An evaluation of operational and research weather forecasts for southern West Africa using observations from the DACCIWA field campaign in June–July 2016

Anke Kniffka[1] | Peter Knippertz[1] | Andreas H. Fink[1] | Angela Benedetti[2] | Malcolm E. Brooks[3] | Peter G. Hill[4] | Marlon Maranan[1] | Gregor Pante[1] | Bernhard Vogel[1]

[1]Research Centre Human Biometeorology, Deutscher Wetterdienst, Freiburg, Germany

[2]European Centre for Medium-Range Weather Forecasts, Reading, UK

[3]Met Office, Exeter, UK

[4]Department of Meteorology, University of Reading, Reading, UK

Correspondence
Anke Kniffka, Research Centre Human Biometeorology, Deutscher Wetterdienst, Freiburg, Germany
Email: anke.kniffka@dwd.de

Funding information
European Union 7th Framework Programme 603502 (EU project DACCIWA)

Abstract

Reliable and accurate weather forecasts, particularly those of rainfall and its extremes, have the potential to improve living conditions in densely populated southern West Africa (SWA). The limited availability of observations has long impeded a rigorous evaluation of current state-of-the-art forecast models. The field campaign of the Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa (DACCIWA) project in June–July 2016 has created an unprecedentedly dense set of measurements from surface stations and radiosondes. Here we present results from a comprehensive evaluation of both numerical model forecasts and satellite products using these data on a regional and local level. Results reveal a substantial observational uncertainty showing considerable underestimations in satellite estimates of rainfall and low-cloud cover with little correlation at the local scale. Models have a dry bias of 0.1–1.9 mm · day^{-1} in rainfall and too low column relative humidity. They tend to underestimate low clouds, leading to excess surface solar radiation of 43 W · m^{-2}. Remarkably, most models show some skill in representing regional modulations of rainfall related to synoptic-scale disturbances, while local variations in rainfall and cloudiness are hardly captured. Slightly better results are found with respect to temperature and for the post-onset rather than for the pre-onset period. Delicate local features such as the Maritime Inflow phenomenon are also rather poorly represented, leading to too cool, dry and cloudy conditions at the coast. Differences between forecast days 1 and 2 are relatively small and hardly systematic, suggesting a relatively quick error saturation. Using explicit convection leads to more realistic spatial variability in rainfall, but otherwise no marked improvement. Future work should aim at improving the subtle balance between the diurnal

cycles of low clouds, surface radiation, the boundary layer and convection. Further efforts are also needed to improve the observational system beyond field campaign periods.

1 | INTRODUCTION

Many developing countries in the Tropics are strongly affected by variations in rainfall, temperature, wind and cloudiness, and have low resilience against weather extremes (e.g., Webster, 2013). Reliable and accurate predictions therefore have the potential to significantly improve the livelihood of these populations. However, current precipitation forecasts have overall poor skill at low latitudes (Haiden *et al.*, 2012). This is particularly true for northern tropical Africa, for which Vogel *et al.* (2018) recently showed essentially no skill in ensemble forecasts by nine global models and thus hardly any potential to improve forecasts through statistical postprocessing beyond simple climatological approaches. Interestingly, this lack of skill may negatively impact medium-range forecasts in neighbouring areas such as the North Atlantic and Europe (Faccani *et al.*, 2009; Fink *et al.*, 2011; Pante and Knippertz, 2019). The current state of weather forecasting in West Africa has recently been summarised by Parker and Diop-Kane (2017). Generally speaking, the use of numerical weather prediction (NWP) is still in its infancy across the region, including densely populated southern West Africa (SWA), the focus of this paper. National weather services in SWA have free access to operational forecast products from global models (e.g. through EUMETCast), but regional NWP systems are rare. One exception is the Nigerian Meteorological Agency that operates the Consortium for Small-scale Modeling (COSMO) model with 7 km grid-spacing (Olaniyan *et al.*, 2015). In addition, private initiatives are now beginning to enter the West African weather market (e.g., https://www.ignitia.se; accessed 30 December 2019).

One key obstacle to evaluate and develop models is the limited availability of observations. The station network in SWA is generally sparse and some existing data do not routinely enter international databases (e.g., Parker *et al.*, 2008; Knippertz *et al.*, 2015). Satellite-based rainfall estimates are known to have sensor-specific biases, particularly in the semi-arid Sahel and in coastal regions of the Tropics with rainfall from warmer clouds (Thiemig *et al.*,

2012; Maggioni *et al.*, 2016). Previous studies have also shown that satellite-derived surface solar irradiance estimates in SWA have large errors due to difficulties in assessing low-level cloud fractions and aerosol contents (Knippertz *et al.*, 2011; Hannak *et al.*, 2017). The lack of adequate observations has impacts on model initialisation from data assimilation, forecast verification and model development. It has been shown that analysis products differ widely over this part of the world due to poor model first guesses and insufficient observational constraints (Roberts *et al.*, 2015). This can even lead to a fundamentally wrong water budget in some NWP products (Meynadier *et al.*, 2010). Additional data from field campaigns (e.g., radiosondes, aircraft dropsondes) can improve analysis fields substantially, but the positive impact on forecasts typically does not last for more than two days (Thorncroft *et al.*, 2003; Faccani *et al.*, 2009; Agustí-Panareda *et al.*, 2010; Karbou *et al.*, 2010), pointing to substantial model error.

What makes weather forecasting in West Africa so challenging? On regional scales, SWA's weather is determined by the complex West African Monsoon (WAM) system and its intraseasonal variations (e.g., due to tropical waves; Schlueter *et al.*, 2019a). Changes in moisture availability, instability and shear, on seasonal but also on shorter time-scales, modulate the conditions for the formation of organised mesoscale convective systems (Maranan *et al.*, 2018; Schlueter *et al.*, 2019b). Numerical models with convective parametrization struggle to realistically represent cold pools that play a crucial role in the organisation process through triggering new systems at the edge of existing ones (Garcia-Carreras *et al.*, 2013; Marsham *et al.*, 2013; Birch *et al.*, 2014). Dust radiative effects can also influence this process (e.g., Shi *et al.*, 2014; Reinares Martínez and Chaboureau, 2018).

Another challenge is the sensitive relationship of deeper clouds with the land surface and the diurnal cycle of the planetary boundary layer (PBL; Couvreux *et al.*, 2014) as well as with the widespread occurrence of extensive decks of low-level clouds (e.g., Schrage and Fink, 2012; van der Linden *et al.*, 2015). During the summer monsoon these clouds typically form at night due to a combination

of cold advection, long-wave radiative cooling and turbulent mixing underneath the nocturnal low-level jet, and then lift and dissolve in the course of the day (Schuster *et al.*, 2013; Adler *et al.*, 2017; Dione *et al.*, 2018; Adler *et al.*, 2019). Hill *et al.* (2018) recently quantified the effect of these clouds on the downwelling short-wave irradiance at the surface to be 35 W · m^{-2}. Given the usually close to moist-neutral state of the atmosphere, small variations in the optical thickness of these clouds can have substantial impacts on rainfall (Kniffka *et al.*, 2019). Climate models struggle to realistically represent these low-level clouds and their diurnal cycle (Knippertz *et al.*, 2011; Hannak *et al.*, 2017), but a comprehensive evaluation for NWP models is lacking.

Day-to-day variations in meteorological conditions are often caused by local features such as soil moisture variations and other surface characteristics (Lavender *et al.*, 2010) or incoming solar radiation (Taylor *et al.*, 2011; Lafore *et al.*, 2017). For the Weather Research and Forecasting (WRF) model, Li *et al.* (2015) showed that the choice of the radiation scheme influences the north–south gradient in surface temperature and thus the strength of the monsoon flow. An important feature along the Guinea Coast is the land–sea breeze (Guedje *et al.*, 2019), which in summer interacts with the monsoon flow to form the Maritime Inflow of the Gulf of Guinea, a stationary front about 30 km from the coast that begins to propagate inland in the late afternoon (Adler *et al.*, 2017). The stationarity is caused by the strong daytime turbulence over land (Deetz *et al.*, 2018).

Representing all these features and their interactions in NWP forecasts is challenging and needs to be evaluated, but so far only a few systematic studies have been published for the region, while climate model evaluations are more common (e.g., Roehrig *et al.*, 2013; Hill *et al.*, 2016; Hannak *et al.*, 2017). Comparing seven NWP models of the TIGGE repository (Bougeault *et al.*, 2010) in ensemble prediction mode to the satellite-derived TRMM dataset for the years 2008–2012 Louvet *et al.* (2016) found satisfactory performance in forecasting regional-scale features, i.e. the seasonal cycle of monsoon precipitation in terms of latitudinal shift of the rain band and onset of the monsoon. Milton *et al.* (2017) systematically investigated 1200 UTC control forecasts for June–September 2012 from five operational centres participating in TIGGE, using data on a 1° grid. Comparisons between 1- and 8-day predictions show significant model drifts in temperature, moisture, pressure and precipitation, with a tendency for rainfall overestimation at longer lead-times. Predictions struggle to reproduce area-mean day-to-day variations with very low correlations for all models. Applying a 5-day smoothing, forecasts become significantly better with the exception of SWA, where correlations are low even for 10-day smoothing.

Consistently, Vogel *et al.* (2018) demonstrated that relatively simple probabilistic forecasts based on observations (rain gauges, satellite estimates) alone can outperform TIGGE ensemble precipitation forecasts for spatial aggregations up to 2° × 5° and temporal aggregations up to five days. However, both Mutemi *et al.* (2007) and Vogel *et al.* (2018) find added value in multi-model ensemble approaches relative to single models.

Many studies document problems with the diurnal cycle of precipitation when using models with parametrized convection (Marsham *et al.*, 2013; Bechtold *et al.*, 2014; Kouadio *et al.*, 2018; Kniffka *et al.*, 2019). Fast physics errors cause significant biases within the first 24 hr and these then affect the northward advection of moisture into the Sahel and the Sahara. This shifts the main rainband latitudinally with the terms of the water budgets acting to reinforce the initial biases (Birch *et al.*, 2014). Söhne *et al.* (2008) compared one month of forecasts using a regional model with 32 km grid spacing with satellite-derived brightness temperatures and found too many/too thick low clouds over SWA (between 5 and 10°N), too low surface temperatures, a too shallow PBL, reduced convective available potential energy (CAPE) and reduced deep convection in agreement with the negative feedback found in the sensitivity study by Kniffka *et al.* (2019). In addition, their model showed overall sharper meridional gradients, a too fast monsoon flow, too little vertical mixing, limiting speed reduction and drying. They also concluded that a forcing by African easterly waves (AEWs) enhances the predictability of high clouds. For a similar grid spacing of 0.5°, Druyan *et al.* (2010) found large overestimation of the low-level moisture advection into the West African continent which is accompanied by excessive rainfall. Finally, forecast studies using explicit convection showed more realistic convective features and reasonable agreement with observations, but a number of issues remain (Beucher *et al.*, 2014; Kouadio *et al.*, 2018; Maurer *et al.*, 2017). The latter work specifically explored the potential of using ensemble approaches at this resolution.

To overcome the notorious lack of observations in SWA, the Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa (DACCIWA) project (Knippertz *et al.*, 2015) organised a major international field campaign in June–July 2016. Amongst other things, the campaign included three research aircraft, three ground supersites and a substantial enhancement of the radiosonde network (Flamant *et al.*, 2018; Kalthoff *et al.*, 2018). In addition, specific efforts were made to obtain station observations from national weather services and research projects, and to digitise data only available on paper. The campaign included the provision of near-real-time forecast information from several different models through a dedicated

webpage (http://dacciwa.sedoo.fr; accessed 30 December 2019), where observations from the DACCIWA supersites and satellite products were made available to aid flight and other campaign planning activities. This combination offers a unique opportunity to obtain new insights into forecast quality over SWA. The detailed and comprehensive evaluation we present here covers a range of meteorological variables using a variety of graphical displays and scores also used in operational services. Given the uncertainty in analysis data shown by various authors, we will concentrate on direct observations from ground, balloon and space for the evaluation. Particular attention will be paid to the coupling of low clouds, radiation, temperature and precipitation, as discussed by Söhne *et al.* (2008) and Kniffka *et al.* (2019).

The article is structured as follows. In Section 2.1 a brief description of the modelling systems participating in this exercise is given. Section 2.2 provides information on the observational datasets used for the evaluation (mostly stations, radiosondes and satellites) and the methods applied. The evaluation results are shown in Section 3, organised by observational source (surface stations, radiosondes) and type of examination (phenomenological or statistical). Finally, a short summary and conclusions are given in Section 4.

2 | DATA AND METHODS

The DACCIWA campaign covered the period from 01 June to 31 July 2016 with some variations in the density of available observations. Nevertheless, evaluation results will be presented for the entire period. As discussed in Knippertz *et al.* (2017), the campaign covered the pre- and post-onset phases of the WAM. During the former, the main rain belt was located south of 7.5°N and then migrated north towards the Sahel during onset. For this reason, some analyses in this paper are divided into the periods 01–21 June (period 1) and 22 June–31 July 2016 (period 2). All forecasts were evaluated within a rectangular region 5°N–10°N; 8°W–8°E which encompasses parts of Côte d'Ivoire, Ghana, Togo, Benin and Nigeria as shown by the yellow box in Figure 1. The following two subsections provide detailed information about the model and observational datasets used in this study.

2.1 | Models

In this evaluation exercise, both operational NWP models and research models which were run to support the DACCIWA field campaign will be compared (five in total). An overview of model type, treatment of
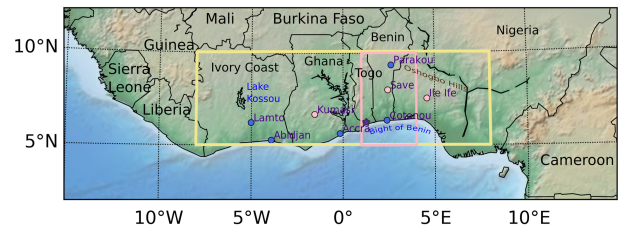

FIGURE 1 Topographic map of SWA. The DACCIWA region is marked with a yellow rectangle. Pink dots indicate the position of the three ground supersites Kumasi, Savé and Ile-Ife, where intense observations of the boundary layer were carried out (Kalthoff *et al.* 2018), while blue dots denote the radiosonde stations and the purple star marks Lomé airport in Togo, where the research aircrafts were based (Flamant *et al.* 2018). The pink rectangle depicts the transect region used for Figures 8–10

convection, model domain and resolution is given in Table 1, while Table 2 provides details on the most important parametrization schemes and the employed initial conditions. All forecasts are deterministic in nature. Three model outputs are from global operational forecasts started from their respective operational analysis using full data assimilation (IFS, UKMO, ICON OPS), while COSMO-ART is a limited-area model initialised with ICON operational runs. Due to the inclusion of aerosol and trace-gas processes, this model is run at a relatively coarse grid spacing of 28 km. One of the models (ICON KIT) has a high enough resolution to allow for explicit deep moist convection in a nest centred on West Africa, where a horizontal grid spacing of 6.6 km is used (Table 1). All models are initialised with prescribed sea-surface temperatures (SSTs) at the start of each forecast with no updates during run time. UKMO and IFS use the OSTIA SST analysis provided by the UK Met Office (Donlon *et al.*, 2011). The research model ICON KIT was initiated with the respective fields from the operational IFS run, and therefore is indirectly also initialised with OSTIA. However, the operational ICON uses its own 3D variational method, where SST is derived using the NCEP analysis together with buoy and ship records. COSMO was initiated with operational runs from ICON and therefore shares the same SST and temperature initialisation. To make the models as comparable as possible, all output was regridded to the same 0.2° grid and model runs were started at 1200 UTC. The first 6 hr were considered to be spin-up time, which means that the validation time periods were: 1800 UTC from the same day to 1800 UTC the following day (forecast day 1) and 1800 UTC at the end of forecast day 1 plus 24 hr (forecast day 2). Table 2 shows a wide range of different parametrizations used in the different models. The crucial aspect of deep moist convection is represented through related approaches in IFS, ICON and

COSMO-ART (mass-flux scheme of Tiedtke and Bechtold; Tiedtke 1989), while UKMO uses the mass-flux scheme by Gregory and Rowntree (1990). Due to the mutual and often nonlinear interactions between the different parametrizations, it is practically impossible to trace back differences in predictive performance to individual aspects of model parametrizations. The following subsections will provide further details of the five model datasets investigated here.

2.1.1 | IFS

Operational forecasts from the Integrated Forecasting System (IFS), which was developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and Météo-France, were used by the DACCIWA team during the field campaign period in June–July 2016. At that time, model cycle 41r2 was operational (ECMWF, 2016). This means that the meteorological fields from the high-resolution deterministic runs possess a horizontal grid-spacing of 9 km with 137 vertical levels. Tailored plots of meteorological, aerosol and chemical variables were produced for the campaign in near-real time. In this study, only meteorological variables are evaluated; a validation of atmospheric chemistry can be found in Menut *et al.* (2018) and Deroubaix *et al.* (2019).

2.1.2 | UKMO

UK Met Office (UKMO) forecasts were provided in near-real time from the operational global NWP version of the Unified Model (Cullen and Davies, 1991; Wood *et al.*, 2014), which is suitable for atmospheric prediction on a wide range of temporal and spatial scales (Brown *et al.*, 2012). The version used has a horizontal resolution of 0.23° longitude by 0.16° latitude, which corresponds to approximately 25 km over SWA. The global model has 70 levels, reaching up to 85 km. The standard meteorological fields are initialised using a hybrid ensemble 4D-Var data assimilation system described in Rawlins *et al.* (2007)

and Clayton *et al.* (2012). In addition, experimental forecasts including additional aerosol species were provided to DACCIWA but are not analysed here.

