitation index. As shown in KF09, the ECMWF model is generally capable of forecasting accurately at much finer scales, but a grid-point based verification of the predominantly convective rainfalls is generally difficult, which has motivated several authors to develop object-based methods, all with their different advantages and problems (e.g. Wernli *et al.*, 2008). Such an approach is beyond the scope of this paper. Instead this study focuses on the question whether the ECMWF model can accurately predict moist episodes during the dry-season on a regional scale, which in our view is sufficient for many of the applications discussed in the Introduction, particularly for early-warning systems. The time period covered is the entire dry-season from 1 November to 31 March.

2.1 Observational data

Daily precipitation estimates used in this paper are from the GPCP 1DD and TRMM 3B 42 Version 6 datasets. The former is largely based on a monthly combined satellite-gauge product, which is used to calibrate daily estimates derived from geostationary and polar-orbiting infrared sensors (Huffman *et al.*, 2001; Roca *et al.,* 2010). Microwave and gauge estimates are not used explicitly owing to sampling limitations. It is provided with a spatial resolution of one degree. The TRMM daily product is derived from a combination of TRMM microwave imager, radar and visibleinfrared scanner data and other satellite infrared observations (e.g. Meteosat; see Huffman *et al.,* 2007). Among other things, the TRMM data are used to produce monthly infrared calibration parameters, which are then applied to 3-hourly precipitation estimates from the other satellite infrared datasets. The daily totals are estimated from the 3-hourly precipitation data between 00Z and 21Z. Finally the daily totals are scaled so that the monthly total matches that of the satellite-gauge TRMM Combination 3B 43 Version 6. These data have a spatial resolution of 0.25 degrees. Both datasets are averaged regionally as explained above and accumulated over five days starting at 0000 UTC each day. This results in 147 pentads for one entire dry season with the first ranging from 1–5 November, the second from 2–6 November and so on, the last covering 27–31 March. This gives 1617 pentads for the 11 dry seasons under study here. Note that for the sake of simplicity, 29 February was ignored for the leap years during the study period 1998/99–2008/09.

A scatter plot of all GPCP pentad precipitation values against their TRMM counterparts (Figure 2) shows a clear positive bias in the GPCP data. Such discrepancies between rainfall estimates illustrate the challenge in evaluating short-term forecasts in the light of large observational uncertainties. Intercomparison studies at 10-day and monthly scales over West Africa (Nicholson *et al.,* 2003a, b; Ali *et al.,* 2005; Lamptey, 2008; Jobard *et al.,* 2011; Paeth *et al.*, 2011) have also found substantial differences between the two datasets. For more details, see the recent comprehensive review by Parker et al. (2011), which also includes daily products. Despite the bias, the correlation between TRMM and GPCP is 0.92. In interpreting this value, however, it should be kept in mind the data points in Figure 2 are not independent due to the overlap of adjacent pentads.

In addition to the precipitation data discussed above, ECMWF ERA-Interim re-analysis (Dee *et al.*, 2011) and Meteosat infrared satellite images are used in section 3 for the discussion of example case studies (e.g. Figure 8).

2.2 Ensemble predictions

The forecast dataset under investigation comes from the operational ECMWF ensemble prediction system (EPS; Buizza *et al.,* 1997; 2007). It was designed to cope with uncertainty in initial conditions and is now also taking into account uncertainties in model physics (Buizza *et al.*, 1999). The EPS became operational in 1992 and has had 50 members since 1996. As a fully operational system, it went through a number of model and configuration changes in the course of our investigation period 1998–2009, so that the forecasts studied here are not a homogeneous dataset in the strictest sense. It is, however, one of the longest and arguably best available EPS datasets to study predictability today. The EPS is currently run twice daily at 0000 and 1200 UTC, but as this was not the case during the first part of our study period, we restrict the analysis to the latter time only. In order to match the TRMM and GPCP pentad data, differences between total precipitation accumulations (convective plus large-scale) at +132h and +12h are considered for each day and then averaged over 7–15°N, 10°W–10°E.

3. Analysis of the ensemble mean

It has been shown that for forecast ranges beyond three days predictions based on the mean of a well-calibrated EPS outperforms a deterministic forecast with the same model after about 3 days (e.g. Buizza *et al.,* 1997). Therefore, this section concentrates on assessing forecast quality based on the ensemble mean only, while section 4 will focus on probabilistic measures. Section 3.1 analyzes the mean seasonal cycle and correlations between the observational and EPS data, while section 3.2 focuses on an event-based verification. Finally, section 3.3 discussed exemplary cases of subjectively selected successful and unsuccessful forecasts.

3.1. Mean seasonal cycle and correlations

Figure 3 shows the seasonal cycle for the EPS, TRMM and GPCP datasets averaged over the 11 dry-seasons 1998/99 to 2008/09. All three datasets show the characteristic decrease from above 2 mm per pentad in early November to very small values in late December and then a gradual increase to values above 3 mm per pentad at the end of the period in late March. Although the overlapping pentad accumulation causes a smoothing of the curves, single significant events still stand out from the 11-year background as for example a period in mid-February. Overall, GPCP shows considerably higher values (on the order of 50%) during the early and late parts of the dry season, while agreement with the other two datasets is better during the middle part. TRMM agrees remarkably well with EPS during the first half of the dry season and shows some tendency for lower values than EPS later on. Averaged over the entire dry season and all years, differences between EPS and TRMM (GPCP) are 0.12 mm (–0.41 mm) per pentad. An analysis of the reasons for these discrepancies is beyond the scope of this paper. They show, however, that there is a considerable uncertainty in the observations, which pose a significant problem to evaluation and model development. Part of this problem is certainly related to the sparse network of surface stations that hampers the calibration and evaluation of satellite retrievals (see also the discussion at the end of section 2.1).

Despite the biases discussed above, the temporal behaviour found in the observational datasets is overall well reproduced by the EPS with linear correlation coefficients reaching a remarkable 0.80 for both observational datasets (Figure 4). Part of this strong relation is certainly associated with the general moistening in February/March (Figure 3), but as we will show below there are significant events that stand out well from the background during all parts of the dry season, which are mostly well reproduced by the ensemble mean. Correlations with positive and negative lags show consistently lower values, indicating that the model does not tend to trigger precipitation too late or too early in a systematic way. It is interesting to note that the regression lines in both panels of Figure 4 are below the diagonal despite the slightly lower mean values in TRMM. This indicates that the EPS shows a general tendency to underestimate the wettest events and overestimate low-intensity events. This finding is consistent with results by KF09 using ERA-40 data.

