Meteorologische Zeitschrift, Vol. **26**, No. 1, 93–110 (published online September 8, 2016)
© 2016 The authors

Predictability of convective precipitation for West Africa: verification of convection-permitting and global ensemble simulations

Vera Maurer*, Norbert Kalthoff and Leonhard Gantner

Institut für Meteorologie und Klimaforschung (IMK-TRO), Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany

(Manuscript received August 28, 2015; in revised form January 11, 2016; accepted May 11, 2016)

Abstract

Within the framework of this investigation, convection-permitting (CP) ensemble forecasts were generated for West Africa by combining different initial and lateral boundary conditions (IBCs) with perturbations that address the uncertainty of land-surface atmosphere interactions (land-surface perturbations). For a multi-analysis setup, IBCs were taken from model analyses of different global models; for a single-model setup, they were selected from the ensemble system of the European Centre for Medium-range Weather Forecasts (ECMWF). The different ensemble setups were assessed using common probabilistic scores as well as by spatial forecast verification of precipitation generated mainly by convective systems during the West African monsoon season. Additionally, it was investigated whether the CP ensemble forecasts were superior to the ECMWF ensemble forecasts.

Probabilistic scores were higher for the single-model ensemble than for the multi-analysis setup, but the latter displayed a larger dispersion and more extreme scenarios. From this, it is concluded that the different model analyses can differ strongly from each other. The land-surface perturbations were able to generate sufficient complementary spread. While the CP simulations showed a stronger negative precipitation bias in the southernmost region near the Guinean coast, the ECMWF simulations exhibited a negative bias further north in the Sahel region, where larger convective systems occur less frequently. Not in all cases did the CP ensemble versions produce better probabilistic scores than the global ensemble forecasts, but they yielded larger spread and less underdispersion. Rank histograms, though, were also influenced by the different structure of the precipitation patterns of the global and CP forecasts. Scores improved when using a later version of the CP model as well as with the skill of the global ensemble forecasts used as IBCs. Altogether, the proposed realization of CP ensemble forecasts is found to be suited for the prediction of convective precipitation in West Africa.

Keywords: Sahel, COSMO model, ECMWF EPS, soil moisture, land-surface perturbations, soil-moisture ensemble

1 Introduction

In tropical regions like West Africa, which are dominated by a monsoon, precipitation is mainly generated by convective systems (Houze, 1981; Le Barbé and Lebel, 1997; Mathon et al., 2002; Frappart et al., 2009). Forecast of convective precipitation is only partly understood and its forecast by models still is a great challenge (Lebel et al., 2000; Guichard et al., 2010). Therefore, it is reasonable that model simulations for West Africa during the monsoon season have significantly improved with increasing horizontal resolution of the forecast models and especially with the step towards convection-permitting (CP) model resolutions (e.g. Beucher et al., 2014; Pearson et al., 2014; Birch et al., 2014). On the other hand, West Africa is a region of very sparse observations (Parker et al., 2008) and it was demonstrated that assimilation of additional information may result in distinct improvements of forecasts (Tompkins et al., 2005; Karbou et al., 2010; Agustí–Panareda et al., 2010). Convection-permitting ensemble simulations for West Africa can incorporate this analysis uncertainty by using different global model analyses as initial and boundary conditions (IBCs).

Over the last years, limited-area and also CP ensemble systems were developed for operational purposes (e.g. Bowler et al., 2008; Gebhardt et al., 2011; Montani et al., 2011; Vié et al., 2011). However, high-resolution (very) short-range ensemble simulations (Iversen, 2011) for regions where precipitation is mainly generated by convective systems are not very widespread yet. For selected convective cases in the United States, such ensemble simulations have been made by e.g. Schumacher et al. (2013). Luo and Chen (2015) investigated the ensemble prediction of a heavy-rain producing mesoscale convective system (MCS) in eastern China. Keil et al. (2014) used an objective method to select, from available forecast days of

*Corresponding author: Vera Maurer, Institut für Meteorologie und Klimaforschung (IMK-TRO), Karlsruher Institut für Technologie (KIT), Postfach 3640, 76021 Karlsruhe, Germany, e-mail: vera.maurer@kit.edu

DOI 10.1127/metz/2016/0728
© 2016 The authors
Gebrüder Borntraeger Science Publishers, Stuttgart, www.borntraeger-cramer.com

the operational convection-permitting ensemble system for Germany (COSMO-DE-EPS), days not dominated by larger-scale forcing. They found that the predictability of precipitation was lower for these days than for all others.

As far as we know, the only study for which regional or mesoscale ensemble forecasts for West Africa were performed was published by Torn (2010). He used an ensemble Kalman filter to produce different model analyses and analyzed the forecast sensitivity with a focus on the development of African Easterly Waves (AEWs). The forecast was sensitive to meridional wind in the lower to middle troposphere as well as to the vertical distribution of equivalent potential temperature. The present paper focuses on the short-range forecast of convective precipitation rather than on the forecast sensitivity of AEWs.

Moist convection in the Sahel region is triggered with higher probability in regions where the sensible heat flux and turbulent mixing are strong (Taylor et al., 2012). This relationship was also confirmed by model case studies (Gantner and Kalthoff, 2010; Klüpfel et al., 2012). Land-surface perturbations, e.g. by varying initial fields of soil moisture, can be used to generate ensemble variability; they influence forecasts by the partitioning of available energy at the Earth's surface into the turbulent fluxes of sensible and latent heat (Kohler et al., 2010). Maurer et al. (2015) found that land-surface perturbations, i.e. variations of initial soil moisture as well as of the static field of soil type, could generate as much ensemble spread as atmospheric perturbations and that CP simulations can provide suitable short-range precipitation forecasts for the Sahel region. Tennant and Beare (2014) showed that perturbations of the soil moisture content could also have benefits for ensemble systems at mid-latitudes. Thus, for this investigation, the land-surface perturbations described by Maurer et al. (2015) are considered an appropriate method to generate ensemble spread. They are combined with the atmospheric perturbations to form different configurations.

The aim of this study is the assessment of the skill of ensemble forecasts of convective precipitation for West Africa using different ensemble setups; in particular, CP ensemble simulations using a multi-analysis and a single-model setup as well as global ensemble simulations for a short forecast range:

(1) By comparing the different setups of CP ensemble versions, it is examined whether the use of a higher number of IBCs can further increase the ensemble spread, as was discussed for example by Marsigli et al. (2014). They found that a single-model limited-area ensemble using a large number of global simulations performs better than a multi-model ensemble using deterministic forecasts from three different global models.

(2) The CP and the global ensemble forecasts will be compared to analyze the expected gain of ensemble simulations at a CP resolution. The reason is that for West Africa, it can be assumed that shortcomings of convection parameterizations as described by Bechtold et al. (2004), e.g., strongly affect the precipitation forecast by global models. Marsham et al. (2013) showed that simulations performed better in this region when using no convection parameterization, even at horizontal resolutions that were substantially coarser than 4 km; Weisman et al. (1997) gave 4 km as the upper limit of horizontal resolution, at which the "mesoscale structure and evolution" of convective systems are reproduced in an adequate way.

The verification of precipitation forecasts is an issue that recently was addressed by many authors; especially for high-resolution forecasts, spatial verification methods have been developed for example by Casati et al. (2004), Roberts and Lean (2008), Wernli et al. (2008), or Keil et al. (2014). These are advantageous in comparison to classical scores. On the other hand, they are not designed specially for probabilistic verification. Hamill et al. (2000) proposed a minimum of metrics that should be used for the evaluation of ensemble systems: Rank histograms, ROC diagrams, and reliability diagrams, containing the relevant attributes. The different ensemble versions will thus be compared by these scores, using gridded observations for the computations. Schwartz et al. (2010) proposed a neighborhood approach to combine the fuzzy verification applied in the fractional skill score (FSS) according to Roberts and Lean (2008) with the classical verification of probability forecasts. This approach was used for the computation of ROC and reliability diagrams. Additionally, an object-based verification method from the set of spatial verification methods will be applied, which is independent of the single grid points.

In the following section, an overview of all evaluated ensemble versions, of the synoptical conditions of the forecast periods as well as of the model setup is given. Thereafter, the methods applied for the verification and comparison of the ensemble versions are described in Section 3. The main verification results follow in Section 4, structured into the comparison of the single-model and multi-analysis ensemble version and of the CP and the global ensemble version. The final section provides a summary and conclusions.

2 Overview of evaluated ensemble simulations and forecast periods

2.1 Ensemble simulations for 2006

The CP ensemble simulations were prepared using two different setups: A multi-analysis setup as in Maurer et al. (2015) and a single-model setup. For the multi-analysis setup, three different global analyses were taken as IBCs: (1) Operational model analyses of the European Centre for Medium-range Weather Forecasts (ECMWF) for the respective simulation period,

(2) ECMWF re-analyses that were specially prepared for the African Monsoon Multidisciplinary Analyses (AMMA) project, which are only available for the monsoon season of 2006 (Agustí-Panareda et al., 2010), and (3) GME analyses (i.e. analyses of the global model of Deutscher Wetterdienst, DWD). Each of these three were taken as IBCs for four simulations by combining them with four different initial conditions for the soil and the land surface, respectively. These initial conditions were mainly different soil moisture fields for the three uppermost levels taken from the AMMA re-analyses, the GME analyses as well as from satellite observations. For a better consistency with the soil model as well as to increase the spread, the soil-type distribution of the integrated forecasting system (IFS) of the ECMWF was used for initialization with soil moisture from the AMMA re-analyses. Satellite soil moisture was adapted by a technique proposed by Reichle and Koster (2004), in combination with two versions of soil heat conductivity used by the soil model. For more details on these perturbations, see Maurer et al. (2015). Hereafter, this ensemble setup with a combination of IBCs taken from different global model ANAlyses and the four land-surface perturbations is referred to as **ANA**.

For the single-model setup, forecasts from the ECMWF ensemble prediction system (EPS) were used for IBCs for the same simulation period. As the ensemble of data assimilation (EDA) improves the forecast skill of the EPS especially in the tropical regions (Isaksen et al., 2010), it was important to use forecasts from ECMWF that employ this scheme. Due to the availability of the AMMA re-analyses for 2006 only, the ANA ensemble as well as the single-model ensemble had to be set up for this year. However, the EDA has been operational since June 2010 only, so that the operational EPS forecasts of 2006 could not be used. As it is computationally extremely expensive to run the EDA, it was only possible to set up ensemble forecasts with the global model of ECMWF (IFS Cycle 38r1, operational from June 2012 to June 2013) using "climatological" EDA perturbations, i.e. perturbations from another year that were added to the operational model analyses for the starting dates of the respective simulations. Operationally, 10 EDA perturbations exist, and 10 ensemble members were generated with the ECMWF EPS. To obtain 12 IBCs for the single-model setup, the IFS control forecast was used twice. The land-surface perturbations were combined randomly with the 12 IBCs. As a result, each land-surface perturbation was used for three simulations, as for the ANA ensemble. Different land-surface perturbations were combined with the two members using the control forecast for IBCs. The ensemble version generated by this combination of 11 different IBCs provided by the ECMWF model and four land-surface perturbations is referred to as **EC**.

The 10 + 1 IFS forecasts generated by using the climatological EDA were also verified and compared with the CP ensemble simulations. This 11-member ensemble was called **ECMWF 2006**. The horizontal resolution of the IFS forecasts is T639 (about 0.28 °).

2.2 Ensemble simulations for 2011

To benefit from the whole spread of the ECMWF EPS, further ensemble simulations were made for a time period for which the EDA was already operational. For detailed information on the ensemble prediction system of ECMWF, it is referred to Leutbecher and Palmer (2008), Isaksen et al. (2010), and Shutts et al. (2011). For the CP ensemble simulations, a time period was selected from the monsoon season of 2011, which was roughly similar to the one of 2006 in terms of the convective activity in the Sahel region.

As the ECMWF EPS consists of 50 + 1 ensemble members, a subset of 16 representative members was selected from these by a clustering method according to Molteni et al. (2001). This method is operationally used for the Limited Area Ensemble Prediction System (LEPS; Montani et al., 2011) developed within the Consortium for Small-scale Modeling (COSMO). It is a hierarchical clustering in which clusters are defined by complete linkage (Wilks, 2011, Ch. 15.2): Vectors consisting of two-dimensional fields of different normalized selected variables at selected time steps and at selected heights are generated for each ensemble member. These vectors are compared pairwise by computing the absolute value $|\mathbf{d}|$ of their difference. The two ensemble members with the smallest $|\mathbf{d}|$ are assigned to the first group. This procedure is repeated until the desired number of groups (the "clusters") is reached, defining $|\mathbf{d}|$ between groups as the largest possible distance between individual ensemble members of the respective groups.