2.1.3 | ICON

The Icosahedral Nonhydrostatic (ICON) model (Zängl *et al.*, 2015) is a global NWP system recently developed by the Max Planck Institute for Meteorology and the German Weather Service. It can be used on a wide temporal and spatial range from large-eddy simulation studies to climate predictions. Here, we evaluate output from both the global NWP model operational in 2016 and from higher-resolution research simulations run at the Karlsruhe Institute of Technology (KIT). The former (ICON OPS hereafter) has a mesh size of 13 km and 90 vertical levels, with 11 levels up to the first 1,000 m. An ensemble assimilation system, a hybrid combination of an ensemble Kalman filter with a variational procedure, is used. The research version (ICON KIT hereafter) is initialised with the IFS operational analysis fields and run for 54 hr. In contrast to the operational configuration, it includes a two-way nested domain over SWA with a grid spacing of 6.6 km that allows for explicit convection.

2.1.4 | COSMO-ART

COSMO-ART (Aerosols and Reactive Trace gases) is an online-coupled extension to the limited-area NWP model COSMO (Baldauf *et al.*, 2011) that calculates tracer dispersion (Vogel *et al.*, 2009). COSMO-ART includes a comprehensive chemistry module to describe the gaseous composition of the atmosphere and secondary aerosol formation. It includes impacts of aerosol particles on radiation, cloud formation and precipitation (e.g., Stanelle *et al.*, 2010; Athanasopoulou *et al.*, 2014; Rieger *et al.*, 2014; Walter *et al.*, 2016). The DACCIWA simulations were initialised with operational, i.e., ICON OPS, forecasts and had a horizontal resolution of 28 km on 50 vertical levels.

TABLE 1 Properties of the evaluated models

Model	Type	Convection	Domain	Horiz. resolution (km)	Vertical levels
IFS	Operational	Parametrized	Global	9	137
UKMO	Operational	Parametrized	Global	25	70
ICON OPS	Operational	Parametrized	Global	13.2	90
ICON KIT	Research	Explicit	Global + nested:	6.6	90
			22°W–28°E, 4°S–28°N		
COSMO-ART	Research	Parametrized	25°W–4°E, 20°S–35°N	28	50

TABLE 2 Overview of the most important parametrizations employed in the models used for this study. The right column provides information about the initial conditions used.

Model	Cloud	Radiation	Boundary layer	Initialisation
IFS	Convection: mass-flux approach (Tiedtke, 1989; Bechtold et al., 2008), diurnal cycle improved by CAPE closure adaption (Bechtold et al., 2014). Cloud and large-scale precipitation scheme: Tiedtke (1993) + improvements (Tompkins et al., 2007; Forbes et al., 2011; Forbes and Ahlgrimm, 2014).	Rapid Radiation Transfer Model (RRTM; Mlawer et al., 1997; Iacono et al., 2008), gaseous absorption with correlated-k scheme, 112 short-wave bands and 140 long-wave bands. Cloud–radiation interaction: Monte Carlo Independent Column Approximation method (Morcrette et al., 2008)	Turbulent mixing scheme by eddy-diffusivity mass-flux, with K-diffusion turbulence closure and mass-flux component (Siebesma et al., 2007; Köhler et al., 2011). Tile-based surface-exchange parametrization with TESSEL (Viterbo and Beljaars, 1995; van den Hurk et al., 2000)	Operational analysis with full data assimilation
UKMO	Convection: mass-flux approach (Gregory and Rowntree, 1990). Cloud and large-scale precipitation scheme: prognostic cloud fraction and condensate (PC2) scheme (Wilson et al., 2008a; 2008b; Morcrette, 2012)	Edwards and Slingo (1996) with configuration based on Cusack et al. (1999) + updates. Gaseous absorption: correlated-k method (6 short-wave bands and 9 long-wave bands).	First-order turbulence closure (Lock et al., 2000) with modifications (Lock, 2001; Brown et al., 2008)	Operational analysis with full data assimilation
ICON OPS	Convection: mass-flux approach (Bechtold et al., 2008). Cloud condensation: saturation adjustment. Microphysics: extension of COSMO-EU five-category prognostic scheme (Doms and Schättler, 2002; Seifert, 2008)	RRTM radiation scheme (Mlawer et al., 1997) with 30 short-wave bands and 16 long-wave bands. Gaseous absorption: correlated-k scheme	Prognostic equation for TKE (Raschendorfer, 2001). Tile-based land-surface scheme TERRA (Heise et al., 2006)	Operational analysis with full data assimilation
ICON KIT	As above but explicit convection	As above	As above	Initialised with operational global IFS analysis
COSMO-ART	Convection: Tiedtke scheme (Tiedtke, 1989). Cloud condensation: saturation adjustment. Microphysics: one-moment bulk scheme (Doms et al., 2011), empirical formulation for subgrid-scale clouds	δ-two-stream code GRAALS by Ritter and Geleyn (1992) for short-wave (3 bands) and long-wave (5 bands) including gaseous absorption	Prognostic 2.3-TKE closure. Surface layer scheme with roughness layer using Mellor–Yamada closure (Doms et al., 2011)	Initialised with ICON operational analysis

They were used to assist in chemical aspects of flight planning during the DACCIWA campaign.

2.2 | Observations and evaluation methods

The measurements considered in this study are mostly from ground stations, radiosondes and satellites. They stem from a range of sources and are heterogeneous in terms of spatial and temporal resolutions. The following subsections give detailed information on each type and on how the data were processed in order to compare them with gridded model data in an optimal way.

2.2.1 | Station observations

Low-level cloud cover $lclc$, 2m temperature T_{2m} and 2m dew point temperatures Td_{2m}, as well as precipitation observations rr stem from various ground-based sources. Among these are stations operated by the National Weather Services and ASECNA (Agence pour la Sécurité de la Navigation Aérienne en Afrique et à Madagascar) but also project-related measurements. Concerning the former, DACCIWA made an effort to collect additional data from the archives of weather services in Côte d'Ivoire, Ghana, Togo, Benin and Nigeria, part of which had to be digitised. rr data were also obtained from AMMA CATCH (http://www.amma-catch.org) for a few dozen stations in the area of the Haute Vallée de l'Oémé in central Benin. The solar surface radiation (SSI) data stem from the DACCIWA supersites Kumasi, Savè and Ile-Ife. Long-term meteorological radiation measurements including the campaign period were available at Lamto, Cotonou and Parakou (Kniffka *et al.*, 2019) as well as at the two AMMA-CATCH stations Nalohou and Djougou (Galle *et al.*, 2018). However, the majority of radiation measurements were located in southwestern Ghana and taken from the Trans-African Hydro-Meteorological Observatory (TAHMO) project database (van de Giesen *et al.*, 2014). The instruments used at the different stations vary, but are all qualified within the World Meteorological Organisation standard for radiation measurements.

The irregular distribution of surface stations poses a challenge to the evaluation strategy. Simply averaging over all stations (in the case of rainfall for example), would put too much weight on Benin due to the high station density in the AMMA CATCH area (Figure 2). A better approach is to divide the evaluation area into boxes, average the station observations within these boxes and then compare to the corresponding model average. In order to find the ideal box size, the DACCIWA region (Figure 1) was

iteratively halved and evaluation results computed. After a certain number of iterations, a convergence should be reached where results do not change much for further divisions. In our case, this was reached after five iterations, corresponding to 64 boxes of about $2° \times 0.75°$.

2.2.2 | Radiosondes

Radiosondes were usually launched 1–4 times per day (sometimes even more often) at the locations Lamto, Abidjan, Accra, Kumasi, Cotonou, Savè and Parakou (Figure 1). Observation periods and total numbers of sondes launched vary from station to station (Table 3). From these upper-air data, profiles of specific humidity q, relative humidity rh, temperature T, wind speed v and wind direction $vdir$ were derived. The data were sent to the Global Telecommunication System (GTS) in real time, thus the operational model versions (Table 1) assimilated available extra radiosondes. Details on the radiosonde campaign can be found in Flamant *et al.* (2018). Collocations of radiosonde data at the main synoptic hours 0000, 0600, 1200 and 1800 UTC and the corresponding model profiles were determined individually for each station. Since not all radiosonde stations were operational from the start of the campaign and due to a few failures, the resulting time series differ in length and in number of missing data. Nevertheless, we decided to use all available launches during the campaign rather than restricting the analysis to times of overlap. Until 24 June, launches occurred only once per day and twice in Abidjan. Only station Lamto had less than 60% coverage and was therefore excluded from this study.

2.2.3 | Satellite data

To complement the station measurements, satellite estimates of $lclc$ and rr were considered. For rr we concentrated on the Global Precipitation Measurement (GPM) IMERG (Integrated Multi-satellite Retrievals for GPM) dataset version 4.4, the successor product of the well-known TRMM (Tropical Rainfall Measuring Mission) 3b43 product. It combines data from the precipitation radar aboard the GPM satellite with microwave and infrared sensors on several low earth orbiting and geostationary satellites (Huffman *et al.*, 2015). The resolution is half hourly on a $0.1° \times 0.1°$ grid.

For $lclc$ we use the Optimal Cloud Analysis (OCA) (Watts *et al.*, 2011) product, which is based on measurements from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) on board the geostationary MeteoSat Second Generation satellite. The OCA product uses simultaneous information from multiple SEVIRI channels in
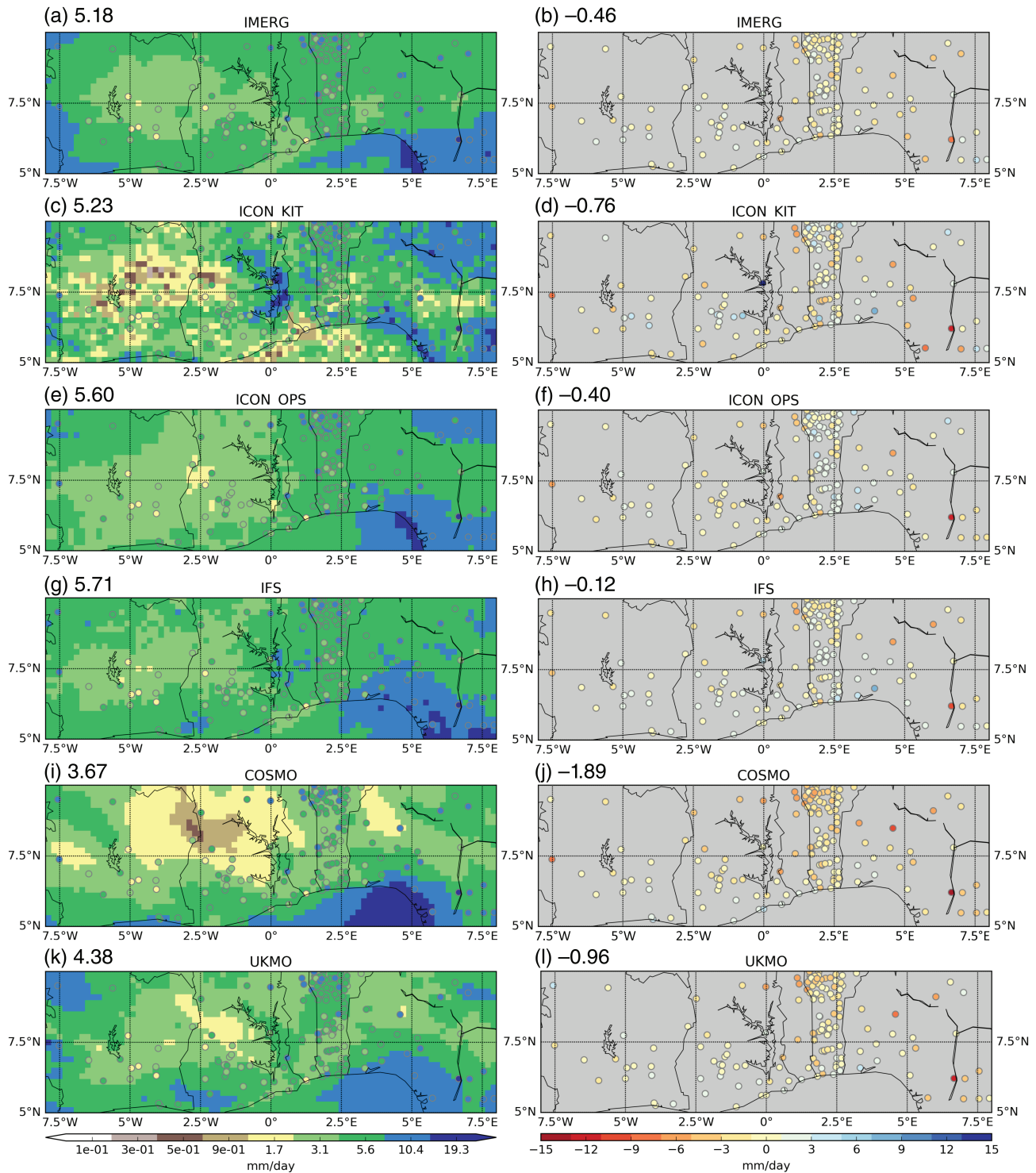
FIGURE 2 Evaluation of rainfall biases for June–July 2016. Left columns show horizontal distributions over the DACCIWA region of average daily rainfall (mm) for the IMERG satellite estimates and the five forecast datasets. Station observations are shown as filled circles using the same colour shading, and area averages are provided as numbers at the top left. The right column shows the corresponding biases relative to the station observations with average biases at the top left. The latter are computed for 64 boxes as explained in Section 2.2

an optimal estimation framework to produce physically consistent estimates of cloud physical properties at high native temporal (15 min) and native spatial (3 km at nadir) resolution. The operational Meteosat satellites are situated at 0°N and 0°E. This point falls into the analysed region, therefore the pixel size of OCA is about 3 × 3 km in the entire domain. The retrieval uses information from 11 channels from 0.6 to 13.4 μm wavelength.

TABLE 3 Radiosonde stations used in this study: observing period in June–July 2016 and total number of launches

Station	Start date	End date	No. of launches
Abidjan	01 June	31 July	135
Accra	16 June	31 July	177
Kumasi	11 June	30 July	131
Cotonou	01 June	21 July	120
Savè	13 June	30 July	145
Parakou	20 June	18 July	97

Additionally, the number of channels used varies from single- to double-layer cloud retrieval mode. This leads to varying error types and magnitudes, which will not be discussed in detail, since only averages over periods are considered in this analysis. As a very advantageous feature, OCA attempts to identify up to two vertical cloud layers for each pixel, which may help to allay some of the problems that other cloud products have with high cloud obscuring low clouds in this region (van der Linden *et al.*, 2015) or with optically thick aerosol layers. Cloud cover from datasets based on polar-orbiting satellites did not deliver a high enough data density for the purpose of a forecast evaluation in a two-monthly time-frame. From the cloud-top pressure data of OCA, we estimated low-level cloud fraction through the following procedure. First, the temporal resolution is adapted to that of the models. This resampling is done to mimic the model output procedure, namely instantaneous fields every 3 hr. This sampling may lead to less certain absolute averages, but enables an as-close-as-possible comparison between models and satellite measurements. As a next step, a mask for low-level clouds is created. A pixel is counted to be filled with low clouds, if the lowest available cloud top is below 796.8 hPa, following the World Meteorological Organisation (WMO) standard for cloud observation (WMO, 2017) and using the US 1976 standard atmosphere (NOAA/NASA, 1976). This procedure counts all clouds in one single- or two-layer mode with tops below the threshold, but misses cases with a large vertical extent. From the cloud mask, a cloud fraction is created using the 16 (4 by 4) original pixels for one output pixel. This way, cloud fractional cover is created for each SEVIRI scene in roughly the models' horizontal grid spacing of 0.2° × 0.2° and processed further like the model data.

It should be noted that for OCA we require a cloud-top pressure greater than 796.8 hPa, while in the station measurements low clouds are defined to have a base below 796.8 hPa. Finally OCA-based *lclc* was aggregated to the model resolution.

3 | RESULTS

In this section, the evaluation of the meteorological forecasts will be presented in three steps. First, systematic model biases are assessed with respect to synoptic station and satellite data. Second, the biases in vertical profiles will be investigated using radiosonde data from the DACCIWA field campaign. Third, a day-to-day forecast evaluation will be shown based on more traditional statistical quality measures.

3.1 | Bias relative to surface stations and satellite data

3.1.1 | Precipitation

Arguably, the most important but at the same time most challenging task is forecasting of *rr*. For June and July 2016 daily *rr* sums measured from 0600 UTC to 0600 UTC from 155 stations were collected. Model data were aggregated accordingly for comparison. The left panels of Figure 2 show mean daily precipitation distribution for the whole of June and July 2016 from the five model datasets, from stations (coloured dots) and from IMERG (a). The corresponding biases relative to the stations are shown in the right panels. Numbers on top of each panel give area averages. For the differences, these were produced from 64 rectangular boxes as described in Section 2.2 in order to give equal weights to different regions.

