Climate Change Impacts on Extreme Rainfall in Eastern Africa in a Convection-Permitting Climate Model

SARAH CHAPMAN,[a] JAMES BACON,[b] CATHRYN E. BIRCH,[a] EDWARD POPE,[b] JOHN H. MARSHAM,[a] HELLEN MSEMO,[a,c] EDSON NKONDE,[e] KENNETH SINACHIKUPO,[e] AND CHARLES VANYA[d]

[a] *School of Earth and Environment, University of Leeds, Leeds, West Yorkshire, United Kingdom*
[b] *Met Office Hadley Centre, Exeter, Devon, United Kingdom*
[c] *Tanzania Meteorological Authority, Dar es Salaam, Tanzania*
[d] *Department of Climate Change and Meteorological Service, Ministry of Forestry and Natural Resources, Blantyre, Malawi*
[e] *Zambia Meteorological Department, Lusaka, Zambia*

(Manuscript received 2 November 2021, in final form 25 August 2022)

ABSTRACT: Climate change is expected to increase the frequency and intensity of rainfall extremes. Understanding future changes in rainfall is necessary for adaptation planning. Eastern Africa is vulnerable to rainfall extremes because of low adaptive capacity and high future population growth. Convection-permitting climate models have been found to better represent moderate (yearly) rainfall extremes than parameterized convection models, but there is limited analysis of rare extremes that occur less frequently than once per year. These events often have the largest socioeconomic impacts. We use extreme value theory and regional frequency analysis to quantify rare rainfall extremes over East Africa in a convection-permitting climate model (CP4A). We compare the results with its parameterized counterpart (P25), the Coordinated Regional Climate Downscaling Experiment for the African region (CORDEX-Africa) ensemble, and observations to understand how the convection parameterization impacts the results. We find that CP4A better matches observations than the parameterized models. With climate change, we find the parameterized convection models have unrealistically high changes in the shape parameter of the extreme value distribution, which controls the tail behavior (i.e., the most extreme events), leading to large increases in return levels of events with a return period of >20 years. This suggests that parameterized convection models may not be suitable for looking at relative changes in rare rainfall events with climate change and that convection-permitting models should be preferred for this type of work. With the more realistic CP4A, RCP8.5 end-of-century climate change leads to 1-in-100-yr events becoming 1-in-23-yr events, which will necessitate serious adaptation efforts to avoid devastating socioeconomic impacts.

SIGNIFICANCE STATEMENT: We use a new, high-resolution climate model to examine how rare extreme rainfall events in East Africa might change in the future with climate change and compare the results with those from standard-resolution climate models. We find that the standard-resolution models have unrealistically large increases in rainfall for events that occur less frequently than every 20 years. The high-resolution model is more realistic and is required to illustrate possible future changes in rare rainfall extremes. Extreme events will become more common with climate change, and in the more realistic model we show that a 1-in-100-yr event may become a 1-in-23-yr event by the end of the century if greenhouse gas emissions are not significantly reduced.

KEYWORDS: Convection; Extreme events; Climate change

1. Introduction

Extreme rainfall can lead to floods, loss of water quality, disease, and damage to property and agriculture as well as loss of life (Tarhule 2005; de Paola et al. 2014; Chamani et al. 2018). In the present climate, extreme rainfall regularly affects large numbers of people in eastern Africa; on average, 1002000 people are impacted each year in Malawi (World Bank 2019), while in 2019 between October and December extreme rainfall caused floods and landslides in eastern Africa, affecting an estimated 2.8 million people (Wainwright et al. 2020). While there is large uncertainty in the sign and magnitude of climate change impacts on mean rainfall in eastern Africa (Rowell and Chadwick 2018; Kendon et al. 2019; Chapman et al. 2020), the intensity of rainfall is expected to increase (Trenberth et al. 2003; Cioffi et l. 2016; Han et al. 2019; Kendon et al. 2019; Finney et al. 2020; Onyutha 2020), suggesting a likely increase in extremes and associated impacts.

Understanding present-day and future rainfall extremes is essential for quantifying and managing risks to lives and livelihoods. The characteristics of extreme rainfall can be studied using a variety of indices, such as the frequency of wet days or percentile-based indices that represent moderate extremes, which occur on a yearly basis (i.e., 90th percentile of daily rainfall), or by looking at rare extremes with return periods of several years or more (Wang et al. 2013). Globally, moderate extremes have been more frequently studied than rare extremes

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0851.s1.

Corresponding author: Sarah Chapman, earsch@leeds.ac.uk

DOI: 10.1175/JCLI-D-21-0851.1

due the difficulties studying rare events in the relatively short observational record (Alexander 2016). However, rare extremes that occur roughly once per 10–500 years are more relevant for infrastructure design than moderate, yearly extremes (Jalbert et al. 2017; Kharin et al. 2018; Do Lago et al. 2019; Wright et al. 2019) and may also have more severe impacts on lives and livelihoods.

The focus of this paper is therefore on rare rainfall extremes, which occur less frequently than once a year. Rare extremes can be studied using extreme value theory (EVT). EVT provides a method for extrapolating beyond the bounds of the available data, whether they are observations or simulations, to estimate the return levels (i.e., amount of rainfall) of events rarer than those observed (Stephenson and Tawn 2004). EVT has previously been applied to parameterized convection climate models (e.g., Kharin et al. 2007; Hanel et al. 2009; Overeem et al. 2009; Knote et al. 2010) including in eastern Africa (de Paola et al. 2014; Garcia-Aristizabal et al. 2015; Cioffi et al. 2016). However, convection parameterized models are known to produce rainfall that is too frequent and light, and insufficient heavy rain (Prein et al. 2015; Kendon et al. 2017). A pan-Africa convection-permitting climate model simulation (CP4A) has been found to better represent rainfall occurrence, intensity, and moderate extremes than its parameterized counterpart (Stratton et al. 2018; Finney et al. 2019; Kendon et al. 2019; Senior et al. 2021). This is due to improved physics that allow better modeling of storm life cycles and mesoscale convective systems and improved coupling between land–sea breezes and convective activity, among other things (Senior et al. 2021). Numerous challenges still remain with convection-permitting model (Senior et al. 2021); however, despite this, CP4A still represents a step-change improvement over its parameterized counterpart, and the differences in physics between it and its parameterized counterpart also lead to different climate change impacts, the causes of which are likely to be relevant for all parameterized models (Senior et al. 2021).

CP4A has not been analyzed for rare extremes, however, it has been found that moderate extremes tend to intensify with climate change more in CP4A than in its parameterized counterpart, in part due to intensification of updrafts, which are underestimated in the parameterized model counterpart (Berthou et al. 2019; Kendon et al. 2019; Finney et al. 2020; Jackson et al. 2020; Misiani et al. 2020). Given that CP4A's explicit representation of convection improves the representation of moderate extremes and affects associated climate change impacts, it may also affect rare extremes. Therefore, we use this convection-permitting model to analyze rare rainfall extremes in eastern Africa and compare the results with its parameterized counterpart and the Coordinated Regional Climate Downscaling Experiment for the African region (CORDEX-Africa) ensemble of regional climate models.

There are few published applications of EVT using convection-permitting models, and those that exist have found that during the season of convective activity parameterized models may overestimate the increase in rainfall with climate change for very rare events in midlatitude regions (Chan et al. 2014; Ban et al. 2020). Previous work has only compared one convection-permitting model with one parameterized model and has only examined European climates. Here, we examine the end-of-century climate change impact (RCP8.5) on rare extremes in all available CORDEX-Africa models, in addition to CP4A and its parameterized counterpart, in order to make our results relevant to parameterized models in general. Further, our work focuses on the tropics, an area where convection-permitting and parameterized models have not previously been compared for rare extremes.

2. Methods

a. Study area

We focus on the East Africa region from 19°S to 0° and from 19° to 42°E (Fig. 1). In parts of northern Tanzania, the rainy season is bimodal (October–December and March–May). Farther south, in Zambia and Malawi, the wet season is from November to April.

b. Models

This work uses the CORDEX-Africa atmosphere-only regional climate models (RCMS; Jones et al. 2011), and a pair of RCM simulations: one convection-permitting (CP4A) and one with parameterized convection (P25; Stratton et al. 2018; Kendon et al. 2019). CP4A and P25 are atmosphere-only simulations and cover the pan-Africa region with a horizontal grid spacing at the equator of 4.5 km × 4.5 km and 26 km × 39 km, respectively (Stratton et al. 2018). Further details on CP4A and P25 are available in Stratton et al. (2018) and Kendon et al. (2019).

The CORDEX-Africa model data are given at 0.44° × 0.44° horizontal resolution (approximately 44 km × 44 km at the equator) and the multimodel ensemble includes 6 RCMs with 11 different GCMs providing initial and boundary driving conditions. The matrix of GCM–RCM combinations is presented in Table 1.

For all models, we compare the "historical" period and the "business-as-usual" end-of-century RCP8.5 scenario. RCP8.5 was selected because it has a strong climate change signal relative to natural climate variability and is the only scenario for which CP4A and P25 simulations are available.

