

## What Is the Added Value of a Convection-Permitting Model for Forecasting Extreme Rainfall over Tropical East Africa?

BETH J. WOODHAMS,<sup>a</sup> CATHRYN E. BIRCH,<sup>a,b</sup> JOHN H. MARSHAM,<sup>a,c</sup> CAROLINE L. BAIN,<sup>b</sup>  
NIGEL M. ROBERTS,<sup>d</sup> AND DOUGLAS F. A. BOYD<sup>b</sup>

<sup>a</sup> *School of Earth and Environment, University of Leeds, Leeds, United Kingdom*

<sup>b</sup> *Met Office, Exeter, United Kingdom*

<sup>c</sup> *National Centre for Atmospheric Science, University of Leeds, Leeds, United Kingdom*

<sup>d</sup> *MetOffice@Reading, Met Office, Reading, United Kingdom*

(Manuscript received 31 December 2017, in final form 18 June 2018)

### ABSTRACT

Forecasting convective rainfall in the tropics is a major challenge for numerical weather prediction. The use of convection-permitting (CP) forecast models in the tropics has lagged behind the midlatitudes, despite the great potential of such models in this region. In the scientific literature, there is very little evaluation of CP models in the tropics, especially over an extended time period. This paper evaluates the prediction of convective storms for a period of 2 years in the Met Office operational CP model over East Africa and the global operational forecast model. A novel localized form of the fractions skill score is introduced, which shows variation in model skill across the spatial domain. Overall, the CP model and the global model both outperform a 24-h persistence forecast. The CP model shows greater skill than the global model, in particular on subdaily time scales and for storms over land. Forecasts over Lake Victoria are also improved in the CP model, with an increase in hit rate of up to 20%. Contrary to studies in the midlatitudes, the skill of both models shows a large dependence on the time of day and comparatively little dependence on the forecast lead time within a 48-h forecast. Although these results provide more motivation for forecasters to use the CP model to produce subdaily forecasts with increased detail, there is a clear need for more in situ observations for data assimilation into the models and for verification. A move toward ensemble forecasting could have further benefits.

### 1. Introduction

Forecasting tropical convection remains a huge challenge for numerical weather prediction (NWP). In particular, there is an urgent need to improve forecasting of high-impact weather, such as heavy convective precipitation, in sub-Saharan Africa. Because of their coarse resolution, traditional global models use parameterization schemes to represent convection. Parameterized models are not designed to produce realistic storm structures; they almost ubiquitously produce too much light rain and are unable to capture the highest-intensity events (Sun et al. 2006; Dai 2006; Stephens et al. 2010). Such models are unable to simulate the diurnal cycle; the convective

maximum tends to occur at midday, rather than late afternoon, as observed (Yang and Slingo 2001; Bechtold et al. 2004). Marsham et al. (2013) and Birch et al. (2014b) showed that an incorrect diurnal cycle may introduce errors into the synoptic-scale flow. In addition, flows over complex topography and those within deep convective storms cannot be well represented at coarse resolutions (Clark et al. 2016).

Low skill in tropical precipitation prediction means that forecasters often use other information from models, such as the synoptic-scale circulation and stability measures, to determine favorable conditions for deep convection. Consequently, predictions are heavily reliant on forecasters' experience and knowledge of the model and meteorology of the region (Lafore et al. 2017).

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Beth Woodhams, eebjw@leeds.ac.uk



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/MWR-D-17-0396.1

© 2018 American Meteorological Society

Unauthenticated | Downloaded 11/08/22 01:51 PM UTC

Convection-permitting (CP) models provide a step change in the representation of convective storms because they explicitly represent the storms themselves (Clark et al. 2016). CP models produce more realistic-looking precipitation fields and have an improved diurnal cycle, with the peak in convection shifted further toward the observed late-afternoon maximum (Lean et al. 2008; Done et al. 2004; Weisman et al. 2008; Weusthoff et al. 2010; Birch et al. 2014b; Prein et al. 2015). Marsham et al. (2013) and Garcia-Carreras et al. (2013) have shown that the finer horizontal grid spacing of CP models allows the simulation of cold pools, which can affect synoptic-scale fluxes in the models and trigger new convection. CP models also better capture the organization and propagation of convection (Weisman et al. 2008; White et al. 2018). The benefits of using CP models are even being utilized for future climate projections to provide more details on regional and local scales (Prein et al. 2015; Stratton et al. 2018).

However, increased realism of rainfall may not translate to increased forecast skill. CP models can produce too much rain, in particular at high intensities, while the proportion of low-intensity events is too small (Lean et al. 2008; Kendon et al. 2012; Marsham et al. 2013; Done et al. 2004). The horizontal grid spacing remains on a scale larger than most convective updrafts, meaning that cloud structures are still underresolved, causing errors in the multiscale interactions, upscale growth, and timing of storms (Clark et al. 2016; Lean et al. 2008). Lean et al. (2008) showed that CP models often perform poorly at the start of runs, since it takes time for the high-resolution detail to “spin up” from the initial fields provided by the driving model.

High-resolution simulations of mesoscale convective systems (MCSs) in various regions of the world have shown that the accuracy of the initial conditions (ICs) and boundary conditions provided by the driving model are key factors responsible for the correct prediction of initial development of a storm (Birch et al. 2013; Guichard et al. 2010; Melhauser and Zhang 2012; Schumacher et al. 2013; Luo and Chen 2015; Vié et al. 2011). Sensitivity to ICs is a particular concern in much of the tropics because of a lack of routine observations for data assimilation into the driving model. These studies also showed that the predictability of a storm may depend on the type of synoptic flow. High-resolution CP models are fundamentally limited by the fact that predictability times are reduced on cloud-resolving scales, compared to synoptic scales. Small-scale errors may grow more quickly and influence larger scales (Lorenz 1969; Hohenegger and Schär 2007).

In recent years, as computing power has increased, CP models have become more feasible and more commonly

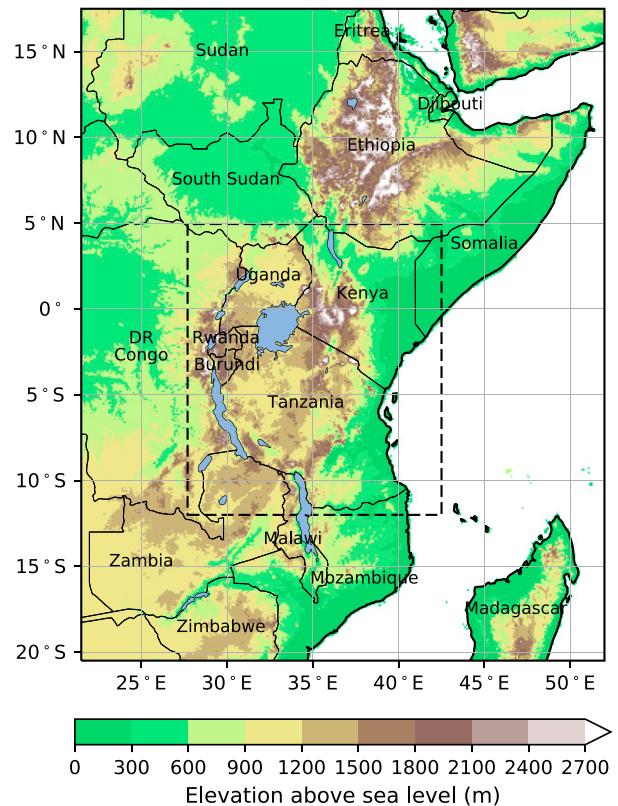


FIG. 1. A map showing the orography over the domain spanned by the current operational Met Office CP model in East Africa. Orography data are from the Global Self-Consistent, Hierarchical, High-Resolution Geography (GSHHG) database (Wessel and Smith 1996). The dashed box encloses the LV subdomain used in the analysis.

used in operational weather forecasts. They have been used at a country level by many national meteorological services for over a decade. The Met Office uses a CP version of the Met Office Unified Model (MetUM) over the United Kingdom and some tropical domains (Tang et al. 2013); the U.S. Weather Research and Forecasting (WRF) Model is CP and used by multiple agencies worldwide (Michalakes et al. 2001), as is the CP version of COSMO, created in Germany (Baldauf et al. 2011). The Application of Research to Operations at Mesoscale (AROME) model developed in France (Seity et al. 2011) and the Japan Meteorological Agency’s Non-hydrostatic Mesoscale Model over Japan (Saito et al. 2006) are also CP models. Many of these models are configured to run operationally beyond their country of origin. Although many CP forecast models are used across the tropics, published verification is limited, especially for an extended period.

Equatorial East Africa (Fig. 1) is a region that receives the majority of its rainfall via deep convective storms. The prediction of severe convection is acutely important in this region, which is at high risk of severe

flooding and drought. It is particularly important to provide accurate forecasts for the Lake Victoria (LV) basin. This densely populated region supports around 35 million people, including 200 000 fishermen. Nocturnal storms over Lake Victoria produce intense precipitation and high winds, which capsize boats. There are an estimated 5000 deaths on the lake each year, with many attributed to severe weather (World Bank DGF 2011). Forecasting convection in this environment with orography and land–lake circulations is a challenge. The localized nature of convection and a lack of observations, especially upper air, to assimilate into forecasts add to the difficulty.

In 2011, the Met Office began running an operational CP forecast model over Lake Victoria, with a horizontal grid spacing of 4.4 km, funded through the Met Office Voluntary Cooperation Programme (VCP) and intended to aid the forecast of severe weather events, in particular over the Lake Victoria basin (Chamberlain et al. 2014). The domain was extended to cover a larger region in February 2014, which currently remains operational. Output from the model is disseminated to operational meteorologists in East Africa (principally, Kenya, Uganda, Tanzania, Rwanda, and Burundi). Model output is available to view on the VCP Africa Web Viewer, a password-controlled site open to African forecasters, which also shows output from the global model and recent satellite imagery and arrival time difference (ATD) lightning.

Some verification and comparison with the global MetUM was performed on the smaller domain by Chamberlain et al. (2014) for spring 2012. Overall, and in agreement with studies of other CP models, both the global and the CP models produced too much light rainfall, especially the global model. The CP model predicted too many intense rainfall events, whereas the global model was unable to produce any of the highest-intensity rainfall rates observed. Objective analysis showed the CP model to have more skill in predicting when a storm would occur, compared to the global model. However, it was also shown to overpredict severe events, leading to more “false alarms.”

Away from NWP, Thiery et al. (2017) developed a prototype of a statistical storm predictor [Lake Victoria Intense Early Warning System (VIEWS)] that uses satellite observations to forecast nocturnal storms over the lake. The model is based on the strong correlation between intense storms over land during the afternoon and intense nocturnal storms over Lake Victoria found by Thiery et al. (2016). While this prototype can achieve high hit rates and low false alarm rates, it has a lead time of only a few hours. NWP models are vital for providing much earlier warnings, with statistical model and now-casting techniques playing an important role in warning confidence and refinement closer to the event.

This paper investigates whether a CP model over East Africa provides additional skill for the forecasting of severe tropical rainfall, compared to a global model. The rainfall field in the model was chosen for verification following meetings with forecasters from East Africa. While forecasters use other model fields, such as surface pressure, relative humidity, and winds, to ensure that the model rainfall field is realistic, the rainfall field is the main tool used to produce forecasts. The most commonly used output is 24-h rainfall accumulations, despite the availability of accumulations at 3-hourly intervals. A particular aim of the paper is to determine whether the CP model provides any added value on subdaily time scales or at specific locations. The effect of forecast lead time and time of day is also investigated. Much of the analysis is performed using the fractions skill score (FSS) proposed by Roberts and Lean (2008), along with an examination of model biases. The analysis builds on that of Chamberlain et al. (2014) by performing more detailed verification on a larger model domain over a period of 2 years. The paper emphasizes the implications of the findings on the operational aspects of the model, including how it is run and how it may best be used by forecasters.

The model configurations, observational data, and verification methods are introduced in section 2. Section 3 presents the characteristics of precipitation and its diurnal cycle in the models, alongside verification of the model skill at different spatial scales and at different times of day. The implications of these results are discussed and conclusions drawn in section 4.

## 2. Methods

This study consists of a comparison of the forecast skill of the MetUM CP model and the MetUM global model for rainfall over East Africa, alongside an analysis of the performance of the models at different forecast lead times and different times of day. The period of study is the 2 years between 18 July 2014 and 17 July 2016. Analysis was performed over the whole model domain, as well as a subdomain centered on LV, shown within the dashed box in Fig. 1 and chosen to include all countries using the CP model.

### a. Models

Both the CP model over East Africa and the global model are subsets of the MetUM and were used operationally in 2017. The start date of the analysis period marks the upgrade of the global model to GA6.1 and the Even Newer Dynamics for General atmospheric modeling of the environment (ENDGame) dynamical core (Wood et al. 2014), with a reduced horizontal grid spacing of

approximately 17 km in the meridional direction by 25 km in the zonal direction (in the tropics). The model is initialized every 6 h, at 0000, 0600, 1200, and 1800 UTC, although only the 0000 and 1200 UTC initializations are considered to allow direct comparison with the CP model. Convection is parameterized in the global model using the mass-flux scheme introduced by [Gregory and Rowntree \(1990\)](#) and with subsequent enhancements.

The CP model has a horizontal grid spacing of 4.4 km in both directions and spans the domain from 20.5°S to 17.5°N and from 21.5° to 52°E ([Fig. 1](#)). The model grid consists of  $762 \times 950$  grid points and has 70 vertical levels up to a lid of 40 km. The model is run with a time step of 100 s. Convection is treated explicitly, although the model also uses a convective available potential energy (CAPE)-dependent closure scheme ([Roberts 2003](#)) used by the Met Office in all operational models with this configuration. This scheme adjusts the time scale over which instability is removed in accordance with the amount of CAPE that is present in order to restrict the parameterized mass flux. This allows explicit convection while trying to parameterize the effects of smaller clouds that the CP model cannot resolve. The parameters within the microphysics and subgrid mixing have not been “tuned” for the tropics, but are the same as those used in the 4-km U.K. (UK4) model ([Eagle et al. 2015](#)). The New Dynamics (ND) dynamical core ([Davies et al. 2005](#)) was replaced with ENDGame [along with some other physics changes detailed in [Eagle et al. \(2015\)](#)] approximately halfway through the study period. [Eagle et al. \(2015\)](#) ran the model in both configurations and found that they produced very similar results, suggesting that the model may be considered to have similar behavior throughout the whole analysis period.