The observations show a relatively homogeneous distribution of rainfall across the DACCIWA region with values on the order of 5 mm · day^{-1} (Figure 2a). The wettest areas are in the southeastern corner of the study region near the Niger Delta and over western Ivory Coast. The driest areas stretch from the Ghanaian coast towards Lake Volta and then west into central Ivory Coast. IMERG appears to underestimate rainfall over almost the entire land area except for the dry region in central Ivory Coast, where an overestimation can be observed. This overestimation is not very pronounced, given that the absolute values of rainfall are rather small in this region. The average difference amounts to –0.46 mm · day^{-1} (Figure 2b), corresponding to about 9% of the average rainfall. A recent study (M. Maranan, personal communication, 2020) conducted a detailed comparison between IMERG and high-resolution station observations in Ghana. They show that, despite the temporal resolution of IMERG of 30 min, the duration of rainfall events is generally overestimated, leading to too much rainfall in cases of weak and short-lived convective events. In addition, an oversensitivity to optically thin clouds leads to frequent false alarms.

On the other hand, events with high rainfall intensities such as mesoscale convective systems or short and strong convective events are underestimated. During the little dry season, when the main rain band has moved to the north of the study region, warm rain undetected by satellite products causes a high number of missed events, thereby creating a negative bias in IMERG.

The overall best agreement with the observations is found for IFS (Figure 2g,h). It largely reproduces the pattern seen in IMERG (Figure 2a) and has a very low bias relative to the stations of –0.12 mm · day^{-1} (Figure 2h). Averaged over the entire region, IFS is wetter by 0.53 mm · day^{-1} relative to IMERG (5.71 versus 5.18 mm · day^{-1}). ICON OPS (Figure 2e,f) shows similarly good results with a largely realistic pattern and mean values ranging between IMERG and the station observations. ICON KIT (Figure 2c,d) shows some structural similarities to its operational counterpart but the higher resolution and explicit representation of convection appears to lead to much more fine structure and overall more extreme maxima and minima. Wettest areas are found along the Nigerian coast and over Lake Volta, while the dry patch over inland Ghana and Ivory Coast is even drier than in observations, leading to an overall dry bias of –0.76 mm · day^{-1} relative to the stations.

The COSMO model shows a realistic land–sea and east–west pattern of precipitation but too wet conditions over ocean and too dry over land. The bias relative to station data (–1.89 mm · day^{-1}) is quite considerable. In the area average COSMO only forecasts 64% of the rainfall produced in IFS.

Finally, the UKMO model exhibits similar rainfall patterns as COSMO, but with less precipitation over the oceans and less extreme dryness west of Lake Volta. The spatial mean difference from IMERG is –0.8 mm · day^{-1} and the bias relative to stations is –0.96 mm · day^{-1}.

Figure 3 splits the biases relative to stations – as just discussed – into pre- and post-onset (periods 1 and 2) as well as into lead-times of one and two days (the latter for models only). For IMERG, the bias clearly increases from period 1 to period 2. During the latter, the main rain-band has moved into the Sahel, creating somewhat drier conditions in the DACCIWA region (4.9 versus 5.5 mm · day^{-1} in IMERG; not shown). The often light (and sometimes warm) rains during this period (Young *et al.*, 2018) appear to be particularly difficult to capture for IMERG (M. Maranan, personal communication, 2019). In contrast, the dry bias on forecast day 1 in the models tends to get smaller from period 1 to period 2. Comparing forecast days 1 and 2, there is some deterioration in bias in IFS and UKMO (only period 2), ICON KIT and OPS (mostly period 2), while COSMO even shows improvement. These results suggest that the model responds quickly
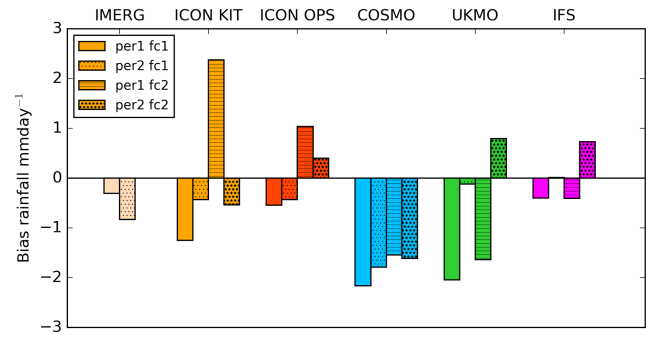

FIGURE 3 Dependence of rainfall bias on monsoon state and forecast lead-time. The biases shown for each model and IMERG are averaged over all stations using 64 gridboxes as in Figure 2, but separated into periods 1 and 2 as well as forecast days 1 and 2. Section 2 gives definitions. No forecast day 2 is plotted for IMERG

to errors in initial conditions or model physics already on day 1.

Figure 4 shows relative frequency distributions of average daily rainfall for the five forecast models and IMERG. First, rainfall data in each grid box in the study area were averaged over periods 1 and 2, and expressed as daily sums for forecast days 1 and 2. This analysis therefore emphasises the degree of spatial variability, as for example related to orography and coastlines as well as to the degree of convective organisation. Since the model data are based on three-hourly accumulated rainfall, IMERG fields were first accumulated over three hours from the half-hourly original data and then averaged and regridded to the model resolution to mimic the model analysis as closely as possible. For period 1 (Figure 4a), IMERG shows a strongly positively skewed distribution with a peak around 4–5 mm · day^{-1} and only few grid cells with low rainfall. This behaviour is most realistically reproduced by IFS but with a slightly lighter tail and little difference between forecast days 1 and 2 (the latter denoted by grey hatching). ICON OPS in contrast displays a too strong concentration on moderate rainfall on forecast day 1 and a drift to heavier rainfall on forecast day 2, indicating potential problems with unbalanced initial conditions. This tendency is also seen for ICON KIT, where imbalances may stem from the initialisation with IFS data (Table 2). The explicit convection employed here appears to lead to a much broader overall distribution (also evident from Figure 2c), which, however, has too many rather dry grid cells on forecast day 1 and too many very moist grid cells on forecast day 2. UKMO overemphasises grid cells with little rain on both forecast days, while COSMO-ART shows a shift of the entire distribution to lower values relative to IMERG (also evident in Figure 2i). Period 2 (Figure 4b), the post-onset phase, is dominated by areas with little rain, particularly in the south of the study area (Knippertz *et al.*, 2017), but
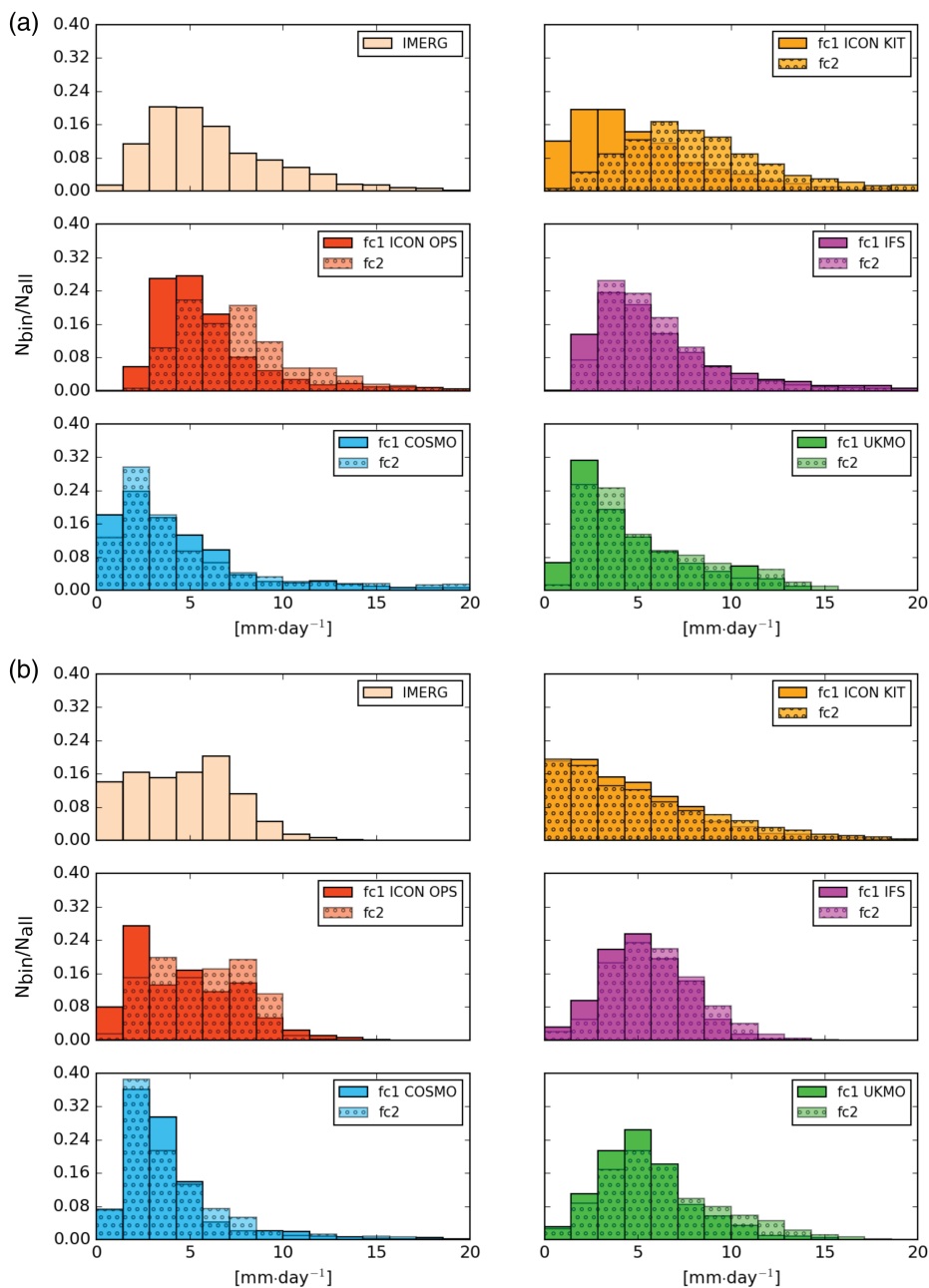
FIGURE 4 Relative frequency distributions of daily rainfall averaged over (a) period 1 and (b) period 2, showing IMERG observations (top left panels) and forecast days 1 and 2 for the five models, the latter with grey stippling. All $0.2° \times 0.2°$ degree grid cells of the study area (yellow box in Figure 1) are considered for the distribution. The bin size is $1.43 \, mm \cdot day^{-1}$

still features many grid cells with around $5 \, mm \cdot day^{-1}$ on average, most likely over northern and hilly areas. The only model with explicit convection, ICON KIT, agrees with IMERGE on the number of low-rainfall grid cells but clearly overestimates the tail of the distribution, consistently on both forecast days. In contrast, the models with parametrized convection all underestimate the dry side of the distribution with a too marked peak at moderate rainfalls. Particularly COSMO-ART overemphasises low rainfall amounts.

The area-averaged diurnal cycle of rainfall is displayed in Figure 5, (a) and (b) containing results for period 1 with forecast days 1 and 2, and (c) and (d) containing period 2 with forecast days 1 and 2. The diurnal cycle is depicted from 1800 to 1500 UTC the following day corresponding to the time-frame of the forecast days (Section 2.1). As indicated by IMERG (light orange lines in Figure 5), the diurnal cycle is relatively flat during both periods with a maximum in the afternoon (1200–1500 UTC). The former suggests a mixture of different types of rainfall with different diurnal cycles such as organised convective systems, vortex rains and land–sea breeze effects (Knippertz *et al.*, 2017; Maranan *et al.*, 2018). The latter is indicative of local convection triggered by diurnal heating. This is consistent with measurements at the DACCIWA supersites (discussed in Kalthoff *et al.*, 2018). Overall drier conditions during period 2 are evident from the diurnal cycles. The models exhibit a higher-amplitude diurnal cycle
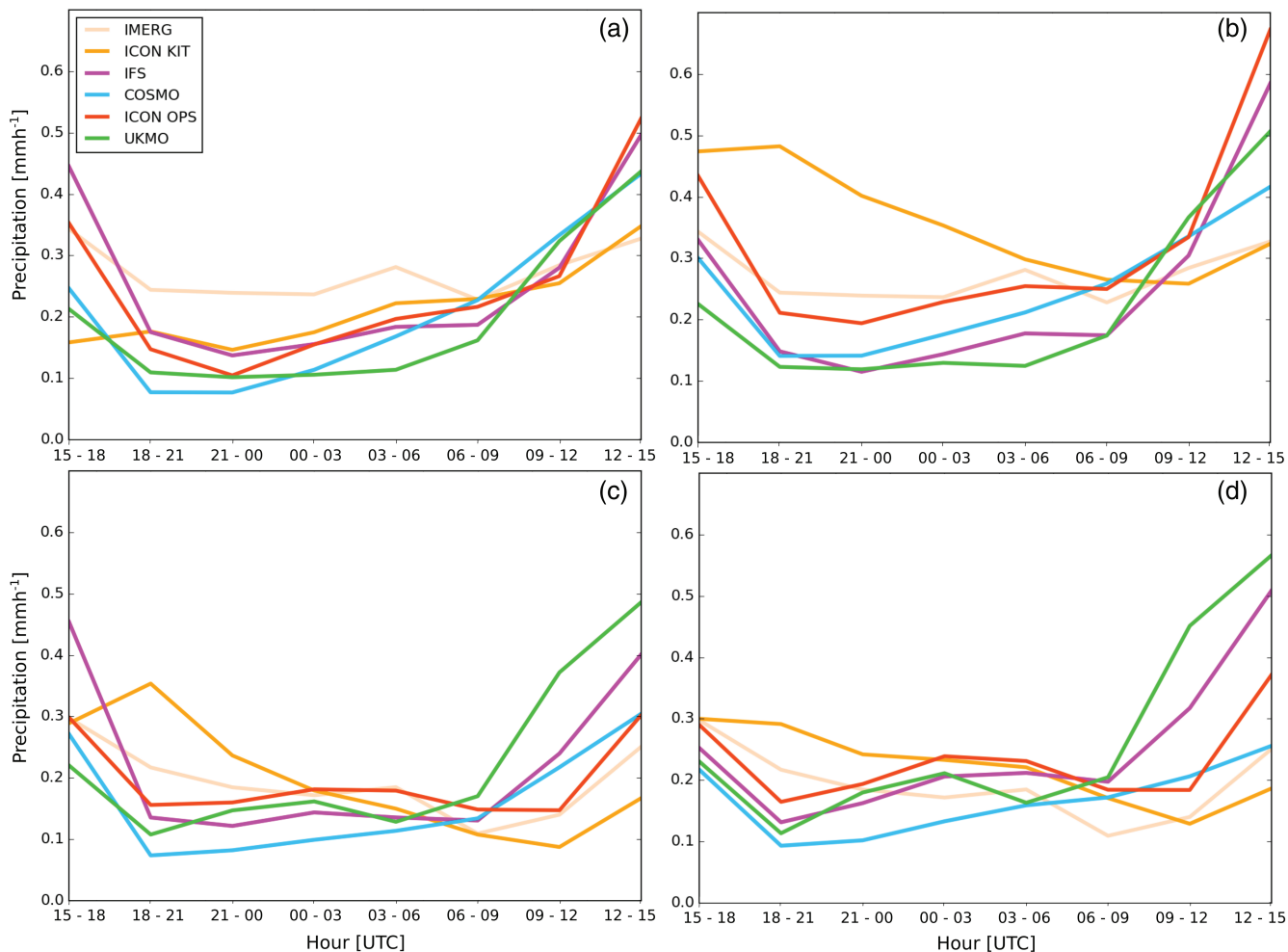
FIGURE 5 Average diurnal cycle of rainfall for IMERG and the five forecast models: (a) period 1 and forecast day 1, (b) period 1 and forecast day 2, (c) period 2 and forecast day 1, and (d) period 2 and forecast day 2

than IMERG. All models with convective parametrization strongly overemphasise the 1200–1500 UTC peak, irrespective of forecast day and period, with a tendency for largest overestimation on forecast day 2. This behaviour is most pronounced for UKMO during period 2. This suggests a too strong emphasis on afternoon convection. There is no clear indication that ICON KIT, the only model with explicit convection, outperforms the other models with respect to the diurnal cycle, as was shown by Marsham *et al.* (2013) and Pante and Knippertz (2019) for the Sahel. This model shows a later start of the evening rainfall, likely caused by delayed triggering of convection. Higher temporal resolution would be needed to make a clearer statement possible. For period 1, surprisingly large differences between forecast days 1 and 2 are found for ICON KIT (Figure 5a,b), as already found for spatial variability in Figure 4. Particularly night-time rains become much too strong on forecast day 2. Period 2 shows less differences between the forecast days overall (Figure 5c,d), but now an overestimation of night-time rainfall on forecast day 1.

3.1.2 | Low-level cloud cover

Figure 6 shows an analysis similar to Figure 2 but for 3-hourly *lclc*. The number of available stations (only those with more than 50% data coverage were retained) is much lower than for rainfall. *lclc* values were taken from the estimated coverage of low cloud in octas. The additional OCA data are created as described in Section 2.2. Differences between model or satellite data and the stations were produced with the same varying box strategy as applied for *rr*.