To determine the impact of time series length on return levels, we examine 10-, 20-, and 30-yr time periods for the CORDEX data, covering 1989–99, 1979–99, and 1969–99 for the historical scenario, and 2089–99, 2079–99, and 2069–99 for the RCP8.5 climate change scenario, hereinafter referred to as COR10, COR20, and COR30. CP4A and P25 only have 10 years of data available, from 1997 to 2006 and 2097 to 2106 (see Table 2). While CP4A and P25 cover only a small portion of the time periods of CORDEX, there is a 100-yr difference between the future and historical periods for all sets of models, and so it is reasonable to compare them.

All analysis was carried out on daily rainfall data from the rainy season (October–May), which for CP4A and P25 means only having nine seasons of data. Discarding January–September of 1997 also allows for spinup.

c. Observations

We used the Climate Hazards Group Infrared Precipitation with Station Data (CHIRPS), version 2.0, and Tropical Rainfall

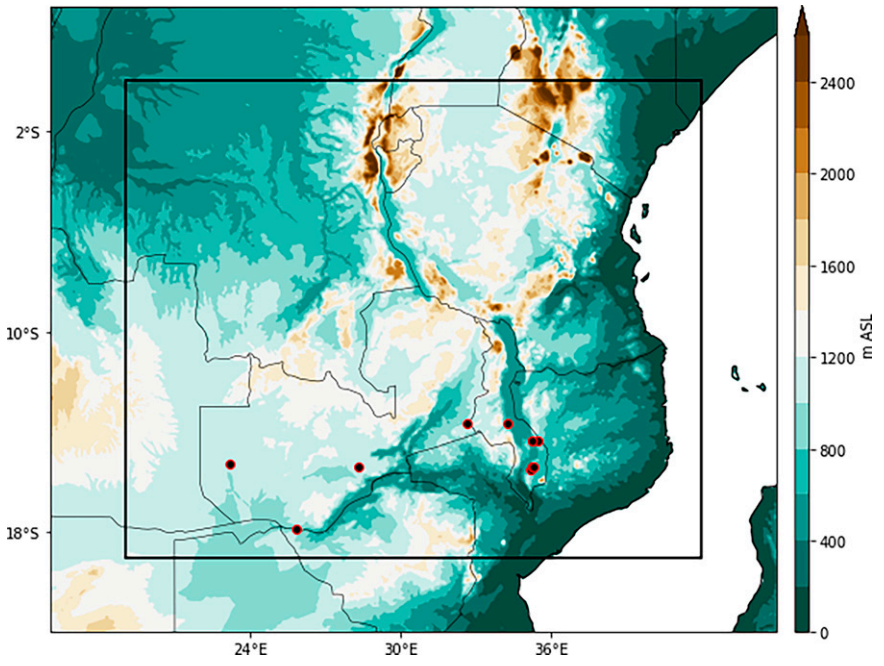

FIG. 1. The study area (black-outlined box). Shading shows the topography, as represented in CP4A (m). Locations of weather stations used are shown as black dots with red outline. The mean average elevations for the stations are 719 m for Malawi and 1089 m for Zambia.

Measuring Mission (TRMM) 3B42 satellite data as a comparison for the model data. The TRMM data combine satellite observations (microwave and infrared) and rain gauges and is available from 1998 to 2019; however, we have only used data up to 2014 as at this point the input data and calibration for the 3B42 dataset changed due to the end of the TRMM satellite mission, which may have introduced inhomogeneities into the dataset (Huffman et al. 2007; Huffman 2019). We used the daily TRMM data at a resolution of 0.25° × 0.25°.

CHIRPS v2.0 is a daily rainfall dataset based on infrared satellite observations and station data and is available at a 0.05° × 0.05° resolution, from 1981 to the present (Funk et al. 2015). Prior to analysis, the CHIRPS data were regridded to the TRMM grid for the purposes of comparison.

The time periods examined for each dataset are shown in Table 2. Multiple 10-yr time series are examined for CHIRPS to ensure overlap with models and with TRMM. The TRMM 10- and 15-yr time slices will hereinafter be referred to as TRMM10 and TRMM15, whereas the CHIRPS 15-, 20-, 30-, and 38-yr time slices will be referred to as CHIRPS15, CHIRPS20, CHIRPS30, and CHIRPS38, respectively. We have not used the GPM satellite rainfall data because of the time period covered.

Other studies indicate that both TRMM and CHIRPS agree well with station data in Africa when it comes to mean rainfall, though both products have difficulty in mountains (Cattani et al. 2016; Kimani et al. 2017; Dinku et al. 2018; Monsieurs et al. 2019; Muthoni et al. 2019). However, there is

TABLE 1. The GCM–RCM combinations that are available for CORDEX Africa for historical and future (RCP8.5) scenarios.

	RCM					
GCM	SMHI-RCA4	CLMcom-CCLM4–8-17	MPI-CSC or GERICS REMO2009	KNMI-RACMO22T	DMI-HIRHAM5	CCCma-CanRCM4
HadGEM2-ES	X	X	X	X		
EC-EARTH	X	X	X	X	X	
MPI-ESM-LR	X	X	X			
CNRM-CM5	X	X				
MIROC5	X		X			
CSIRO Mk3.6.0	X					
IPSL-CM5A-MR	X					
IPSL-CM5A-LR			X			
CanESM2	X					X
GFDL-ESM2M	X					
NorESM1-M	X					

TABLE 2. Historical time periods used for each model and observational dataset. Only data for the rainy season (October–May) were used for analysis. Note that, for row 1 (10 years), CP4A/P25 only has data for 9 full rainy seasons, rather than 10.

Years	CORDEX	CP4A/P25	TRMM	CHIRPS
10 years	—	1997–2006	—	1997–2006
	—	—	1998–2008	1998–2008
	1989–99	—	—	1989–99
15 years	—	—	1998–2014	1998–2014
20 years	1979–99	—	—	—
	—		—	1981–2001
30 years	1969–99	—	—	—
	—	—	—	1981–2011
38 years	—	—	—	1981–2019

more disagreement among the satellite products for extreme rainfall (Sylla et al. 2013; Shiferaw et al. 2018; Timmermans et al. 2019), so in this study we use two observational gridded datasets rather than relying on only one. It is for this reason that we also perform EVT analysis on individual rain gauge data from stations within the study region (see Fig. 1 for station locations), which allows us to compare how results differ between the gridbox scale and the point location scale. Extremes at point locations are expected to be systematically higher than gridbox aggregates (e.g., Mannshardt-Shamseldin et al. 2010), and knowledge of how rainfall extremes spatially downscale in eastern Africa would be required to combine EVT with climate projections in real-world applications, such as the design of climate resilient infrastructure projects. Station data for the southern half of the study region were privately shared with us by the Department of Climate Change and Meteorological Services (DCCMS) and the Zambia Meteorological Department (ZMD) for Malawi (5 stations) and Zambia (4 stations), respectively. The mean length of the station rain gauge time series is 73 years in Malawi, 38 years in Zambia, and 59 years across all stations.

d. Analysis

EVT shows that extremes can be described by one of three related types of statistical distribution, Gumbel, Fréchet, or Weibull, which can be expressed as the generalized extreme value distribution (GEVD) for block maxima (i.e., annual maxima), or generalized pareto distribution (GPD) for threshold exceedances (Coles 2001; Gilleland and Katz 2006). The fitted GEVD or GPD gives the estimated relationship between the return period and return value, that is, the average time interval between events of a given magnitude (Knote et al. 2010).

We examine return levels of extreme rainfall events using the block-maxima approach (BM). Using BM, the GEVD [see Eq. (1)] is fitted to the maximum daily rainy season rainfall, using maximum likelihood estimation (MLE). A drawback of the BM approach is that it does not retain as much data as the peaks-over-threshold (POT) approach; however, an advantage is that it does not require a threshold to define extremes, unlike POT. Selecting a threshold can be a subjective and time-consuming process, particularly for a large ensemble of climate models (Scarrott and MacDonald 2012). With a poor threshold choice, the GPD will not provide as good an estimate of the distribution as a GEV (Gilleland and Katz 2006; Knote et al. 2010).

The GEVD is defined by the shape, scale, and location parameters (Coles 2001). The location parameter is a measure of central tendency of the GEVD, the scale parameter is a measure of variance, and the shape parameter defines the tail behavior and is analogous to skewness:

$$F(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (1)$$

defined on x: $1 + \xi(x - \mu)/\sigma > 0$, where μ, ξ, and σ are the location, shape, and scale parameters, respectively.

The behavior of the GEVD is strongly determined by the shape parameter ξ. If $\xi < 0$, the distribution follows the Weibull distribution and is bounded, meaning that the extremes tend asymptotically to a maximum limit. In contrast, if $\xi > 0$ the Fréchet distribution is obtained, and the distribution has no upper limit. The Gumbel distribution is obtained in the limit that $\xi \to 0$ (Gilleland and Katz 2006). The Gumbel distribution gives a light tail, and the probability of obtaining extremely high values decays exponentially. By comparison, the Fréchet distribution is heavy tailed, and the probability of obtaining extremely high values decays polynomially so that higher values are obtained with greater probability than would be obtained with a light tail (Gilleland and Katz 2006). Accurately estimating the shape parameter is, therefore, important for obtaining physically realistic values of rare extremes.

e. Regionalization

A drawback of using CP4A and P25 is that only a short time series (i.e., 10 years) is available. Larger sample sizes (or longer time series) of extremes tend to offer more accurate estimates of the parameters of the GEVD/GPD because the variance of an estimator decreases as the sample size increases (Papalexiou and Koutsoyiannis 2013). The length of the time series influences the estimates of all the parameters of the GEVD, but the shape parameter is of particular note as it controls the tail behavior of the GEVD and the return levels of events with a high return period. With shorter time series, the uncertainty in the estimate of the shape parameter is higher, which increases the uncertainty in estimates of events with a higher return period. Therefore, for example, with 30 years of data we might reasonably say something about a 1-in-50-yr event, but beyond that the uncertainty is so large it is not useful to say something about a 1-in-200-yr event. Likewise, with a 10-yr time series we might be able to say something about a 1-in-20-yr event, but not a 1-in-50- or 1-in-100-yr event.