The CP model is initialized twice per day, at 2100 and 0900 UTC, using 3-h forecasts from the global model (i.e., ICs are taken from 3 h into the global model 1800 and 0600 UTC runs, respectively). The models are allowed 3 h for spin up until the first diagnostics are outputted at 0000 and 1200 UTC, respectively. Forecasts are outputted up to a lead time of 48 h from the first diagnostics ([Chamberlain et al. 2014](#)). The two initializations are referred to as the 0000 and 1200 UTC initializations since these are the times of the first available forecasts. The model is forced by lateral boundary conditions from the global model, updated every 3 h. Smoothing is applied at the edges to remove discontinuities between the models.

In both models, lake surface temperatures (LSTs) are prescribed as the foundation water surface temperature (temperature below the diurnal warm layer) taken from daily Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA), available on a  $1/20^\circ$  ( $\sim 6$  km) grid

([Fiedler et al. 2014](#)). Observations are obtained from in situ data received via the Global Telecommunication System (GTS; although no in situ observations existed over Lake Victoria during this study) and satellite sea surface temperature (SST) data from the Group for High-Resolution SST (GHRSSST). Only overnight observations and daytime observations for wind speeds greater than  $6 \text{ m s}^{-1}$  are used, such that the effect of the diurnal warm layer above the sea surface is not included. These observations are assimilated onto a background field of the analysis from the previous day, slightly relaxed toward the [MacCallum and Merchant \(2011\)](#) ARC-Lake nighttime climatology ([Donlon et al. 2012](#)). Using OSTIA, the temperature in the model is updated once per day to the predawn value. When observations are unavailable, the LST in OSTIA will relax toward the ARC-Lake climatology over a period of 30 days ([Fiedler et al. 2014](#)).

At the time of the study, the OSTIA system included no lake-specific processing; satellite retrievals were optimized for over oceans, meaning that the different wind and cloud regimes and the elevations and continental locations of some lakes may have introduced errors into the satellite retrievals ([Fiedler et al. 2014](#)). However, [Fiedler et al. \(2014\)](#) note that Lake Victoria performs very well, compared to other lakes across the globe, likely due to its large size, position on the equator, and relatively low elevation. This is important since [Thiery et al. \(2015\)](#) and [Argent et al. \(2015\)](#) showed that accurate LSTs are crucial for reproducing observed precipitation patterns over and around Lake Victoria. In addition, Lakes Malawi and Tanganyika (also in the model domain) are said to perform well. In JJA 2009, the lakes in this region were found to have an average of several hundred observations per day due to satellite overpasses ([Fiedler et al. 2014](#)). However, since JJA is a dry period, the number of observations during the wet seasons may be decreased as sampling issues due to cloud cover increase.

For some analysis purposes, a time series of model data was required. For both models, a set of four different time series was formed by stitching together data from either forecast lead times between  $T + 12$  and  $T + 33$  h or between  $T + 24$  and  $T + 45$  h, for both the 0000 and 1200 UTC initializations. In addition, four different 24-h accumulation periods were defined using the same lead time bounds and the two initialization times.

### *b. Observations*

Few in situ observations are available for the region during the analysis period. There were no working radars and very few weather stations, forcing a reliance on satellite-derived data for observations. Precipitation

intensity observations are sourced from the Global Precipitation Measurement (GPM) mission, in particular the IMERG Final Precipitation version 5 (V05) level 3 product at  $0.1^\circ$  (Huffman 2017; Huffman et al. 2018). The product is available at 30-min time steps, but only times matching the model output were used. GPM IMERG is the successor to the Tropical Rainfall Measuring Mission (TRMM) 3B42 V7 product (Liu et al. 2012; TRMM 2011). IMERG takes observations from a network of satellites in the GPM constellation and unifies them to create a gridded product. In particular, the GPM *Core Observatory* satellite hosts a dual-frequency precipitation radar (DPR) and a conical-scanning multichannel microwave imager (GMI) (Hou et al. 2014). These two instruments are used as a reference to intercalibrate the passive microwave (PMW) precipitation estimates from other satellites in the GPM constellation using the method developed for TRMM by Huffman et al. (2007) and including calibration against the Global Precipitation Climatology Centre (GPCC) gauge analysis by Schneider et al. (2008) (Huffman et al. 2018; Hou et al. 2014). Since the PMW sensors do not have complete coverage of the ground, the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center morphing technique with Kalman filter (CMORPH-KF) (Joyce et al. 2004; Joyce and Xie 2011) is applied, which estimates precipitation outside the sensed area by propagation of PMW estimates with motion vectors derived from geosynchronous IR satellite imagery. For even further coverage, the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Cloud Classification System (PERSIANN–CCS) uses IR retrievals, calibrated against PMW retrievals, to estimate rainfall (Hong et al. 2004; Sorooshian et al. 2000). For brevity, the GPM IMERG product is referred to as GPM for the remainder of the paper.

Many studies in different parts of the world, including many with complex topography, have shown that the GPM product outperforms its predecessor, the TRMM 3B42 V7 product, on various spatial and temporal scales (Tang et al. 2016a,b; Kim et al. 2017; Xu et al. 2017; Wang et al. 2017b; Sharifi et al. 2016; Prakash et al. 2018). In particular, GPM is better able to detect low-intensity rainfall due to four more high-frequency channels on the GMI instrument, compared to the corresponding instrument for TRMM. From studies over mainland China, Tang et al. (2016a) suggested that GPM required improvement in dry climates and high altitudes. This could be a cause for concern over the Horn of Africa and East African highlands. Kim et al. (2017) noted uncertainties with orographic convection, and Xu et al. (2017) found that GPM has issues for

orography greater than 4500 m. O and Kirstetter (2018) found an underestimation of the diurnal variation over mountains in the United States. Further studies have shown that GPM underestimates high-intensity precipitation events (Wang et al. 2017a; O et al. 2017).

For comparison with GPM, the bias-corrected V1.0 CMORPH precipitation intensity dataset on an 8-km grid (NCEP 2017) is also used. The CMORPH algorithm by Joyce et al. (2004) is also used to produce GPM as described above, but for CMORPH, V1.0, is used without the Joyce and Xie (2011) Kalman filter. As for GPM, CMORPH is available every 30 min, but only times matching the model output are used. The performance of CMORPH has been shown to be very variable, both spatially and temporally (Zeweldi and Gebremichael 2009; Habib et al. 2012; Haile et al. 2013). Studies have shown that CMORPH is often unable to capture the highest-intensity events and therefore overestimates the frequency of lower-intensity rain events (Kumar et al. 2016; Habib et al. 2012). However, over complex terrain in Mexico, Nesbitt et al. (2008) found an overestimation of precipitation within deep convective systems. Several studies over Ethiopia found a general underestimation of rainfall rates (Haile et al. 2013; Romilly and Gebremichael 2011; Hirpa et al. 2010). Romilly and Gebremichael (2011) note that over northwestern Ethiopia, where the ITCZ has a strong effect and the climate is humid, CMORPH tends to overestimate rainfall at low elevations but performs well at higher elevations. Contrastingly, in the northeast, Hirpa et al. (2010) found an underestimation at high elevations. This shows the variability in performance of CMORPH over even a small area.

All model and observational data were interpolated onto the same grid for analysis. The global model has the coarsest resolution ( $\sim 17 \text{ km} \times 25 \text{ km}$ ), but square grid boxes were required for some of the analysis; therefore, all data were interpolated onto a regular  $0.25^\circ$  grid.

### c. Verification methods

#### 1) OBJECTIVE ANALYSIS

Objective analysis was performed to compare the skill of the models in forecasting storms over Lake Victoria. A storm was identified if rainfall occurred with a minimum given intensity over a minimum given area. A range of size and intensity thresholds was sampled. Stormy 3-h periods were identified in both the models and observations to obtain the number of hits (storm in both the observations and forecast), false alarms (storm was forecast but did not occur), misses (storm occurred but was not forecast), and correct negatives (storm was not forecast and did not occur). These were then used to compute the hit rate [hits/(hits + misses)] and false

alarm ratio [false alarms/(hits + false alarms)]. The false alarm rate [false alarms/(correct negatives + false alarms)] was also computed.

Using a fixed size and intensity threshold to define a storm over the lake in the observations, the hit rate was plotted against the false alarm rate for a variety of forecast intensity thresholds (keeping the size threshold the same as in the observations) to produce a receiver operating characteristic (ROC) curve (Swets 1973; Mason 1982). A good forecast should maximize the hit rate and minimize the false alarm rate; hence, the ROC curve should lie in the upper-left half of the ROC diagram. The closer the curve lies to false alarm rate = 0 and hit rate = 1, the greater the skill. This skill was captured by computing the area under the curve (AUC), which is greater than 0.5 for a skillful forecast and equal to 1 for a perfect forecast (Mason and Graham 2002; Wilks 2011).

## 2) FRACTIONS SKILL SCORE

The ability of the models to forecast storms at the correct time and in the correct location was assessed. The small horizontal grid spacing of the CP model meant that traditional point-to-point verification methods were not appropriate because small-scale errors can be heavily penalized. If the forecast position of the storm is offset from the true position of the storm, point-to-point verification applies a “double penalty”: where the storm was forecast will record false alarms, and where the storm did occur will record misses, resulting in a low skill score (Roberts and Lean 2008; Mittermaier 2012). Forecasters using the model should be aware of this and use judgement to forecast storms over a region, rather than at a specific location.

For this reason, the FSS, developed by Roberts and Lean (2008), is used in this study. This metric aims to measure the variation in skill with spatial scale and hence the smallest scale at which a model has skill. The verification process is as follows: first, a rainfall rate threshold is chosen. At each grid box, the fraction of points that exceeds this threshold within a surrounding  $n \times n$  gridpoint “neighborhood” is recorded for both the model and observations for increasing values of  $n$ . The square of the difference between the fraction in the model  $M_{(n)}$  and the fraction in the observations  $O_{(n)}$  is averaged over all grid points at a given time to compute the mean squared error (MSE):

$$\text{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)i,j} - M_{(n)i,j}]^2, \quad (1)$$

where  $N_x$  and  $N_y$  are the number of longitude and latitude points, respectively. Since the MSE has a high dependence on the frequency of the event, it is compared

against the MSE of a reference forecast with low-skill  $\text{MSE}_{(n)\text{ref}}$ , defined in Murphy and Epstein (1989) as

$$\text{MSE}_{(n)\text{ref}} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)i,j}^2 + M_{(n)i,j}^2], \quad (2)$$

to obtain the fractions skill score:

$$\text{FSS}_{(n)} = 1 - \frac{\text{MSE}_{(n)}}{\text{MSE}_{(n)\text{ref}}}. \quad (3)$$

A score of 1 implies that the forecast has a perfect match for the neighborhood size used, whereas a score of 0 implies that the forecast has no skill. The skill score that would be achieved, on average, by a random forecast with the same fraction of events  $f_0$  over the domain as the observations is given by  $f_0$ . A “target” or uniform forecast is defined as a forecast in which the model fraction  $M_{(n)}$  at each grid point is equal to  $f_0$ . The score for such a forecast is given by  $0.5 + (f_0/2)$  and can be approximated to 0.5 for comparisons when the frequencies are small (Roberts and Lean 2008; Roberts 2008). In fact, if the FSS is being used as a measure of spatial displacement, it is the value of 0.5 that should be used (Skok 2015; Skok and Roberts 2016). The spatial scale at which the model reaches this skill can be interpreted as the spatial scale on which the model has skill. More information regarding the details of the FSS calculation is given in the appendix.

The traditional FSS generates a skill score at each time step to describe the whole domain and allow variation in the model performance with time to be studied. However, FSS cannot be used to understand how the skill of the model varies across the spatial domain. An adapted version of the FSS, termed the localized fractions skill score (LFSS), was used to do this. The mean squared error between the fraction of grid points exceeding the threshold in the model and observations is found by taking the mean over time for a given neighborhood size, rather than over the spatial domain. In this version of the score, Eqs. (1) and (2) are therefore replaced by

$$\text{MSE}_{(n)} = \frac{1}{N_t} \sum_{k=1}^{N_t} [O_{(n)k} - M_{(n)k}]^2, \quad (4)$$

and

$$\text{MSE}_{(n)\text{ref}} = \frac{1}{N_t} \sum_{k=1}^{N_t} [O_{(n)k}^2 + M_{(n)k}^2], \quad (5)$$

respectively, where  $N_t$  is the number of time steps. Rather than a skill score per time, a skill score is instead obtained per grid point over a period of time for the

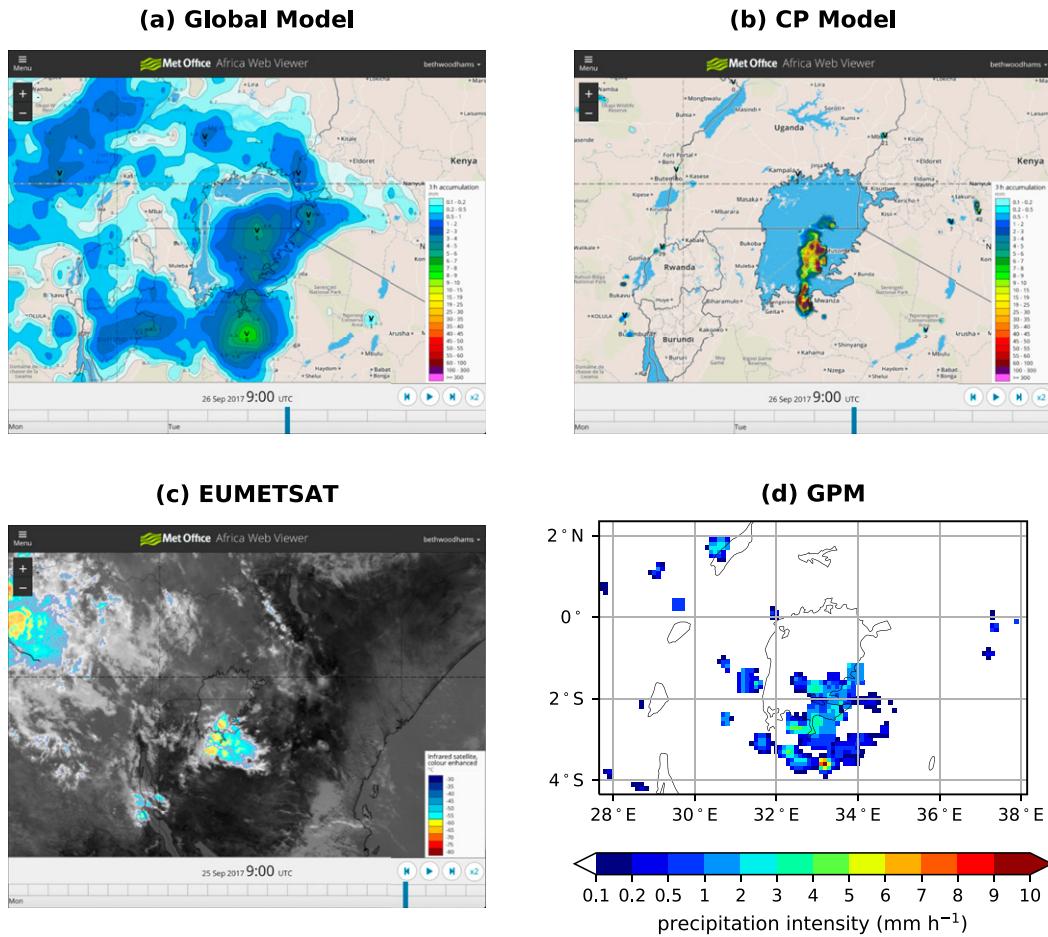


FIG. 2. The  $T + 33$  h 3-h precipitation accumulation forecasts from (a) the global model and (b) the CP model, and (c) IR observations from EUMETSAT for 0900 UTC 26 Sep 2017; screenshots taken from the Met Office VCP Africa Web Viewer. (d) Instantaneous precipitation intensity at 0900 UTC from GPM.

neighborhood size of interest. Because the neighborhood size is still in the spatial domain, the LFSS cannot have a “target” score to find the displacement of storms in time, which would require the size of the neighborhood to be in the time dimension. The LFSS is therefore still used to understand the spatial skill. While the FSS should be used to quantitatively find a neighborhood size on which the model has skill, the LFSS should be used more qualitatively to find regions that have greater or lesser skill relative to the whole domain.