3.2 Event-based verification

Given the strong seasonal cycle displayed in Figure 3, an event definition based on absolute values (either total or anomaly thresholds) is problematic. To circumvent this problem, KF09 defined events as anomalies of more than 200% relative to the mean seasonal cycle. This method biases the event identification to the driest part of the season when such large values are reached with much smaller absolute precipitation amounts. To avoid this effect, a new approach is proposed here that involves a moving block of 10 pentads in the following way:

- Identify the maximum of the first 10 pentad accumulations of the season (1–5 November, 2–6 November, 3–7 November, and so on until 10–14 November) of all 11 dry-seasons (110 values in total).
- Shift by 5 days and identify maximum of the 110 pentads 6–10 November, 7–11 November, 8–12 November and so on until 15–19 November of all dry-seasons.
- Repeat shift by 5 days 26 times until the end of the dry-season. The last 10-pentad block contains 16–20 March, 17–21 March, 18–22 March and so on until 25–29 March. This way the last two days, 30 and 31 March, cannot be considered.

In principle this procedure finds 28 maxima evenly distributed across the dry season. However, the overlap between the 10-pentad blocks used to find maxima leads to double counting, giving 20 actual maxima in TRMM and 21 in GPCP and EPS. In addition, if identified maxima are four days or less apart from each other in time, they are considered part of the same meteorological event and only the pentad with the larger precipitation value is retained. This reduces the final numbers of events to 16 (TRMM) and 17 (GPCP and EPS), thus about one and a half event per dry-season (results are listed in Table I). The two dry-seasons 2000/01 and 2004/05 stand out as the ones without any events in any of the datasets. Remarkably, TRMM and GPCP only agree on 12 events (75% and 71% of all, respectively), underlining the substantial observational uncertainty.

The definition of events as explained above enables the identification of forecast hits, missed events and false alarms. In matching events from two different datasets, a time shift of 1 or 2 days was tolerated. Table II gives the results of this analysis. 7 events were found with all three datasets matching, while additional 2 had at least a match between EPS and TRMM (see shadings in Table I). Given the observational uncertainty, these can be regarded as 9 hits. 5 events are identified in both TRMM and GPCP, but not in the EPS data and therefore qualify as missed events. Interestingly these are predominantly during the early part of the study period (1998/99, 1999/2000, 2 in 2001/02, 2008/09). It may indicate that improvements to the operational system in the course of the study period have led to a reduction of misforecasts. There are 7 situations in which only one of the two observational datasets shows an event. These could be interpreted as partial misses, but we would argue that they should be considered correct negatives in the light of the observational uncertainty. The biggest problem with the event forecasts is the high number of false alarms (8 during the 11 seasons). Some of these clearly stand out as significant outliers in Figure 4. Given the impacts of dry-season rainfalls as discussed in the introduction, it is probably tolerable to have more false alarms than missed events, but such a high number points to some fundamental problem in terms of predicting rainfall quantities.

Given the relatively small number of events and the particular nature of the identification routine that is designed to give fairly equal numbers for each datasets, the authors decided not to take this analysis any further by computing standard verification measures such as frequency bias, hit rate, false-alarm ratio etc. However, it is interesting to view the results for the categorical evaluation of the ensemble mean in a more probabilistic sense. To do this, it is necessary to see to what extent the false alarms and missed events discussed above are consistent with the ensemble spread for the given period. Situations where the observations lie outside of the spread are undesirable and should occur rarely for a well-tuned EPS. Results show that four out of five missed events (all but that in 1998/99) fall inside of the ensemble spread with one even inside the interquartile range. This suggests that most of these cases are only ‘missed’ in the sense of the event definition based on the ensemble mean, but that they can still be considered successful forecasts in a probabilistic sense. Five of the eight false alarms fall inside of the spread with three inside the interquartile range. The remaining three can be considered as misforecasts. This simple analysis suggests that probabilistic measures as discussed in section 4 will most likely evaluate the performance of the EPS more positively than the event-based one presented here.

3.3 Example cases

To illustrate this further, Figure 5 shows time series of the three considered datasets for four selected example dry seasons. TRMM and GPCP data are plotted as lines with crosses; the EPS data is depicted with box-and-whiskers plots (see caption for more details). Identified events are marked with bold horizontal lines. There is a generally very low ensemble spread during many observed dry periods, which could be capitalized on in terms of impacts.

The first season 2001/02 shows four conspicuous rainfall periods (Figure 5a):

- The event in mid-November qualifies as a miss, but at least the EPS forecasts unseasonal rainfall around the right time with all members above 2 mm per pentad and TRMM estimates well within the ensemble spread. The quantitative disagreement between GPCP and TRMM is remarkable.

- The weaker event in late November is only flagged by GPCP, while TRMM is well within the interquartile range of the EPS, which starts precipitation too early in this case, possibly related to problems with representing shallow stable layers (see KF09).
- The event in early January is a clear hit with remarkable agreement between the three datasets. This case brought heavy rainfall across large parts of West Africa and severe flooding in Senegal and Mauritania (outside of our study region; see Knippertz and Martin, 2005; Fall *et al.*, 2007). Figure 6 shows the horizontal distribution of rainfall for the three datasets. The overall spatial agreement is good, but GPCP tends to extend very light rainfall far into the Sahel and Sahara and smears out the localized maximum over the Ivory Coast (Figure 6b). The ensemble-mean forecast (Figure 6c) is expectedly rather smooth and shows the advance of the rainfalls into the southwestern part of the domain very clearly. Meier and Knippertz (2009) also noted the high predictability of this case in their model sensitivity experiments.
- The event in the first half of March qualifies as a miss. This event brought unusual early rains in central Benin between 9 and 11°N (Fink et al., 2006). Again, there is a large disagreement between GPCP and TRMM, which falls just inside the interquartile range of the EPS. An underestimation of March rainfalls is not observed for most other years, which might be connected to a soil moisture or vegetation bias after the unusual rainfalls in January.

The dry season 2003/04 had four wet events (Figure 5b). The first (early November), third (late January) and fourth (early March) are classified as hits, despite a significant underestimation for the former two (recall that event identification occurs relative to each dataset). This is clearly a disadvantage of the event identification proposed here, suggesting that it should only be used in combination with other, more continuous measures. The event in mid-January (Figure 7) is the case discussed in detail in KF08, which first instigated research into this phenomenon. TRMM clearly shows very unusual, heavy precipitation reaching from Ivory Coast into Nigeria (Figure 7a). GPCP shows a coarse-grained and somewhat weaker pattern, again with light rains spreading far away from the main rainfall event (Figure 7b). The ensemble mean gives a clear indication of a general northward shift of the rain zone, but fails to reflect the full detail and magnitude (Figure 7c), largely consistent with the ERA-40 forecasts in KF08. The event in mid-November 2003 (Figure 5b) and the very last pentad of this dry season (not identified as an event) show a remarkable disagreement between GPCP and the other two datasets..