We adapted the settings of Montani et al. (2011) to the synoptical conditions in West Africa and chose four variables (both horizontal wind components, specific humidity, and geopotential) on three pressure levels (925 hPa, 700 hPa, and 500 hPa) and the forecast hours 12, 36, and 48 for the clustering. These 16 ensemble members also were randomly combined with the land-surface perturbations, as for the EC 2006 version. The resulting ensemble version was termed **EC 2011**. The ensemble forecasts of EC 2011 were compared to the whole ECMWF EPS for the corresponding time period (hereafter called **ECMWF 2011**). An overview of all five ensemble versions is given in Table 1.

2.3 Forecast periods

For the CP simulations for both years, a 10-day period was selected. The selection criterion was that large convective systems had to occur in the Sahel region, where land-surface atmosphere interactions are most effective (e.g. Taylor et al., 2012). This is usually the case during July and August, when the monsoon rain belt is at its northernmost position. A short synoptical overview for the periods from 2006 and 2011 is given in Figure 1. In West Africa, easterly winds prevail in

Table 1: Overview of ensemble simulations made with the COSMO model and of evaluated ensemble forecasts from the Integrated Forecasting System (IFS) at ECMWF.

	type of forecasts	**perturbations**	**members**	**evaluated forecast period**
ANA (multi-analysis setup)	COSMO 4.18, two nested domains with grid-point distances of 14 km and 2.8 km	3 different IBCs (model analyses) combined with 4 land-surface perturbations	12	23/07–01/08 2006
EC (single-model setup)	as ANA	11 different IBCs (ECMWF 2006) with 4 randomly added land-surface perturbations	12	23/07–01/08 2006
ECMWF 2006 (global ensemble)	ECMWF EPS (experimental IFS forecasts) with grid-point distance of ≈ 30 km (T639)	climatological EDA	10 + 1	23/07–01/08 2006
EC 2011 (single-model setup)	COSMO 4.26, two nested domains as for 2006	16 different IBCs (16 members of ECMWF 2011) with 4 randomly added land-surface perturbations	16	12/07–21/07 2011
ECMWF 2011 (global ensemble)	ECMWF EPS (operational 1200-UTC IFS forecasts, evaluated forecast hours: 24–48) with grid-point distance of ≈ 30 km (T639)	singular vectors and EDA	50 + 1	12/07 – 21/07 2011

the mid-troposphere due to strong meridional baroclinicity between the region of the Saharan heat low and the cooler monsoon air further south. The wind maximum at 700–600 hPa is the so-called African Easterly Jet (AEJ), which makes convective systems in the Sahel propagate westwards as is evident from both zonal Hovmöller diagrams of precipitation for this region (Figure 1a and b for 2006 and 2011, respectively). During both 10-day periods, about three to four rain bands can be observed. They alternate with rain-free times at a frequency of about two to three days. A typical synoptical feature in West Africa are AEWs (e.g. Reed et al., 1977). These disturbances of the AEJ are subject to interactions with the convective systems there (e.g. Thorncroft et al., 2008). Convection is often observed in front of or near the axis of an AEW trough, i.e. at locations, where the wind at 600 hPa has northerly components or changes from northerly to southerly ones. This is also evident from the zonal Hovmöller diagrams for both time periods (Figure 1a and b). On the large scale, it is also the northward penetration of the moist and cool monsoon air, which is related to convective activity. The humidity and zonal wind in the monsoon layer, i.e., the lower part of the troposphere, are good indicators of the northern boundary of the monsoon (meridional Hovmöller diagrams, Figure 1c and d), as westerly wind dominates in the monsoon layer and easterly wind further north (e.g. Janicot et al., 2008). For the two periods, the northern boundary of the south-westerly monsoon flow is also very similar (between 22 °–24 ° N), as is the transition region from very moist to drier air masses. The main difference is that the monsoon layer is by about 0.5–1 g kg^{-1} more humid for the time period in 2006 than for the one in 2011. From this, it can be concluded that the two periods are sufficiently similar to say that the results of the planned ensemble verification will not be influenced strongly by large-scale differences.

2.4 Model setup

The model setup was the same for all CP ensemble versions: For a first nest, simulations of a horizontal resolution of 14 km were conducted on a domain of 2 °–25 ° N/15 ° W–20 ° E and run for a forecast time of 48 h. They were started daily at 1200 UTC, i.e. at a time when deep convective activity is still weak. For the 14-km simulations, the COSMO model was used with its operational convection parameterization according to Tiedtke (1989). Into these, the simulations on the convection-permitting scale of 2.8 km were nested. They were initialized daily at 0000 UTC using the 12-h forecasts of the 14-km simulations. By initializing at 0000 UTC, a time of day was chosen when no boundary-layer mixing occurs. The 2.8-km simulations were run for 36 h. The simulation domain was from 8 °–22 ° N/8 ° W–12 ° E., i.e. about 1600×2200 km^2. The evaluation domain (Figure 2) was one degree smaller in each direction to exclude the areas that can be affected by numerical boundary effects. For the time period of 2006, the COSMO model version 4.18 was used and for 2011, the version 4.26. Due to a unification with the climate version of COSMO, a climatological albedo could be used in the version for 2011 which is derived from satellite measurements. In the older version, surface albedo only depended on the soil type, soil moisture, and vegetation. In comparison, the albedo was much lower in areas with sparse vegetation in the older version (e.g. 0.26 as compared to 0.4 in the Sahara). This may influence mean precipitation sums in the respective area, but it should not affect the probabilistic scores.

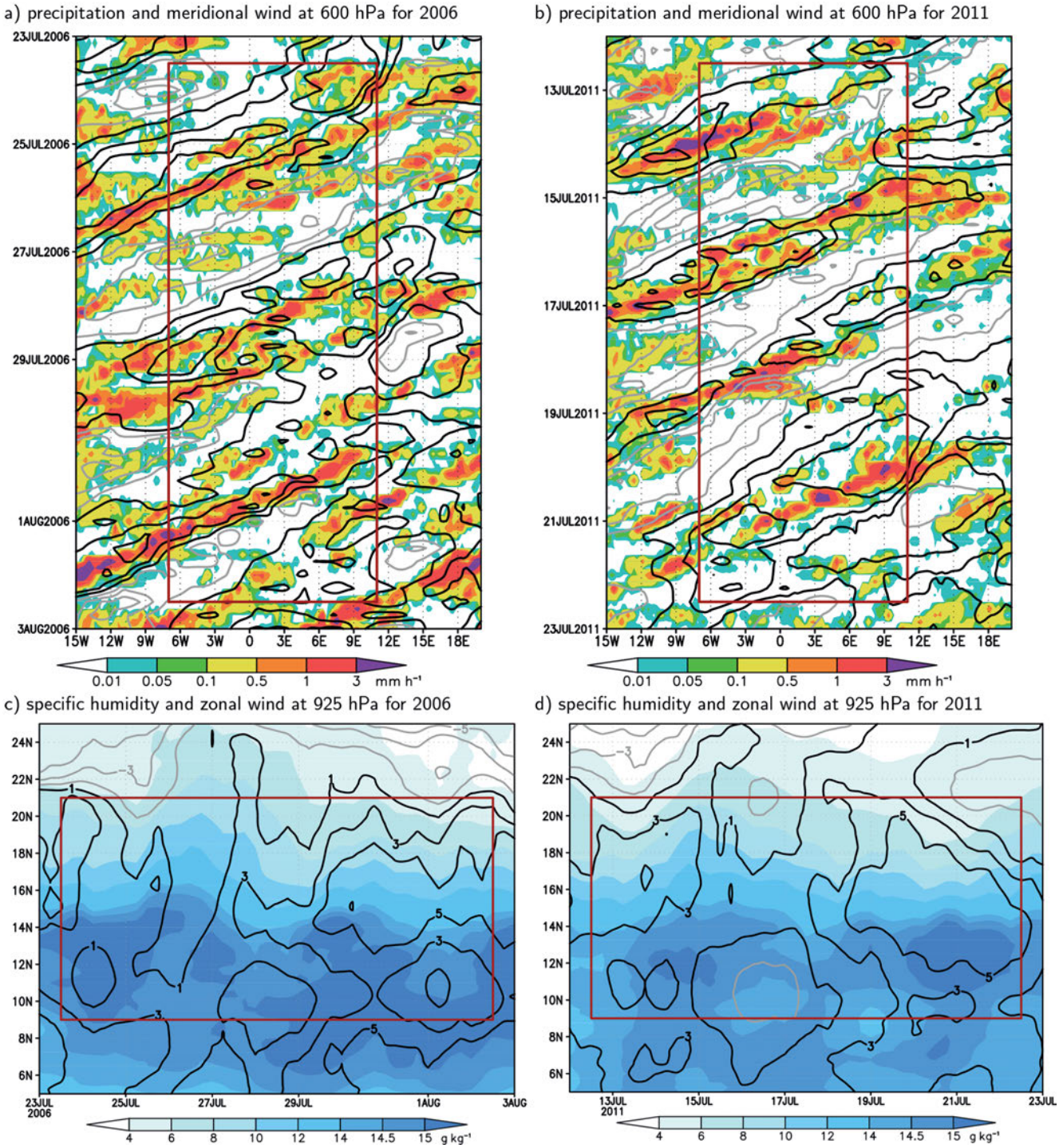

Figure 1: (a) and (b): Zonal Hovmöller diagrams (average 9 °–17 °N) of precipitation (TMPA; shading) and of the meridional wind component at 600 hPa (ECMWF AMMA re-analyses for 2006 (a), operational analyses for 2011 (b); gray isolines for southerly wind, black for northerly, intervals of 2 m s^{-1} from −5 m s^{-1} to 5 m s^{-1}). (c) and (d): Meridional Hovmöller diagrams (average 7 °W–11 °E) of 24-h running means of specific humidity (shading) and the zonal wind component at 925 hPa (gray isolines for easterly wind, black for westerly, same intervals as in (a) and (b); both variables were taken from AMMA re-analyses (a) and analyses (b), respectively). Dark red rectangles denote the zonal-temporal and the meridional-temporal extent of the evaluation domain and period, respectively.

3 Methods

3.1 Data used for forecast verification

Mainly two gridded data sets were used for verification: The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007) and the 10.8 µm brightness temperatures T_B as observed by Seviri on Meteosat Second Generation (MSG-1 for 2006 and MSG-2 for 2011). As precipitation in West Africa is almost exclusively produced by convective systems during the monsoon season, pre-