Station observations show widespread low-level cloudiness with values on the order of 50%. Thickest cloud decks are seen inland with decreases towards the coast and the Sahel, in agreement with van der Linden *et al.* (2015). OCA clearly struggles to capture this behaviour with an area average of 15%, about 32% less than the stations. As indicated by the stippling in Figure 6b, this is likely due to the high amount of obscuring higher clouds despite OCA's general ability to detect more than one cloud layer. No low-cloud cover will be recorded in cases where there
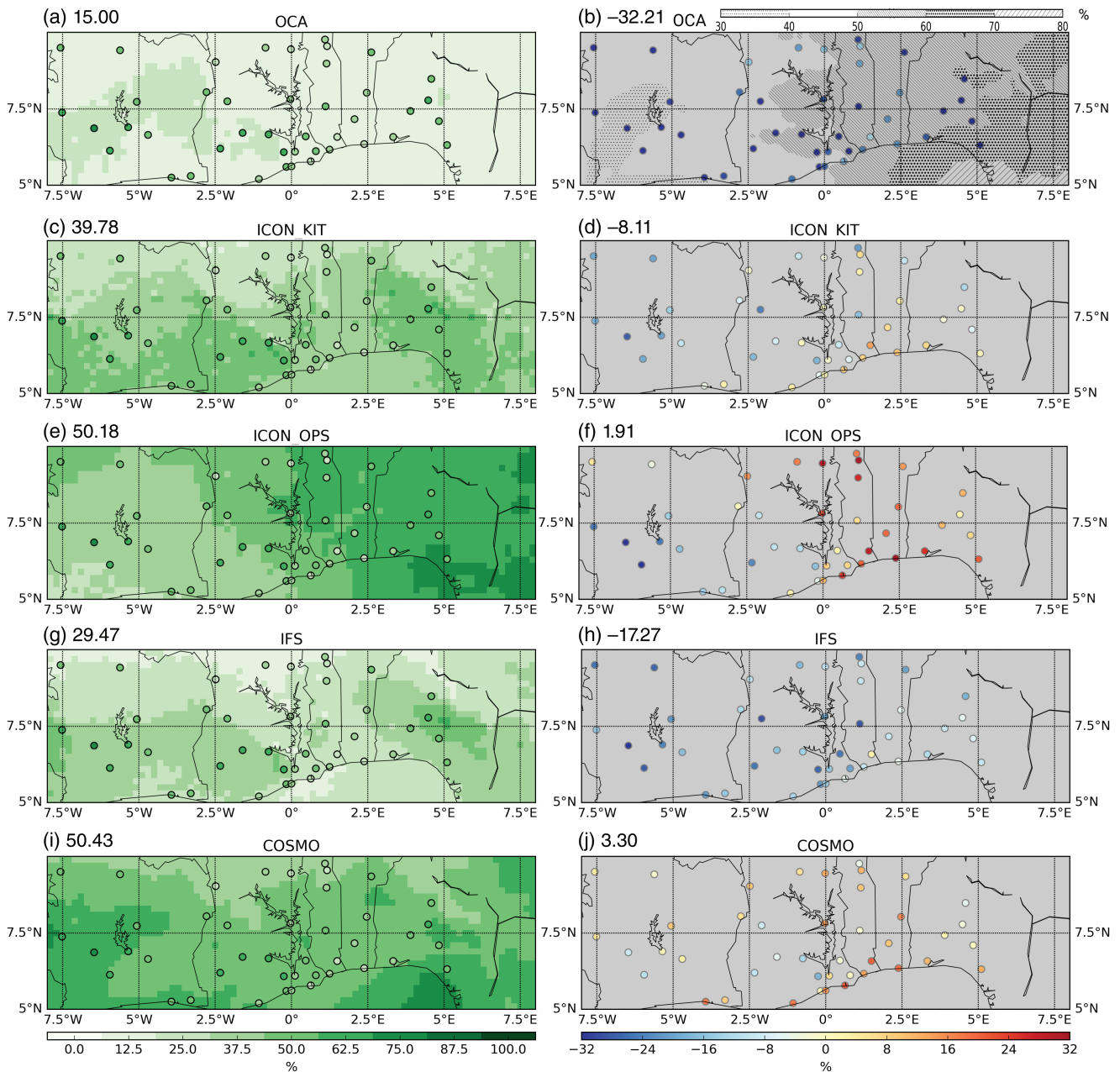
FIGURE 6 Evaluation of low cloud biases for June–July 2016. Left columns show horizontal distributions over the DACCIWA region of average daily *lclc* for OCA satellite estimates and four forecast datasets. Station observations are shown as filled circles using the same colour shading and area averages are provided as numbers at the top left. The right column shows the corresponding biases relative to the station observations with average biases printed at the top left. The latter are computed for 64 boxes as explained in Section 2.2. (b) includes a stippling to denote the percentage of high clouds without a successful low-cloud retrieval underneath

is substantial cover of high clouds or vertically extensive cloud cover. This general issue was also discussed extensively in van der Linden *et al.* (2015). Note that no distinction is made between daytime and night-time errors or single- and double-layer retrieval errors as in the original OCA retrieval, since period averages are considered.

With respect to the spatial distribution of the low clouds, IFS and ICON KIT show fairly realistic patterns, although both underestimate the cloud cover (by 17 and

8%, respectively), with ICON KIT struggling somewhat with the penetration of the low clouds inland and towards the west (Figure 6c,d,g,h). In terms of overall bias, the best results are generated by ICON OPS and COSMO with slight overestimations relative to the stations of 2–3%. COSMO shows an overestimation over coastal stations and relatively large cloud cover in the precipitation maximum over the ocean, while the cloud minimum in the Lake Volta region is not well reproduced (Figure 6i,j). ICON OPS

shows a surprisingly large west–east gradient in cloudiness in contrast to observations, and a clear overestimation of clouds along the coast (Figure 6e,f). The problems with coastal stations are likely related to an inability to resolve land–sea breeze type circulations that can lead to clearing in the afternoon (e.g. Guedje *et al.*, 2019). Unfortunately *lclc* was not available for UKMO.

3.1.3 | Radiation

As shown by Knippertz *et al.* (2011) and Hannak *et al.* (2017), differences in low-cloud cover can be expected to significantly impact on surface solar irradiance. Figure 7 shows a comparison of the four forecast models that provide this parameter with all available station observations. As evident from Figure 7, the spatial distribution of the observations is very heterogeneous with most radiometers located in Ghana and Benin, one in Ivory Coast and one in Nigeria. They show a clear indication of the shading by low clouds reducing *SSI* to values around $150\,\mathrm{W}\cdot$ m^{-2}. ICON KIT shows the overall best agreement with these station observations (Figure 7a,b). The overall pattern shows a clear imprint of the low-level cloud distribution (Figure 6a,b) with an area average of $166.5\,\mathrm{W}\cdot\mathrm{m}^{-2}$, which is only slightly higher than the station observations. IFS (Figure 7e,f) shows a similar pattern and again a clear correspondence to its low-cloud fields, but the area average ($201\,\mathrm{W}\cdot\mathrm{m}^{-2}$) and bias ($51\,\mathrm{W}\cdot\mathrm{m}^{-2}$) are much higher, pointing to too low coverage, possibly combined with too low optical thickness of the clouds. As already discussed for clouds, ICON OPS (Figure 7c,d) tends to have an unrealistic east–west gradient. The relatively large radiation bias of $43\,\mathrm{W}\cdot\mathrm{m}^{-2}$ in combination with little low-cloud bias points to problems with clouds at other levels or with cloud optical thickness in ICON OPS. Finally, UKMO (for which no no cloud information is available) shows the largest overestimation with $213.7\,\mathrm{W}\cdot\mathrm{m}^{-2}$ in the area average and very little structure across the region (Figure 7g,h).

3.1.4 | Latitudinal transects

A key ingredient of the WAM is the development of a near-surface gradient in temperature between the Atlantic cold tongue and the Saharan heat low. Here we evaluate to what extent the forecast models investigated are capable of reproducing this feature over SWA. We concentrate on the area from 1°E to 4°E, as the station density is large here. Mean values of $T_{2\mathrm{m}}$ and $Td_{2\mathrm{m}}$ from individual stations along this transect are compared to corresponding zonal averages from the models. As we expect the monsoon onset to have a marked influence on the north–south

distribution of temperature and moisture, the pre- and post-onset periods are displayed separately. In addition, forecast days 1 and 2 are shown to detect possible drifts with increasing lead-time.

For period 1, the coastal upwelling is not yet established (discussion in Knippertz *et al.*, 2017), leading to overall lower temperatures over land than over the ocean (Figure 8a,b). This agrees with Guedje *et al.* (2019), who investigated measurements from a buoy off the Beninese coast. Surprisingly, there is a difference in temperatures over sea of more than 0.6 C between the models, with UKMO being the warmest and IFS and ICON OPS the coldest. Taking into account that ICON KIT and IFS as well as ICON OPS and COSMO share the same intput data, it could be expected that those models are close to each other over sea, but in fact, they are differently grouped. This indicates that the $T_{2\mathrm{m}}$ differences arise from in-model treatment of surface exchange coefficients and the boundary layer development which are listed in Table 2. Some of these differences appear to persist inland through the dominating onshore flow during this season. The agreement with the stations is largely satisfactory on forecast day 1. Cotonou (Co) is right on the coast and often sunny during the day once the land–sea breeze has passed, which is not fully captured by the models. Atakpame (At) is elevated and therefore cooler than neighbouring stations. The reason for the warm conditions in Kara (Ka) in Togo are not clear and they differ quite markedly from nearby Niamtougou (Ni). On forecast day 2, the model behaviour near the coast changes little but the spread between the models increases markedly inland, reaching about 2 C with ICON KIT being the coolest and UKMO the warmest model. These differences are in close agreement with the differences in low-level cloudiness and *SSI* shown in Figures 6 and 7. The best match on forecast day 2 is found for ICON OPS

The situation changes markedly in period 2 (Figure 8c,d). The coastal upwelling is now established and reduces the temperature through the entire transect. While agreement inland is reasonable on forecast day 1, $T_{2\mathrm{m}}$ at the stations nearer the coast are much too low in the models. On forecast day 2, conditions at the coast remain largely unchanged but COSMO and ICON KIT develop a cold bias inland.

Figure 9 shows the corresponding analysis for $Td_{2\mathrm{m}}$. Both models and observations show an increasing drying from the ocean inland for all periods and forecast days. As for $T_{2\mathrm{m}}$, the overall agreement is better during period 1 than period 2, but with a larger spread than for temperature. For period 1 (Figure 9a,b), there is a mild increase in model spread from forecast day 1 to 2 with ICON KIT developing a dry bias and COSMO showing a stable wet bias, which is surprising giving the little precipitation
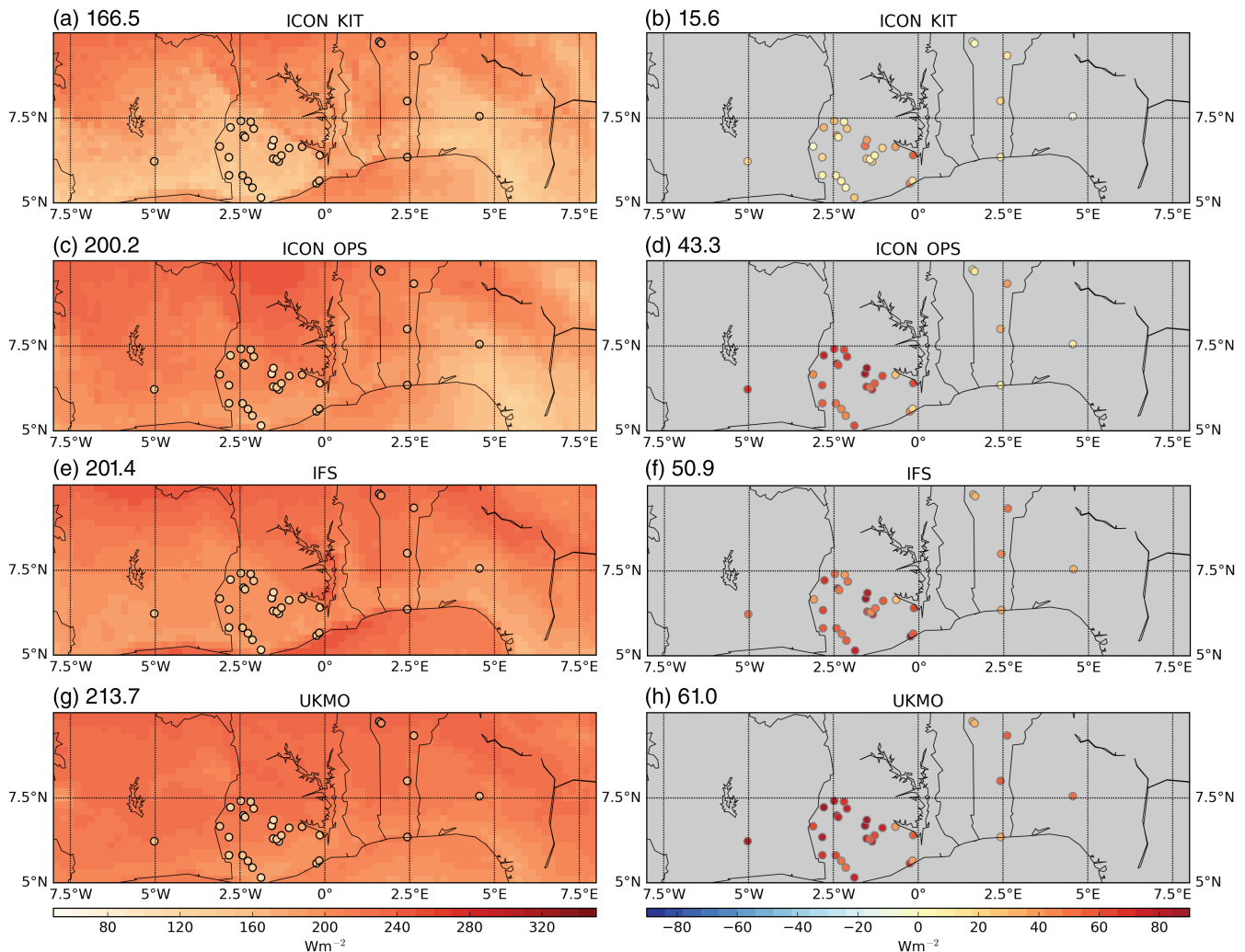
FIGURE 7 Evaluation of solar surface irradiance for June–July 2016. Left columns show horizontal distributions over the DACCIWA region for four forecast datasets. Station observations are shown as filled circles using the same colour shading and area averages are provided at the top left. The right column shows the corresponding biases relative to the station observations with averages at the top left. The latter are computed for 64 boxes as explained in Section 2.2

simulated by this model (Figure 2). The latter may indicate that triggering convection is too difficult in the model, possibly due to the high amount of low cloud (Figure 6) preventing surface heating (Figure 8). Such a behaviour is consistent with the findings of Söhne *et al.* (2008). In period 2 (Figure 9c,d), Td_{2m}s are markedly too low in the coastal areas in all models, where temperatures are also underestimated (Figure 8c,d). ICON OPS and IFS show a stable dry bias inland. Given the equally similar behaviour in temperature (Figure 8c,d) of these models, it is surprising to see such large differences in low-level cloudiness (Figure 6c,e). Difference between forecast days 1 and 2 are rather small, apart maybe from a marked drying in ICON OPS during period 1.

What creates the large coastal biases during period 2? An inspection of the diurnal cycle in T_{2m} reveals good agreement at night-time (not shown), but large differences

at 1200 UTC (Figure 10a) and lasting until 1800 UTC. This signal appears to be related to an overestimation of low cloud in the models at 0600 and 1200 UTC (Figure 10c), implying too weak surface heating during the morning hours. The models likely have difficulties to represent the so-called Maritime Inflow, a quasi-stationary front that moves inland during the night (Adler *et al.*, 2017; Deetz *et al.*, 2018) with clearing often observed on the seaward side of the front. Td_{2m} also show largest errors near the coast, revealing an overall underestimation in the models (Figure 10b). The reasons for this are not straightforward to understand and call for further study. Possible explanations include too much downward mixing of dry free-tropospheric air over the ocean, which is then advected onshore, or a moistening through showers or drizzle being triggered at the Maritime Inflow front.
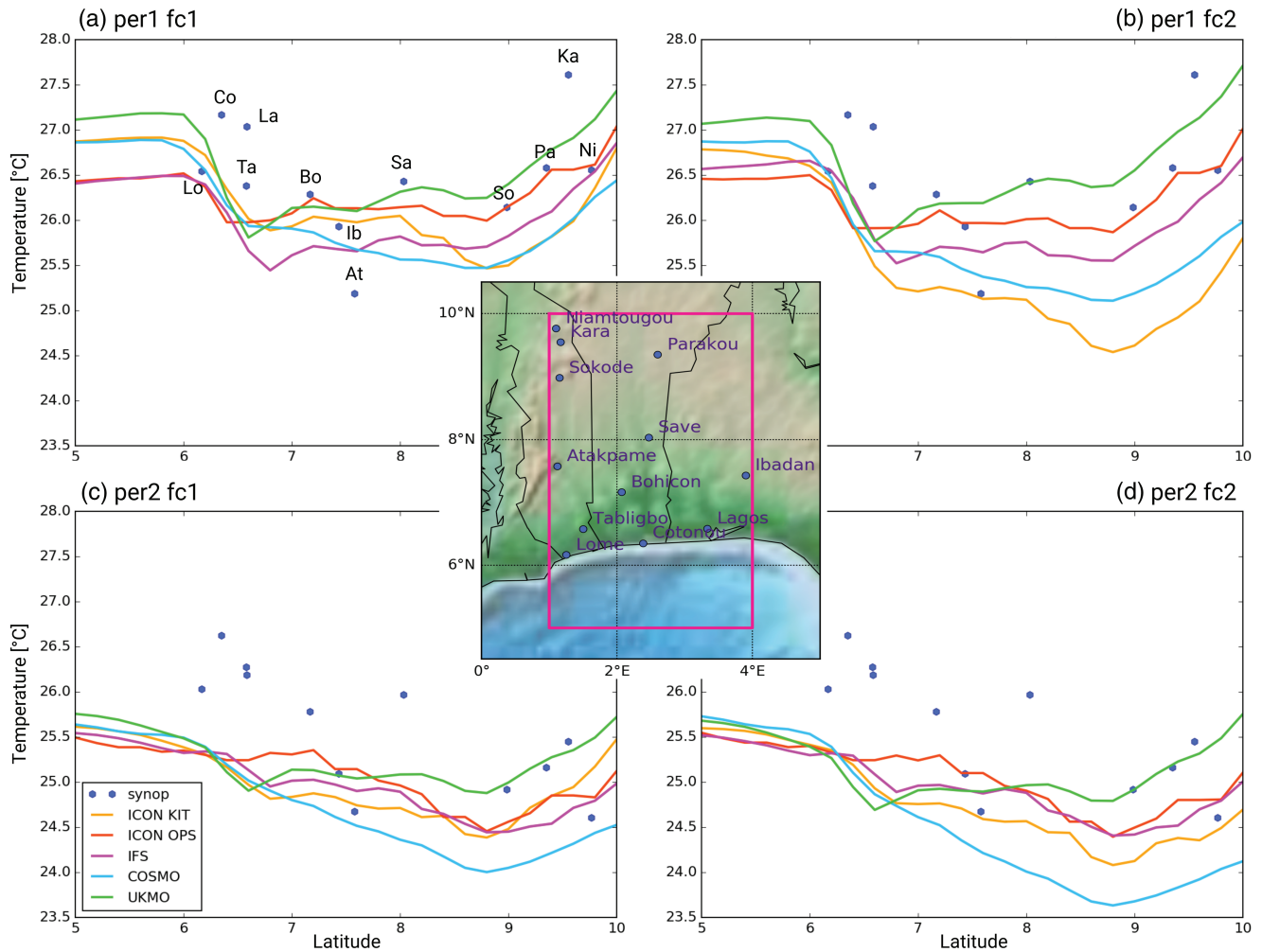
FIGURE 8 Evaluation of the near-surface meridional temperature gradient. The chosen area stretches from central Togo to western Nigeria and is characterised by a particularly high density of stations (inset map and Figure 1), for which individual values of T_{2m} are displayed as points (labelled as "synop"). Abbreviations of the station names are given in (a). For the models, displayed as coloured lines, zonal averages from 1°E to 4E, i.e., across the box shown in the inset map, are computed. Periods 1 and 2 as well as forecast days 1 and 2 are distinguished as in Figure 3