### 3. Results

#### a. Forecast example

Figure 2 shows an example of a 3-h rainfall accumulation forecast over Lake Victoria from (Fig. 2a) the global model and (Fig. 2b) the CP model, as would be seen by a forecaster using the Met Office VCP Africa Web Viewer. This example is from 0900 UTC 26 September 2017

(1200 LT; LT = UTC + 3 h). This case was chosen as an exemplar case of a good forecast by the CP model, which exhibits many of the characteristics typical of the model performance for a significant storm event over the lake. Both models correctly predicted an event over the lake, with the CP model predicting a more realistic storm structure. Although the global model forecast did show some structure in the precipitation field, it did not produce an organized storm. Rather, it forecast rain over most of the lake. The CP model forecast a linear storm structure, with a region of heavy precipitation over the south of the lake and another region over land on the southern shore. IR observations available on the Africa Web Viewer in Fig. 2c show that the regions of rainfall forecast by the CP model coincided well with areas of low IR. Precipitation rates from GPM (Fig. 2d) show that the heaviest region of precipitation on the southern shore collocated fairly well with the area of highest accumulations in the CP model forecast, although it was slightly

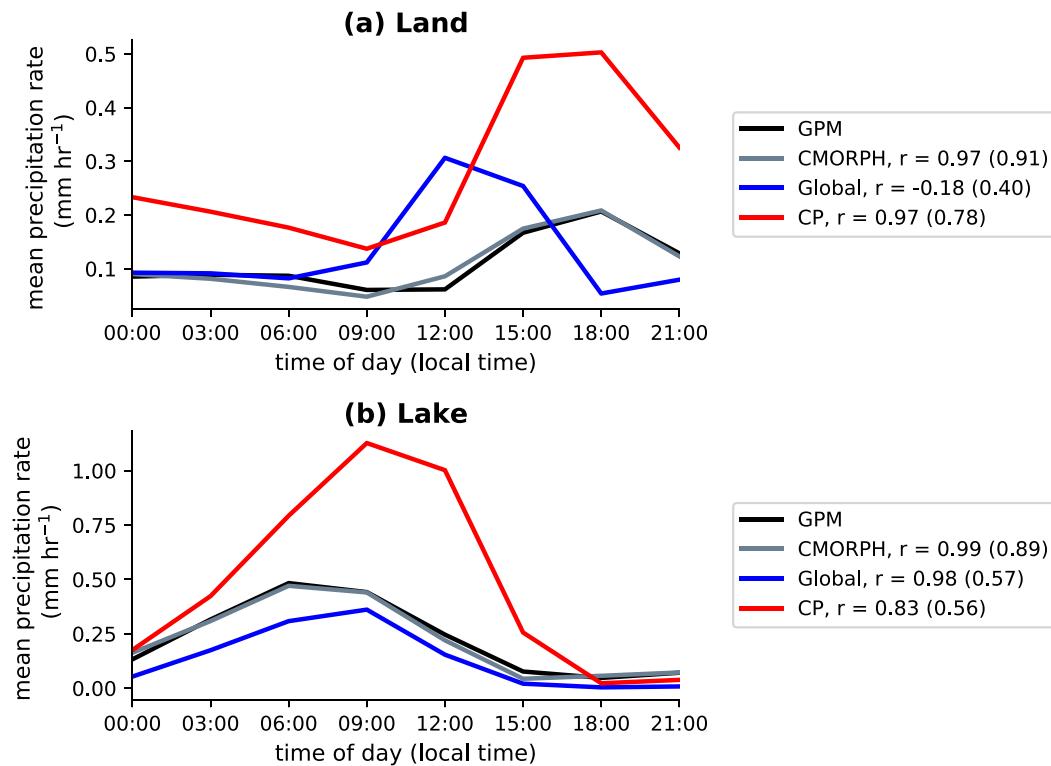


FIG. 3. Mean diurnal cycle of precipitation rates over (a) the land grid points within the LV subdomain (dashed black line in Fig. 1) and (b) the lake grid points of LV for GPM, CMORPH, and the global and CP models. The Pearson's  $r$  value for the correlation of the mean diurnal cycle with that of GPM is given in the legend. The correlation over the full 2-yr time series is also given in brackets. The four model time series (comprising different initialization and lead times as described in section 2a) are used for the global and CP model data, so the results reflect an average across these datasets. This is true in all figures unless stated otherwise.

farther south. The observed rainfall rates suggest that the forecast accumulation was too high and also too localized.

### b. Diurnal cycle

The mean diurnal cycles over (Fig. 3a) the land grid points within the LV subdomain and (Fig. 3b) Lake Victoria (lake grid points only) are shown for GPM observations (and CMORPH for comparison) and for the global and CP models. Over land, the observations show that the convective maximum occurs at 1800 LT. During the day, the land surrounding the lake heats up faster than the lake itself, causing divergence and hence subsidence over the lake (Chamberlain et al. 2014). Onshore winds trigger storms over the mountains to the east of the lake, which reach peak intensity late in the afternoon. However, the global model shows a mean rainfall maximum at 1200 LT, which is larger than the peak in the observations by a factor of 1.5. There is actually a weak negative correlation between the mean diurnal cycle in the model and observations (Pearson's  $r$  value of  $-0.18$ ). Comparing the full 2-yr time series from the global model to that of the observations, the

correlation coefficient is 0.40, showing an improved relationship compared to the mean. The erroneous mid-day maximum is in agreement with previous studies of global models (Yang and Slingo 2001; Bechtold et al. 2004). The peak in rainfall in the CP model occurs at 1800 LT, in agreement with the observations, although the mean rainfall rate at 1500 LT is nearly as great. The CP model overpredicts the maximum rainfall rate by a factor of 2.4, consistent with known biases in CP MetUM (Lean et al. 2008; Kendon et al. 2012), but improved in recent research configurations (Aranami et al. 2015; Zerroukat and Shipway 2017). Despite this, correlation coefficients of 0.97 and 0.78 are achieved by the CP model for the mean diurnal cycle and the full 2-yr time series, respectively. CMORPH exhibits a similar diurnal cycle to GPM over land, with a correlation coefficient of 0.97 for the mean diurnal cycle and 0.91 across the whole time series.

At night, the lake remains warmer than the land, such that convergence and, hence, convection occurs over the lake, and the observed precipitation peaks at 0600 LT (Song et al. 2004). The maximum rainfall occurs later, at

0900 LT, in the global and CP models. The global model slightly underpredicts the maximum mean rainfall rate over the lake, whereas the rate in the CP model is a factor of 2.3 greater than GPM. The correlation between the mean diurnal cycle and the models is greater for the global model (0.98) than the CP model (0.83), although both show strong correlations. The correlation coefficients between both models and the observations over the full time series are very similar: 0.57 and 0.56 for the global and CP model, respectively. The ability of the global model to capture the correct diurnal cycle over the lake, when it is unable to do so over land, suggests that the forcings that lead to storms over the lake are particularly strong and on a scale greater than the grid spacing of the global model. The consistently warm temperature of the lake relative to the land may allow storms to be triggered in the parameterization scheme, without the need for solar insolation, while the positioning of the rain is constrained by the location of the lake. CMORPH correlates very well with GPM, with a correlation coefficient of 0.99 for the mean diurnal cycle and 0.89 for the whole time series over the lake.

### c. Precipitation rates

The precipitation rates corresponding to different percentiles are shown in Fig. 4 for the observational data (GPM) and the two models over the full domain (solid line) and over the LV subdomain (dashed line). Precipitation rates for the additional observations from CMORPH are also shown. The corresponding rainfall rates were found using the data from all times and all grid points within the analysis period and corresponding domain. Most of the rainfall intensities within this time were  $0 \text{ mm h}^{-1}$ . This is reflected in Figs. 4a and 4b, which show the rainfall rates as a function of percentile, computed including times of no rain. Figure 4a demonstrates that while the models do predict a large proportion of dry events, they predict too many rainfall events, compared to observations. While 90.9% of all data points in the full domain have no rain in the observations, this is true of 75.8% and 81.1% of data points in the global and CP models, respectively. Similar results hold over the LV subdomain. Figure 4b shows that above the 99.9th percentile, the rain rates in the global model are 15% and 20% lower than the observations over the full domain and LV subdomain, respectively. However, the CP model produces far too much heavy rain, over a factor of 3.5 greater than the observations for the 99.9th percentile, and increasing with greater percentiles. However, studies such as Wang et al. (2017a) and O et al. (2017) suggest that GPM may underestimate high rainfall events, so the difference between the CP model and true rainfall amount may not be so large. Overall, there

is little difference between the full domain and LV subdomain.

After removing data points with no rain, Figs. 4c and 4d show how the rainfall rates are distributed. Figure 4c shows that of the rain that was produced by both models (and over both domains), the rain was too light for the majority of data points. Figure 4d shows that the global model is unable to predict the highest-intensity events, while the CP model has a disproportionate amount of unrealistically extreme events, with the precipitation rate around a factor of 3 too large above the 99.9th percentile.

The contribution of rain of different intensities to the mean rainfall intensity is shown in Fig. 4e for GPM and the two models. Again, this shows that the global model produces light precipitation too often, compared to GPM. For the very highest intensities, the global model matches the observations well. Despite being predicted relatively too often, the lightest-intensity precipitation contributes to the mean rainfall too infrequently in the CP model, because rain rates above  $5 \text{ mm h}^{-1}$  are produced far too frequently.

Comparing the two observational datasets, Fig. 4a shows that CMORPH is drier than GPM. GPM has higher extreme rainfall rates over the whole domain, but CMORPH has the highest rain rates over the LV subdomain (Figs. 4b,d). This difference could be related to the performance of the two observational datasets over the complex topography within the subdomain. Low rainfall rates contribute to the mean rainfall less in CMORPH than in GPM, but intermediate rainfall rates (between 6 and  $19 \text{ mm h}^{-1}$ ) contribute more (Fig. 4e). Overall, the differences between the characteristics of the observational datasets are smaller than the differences between the models and observations. Therefore, only results using GPM observations are presented in the remainder of the paper. Much of the following work was reproduced using CMORPH and yielded very similar results.

### d. Lake Victoria objective analysis

Because of the dangers faced by fishermen when a storm occurs over Lake Victoria, the ability of the models to forecast storms over the lake itself is extremely important. Objective analysis was performed to compute the hit rate, false alarm ratio, and ROC AUC for a range of minimum storm sizes and intensities over the lake. The intensity threshold was defined by a percentile threshold (calculated using only lake grid points) to avoid issues with the different distributions of rain rates between the models and observations.