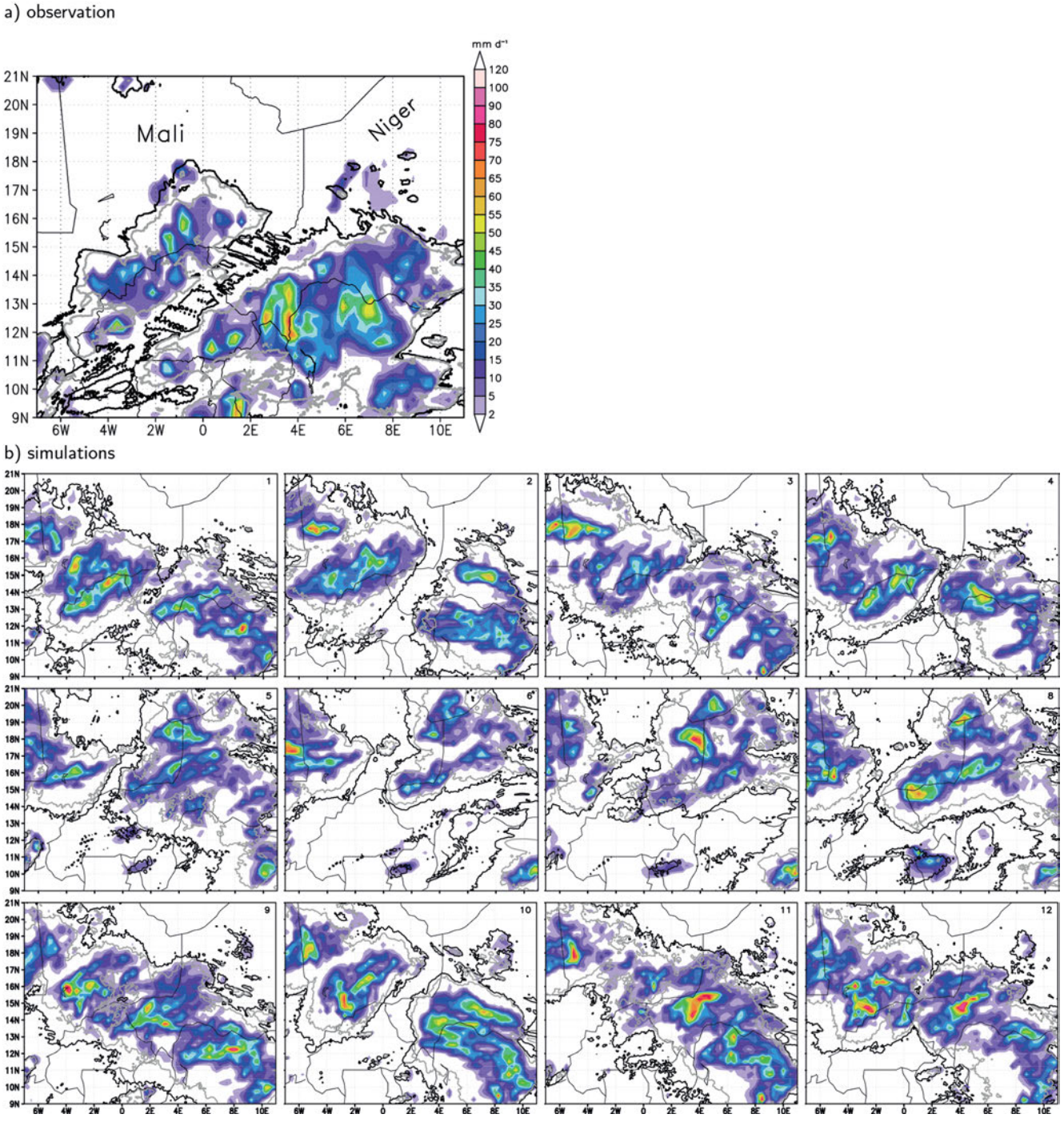

Figure 2: Example of observed (a) and simulated spatial precipitation distributions for the ANA ensemble (b; sums between 30 July 2006, 1200 UTC and 31 July 2006, 1200 UTC; shading) in the evaluation domain with isolines of 213 K and 233 K of daily minimum brightness temperature (gray and black, respectively).

cipitation patterns mainly indicate their paths and their intensities. The TMPA also has the advantage that it is completely independent of all simulations, because it is not part of the data assimilation at ECMWF or DWD. The additional verification of IR temperatures which can be derived from the model atmosphere against the MSG brightness temperatures was conducted to account for possible uncertainties in the TMPA precipitation product. All grid-point comparisons were computed on the grids of these two products. For this, modeled fields were interpolated to the horizontal grid-point distance of these two, i.e. to 0.25 ° and 0.05 °, respectively. Station observations were not used, as they would have been too sparse. Furthermore, the verification was mainly carried out for daily values. In Maurer et al. (2015) it was shown that the diurnal cycle is generally correctly reproduced by the CP simulations, which is not the case for the global simulations (Bechtold et al., 2004). In this way, mainly the daily precipitation location as well as the amount of precipitation are verified

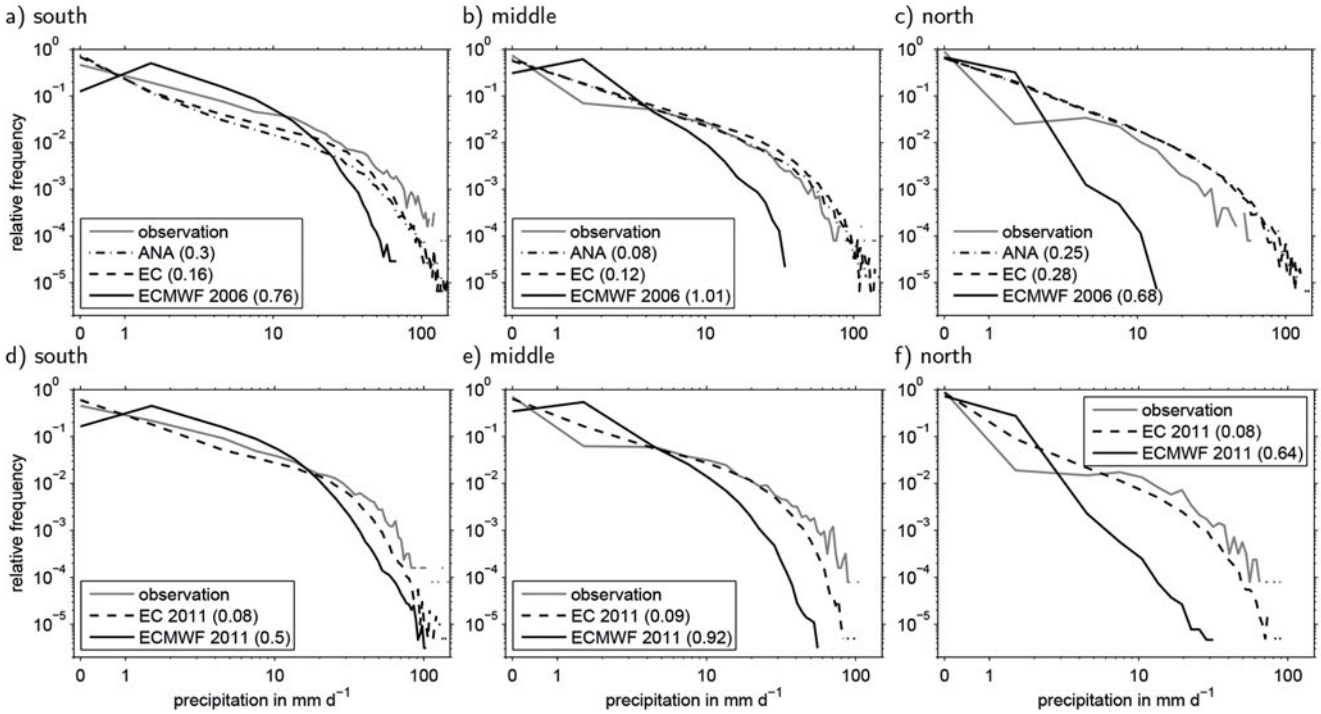

Figure 3: Frequency distributions of observed and simulated precipitation for 2006 (a–c) and 2011 (d–f); values of the Kullback-Leibler distance are given in parentheses.

and not the exact timing. For the 2.8-km simulations, only the forecast range from 12–36 h was evaluated so that the exact verification time periods are 23 July 2006, 1200 UTC to 02 August 2006, 1200 UTC and 12 July 2011, 1200 UTC to 22 July 2006, 1200 UTC. The "daily" sums below refer to the 1200–1200 UTC sums. The reason is the diurnal cycle of the convective systems producing most of the precipitation at about 1800 UTC and the least from about morning to noon, as shown e.g. in Maurer et al. (2015). As a measure that is comparable to the daily precipitation sums, the diurnal minimum of T_B was determined, which is more meaningful than daily averages of cloud-top temperatures. The fields of T_B cannot be averaged directly as they contain also much warmer temperatures at locations where no clouds are present. By computing the diurnal minimum, all regions that were covered by high clouds on the respective day can be identified easily without the need to define a prior threshold of T_B corresponding to high clouds. An example of a comparison of TMPA daily precipitation sum and daily minimum T_B is given for 30 July 2006 (Figure 2a). The same quantities can be derived from the simulations (Figure 2b). The precipitation and T_B patterns on this day indicate the location of an MCS in the south of Niger and a smaller one northwest of it in Mali. Subjectively, some of the ensemble members (1, 2, and 10) also reproduce two MCSs at roughly the correct location, while other members display larger location errors or different patterns. Although the precipitation sums seem to be low for some members, all of them display deep convection on this day.

3.2 Verification metrics

For most verifications, the evaluation domain was divided into three zonal subdomains of equal size (9 °–13 ° N, 13 °–17 ° N, and 17 °–21 ° N) that are approximately located within the three climate zones of the Guinean coast, the Sahel, and the southern Sahara (from south to north). The exact spatial specification may vary in literature (e.g. Fink, 2006; Lafore et al., 2011; Nicholson, 2013). In the following sections, they are referred to as the **southern**, the **middle**, and the **northern subdomain**, respectively. The separation into these three subdomains was done, because they display different mean precipitation characteristics: In the southern subdomain, largest precipitation sums were observed for the considered time periods (on spatial and temporal average 5.6 and 6.6 mm d^{-1} for the periods of 2006 and 2011, respectively, compared to 3.7/2.8 mm d^{-1} for the middle subdomain and 1.2/0.9 mm d^{-1} for the northern subdomain). The southern subdomain is the region with maximum convective activity. The middle subdomain also is a region with convective activity, but it is less frequent. No larger MCSs cross the northern subdomain and in less than 10 % of all cases, observed daily precipitation sums are larger than 2 mm d^{-1}. In comparison, this applies to almost 40 % and 25 % of all cases in the southern and middle subdomain, respectively.

The frequency distributions of precipitation were computed for the two time periods separately, using all members for each ensemble version (Figure 3). For an objective comparison with the observed frequency

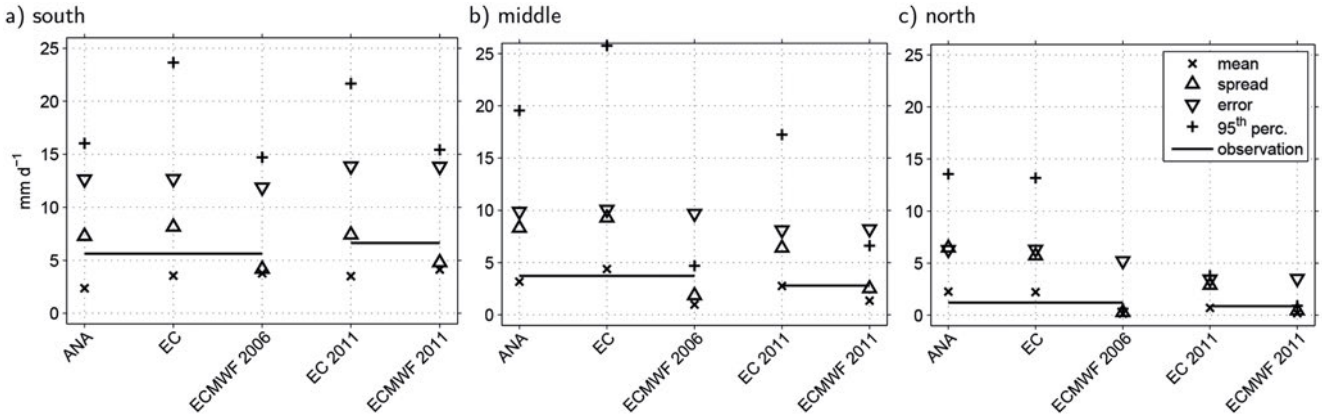

Figure 4: Mean precipitation sums for the 10 days for all ensemble versions in the three subdomains as well as observed sums, ensemble spread, and error (standard deviation and root-mean squared error of ensemble mean, respectively) and 95th percentile of the frequency distributions of precipitation for all ensemble versions.

distributions, the Kullback-Leibler distance KL (Kullback and Leibler, 1951) is analyzed, quantifying the difference between two distributions. It was calculated according to

$$KL(p,q) = \sum_{i=1}^{N} p_i \, log_2(p_i/q_i). \qquad (3.1)$$

Here, p and q are discrete frequency distributions of observed and forecast precipitation, respectively, each consisting of the relative frequencies p_i and q_i in bin i. For $i = 1$, all events with precipitation ≤ 0.1 mm d^{-1} were counted, i.e. daily sums of zero and very small values that are different from zero due to numerical reasons only. The upper edges of all remaining bins were $3(i-1)$ mm d^{-1}.