Given the short time series for CP4A and P25 we use regional frequency analysis (RFA) and "trade space for time" (Hosking and Wallis 1997). RFA assumes that the scale and shape parameters of the GEVD are fixed by region, and that only the location parameter varies spatially within a region. Therefore, prior to estimating the GEVD for the spatially

pooled data we split our study area into regions with similar extreme rainfall characteristics using cluster analysis. Pooling data across regions with similar rainfall characteristics should increase the accuracy of parameter estimates and allow a shorter time series to be used to estimate the GEVD (Overeem et al. 2009). However, selecting regions is also a fairly subjective step, and there is no consensus in the literature about the best method (Badr et al. 2015). Other methods are available that involve spatial pooling, such Bayesian hierarchical analysis, which does not require the subjective step of defining regions; however, as far as we are aware this method has not been applied to a time series as short as 9 years, whereas RFA has (Overeem et al. 2009; Jalbert et al. 2017). For RFA, Hosking and Wallis (1997) recommend clustering methods that give clusters of roughly similar size, such as Ward's, as preferable to methods that give very large or very small clusters. We used Ward's minimum variance method, which minimizes the total within cluster variance by minimizing the sum of squares of the difference between clusters and creates compact clusters. This approach split the data into 30 regions based on a time series of maximum daily rainfall in each rainy season in the historical period (see Table 1), applied using the R HiClimR software package (R Core Team 2013; Badr et al. 2015). We used the longest time series for each model and each set of satellite observations to determine regions. We used maximum daily rainy season rainfall as we are looking for regions with homogenous extreme rainfall characteristics. Because of the number and size of the regions, they do not entirely blend into the pattern of local topography; however, they do distinguish between the eastern and southern African climate regimes, coastal regions, and humid and semiarid inland regions.

An assumption of previous applications of RFA is that 10 years of daily data are sufficient with spatial pooling to calculate return levels of rare extremes, in some cases up to 1-in-100-yr events (Groupe de Recherche en Hydrologie Statistique 1996). To test the idea that 10 years of spatially pooled data are adequate for EVT, we also use the CORDEX-Africa ensemble of regional climate models to examine the impact of 30, 20, and 10 years of data on the estimates of return-levels, and the associated confidence intervals.

f. *Estimation of return levels*

We applied maximum likelihood estimation (MLE) to each region to estimate the GEVD. We have selected MLE over other popular methods, such as L-moments, because while L-moments lead to less biased estimates of the GEV parameters than MLE for small to moderate sample sizes (Martins and Stedinger 2000), Overeem et al. (2009) show that, for a very small sample size such as our 9-yr sample for CP4A and P25, MLE may be less biased than L-moments.

We first applied MLE to pooled data for each region and estimated the shape and scale parameters. We then reapplied MLE to each grid cell, holding the scale and shape parameters constant and allowing only the location parameter to vary. This method is also known as the index-flood method in hydrology and assumes that variables within a homogenous region are identically distributed after scaling with a site-specific factor, the index flood (Hanel et al. 2009). The index flood is estimated using measures of central tendency, such as the mean, median or trimmed mean (Hosking and Wallis 1997). For estimating the GEVD, the location parameter, which is a measure of central tendency for the GEVD, is used as the index flood (Hanel et al. 2009). Biases will be introduced into the analysis if the assumption of homogeneity within regions is not met; however, simulation studies have shown that RFA with moderate amounts of heterogeneity can still be more accurate than analysis using information only from individual sites or grid cells (Hosking and Wallis 1997; Hanel et al. 2009). The assumption that the shape parameter is constant within regions is reasonable, given research has found that it is constant over large areas (i.e., small countries), though may vary in mountainous regions (Ragulina and Reitan 2017). The scale parameter also appears to vary slowly over space, though as far as we are aware there is less specific research on this than on spatial variation in the shape parameter (Acero et al. 2011; Golroudbary et al. 2016).

To test goodness of fit of our annual maximums to the GEVD we used the Anderson–Darling test. The null hypothesis is that the data are drawn from a GEVD. The majority of the seasonal maximum rainfall for individual grid cells fit the GEVD for both models and observations. On average, 90% of grid cells fit the GEVD for the observations and the 10-yr time series from the climate models. For each model within COR20 and COR30, on average 89% and 87% of grid cells respectively fit the GEVD. For the rain gauge data, all stations fit the GEVD.

An assumption of EVT is that the data are stationary. We tested for stationarity using the augmented Dickey–Fuller test and confirmed that in all models and all regions that the daily rainy season rainfall data were stationary at $p < 0.05$. This is not surprising as observational datasets show no significant trend for extreme rainfall indices for the study area between 1983 and 2013 (Harrison et al. 2019).

g. *Bootstrap samples*

MLE provides confidence intervals for parameter estimates; however, MLE assumes the data are spatially independent, which is not the case for climate data, and so these confidence intervals may be underestimated (Overeem et al. 2009). Instead, we have used bootstrapping (40 samples) with replacement to create additional samples drawn from the original data and used these to calculate confidence intervals (Knote et al. 2010). To maintain the same spatial dependence in the bootstrap sample as in the original data, we use all grid cells for each bootstrap sample, rather than select individual grid cells (Hanel et al. 2009). We obtained bootstrap samples as follows:

- Draw a random sample with replacement from the series of year numbers, that is, 1997 . . . 2006 to obtain a set of years with the same time series length for each model.
- Select the rainfall maxima from all grid cells for the sampled year numbers. This leads to a sample of the same length (i.e., 10 maxima for COR10 models, 20 maxima for COR20 models, etc.), and the same number of grid cells as the original sample.

TABLE 3. Average location (mm day^{-1}), scale (mm day^{-1}), and shape parameters for TRMM and CHIRPS data. The standard deviation across the study area is shown in parentheses. See Table 1 for information on years covered by each set of observations. One outlier region was removed from the CHIRPS (TRMM years) set (see the text for details).

	10 years				15 years				
Parameter	TRMM	CHIRPS (TRMM years)	CHIRPS (CORDEX years)	CHIRPS (CP4 years)	TRMM	CHIRPS	20 years CHIRPS	30 years CHIRPS	38 years CHIRPS
Location	61.7 (12.02)	33.56 (9.21)	33.93 (8.86)	33.01 (9.17)	58.8 (10.96)	32.90 (8.6)	34.30 (7.36)	33.76 (7.89)	33.76 (8.10)
Scale	17.7 (4.03)	8.80 (5.46)	9.53 (5.65)	8.60 (5.25)	17.74 (4.06)	8.62 (5.15)	9.86 (4.32)	9.48 (4.66)	9.49 (4.96)
Shape	0.05 (0.04)	0.06 (0.09)	0.10 (0.08)	0.04 (0.08)	0.04 (0.03)	0.06 (0.08)	0.14 (0.10)	0.14 (0.10)	0.14 (0.09)

- Repeat this 40 times for each model, scenario and time period.
- Apply MLE to each region to estimate the scale and shape parameters for each bootstrap sample, and then apply MLE to each grid cell to estimate the location parameter, as with the original sample.

Testing showed that 40 bootstrap samples are a sufficient number to explore sampling uncertainty (see the online supplemental material; bootstrap resampling). We analyzed 40 bootstrap samples for COR20 and COR30, TRMM15, CHIRPS15, and CHIRPS 38. We then calculated confidence intervals for each individual model.

h. Comparison with station point data

We compare the gridcell GEVD parameter and return-level estimates with point-location (station) estimates to explore how daily rainfall extremes may spatially downscale in eastern Africa. We do not apply RFA methods to the station data because the data are spatially uneven, and site-specific factors that affect rain gauge readings are likely to be more prevalent than at the gridcell scale (Sieck et al. 2007). The length of the station time series is also not as limiting for EVT as it is for some of the gridded products (Table 2). We subset the rain gauge time series data from 1981 to 2019, making the data comparable to the longest gridded dataset (38 years in CHIRPS; see Table 2). We otherwise prepare the station data using the same method described above and use MLE to estimate the GEVD parameters from the block maxima of daily rainfall at each station. We compare the station-derived parameter and return-level estimates with estimates from corresponding grid cells in TRMM and CHIRPS to test for a spatial scaling factor between daily rainfall extremes in our study region.