Figures 5a and 5b show the hit rates achieved by the global and CP models, respectively, as a function of minimum size and intensity, and Fig. 5c shows the difference in

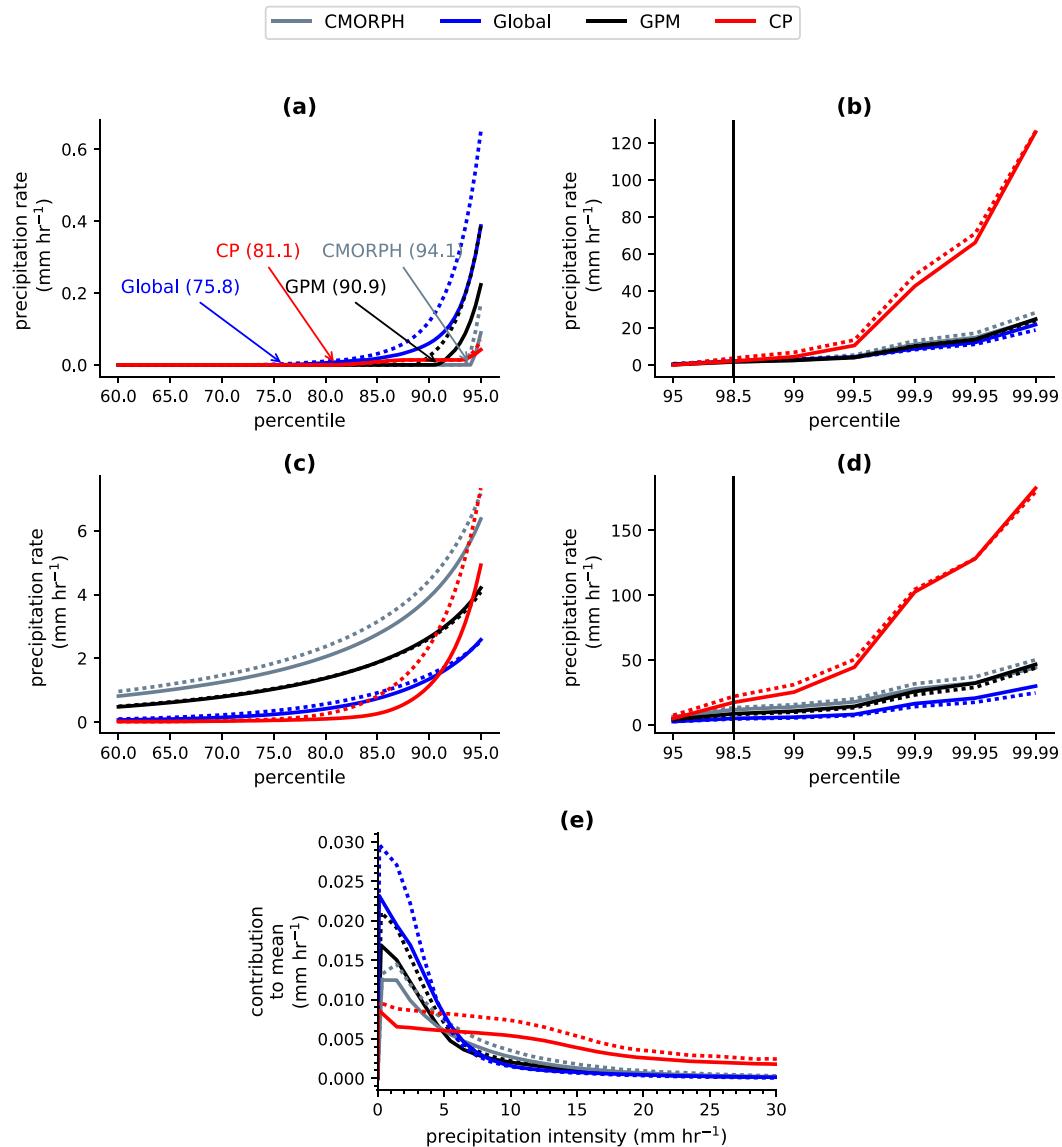


FIG. 4. (a)–(d) Precipitation rates corresponding to different percentile thresholds for GPM, CMORPH, and the global and CP models. (a),(b) The corresponding precipitation rates including points with precipitation rates of  $0 \text{ mm h}^{-1}$ ; (c),(d) neglect these points. (e) Contribution to the mean rainfall intensity, computed by binning rainfall rates into  $1 \text{ mm h}^{-1}$  bins and multiplying the average rainfall rate within the bin by the probability of the rainfall rate lying within that bin. The area under the curve is equal to the mean rainfall. For all panels, the solid lines show results for the full domain, and the dashed lines show results for the LV subdomain.

skill between the two models. For all three metrics, results are only plotted when a storm of the specified size and intensity was detected in the observations for at least 2.5% of the time steps (corresponding to over 500 storm events) in order to have a large sample size for more robust statistics. The sizes and intensities where this criterion was not met are hatched. The hit rate for both the global and CP models decreases as the minimum size and intensity of the storm increases. Overall, the CP model has a higher hit rate than the global model, and for the lowest intensities, it

predicts up to 20% more storms correctly. For a minimum storm size of 20 grid points (approximately  $12\,500 \text{ km}^2$ , 1/5 of the size of Lake Victoria) and rainfall above the 92nd percentile, the CP model has a hit rate of 57%, compared to 33% in the global model.

Similar plots of the false alarm ratios are shown in Figs. 5d–f. For both models, the false alarm ratio increases with increasing size and intensity. The global model has a smaller false alarm ratio for the majority of storm sizes and intensities, especially for small, low-intensity storms.

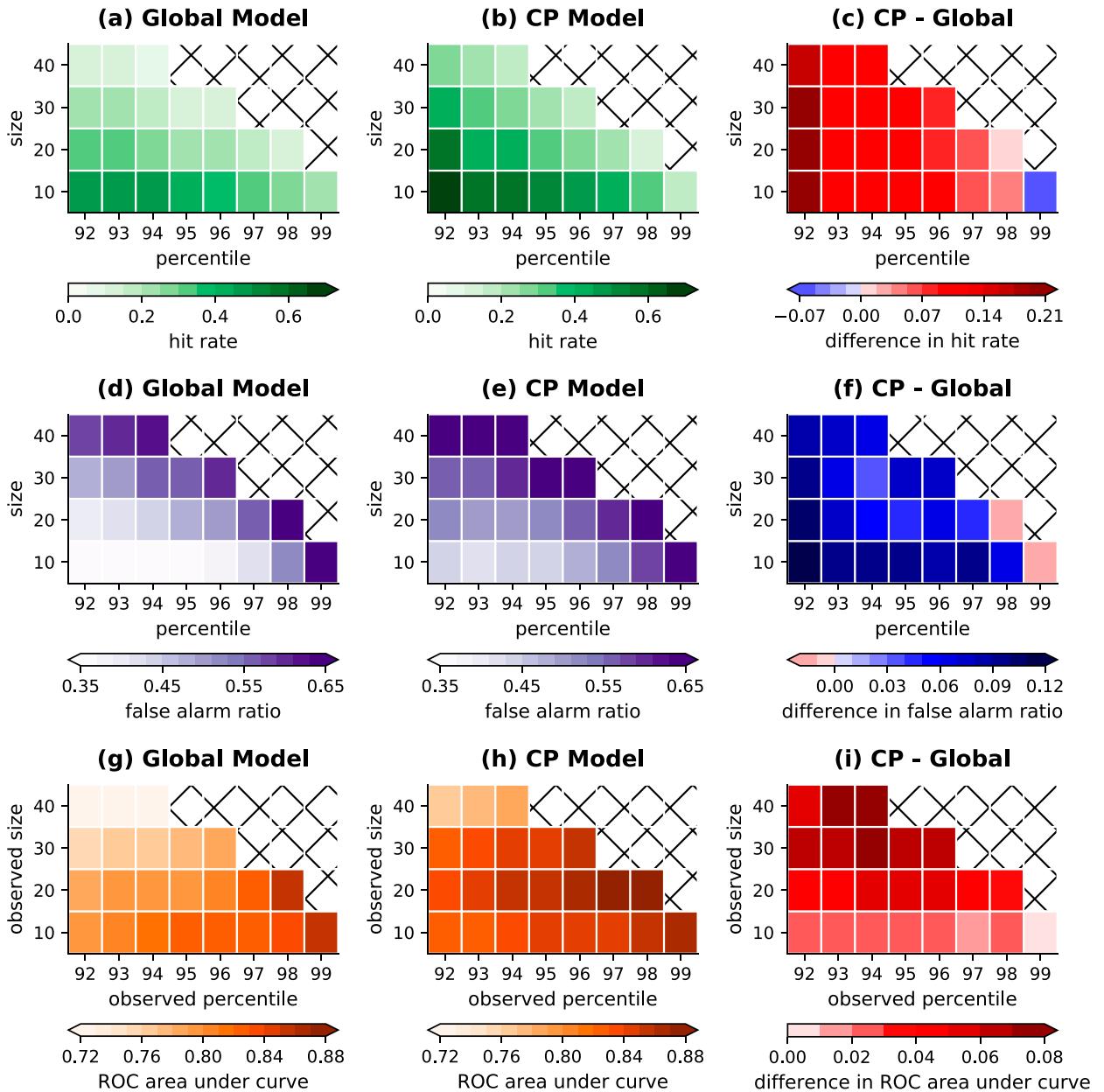


FIG. 5. (a)–(c) Hit rate, (d)–(f) false alarm ratio, and (g)–(i) ROC AUC for storms of a minimum intensity over a minimum size (number of grid points) over LV (lake grid points only) on 3-h intervals. The first column shows results for the global model, the second column for the CP model, and the third column is the difference between the CP and global models. In the difference plots, red shows where the CP model has increased skill relative to the global model. The hatched areas show where a storm of the given size or intensity was observed over less than 2.5% of the time steps.

For storms with a minimum size of 20 grid points and rainfall above the 92nd percentile, the global and CP models have false alarm rates of 51% and 41%, respectively.

The ROC AUC of both models generally decreases as the size of the observed storm increases (Figs. 5g,h). However, the AUC generally increases as the intensity of the observed storm increases. Figure 5i shows that for

all intensities and sizes, the CP model scores more highly than the global model, especially for large storms. The ROC AUC for storms with a minimum size of 20 grid points and rainfall above the 92nd percentile is 0.28 for the CP model and 0.22 for the global model. The greater skill of the CP model suggests that its greater hit rate outweighs the increased fraction of false alarms.

### e. FSS for 24-h accumulation rates

To verify the model over a long time period, the FSS was used. The rainfall threshold was chosen to be the 98.5th percentile for reasons discussed in the [appendix](#). Over the full domain (LV subdomain), the 98.5th percentile corresponded to 30.2 (32.0) mm day<sup>-1</sup> in the observations and, taking an average of the four combinations of forecast initializations and forecast lead time periods, 29.0 (27.6) mm day<sup>-1</sup> in the global model and 96.2 (109.3) mm day<sup>-1</sup> in the CP model.

Figures 6a and 6b show the mean FSS as a function of spatial scale for 24-h accumulation forecasts from the different models and a 24-h persistence forecast for the full domain and the LV subdomain, respectively. All three forecasts perform better than a random forecast over both domains, although not by much at the grid scale. Both models improve upon the persistence forecast at all spatial scales. Over the full domain, the two models show very similar skill, only reaching the target skill for spatial scales greater than almost 400 km. Over the LV subdomain, the CP model has increased skill, compared to the global model, reaching the target skill at a spatial scale of around 350 km, compared to just over 375 km for the global model.

#### 1) LOCALIZED FSS

Maps of LFSS, computed for 24-h accumulations at a spatial scale of 425 km ( $n = 17$ , because this is the nearest neighborhood size at which the models first become skillful over the whole domain, according to [Fig. 6](#)) and for a threshold of the 98.5th percentile, are shown in [Figs. 7a and 7b](#) for the global and CP models, respectively. The resulting spatial patterns are consistent across all spatial scales and for all percentile thresholds greater than the 95th percentile, although the magnitude of the score does change. Both models show enhanced skill over Lake Victoria and many of the regions of high orography, especially the Ethiopian highlands and the mountain ridges on either side of Lake Victoria. The global model shows particularly poor performance along the Somalian coastline, compared to the CP model. The green contour encloses regions where the mean rainfall is in the top 25% in both the observations and models. Although some of the regions of enhanced LFSS coincide with overlapping regions of the heaviest rainfall, this is not exclusive, suggesting that the model does not only perform well where rainfall is heaviest and most common. As shown in [Fig. 7c](#), the CP model generally shows greater skill over the land and lake than the global model, but the global model shows higher skill over the ocean. Since only a small proportion of the sea is included in the LV subdomain, this explains why the skill of the global

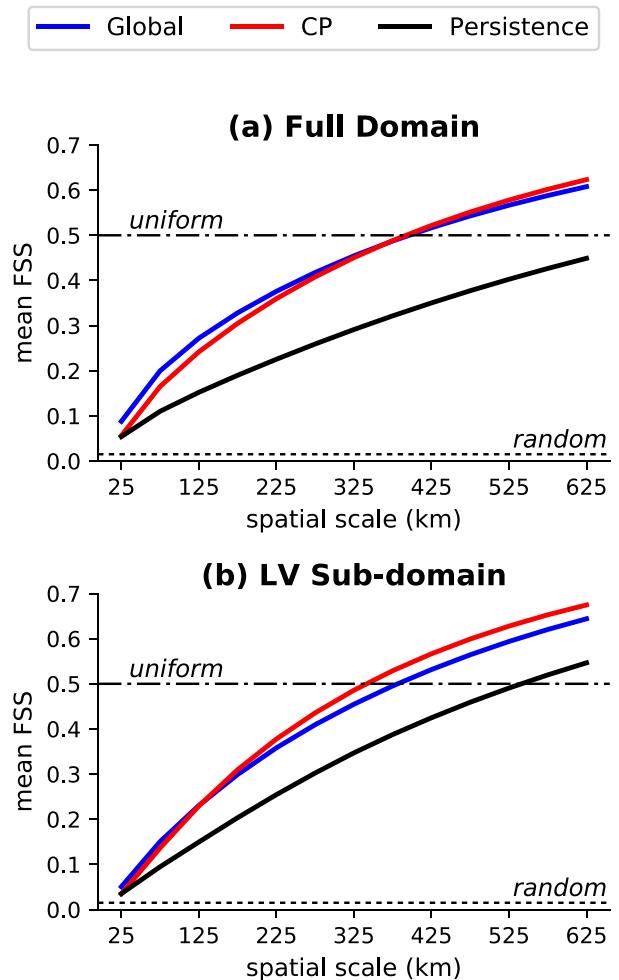


FIG. 6. Mean FSS as a function of neighborhood size for 24-h rainfall accumulations above the 98.5th percentile, shown for the global and CP models and a 24-h persistence forecast over (a) the full domain and (b) the LV subdomain.

model relative to the CP model drops over the LV subdomain ([Fig. 6](#)). Over land and on Lake Victoria are the most important places to forecast correctly, since these are where people live and work. The ubiquitously high scores over the ocean in the global model suggest that the LV subdomain should be used for further analysis to avoid contamination by ocean grid points. As such, for the remainder of the paper, only FSSs over the LV subdomain are presented.