Basic statistics of the different ensemble versions are then shown on the basis of ensemble mean, spread, and error (Figure 4). These quantities were computed as area averages of the three subdomains, using the daily precipitation sums. The spread corresponds to the ensemble standard deviation on a grid-point basis and the error to the root-mean squared error of the ensemble mean compared to the observation. Additionally, the 95th percentile of the overall sample of daily precipitation data for each ensemble version is given to illustrate the range of values. The area-averaged observed precipitation sums are given, too. The distance between the ensemble mean and the observed sum is equivalent to the bias.

Rank histograms as discussed by Hamill (2001), e.g., were also computed on a grid-point basis for the diurnal precipitation sums in the three subdomains (Figure 5). Sums of less than 0.1 mm d^{-1} were set to zero. As an independent data set, the MSG data were used. Rank histograms of daily minimum T_B were computed for the high-resolution ensemble versions for 2006 setting all values above 233 K to the constant value of 235 K, as they do not contribute to the distribution of high clouds. As the relative frequencies are not strictly comparable for different ensemble sizes, two versions of rank histograms are given for the ensemble version ECMWF 2011: In the first version (Figure 5e), the rank histogram was computed using the representative members, i.e. the IBCs for the high-resolution ensemble; in the second version (Figure 5f), the whole ECMWF EPS was used. For a clearer presentation, frequencies were aggregated for every fourth rank.

ROC diagrams (e.g. Mason and Graham, 1999; Wilks, 2011) were computed for the three subdomains using the precipitation thresholds of 2 mm d^{-1}, 5.5 mm d^{-1}, 10 mm d^{-1}, and 15 mm d^{-1}. As can be seen in the precipitation fields for the example of 30 July 2006 (Figure 2), 2 mm d^{-1} is a relatively low precipitation sum which also occurs in the larger vicinity of MCSs. The threshold of 5.5 mm d^{-1} is the median of the overall sample of observed precipitation in the 10-day period chosen for 2006. The two largest thresholds confine the precipitation patterns generated by MCSs more closely. Scores for extreme thresholds were not compared as this would be disadvantageous for the global ensemble versions. The ROC areas (Figure 6a–c) were calculated as an area under the curve spanned by the individual points defined by the hit rate and false alarm rate found for different probability thresholds P_i. Different details of the ROC diagrams will be discussed to explain the ROC areas, but the individual diagrams are shown for selected examples only for the sake of brevity (e.g. Figure 7a, southern subdomain for a precipitation threshold of 5.5 mm d^{-1}).

ROC diagrams were computed using a neighborhood approach as proposed by Schwartz et al. (2010): Traditionally, the forecast probabilities needed for the computation are calculated independently for each grid point which equals the ratio of ensemble members forecasting the given event. Using the neighborhood approach, probabilities are computed as the ratio of forecast events by all members within a given neighborhood around the considered grid point. Considering the relatively low predictability of the MCSs, we chose a large neighbor-

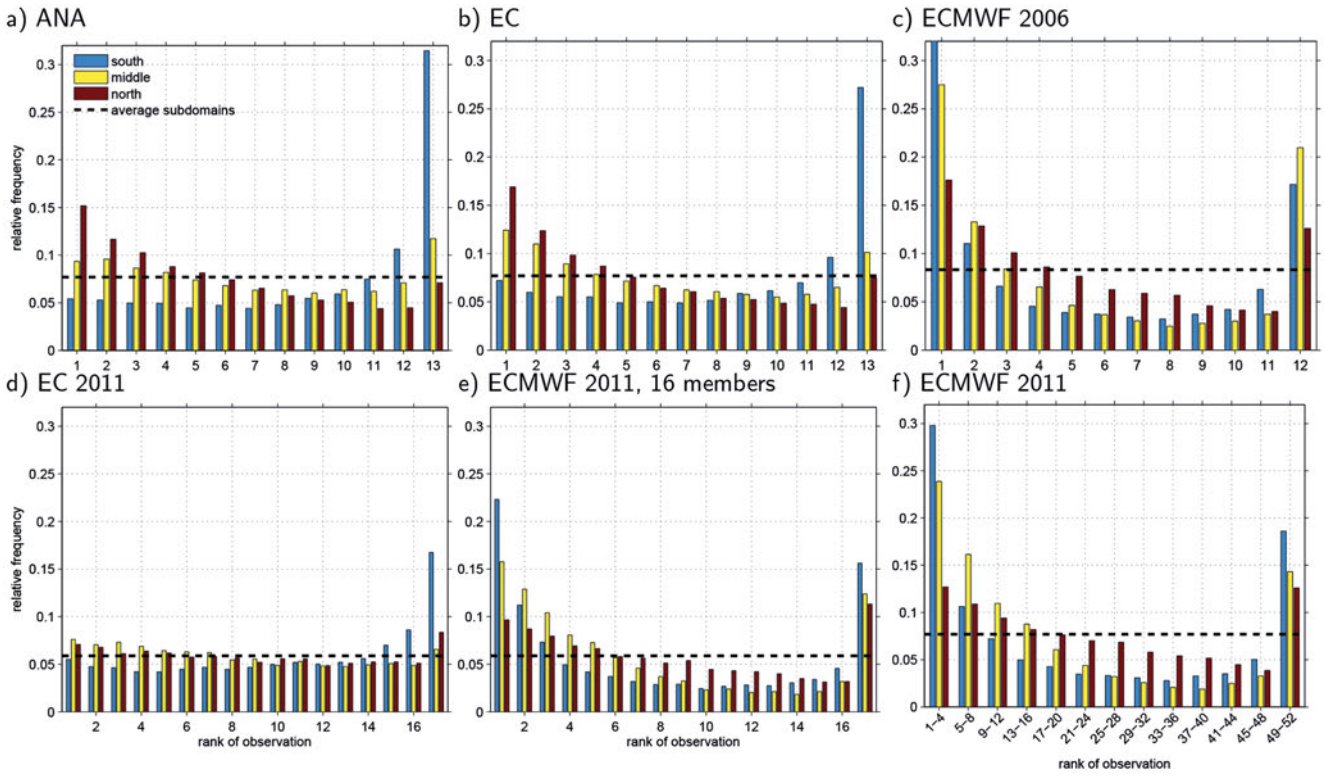

Figure 5: Rank histograms for all ensemble versions for 24-h precipitation sums, for the three subdomains for each ensemble version; in (e), the ranks are shown for the 16 ensemble members of ECMWF 2011 used as IBCs for EC 2011 and in (f), for the full ECMWF EPS.

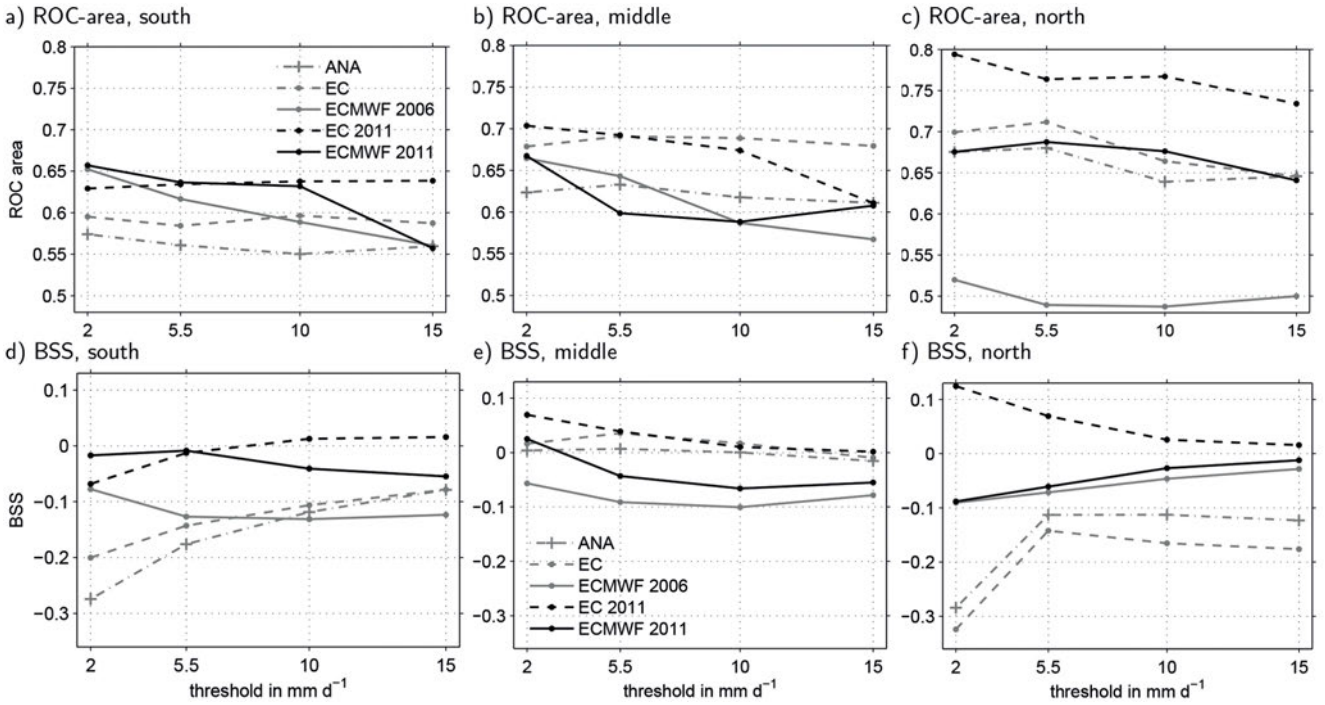

Figure 6: ROC area and Brier skill score (BSS) as a function of precipitation threshold for all ensemble versions in the three subdomains.

hood of about ±125 km (i.e., more than a tenth of the zonal extent) in both *x*- and *y*-direction. This also means that the areas of the three subdomains overlap when using the neighborhood approach.

Finally, reliability diagrams (Wilks, 2011) were computed using the same precipitation thresholds as for the ROC diagrams as well as the neighborhood approach. The Brier skill scores (BSS, Figure 6d–f)

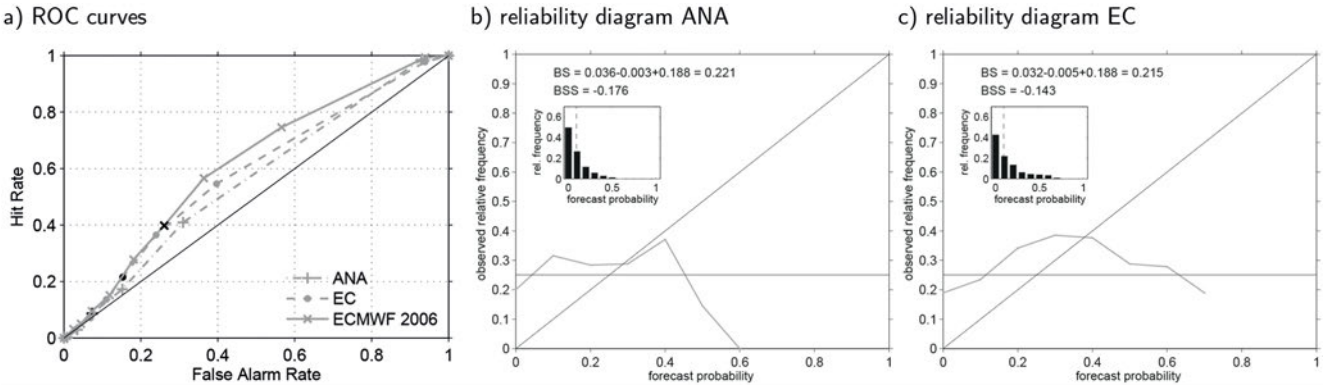

Figure 7: ROC curves for the three ensemble versions for 2006 (a) and reliability diagrams for ANA as well as EC (b and c), all for a precipitation threshold of 5.5 mm d^{-1} and the southern subdomain; in (a), the points corresponding to $P_i > 0.3$ are denoted by black markers; values of Brier score (BS) calculated as described by Murphy (1973) and of BSS as well as frequency distributions of forecast probabilities (small histograms) are given in (b) and (c).