We also compare parameter estimates from the stations with the corresponding estimates from the climate models. While we do not expect the models to represent extremes true to the observations, it has been argued that the modeled change in extremes may still be robust (Fowler et al. 2010). Under this assumption, we apply a delta change (DC) to the station parameter estimates:

$$\theta_{\text{station future}} = \theta_{\text{station historical}} + \theta_{\text{DC}}, \tag{2}$$

where θ represents a GEV parameter estimate (either location, scale, or shape), and θ_{DC} is the change in the corresponding parameter from the climate model, defined as

$$\theta_{\text{DC}} = \theta_{\text{climate model future}} - \theta_{\text{climate model historical}}. \tag{3}$$

We adapt this method from (Fontolan et al. 2019). The delta change method is essentially a form of bias correction and statistical downscaling, which gives a more realistic indication of future change in local-scale daily rainfall extremes than using the model alone We calculate confidence intervals for future return levels at each of the stations by bootstrap sampling from the adjusted GEVDs in the extRemes R package (Gilleland and Katz 2006), using 200 bootstrap samples for each return level.

3. Results

a. Parameter estimates

The mean GEVD parameters from the satellite and station observations are shown in Tables 3 and 4. The choice of satellite product has a much larger impact on the parameters than the length of the time series (Table 3). One of the 30 regions has been removed from the CHIRPS TRMM years set as it was an extreme outlier and had a negative location parameter. This may be due to an unrealistic rainfall event included in this time period, or the regions not fitting well for this time period. The shape parameter increases for CHIRPS with increasing time series length, though relative to the standard deviation this increase is small. The increasing shape parameter may be due to capturing rarer extremes with a longer time series and is in line with other observational work (e.g., Papalexiou and Koutsoyiannis 2013). However, the increasing shape parameter with time series length may also be due to the 10–15-yr time series missing the 1981–96 period, during which there were strong El Niño events, 1982/83 and 1987/88 (Climate Prediction Center 2020), which may be dominating the extremes for CHIRPS.

In comparing the GEVD parameters from TRMM, CHIRPs, and CP4A with the station data, we find that the parameters estimated TRMM grid cells at the station locations are largely in agreement with the corresponding station-derived estimates (Table 4). In contrast, we find that CHIRPS grid cells have small location and scale parameters and large shape parameters relative to TRMM and the station data. For the historical period of the CP4A climate model, the location and scale parameters are also comparable to station and TRMM estimates, although the shape parameters are generally larger. There is

TABLE 4. Average location (mm day^{-1}), scale (mm day^{-1}), and shape parameters for station data and at TRMM15, CHIRPS38, and CP4A grid cells that correspond to the coordinates of each station. The standard deviation across the study area is shown in parentheses. We use the longest time series available for each TRMM and CHIRPS to get best estimates of the GEV parameters for each dataset. We tabulate the full range of parameter estimates across all stations and station grid cells in Table S4 in the online supplemental material.

	Parameter	Station	TRMM15 (station grid cells)	CHIRPS38 (station grid cells)	CP4A (station grid cells)
All stations	Location	66.0 (7.6)	55.6 (8.6)	35.7 (4.5)	68.6 (10.8)
	Scale	19.0 (3.3)	15.8 (2.5)	9.8 (2.2)	22.9 (8.7)
	Shape	0.03 (0.17)	0.02 (0.02)	0.11 (0.06)	0.08 (0.10)
Malawi	Location	67.8 (7.9)	61.5 (3.2)	38.7 (2.2)	75.9 (5.9)
	Scale	18.9 (3.8)	17.5 (1.1)	10.9 (1.9)	27.1 (9.1)
	Shape	0.07 (0.20)	0.02 (0.02)	0.13 (0.05)	0.12 (0.10)
Zambia	Location	63.2 (7.0)	46.7 (5.4)	31.3 (3.0)	57.5 (4.2)
	Scale	19.1 (2.8)	13.2 (1.3)	8.1 (1.4)	16.6 (1.3)
	Shape	−0.03 (0.14)	0.03 (0.03)	0.07 (0.06)	0.02 (0.04)

spatial variation in the parameters describing the extreme rainfall distributions (Table 4); this is expected, as factors such as geography and elevation will influence the characteristics of rainfall and rainfall extremes.

The mean of the location, scale, and shape parameters for the observations and climate models is shown in Fig. 2. There is a large range in all parameters among the CORDEX models. CP4A is close to the TRMM and station observations for

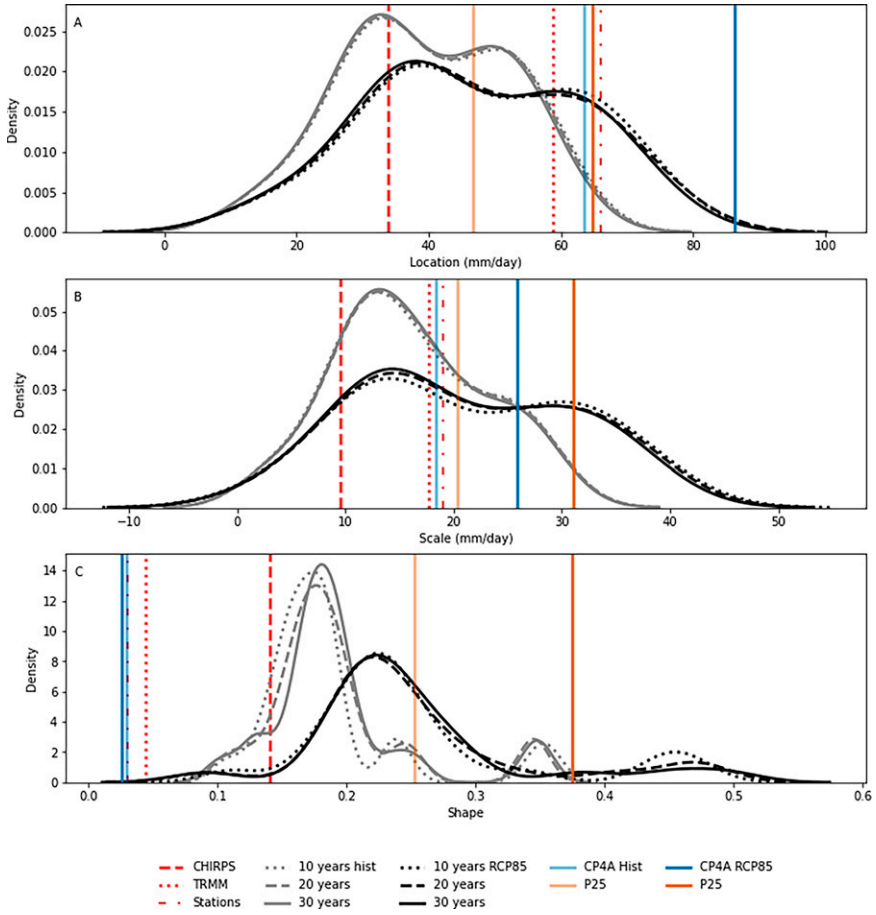

FIG. 2. Average location, scale, and shape parameters for the study area, determined from regional MLE in 10-, 20-, and 30-yr CORDEX models (black and gray curves), CP4A, and P25 for historical and RCP8.5 scenario. TRMM15 (dotted line), CHIRPS38 (dashed line), and stations (dot–dash line) are also shown. CORDEX is an ensemble of 23 model simulations, which is why it is presented as a curve, and the observations, P25, and CP4A are shown as vertical lines. Kernel density is estimated from the regional mean from individual models.

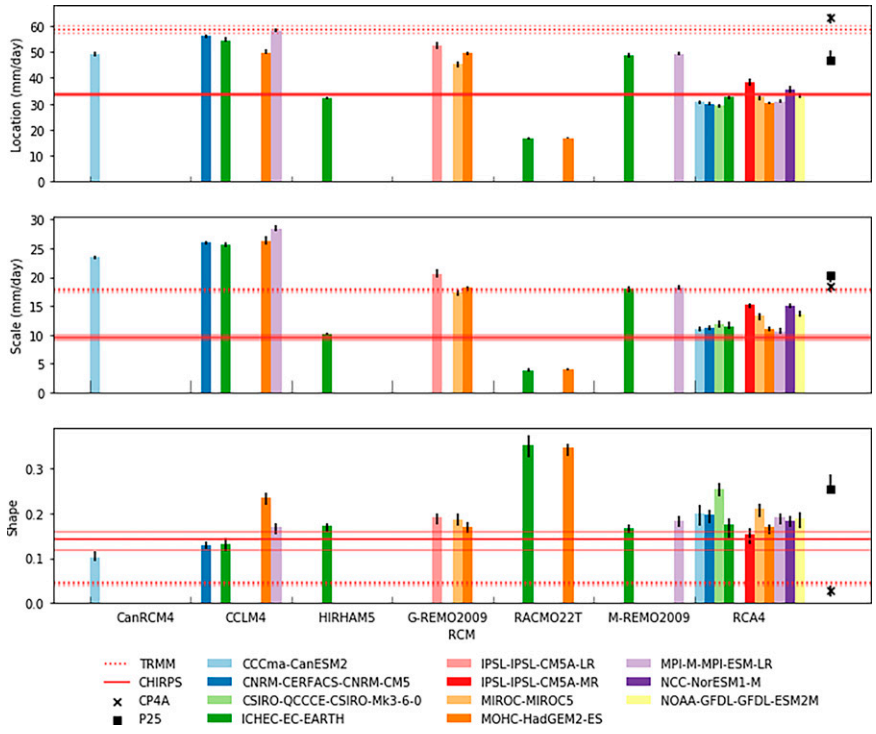

FIG. 3. Average present-day shape, location, and scale parameters across the study area in COR30. Colors show different GCMs; the RCM is given by the *x* axis. CP4 and P25 values are also shown to the right (black symbols). TRMM15 and CHIRPS38 region average are shown as dotted and solid red lines, respectively. Confidence intervals (5th and 95th percentile) from 40 bootstrap samples are shown as vertical black lines for the climate models and as fainter lines for observations.

the location and scale parameters (Figs. 2a,b). For the shape parameter, which is critical for very rare extremes, CP4A is closest to the observations and toward the lower end of the range of the models, while P25 is toward the higher end. While CHIRPS has a higher shape parameter than the stations and TRMM, it is still toward the lower end of the range from the parameterized models.