#### 2) LFSS SEASONAL VARIABILITY

Figures 8 and 9 show how the LFSS, computed separately over the different seasons, differs from the LFSS computed over the whole time period for the global and CP models, respectively. Although quite noisy, the results are fairly similar for both models. This is likely because the annual cycle of convection is controlled by

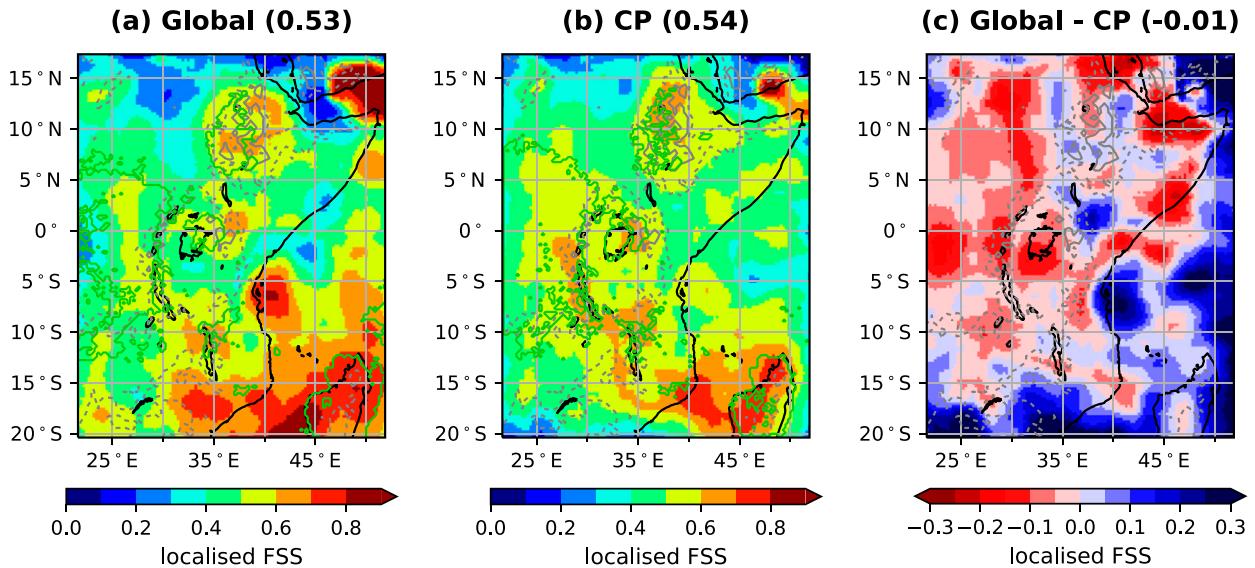


FIG. 7. Map of LFSS, computed for (a) the global and (b) the CP model, using a threshold of the 98.5th percentile and a neighborhood size of 425 km ( $n = 17$ ) for 24-h accumulations. (c) Map of the difference between the LFSS in the global model and CP model. The solid green contour encloses regions with mean rainfall in the top 25% in both the observations and the model. The gray dashed and solid contours mark orography of height 1 and 2 km, respectively. The numbers in brackets give the area average within the LV subdomain.

the intertropical convergence zone (ITCZ), the location of which should be similar in both models, since the global model fields are used to initiate the CP model.

During the rainy seasons (MAM and OND), when the ITCZ is located over equatorial East Africa, the variability in skill in both models is generally small around the Lake Victoria basin, with areas of both moderately increased and decreased skill. The skill over the ocean is decreased during the short rains (OND), but increased during the long rains (MAM). The dry seasons in the equatorial region correspond to when the ITCZ and main band of rainfall are to the north of the region in June–September (JJAS) and to the south in January–February (JF). For both models, there is broadly a large negative perturbation in skill over the opposite region of the domain to where the ITCZ sits. However, within the large areas of reduced skill, there are small areas of highly increased skill. Some of these are located in regions of high orography or over some of the small lakes or parts of the coastline. Within a contour enclosing the regions for which the top 25% of mean rainfall overlap in the model and observations, the mean LFSS perturbation is close to zero for all seasons and both models, showing that the model does not just perform well where rainfall is heaviest or most common.

*f. FSS for 3-hourly precipitation rates*

Figure 10 shows how the FSS for the global and CP models and a 24-h persistence forecast varies with spatial scale for the forecast rainfall rate at 3-h intervals

throughout the day. Similar to the analysis on 24-h accumulations, a threshold of the 98.5th percentile was chosen, corresponding to  $2.1 \text{ mm h}^{-1}$  in the observations and (taking an average of the four combinations of forecast initializations and forecast lead time periods)  $2.3 \text{ mm h}^{-1}$  in the global model and  $3.8 \text{ mm h}^{-1}$  in the CP model.

All three forecasts consistently beat the random forecast skill. The CP model outperforms the global model and persistence forecast at almost all spatial scales and at all times of day. The global model generally has greater skill than the persistence forecast, except at 1800 LT. Before 1200 LT, the target “uniform” skill is generally not achieved by any of the forecasts within spatial scales below 425 km. In particular, the skill of the CP model increases after 1200 LT and is very high at 1500 and 1800 LT, reaching the target skill at a spatial scale of approximately 275 km. These times of day correspond to the convective maximum over land.

Figure 11 shows how the skill of (Fig. 11a) the global model and (Fig. 11b) the CP model at different times of day varies with forecast lead time, using the 98.5th percentile as a threshold. Plots are only shown for a neighborhood size of 475 km ( $n = 19$ , the spatial scale at which the models have reached the target score for almost all times of day), but trends seen are broadly consistent for all spatial scales and percentiles greater than the 95th percentile. Each line corresponds to a different time of day, being formed of three points, since each time of day is forecast from three different initializations (the

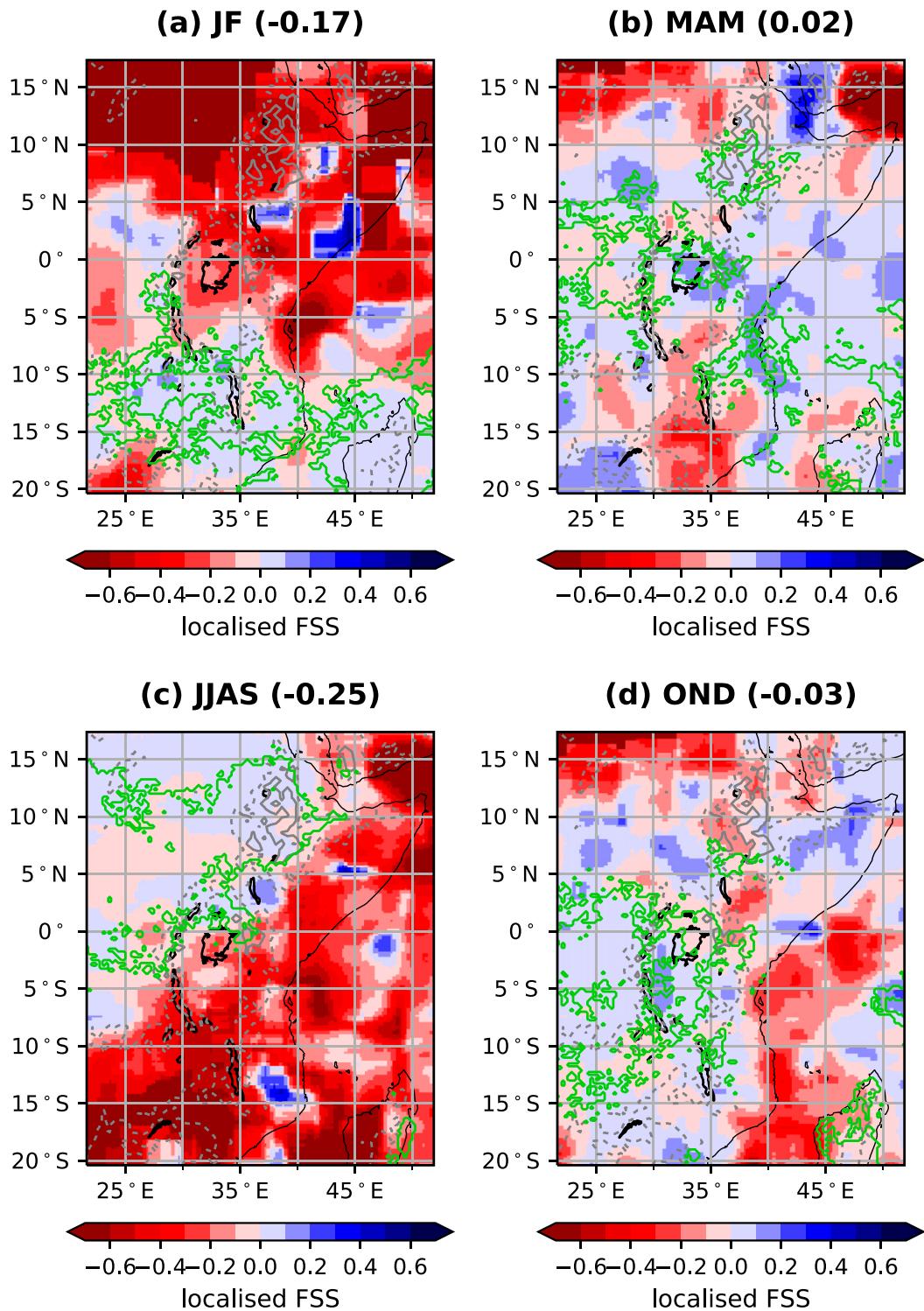


FIG. 8. Maps of perturbation in LFSS from Fig. 7a for different seasons, computed for the global model, using a threshold of the 98.5th percentile and a neighborhood size of 425 km ( $n = 17$ ) for 24-h accumulations. The solid green contour encloses regions with mean rainfall in the top 25% in both the observations and the model for the given season. The gray dashed and solid contours mark orography of height 1 and 2 km, respectively. The numbers in brackets give the area average within the LV subdomain.

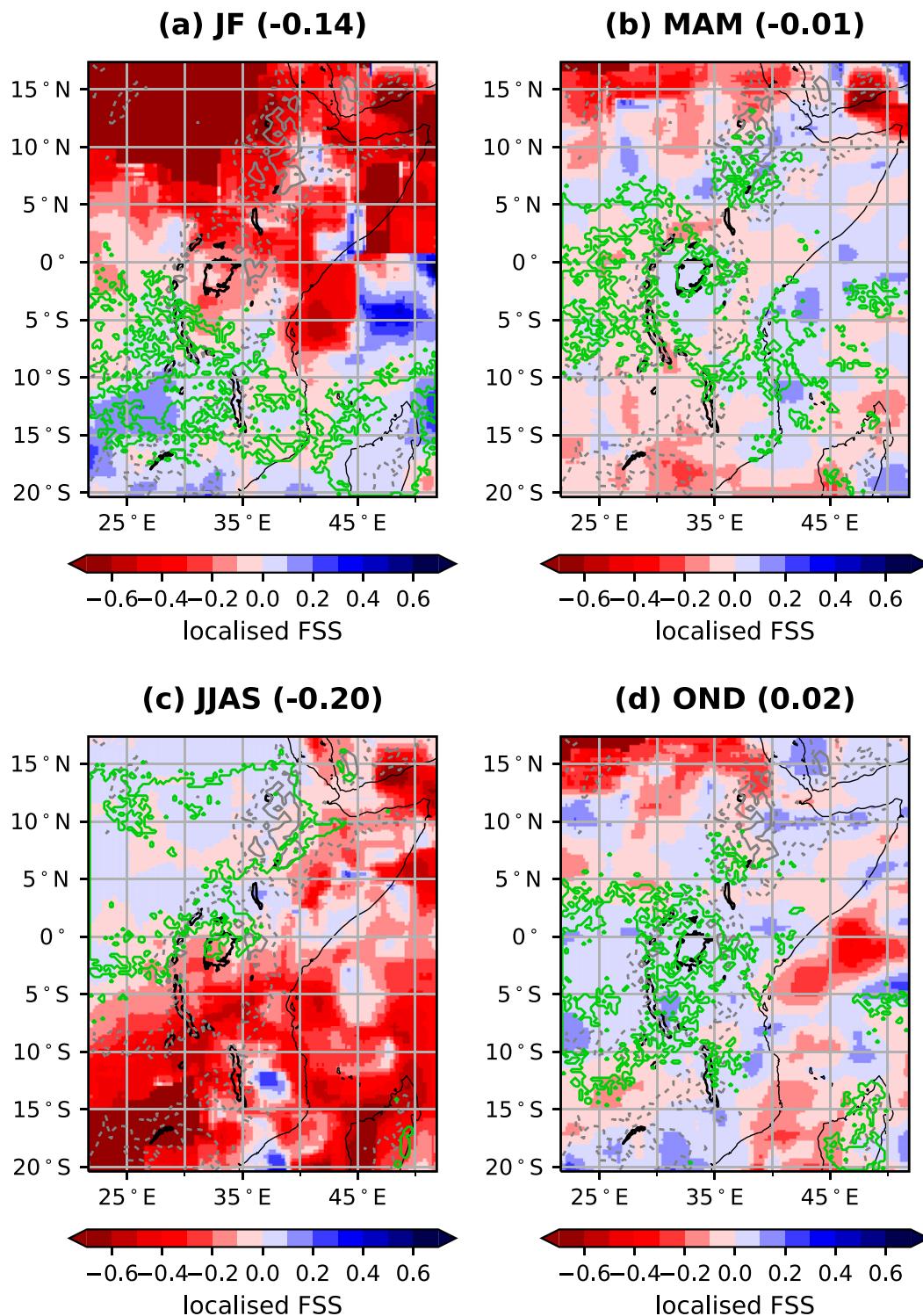


FIG. 9. As in Fig. 8, but computed for the CP model and therefore showing the perturbation in LFSS from Fig. 7b.

first 12 h of forecast are discarded due to spin up). Although there is small degradation in the skill of both models as forecast lead time increases, this is almost negligible, compared to the variation in skill between the

different times of day, especially in the CP model. These results contrast with those of Roberts (2008) for U.K. precipitation, who showed no dependence of the score on time of day. This is because rainfall in the midlatitudes is