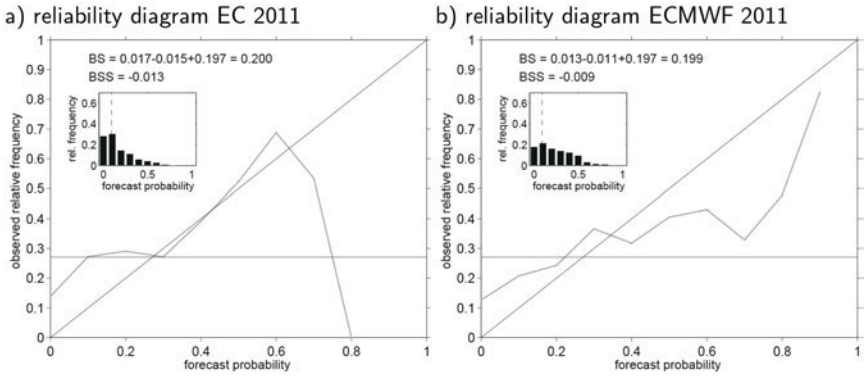

Figure 8: Reliability diagrams for the two ensemble versions for 2011, for a precipitation threshold of 5.5 mm d^{-1} and the southern subdomain.

were calculated using "climatology" as a reference, i.e. the reference Brier score is defined as the uncertainty $\overline{o}(1-\overline{o})$, with $\overline{o}$ being the relative observed frequency of the considered event. As for the ROC diagrams, mainly values of BSS are given without showing all reliability diagrams for all ensemble versions, subdomains and precipitation thresholds (examples of reliability diagrams are given in Figs. 7 and 8).

Apart from the probabilistic methods, the Structure, Amplitude and Location (SAL) index (Wernli et al., 2008) was applied. The SAL index was chosen from different spatial verification methods (Gilleland et al., 2009), because it is an object-based method that is well applicable to large precipitation objects generated by MCSs. The *S* component gives information about the mean intensity of the simulated precipitation objects compared to the observed ones. Originally, *S* was designed to identify forecast errors of simulations reproducing stratiform precipitation in convective cases (too large and flat objects; positive *S*) or vice versa (negative *S*). Here, we mainly deal with convective events in both forecasts and observations. Still, *S* helps to determine whether the simulated convective systems were appropriate in size and intensity. The amplitude *A* compares the absolute values of simulated sums to the observed ones. The *L* component describes how well the location of the simulated precipitation patterns agrees with the observed ones. For the verification of the ensemble versions of this investigation, mainly the *S* and *L* components were used as the *A* component corresponds to the mean bias already discussed in the analysis of ensemble mean, spread and error.

To compute *S* and *L* for the daily precipitation sums, precipitation objects were identified first. For this, a threshold value of 1/15 of the 95$^{\text{th}}$ percentile of the precipitation field was chosen, as proposed by Wernli et al. (2009). In contrast to the probabilistic scores, the components were calculated for the whole evaluation domain, as the subdomains would have been too small. *S* is based on the "scaled volume", which is defined as the precipitation sum of an object scaled by its maximum. It is normalized so that it ranges from −2 to +2. *L* is equal to the normalized difference between the centers of mass of the observed and simulated spatial precipitation distribution plus the normalized difference of the mean distance of all precipitation objects from the center of mass of the respective field. For this reason, the mean location of the simulated precipitation objects agrees better with the observed one for smaller values of *L*.

4 Verification results

4.1 Single-model vs. multi-analysis ensemble version

The single-model version (EC) will now be compared with the multi-analysis version (ANA) for the three subdomains.

In the southern subdomain, both ensemble versions display a negative precipitation bias, which is smaller for EC than for ANA (-2.1 and $-3.3\,\mathrm{mm\,d^{-1}}$, respectively, Figure 4a). This is also reflected by the frequency distribution of each ensemble version (Figure 3a): They indicate lower relative frequencies of precipitation than observed throughout the whole range of values larger than zero, more markedly for ANA than for EC ($KL = 0.30$ compared to 0.16). In EC, the spread is also higher and the 95$^{\mathrm{th}}$ percentile, i.e. the extreme values, are larger than in ANA. For both, the spread is clearly lower than the error. As the mean values, the rank histograms (Figure 5a and b, blue bars) also indicate a smaller bias for EC (lower frequency of the highest rank). The ROC area is higher for EC (Figure 6a), as the hit rate increases faster with decreasing probability thresholds P_i; for example for $P_i > 0.3$ and the precipitation threshold of $5.5\,\mathrm{mm\,d^{-1}}$, the hit rate is only 0.09 for ANA, but 0.22 for EC (Figure 7a). For both, the corresponding false alarm rate is < 0.2. The BSS is negative and slightly higher for EC for all but the highest precipitation threshold (Figure 6d). The sharpness is better in EC, i.e. high P_i values occur more often than in ANA (Figure 7b and c).Yet, both ensemble versions are highly overconfident: The observed frequency in cases with $P_i = 0$ is about 0.2, for example, and lower than 0.3 for $P_i = 0.5$ (Figure 7b and c); i.e., precipitation is observed in 20 % of all cases when precipitation is forecast by none of the members, and in less than 30 % when it is forecast by half of the members. This is attributable to the negative bias.

In contrast to the southern subdomain, the bias for both ANA and EC **in the middle subdomain** is low (-0.5 and $+0.7\,\mathrm{mm\,d^{-1}}$, respectively, Figure 4b). This is also reflected by a much smaller difference of the frequency distributions to the observed one (Figure 3b). Furthermore, the spread is only slightly smaller than the error. The spread and the extreme values again are higher in EC than in ANA. Yet, the rank histograms indicate a higher underdispersion of EC than of ANA (Figure 5a and b, respectively): The mean relative deviation of all ranks from their mean is 24 % for EC and 18 % for ANA. At the same time, the highest rank is less frequent in EC, indicating that precipitation is less often underestimated by all ensemble members. This is consistent with a slightly higher precipitation sum than in the observations in EC and a slightly smaller sum in ANA. On the other hand, the lowest ranks are more frequent in EC, with a mean observed precipitation of zero for the lower ranks. These two facts together indicate that more often than in ANA, a high number of members of EC forecasts precipitation where none was observed. With this, it can be assumed that the false alarm rate is higher in EC than in ANA. However, the ROC area is consistently higher in EC than in ANA (Figure 6b), which is again due to a faster increasing hit rate. Smaller false alarm rates in ANA do not influence the ROC area, as they occur in combination with a very low hit rate. For $P_i > 0.8$ for example, both rates are still zero for ANA, while the hit rate is 0.034 and the false alarm rate is 0.01 for EC. The BSS is also higher for EC than for ANA (Figure 6e).

To sum up, more members of EC than of ANA reproduce precipitation in cases of no precipitation being observed in the middle subdomain, which results in higher false alarm rates in EC. As the hit rates are higher in EC, the ROC area nevertheless is higher. As the majority of ranks (about 70 %) corresponds to observed precipitation sums of zero, however, the false alarms have a high weight in the rank histogram and it deviates more strongly from a flat frequency distribution for EC than for ANA. The underdispersion in EC is, thus, mainly caused by cases of a large number of members reproducing precipitation at the wrong location.

In turn, there is a positive precipitation bias **in the northern subdomain** for both EC and ANA ($+1.0\,\mathrm{mm\,d^{-1}}$ for both, Figure 4c). The relative frequency of both lower ($< 5\,\mathrm{mm\,d^{-1}}$) and higher precipitation sums ($> 10\,\mathrm{mm\,d^{-1}}$) is clearly higher than observed for both ensemble versions (Figure 3c). Still, the sums are lower than in the middle subdomain in both observations and simulations and, hence, the error is smaller. At the same time, the spread is large, so that spread and error are almost equal. The spread is slightly smaller in EC, which agrees with the distribution of ranks (Figure 5a and b): The lowest and highest ranks for EC are more frequent than for ANA, indicating stronger underdispersion of EC. Even though a higher frequency of low ranks in EC than in ANA might also suggest a larger positive bias, this is not confirmed by the mean sums (Figure 4c). As in the southern and middle subdomains, the ROC area is larger for EC (Figure 6c). For both versions, it is larger than in the other subdomains for the low thresholds at least. The higher ROC in the northern subdomain is due to a lower false alarm rate there (not shown). In spite of relatively high ROC areas, the BSS (Figure 6f) is extremely low for both versions due to high overconfidence: In cases of high P_i, the observed frequency is much lower (e.g. about 0.2 for $P_i = 0.6$ for the precipitation threshold of $2\,\mathrm{mm\,d^{-1}}$). Moreover, it is higher for ANA than for EC. The reason of this opposite behavior of ROC and BSS is that the false alarm rate gives the ratio of incorrect forecasts for all cases that were not observed, while the reliability is based on the rate of observed events for all cases that were forecast with a particular P_i. In the northern subdomain, the ratio of cases not observed is 0.9 for a precipitation threshold of $2\,\mathrm{mm\,d^{-1}}$ (in 0.9 of all observed cases, precipitation was lower than $2\,\mathrm{mm\,d^{-1}}$). In contrast, the ratio of $P_i > 0.5$, which contributes to about 80 % of the reliabil-

ity in this case, is 0.11 for EC and 0.10 for ANA. Hence, while a large part of the data samples contributes to larger ROC areas, a smaller part contributes to the high Brier score and, consequently, lower BSS. This means that in this case, not too much importance should be attached to the low BSS.

The rank histograms for T_B (not shown) mainly confirm the smaller negative bias (too few high clouds) for EC in the south, higher underdispersion in the middle, and a larger mean deviation from the mean in the north. However, they also indicate a general overestimation of high clouds, which is a known bias of the COSMO model (Böhme et al., 2011; Eikenberg et al., 2015).

The comparison of ANA and EC clearly shows that both ensemble versions display similar biases: A negative precipitation bias in the southern subdomain and a positive bias in the northern one. The spread is, thus, smaller than the error, but much less in the middle subdomain, where the bias is much smaller. The hit rates are slightly higher for EC. High P_i are more frequent, indicating better sharpness, as $\overline{o}$ is low. On the other hand, underdispersion is stronger for EC than for ANA in the middle; according to a low $\overline{o}$, the rank histograms generally are strongly influenced by large areas where no precipitation was observed. As less members agree with the observation in these cases in EC, the lowest ranks are more frequent than in ANA. Overall, the EC ensemble displays higher scores than ANA and is more skillful, with slightly higher spread due to more extreme values in two of the three subdomains, but also higher underdispersion due to more false alarms at low precipitation thresholds. The findings indicate that the higher number of IBCs used for EC does not definitely generate a larger spread than in ANA. The reason may be that the global ensemble (ECMWF 2006) itself, which provides the IBCs for EC, displays a small spread only. Therefore, we conclude that the analyses used for ANA differ more from each other. As the rank histograms for ANA do not indicate that the ensemble members are grouped by their IBCs as shown for a comparable ensemble setup (COSMO-DE-EPS; Theis and Gebhardt, 2011), we may additionally conclude that the applied land-surface perturbations generate sufficient complementary spread within the ANA ensemble.

4.2 Convection-permitting vs. global ensemble version for 2006

The comparison of the two different setups of CP ensemble versions suggested that the global ensemble ECMWF 2006, which provided initial and boundary conditions for one of them, exhibits low spread. Thus, we will now analyze the global ensemble in comparison to the CP ensemble versions.

In the southern subdomain, the global ensemble also displays a negative precipitation bias, but it is smaller than for the CP versions (Figure 4a). The frequency distribution reveals that mainly the higher sums ($> 25\,\mathrm{mm\,d^{-1}}$) have a lower frequency than observed, while the lower sums have a higher frequency (Figure 3a). This distribution implies too broad precipitation patterns with too weak maxima, which are typical for models with coarse horizontal resolution. It results in a much larger value of KL than for the CP version (0.76 compared to 0.16), even if the latter has a strong bias. The error of the CP versions is, as the negative bias, slightly higher than for the global ensemble, and the spread is much higher (Figure 4a). The higher spread of the CP versions can be attributed partly to higher and less frequent extreme values. Still, it can be assumed that the spread is indeed particularly low for the global ensemble, because there also is a strong underdispersion, as indicated by the increased frequencies of the outer ranks (relative frequencies of 0.32 and 0.17 compared to the average of 0.08, Figure 5c). The mean relative deviation from the mean is 70 %, while it is 53 % and 43 % for the CP versions. Except for the highest precipitation threshold, the ROC area is smaller for the CP versions (Figure 6a), because the hit rates for the medium-high and low P_i are lower: e.g. at $P_i > 0.3$ and the precipitation threshold of $5.5\,\mathrm{mm\,d^{-1}}$, the hit rate is smaller than 0.3 for both CP versions, while it is > 0.4 for the global version (Figure 7a). With increasing precipitation thresholds, the difference between the ROC areas obtained with the CP and global ensemble versions becomes smaller. Together with comparably high false alarm rates in the global ensemble, this may be caused by the too broad and flat precipitation patterns. Likewise, the BSS is lower for the CP versions for the smallest threshold of $2\,\mathrm{mm\,d^{-1}}$ (Figure 6d). For a threshold of $10\,\mathrm{mm\,d^{-1}}$, it is similar for all versions.