3.2 | Biases in vertical profiles

This section is dedicated to expand the evaluation into the troposphere up to 300 hPa. The rich dataset of radiosonde measurements from Abidjan (abbreviated abi hereafter), Accra (acc), Kumasi (kum), Cotonou (cot), Savè (sav) and Parakou (par) as indicated in Figure 1 are compared to the corresponding model output. To illustrate the details of this comparison, Figure 11 shows campaign averages and their standard deviations of five meteorological variables from the Kumasi station.

For q (Figure 11a), the day-to-day variation in the observations is relatively small compared to the natural decrease with height. However, there is enhanced spread around 400 hPa (compared to an already low q), which is between the height of mid- and high-level clouds (see rh in Figure 11c). This signal is also found in Abidjan, Accra

and Cotonou but only weakly in the more northern station Savè and not at all in Parakou (not shown). The physical reason for this marked increase in spread is not clear. All models tend to underestimate moisture throughout most of the vertical column. Moreover, the models have a markedly smaller standard deviation. The latter may be the result of a point-to-gridbox comparison, but could also be related to the inability of models with parametrized convection to generate larger organised systems. This is consistent with the fact that ICON KIT, the only model with explicit convection, has the largest spread at midlevels. With respect to temperature (Figure 11b), agreement between observations and models is good with very small standard deviations in all datasets. Consequently, the profiles in relative humidity (Figure 11c) mostly show a reflection of the q signals. The models are mostly too dry up to about 500 hPa with largest underestimations in the
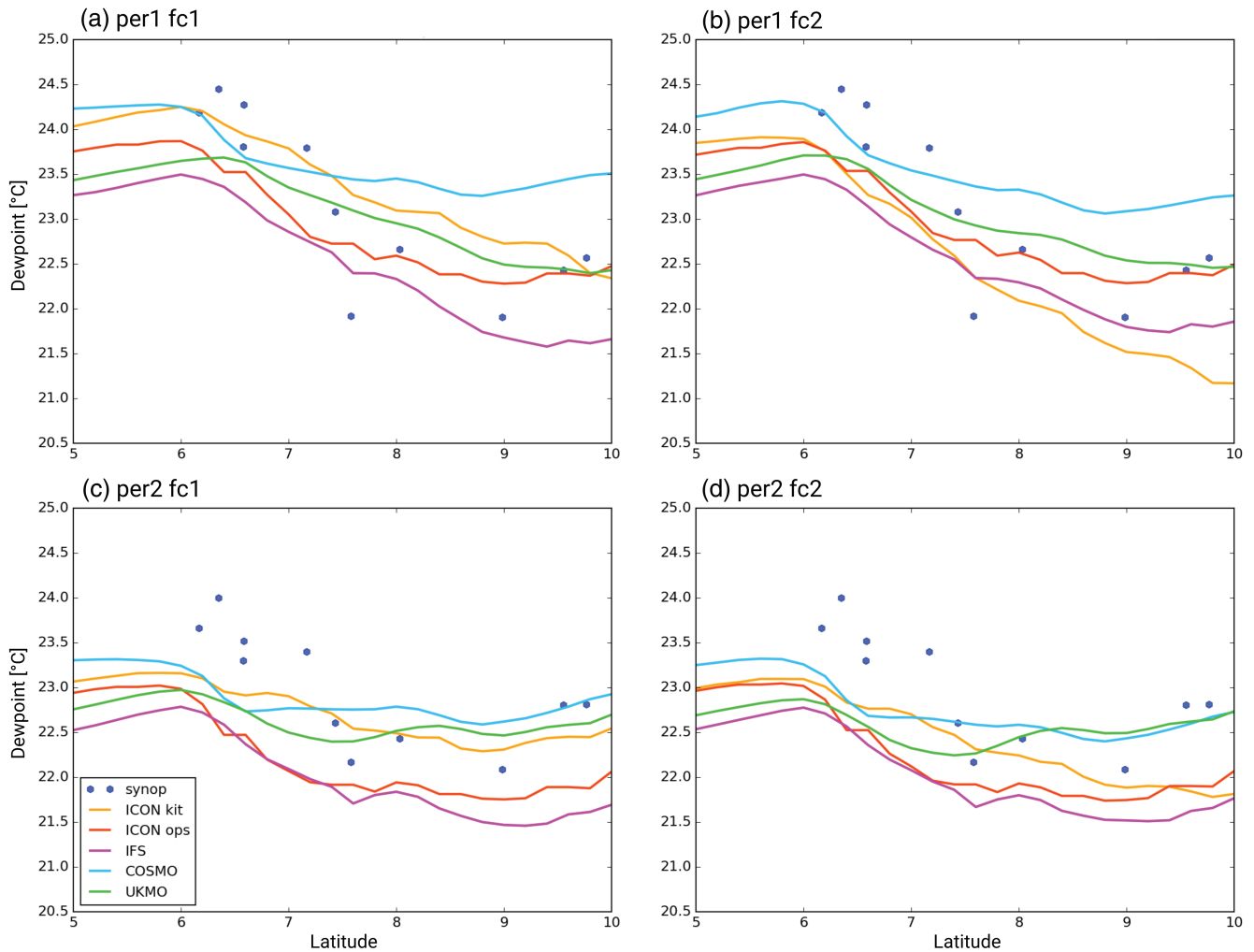
FIGURE 9 As Figure 8, but for Td_{2m}

mid-levels followed by the low-level *rh* (and thus cloud) maximum. In contrast, models tend to be too moist in the *rh* minimum around 400 hPa. This suggests that the models tend to mix moisture too evenly through the vertical column. As for *q*, the standard deviation is large in the radiosonde data almost everywhere.

Finally, Figure 11d,e show *v* and *vdir*. Both mean values and standard deviations are largely well captured by the models, although COSMO tends to overestimate the low-level jet and the mid-level easterlies at the southern side of the African Easterly Jet. There is a general tendency of all models to show more easterly flow at this level and to underestimate the southerly component in the observations. It is conceivable that this is a consequence of convective momentum transport from the monsoon layer, which may well be underrepresented in the models. The level of agreement seen here is much closer than was documented for state-of-the-art climate models by Hannak *et al.* (2017).

To expand the analysis to the other five radiosonde stations, an integrative measure for the station–model deviation is created from the individual profiles of each station and model. For each height level, the difference and the absolute difference between model and station are calculated and then averaged, weighted with the respective layer thickness. Figure 12 shows these values for *q*, *T*, *rh* and *v* separated into forecast days 1 and 2.

As already discussed for Kumasi, models tend to be too dry on forecast day 1, particularly ICON KIT and UKMO (Figure 12a). This tendency appears to increase from coastal Abidjan and Accra to the northern station of Parakou. Interestingly, this bias reduces a little on forecast day 2 and even becomes positive for Abidjan in ICON OPS, IFS and COSMO. Biases for temperature (Figure 12b) are rather heterogeneous and hardly change from forecast day 1 to day 2, indicating the very fast response of the models to potential model or initial condition errors, likely through changes in surface fluxes, clouds, convection and radiation. Savè and Accra stand out as stations with warm biases, while Kumasi shows a cold bias. ICON OPS and COSMO are the warmest models with UKMO the coldest.
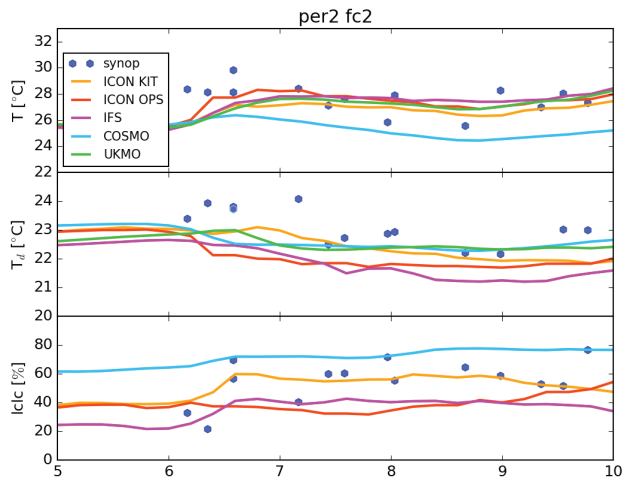
FIGURE 10 Meridional transects as in Figure 8, showing T, Td and *lclc* at 1200 UTC for the post-onset period and forecast day 2. Due to the restriction to 1200 UTC, the station selection is similar but not identical to Figure 8

The consequences of these biases on *rh* are displayed in Figure 12c. Most striking are the consistent dry bias in northernmost Parakou and the moist bias in coastal Abidjan. Particularly the latter increases on forecast day 2 and is then also visible in Accra. In Section 3.1 it was shown that the models tend to be too cold and dry near the surface at the coast as a reflection of too cloudy conditions there. The latter may be partly related to the column moist bias we see at Abidjan and Accra.

An inspection of absolute forecast errors (not shown) shows an increase from forecast day 1 to forecast day 2 for all variables, particularly for wind speed. This effect causes problems for the models further inland, because the properties of the atmosphere at the coast are not transported correctly with the mean wind to the north, which disturbs the diurnal development of the boundary layer in the northern half of the study region. The models differ among each other, but not always in a systematical way.

3.3 | Evaluation of day-to-day forecast

In this section, the ability of the five models to forecast local and regional day-to-day variations is studied with the help of traditional statistical measures such as centred root mean squared error (CRMSE), correlation and standard deviation. Data for *rr*, *lclc* and *T* are compiled as model–observation pairs of single measurements and then averaged over the DACCIWA domain. Results are summarised in the form of Taylor diagrams. This analysis is not done for radiation, as the different time resolution and grid box sizes make it difficult to generate a fair comparison

between the point measurements, which can be strongly affected by single clouds, and the grid-cell values.

3.3.1 | Precipitation

The evaluation of precipitation is based on the 155 stations introduced in Section 2.2. Here, the forecasts from 0600 UTC to 0600 UTC of the following day were summed up as daily *rr* values in order to emulate the station measurements as accurately as possible. In this way, only full days could be taken into account and no comparison between forecast days 1 and 2 is done here.

With respect to the regional picture, Figure 13 shows averaged *rr* time series of the models, IMERG satellite estimates and station observations for periods 1 and 2. The pre-onset phase is characterised by more regular precipitation events than the post-onset-phase. However, the average amounts differ only by 0.22 mm · day^{-1} (from 5.65 to 5.43 mm · day^{-1}) in the station observations and by 0.59 mm · day^{-1} (from 5.49 to 4.90 mm · day^{-1}) in IMERG. This discrepancy is partly related to the fact that the northern part of Benin, which is more often affected by the main rain band even after monsoon onset, is over-represented in the station observations (compare Figure 2). In addition, a substantial fraction of the rainfall during period 2 was related to the unusually wet conditions during the end of July 2016 (Knippertz *et al.*, 2017). During period 1 (Figure 13a), the time-series of IMERG and the station observations follow each other closely, but there are sometimes differences in the height of peaks and also some temporal shifts. The overall agreement between models and observations is rather low, even at this relatively large regional scale. This is reflected in the overall low correlations shown in Table 4, where IMERG displays a correlation coefficient of 0.95 while the models reach only 0.62 at best. However, some individual peaks are captured well by some models. The three clearest examples are those around 9–10, 12 and 15 June 2016. The DACCIWA campaign overview paper by Knippertz *et al.* (2017) shows that these precipitation maxima are connected to synoptic-scale vortices that cross the DACCIWA region from east to west (labelled A, B and C in that paper). It appears that these structures are coherent enough to be represented in forecast models, this way generating some skill on the regional scale, although some models clearly underestimate the magnitude of the rainfall enhancement by the disturbances. Consistent with this idea, the first rainfall peak, which is only captured by ICON KIT, is not associated with a coherent feature in Knippertz *et al.* (2017).

In period 2, the main rainband shifts inland and most of the precipitation modulation of coherent disturbances
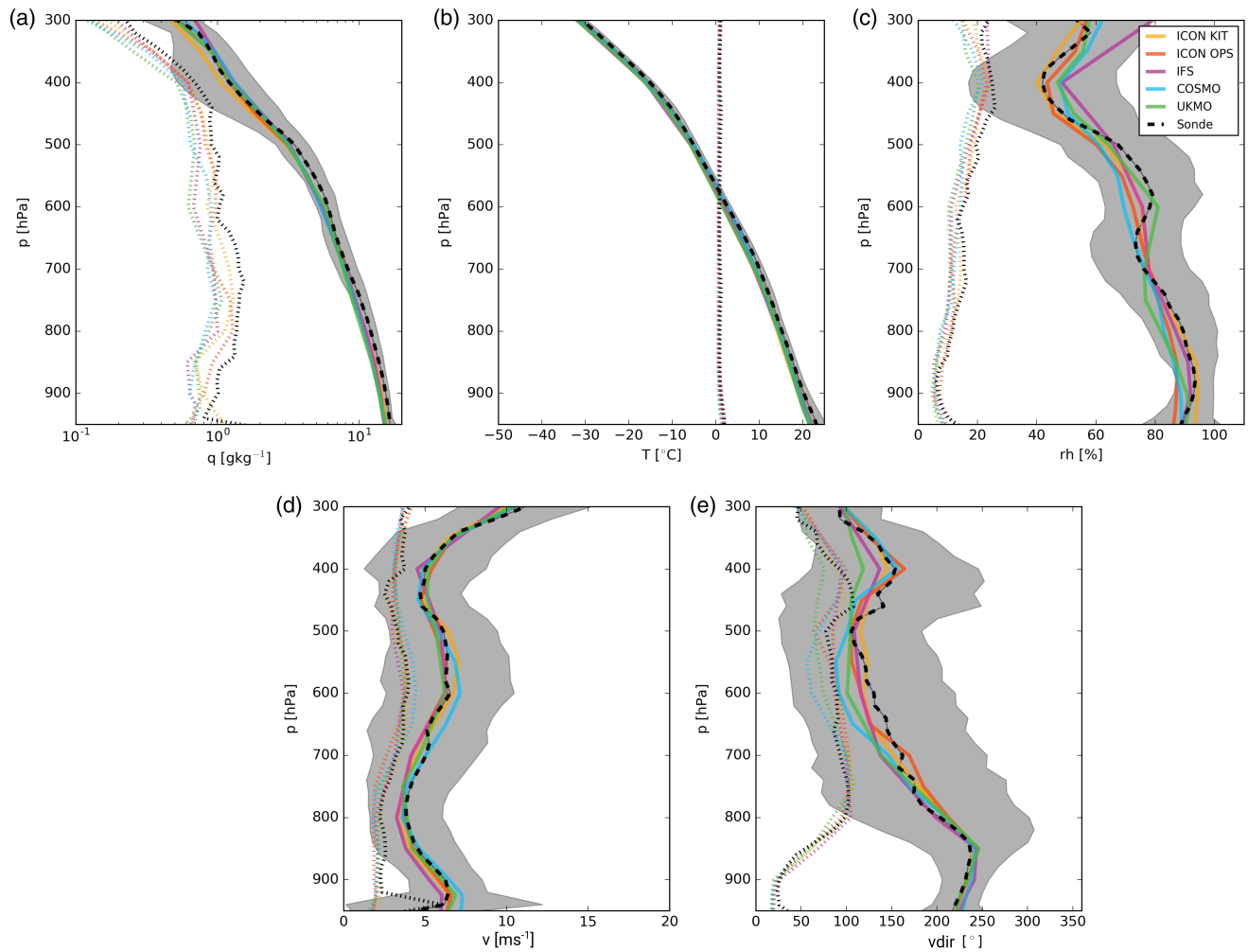
FIGURE 11 Evaluation of the vertical profiles of (a) *q*, (b) *T*, (c) *rh*, (d) *v* and (e) *vdir* at Kumasi (Ghana). The thick dashed lines show mean profiles from all radiosondes during the entire campaign. The model output was subsampled to match the radiosondes launch times. Profiles were created from the model columns containing the respective launching sites. The standard deviation is displayed as grey shading around the mean profiles and as dotted lines. The coloured lines show the corresponding means and standard deviations of the five NWP models using data from forecast day 1. Note the logarithmic axis in (a) showing specific humidity

TABLE 4 Correlations of models and IMERG with station observations based on daily sums of regional averages

Model	Period 1	Period 2
IMERG	0.95	0.93
ICON KIT	0.62	0.60
ICON OPS	0.46	0.49
IFS	0.56	0.49
COSMO	−0.03	0.16
UKMO	0.26	0.59

takes place to the north of about 8°N. Nevertheless some skill to forecast major rainfall peaks is evident, e.g., around 10, 16 and 24 July 2016 (Figure 13b). These three dates are

associated with disturbances H, I and J of Knippertz *et al.* (2017). As in period 1, the magnitude of rainfall modulation is mostly underestimated, while some overestimation takes place in the less active periods. The correlation coefficients for the period 2 in Table 4 are comparable to period 1.