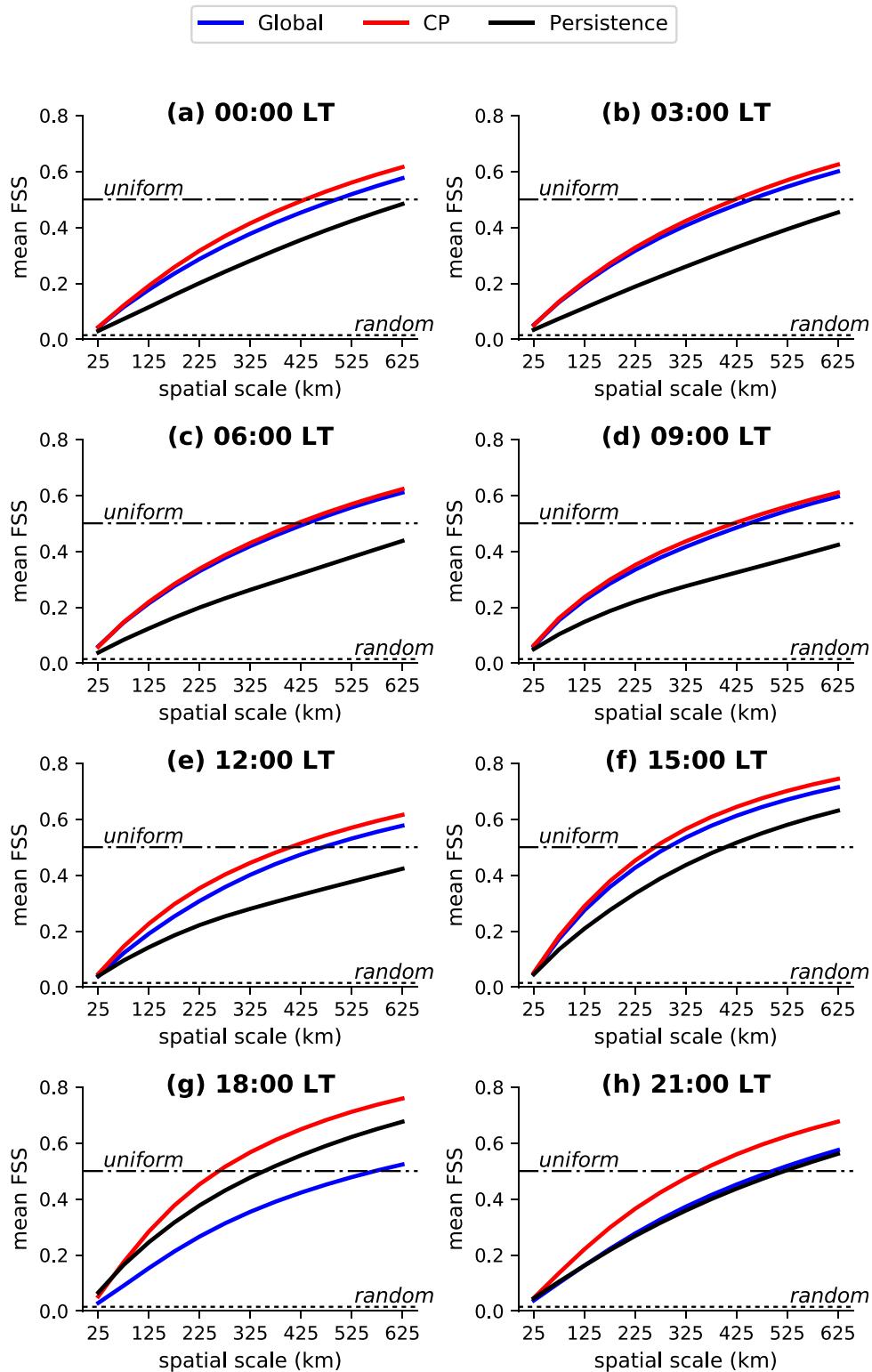


FIG. 10. Mean FSS over the LV subdomain as a function of neighborhood size for rainfall above the 98.5th percentile, computed at 3-h intervals (corresponding to the model diagnostic output times) for the global and CP models and a 24-h persistence forecast.

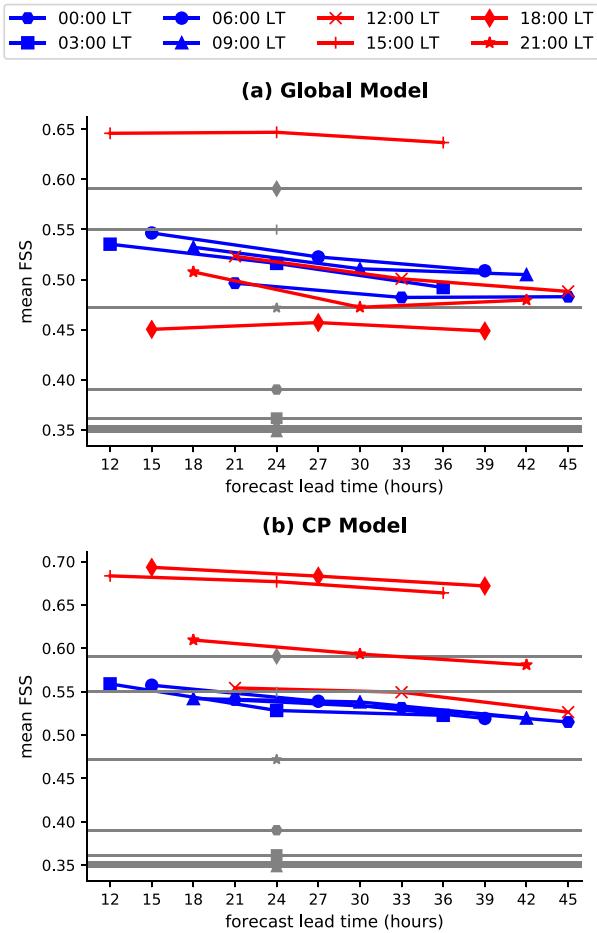


FIG. 11. Mean FSS as a function of forecast lead time, using the 98.5th percentile for (a) the global model and (b) the CP model over the LV subdomain. A neighborhood size of 475 km ( $n = 19$ ) is used, but results are consistent for all spatial scales. Each line corresponds to a different model diagnostic output time and is formed of three points (e.g., for 0000 LT, the first and last points are the  $T + 21$  and  $T + 45$  forecasts from 0000 UTC initializations and the middle point is the  $T + 33$  forecast from 1200 UTC initializations). The gray lines show the skill of a 24-h persistence forecast.

generally controlled by frontal systems, whereas convection in the tropics is strongly forced by the diurnal cycle.

Figure 11a shows that the skill of the global model is higher at 1500 LT by over 1/4, compared to other times of day. The CP model has increased skill at 1500 and 1800 LT by around 1/4 relative to most other times of day (Fig. 11b), and scores at 2100 LT are also elevated. These times correspond to the period when convection is at a maximum over land.

Figure 12a shows the probability distribution of the 98.5th percentile of precipitation rates in the GPM observations obtained from all time steps. Except between 1500 and 2100 LT, there is a high likelihood that the 98.5th percentile of the rainfall intensity is small.

Figures 12b and 12c show the mean FSS associated with a given 98.5th percentile of rainfall intensity for the global and CP models, respectively, for different times of day. Except at 1800 LT in the global model, the smallest intensities score the lowest FSS at all times of day in both models. The high skill for very low rainfall at 1800 LT in the global model is likely because the global model always forecasts such little rain at this time, so any time steps that occasionally do have little rainfall at 1800 LT will automatically score highly.

One reason for the increased skill in the global model at 1500 LT is that rainfall rates between 1 and 3.5 mm h<sup>-1</sup> are likely to have a greater FSS than at other times of day (Fig. 12b). Furthermore, the likelihood of the lowest rainfall rates (which generally have reduced scores) is decreased (Fig. 12a). Since this time lies between the midday convective maximum in the global model and the observed 1800 LT maximum, this increased skill is likely an artifact of storms over land nearing the end of their lives in the model coincident with growing storms in the observations.

Increased skill in the CP model between 1500 and 1800 LT occurs because the scores associated with intermediate rainfall rates have an increased mean FSS, compared to other times of day (Fig. 12c). In addition, the likelihood of the 98.5th percentile being below 0.5 mm h<sup>-1</sup> is greatly reduced. This is also true at 2100 LT, which explains the increased skill at this time of day. It is speculated that increased skill in the CP model coincident with the convective maximum over land is due to the model's ability to more accurately predict the location of storms at this time of day, not simply because there is more heavy rain and therefore a reduced number of low-scoring, low-intensity events.

#### 4. Discussion and conclusions

The precipitation forecast produced by a CP model, which was run operationally for East Africa by the Met Office in 2017, was verified over a 2-yr period from July 2014 to July 2016 using observations from GPM. Its performance was also compared to that of the operational global MetUM in order to understand the value added by a CP forecast. Verification was performed for the prediction of 24-h accumulations, as well as for rainfall rates at 3-hourly time intervals. Much of this verification used the fractions skill score, including a novel form of the skill score, termed the localized fractions skill score, to look at spatial variations in performance. Although extended assessments of CP models have been performed in the midlatitudes (Mittermaier et al. 2013), such an assessment has not been performed in the tropics before.

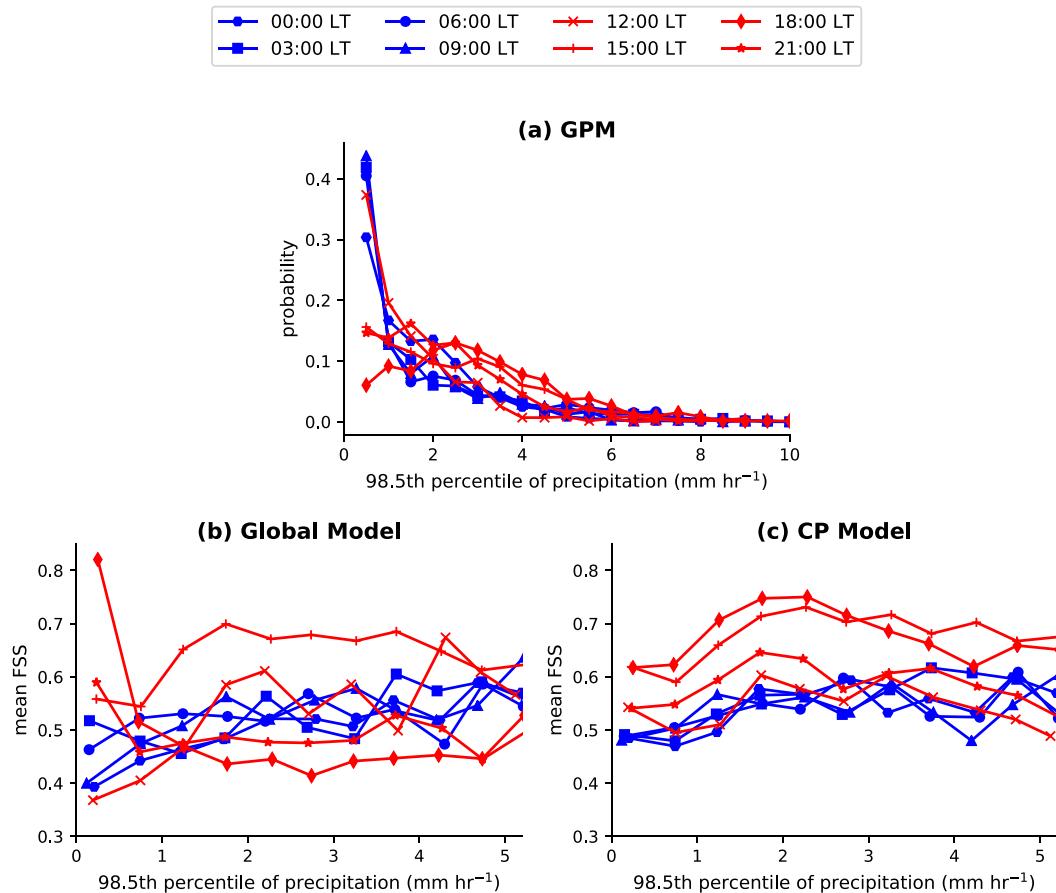


FIG. 12. (a) A probability distribution for the precipitation rate at the 98.5th percentile according to GPM observations. Mean FSS as a function of the 98.5th percentile of observed precipitation intensity across the LV subdomain for 3-h intervals (corresponding to the model diagnostic output times) for (b) the global model and (c) the CP model. The intensity is only plotted within an intensity band where there is a high-enough sample of data. A neighborhood size of 475 km ( $n = 19$ ) is used, with consistent results across all spatial scales.

The models show large biases in the distribution of rainfall rates, compared to the observations. In particular, both models are too wet, and the CP model has excessive rain rates. Therefore, percentile thresholds were chosen to define events in both the objective and FSS analysis instead of absolute intensities. Since there is uncertainty associated with the GPM observations, there is uncertainty about the “truth” against which the models are compared, but using a percentile threshold goes some way to reducing this problem.

The generally increased skill in the CP forecast shown over Lake Victoria is encouraging, since the primary reason for the existence of the model is to improve safety on the lake. Objective analysis shows that the CP model is better able to predict whether a storm will occur, compared to the global model. Despite a higher hit rate than the global model, the CP model does produce more false alarms, in agreement with Chamberlain et al.

(2014). Since the livelihoods of fishermen are dependent on their ability to go out on the lake, too many false alarms may lower trust in the forecast and thus reduce adherence to future warnings, even if issued correctly. Higher ROC AUC scores for the CP model suggest that the increased hit rate does outweigh the negative effect of increased false alarms, but the benefit of this does depend on the needs of the lake users.

Using the FSS, both the global and the CP model show greater skill than a 24-h persistence forecast, demonstrating that the models are valuable as forecasting tools. The CP model improves upon the forecast produced by the global model, especially on subdaily time scales; severe storms are more realistic, the diurnal cycle is improved, and this is reflected in generally elevated FSS for the CP model. Presently, forecasters in East Africa tend to use 24-h accumulation values produced by the models in their forecasts (Kenya Meteorological

Department; Tanzania Meteorological Agency; Uganda National Meteorological Authority, 2016, personal communications). Given that the CP model performs better on 3-hourly time intervals, the provision of forecasts with diurnal detail should be increased to provide more guidance to users of the lake as to exactly when it may be hazardous.

The CP model has the greatest benefit in the prediction of storms over land relative to the global model, but the global model performs better over the ocean. Birch et al. (2014a) show that convective initiation in explicit models usually corresponds to low-level convergence, whereas parameterized models tend to be unresponsive to convergence, and convection is generally triggered by instability. Regions of convergence are more likely over land than ocean due to the orography and influence of the lakes. Therefore, the increased scores over land may be due to the increased ability of the CP model to respond to convergence. This hypothesis is also supported by the fact that FSSs for the CP model peak at 1800 LT and are also elevated at 1500 and 2100 LT. These correspond to times of storm initiation over land, suggesting that the model is better able to forecast initiation, but is unable to capture storm propagation. The global model performs poorly along the Somalian coastline, suggesting that the model cannot capture the response of convection to the sea breeze (Birch et al. 2015), but the CP model shows improved performance here.