In contrast to the CP versions, the global ensemble also displays a negative precipitation bias **in the middle subdomain** (Figure 4b). It is even higher than in the southern subdomain, which is also reflected by the higher KL (Figure 3b). The error again is slightly higher for the CP versions than for the global ensemble, while the spread as well as the extreme values are much higher. Low extreme values in the global ensemble are in agreement with the negative bias. In the rank histogram, though, frequencies of both lowest and highest ranks are very high (Figure 5c), which generally means a strong underdispersion. Due to the mean negative bias, it must be assumed that the overestimated precipitation sums, which generate the high frequency of the lowest ranks, are low. If the simulated sums for these cases were as high as the thresholds considered for calculating skill scores (Figure 6), this would be reflected by high false alarm rates and, thus, small ROC areas for the global ensemble. However, the ROC area is as large as for the EC ensemble for the lowest precipitation threshold of $2\,\mathrm{mm\,d^{-1}}$ and the false alarm rate is not very high. Thus, the majority of false alarms causing the high frequency of the lowest ranks is due to sums lower than $2\,\mathrm{mm\,d^{-1}}$ in the global ensemble. Still, the ROC area decreases strongly with increasing precipitation thresholds. For the threshold of $10\,\mathrm{mm\,d^{-1}}$ for example, the hit rate for $P_i > 0$ is only 0.2, which is equivalent to

a relative frequency of missed hits of 0.8. This means that in a large number of cases, none of the ensemble members of the global ensemble reproduces the larger MCSs observed in the middle subdomain, which again is in agreement with the negative precipitation bias. Accordingly, the BSS is clearly higher for the CP versions (Figure 6e).

In the northern subdomain, the mean precipitation of the global ensemble is almost zero with the 95th percentile of 0.63 mm d^{-1}, while the mean observed precipitation is 1.2 mm d^{-1} (Figure 4c). This negative bias of the global ensemble is also indicated by the frequency distribution displaying a very strong underestimation of relative frequencies of all precipitation sums > 2.5 mm d^{-1}. The ranks (Figure 5c) indicate that small sums are still simulated in regions where no precipitation was observed, as was the case in the middle subdomain (u-shaped histogram instead of an overestimation of the lowest ranks only). The ROC area is dominated by the very large number of missed hits and it is much larger for the CP versions (Figure 6c). Nevertheless, the BSS is lower for the CP versions. The reason is that almost only $P_i = 0$ exists for the global versions. For the CP versions, the BSS is very low, because a small number of cases with higher P_i contributes strongly, as was discussed for the northern subdomain in Section 4.1.

In conclusion, the results of the comparison of the CP and the global ensemble versions for 2006 are strongly influenced by the fact that the precipitation patterns are smoother in the global simulations. Smooth precipitation patterns partly cause the strong underdispersion indicated by the rank histograms for the global ensemble: Ensemble members reproduce light precipitation (0.1–1 mm d^{-1}) in the larger vicinity of MCSs, increasing the frequency of the lower ranks. Simultaneously, peaks are too flat, which contributes to the high frequency of the higher ranks. Furthermore, skill scores decrease for increasing precipitation thresholds when the maxima are too weak. On the other hand, an underdispersion is also found for the global ensemble in the southern subdomain where precipitation sums are high. A strong underforecast of MCSs is found in the middle subdomain. In comparison, the CP ensemble versions have a strong negative bias in the south so that only for the higher precipitation thresholds, scores are better than for the global ensemble. Overall, it can be confirmed that the spread is low for the global ensemble. For the CP versions, it is clearly higher, underdispersion is lower and the scores are mostly higher.

4.3 Convection-permitting vs. global ensemble version for 2011

As for 2006, the CP and the global ensemble versions for 2011 (EC and ECMWF 2011) are compared for the three subdomains.

Again, there is a negative precipitation bias **in the southern subdomain** for both the CP and the global ensemble versions, which is higher for the CP ensemble (-3.1 compared to -2.5 mm d^{-1}, Figure 4a). As for 2006, the frequency distribution of the CP ensemble displays overall lower frequencies than observed, but with a much smaller value of KL (Figure 3d). The global ensemble again overestimates the frequencies of the lower sums and underestimates them for the higher sums. The error is similar for both versions, but both extreme values and spread are higher for the CP than for the global ensemble. As in 2006, the high frequency of the largest ranks confirms the negative bias for the CP ensemble (Figure 5d). Concerning the global ensemble, the rank histograms for both the subset of 16 members as well as the full ECMWF EPS 2011 reveal strong underdispersion (Figure 5e and f). The ROC area is smaller for the CP than for the global ensemble at the lowest precipitation threshold of 2 mm d^{-1} (Figure 6a). At this threshold, the global ensemble has especially high hit rates for low P_i, meaning that for all cases when precipitation was observed, a small number of members mostly reproduces at least this sum of precipitation. As for 2006, the ROC area for the global ensemble decreases with the threshold. For the higher thresholds, it is smaller than for the CP ensemble. For the CP ensemble, in comparison, the largest contributions to the ROC area are found for high P_i, with false alarm rates increasing more slowly for higher precipitation thresholds. Smoother patterns for the global simulations and decreasing false alarm rates with increasing precipitation thresholds for the CP ensemble explain why the scores are better for the CP ensemble at higher precipitation thresholds despite the strong negative bias. The fact that lower P_i ($P_i < 0.3$) are more reliable for the global ensemble, while higher P_i ($0.3 < P_i < 0.7$) are more reliable for the CP ensemble, is confirmed by the reliability diagrams (Figure 8). The BSS for both versions behaves similarly to the ROC area regarding different precipitation thresholds (Figure 6d).

In the middle subdomain, a small negative precipitation bias for the global ensemble is still present, but no bias for the CP ensemble (Figure 4b). Values of KL confirm that also the frequency distribution of precipitation matches better the observed one for the CP than for the global ensemble (0.09 compared to 0.92; Figure 3e). The error is similar for both versions, and again, spread and extreme values are much higher for the CP ensemble. The rank histogram shows underdispersion of the global ensemble (Figure 5e and f). The CP ensemble, though, displays an almost flat rank histogram (Figure 5d), indicating well-distributed probability forecasts. The remaining scores (Figure 6b and e) show consistently higher values for the CP than for the global ensemble.

Finally, the bias for the CP ensemble is small **in the northern subdomain** (Figure 4c). For the global ensemble, it is larger (-78 % compared to the observed sum) and it is evident from the frequency distribution that the global ensemble underestimates all frequencies of precipitation sums of more than 3 mm d^{-1} (Figure 3f). The rank histograms (Figure 5d–f) indicate a minor negative

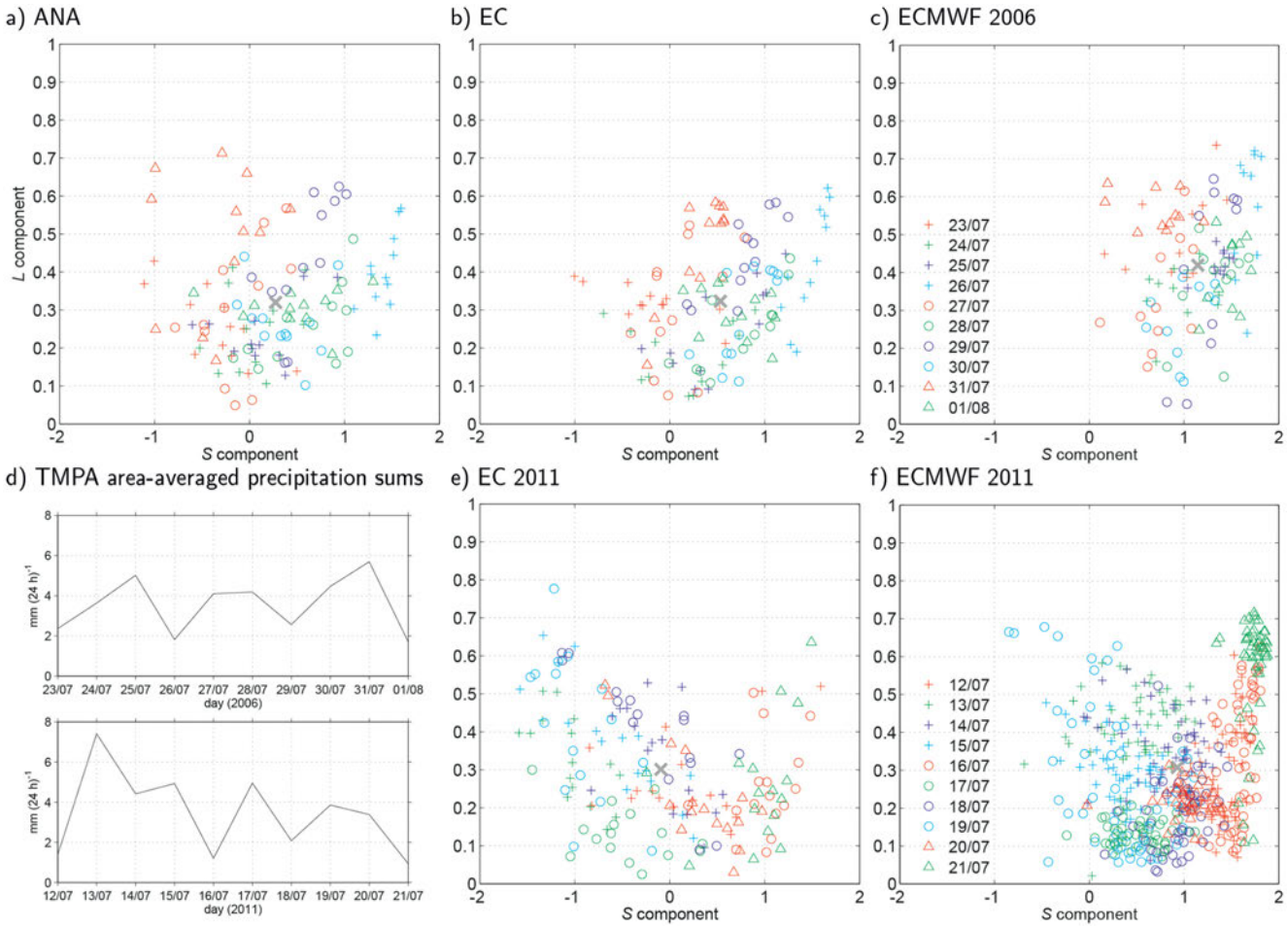

Figure 9: *L* and *S* components for all members and all days of each ensemble version (a–c, e, and f), with gray crosses denoting the mean values; area-averaged 24-h precipitation sums of TMPA (d) for both 10-day periods.

bias for the CP ensemble and underdispersion for the global ensemble. The ROC area is large for the CP ensemble and much smaller for the global ensemble, which is due to a large amount of missed hits (almost 90 % for a threshold of 5.5 mm d^{-1}). The BSS is much higher for the CP than for the global ensemble for the lower precipitation thresholds, with a decreasing difference towards the higher thresholds (Figure 6f). The increasing BSS with the precipitation threshold for the global ensemble is caused by the decreasing frequency of cases with $P_i > 0$. For a threshold of 10 mm d^{-1} for example, $P_i = 0$ in 99.9 % of all cases. For the CP ensemble, the decrease of frequency of higher P_i is less distinct, but higher P_i become less reliable, so that the BSS decreases with the threshold. On the other hand, the observation has a frequency of less than 3 % for the threshold of 10 mm d^{-1}, and the statistics are possibly affected here because of the relatively small sample size.

Altogether, the scores for the global ensemble are influenced by the fact that the precipitation patterns are too smooth, as was the case in 2006; at the same time, the underdispersion in the south is so strong that it cannot be caused entirely by this effect. In the south, the CP ensemble has a larger negative precipitation bias than the global ensemble, but the CP ensemble has not clearly lower scores there. In the middle subdomain, the CP ensemble is consistently more skillful than the global ensemble and it also shows skillful forecasts in the north. The gain of the CP version is thus evident for 2011. All of the scores indicate a better skill for ECMWF 2011 than for ECMWF 2006, which may be due to the fact that ECMWF 2006, as an experimental ensemble, had no optimal setup.