The dry season 2005/06 (Figure 5c) again shows large disagreement between the two observational datasets over longer periods, particularly in November, December and March. The first event in late January is remarkably well forecast, but the second one in mid-February is the most significant false alarm of the study period with both observational datasets well below the driest ensemble members. The synoptic situation during this event was characterised by a very pronounced, strongly tilted upper-level trough located over northwestern Africa and the adjacent Atlantic Ocean, which is associated with an area of low surface pressure reaching from Burkina Faso to southwestern Algeria (Figure 8a). As a consequence the ITD shifts northward (thick line in Figure 8a). The satellite image on 17 February 2006 (Figure 8b) shows a subtropical cyclone-like cloud structure over the Sahara with some patchy convection to the south of it in the Tropics, which locally brought precipitation exceeding 50 mm (e.g. on 15

February 2006 around 9.5°N, 2°E; Pospichal et al., 2010). According to TRMM there are two separated areas of rainfall of moderate intensity: one associated with the cyclonic feature (mostly outside the study area) and one associated with the northward shifted ITD (Figure 8c). GPCP shows a coarse-grained version of this with very widespread light precipitation over the Sahel and Sahara (Figure 8d). Surprisingly, the main band over the Sahara is shifted eastwards in the GPCP data with respect to TRMM. In the EPS mean the two precipitation areas are connected, leading to too much and too widespread precipitation (Figure 8e). It appears that the model triggers convection too easily in this situation of enhanced low-level moisture and supposedly upper-level forcing for uplift. It is also conceivable that evaporation of precipitation in the dry desert air is not handled well in the model. The frequent occurrence of false alarms suggests that these problems are potentially systematic.

The last example, the dry season 2008/09, was one of the most active seasons with 5 identified events (3 hits, one missed event and one false alarm; Figure 5c). The latter underlines again the problems with the event definition, which can indicate a misforecast, although the absolute precipitation amounts agree rather well with each other. The most remarkable event in this season is the heavy rainfall in mid February 2009 (see Waliser *et al.* 2012). Other events mentioned in that paper are 5–6 December 2008 and 8–9 January 2009.

4. Probabilistic analysis

In this section the full probabilistic information content from the EPS is explored with two standard evaluation methods, using the available TRMM and GPCP data for the

11 dry seasons 1998/99 to 2008/09 (as described in section 2.1) as the observed 'truth'. This analysis complements the categorical evaluation based on extreme values in the ensemble mean presented in section 3. It is assumed here that each ensemble member carries the same probability of occurrence.

The first method is the relative operating characteristic (ROC) diagram (e.g. Joliffe and Stephenson, 2003). It is constructed using a set of simple four-cell contingency tables. For the observations an event is defined through exceedance of a certain precipitation threshold. For the EPS, the event/no event decision is made based on a given forecast probability threshold (here 2, 20, 40, 60, 80 and 98%) for the chosen precipitation amount. The hit rate, H, is then defined as the ratio of the number of correct forecasts divided by the total number of observed events. The false alarm rate, F, is defined as the ratio of the number of false alarms divided by the total number of non-events in the observations. In this way, a single point can be plotted on a graph of H against F. Plotting this for a set of probability thresholds creates a ROC curve, which has several important characteristics. The bottom left corner represents a situation of no warnings at all and therefore $H = F = 0$. The top right corner describes a situation of always warnings and therefore $H = F = 1$. Typically the area underneath the ROC curve is taken as a measure of skill (Buizza *et al.*, 1999). A perfect forecast will have $H = 1$ and $F = 0$ and therefore a ROC area of 1.

The reliability of an EPS is its ability to forecast accurate probabilities (Palmer, 1999). This can be simply tested by plotting forecast probability against observed frequency, again for a given precipitation threshold. The diagonal in this diagram indicates perfect reliability. Circles of varying sizes represent the frequency of forecast probabilities. Largest circles in the centre of the diagram indicate clustering around the climatological average and therefore low predictability. The property of an EPS to spread away from the climatological average is called “sharpness”.

Examples of ROC and reliability diagrams for different thresholds are shown in Figure 9 for both TRMM and GPCP observations. For a threshold of 0.5 mm per pentad both datasets are relatively close to the diagonal indicating good reliability (Figure 9a). There is a general tendency to underestimate observed frequencies for low forecast probability, particularly in GPCP. This could be partly related to the spurious widespread low-intensity rainfall evident from Figures 6b, 7b and 8d. On the other hand observed frequencies are overestimated for high forecast probability. This might be a reflection of the EPS triggering convection too easily if the general conditions are favourable consistent with the high number of false alarms discussed in the previous section. The positive bias of GPCP with respect to TRMM leads to a general upward shift to higher observed frequencies in the diagram. The differences in behaviour between low and high observed frequencies do not make it possible to improve reliability greatly through a general bias correction. The distribution has good sharpness with largest circles in the top right corner, representing situations of successful forecasts where all ensemble members exceed the threshold. The corresponding ROC diagram (Figure 9b) shows good skill for both observational datasets with ROC areas of 0.90 (TRMM) and 0.91 (GPCP), with differences again reflecting the different biases as discussed previously.

Figures 9c and 9d show the corresponding analyses for a precipitation threshold of 1 mm per pentad. Overall the results are very stable with only a slight shift to more frequent events of low forecast probability as expected for a higher precipitation threshold. This general behaviour continues for higher thresholds, so that for 3 mm per pentad only few events in the top right corner of the reliability diagram are recorded (Figure 9e). However, ROC scores continue to remain above 0.9, even for these relatively extreme events (Figure 9f). Overall this analysis shows that EPS forecast of dry-season precipitation are in fact of high quality and usefulness.

5 Discussion and conclusions

Dry-season precipitation events in tropical West Africa are rare, but have important ramifications for the local population. This work extended previous studies on this subject by KF08 and KF09 in two ways: (A) Forecasts of these events considered here are from operational ECMWF ensemble predictions that allow an assessment of predictability. (B) More recent and high-resolution precipitation datasets are used for evaluation. The study region corresponds to that of KF09 and spans 7–15°N, 10°W–10°E. The 11 dry seasons (November–March) 1998/99–2008/09 were investigated. EPS forecasts and observations were compared on the basis of pentads, using +132h minus +12h predictions. Evaluations are done both for the ensemble mean and using probabilistic methods. The most important conclusions from this work are:

- There is a considerable observational uncertainty for this region during this time of year. GPCP has a substantial positive bias with respect to TRMM and tends to show widespread low-intensity rainfall. Although the overall temporal correlation is 0.92, deviations during single events can be remarkably high, practically impeding a forecast evaluation for some individual cases.

- The agreement of EPS is generally better with TRMM than with GPCP. This holds for the mean seasonal cycle, temporal correlations, event evaluation and ROC scores.
- Temporal correlations between EPS mean and the observational datasets reach 0.8.
- ROC scores are on the order of 0.9. Sharpness and reliability are satisfactory with a general tendency of too high (low) forecast probabilities for high (low) observed frequencies.
- Categorical evaluation of extreme events identified from the ensemble mean is much more sensitive to small variations in precipitation amounts and therefore indicates less skill. There is a moderate number of missed events, but the biggest problems are too many false alarms and a tendency of the EPS to start precipitation too early. Both may indicate that convection is triggered too easily in the typical dry-season rainfall situation with upper-level forcing and high low-level moisture.