Climate change moves all parameters higher for the majority of models, which means extremes would also become more severe. The change in the shape parameter with climate change is very small for CP4A relative to P25 and the majority of CORDEX models (Fig. 2). The length of the time series has only a minor impact on the estimated parameters (and change in parameters) for the CORDEX models.

Figure 3 shows the range in the parameters of the CORDEX models by GCM for the historical scenario, as well as the confidence intervals (CI) from the bootstrap resampling (5%–95%). The range in the parameters of the CORDEX models is mainly due to the RCM, rather than the driving GCM (Fig. 3). Given the large domain covered by the CORDEX-Africa models, the RCM dynamics may have a larger influence than lateral boundary conditions on not only fine-scale processes (such as those related to topography) but also large-scale processes (Mariotti et al. 2011; Dosio et al. 2019). The importance of RCM physics is also why CP4 and P25 give such different results, even though their driving GCM and boundary conditions are the same. All CORDEX models and P25 give shape parameter estimates higher than TRMM, and the majority give shape parameter estimates higher than CHIRPS. For some CORDEX models, the shape parameter is over double that found from TRMM, while the CP4A shape parameter is in line with TRMM (Fig. 3).

b. *Return levels*

The choice of satellite product has more of an impact on return levels than time series length (Fig. 4, comparing TRMM with CHIRPs), which we also found for the GEVD parameters in section 3a. The TRMM return levels are generally higher than the CHIRPS return levels (80%–92%, with larger differences at higher return levels), except along the coast of Tanzania, which includes the coastal plain and mountainous areas, due to the higher location and scale parameters. This is in line with previous work looking at moderate extremes that found that TRMM had higher-intensity rainfall than CHIRPS (Shiferaw et al. 2018). As the return period increases, CHIRPS becomes wetter than TRMM in eastern Tanzania due to the higher shape parameter. In contrast to most of the study area, TRMM has much lower return levels than CHIRPS38 in

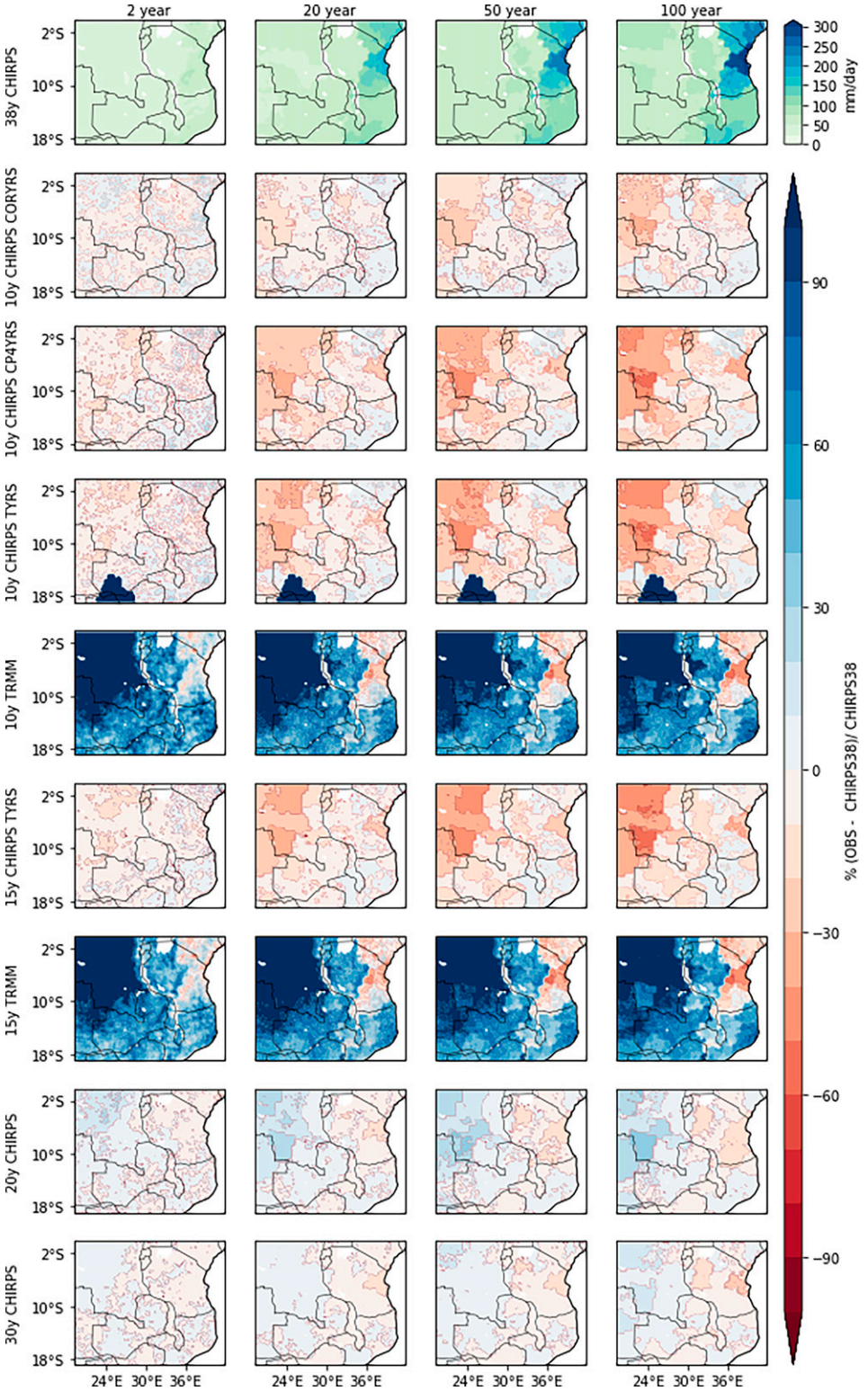

FIG. 4. Percentage difference between estimated return levels from TRMM, and CHIRPS10, 15, 20, and 30 and the return levels estimated from CHIRPS38.