4.4 Comparison of ensemble versions by spatial forecast verification

It was shown that the differences of the probabilistic methods used here partly reflect the different structures of the precipitation patterns in the global and CP forecasts. Additionally, the *S* and *L* components of the SAL index were calculated for each ensemble member and each daily precipitation sum.

The *S* component confirms the assumptions concerning the structure of the precipitation patterns (Figure 9): While values of *S* are closer to 0 for the CP versions (Figure 9a, b, and e), they are mostly positive for the global ones (Figure 9c and f), meaning that the objects

there are indeed too flat and/or large on the average. This implies that the negative precipitation bias found for the global ensemble versions in all three subdomains can be attributed partly to too small precipitation maxima and must not necessarily be caused by a large number of missed MCSs. On the other hand, more negative values of S exist in ANA than in EC (Figure 9a and b). The negative values correspond mainly to 31 July (red triangles). Examination of the individual precipitation fields reveals that too small and peaked precipitation objects are simulated by members of the ANA ensemble (not shown). Members of ANA with lowest values are those with IBCs from GME, i.e. members that can provide forecasts that are very different from EC forecasts. Apart from the S values for this day, they are mostly positive for ANA and EC – even if closer to 0 than for ECMWF 2006 and 2011 – indicating too flat precipitation objects, which agrees with the strong negative bias in the southern subdomain, where most precipitation is observed. For 2011 (Figure 9e and f), values of S are lower for both the CP and the global ensemble than for 2006.

The L components are below 0.7 in most cases (Figure 9). The values are generally low, because the precipitation objects are large compared to the size of the evaluation domain, which causes the centers of mass to be mainly located in the middle of the domain. For the EC version, L is even lower than 0.6 in most cases, but the mean value of L, $\overline{L}$, is 0.32 for both ANA and EC (Figure 9a and b). The reason is that more maximum values of L occur for ANA than for EC, indicating that also better forecasts than in EC must exist to result in the same mean value. Consequently, even if EC has higher scores, ANA seems to exhibit a larger spread concerning the location of convective systems, at the cost of containing ensemble members that reproduce convection at a wrong location. This is in agreement with very small hit rates in ANA for the higher P_i as well as a lack of high P_i and is attributed to the different IBCs of ANA that can yield larger differences than those of EC, as already found in Section 4.1.

For the ECMWF 2006 ensemble version (Figure 9c), $\overline{L}$ is larger than for the CP ensemble versions (0.42 compared to 0.32), indicating that the location of the precipitation patterns is better reproduced by the CP ensemble on the average. This could not be inferred from the probabilistic scores.

For EC 2011 and ECMWF 2011 (Figure 9e and f), $\overline{L}$ is almost similar (0.30 for EC 2011 and 0.31 for ECMWF 2011), whereas probabilistic scores were mostly better for the CP ensemble in the middle and northern subdomain. One reason is that these scores were calculated for the three subdomains separately, while the whole evaluation domain was used for the SAL index. The highest precipitation sums were observed in the southern subdomain and the computation of L is based on the center of mass. Consequently, the southern subdomain is given the highest weight. As not all scores were better for the CP ensemble in the southern subdomain, the weighting may be the reason why $\overline{L}$ is similar for the whole domain in both ensemble versions. Apart from that, the mean location of ECMWF 2011 is clearly better than that of the ECMWF 2006.

Comparing S and L for the different members within the individual ensemble versions, it can be said that they are roughly grouped by day. This is partly attributable to the properties of S and L: Very high values of L for 19 July 2011 (light blue circles in Figure 9e and f), for example, can be explained by the structure of the precipitation fields. The main precipitation lies in the southeastern corner of the evaluation domain. L can easily become higher when the center of mass of the observation field is moved towards the border of the evaluation domain. For the 2006 ensemble versions, for example, highest values of S and high values of L are obtained for 26 July (light blue crosses in Figure 9a–c). On this day, no large convective system was observed and the precipitation sum was low (Figure 9d), but the simulations reproduce some precipitation in the northern subdomain (not shown), which causes the poor values of S and L. The values for A are close to 0 for this day (not shown), meaning that the absolute sum is adequate.

This means that the day-to-day variability of S and L also depends on the observed precipitation pattern and that the values of the scores are not absolutely comparable for all patterns. This helps understand why the variability is similar for the CP and the global ensemble versions.

5 Summary and conclusions

Different ensemble forecasts of convective precipitation for West Africa were assessed and compared: (1) For a period at the end of July 2006, two versions of convection-permitting (CP) ensemble simulations were performed using different combinations of multiple IBCs and specially generated land-surface perturbations. It was investigated whether the single-model ensemble generated larger spread than the multi-analysis one, as a larger number of IBCs was used. (2) These two versions were then compared to the global forecasts providing IBCs for the single-model version. Additionally, a second ensemble forecast with the single-model setup was performed for a period of July 2011, when the ensemble of data assimilations was already operational at ECMWF. Sixteen members of the ECMWF EPS were selected by clustering and used as IBCs for the CP ensemble. This CP ensemble forecast was compared to the ECMWF EPS.

The result of (1) is that the single-model setup has higher scores than the multi-analysis ensemble. Both display a negative precipitation bias in the southern subdomain, which is slightly higher for the multi-analysis version. In the middle, the multi-analysis version has a higher dispersion. On the other hand, the single-model version is more reliable and has better sharpness; high

forecast probabilities occur more frequently than in the multi-analysis version. Moreover, the single-model version has a lower percentage of missed hits. The location component L of the SAL index shows that members of the multi-analysis ensemble yield the best forecasts of the precipitation patterns, while members with a low skill exist, too. This must be attributed to the fact that the different model analyses can differ strongly from each other and agrees with the finding that high forecast probabilities of the investigated precipitation events are very rare. In conclusion, the comparison of the single-model and the multi-analysis CP ensemble version reveals that a larger number of IBCs (as done in the single-model ensemble) cannot clearly increase the ensemble spread, even if this was achieved for a regional ensemble for Europe (Marsigli et al., 2014). The reason may be that the model analyses comprise larger uncertainties for West Africa than for Europe. Lower skill of the multi-analysis ensemble implies at the same time that the global model analyses occasionally provide less realistic synoptic fields than the ECMWF forecasts used as IBCs for the single-model version. On the other hand, the rank histograms do not imply that ensemble members of the multi-analysis ensemble are grouped by the different IBCs. We can thus conclude that the applied land-surface perturbations generate enough spread to be useful in such an ensemble setup, which is consistent with the finding of Maurer et al. (2015) that land-surface perturbations in the multi-analysis ensemble generate as much spread as the IBCs from the three chosen model analyses. With respect to (2), the global ensemble versions have a smaller negative bias than the CP versions in the southern subdomain, but distinct negative biases in the middle and northern subdomains. The spread is much smaller in the global than in the CP ensemble versions, which is not only attributable to lower extreme values. The rank histograms show a strong underdispersion for the global ensemble versions, too. The underdispersion is partly caused by the smoothed structure of patterns in simulations of coarser horizontal resolution. The large difference of the patterns between the CP and the global ensemble versions is confirmed by the structure component of the SAL index. It is, thus, notable that the rank histograms for precipitation have to be interpreted carefully as they are sensitive to different frequency distributions. In the southern subdomain, the considered skill scores are not consistently higher for global or CP versions, as they also reflect the different structure of the precipitation patterns and behave differently for changing precipitation thresholds. In the middle and in the north, the CP versions show better scores in almost all cases and it could be explained by the issue of sample size for the cases when this was not true.

The gain of the CP ensemble compared to the global ensemble is impaired by the strong negative precipitation bias in the southern subdomain. On the other hand, the CP simulations compensate the negative bias of the ECMWF simulations in the middle subdomain (Sahel) as well as in the northern one (southern Sahara). Apart from the removed influence of smoothed structures that were encountered in the coarser-resolution forecasts, the CP ensemble versions also display larger spread.

For the evaluated period of the full ECMWF EPS (referred to as ECMWF 2011 in this investigation), better scores were found than for the ECMWF version for 2006. The scores for the CP ensemble for 2011 (EC 2011) were also improved compared to the CP ensemble versions for 2006, which means that also the skill of the IBCs had an influence. The overall skill of EC 2011 is satisfying, except for the negative precipitation bias in the southern subdomain. From all five considered ensemble versions, it is the only one that gives objectively a real improvement over "climatology" (as indicated by the positive Brier skill score), reflecting that there is still a huge potential to advance short-range quantitative precipitation forecasts for West Africa. To conclude, drawing IBCs from the ECMWF EPS and adding additional perturbations on the land surface is an appropriate method to generate CP ensemble simulations for West Africa during the peak of the monsoon season in the Sahel. Furthermore, it was shown that it is necessary to perform CP simulations to achieve reliable probability forecasts of precipitation generated by MCSs in the Sahel region.

Acknowledgments

The authors would like to thank Andrea Montani and the COSMO-LEPS group (ARPA-SIMC, Bologna) for their kind assistance and especially for making available the clustering tools. Moreover, we are grateful to Folke Olesen (IMK-ASF) for providing the MSG IR satellite data. We also want to thank Simon Lang at ECMWF for conducting the experimental IFS 10-member ensemble simulations for our forecast period for 2006 by applying a climatological EDA. We also acknowledge the use of ECMWF's computing and archive facilities through the special project 'Convection-permitting ensemble simulations for West Africa based on different soil moisture fields'. The TMPA data used in this effort were acquired as part of the activities of NASA's Science Mission Directorate, and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.

References

Agustí-Panareda, A., A. Beljaars, M. Ahlgrimm, G. Balsamo, O. Bock, R. Forbes, A. Ghelli, F. Guichard, M. Köhler, R. Meynadier, J.-J. Morcrette, 2010: The ECMWF re-analysis for the AMMA observational campaign. – Quart. J. Roy. Meteor. Soc. **136**, 1457–1472, DOI: 10.1002/qj.662.

Bechtold, P., J.-P. Chaboureau, A. Beljaars, A.K. Betts, M. Köhler, M. Miller, J.-L. Redelsperger, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. – Quart. J. Roy. Meteor. Soc. **130**, 3119–3137, DOI: 10.1256/qj.03.103.

Beucher, F., J.-P. Lafore, F. Karbou, R. Roca, 2014: High-resolution prediction of a major convective period over West Africa. – Quart. J. Roy. Meteor. Soc. **140**, 1409–1425, DOI: 10.1002/qj.2225.

Birch, C.E., D.J. Parker, J.H. Marsham, D. Copsey, L. Garcia-Carreras, 2014: A seamless assessment of the role of convection in the water cycle of the West African Monsoon. – J. Geophys. Res. Atmos. **119**, 2890–2912, DOI: 10.1002/2013JD020887.

Bowler, N.E., A. Arribas, K.R. Mylne, K.B. Robertson, S.E. Beare, 2008: The MOGREPS short-range ensemble prediction system. – Quart. J. Roy. Meteor. Soc. **134**, 703–722, DOI: 10.1002/qj.234.

Böhme, T., S. Stapelberg, T. Akkermans, S. Crewell, J. Fischer, T. Reinhardt, A. Seifert, C. Selbach, N. Van Lipzig, 2011: Long-term evaluation of COSMO forecasting using combined observational data of the GOP period. – Meteorol. Z. **20**, 119–132, DOI: 10.1127/0941-2948/2011/0225.

Casati, B., G. Ross, D.B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. – Meteorol. Apps. **11**, 141–154, DOI: 10.1017/S1350482704001239.

Eikenberg, S., C. Köhler, A. Seifert, S. Crewell, 2015: How microphysical choices affect simulated infrared brightness temperatures. – Atmos. Res. **156**, 67–79, DOI: 10.1016/j.atmosres.2014.12.010.

Fink, A.H., 2006: The West African monsoon system. – promet **32**, 114–122.

Frappart, F., P. Hiernaux, F. Guichard, E. Mougin, L. Kergoat, M. Arjounin, F. Lavenu, M. Koité, J.-E. Paturel, T. Lebel, 2009: Rainfall regime across the Sahel band in the Gourma region, Mali. – J. Hydrol. **375**, 128–142, DOI: 10.1016/j.jhydrol.2009.03.007.