Overall the results presented here indicate a general ability of the ECMWF EPS to provide reliably forecast information of dry-season rain events in tropical West Africa on the pentad timescale. It would be interesting to explore whether there is also some seasonal predictability, for example related to the influence of El Niño on upper-level troughs and tropical plumes over the eastern North Atlantic (Luise Fröhlich, University of Cologne, pers. comm., 2012). One of the important limitations of this study is the large observational uncertainty related to the disagreement of different precipitation products on daily timescales (see Parker et al., 2011). The evaluation of rainfall products is generally focused on the rainy season, such that biases during the dry season are particularly large (Adeyewa and Nakamura, 2003). More targeted efforts are needed to understand the origin of such biases and to explore retrievals designed for precipitation outside of the main rainy season. Other research in the future should explore how the predictability found in this study can be used to inform decision makers in West Africa, particularly in the health, agriculture and water sectors. This will most likely require investigations on finer spatial scales than used in this study and include the identification of optimal forms of communicating uncertainty in ensemble predictions to a given end-user community.

Acknowledgments

This work was funded by the German Science Foundation (DFG) Emmy Noether program (grant KN 581/3-2). The TRMM and GPCP data were provided by the NASA/Goddard Space Flight Center's Laboratory for Atmospheres. We thank the UK Met Office for providing access to the ECMWF archive and the Dundee Satellite Receiving Station for providing access to Meteosat imagery. We are grateful to two anonymous reviewers whose comments helped to substantially improve an earlier version of this manuscript.

References

Ali A, Amani A, Diedhiou A, Lebel T. 2005. Rainfall estimation in the Sahel. Part 2: Evaluation of raingauge networks in the CILSS countries and objective intercomparison of rainfall products. *J. Appl. Meteorol.* **44:** 1707–1722.

Adeyewa ZD, Nakamura K. 2003. Validation of TRMM radar rainfall data over major climatic regions in Africa. *J. Appl. Meteorol.* **42:** 331–347.

Buckle C. 1996. *Weather and Climate in Africa*. Longman, London, UK, 312pp.

Buizza R. 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction System. *Mon. Weather Rev.* **125:** 99–119.

Buizza R, Hollingsworth A, Lalaurette F, Ghelli A. 1999. Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Weather Forecast.* **14**: 168–189.

Buizza R, Bidlot J-R, Wedi N, Fuentes M, Hamrud M, Holt G, Vitart F. 2007. The new ECMWF VAREPS. *Q. J. R. Meteorol. Soc.* **133:** 681–695.

Dee DP and Coauthors. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137:** 553–597.

Fall S, Niyogi D, Mohanty UC, Kumar A. 2007. Application of weather prediction models for hazard mitigation planning: A case study of heavy off-season rains in Senegal. *Nat. Hazards* **41:** 227–243.

Fink, AH, Vincent DG, Ermert V. 2006. Rainfall types in the West African Soudanian Zone during the summer monsoon 2002. *Mon. Weather Rev*. **134:** 2143–2164.

Fröhlich L, Knippertz P. 2008. Identification and global climatology of upper-level troughs at low latitudes. *Meteorol. Z.* **17(5):** 565–573.

Hastenrath S. 1991. *Climate dynamics of the tropics*. Kluwer, Dordrecht, Netherlands, 512pp.

Hell K, Mutegi C. 2011. Aflatoxin control and prevention strategies in key crops of Sub-Saharan Africa. *Afr. J. Microbiol. Res.* **5(5):** 459–466.

Huffman GJ, Adler RF, Morrissey M, Bolvin DT, Curtis S, Joyce R, McGavock B, Susskind J. 2001. Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeor.* **2:** 36–50.

Huffman, GJ, Adler RF, Bolvin DT, Gu G, Nelkin EJ, Bowman KP, Stocker EF, Wolff DB. 2007. The TRMM multi-satellite precipitation analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol*. **8:** 38–55.

Jobard I, Chopin F, Berges JC, Roca R. 2011. An intercomparison of 10-day satellite precipitation products during West African monsoon. *Int. J. Remote Sensing* **32:** 2353–2376.

Joliffe IT, Stephenson DB. 2003. *Forecast verification: a practitioner's guide in atmospheric science*. Wiley and Sons, Chicheseter, UK, 240 pp.

Knippertz, P., 2005: Tropical–extratropical interactions associated with an Atlantic tropical plume and subtropical jet streak. *Mon. Weather Rev.* **133:** 2759–2776.

Knippertz P, Fink AH. 2008. Dry-season precipitation in tropical West Africa and its relation to forcing from the extratropics. *Mon. Weather Rev.* **136:** 3579–3596.

Knippertz P, Fink AH. 2009. Prediction of dry-season precipitation in tropical West Africa and its relation to forcing from the extratropics. *Weather Forecast.* **24:**1064–1084.

Knippertz P, Martin JE. 2005. Tropical plumes and extreme precipitation in subtropical and tropical West Africa. *Q. J. R. Meteorol. Soc.* **131:** 2337–2365.

Lamptey B. 2008. Comparison of gridded multisatellite rainfall estimates with gridded gauge rainfall over West Africa. *J. Appl. Meteorol. Climatol*. **47:** 185–205.

Nicholson S and Coauthors. 2003a. Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part 1: Validation of GPCC rainfall product and pre-TRMM satellite and blended products. *J. Appl. Meteorol.* **42:** 1337–1354.

Nicholson S and Coauthors. 2003b. 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. Meteorol.* **42:** 1355–1368.

Meier F, Knippertz P. 2009. Dynamics and predictability of a heavy dry-season precipitation event over western Africa - Sensitivity experiments with a global model. *Mon. Weather Rev.* **137:** 189–206.

Paeth H, Fink AH, Pohle S, Keis F, Mächel H, Samimi C. 2011. Meteorological characteristics and potential causes of the 2007 flood in sub-Saharan Africa. *Int. J. Climatol*. **31:** 1908–1926.

Palmer TN, 1999. Predicting uncertainty in forecasts of weather and climate. *ECMWF Technical Memorandum No. 294*, available online: http://www.ecmwf.int/newsevents/training/rcourse_notes/pdf_files/Uncertainty_prediction.pdf

Parker D, Good E, Chadwick R. 2011. Reviews of observational data available over Africa for monitoring, attribution and forecast evaluation. *Hadley Centre Technical Note* **86**, available at: www.metoffice.gov.uk/media/pdf/b/o/HCTN 86 revised.pdf.

Pospichal B, Bou Karam D, Crewell S, Flamant C, Hünerbein A, Bock O, Säıd F. 2010. The diurnal cycle of lower boundary-layer wind in the West African monsoon. *Q. J. R. Meteorol. Soc.* **136(s1):** 92–106.