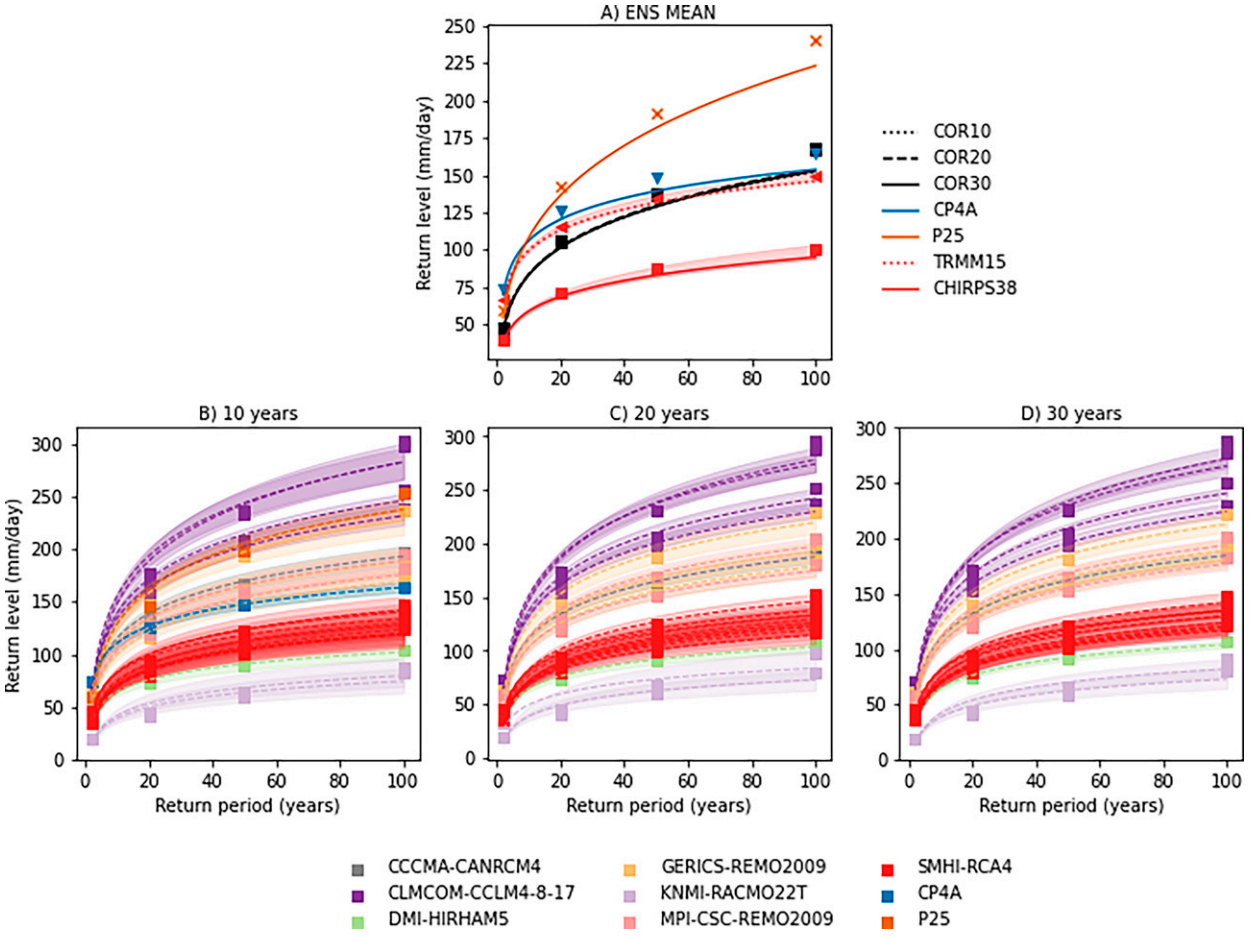

FIG. 5. (a) Return levels in the present day for the ensemble mean for COR30, 20, and 10, CP4A and P25, and TRMM15 and CHIRPS38 for return periods of 2–100 years, and confidence intervals for TRMM15 and CHIRPS38. Return periods of 2, 20, 50, and 100 are calculated on the basis of gridcell GEV. The lines are indicative of return levels and are based on GEV applied to region-level parameters. (b)–(d) Confidence intervals for individual CORDEX models (5th and 95th percentile) from 40 bootstrap samples, colored by RCM. All data regridded to CORDEX 0.44° × 0.44° grid for the purposes of comparison. Points show return levels calculated from gridcell-level GEV; the lines are estimated between points using a logarithmic function.

coastal areas. This region also includes mountainous areas. Satellite products in eastern Africa tend to struggle in mountainous areas, with many products underestimating rainfall in this area (Cattani et al. 2016). Difficulties with orographic rainfall may be why the relationship between TRMM and CHIRPS reverses in mountainous regions.

For the CHIRPS10 TYRS (TRMM years) time slice, there is a region in the south that is much wetter than CHIRPS38. This clear outlier may be due to the regionalization, which was based on CHIRPS38, not working well for this particular region and year, or unusual climate conditions in that region and year, which CHIRPS10 TYRS may have been sensitive to due to the shortness of the time slice.

The return levels from the climate models are shown in Fig. 5. The length of the time series has a negligible overall influence on the CORDEX ensemble mean of the return levels (i.e., the mean of the return levels calculated from individual models, black lines, Fig. 5a), although some minor difference is apparent when looking at individual models (Figs. 5b–d). The influence of time series length is more apparent in the width of the confidence intervals, which get smaller with increasing time series length. The relationship between CP4A and the parameterized models remains the same when grid cells in mountainous areas and at the coast are removed, both of which have higher return levels (not shown).

For events with a return period of 2 years, the CP4A model gives higher return levels than CORDEX and P25 (Fig. 5a). This is in line with previous work showing that for extreme indices (i.e., 95th percentile), CP4A is wetter than P25 (Kendon et al. 2019). However, for events with a return period of 20 years or higher, the return levels in CP4A are lower than in P25. At events with a return period of 100 years, CP4A and CORDEX converge, as the CORDEX models have a higher shape parameter than CPA, so the probability of events with very high return levels occurring is higher than in CP4A. P25

TABLE 5. Average return levels (mm day^{-1}) for return periods of 2, 20, 50, and 100 years. Estimates are provided for the station data and for TRMM15, CHIRPS38, and CP4A grid cells that correspond to the coordinates of each station. Confidence intervals for the station return-level estimates are tabulated in Table S5 in the online supplemental material.

	All stations				Malawi				Zambia			
	2 yr	20 yr	50 yr	100 yr	2 yr	20 yr	50 yr	100 yr	2 yr	20 yr	50 yr	100 yr
Station	73	127	149	168	75	132	158	179	70	119	137	150
TRMM15	61	104	121	133	68	115	133	146	52	88	102	113
CHIRPS38	39	70	84	95	43	78	94	107	34	59	69	77
CP4A	77	153	188	218	86	183	230	271	64	109	126	139

sits toward the upper end of the range of the CORDEX models, and gives return levels higher than the CORDEX ensemble mean.

In comparing the station-derived return-level estimates with the satellite-derived estimates (Table 5), we find that station return levels are on average 19%–26% larger than in TRMM across all return periods. This systematic difference is primarily related to the higher average shape and location parameters at the stations in comparison with TRMM (Table 4). For CHIRPS, we find that daily rainfall extremes are underestimated by an average of 77%–85% in comparison with station-derived estimates, which is related to the much smaller average location parameters in CHIRPS.

In CP4A, return levels are progressively overestimated relative to the observations as return period increases in Malawi. This is because of the larger shape parameter (Table 4), which likely arises from the moderate wet bias in extreme precipitation over eastern Africa, and Malawi in particular, in CP4A (Kendon et al. 2019), in combination with the short time series length of CP4A. This overestimation of extreme rainfall in Malawi may also be partly due to the geography of Malawi. Both CP4A and P25 have the tendency to overestimate extreme rainfall over mountains and lakes (Kendon et al. 2019), and large portions of Malawi are covered by both; the Great Rift Valley runs through Malawi from north to south, while Lake Malawi makes up a large portion of Malawi's eastern boundary. Return-level estimates for Zambia in CP4A more closely align with station-derived estimates but are slightly underestimated.

c. *Climate change impacts on return levels*

Climate change increases the regional mean return levels for the CORDEX ensemble and for CP4A and P25 (Fig. 6). The increase in return levels with climate change of very rare events (return period $>$ 20 years) is overestimated in the parameterized models relative to CP4A because of large changes in the shape parameter with climate change. However, the increase in extreme rainfall is still large in CP4A, with return levels increasing by 36% on average. This means that 1-in-100-yr events in the present day becoming 1-in-23-yr events at the end of the century. For all models, there are some parts of the study area with decreases in return levels due to rainfall declines (not shown).

The difference between CP4A and the parameterized models is more apparent when looking at the climate change impacts on return levels than at the historical return levels. For events with a 2-yr return period, the absolute increase in return levels with climate change is higher for CP4A than for the CORDEX ensemble and P25 (consistent with comparison of previous analysis of CP4A vs P25 for return periods of around 1 yr; e.g., Kendon et al. 2019; Finney et al. 2020), but this reverses for higher return periods with CP4A having smaller increases in return levels for return periods of 50 and 100 years than CORDEX and P25. The difference between CP4A and P25 and CORDEX is likely due to the shape parameter. Climate change has little impact on the shape parameter for CP4A (absolute mean change $<$ 0.001), whereas the shape parameter increases for the majority of CORDEX models and for P25, meaning rarer events become more frequent in these models with climate change. This may be an unrealistic response given that Wilson and Toumi (2005) argue that the shape parameter should be largely unaffected by climate change given moisture conservation and also that the shape parameter is invariant with latitude.

The climate change impact on return levels in CP4A is fundamentally different to the parameterized models. Even those models with realistic (i.e., similar to satellite or station observations) present-day shape parameters in the parameterized convection models have large changes in the shape parameter with climate change, while CP4A has near-zero mean change in the shape parameter (Figs. S2 and S3 in the online supplemental material). Note that, while the mean change in CanRCM4 is near zero, the change in the shape parameter is large and varies spatially (between 0.23 and -0.29), while in contrast the change in the shape parameter in CP4A is small everywhere (between 0.08 and -0.11). Similarly, the parameterized models in Fig. S2 that have a ratio near one like CP4A, meaning that the percentage change in 100-yr events is similar to percentage change in 2-yr events, have large spatial variation in the ratio, which is disguised by taking the mean (std dev CP4A = 4.2, CORDEX ensemble mean std dev = 193, with a minimum std dev of 7). The models with the largest present-day shape parameter also generally have the largest change in the return levels for 1-in-100-yr events relative to the change in 1-in-2-yr events (Fig. S2), and large changes in the shape parameter with climate change (Fig. S3).

d. *Climate change impact on return levels at station locations*

Given CP4A better represents observed daily rainfall relative to the other climate models examined, and better captures the essential physics of intensification of rainfall under