Despite the superior skill of the models—especially the CP model—relative to a persistence forecast, the level of skill is still fairly poor. Roberts and Lean (2008) suggest that an FSS of at least 0.5 should be the “target,” above which a model may be considered skillful. Figure 6 shows that neither the CP model nor the global model reaches this skill within a neighborhood size of over 350 km for 24-h accumulations. On subdaily time scales, the skillful target is reached by the CP model at a spatial scale of around 275–350 km, between 1500 and 2100 LT, but at other times of day, the target is not reached below approximately 425 km. This shows how heavier rain is easier to forecast, in the CP model in particular, when convection is most widespread. It should be noted that these results are only for the heaviest 1.5% of events, and for lower-percentile thresholds that represent lighter and more extensive rain, the spatial scale at which useful skill is reached is improved. Clark et al. (2016) note that the scales on which CP models have reasonable skill is large, compared to the grid spacing of the model, as well as the horizontal extent of convective rainfall events. Similar results are also shown by Mittermaier et al. (2013) for very high-intensity events in the UK4 model. Results such as these indicate that a CP model should not be interpreted deterministically; rather, an “ensemble” of the model should be run in order to obtain probabilistic forecasts (Clark et al. 2016). However, these are expensive and require greater computing power.

Another reason for such low skill, compared to CP models in the midlatitudes, is the lack of observations available for data assimilation in the tropics. Given the sensitivity of CP models to initial conditions (Birch et al. 2013; Guichard et al. 2010; Melhauser and Zhang 2012; Schumacher et al. 2013; Luo and Chen 2015; Vié et al. 2011), an increase in in situ observations, such as radiosondes and observations of lake surface temperature, has the potential to greatly increase the skill of the model.

Thiery et al. (2015), Argent et al. (2015), and Anyah and Semazzi (2004) showed the importance of accurate LSTs for the prediction of rainfall over Lake Victoria. Since LST observations are only available once per day, and the model contains no interactive lake model, the temperature of the lake will deviate from the truth throughout the duration of the simulation. Thiery et al. (2015) showed that a one-dimensional lake model—in this case, the Freshwater Lake (FLake) model (Mironov 2008; Mironov et al. 2010)—did add value to climate simulations using the COSMO model in climate mode. The FLake model has been used with MetUM (Rooney and Bornemann 2013) before; however, given that the diurnal cycle is fairly well simulated in the current model, the additional computational expense of running the lake model may outweigh the benefits. It would be interesting to see if an interactive lake model could shift the peak in precipitation over the lake slightly earlier to match the observations.

The forecast skill shows a much stronger dependence on the diurnal cycle than the forecast lead time, in contrast with similar studies performed in the midlatitudes. The results suggest that a good use of computer resources could be to initialize the model only once per day, instead of twice, and run out to a longer lead time. This would allow for more advance warnings to be given to users of the lake. However, further tests would need to be run to investigate model performance at lead times greater than 48 h. An immediate solution could be to consider forecasts from the multiple sequential initializations as an ensemble and maintain access to them on the VCP Africa Web Viewer, rather than older forecasts being rapidly replaced by newer ones.

The excessive rainfall rates for the highest-intensity events in the CP model are a cause for concern. This bias exists because semi-Lagrangian advection (as used in MetUM) does not conserve mass. Mass restoration schemes, such as Priestley (1993) and Zerroukat (2010), may easily be applied to global models, but they cannot directly be applied to limited-area models (LAMs) without knowledge of flux through the domain boundaries (Aranami et al. 2015). Aranami et al. (2015) developed a mass restoration scheme for LAMs, which reduces the excessive rainfall rates. This scheme is very

computationally expensive, but was followed by a less computationally intense scheme by [Zerroukat and Shipway \(2017\)](#), which requires no computation of lateral fluxes. This scheme is used in the latest CP MetUM configuration for the tropics, called RA1-T, which also includes further changes, such as tuning of the microphysical parameters to the tropics.

CP models are still fairly novel, and the way in which they must be used and interpreted is extremely different to global models. The uses and limitations of global models are well known, since these models have been used for many years. For a forecaster to feel comfortable using a CP model, time is required to adapt to the different interpretation and to learn its strengths and weaknesses through experience. In essence, a forecaster must learn to “trust” the model. This can be accelerated by more assessment of operational CP models and a greater understanding of when they provide increased skill over a global model forecast.

*Acknowledgments.* This work was supported by the NERC SPHERES DTP (Grant NE/L002574/1) and the HyCRISTAL project (Grant NE/M02038X/1). This work was also supported by U.K. Research and Innovation as part of the Global Challenges Research Fund, Grant NE/P021077/1 (GCRF African SWIFT). The GPM IMERG data were provided by the NASA Goddard Space Flight Center’s Precipitation Measurement Missions Science Team and Precipitation Processing System, which develop and compute the GPM IMERG as a contribution to GPM, including data archiving at the NASA GES DISC. The CMORPH data were provided by the NOAA Climate Prediction Center. The authors thank Humphrey Lean for his useful discussion and comments on the work, Stuart Webster for helpful discussions regarding CP MetUM configurations, and Peter Willets for help in processing CMORPH data. The authors would also like to thank forecasters at the Kenya Meteorological Department, Tanzania Meteorological Agency and Uganda National Meteorological Authority for their helpful discussions and Karen McCourt and Chris Tubbs for facilitating these meetings. The authors are grateful to the anonymous reviewers for improving the clarity and content of the paper.

## APPENDIX

### FSS Computational Details and Choice of Rainfall Threshold

As described in [section 2c\(2\)](#), the FSS relies on a rainfall threshold, above which an “event” is deemed to have occurred. However, [Fig. 4](#) shows that the rainfall

intensities produced by CP models and global models can be vastly different from one another and from observations. For many forecasting applications, the magnitude of the intensity may be bias corrected, and it is more important for the model to forecast whether or not a storm will occur. Therefore, in the computation of the FSS, the precipitation intensity corresponding to a given percentile is used instead of an absolute threshold. Consequently, the models and observations may have different minimum rainfall rates above which an event is defined. For the  $N$ th rainfall percentile, at each time step, the top  $(100 - N)\%$  of grid points (including dry grid points) in the model and observations are considered as “events.” If it is very dry at a time step in either the model or observations, such that there are not  $(100 - N)\%$  of grid points that receive rainfall, these fields are still included in the comparison. This means that the effective percentile threshold is increased for some time steps. The reason for doing this is to include comparisons in which the percentile threshold is just missed (not quite enough rain), but a meaningful comparison can still be made. If time steps with little rain were not included, the sample size would drop considerably because there are many times when the fractional rainfall coverage over the domain is very low.

An issue arises when either the model or observations is much drier than the other, and at least one has less than  $(100 - N)\%$  coverage, such that the ratio of the number of grid points being compared in the model and observations is not equal to one. This leads to a frequency bias, which can become very large when the ratio diverges from one. [Roberts and Lean \(2008\)](#) showed that a frequency bias can greatly reduce the FSS. Therefore, if the frequency bias is more than a factor of 3, the percentile is altered in the wetter dataset to keep the frequency bias within a factor of 3. Again, this means that the effective percentile threshold is increased (spatial coverage is reduced) for some times.

To ensure that only the highest-intensity events are considered, a further constraint is imposed, such that the rainfall must also be above the  $(100 - N)$ th percentile, computed using data from all time steps. This prevents the inclusion of light rainfall at time steps with light rain over large areas.

[Figure A1](#) shows how the mean FSS over the domain varies as a function of percentile threshold. A neighborhood size of 425 km ( $n = 17$ ) is used, but a similar pattern of decline in FSS as percentile threshold increases emerges at all spatial scales. At lower thresholds, the decline is gradual but becomes rapid at around the 95th percentile, because it is more difficult to get the correct positioning of more localized rain. In this study, a threshold of the 98.5th percentile is used to verify the performance of the models in predicting the most

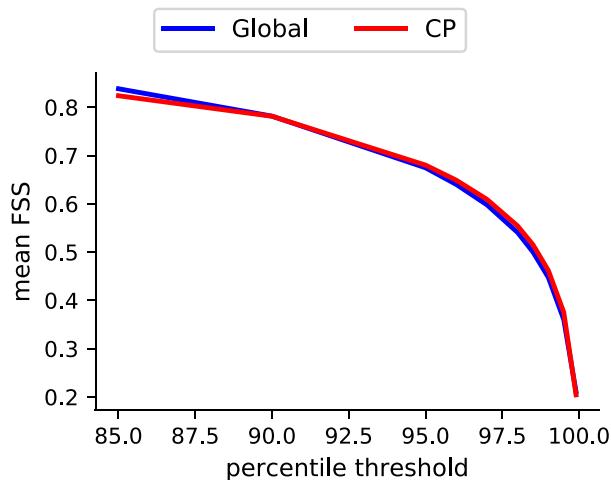


FIG. A1. Mean FSS as a function of percentile threshold for a neighborhood size of 425 km ( $n = 17$ ) for 24-h rainfall accumulations, shown for the global and CP models over the full domain.

extreme events. Although this percentile lies in the region of Fig. A1, in which there is a rapid decline in skill, a high percentile is necessary to focus on the more extreme events, while not so high that a meaningful measure of skill is lost. However, it should be noted that low skill scores are to be expected in comparison to similar studies over the United Kingdom by Roberts and Lean (2008), Roberts (2008), and Mittermaier (2012).

The time steps that do not have coverage of at least 1.5% and, therefore, have a higher effective percentile threshold will be at a disadvantage, given the rapid decline in average score for high-percentile thresholds shown in Fig. A1. To reduce this unfair effect on the FSS, only time steps with at least 0.25% coverage in both datasets are included.

#### REFERENCES

- Anyah, R., and F. Semazzi, 2004: Simulation of the sensitivity of Lake Victoria basin climate to lake surface temperatures. *Theor. Appl. Climatol.*, **79**, 55–69, <https://doi.org/10.1007/s00704-004-0057-4>.
- Aranami, K., T. Davies, and N. Wood, 2015: A mass restoration scheme for limited-area models with semi-Lagrangian advection. *Quart. J. Roy. Meteor. Soc.*, **141**, 1795–1803, <https://doi.org/10.1002/qj.2482>.
- Argent, R., X. Sun, F. Semazzi, L. Xie, and B. Liu, 2015: The development of a customization framework for the WRF Model over the Lake Victoria basin, eastern Africa on seasonal timescales. *Adv. Meteor.*, **2015**, 653473, <https://doi.org/10.1155/2015/653473>.
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Bechtold, P., J.-P. Chaboureaud, A. Beljaars, A. Betts, M. Köhler, M. Miller, and J.-L. Redelsperger, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. *Quart. J. Roy. Meteor. Soc.*, **130**, 3119–3137, <https://doi.org/10.1256/qj.03.103>.
- Birch, C. E., D. J. Parker, A. O’Leary, J. H. Marsham, C. M. Taylor, P. Harris, and G. Lister, 2013: Impact of soil moisture and convectively generated waves on the initiation of a West African mesoscale convective system. *Quart. J. Roy. Meteor. Soc.*, **139**, 1712–1730, <https://doi.org/10.1002/qj.2062>.
- , J. H. Marsham, D. J. Parker, and C. M. Taylor, 2014a: The scale dependence and structure of convergence fields preceding the initiation of deep convection. *Geophys. Res. Lett.*, **41**, 4769–4776, <https://doi.org/10.1002/2014GL060493>.
- , D. J. Parker, J. H. Marsham, D. Copsey, and L. Garcia-Carreras, 2014b: A seamless assessment of the role of convection in the water cycle of the West African Monsoon. *J. Geophys. Res. Atmos.*, **119**, 2890–2912, <https://doi.org/10.1002/2013JD020887>.
- , M. J. Roberts, L. Garcia-Carreras, D. Ackerley, M. J. Reeder, A. P. Lock, and R. Schiemann, 2015: Sea-breeze dynamics and convection initiation: The influence of convective parameterization in weather and climate model biases. *J. Climate*, **28**, 8093–8108, <https://doi.org/10.1175/JCLI-D-14-00850.1>.
- Chamberlain, J., C. Bain, D. Boyd, K. McCourt, T. Butcher, and S. Palmer, 2014: Forecasting storms over Lake Victoria using a high resolution model. *Meteor. Appl.*, **21**, 419–430, <https://doi.org/10.1002/met.1403>.
- Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, <https://doi.org/10.1002/met.1538>.
- Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, <https://doi.org/10.1175/JCLI3884.1>.
- Davies, T., M. J. Cullen, A. J. Malcolm, M. Mawson, A. Staniforth, A. White, and N. Wood, 2005: A new dynamical core for the Met Office’s global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759–1782, <https://doi.org/10.1256/qj.04.101>.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) Model. *Atmos. Sci. Lett.*, **5**, 110–117, <https://doi.org/10.1002/asl.72>.
- Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158, <https://doi.org/10.1016/j.rse.2010.10.017>.
- Eagle, C., H. Lean, and S. Webster, 2015: Convective scale NWP model configurations for the Lake Victoria region. Met Office Tech. Rep., 27 pp.
- Fiedler, E. K., M. J. Martin, and J. Roberts-Jones, 2014: An operational analysis of lake surface water temperature. *Tellus*, **66**, 21247, <https://doi.org/10.3402/tellusa.v66.21247>.
- Garcia-Carreras, L., and Coauthors, 2013: The impact of convective cold pool outflows on model biases in the Sahara. *Geophys. Res. Lett.*, **40**, 1647–1652, <https://doi.org/10.1002/grl.50239>.
- Gregory, D., and P. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483–1506, [https://doi.org/10.1175/1520-0493\(1990\)118<1483:AMFCSW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2).