Roca R, Chambon P, Jobard I, Kirstetter P-E, Gosset M, Bergès JC. 2010. Comparing satellite and surface rainfall products over West Africa at meteorologically relevant scales during the AMMA campaign using error estimates. *J. Appl. Meteorol. Climatol*. **49:** 715–731.

Sultan B, Labadi K, Guégan J-F, Janicot S. 2005. Climate drives the meningitis epidemics onset in West Africa. *PLoS Med.* **2:** 43–49.

Thomson MC, Molesworth AM, Mamodou, HD, Yameogo KR, Belanger F, Cuevas LE. 2006. Potential of environmental models to predict meningitis epidemics in Africa. *Trop. Med. Int. Health* **11:** 781–788.

Uppala S and Coauthors. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131:** 2961–3012.

Waliser DE and Coauthors. 2012. The "Year" of Tropical Convection (May 2008 to April 2010): Climate variability and weather highlights. *Bull. Amer. Meteorol. Soc.*, in press, doi:10.1175/2011BAMS3095.1.

Wernli H, Paulat M, Hagen M, Frei C. 2008. SAL – A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Weather Rev.* **136:** 4470–4487.

Xie P-P, Janowiak JE, Arkin PA, Adler RF, Gruber A, Ferraro R, Huffman GJ, Curtis S. 2003. GPCP pentad precipitation analyses: An experimental data set based on gauge observations and satellite estimates. *J. Climate* **16:** 2197–2214.

Tables

Table I: List of identified extreme pentads in the three datasets considered. Light shading indicates agreement between TRMM and GPCP, intermediate shading between EPS and TRMM and dark shading between all three datasets.

EPS	TRMM	GPCP
14–18 MAR 1999	06–10 DEC 1998	05–09 DEC 1998
17–21 NOV 1999	15–19 FEB 1999	30 DEC 1999–03 JAN 2000
23–27 NOV 1999	25–29 NOV 1999	16–20 MAR 2000
01–05 JAN 2000	31 DEC 1999–04 JAN 2000	11–15 NOV 2001
05–09 JAN 2002	16–20 MAR 2000	23–27 NOV 2001
03–07 NOV 2002	11–15 NOV 2001	06–10 JAN 2002
06–10 DEC 2002	06–10 JAN 2002	07–11 MAR 2002
07–11 NOV 2003	07–11 MAR 2002	13–17 FEB 2003
19–23 JAN 2004	05–09 NOV 2003	05–09 NOV 2003
01–05 MAR 2004	19–23 JAN 2004	16–20 NOV 2003
26–30 JAN 2006	27 FEB–03 MAR 2004	19–23 JAN 2004
13–17 FEB 2006	26–30 JAN 2006	26 FEB–02 MAR 2004
07–11 FEB 2007	01–05 DEC 2008	22–26 DEC 2004
03–07 DEC 2008	14–18 DEC 2008	26–30 JAN 2006
15–19 DEC 2008	02–06 FEB 2009	20–24 MAR 2008
16–20 FEB 2009	25–29 MAR 2009	13–17 DEC 2008
16–20 MAR 2009		01–05 FEB 2009

Table II: Results from the event-based evaluation. Situations where the two observational datasets disagree can be interpreted in two different ways. For more details see section 3.2.

Event in	All three	EPS TRMM	EPS GPCP	EPS	TRMM GPCP	TRMM	GPCP
Number	7	2	0	8	5	2	5
Interpretation	hit	hit / false alarm		false alarm	miss	miss / correct neg.	

Figures

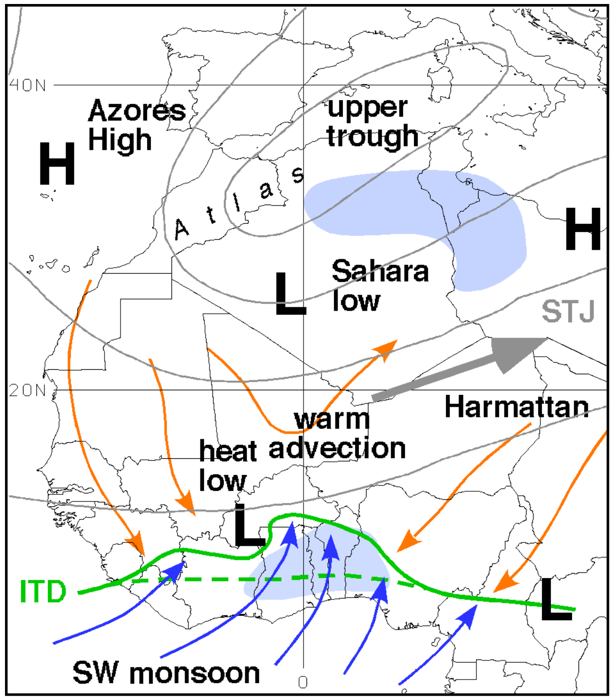

Figure 1. Schematic depiction of the mechanism of tropical-extratropical interaction responsible for many dry-season precipitation events (modified version of Figure 14 in KF08). Thin solid lines depict upper-level geopotential height with the subtropical jet (STJ) marked with a thick arrow. Thin arrows depict the predominantly northerly (southerly) low-level flow of dry Saharan (moist tropical) air masses. 'H' and 'L' mark high- and low-pressure centres, respectively. The thick dashed and solid lines show the climatological and the actual positions of the Intertropical Discontinuity (ITD). Areas of precipitation are shaded.

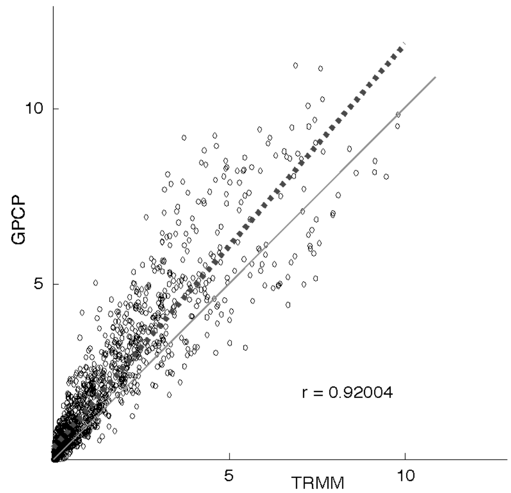

Figure 2. Comparison between GPCP and TRMM rainfall estimates. The scatter diagram shows pentad precipitation (mm) for each day during the 11 dry seasons (November–March) 1998/99–2008/09 averaged over the study area (7–15°N, 10°W–10°E). The linear regression line (thick dashed), the diagonal (thin solid) and the linear correlation coefficient are also given in the plot.