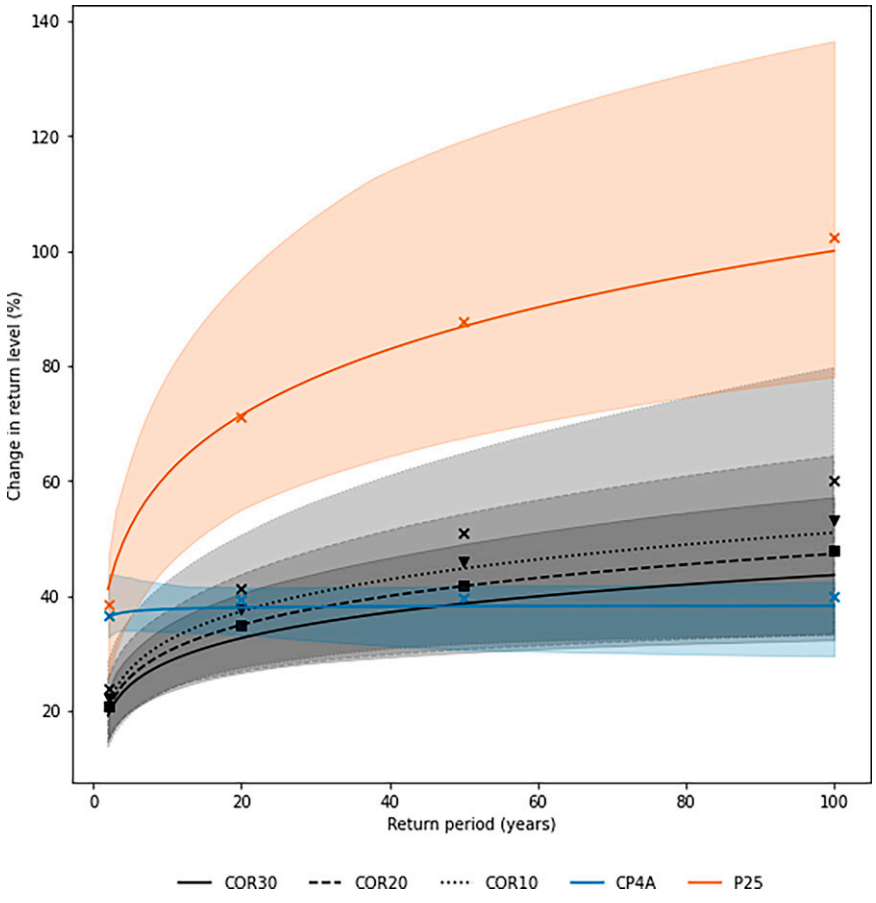

FIG. 6. Percentage change in return levels with end-of-century climate change (RCP8.5) for COR30, COR20, COR10, CP4, and P25. Confidence intervals (95th and 5th percentile) for change in return levels with climate change based on 40 bootstrap samples are also shown. One region from one bootstrap sample was removed from the CI calculation for CP4A as it had a large negative location parameter. Points show return levels based on gridcell GEV; lines are indicative of return levels on the basis of GEV applied to region-level parameters.

climate change, such as increases in updraft intensity (Jackson et al. 2020) and response to changing mesoscale forcings (Finney et al. 2020), we apply the future shift in GEVD parameters from CP4A to the historical station-derived parameters using the delta change method previously described to estimate the future change in return levels at the station locations [Eq. (2)].

The return-level estimates for the proxy future station data ($RCP8.5_{DC}$) are on average smaller than those in CP4A under RCP8.5 (Table 6), indicating that the moderate wet bias in CP4A is partly accounted for using the delta change method. These return-level estimates also partly account for the difference in extreme rainfall characteristics between the gridcell scale and point location scale. Applying the 100-yr climate change signal from CP4A to the station data and taking the all-station average (Table 6), return levels increase by approximately 37% across all return periods under $RCP8.5_{DC}$ relative to historic return levels (Table 5). However, there is considerable regional variability in the future estimates, and the confidence intervals for the future estimates can be wide (Table S5 in the online supplemental material).

TABLE 6. Average future return levels (mm day^{-1}) for station grid cells in CP4A under the RCP8.5 emissions scenario, and projected return levels at station point locations ($RCP8.5_{DC}$). Confidence intervals for the future station return-level estimates are tabulated in Table S5 in the online supplemental material.

	All stations				Malawi				Zambia			
	2 yr	20 yr	50 yr	100 yr	2 yr	20 yr	50 yr	100 yr	2 yr	20 yr	50 yr	100 yr
CP4A (RCP8.5)	102	203	247	289	116	237	293	341	82	152	179	200
Station ($RCP8.5_{DC}$)	98	175	206	232	104	186	221	250	88	159	185	205

TABLE 7. Average percentage increase in return level per kelvin of global warming in CP4A (5.2 K under RCP8.5) for the all-station average. We do not compute this at the regional level because of data sparsity and because local temperature increases do not match global temperature increases. The typical Clausius–Clapeyron relationship between rainfall and global temperature change only emerges when results are averaged over large areas.

	2 yr	20 yr	50 yr	100 yr
Percent increase per kelvin (RCP8.5)	6.2	6.2	6.1	5.9
Percent increase per kelvin ($RCP8.5_{DC}$)	6.5	7.3	7.3	7.4

Table 7 shows the average percentage increase in return level per kelvin of global temperature rise for station grid cells in CP4A, and for the future station data under $RCP8.5_{DC}$. The modeled increase in global mean air temperature at 1.5 m for the CP4A future period is 5.2 K under RCP8.5 (Kendon et al. 2019). The average percentage increase in return level per kelvin across all return periods is approximately 7.1% for the station data, which is larger than the change modeled at the gridcell scale in CP4A (6.1%). This is close to the upper limit of the Clausius–Clapeyron relation, which states that air can hold around 6%–7% more moisture per kelvin of warming at Earth's surface (Chan et al. 2016), suggesting that the station return-level estimates under $RCP8.5_{DC}$ are physically plausible and qualitatively consistent with other studies (e.g., Shongwe et al. 2009; Pinto et al. 2016; Kendon et al. 2019).

4. Discussion and conclusions

We used EVT to examine the climate change impact on rare extreme rainfall events in a novel, convection-permitting climate model (CP4A) and compared this with its parameterized counterpart (P25), the CORDEX-Africa ensemble and observations. We used the CORDEX-Africa ensemble to examine the impact of time series length (10, 20 and 30 years) on the EVT analysis. In CP4A, which had the best agreement with the observations, we found that return levels increase by on average 36% in the future, which results in a 1-in-100-yr event becoming a 1-in-23-yr event.

We found large differences in the behavior between CP4A and the parameterized convection models due to differences in the shape parameter, which controls the tail behavior of the GEVD. The parameterized models we examined have shape parameters much higher than TRMM and CP4A, and most models also had higher estimates than CHIRPS as well. The TRMM and CHIRPS satellites had shape parameters between −0.08 and 0.14, depending on time series length, and the stations had estimates between −0.29 and 0.29. Other studies based on rain gauges estimate the shape parameter for daily rainfall to be between 0 and 0.23 (Papalexiou and Koutsoyiannis 2013; Ragulina and Reitan 2017). With sufficient time series length (>100 yr), a negative shape parameter is highly unlikely (Papalexiou and Koutsoyiannis 2013). The mean shape parameter for CP4A in the present day was 0.03, and for the CORDEX models was between 0.11 and 0.36 depending on time series length. Further, all CORDEX models (even those with realistic present-day shape parameters) had large changes in the shape parameter with climate change, relative to the change in the shape parameter in CP4A with climate change. Large changes in the shape parameter may be unrealistic as it should be largely unaffected by climate change due to moisture conservation and as it is invariant with latitude in the present day (Wilson and Toumi 2005). With large changes in the shape parameter, also came large increases in the return levels for very rare events (return period > 20 years) with climate change. The change in return levels for CP4A was mainly due to changes in the location and scale parameters, not the shape parameter. For events with return periods > 20 years, the increase in return level with climate change in CP4A was lower than for the CORDEX ensemble mean.

Previous applications of EVT to rainfall have made extensive use of parameterized GCMs and RCMs (e.g., Semmler and Jacob 2004; Frei et al. 2006; Kharin et al. 2007; Fowler and Ekstrom 2009; Schliep et al. 2010). The assumption of this work is that even if the present-day extremes are not well represented, parameterized models can still represent the relative change in extremes with climate change (Fowler et al. 2010). However, recent work has shown how convection-permitting models, by explicitly modeling convective updrafts, capture the intensification of updrafts under climate change more realistically and give an improved physical basis for modeling changes in extremes, with larger changes in moderate (approximately annual) extremes (Kendon et al. 2019; Finney et al. 2020; Jackson et al. 2020). We analyze rarer extremes, and our results show large differences in the climate change impact on the parameters of the extreme value distribution between CP4A and the parameterized convection models, and therefore large differences in the climate change impact on return levels. Even the more realistic parameterized convection climate models for the present day may have unrealistic climate change impacts.

The same processes that lead to unrealistic present-day shape parameters may lead to unrealistic responses of extreme rainfall to climate change in parameterized models. As EVT is a method for looking at extremes, it is sensitive to outliers. An assumption of EVT is that the maximum value found within a time slice is representative for a return period of the length of the time slice, that is, the maximum value found within a 30-yr time period is representative of a 1-in-30-yr event (Knote et al. 2010). In parameterized convection models, unrealistic high-intensity gridpoint storms occur (Chan et al. 2014; Ban et al. 2015; Thomassen et al. 2021). Outliers have a lot of weight when fitting the GEVD (Knote et al. 2010), and these unrealistic storms therefore have a big impact on the estimation of parameters. There is limited work on EVT in convection-permitting models, however, those few other studies have also found large differences in the climate change signal between convection-permitting and parameterized models, particularly in seasons when convective activity is most important (Chan et al. 2014; Kendon et al. 2017; Cannon and Innocenti 2019), likely linked to the ability of convection-permitting models to model intensification of

updrafts under climate change (Jackson et al. 2020). In seasons when convective precipitation is less important and large-scale frontal ascent is more dominant, the differences in extreme precipitation between convection-permitting and parameterized models are smaller (Chan et al. 2014; Kendon et al. 2017). Our results suggest that most parameterized convection models do not capture the behavior of extremes in the present day, and that for even those models that do capture present-day behavior, relative changes in extremes (return period > 20 years) with climate change may be overestimated. For moderate (yearly) extremes, parameterized convection models may underestimate climate change impacts (Kendon et al. 2019). It is likely that the CP4A response to climate change is more in line with reality than the parameterized models, and thus, most parameterized convection climate models are likely not suitable for looking even at relative changes in extreme events in seasons when convective activity is important to rainfall.