- Guichard, F., and Coauthors, 2010: An intercomparison of simulated rainfall and evapotranspiration associated with a meso-scale convective system over West Africa. *Wea. Forecasting*, **25**, 37–60, <https://doi.org/10.1175/2009WAF2222250.1>.
- Habib, E., A. T. Haile, Y. Tian, and R. J. Joyce, 2012: Evaluation of the high-resolution CMORPH satellite rainfall product using dense rain gauge observations and radar-based estimates. *J. Hydrometeorol.*, **13**, 1784–1798, <https://doi.org/10.1175/JHM-D-12-017.1>.
- Haile, A. T., E. Habib, and T. Rientjes, 2013: Evaluation of the Climate Prediction Center (CPC) morphing technique (CMORPH) rainfall product on hourly time scales over the source of the Blue Nile River. *Hydrol. Processes*, **27**, 1829–1839, <https://doi.org/10.1002/hyp.9330>.
- Hirpa, F. A., M. Gebremichael, and T. Hopson, 2010: Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *J. Appl. Meteor. Climatol.*, **49**, 1044–1051, <https://doi.org/10.1175/2009JAMC2298.1>.
- Hohenegger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1794, <https://doi.org/10.1175/BAMS-88-11-1783>.
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteor. Climatol.*, **43**, 1834–1853, <https://doi.org/10.1175/JAM2173.1>.
- Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Huffman, G. J., 2017: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree  $\times$  0.1 degree V05. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 20 December 2017, <https://doi.org/10.5067/GPM/IMERG/3B-HH/05>.
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- , and Coauthors, 2018: NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). NASA Tech. Rep., 35 pp., [https://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V5.2.pdf](https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2.pdf).
- Joyce, R. J., and P. Xie, 2011: Kalman filter-based CMORPH. *J. Hydrometeorol.*, **12**, 1547–1563, <https://doi.org/10.1175/JHM-D-11-022.1>.
- , J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. *J. Climate*, **25**, 5791–5806, <https://doi.org/10.1175/JCLI-D-11-00562.1>.
- Kim, K., J. Park, J. Baik, and M. Choi, 2017: Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmos. Res.*, **187**, 95–105, <https://doi.org/10.1016/j.atmosres.2016.12.007>.
- Kumar, B., K. C. Patra, and V. Lakshmi, 2016: Daily rainfall statistics of TRMM and CMORPH: A case for trans-boundary Gandak River basin. *J. Earth Syst. Sci.*, **125**, 919–934, <https://doi.org/10.1007/s12040-016-0710-1>.
- Lafore, J. P., and Coauthors, 2017: West African synthetic analysis and forecast. *Meteorology of Tropical West Africa: The Forecasters' Handbook*, D. Parker and M. Diop-Kane, Eds., Wiley-Blackwell, 423–451.
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, **136**, 3408–3424, <https://doi.org/10.1175/2008MWR2332.1>.
- Liu, Z., D. Ostrenga, W. Teng, and S. Kempler, 2012: Tropical Rainfall Measuring Mission (TRMM) precipitation data and services for research and applications. *Bull. Amer. Meteor. Soc.*, **93**, 1317–1325, <https://doi.org/10.1175/BAMS-D-11-00152.1>.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Luo, Y., and Y. Chen, 2015: Investigation of the predictability and physical mechanisms of an extreme-rainfall-producing meso-scale convective system along the Meiyu front in East China: An ensemble approach. *J. Geophys. Res. Atmos.*, **120**, 10 593–10 618, <https://doi.org/10.1002/2015JD023584>.
- MacCallum, S., and C. Merchant, 2011: ARC-lake: Data product description. University of Edinburgh Tech. Rep., [http://www.geos.ed.ac.uk/arclake/ARCLake\\_DPD\\_v1\\_v1\\_1](http://www.geos.ed.ac.uk/arclake/ARCLake_DPD_v1_v1_1).
- Marshall, J. H., N. S. Dixon, L. Garcia-Carreras, G. Lister, D. J. Parker, P. Knippertz, and C. E. Birch, 2013: The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations. *Geophys. Res. Lett.*, **40**, 1843–1849, <https://doi.org/10.1002/grl.50347>.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30** (4), 291–303.
- Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, <https://doi.org/10.1256/003590002320603584>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock, 2001: Development of a next-generation regional Weather Research and Forecast Model. *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*, World Scientific, 269–276.
- Mironov, D. V., 2008: Parameterization of lakes in numerical weather prediction. Part 1: Description of a lake model. Tech. rep., Consortium for Small-Scale Modelling Tech. Rep. 11, 47 pp., <http://www.cosmo-model.org/content/model/documentation/techReports/docs/techReport11.pdf>.
- , E. Heise, E. Kourzeneva, B. Ritter, N. Schneider, and A. Terzhevik, 2010: Implementation of the lake parameterisation scheme Flake into the numerical weather prediction model COSMO. *Boreal Environ. Res.*, **15**, 218–230, <http://www.borenv.net/BER/pdfs/ber15/ber15-218.pdf>.
- Mittermaier, M., 2012: Verification. Met Office Tech. Rep. 17.6, MOSAC-17, Met Office, 10 pp., [https://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/f/3/mosac\\_17.6\\_mittermaier.pdf](https://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/f/3/mosac_17.6_mittermaier.pdf).
- , N. Roberts, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteor. Appl.*, **20**, 176–186, <https://doi.org/10.1002/met.296>.
- Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–582, [https://doi.org/10.1175/1520-0493\(1989\)117<0572:SSACCI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2).

- NCEP, 2017: CMORPH bias corrected V1.0. NCEP, accessed 24 November 2017, [http://ftp.cpc.ncep.noaa.gov/precip/CMORPH\\_V1.0/CRT/8km-30min/](http://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/8km-30min/).
- Nesbitt, S. W., D. J. Gochis, and T. J. Lang, 2008: The diurnal cycle of clouds and precipitation along the Sierra Madre Occidental observed during NAME-2004: Implications for warm season precipitation estimation in complex terrain. *J. Hydrometeorol.*, **9**, 728–743, <https://doi.org/10.1175/2008JHM939.1>.
- O, S., and P.-E. Kirstetter, 2018: Evaluation of diurnal variation of GPM IMERG-derived summer precipitation over the contiguous US using MRMS data. *Quart. J. Roy. Meteor. Soc.*, <https://doi.org/10.1002/qj.3218>, in press.
- , U. Foelsche, G. Kirchengast, J. Fuchsberger, J. Tan, and W. A. Petersen, 2017: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrol. Earth Syst. Sci.*, **21**, 6559–6572, <https://doi.org/10.5194/hess-21-6559-2017>.
- Prakash, S., A. K. Mitra, A. AghaKouchak, Z. Liu, H. Norouzi, and D. Pai, 2018: A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.*, **556**, 865–876, <https://doi.org/10.1016/j.jhydrol.2016.01.029>.
- Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, <https://doi.org/10.1002/2014RG000475>.
- Priestley, A., 1993: A quasi-conservative version of the semi-Lagrangian advection scheme. *Mon. Wea. Rev.*, **121**, 621–629, [https://doi.org/10.1175/1520-0493\(1993\)121<0621:AQCVOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<0621:AQCVOT>2.0.CO;2).
- Roberts, N. M., 2003: The impact of a change to the use of the convection scheme to high resolution simulations of convective events (stage 2 report from the storm scale numerical modelling project). Met Office Tech. Rep. 407, 30 pp.
- , 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, <https://doi.org/10.1002/met.57>.
- , and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Romilly, T. G., and M. Gebremichael, 2011: Evaluation of satellite rainfall estimates over Ethiopian river basins. *Hydrol. Earth Syst. Sci.*, **15**, 1505–1514, <https://doi.org/10.5194/hess-15-1505-2011>.
- Rooney, G. G., and F. J. Bornemann, 2013: The performance of Flake in the Met Office Unified Model. *Tellus*, **65A**, 21363, <https://doi.org/10.3402/tellusa.v65i0.21363>.
- Saito, K., and Coauthors, 2006: The operational JMA non-hydrostatic mesoscale model. *Mon. Wea. Rev.*, **134**, 1266–1298, <https://doi.org/10.1175/MWR3120.1>.
- Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf, 2008: Global precipitation analysis products of the GPCP. Tech. rep., Deutscher Wetterdienst Tech. Rep., 12 pp., [http://www.mapcruzin.com/environmental-shapefile-maps/water/precipitation/GPCP\\_intro\\_products\\_2008.pdf](http://www.mapcruzin.com/environmental-shapefile-maps/water/precipitation/GPCP_intro_products_2008.pdf).
- Schumacher, R. S., A. J. Clark, M. Xue, and F. Kong, 2013: Factors influencing the development and maintenance of nocturnal heavy-rain-producing convective systems in a storm-scale ensemble. *Mon. Wea. Rev.*, **141**, 2778–2801, <https://doi.org/10.1175/MWR-D-12-00239.1>.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, <https://doi.org/10.1175/2010MWR3425.1>.
- Sharifi, E., R. Steinacker, and B. Saghaian, 2016: Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: Preliminary results. *Remote Sens.*, **8**, 135, <https://doi.org/10.3390/rs8020135>.
- Skok, G., 2015: Analysis of fraction skill score properties for a displaced rainband in a rectangular domain. *Meteor. Appl.*, **22**, 477–484, <https://doi.org/10.1002/met.1478>.
- , and N. Roberts, 2016: Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, <https://doi.org/10.1002/qj.2849>.
- Song, Y., F. H. Semazzi, L. Xie, and L. J. Ogallo, 2004: A coupled regional climate model for the Lake Victoria basin of East Africa. *Int. J. Climatol.*, **24**, 57–75, <https://doi.org/10.1002/joc.983>.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSSE>2.3.CO;2).
- Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, <https://doi.org/10.1029/2010JD014532>.
- Stratton, R. A., and Coauthors, 2018: A pan-African convection-permitting regional climate simulation with the Met Office Unified Model: CP4-Africa. *J. Climate*, **31**, 3485–3508, <https://doi.org/10.1175/JCLI-D-17-0503.1>.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, <https://doi.org/10.1175/JCLI3672.1>.
- Swets, J. A., 1973: The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, **182**, 990–1000, <https://doi.org/10.1126/science.182.4116.990>.
- Tang, G., Y. Ma, D. Long, L. Zhong, and Y. Hong, 2016a: Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over mainland China at multiple spatiotemporal scales. *J. Hydrol.*, **533**, 152–167, <https://doi.org/10.1016/j.jhydrol.2015.12.008>.
- , Z. Zeng, D. Long, X. Guo, B. Yong, W. Zhang, and Y. Hong, 2016b: Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7? *J. Hydrometeorol.*, **17**, 121–137, <https://doi.org/10.1175/JHM-D-15-0059.1>.
- Tang, Y., H. W. Lean, and J. Bornemann, 2013: The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteor. Appl.*, **20**, 417–426, <https://doi.org/10.1002/met.1300>.
- Thiery, W., E. L. Davin, H.-J. Panitz, M. Demuzere, S. Lhermitte, and N. van Lipzig, 2015: The impact of the African Great Lakes on the regional climate. *J. Climate*, **28**, 4061–4085, <https://doi.org/10.1175/JCLI-D-14-00565.1>.
- , —, S. I. Seneviratne, K. Bedka, S. Lhermitte, and N. P. van Lipzig, 2016: Hazardous thunderstorm intensification over Lake Victoria. *Nat. Commun.*, **7**, 12786, <https://doi.org/10.1038/ncomms12786>.
- , L. Gudmundsson, K. Bedka, F. H. Semazzi, S. Lhermitte, P. Willems, N. P. van Lipzig, and S. I. Seneviratne, 2017: Early warnings of hazardous thunderstorms over Lake Victoria. *Environ. Res. Lett.*, **12**, 074 012, <https://doi.org/10.1088/1748-9326/aa7521>.

- TRMM, 2011: TRMM (TMPA) rainfall estimate L3 3 hour 0.25 degree  $\times$  0.25 degree V7. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 16 November 2016, <https://doi.org/10.5067/TRMM/TMPA/3H7>.
- Vié, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, **139**, 403–423, <https://doi.org/10.1175/2010MWR3487.1>.
- Wang, R., J. Chen, and X. Wang, 2017a: Comparison of IMERG Level-3 and TMPA 3B42V7 in estimating typhoon-related heavy rain. *Water*, **9**, 276, <https://doi.org/10.3390/w9040276>.
- Wang, Z., R. Zhong, C. Lai, and J. Chen, 2017b: Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmos. Res.*, **196**, 151–163, <https://doi.org/10.1016/j.atmosres.2017.06.020>.
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model. *Wea. Forecasting*, **23**, 407–437, <https://doi.org/10.1175/2007WAF2007005.1>.
- Wessel, P., and W. H. Smith, 1996: A global, self-consistent, hierarchical, high-resolution shoreline. *J. Geophys. Res.*, **101**, 8741–8743, <https://doi.org/10.1029/96JB00104>.
- Weusthoff, T., F. Ament, M. Arpagaus, and M. W. Rotach, 2010: Assessing the benefits of convection-permitting models by neighborhood verification: Examples from MAP D-PHASE. *Mon. Wea. Rev.*, **138**, 3418–3433, <https://doi.org/10.1175/2010MWR3380.1>.
- White, B., A. Buchanan, C. Birch, P. Stier, and K. Pearson, 2018: Quantifying the effects of horizontal grid length and parameterized convection on the degree of convective organization using a metric of the potential for convective interaction. *J. Atmos. Sci.*, **75**, 425–450, <https://doi.org/10.1175/JAS-D-16-0307.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140**, 1505–1520, <https://doi.org/10.1002/qj.2235>.
- World Bank DGF, 2011: Towards a regional framework for weather and climate services for food aid, food security, maritime transport safety contributing to disaster risk reduction in Lake Victoria region. World Bank Development Grant Facility Tech. Rep., 17 pp., [https://www.wmo.int/pages/prog/www/DPFS/Meetings/RAI-EA-SWFDP-RSMT\\_Nairobi2011/documents/DGF-Lake%20Victoria-scoping.pdf](https://www.wmo.int/pages/prog/www/DPFS/Meetings/RAI-EA-SWFDP-RSMT_Nairobi2011/documents/DGF-Lake%20Victoria-scoping.pdf).
- Xu, R., F. Tian, L. Yang, H. Hu, H. Lu, and A. Hou, 2017: Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. *J. Geophys. Res. Atmos.*, **122**, 910–924, <https://doi.org/10.1002/2016JD025418>.
- Yang, G.-Y., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801, [https://doi.org/10.1175/1520-0493\(2001\)129<0784:TDCITT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2).
- Zerroukat, M., 2010: A simple mass conserving semi-Lagrangian scheme for transport problems. *J. Comput. Phys.*, **229**, 9011–9019, <https://doi.org/10.1016/j.jcp.2010.08.017>.
- , and B. Shipway, 2017: ZLF (Zero Lateral Flux): A simple mass conservation method for semi-Lagrangian-based limited-area models. *Quart. J. Roy. Meteor. Soc.*, **143**, 2578–2584, <https://doi.org/10.1002/qj.3108>.
- Zeweldi, D. A., and M. Gebremichael, 2009: Evaluation of CMORPH precipitation products at fine space–time scales. *J. Hydrometeorol.*, **10**, 300–307, <https://doi.org/10.1175/2008JHM1041.1>.