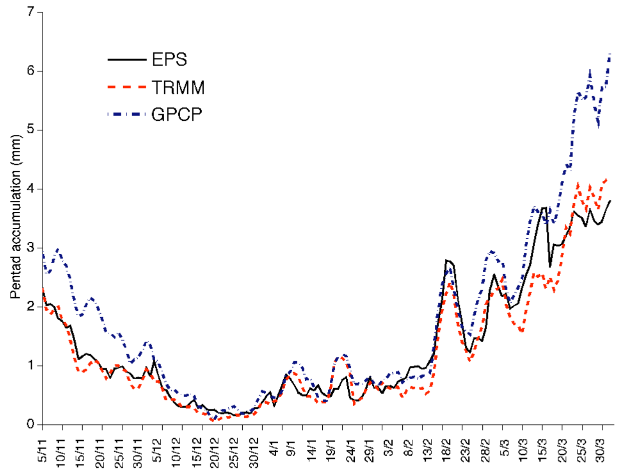

Figure 3. Mean seasonal cycle in EPS ensemble means, TRMM and GPCP data. Shown are pentad precipitation values for each dry-season day (November–March) averaged temporally over 1998/99–2008/09 and spatially over the study area (7–15°N, 10°W–10°E). The dates underneath the x-axis give the end date of the respective pentad.

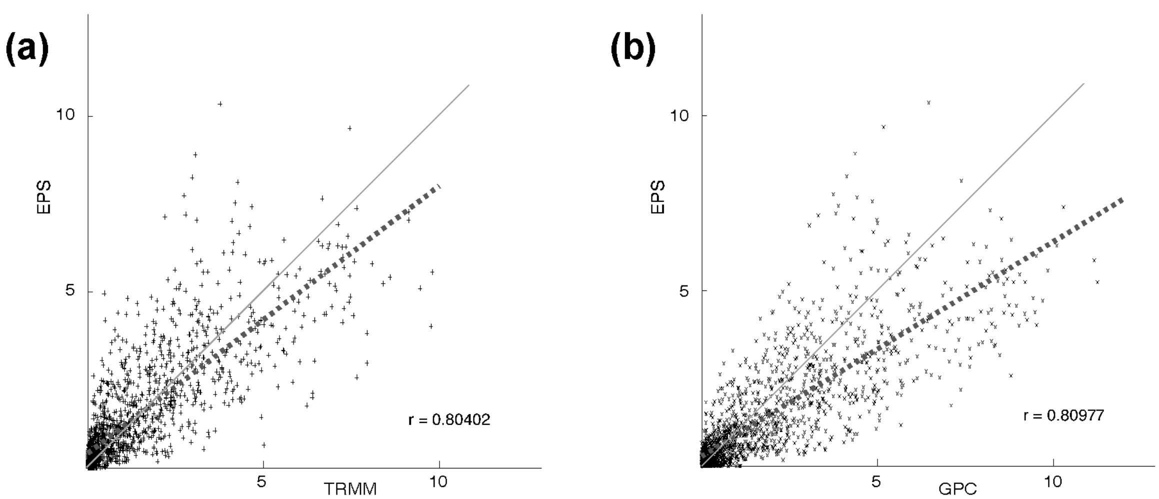

Figure 4. As Figure 2 but for EPS ensemble mean precipitation forecasts and (a) TRMM and (b) GPCP.

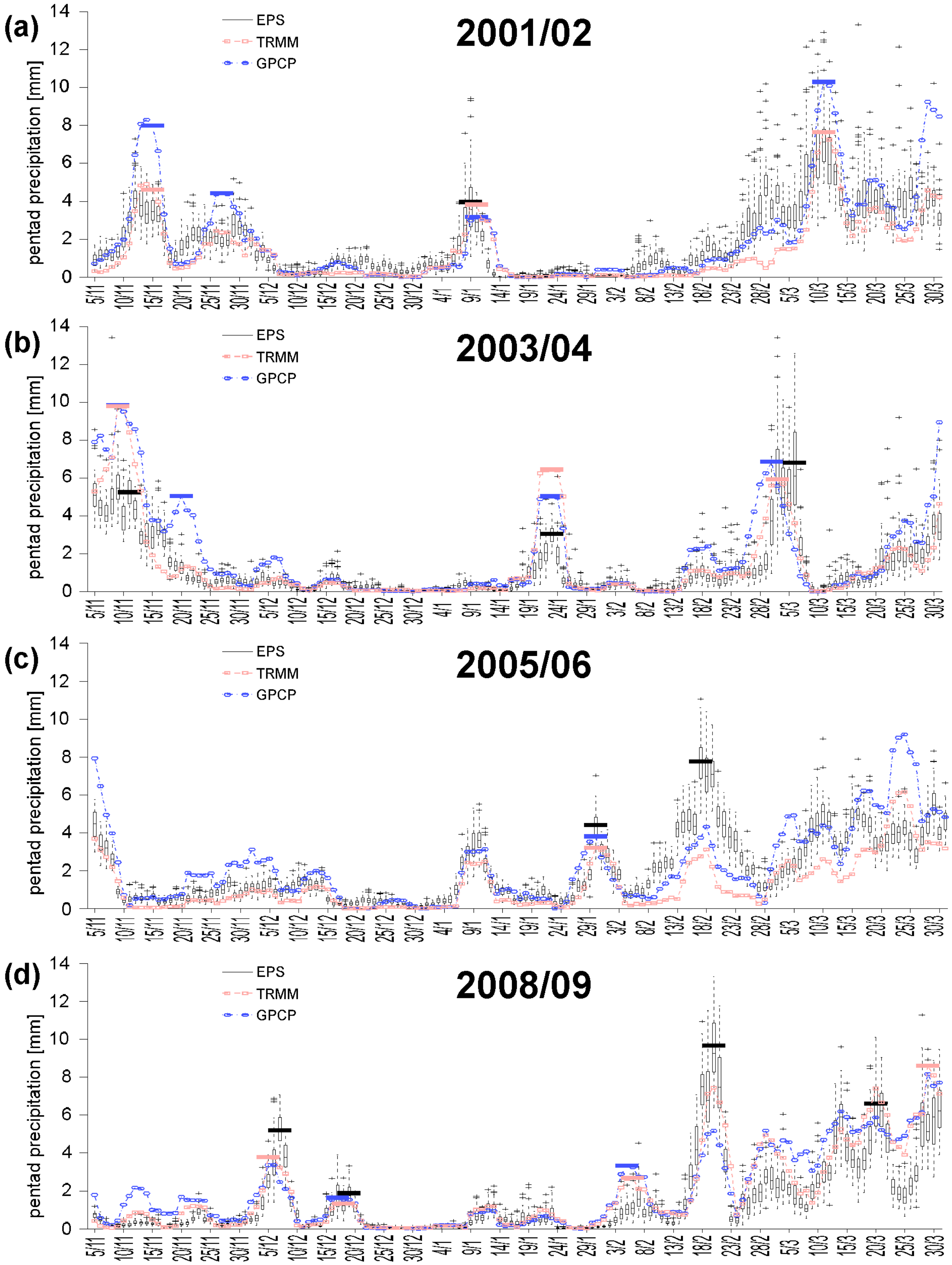

Figure 5. Time series of selected dry seasons (November–March): (a) 2001/02, (b) 2003/04, (c) 2005/06 and (d) 2008/09. Shown are daily pentad precipitation values

averaged over the study area (7–15°N, 10°W–10°E) with the end dates given underneath the x-axis. EPS forecasts are depicted as box-and-whisker plots (the box indicates the interquartile range, the central line is the median, and ‘+’ represent outliers). Ellipses (rectangles) represent GPCP (TRMM) observations. The heavy horizontal lines represent pentad rainfall for extreme events, as further detailed in section 3.2.