Gantner, L., N. Kalthoff, 2010: Sensitivity of a modelled life cycle of a mesoscale convective system to soil conditions over West Africa. – Quart. J. Roy. Meteor. Soc. **136**(S1), 471–482, DOI: 10.1002/qj.425.

Gebhardt, C., S. Theis, M. Paulat, Z.B. Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. – Atmos. Res. **100**, 168–177, DOI: 10.1016/j.atmosres.2010.12.008.

Gilleland, E., D. Ahijevych, B.G. Brown, B. Casati, E.E. Ebert, 2009: Intercomparison of spatial forecast verification methods. – Wea. Forecast. **24**, 1416–1430, DOI: 10.1175/2009WAF2222269.1.

Guichard, F., N. Asencio, J.-L. Redelsperger, J.-P. Lafore, M. Nuret, A. Boone, B. Decharme, C. Peugeot, O. Bock, X. Cui, A. Morse, M. Garvert, M.A. Gaertner, B. Lamptey, E. Orlandi, J. Sander, S.C. Jones, F. Fierli, G. Balsamo, P. de Rosnay, P.P. Harris, J.-C. Bergès, 2010: An Intercomparison of Simulated Rainfall and Evapotranspiration Associated with a Mesoscale Convective System over West Africa. – Wea. Forecast. **25**, 37–60, DOI: 10.1175/2009WAF2222250.1.

Hamill, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. – Mon. Wea. Rev. **129**, 550–560, DOI: 10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hamill, T.M., S.L. Mullen, C. Snyder, D.P. Baumhefner, Z. Toth, 2000: Ensemble forecasting in the short to medium range: Report from a workshop. – Bull. Amer. Meteor. Soc. **81**, 2653–2664, DOI: 10.1175/1520-0477(2000)081<2653:EFITST>2.3.CO;2.

Houze, R.A., 1981: Structures of atmospheric precipitation systems: A global survey. – Radio Sci. **16**, 671–689, DOI: 10.1029/RS016i005p00671.

Huffman, G., R. Adler, D. Bolvin, G. Gu, E. Nelkin, K. Bowman, Y. Hong, E. Stocker, D. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. – J. Hydrometeorol. **8**, 38–55, DOI: 10.1175/JHM560.1.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M.Leutbecher, L. Raynaud, 2010: Ensemble of Data Assimilations at ECMWF. – Technical Memorandum **636**.

Iversen, T., 2011: Tellus A Special Issue on probabilistic short-range weather forecasting. – Tellus A **63**, 371–372, DOI: 10.1111/j.1600-0870.2011.00520.x.

Janicot, S., C.D. Thorncroft, A. Ali, N. Asencio, G. Berry, O. Bock, B. Bourles, G. Caniaux, F. Chauvin, A. Deme, L. Kergoat, J.-P. Lafore, C. Lavaysse, T. Lebel, B. Marticorena, F. Mounier, P. Nedelec, J.-L. Redelsperger, F. Ravegnani, C.E. Reeves, R. Roca, P. de Rosnay, H. Schlager, B. Sultan, M. Tomasini, A. Ulanovsky, ACMAD forecasters team, 2008: Large-scale overview of the summer monsoon over West Africa during the AMMA field experiment in 2006. – Ann. Geophys. **26**, 2569–2595.

Karbou, F., F. Rabier, J.-P. Lafore, J.-L. Redelsperger, O. Bock, 2010: Global 4dvar assimilation and forecast experiments using amsu observations over land. part ii: Impacts of assimilating surface-sensitive channels on the african monsoon during amma. – Wea. Forecast. **25**, 20–36, DOI: 10.1175/2009WAF2222244.1.

Keil, C., F. Heinlein, G.C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. – Quart. J. Roy. Meteor. Soc. **140**, 480–490, DOI: 10.1002/qj.2143.

Klüpfel, V., N. Kalthoff, L. Gantner, C.M. Taylor, 2012: Convergence zones and their impact on the initiation of a mesoscale convective system in West Africa. – Quart. J. Roy. Meteor. Soc. **138**, 950–963, DOI: 10.1002/qj.979.

Kohler, M., N. Kalthoff, C. Kottmeier, 2010: The impact of soil moisture modifications on CBL characteristics in West Africa: A case study from the AMMA campaign. – Quart. J. Roy. Meteor. Soc. **136**, 442–455, DOI: 10.1002/qj.430.

Kullback, S., R.A. Leibler, 1951: On information and sufficiency. – Ann. Math. Statist. **22**, 79–86, DOI: 10.1214/aoms/1177729694.

Lafore, J.-P., C. Flamant, F. Guichard, D.J. Parker, D. Bouniol, A.H. Fink, V. Giraud, M. Gosset, N. Hall, H. Höller, S.C. Jones, A. Protat, R. Roca, F. Roux, F. Saïd, C. Thorncroft, 2011: Progress in understanding of weather systems in West Africa. – Atmos. Sci. Lett. **12**, 7–12, DOI: 10.1002/asl.335.

Le Barbé, L., T. Lebel, 1997: Rainfall climatology of the HAPEX-Sahel region during the years 1950–1990. – J. Hydrol. **188–189**, 43–73, DOI: 10.1016/S0022-1694(96)03154-X.

Lebel, T., F. Delclaux, L. Le Barbé, J. Polcher, 2000: From gcm scales to hydrological scales: rainfall variability in west africa. – Stochastic Environ. Res. Risk Assess. **14**, 275–295, DOI: 10.1007/s004770000050.

Leutbecher, M., T. Palmer, 2008: Ensemble forecasting. – J. Comput. Phys. **227**, 3515–3539, DOI: 10.1016/j.jcp.2007.02.014.

Luo, Y., Y. Chen, 2015: Investigation of the predictability and physical mechanisms of an extreme-rainfall-

producing mesoscale convective system along the meiyu front in east china: An ensemble approach. – J. Geophys. Res. **120**, 10,593–10,618, DOI: 10.1002/2015JD023584 2015JD023584.

Marsham, J.H., N. Dixon, L. Garcia-Carreras, G.M. Lister, D.J. Parker, P. Knippertz, C. Birch, 2013: The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations. – Geophys. Res. Lett. **40**, 1843–1849, DOI: 10.1002/grl.50347.

Marsigli, C., A. Montani, T. Paccagnella, 2014: Perturbation of initial and boundary conditions for a limited-area ensemble: multi-model versus single-model approach. – Quart. J. Roy. Meteor. Soc. **140**, 197–208, DOI: 10.1002/qj.2128.

Mason, S.J., N.E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. – Wea. Forecast. **14**, 713–725.

Mathon, V., H. Laurent, T. Lebel, 2002: Mesoscale convective system rainfall in the Sahel. – J. Appl. Meteor. **41**, 1081–1092, DOI: 10.1175/1520-0450(2002)041<1081:MCSRIT>2.0.CO;2.

Maurer, V., N. Kalthoff, L. Gantner, 2015: Predictability of convective precipitation for West Africa: Does the land surface influence ensemble variability as much as the atmosphere?. – Atmos. Res. **157**, 91–107, DOI: 10.1016/j.atmosres.2015.01.016.

Molteni, F., R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. – Quart. J. Roy. Meteor. Soc. **127**(576), 2069–2094, DOI: 10.1002/qj.49712757612.

Montani, A., D. Cesari, C. Marsigli, T. Paccagnella, 2011: Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. – Tellus A **63**, 605–624, DOI: 10.1111/j.1600-0870.2010.00499.x.

Murphy, A., 1973: A new vector partition of the probability score. – J. Appl. Meteor. **12**, 595–600, DOI: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

Nicholson, S.E., 2013: The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability. – ISRN Meteor. **2013**, 453521, 32 pp. DOI: 10.1155/2013/453521.

Parker, D.J., A. Fink, S. Janicot, J.-B. Ngamini, M. Douglas, E. Afiesimama, A. Agustí-Panareda, A. Beljaars, F. Dide, T. Diedhiou, A.and Lebel, J.-L. Polcher, J. Redelsperger, C.D. Thorncroft, G.A. Wilson, 2008: The AMMA radiosonde program and its implications for the future of atmospheric monitoring over Africa.. – Bull. Amer. Meteor. Soc. **89**, 1015–1027, DOI: 10.1175/2008BAMS2436.1.

Pearson, K.J., G.M.S. Lister, C.E. Birch, R.P. Allan, R.J. Hogan, S.J. Woolnough, 2014: Modelling the diurnal cycle of tropical convection across the "grey zone". – Quart. J. Roy. Meteor. Soc. **140**, 491–499, DOI: 10.1002/qj.2145.

Reed, R.J., D.C. Norquist, E.E. Recker, 1977: The structure and properties of African wave disturbances as observed during Phase III of GATE. – Mon. Wea. Rev. **105**, 317–333, DOI: 10.1175/1520-0493(1977)105<0317:TSAPOA>2.0.CO;2.

Reichle, R.H., R.D. Koster, 2004: Bias reduction in short records of satellite soil moisture. – Geophys. Res. Lett. **31**, L19501, DOI: 10.1029/2004GL020938.

Roberts, N.M., H.W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. – Mon. Wea. Rev. **136**, 78–97, DOI: 10.1175/2007MWR2123.1.

Schumacher, R.S., A.J. Clark, M. Xue, F. Kong, 2013: Factors influencing the development and maintenance of nocturnal heavy-rain-producing convective systems in a storm-scale ensemble. – Mon. Wea. Rev. **141**, 2778–2801, DOI: 10.1175/MWR-D-12-00239.1.

Schwartz, C.S., J.S. Kain, S.J. Weiss, M. Xue, D.R. Bright, F. Kong, K.W. Thomas, J.J. Levit, M.C. Coniglio, M.S. Wandishin, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. – Wea. Forecast. **25**, 263–280, DOI: 10.1175/2009WAF2222267.1.

Shutts, G., M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, M. Bonavita, 2011: Representing model uncertainty: stochastic parametrizations at ECMWF. – ECMWF Newsletter **129**, 19–24.

Taylor, C.M., R.A. de Jeu, F. Guichard, P.P. Harris, W.A. Dorigo, 2012: Afternoon rain more likely over drier soils. – Nature Geosci. **489**, 423–426, DOI: 10.1038/nature11377.

Tennant, W., S. Beare, 2014: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. – Quart. J. Roy. Meteor. Soc. **140**, 1150–1160, DOI: 10.1002/qj.2202.

Theis, S., C. Gebhardt, 2011: Probabilistische Wettervorhersage, 6: COSMO-DE-EPS: Das „konvektionserlaubende" Ensemble am DWD. – promet **37**, 53–61, ISSN 2194-5950.

Thorncroft, C.D., N.M. Hall, G.N. Kiladis, 2008: Three-dimensional structure and dynamics of African easterly waves. Part III: Genesis. – J. Atmos. Sci. **65**, 3596–3607, DOI: 10.1175/2008JAS2575.1.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. – Mon. Wea. Rev. **117**, 1779–1800, DOI: 10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Tompkins, A.M., A. Diongue-Niang, D.J. Parker, C.D. Thorncroft, 2005: The African easterly jet in the ECMWF Integrated Forecast System: 4D-Var analysis. – ISSN 2194-5950 **131**, 2861–2885.

Torn, R.D., 2010: Ensemble-based sensitivity analysis applied to African easterly waves. – Wea. Forecast. **25**, 61–78.

Vié, B., O. Nuissier, V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions. – Mon. Wea. Rev. **139**, 403–423, DOI: 10.1175/2010MWR3487.1.

Weisman, M.L., W.C. Skamarock, J.B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. – Mon. Wea. Rev. **125**, 527–548, DOI: 10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Wernli, H., M. Paulat, M. Hagen, C. Frei, 2008: SAL–a novel quality measure for the verification of quantitative precipitation forecasts. – Mon. Wea. Rev. **136**, 4470–4487, DOI: 10.1175/2008MWR2415.1.

Wernli, H., C. Hofmann, M. Zimmer, 2009: Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL technique. – Wea. Forecasting **24**, 1472–1484, DOI: 10.1175/2009WAF2222271.1.

Wilks, D.S., 2011: Statistical methods in the atmospheric sciences. International geophysics series; 100. – Elsevier Acad. Press, Amsterdam, 676 pp.