An interesting question is now whether the models are also capable of reproducing some of the local variability reflected in a direct gridpoint to station comparison. To investigate this, Figure 14a,b show Taylor diagrams for the *rr* time series during periods 1 and 2, respectively, for all valid station–model and station–IMERG data pairs. For period 1 (Figure 14a), the comparison between stations and IMERG (star symbol) shows a Pearson correlation coefficient of only 0.52 and an underestimation of the standard deviation by the satellite product (compare to yellow half circle). While the latter was expected due to
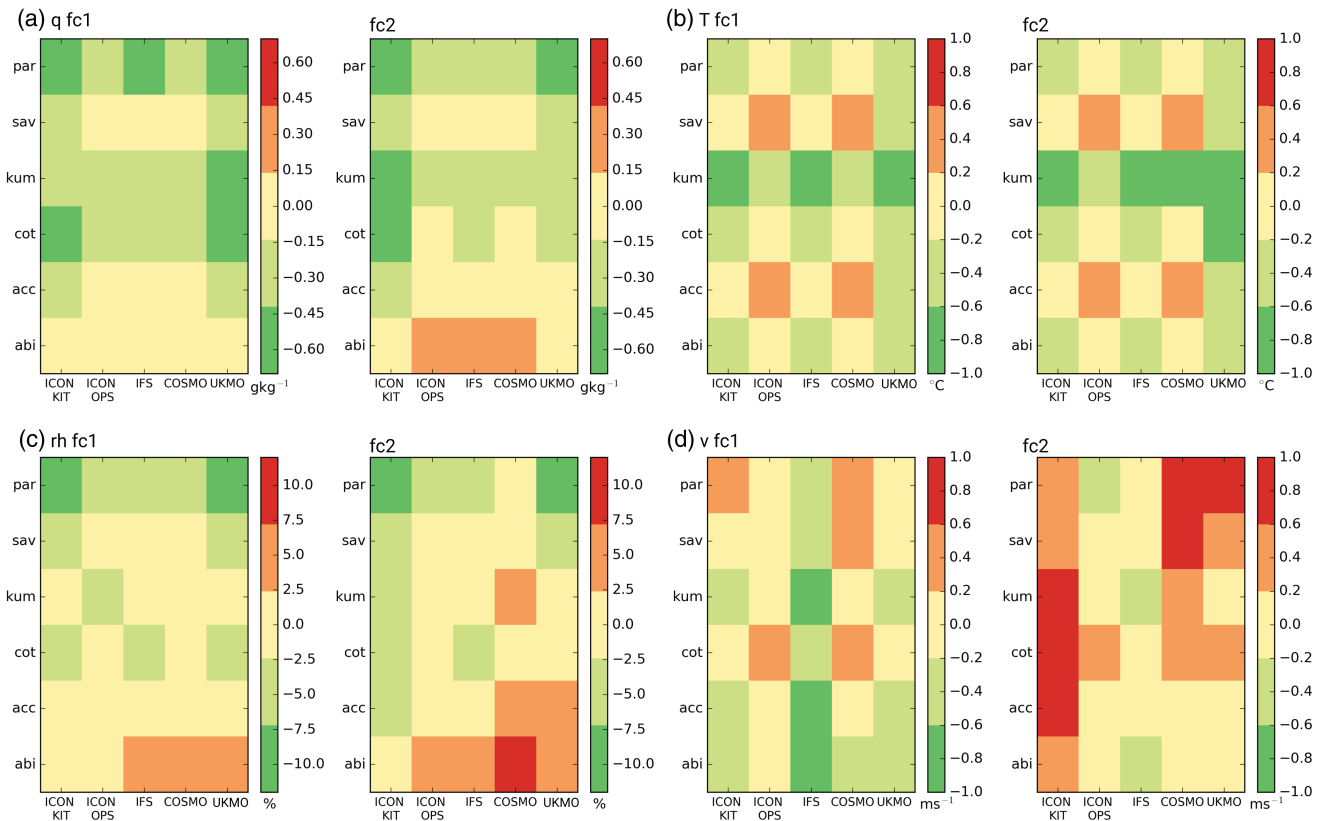
FIGURE 12 Evaluation of the vertical profile at all radiosonde stations. Coloured boxes show vertically integrated and campaign-averaged differences for Abidjan (abi), Accra (acc), Cotonou (cot), Kumasi (kum), Savè (sav) and Parakou (par) from the radiosonde–model pairs (compare Figure 11). Displayed variables are (a) q, (b) T, (c) rh and (d) v

the point–pixel comparison, the former demonstrates that the information fed into IMERG is not enough to clearly distinguish rainy from non-rainy pixels. This needs to be kept in mind for future evaluation, for which such a great density of stations will hardly be available.

Potential reasons for the unsatisfactory performance of IMERG include persistent cover with non-precipitating high clouds, frequent small-scale and short-lived warm rain showers and limited availability of ground stations for calibration (M. Maranan, personal communication, 2019). Comparison of IMERG validation in other regions of the world show that the correlation coefficients vary strongly with the analysed region. Guo *et al.* (2016) carried out a validation based on rain gauges for China. Correlations for one year of daily measurements regridded to $0.25° \times 0.25°$ vary from 0.45 to 0.94 for different subregions, which is attributed to complicated surface structures and poorly represented snowfall. Nevertheless, many studies found that the overall performance of IMERG exceeds that of the predecessor TMPA (TRMM Multisatellite Precipitation Analysis) (e.g., Guo *et al.*, 2016; Prakash *et al.*, 2016; Tan *et al.*, 2016; Asong *et al.*, 2017; Dezfuli *et al.*, 2017). The performance of the NWP models is rather sobering. All five consistently show low correlations of about 0.2, which is

likely due to the existence of synoptic-scale vorticies as discussed with respect to Figure 13a. ICON KIT is the only model able to reproduce the standard deviation observed at the stations, while all other models dramatically underestimate variability by about 50%. This demonstrates the large impact of explicit convection on small-scale variability, as also evident from Figure 2. However, due to the low correlation, the resulting CRMSE is larger for ICON KIT ($18.5\,\text{mm}\cdot\text{day}^{-1}$) than for the other models (13–$15\,\text{mm}\cdot\text{day}^{-1}$) (grey circles in Figure 14).

In period 2 (Figure 14b), the overall low correlations continue (including that with IMERG) but the underestimation of variability is even larger than in period 1, leading to overall similar CRMSE. ICON KIT's variability is now also an underestimate reducing CRMSE to $16.9\,\text{mm}\cdot\text{day}^{-1}$. This behaviour may be related to the frequent occurrence of isolated showers during the little dry season in boreal summer as described by Maranan *et al.* (2018).

3.3.2 | Low-level cloud cover

For a corresponding evaluation of forecasts of low-level cloudiness, SYNOP observations from 46 stations are
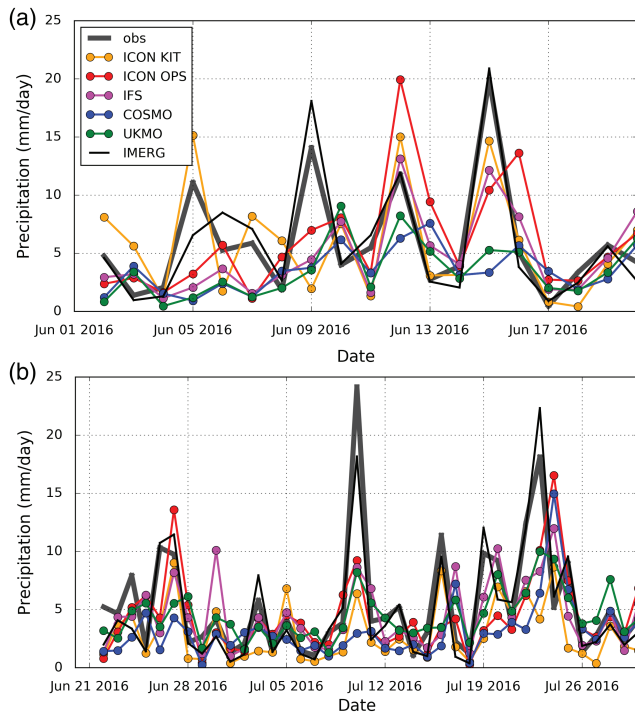
FIGURE 13 Timeseries daily *rr* spatially averaged over the DACCIWA region shown in Figure 1 for all models, IMERG satellite estimates and station observations. Only valid data pairs per location and time are included in this analysis. (a) Pre-onset period 1 and (b) post-onset period 2

used together with OCA satellite-based data (Section 2.2). Again, only valid station–model and station–satellite pairs are taken into account. Figure 14c–f show the respective Taylor diagrams split into periods 1 and 2 for the whole day and specifically for 0600 UTC, the time when the nocturnal stratus is usually well developed (van der Linden *et al.*, 2015). Forecast days 1 and 2 are depicted by different symbols in each panel. For the whole day during period 1 (Figure 14c), correlations are again low, ranging between about 0 and 0.4. Surprisingly, the OCA product shows very low correlation with the station observations and too low variability, underlining the challenge to detect low clouds from space. The models are more successful in reproducing the observed standard deviation. A marked exception is the very high variability in ICON OPS, the reason for which is not clear. Differences between forecast days 1 and 2 are typically rather small (and smaller than those between models) with mild deterioration for COSMO and ICON KIT and even a small improvement for ICON OPS. If one restricts the analysis to 0600 UTC (Figure 14e), all models but ICON OPS (and OCA) reproduce the day-to-day variability, but correlations are even lower. This is an indication that the subtle balances in the stable night-time PBL are even harder to represent in the models than those during the day. The very low correlation

with OCA indicates that low-cloud detection is very challenging in the morning hours, when the sun is just rising, casting some doubt on the usefulness of this dataset for the given purpose. As before, changes from forecast day 1 to 2 are neither large nor systematic.

The corresponding analysis for period 2 (Figure 14d,f) shows a slight improvement in correlations (0.5 for IFS) for the whole day but little change otherwise. The strikingly high variability and low correlations in ICON OPS remain. Correlations for OCA are slightly improved. These slightly more encouraging results may be related to the fact that in period 2 the stratus is more established (Knippertz *et al.*, 2017).

3.3.3 | 2m temperature

A corresponding analysis for T_{2m} within the transect discussed in Section 3.1 (Figure 14g,h) confirms some of the conclusions drawn for *rr* and *lclc*. For period 1 (Figure 14g) correlations range around 0.4 for all models with only a slight underestimation of day-to-day variability. A consistent deterioration from forecast day 1 to 2 can be seen for all five models. The CRMSE ranges around 1.5 K and thus on the order of the standard deviation in the surface observations. Overall, these results suggest a more consistent and successful representation of temperature variations as compared to *rr* or *lclc*. For period 2, correlations increase to values around 0.5, variability is still only slightly underestimated, but the deterioration from forecast day 1 to 2 is less clear. Overall this leads to a reduction of CRMSE to around 1.0 K, again of the same order as the standard deviation in the station observations.

4 | CONCLUSIONS

In June and July 2016 the comprehensive DACCIWA field campaign was conducted in SWA (Flamant *et al.*, 2018). Observations were taken from three research aircraft and three highly instrumented supersites, and radiosondes were launched regularly from stations in Ivory Coast, Ghana, Togo, Benin and Nigeria. Each day of the campaign, a range of forecast products from operational and research models was consulted to aid the planning of the field work. Here these model outputs were evaluated comprehensively using radiosondes, ground-based measurements and satellite data. Observations from a large number of stations were acquired through collaboration with West African weather services and other research projects, including digitisation of weather records on paper. Together this has created an unprecedented density of observations for this evaluation study. In the
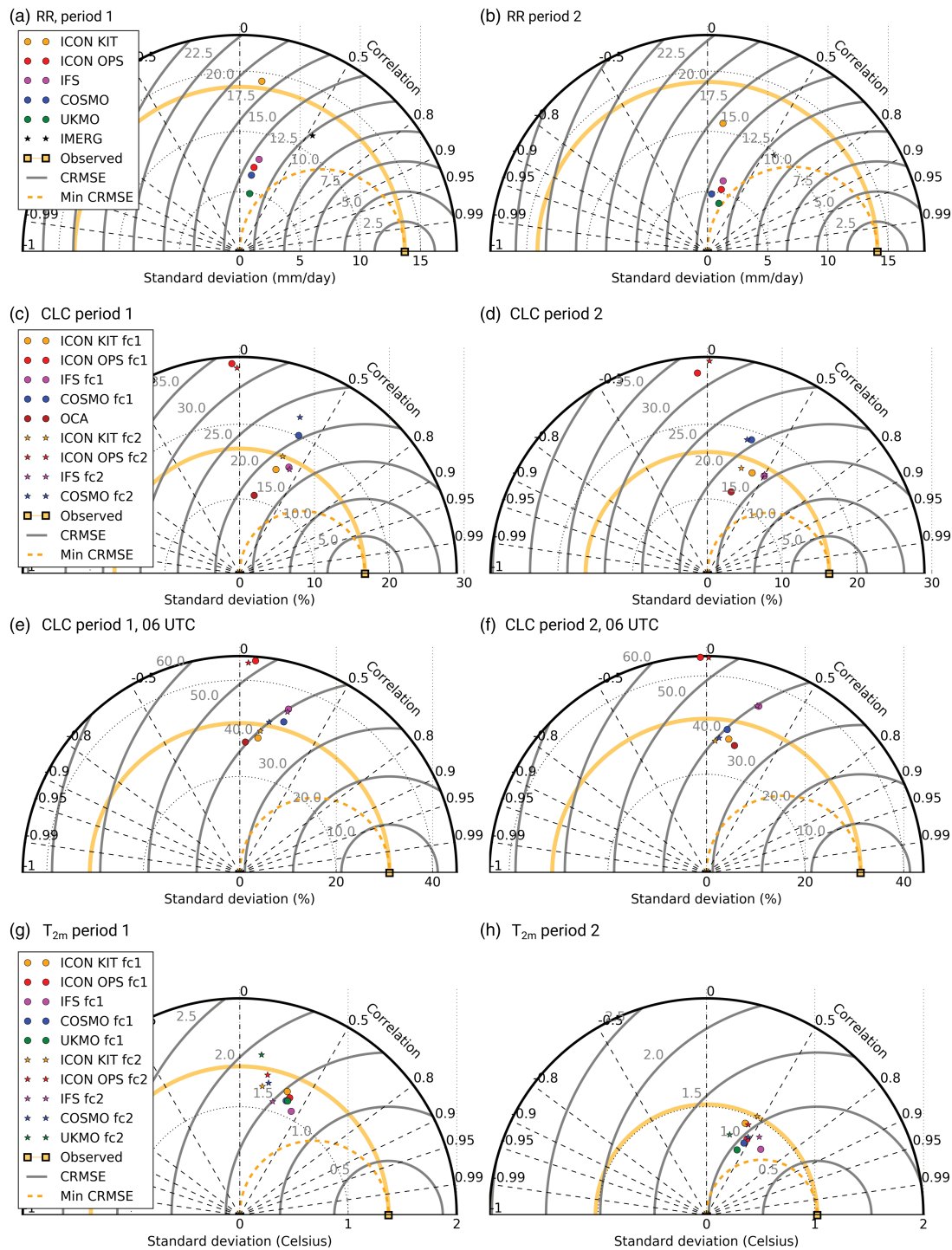
FIGURE 14 Taylor diagrams for (a, b) rain rate *rr*, (c–f) low-level cloud cover *lclc* and (g, h) 2 m temperature T_{2m} divided into (a, c, e, g) period 1 and (b, d, f, h) period 2. The plots are based on all valid station–model and station–satellite data pairs, respectively. (a, b) show daily *rr* (0600–0600 UTC) including a comparison to IMERG (stars). (c, d) show *lclc* for all SYNOP hours with dots denoting forecast day 1 and stars forecast day 2. As an additional observational reference, OCA low-level cloud data are shown. (e, f) show *lclc* as (c, d), but for 0600 UTC only, using the same legend. (g, h) show T_{2m} within the transect discussed in Section 3.1

analysis, an emphasis was put on aspects such as rainfall, low-level cloudiness, surface radiation, temperature and wind, which have been identified as crucial parameters in previous studies (e.g., Kniffka *et al.*, 2019). For

some investigations, the analysis was split into the pre- and post-onset phases as defined in Knippertz *et al.* (2017). Evaluation was done both on the regional level (5°N–10°N, 8°W–8°E) and for individual station–gridbox

pairs. Overall, the evaluation showed a significant level of diversity between the model behaviours that prohibits a clear ranking.