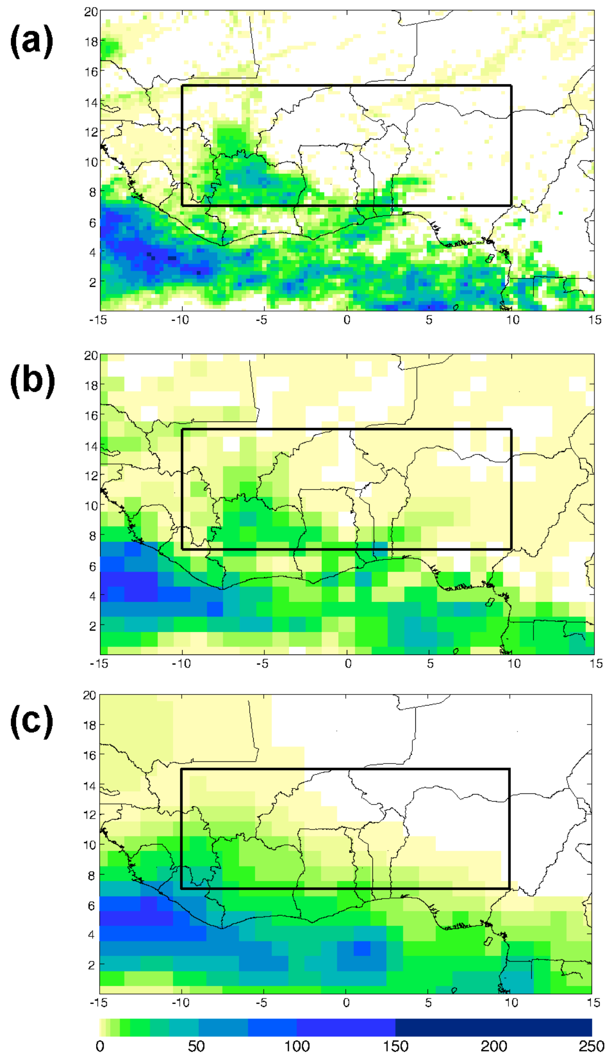

Figure 6. Example case I. Pentad precipitation (mm) for 5–9 January 2002 showing (a) TRMM, (b) GPCP and (c) EPS mean. The black boxes indicate the study area used for averaging in Figures 2–5.

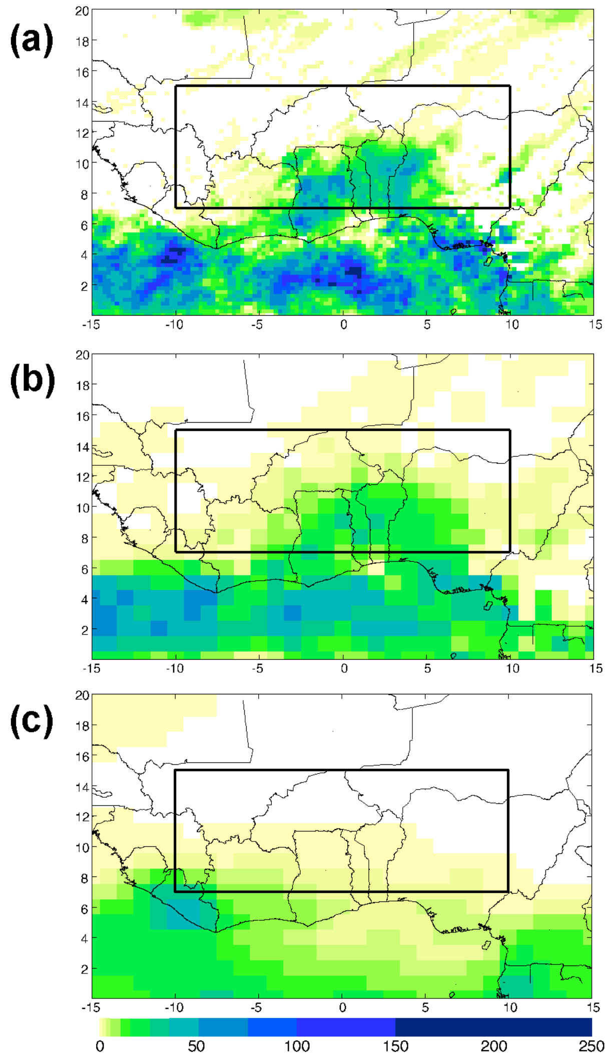

Figure 7. Example case II. As Figure 6 but for 19–23 January 2004.