The main conclusions from this comprehensive evaluation are:

- There is a substantial level of observational uncertainty. A comparison of satellite-based estimates of rainfall and low-level cloud cover with station observations reveals underestimations of $-0.46\,\text{mm} \cdot \text{day}^{-1}$ (IMERG) and more than 30% (OCA), respectively. The latter is related to the frequent occurrence of obscuring high clouds (also van der Linden *et al.*, 2015) and different detection procedures (e.g., station observers may report a low cloud even in cases of cumulonimbus). Local day-to-day variations in low-cloud cover and rainfall from satellite estimates show low correlations to station observations. Overall, this illustrates the difficulties for a long-term evaluation, when only a much less dense network of stations is available.

- All five models underestimate rainfall by $0.1-1.9\,\text{mm} \cdot \text{day}^{-1}$. Rainfall biases improve slightly from the pre- to the post-onset period. Consistent with this, specific and relative humidity is mostly underestimated (apart from the immediate coast) when compared to radiosonde observations, while temperature biases are rather unsystematic when the entire tropospheric column is considered.

- Models tend to underestimate low-cloud cover (range +3 to −17%). Together with errors in higher clouds and in cloud optical thickness, this leads to a positive bias in solar radiation at the surface of $43\,\text{W} \cdot \text{m}^{-2}$ averaged over all models (range $16-61\,\text{W} \cdot \text{m}^{-2}$). These results deviate significantly from Söhne *et al.* (2008), who find enhanced clouds and reduced surface heating. As demonstrated in Kniffka *et al.* (2019), the enhanced surface radiation leads to an increase in vertical mixing and changes the daily PBL evolution. However, in contrast to the sensitivity experiments presented in Kniffka *et al.* (2019), the increased solar radiation does not enhance convective instability and thus rainfall in the NWP models.

- Despite these biases, most models show some skill in representing regional modulations of rainfall (timing and duration) related to synoptic-scale disturbances (Knippertz *et al.*, 2017), although the magnitude is underestimated. This agrees with findings by Louvet *et al.* (2016) and Söhne *et al.* (2008), the latter specifically relating to African easterly waves. Correlations with station observations at the local level are very low for both rainfall and low-level clouds (order 0.2), while for temperature correlations are a little higher with 0.4–0.5,

particularly during the post-onset period, and model spread is reduced. Interestingly, day-to-day variability in cloud cover and temperature is largely realistic in models, while that for rainfall tends to be underestimated. The low skill in local rainfall predictions is consistent with Vogel *et al.* (2018), who evaluated global ensemble prediction systems.

- Models appear to particularly struggle with representing the conditions along the Guinea Coast, which is affected by land–sea-breeze effects (Guedje *et al.*, 2019) and the Maritime Inflow phenomenon as described in Adler *et al.* (2017) and Deetz *et al.* (2018). In this area models tend to overestimate clouds, associated with too dry and too cold conditions near the surface, particularly after the monsoon onset, when the main rain band has moved northwards towards the Sahel and when coastal upwelling is well established. Consistent with this, there is a positive bias in *rh* at coastal stations compared to radiosondes.

- Differences between forecast days 1 and 2 are relatively small and hardly systematic, suggesting that forecasts respond quickly to errors in initialisation data and model physics and then change less as the forecast continues. Exceptions are notable degradations in the wind profile and in local temperature variations during the pre-onset period.

- The only model with explicit convection, ICON KIT, shows a more realistic spatial variability, particularly in rainfall, but otherwise forecasts are not significantly better. Particularly, the former aspect agrees with findings by Beucher *et al.* (2014) and Maurer *et al.* (2017). It has to be noted though that ICON KIT was initialised with IFS analyses, which is not optimised for ICON. Nevertheless, this demonstrates that convective parametrization is not the sole cause of model issues in this region with its very subtle couplings between the surface, the PBL and the free troposphere (e.g., Couvreux *et al.*, 2014; Kniffka *et al.*, 2019) and where convective organisation plays an overall smaller role than, for example, in the Sahel (Maranan *et al.*, 2018).

All in all, the investigated models showed a reasonable performance in capturing the large-scale dynamic state of the atmosphere over SWA but local variations are not well represented and some considerable biases remain. Future efforts to improve models in this part of the world should concentrate on a better representation of low-level clouds and their diurnal cycle, as these control surface solar radiation and thus the diurnal evolution of the PBL and ultimately rainfall. It would also be desirably to improve the representation of features determining conditions at

the very densely populated Guinea Coast. The skill related to synoptic-scale features shown here is promising and should be explored further, e.g. for other years and seasons. Another important outcome of this study is that further efforts are needed to improve the observational network over SWA and to make observations available to the international community, as satellite data alone are insufficient to represent all relevant features. Given the past, present and likely future dearth of observations in West Africa, the unprecedented validation dataset collected for June–July 2016 is a valuable and rich source of information in future forecast and climate model validations. It has therefore been made available for use through the DACCIWA database at http://baobab.sedoo.fr/DACCIWA (accessed 30 December 2019).

ORCID

Anke Kniffka https://orcid.org/0000-0002-9955-5262
Peter Knippertz https://orcid.org/0000-0001-9856-619X
Andreas H. Fink https://orcid.org/0000-0002-5840-2120
Marlon Maranan https://orcid.org/0000-0002-0324-8859

REFERENCES

Adler, B., Babić, K., Kalthoff, N., Lohou, F., Lothon, M., Dione, C., Pedruzo-Bagazgoitia, X. and Andersen, H. (2019) Nocturnal low-level clouds in the atmospheric boundary layer over southern West Africa: an observation-based analysis of conditions and processes. *Atmospheric Chemistry and Physics*, 19, 663–681.

Adler, B., Kalthoff, N. and Gantner, L. (2017) Nocturnal low-level clouds over southern West Africa analysed using high-resolution simulations. *Atmospheric Chemistry and Physics*, 17, 899–910.

Agustí-Panareda, A., Beljaars, A., Cardinali, C., Genkova, I. and Thorncroft, C. (2010) Impact of assimilating AMMA soundings on ECMWF analyses and forecasts. *Weather and Forecasting*, 25, 1142–1160.

Asong, Z.E., Razavi, S., Wheater, H.S. and Wong, J.S. (2017) Evaluation of integrated multisatellite retrievals for GPM (IMERG) over southern Canada against ground precipitation observations: a preliminary assessment. *Journal of Hydrometeorology*, 18, 1033–1050.

Athanasopoulou, E., Rieger, D., Walter, C., Vogel, H., Karali, A., Hatzaki, M., Gerasopoulos, E., Vogel, B., Giannakopoulos, C., Gratsea, M. and Roussos, A. (2014) Fire risk, atmospheric chemistry and radiative forcing assessment of wildfires in eastern Mediterranean. *Atmospheric Environment*, 95, 113–125.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, 139, 3887–3905.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M.J., Vitart, F. and Balsamo, G. (2008) Advances in simulating atmospheric variability with the ECMWF model: from synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 134, 1337–1351.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A. and Bormann, N. (2014) Representing equilibrium and non-equilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, 71, 734–753.

Beucher, F., Lafore, J.-P., Karbou, F. and Roca, R. (2014) High-resolution prediction of a major convective period over West Africa. *Quarterly Journal of the Royal Meteorological Society*, 140, 1409–1425.

Birch, C.E., Parker, D.J., Marsham, J.H., Copsey, D. and Garcia-Carreras, L. (2014) A seamless assessment of the role of convection in the water cycle of the West African monsoon. *Journal of Geophysical Research*, 119, 2890–2912.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D.H., Ebert, B., Fuentes, M., Hamill, T.M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Dias, P.S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L. and Worley, S. (2010) The THORPEX Interactive Grand Global Ensemble. *Bulletin of the American Meteorological Society*, 91, 1059–1072.

Brown, A.R., Milton, S.F., Cullen, M.J.P., Golding, B., Mitchell, J. and Shelly, A. (2012) Unified modeling and prediction of weather and climate: a 25-year journey. *Bulletin of the American Meteorological Society*, 93, 1865–1877.

Brown, A.R., Beare, R.J., Edwards, J.M., Lock, A.P., Keogh, S.J., Milton, S.F. and Walters, D.N. (2008) Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. *Boundary-Layer Meteorology*, 128, 117–132.

Clayton, A., Lorenc, A.C. and Barker, D.M. (2012) Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139, 1445–1461.

Couvreux, F., Guichard, F., Gounou, A., Bouniol, D., Peyrillé, P. and Köhler, M. (2014) Modelling of the thermodynamical diurnal cycle in the lower atmosphere: a joint evaluation of four contrasted regimes in the tropics over land. *Boundary-Layer Meteorology*, 150, 185–214.

Cullen, M.J.P. and Davies, T. (1991) A conservative split-explicit integration scheme with fourth-order horizontal advection. *Quarterly Journal of the Royal Meteorological Society*, 117, 993–1002.

Cusack, S., Edwards, J.M. and Crowther, J.M. (1999) Investigating k distribution methods for parameterizing gaseous absorption in the Hadley Centre climate model. *Journal of Geophysical Research*, 104, 2051–2057.

Deetz, K., Vogel, H., Knippertz, P., Adler, B., Taylor, J., Coe, H., Bower, K., Haslett, S., Flynn, M., Dorsey, J., Crawford, I., Kottmeier, C. and Vogel, B. (2018) Numerical simulations of aerosol radiative effects and their impact on clouds and atmospheric dynamics over southern West Africa. *Atmospheric Chemistry and Physics*, 18, 9767–9788.

Deroubaix, A., Menut, L., Flamant, C., Brito, J., Denjean, C., Dreiling, V., Fink, A., Jambert, C., Kalthoff, N., Knippertz, P., Ladkin, R., Mailler, S., Maranan, M., Pacifico, F., Piguet, B., Siour, G. and Turquety, S. (2019) Diurnal cycle of coastal anthropogenic pollutant transport over southern West Africa during the DACCIWA campaign. *Atmospheric Chemistry and Physics*, 19, 473–497.

Dezfuli, A.K., Ichoku, C.M., Huffman, G.J., Mohr, K.I., Selker, J.S., van de Giesen, N., Hochreutener, R. and Annor, F.O. (2017) Validation of IMERG precipitation in Africa. *Journal of Hydrometeorology*, 18, 2817–2825.

Dione, C., Lohou, F., Lothon, M., Adler, B., Babić, K., Kalthoff, N., Pedruzo-Bagazgoitia, X., Bezombes, Y. and Gabella, O. (2018) Low-level cloud and dynamical features within the southern West African monsoon. *Atmospheric Chemistry and Physics Discussions*, 19, 8979–8997. https://doi.org/10.5194/acp-19-8979-2019

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P. and Vogel, G. (2011) *A Description of the non-hydrostatic regional COSMO model, Part II : Physical parameterization*. DWD, Offenbach, Germany. Available at: http://www.cosmo-model.org/content/model/documentation/core/cosmoPhys Paramtr.pdf; accessed 30 December 2019.

Doms, G. and Schättler, U. (2002) *A description of the non-hydrostatic regional model LM. Part I: Dynamics and numerics*. DWD, Offenbach, Germany.

Donlon, C.J., Martin, M., Stark, J.D., Roberts-Jones, J., Fiedler, E. and Wimmer, W. (2011) The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, 116, 140–158.

Druyan, L.M., Feng, J., Cook, K.H., Xue, Y., Fulakeza, M., Hagos, S.M., Konaré, A., Moufama-Okia, W., Powell, D.P., Vizy, E.K. and Ibrah, S.S. (2010) The WAMME regional model intercomparison study. *Climate Dynamics*, 35, 175–192.

ECMWF (2016) *IFS Documentation. CY41R2, Parts III–VI*. ECMWF, Reading, UK. Available at: http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation; accessed 30 December 2019.

Edwards, J.M. and Slingo, A. (1996) Studies with a flexible new radiation code. I: choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, 122, 689–719.

Faccani, C., Rabier, F., Fourrié, N., Agustí-Panareda, A., Karbou, F., Moll, P., Lafore, J.P., Nuret, M., Hdidou, F.Z. and Bock, O. (2009) The impact of the AMMA radiosonde data on the French global assimilation and forecast system. *Weather and Forecasting*, 24, 1268–1286.

Fink, A.H., Agustí-Panareda, A., Parker, D.J., Lafore, J.-P., Ngamini, J.-B., Afiesimama, E., Beljaars, A., Bock, O., Christoph, M., Didé, F., Faccani, C., Fourrié, N., Karbou, F., Polcher, J., Mumba, Z., Nuret, M., Pohle, S., Rabier, F., Tompkins, A.M. and Wilson, G. (2011) Operational meteorology in West Africa: observational networks, weather analysis and forecasting. *Atmospheric Science Letters*, 12, 135–141.

Flamant, C., Knippertz, P., Fink, A.H., Akpo, A., Brooks, B., Chiu, C., Coe, H., Danuor, S., Evans, M., Jegede, O., Kalthoff, N., Konaré, A., Liousse, C., Lohou, F., Mari, C., Schlager, H., Schwarzenboeck, A., Adler, B., Amekudzi, L., Aryee, J., Ayoola, M., Batenburg, A.M., Bessardon, G., Borrmann, S., Brito, J., Bower, K., Burnet, F., Catoire, V., Colomb, A., Denjean, C., Fosu-Amankwah, K., Hill, P.G., Lee, J., Lothon, M., Maranan, M., Marsham, J.H., Meynadier, R., Ngamini, J.-B., Rosenberg, P., Sauer, D., Smith, V., Stratmann, G., Taylor, J.W., Voigt, C. and Yoboué, V. (2018) The dynamics-aerosol-chemistry-cloud interactions in the West Africa field campaign: overview and research highlights. *Bulletin of the American Meteorological Society*, 99, 83–104.

Forbes, R.M. and Ahlgrimm, M. (2014) On the representation of high-latitude boundary-layer mixed-phase cloud in the ECMWF global model. *Monthly Weather Review*, 142, 3425–3445.

Forbes, R.M., Tompkins, A.M. and Untch, A. (2011) *A new prognastic bulk microphysics scheme for the IFS*. Technical Memorandum 649, ECMWF, Reading, UK.

Galle, S., Grippa, M., Peugeot, C., Moussa, I.B., Cappelaere, B., Demarty, J., Mougin, E., Panthou, G., Adjomayi, P., Agbossou, E., Ba, A., Boucher, M., Cohard, J.-M., Descloitres, M., Descroix, L., Diawara, M., Dossou, M., Favreau, G., Gangneron, F., Gosset, M., Hector, B., Hiernaux, P., Issoufou, B.-A., Kergoat, L., Lawin, E., Lebel, T., Legchenko, A., Abdou, M.M., Malam-Issa, O., Mamadou, O., Nazoumou, Y., Pellarin, T., Quantin, G., Sambou, B., Seghieri, J., Séguis, L., Vandervaere, J.-P., Vischel, T., Vouillamoz, J.-M., Zannou, A., Afouda, S., Alhassane, A., Arjounin, M., Barral, H., Biron, R., Cazenave, F., Chaffard, V., Chazarin, J.-P., Guyard, H., Koné, A., Mainassara, I., Mamane, A., Oi, M., Ouani, T., Soumaguel, N., Wubda, M., Ago, E., Alle, I., Allies, A., Arpin-Pont, F., Awessou, B., Cassé, C., Charvet, G., Dardel, C., Depeyre, A., Diallo, F., Do, T., Fatras, C., Frappart, F., Gal, L., Gascon, T., Gibon, F., Guiro, I., Ingatan, A., Kempf, J., Kotchoni, D., Lawson, F., Leauthaud, C., Louvet, S., Mason, E., Nguyen, C., Perrimond, B., Pierre, C., Richard, A., Robert, E., Román-Cascón, C., Velluet, C. and Wilcox, C. (2018) AMMA-CATCH, a critical zone observatory in West Africa monitoring a region in transition. *Vadose Zone Journal*, 17, 1–24. https://doi.org/10.2136/vzj2018.03.0062

Garcia-Carreras, L., Marsham, J.H., Parker, D.J., Bain, C.L., Milton, S.F., Saci, A., Salah-Ferroudj, M., Ouchene, B. and Washington, R. (2013) The impact of convective cold pool outflows on model biases in the Sahara. *Geophysical Research Letters*, 40, 1647–1652.

Gregory, D. and Rowntree, P.R. (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. *Monthly Weather Review*, 118, 1483–1506.

Guedje, F.K., Arnaud Houeto, V.V., Houngninou, E., Fink, A.H. and Knippertz, P. (2019) Climatology of coastal wind regimes in Benin. *Meteorologische Zeitschrift*, 28, 23–39.

Guo, H., Chen, S., Bao, A., Behrangi, A., Hong, Y., Ndayisaba, F., Hu, J. and Stepanian, P.M. (2016) Early assessment of integrated multi-satellite retrievals for global precipitation measurement over China. *Atmospheric Research*, 176-177, 121–133.

Haiden, T., Rodwell, M.J., Richardson, D.S., Okagaki, A., Robinson, T. and Hewson, T. (2012) Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Monthly Weather Review*, 140, 2720–2733.

Hannak, L., Knippertz, P., Fink, A.H., Kniffka, A. and Pante, G. (2017) Why do global climate models struggle to represent low-level clouds in the West African summer monsoon?. *Journal of Climate*, 30, 1665–1687.

Heise, E., Ritter, B. and Schrodin, R. (2006). *Operational implementation of the multilayer soil model TERRA*. COSMO Technical Report 9. Available at: http://www.cosmo-model.org/content/model/documentation/techReports/ default.htm; accessed 30 December 2019.

Hill, P.G., Allan, R.P., Bodas-Salcedo., A. and Chiu, J.C. (2018) Quantifying the contribution of different cloud types to the radiation budget in southern West Africa during the monsoon season. *Journal of Climate*, 31, 5273–5291.

Hill, P.G., Allan, R.P., Chiu, J.C. and Stein, T.H.M. (2016) A multi-satellite climatology of clouds, radiation and precipitation in southern West Africa and comparison to climate models. *Journal of Geophysical Research*, 121, 10,857–10,879.

Huffman, G., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J. and Xie, P. (2015). NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD) version 4.5. Available at https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf; accessed 30 December 2019.

Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A. and Collins, W.D. (2008) Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *Journal of Geophysical Research*, 113, D13103

Kalthoff, N., Lohou, F., Brooks, B., Jegede, G., Adler, B., Babic, K., Dione, C., Ajao, A., Amekudzi, L.K., Aryee, J.N.A., Ayoola, M., Bessardon, G., Danuor, S.K., Handwerker, J., Kohler, M., Lothon, M., Pedruzo-Bagazgoitia, X., Smith, V., Sunmonu, L., Wieser, A., Fink, A.H. and Knippertz, P. (2018) An overview of the diurnal cycle of the atmospheric boundary layer during the West African monsoon season: results from the 2016 observational campaign. *Atmospheric Chemistry and Physics*, 18, 2913–2928.

Karbou, F., Rabier, F., Lafore, J., Redelsperger, J. and Bock, O. (2010) Global 4DVAR assimilation and forecast experiments using AMSU observations over land. Part I: Impacts of assimilating surface-sensitive channels on the African monsoon during AMMA. *Weather and Forecasting*, 25, 20–36.

Kniffka, A., Knippertz, P. and Fink, A.H. (2019) The role of low-level clouds in the West African monsoon system. *Atmospheric Chemistry and Physics*, 19, 1623–1647.

Knippertz, P., Coe, H., Chiu, J.C., Evans, M.J., Fink, A.H., Kalthoff, N., Liousse, C., Mari, C., Allan, R.P., Brooks, B., Danour, S., Flamant, C., Jegede, O.O., Lohou, F. and Marsham, J.H. (2015) The DACCIWA project: dynamics-aerosol-chemistry-cloud interactions in West Africa. *Bulletin of the American Meteorological Society*, 96, 1451–1460.

Knippertz, P., Fink, A.H., Deroubaix, A., Morris, E., Tocquer, F., Evans, M., Flamant, C., Gaetani, M., Lavaysse, C., Mari, C., Marsham, J.H., Meynadier, R., Affo-Dogo, A., Bahaga, T., Brosse, F., Deetz, K., Guebsi, R., Latifou, I., Maranan, M., Rosenberg, P.D. and Schlueter, A. (2017) A meteorological and chemical overview of the DACCIWA field campaign in West Africa in June–July 2016. *Atmospheric Chemistry and Physics*, 17, 10893–10918.

Knippertz, P., Fink, A.H., Schuster, R., Trentmann, J., Schrage, J.M. and Yorke, C. (2011) Ultra-low clouds over the southern West African monsoon region. *Geophysical Research Letters*, 38(L21808). https://doi.org/10.1029/2011GL049278

Köhler, M., Ahlgrimm, M. and Beljaars, A. (2011) Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, 137, 43–57.

Kouadio, K., Bastin, S., Konare, A. and Ajayi, V.O. (2018) Does convection-permitting simulate better rainfall distribution and extreme over the Guinean coast and surroundings?. *Climate Dynamics*, 1–22.

Lafore, J.-P., Chapelon, N., Diop-Kane, M., Gueye, B., Largeron, Y., Lepape, S., Ndiaye, O., Parker, D.J., Poan, E., Roca, R., Roehrig, R. and Taylor, C. (2017). Deep convection, Chapter 3 in Meteorology of Tropical West Africa: The Forecasters' Handbook. Parker, D.J., Diop-Kane, M. (eds), Wiley, Chichester, UK.

Lavender, S.L., Taylor, C.M. and Matthews, A.J. (2010) Coupled land–atmosphere intraseasonal variability of the West African monsoon in a GCM. *Journal of Climate*, 23, 5557–5571.

Li, R., Jin, J., Wang, J. and Gillies, R.R. (2015) Significant impacts of radiation physics in the Weather Research and Forecasting model on the precipitation and dynamics of the West African monsoon. *Climate Dynamics*, 44, 1583–1594.

Lock, A.P. (2001) The numerical representation of entrainment in parametrizations of boundary-layer turbulent mixing. *Monthly Weather Review*, 129, 1148–1163.

Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. and Smith, R.N.B. (2000) A new boundary-layer mixing scheme. Part I: scheme description and single-column model tests. *Monthly Weather Review*, 128, 3187–3199.

Louvet, S., Sultan, B., Janicot, S., Kamsu-Tamo, P.H. and Ndiaye, O. (2016) Evaluation of TIGGE precipitation forecasts over West Africa at intraseasonal timescale. *Climate Dynamics*, 47, 31–47.

Maggioni, V., Meyers, P.C. and Robinson, M.D. (2016) A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *Journal of Hydrometeorology*, 17, 1101–1117.

Maranan, M., Fink, A.H. and Knippertz, P. (2018) Rainfall types over southern West Africa: objective identification, climatology and synoptic environment. *Quarterly Journal of the Royal Meteorological Society*, 144, 1628–1648.

Maranan, M., Fink, A.H., Knippertz, P., Amekudzi, L.K., Atiah, W.A. and Stengel, M. (2020) A source-, rainfall-type andcloud-property based validation of GPM IMERG for the West African forest zone. (Personal communication, 2020).

Marsham, J.H., Dixon, N.S., Garcia-Carreras, L., Lister, G.M.S., Parker, G.M.S., Knippertz, P. and Birch, C.E. (2013) The role of moist convection in the West African monsoon system – insights from continental-scale convection-permitting simulations. *Geophysical Research Letters*, 40, 1843–1849.

Maurer, V., Kalthoff, N. and Gantner, L. (2017) Predictability of convective precipitation for West Africa: verification of convection-permitting and global ensemble simulations. *Atmospheric Chemistry and Physics*, 26, 93–110.

Menut, L., Flamant, C., Turquety, S., Deroubaix, A., Chazette, P. and Meynadier, R. (2018) Impact of biomass burning on pollutant surface concentrations in megacities of the Gulf of Guinea. *Atmospheric Chemistry and Physics*, 18, 2687–2707.

Meynadier, R., Bock, O., Gervois, S., Guichard, F., Redelsperger, J.-L., Agustí-Panareda, A. and Beljaars, A. (2010) West African monsoon water cycle: 2. Assessment of numerical weather prediction water budgets. *Journal of Geophysical Research*, 115(D19107). https://doi.org/10.1029/2010JD013919

Milton, S.F., Diongue-Niang, A., Lamptey, B., Bain, C., Birch, C. and Bougeault, P. (2017). Numerical weather prediction over Africa, Chapter 10 in Meteorology of Tropical West Africa: The Forecasters' Handbook. Parker, D.J., Diop-Kane, M. (eds). Wiley, Chichester, UK.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, 102, 16663–16682.

Morcrette, C.J. (2012) Improvements to a prognostic cloud scheme through changes to its cloud erosion parametrization. *Atmospheric Science Letters*, 13, 95–102.

Morcrette, J.-J., Barker, H.W., Cole, J.N.S., Iacono, M.J. and Pincus, R. (2008) Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, 136, 4773–4798.

Mutemi, J.N., Ogallo, L.A., Krishnamurti, T.N., Mishra, A.K. and Kumar, T.S.V.V. (2007) Multimodel-based superensemble forecasts for short- and medium-range NWP over various regions of Africa. *Meteorology and Atmospheric Physics*, 95, 87–113. https://doi.org/10.1007/s00703-006-0187-6

NOAA/NASA (1976) U.S. Standard Atmosphere. NASA-TM-X-74335, NOAA-S/T-76-1562, NASA, Washington DC

Olaniyan, E., Afiesimama, E., Oni, F. and Lawal, K.A. (2015) Simulating the daily evolution of the West African monsoon using the high-resolution regional COSMO model: a case study of the first half of 2015 over Nigeria. *Journal of Climatology and Weather Forecasting*, 3, 142.

Pante, G. and Knippertz, P. (2019) Resolving Sahelian thunderstorms improves global weather forecasts. *Nature Communications*, 10(3487). https://doi.org/10.1038/s41467-019-11081-4

Parker, D.J. and Diop-Kane, M. (Eds.) (2017) *Meteorology of Tropical West Africa: The Forecasters' Handbook*. Wiley, Chichester, UK.

Parker, D.J., Fink, A.H., Janicot, S., Ngamini, J.B., Douglas, M., Afiesimama, E., Agustí-Panareda, A., Beljaars, A., Dide, F., Diedhiou, A., Lebel, T., Polcher, J., Redelsperger, J.-L., Thorncroft, C. and Wilson, G.A. (2008) The AMMA radiosonde program and its implications for the future of atmospheric monitoring over Africa. *Bulletin of the American Meteorological Society*, 89, 1015–1027.

Prakash, S., Mitra, A.K., Pai, D.S. and AghaKouchak, A. (2016) From TRMM to GPM: how well can heavy rainfall be detected from space?. *Advances in Water Resources*, 88, 1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Raschendorfer, M. (2001) The new turbulence parameterization of LM. *COSMO Newsletter*, 1, 89–97. Available at: http://www.cosmo-model.org; accessed 30 December 2019.

Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C. and Payne, T.J. (2007) The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133, 347–362.

Reinares Martínez, I. and Chaboureau, J. (2018) Precipitation and mesoscale convective systems: radiative impact of dust over northern Africa. *Monthly Weather Review*, 146, 3011–3029.

Rieger, D., Bangert, M., Kottmeier, C., Vogel, H. and Vogel, B. (2014) Impact of aerosol on post-frontal convective clouds over Germany. *Tellus B*, 66(1). https://doi.org/10.3402/tellusb.v66.22528

Ritter, B. and Geleyn, J.-F. (1992) A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review*, 120, 303–325.

Roberts, A.J., Marsham, J.H. and Knippertz, P. (2015) Disagreements in low-level moisture between (re)analyses over summertime West Africa. *Monthly Weather Review*, 143, 1193–1211.

Roehrig, R., Bouniol, D., Guichard, F., Hourdin, F. and Redelsperger, J.-L. (2013) The present and future of the West African monsoon: a process-oriented assessment of CMIP5 simulations along the AMMA transect. *Journal of Climate*, 6, 6471–6505.

Schlueter, A., Fink, A.H. and Knippertz, P. (2019b) A systematic comparison of tropical waves over northern Africa. Part II: dynamics and thermodynamics. *Journal of Climate*, 32, 2605–2625.

Schlueter, A., Fink, A.H., Knippertz, P. and Vogel, P. (2019a) A systematic comparison of tropical waves over northern Africa. Part I: influence on rainfall. *Journal of Climate*, 32, 1501–1523.

Schrage, J.M. and Fink, A.H. (2012) Nocturnal continental low-level stratus over tropical West Africa: observations and possible mechanisms controlling its onset. *Monthly Weather Review*, 140, 1794–1809.

Schuster, R., Fink, A.H. and Knippertz, P. (2013) Formation and maintenance of nocturnal low-level stratus over the southern West African monsoon region during AMMA 2006. *Journal of the Atmospheric Sciences*, 70, 2337–2355.

Seifert, A. (2008) A revised cloud microphysical parameterization for COSMO–IME. *COSMO Newsletter*, 8, 25–28. Available at: http://www.cosmo-model.org; accessed 30 December 2019.

Shi, J.J., Matsui, T., Tao, W.-K., Tan, Q., Peters-Lidard, C., Chin, M., Pickering, K., Guy, N., Lang, S. and Kemp, E.M. (2014) Implementation of an aerosol–cloud-microphysics–radiation coupling into the NASA unified WRF: simulation results for the 6–7 August 2006 AMMA special observing period. *Quarterly Journal of the Royal Meteorological Society*, 140, 2158–2175.

Siebesma, A.P., Soares, P.M.M. and Teixeira, J. (2007) A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the Atmospheric Sciences*, 64, 1230–1248.

Söhne, N., Chaboureau, J.-P. and Guichard, F. (2008) Verification of cloud cover forecast with satellite observation over West Africa. *Monthly Weather Review*, 136, 4421–4434.

Stanelle, T., Vogel, B., Vogel, H., Bäumer, D. and Kottmeier, C. (2010) Feedback between dust particles and atmospheric processes over West Africa during dust episodes in March 2006 and June 2007. *Atmospheric Chemistry and Physics*, 10, 10771–10788.

Tan, J., Petersen, W.A. and Tokay, A. (2016) A novel approach to identify sources of errors in IMERG for GPM ground validation. *Journal of Hydrometeorology*, 17, 2477–2491.

Taylor, C.M., Gounou, A., Guichard, F., Harris, P.P., Ellis, R.J., Couvreux, F. and De Kauwe, M. (2011) Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns. *Nature Geoscience*, 4, 430–433.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V. and de Roo, A. (2012) Validation of satellite-based precipitation products over sparsely gauged African river basins. *Journal of Hydrometeorology*, 13, 1760–1783.

Thorncroft, C.D., Parker, D.J., Burton, R.R., Diop, M., Ayers, J.H., Barjat, H., Devereau, S., Diongue, A., Dumelow, R., Kindred, D.R., Price, N.M., Saloum, M., Tayor, C.M. and Tompkins, A.M. (2003) The JET2000 project: aircraft observations of the African easterly jet and African easterly waves. *Bulletin of the American Meteorological Society*, 84, 337–352.

Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117, 1779–1800.

Tiedtke, M. (1993) Representation of clouds in large-scale models. *Monthly Weather Review*, 121, 3040–3061.

Tompkins, A.M., Gierens, K. and Radel, G. (2007) Ice supersaturation in the ECMWF integrated forecast system. *Quarterly Journal of the Royal Meteorological Society*, 133, 53–63.

van de Giesen, N., Hut, R. and Selker, J. (2014) The trans-African hydro-meteorological observatory (TAHMO). *WIREs Water*, 1, 341–348.

Van den Hurk, B., Viterbo, P., Beljaars, A.C.M. and Betts, A.K. (2000). Offline validation of the ERA-40 surface scheme. Technical Memorandum 295, ECMWF, Reading, UK.

van der Linden, R., Fink, A.H. and Redl, R. (2015) Satellite-based climatology of low-level continental clouds in southern West Africa during the summer monsoon season. *Journal of Geophysical Research*, 120, 1186–1201.

Viterbo, P. and Beljaars, A.C.M. (1995) An improved land surface parameterization scheme in the ECMWF model and its validation. *Journal of Climate*, 8, 2716–2748.

Vogel, B., Vogel, H., Bäumer, D., Bangert, M., Lundgren, K., Rinke, R. and Stanelle, T. (2009) The comprehensive model system COSMO-ART – Radiative impact of aerosol on the state of the atmosphere on the regional scale. *Atmospheric Chemistry and Physics*, 9, 8661–8680.

Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A. and Gneiting, T. (2018) Skill of global raw and postprocessed ensemble, predictions of rainfall over northern tropical Africa. *Weather and Forecasting*, 33, 369–388.

Walter, C., Freitas, S.R., Kottmeier, C., Kraut, I., Rieger, D., Vogel, H. and Vogel, B. (2016) The importance of plume rise on the concentrations and atmospheric impacts of biomass-burning aerosol. *Atmospheric Chemistry and Physics*, 16, 9201–9219.

Watts, P.D., Bennartz, R. and Fell, F. (2011) Retrieval of two-layer cloud properties from multispectral observations using optimal estimation. *Journal of Geophysical Research*, 116, D16203.

Webster, P.J. (2013) Improve weather forecasts for the developing world. *Nature*, 493, 17–19.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D. and Morcrette, C.J. (2008a) PC2: a prognostic cloud fraction and condensation scheme. I: scheme description. *Quarterly Journal of the Royal Meteorological Society*, 134, 2093–2107.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D., Morcrette, C.J. and Bodas-Salcedo, A. (2008b) PC2: a prognostic cloud fraction and condensation scheme. II: climate model simulations. *Quarterly Journal of the Royal Meteorological Society*, 134, 2109–2125.

WMO (2017) *International Cloud Atlas, Manual on the Observation of Clouds and Other Meteors*. WMO-No. 407, volume 3. Available at: https://cloudatlas.wmo.int/home.html; accessed 30 December 2019.

Wood, N., Staniforth, A., White, A.A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S.P., Zerroukat, M. and Thuburn, J. (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140, 1505–1520.

Young, M.P., Chiu, J.C., Williams, C.J.R., Stein, T.H.M., Stengel, M., Fielding, M.D. and Black, E. (2018) Spatio-temporal variability of warm rain events over southern West Africa from geostationary satellite observations for climate monitoring and model evaluation. *Quarterly Journal of the Royal Meteorological Society*, 144, 2311–2330.

Zängl, G., Reinert, D., Rípodas, P. and Baldauf, M. (2015) The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141, 563–579.