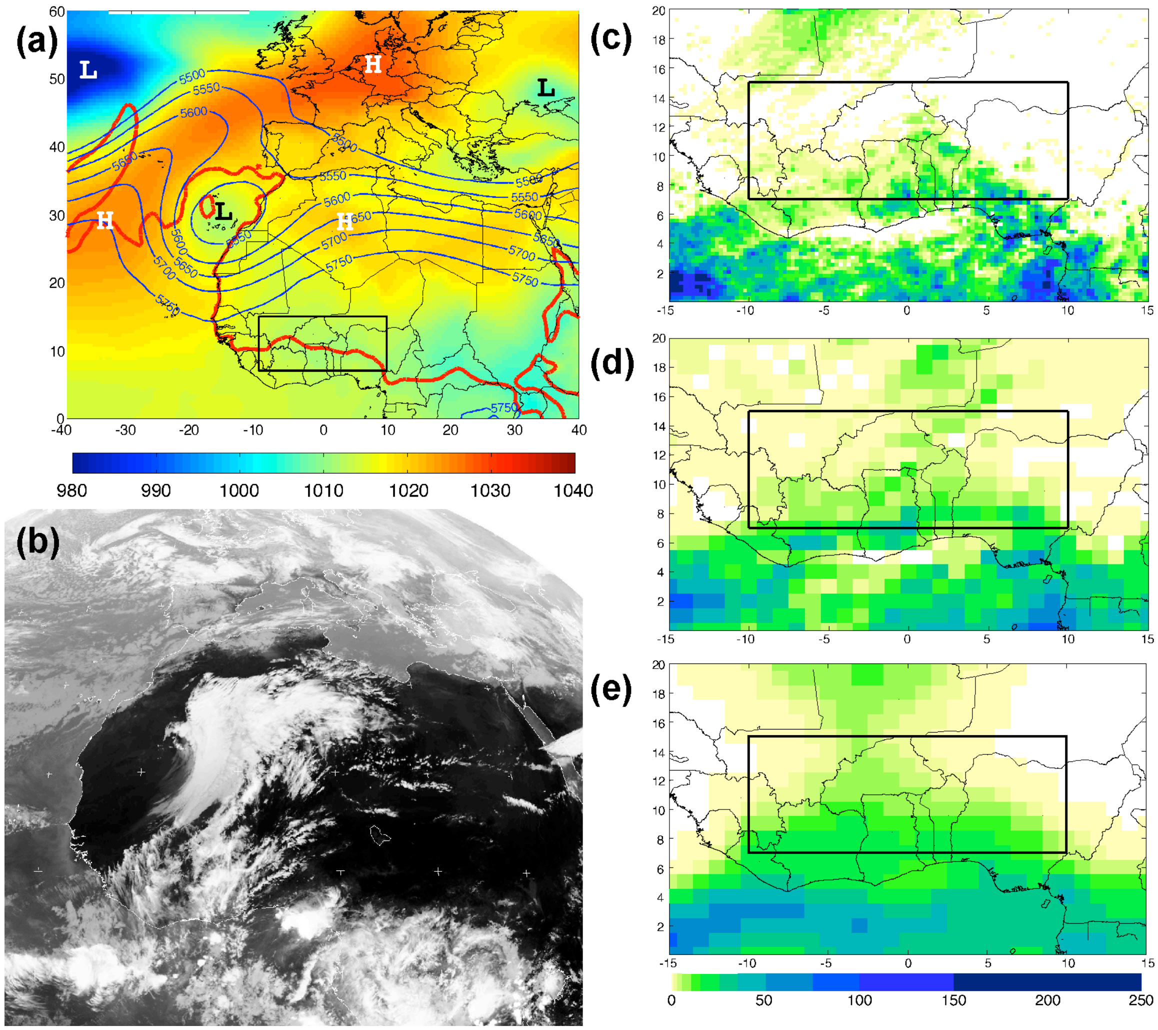

Figure 8. Example case III. (a) Geopotential height at 500 hPa (contours every 50m), mean sea-level pressure (hPa, shading) and the position of the ITD as indicated by the 14°C contour of the 2m dewpoint (dashed line) at 0000 UTC 15 February 2006. 'H' and 'L' mark high- and low-pressure centres, respectively. (b) Meteosat infrared satellite image at 1200 UTC 17 February 2006. (c)–(d) as in Figure 6 but for 13–17 February 2006.

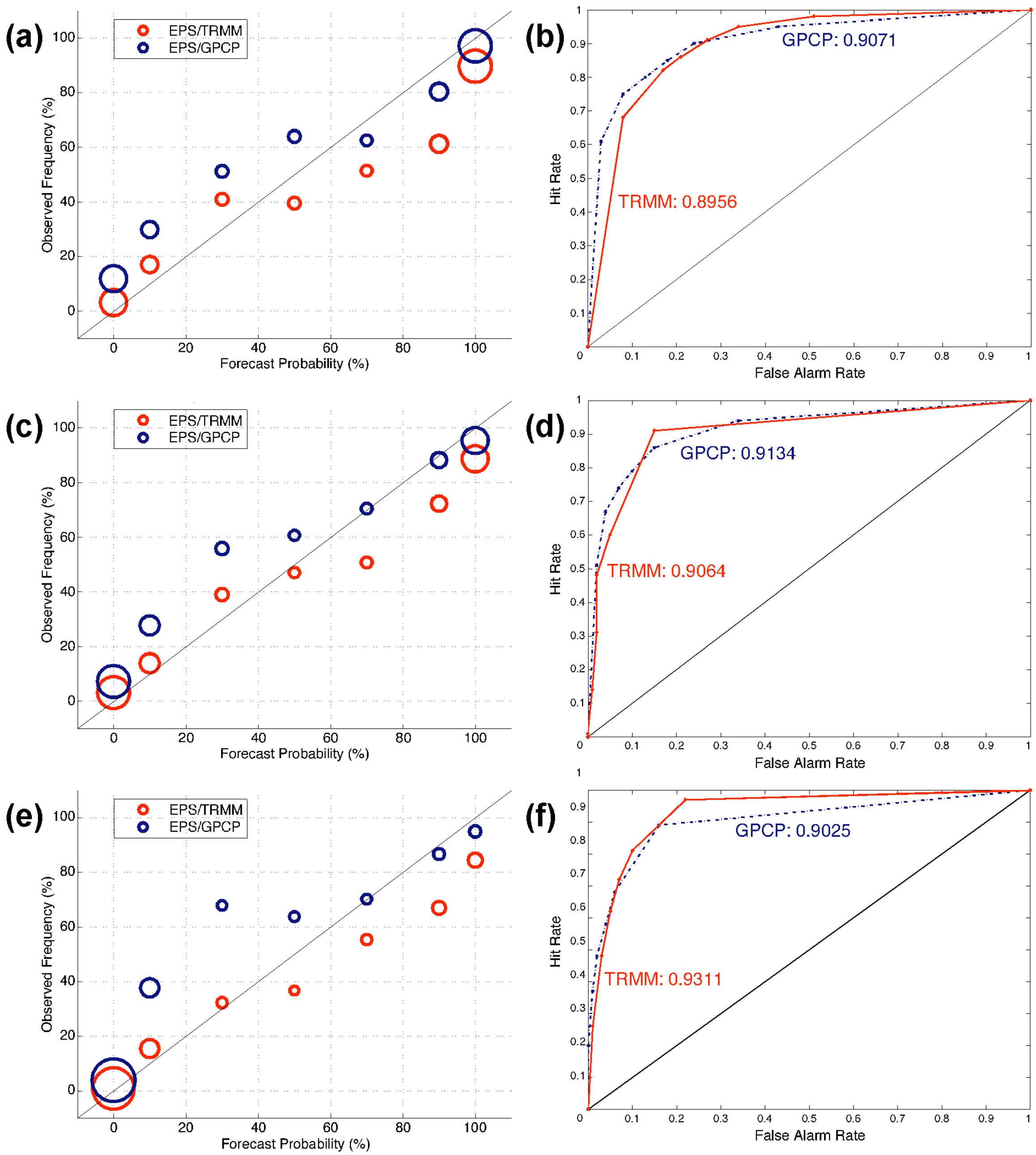

Figure 9: Probabilistic forecast evaluation. Reliability (left) and ROC (right) diagrams for TRMM and GPCP pentad precipitation values averaged over the study area (7–15°N, 10°W–10°E) for the 11 dry seasons 1998/98–2008/09. Thresholds are (a), (b) 0.5 mm, (c), (d) 1 mm and (e), (f) 3 mm. The size of the circles in the reliability diagrams indicates the number of cases in each bin (e.g. 1088 for the largest circles in Fig. 9e). ROC areas are also given in the plots.