We found large differences between the TRMM and CHIRPS satellite data. CP4A parameters and return levels were closer to TRMM than TRMM was to CHIRPS, while the parameterized models were different to all observations and CP4A. This large difference in satellite-derived estimates would make it difficult to use satellite data as the basis for improving model performance. Station-derived parameters and return levels were also found to be close to estimates made form corresponding grid cells in TRMM and CP4A. Differences in satellite observations have previously been found when looking at moderate extremes using satellite data, and when comparing satellites and rain gauges for extremes, particularly in mountainous areas (Islam 2018; Timmermans et al. 2019). A recent evaluation of satellites in East Africa found that all satellites struggle with detecting daily extreme rainfall events, and that CHIRPS may have particularly poor performance in this area (Ageet et al. 2022). The large differences between TRMM and CHIRPS raises the question as to whether satellite rainfall products are appropriate or useful for examining rare extreme rainfall events (Timmermans et al. 2019), although if rain gauge data are inaccessible, of poor quality, or spatially sparse there may not be many other options. As CP4A is already closer to TRMM than TRMM is to CHIRPS, it would be difficult to use these satellite products as a basis for improving CP4As performance. This disagreement between observational products highlights the importance of collection and accessibility of in situ data rainfall for informing adaptation, including digitization of old paper rainfall records.

The time period examined impacted the return levels estimated from satellite data; though these differences were small relative to the differences between CHIRPS and TRMM. There were differences in the estimates of return level for CHIRPS depending on which 10-yr time period we looked at. Even with regional frequency analysis (RFA), with a time period as short as 10 years the results may be sensitive to the number of ENSO or IOD events, which impact rainfall in the region (Black et al. 2003; Kijazi and Reason 2005). Even with RFA, using a time period longer than 10 years would be preferable, as pooling data does not account for the region being systematically wetter or drier due to the phase of key teleconnections. However, the influence of ENSO and the IOD on daily rainfall extremes in eastern Africa is beyond the scope of this paper and requires further analysis.

A further limitation is that we have not examined the impact of spatial resolution on results. In comparing gridded output to station output, each grid cell represents rainfall averaged over a greater surface area, which will mean that extremes are damped and the return levels will be smaller. A comprehensive analysis of the effects of resolution on extremes is outside the scope of this study (our principal focus is on convection-permitting versus parameterized models); however, future studies should examine this.

We find that there is large variation in the scaling of return-level estimates between the gridbox scale and the point location scale, with variation occurring between stations, datasets, and return periods. On average, we find that station return levels are 19%–26% larger than at corresponding grid cells in TRMM across all return periods and 77%–85% larger than in CHIRPS. This suggests that there is no constant downscaling factor that can be applied to satellite-derived return-level estimates in the study region, and that site-specific analysis is likely required if stakeholders seek to incorporate extreme rainfall return levels into the design of large infrastructure projects. Our work further motivates the need for ensembles of convection-permitting models for Africa (Senior et al. 2021) and synthesizing information from observations and convection-permitting and global models to inform decisions (e.g., Klein et al. 2021; Mittal et al. 2021).

Acknowledgments. This work was supported by the Biotechnology and Biological Sciences Research Council through U.K. Research and Innovation as part of the Global Challenges Research Fund, AFRICAP program, Grant BB/P027784/1. Author Marsham was funded by HyCRISTAL, IMPALA, and the NCAS ACREW program.

Data availability statement. The climate model and satellite rainfall data used in this paper are freely available online. The CP4A and P25 data are available from the CEDA archive. The CORDEX climate model data are available from ESGF. The weather station data were privately shared by the in-country meteorological agencies, and so we are unable to share them without authorization. All other analyses are available upon reasonable request to the corresponding author.

REFERENCES

Acero, F. J., J. A. García, and M. C. Gallego, 2011: Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *J. Climate*, **24**, 1089–1105, https://doi.org/10.1175/2010JCLI3627.1.

Ageet, S., A. H. Fink, M. Maranan, J. E. Diem, J. Hartter, A. L. Ssali, and P. Ayabagabo, 2022: Validation of satellite rainfall estimates over equatorial East Africa. *J. Hydrometeor.*, **23**, 129–151, https://doi.org/10.1175/JHM-D-21-0145.1.

Alexander, L. V., 2016: Global observed long-term changes in temperature and precipitation extremes: A review of progress

and limitations in IPCC assessments and beyond. *Wea. Climate Extremes*, **11**, 4–16, https://doi.org/10.1016/j.wace.2015.10.007.

Badr, H. S., B. F. Zaitchik, and A. K. Dezfuli, 2015: A tool for hierarchical climate regionalization. *Earth Sci. Inf.*, **8**, 949–958, https://doi.org/10.1007/s12145-015-0221-7.

Ban, N., J. Schmidli, and C. Schär, 2015: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophys. Res. Lett.*, **42**, 1165–1172, https://doi.org/10.1002/2014GL062588.

——, J. Rajczak, J. Schmidli, and C. Schär, 2020: Analysis of alpine precipitation extremes using generalized extreme value theory in convection-resolving climate simulations. *Climate Dyn.*, **55**, 61–75, https://doi.org/10.1007/s00382-018-4339-4.

Berthou, S., E. J. Kendon, D. P. Rowell, M. J. Roberts, S. Tucker, and R. A. Stratton, 2019: Larger future intensification of rainfall in the West African Sahel in a convection-permitting model. *Geophys. Res. Lett.*, **46**, 13 299–13 307, https://doi.org/10.1029/2019GL083544.

Black, E., J. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94, https://doi.org/10.1175/1520-0493(2003)131<0074:AOSOTR>2.0.CO;2.

Cannon, A. J., and S. Innocenti, 2019: Projected intensification of sub-daily and daily rainfall extremes in convection-permitting climate model simulations over North America: Implications for future intensity-duration-frequency curves. *Nat. Hazards Earth Syst. Sci.*, **19**, 421–440, https://doi.org/10.5194/nhess-19-421-2019.

Cattani, E., A. Merino, and V. Levizzani, 2016: Evaluation of monthly satellite-derived precipitation products over East Africa. *J. Hydrometeor.*, **17**, 2555–2573, https://doi.org/10.1175/JHM-D-15-0042.1.

Chamani, R., D. Monkam, and Z. Yepdo Djomou, 2018: Return times and return levels of July–September extreme rainfall over the major climatic sub-regions in Sahel. *Atmos. Res.*, **212**, 77–90, https://doi.org/10.1016/j.atmosres.2018.04.026.

Chan, S. C., E. J. Kendon, H. J. Fowler, S. Blenkinsop, N. M. Roberts, and C. A. T. Ferro, 2014: The value of high-resolution Met Office regional climate models in the simulation of multi-hourly precipitation extremes. *J. Climate*, **27**, 6155–6174, https://doi.org/10.1175/JCLI-D-13-00723.1.

——, ——, N. M. Roberts, H. J. Fowler, and S. Blenkinsop, 2016: Downturn in scaling of UK extreme rainfall with temperature for future hottest days. *Nat. Geosci.*, **9**, 24–28, https://doi.org/10.1038/ngeo2596.

Chapman, S., C. E. Birch, E. Pope, S. Sallu, C. Bradshaw, J. Davie, and J. H. Marsham, 2020: Impact of climate change on crop suitability in sub-Saharan Africa in parameterized and convection-permitting regional climate models. *Environ. Res. Lett.*, **15**, 094086, https://doi.org/10.1088/1748-9326/ab9daf.

Cioffi, F., F. Conticello, and U. Lall, 2016: Projecting changes in Tanzania rainfall for the 21st century. *Int. J. Climatol.*, **36**, 4297–4314, https://doi.org/10.1002/joc.4632.

Climate Prediction Center, 2020: Cold & warm episodes by season. NOAA/National Weather Service, accessed 24 February 2021, https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 209 pp.

De Paola, F., M. Giugni, M. E. Topa, and E. Bucchignani, 2014: Intensity-duration-frequency (IDF) rainfall curves, for data series and climate projection in African cities. *SpringerPlus*, **3**, 133, https://doi.org/10.1186/2193-1801-3-133.

Dinku, T., C. Funk, P. Peterson, R. Maidment, T. Tadesse, H. Gadain, and P. Ceccato, 2018: Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quart. J. Roy. Meteor. Soc.*, **144** (Suppl. 1), 292–312, https://doi.org/10.1002/qj.3244.

Do Lago, C., E. Mendiondo, F. Olivera, and M. Giocomoni, 2019: Application of a disaggregation method for the generation of climate changed intensity-duration-frequency curves for predicting future extreme rainfall impacts on transportation infrastructure. *MATEC Web Conf.*, **271**, 04002, https://doi.org/10.1051/matecconf/201927104002.

Dosio, A., R. G. Jones, C. Jack, C. Lennard, G. Nikulin, and B. Hewitson, 2019: What can we know about future precipitation in Africa? Robustness, significance and added value of projections from a large ensemble of regional climate models. *Climate Dyn.*, **53**, 5833–5858, https://doi.org/10.1007/s00382-019-04900-3.

Finney, D. L., and Coauthors, 2019: Implications of improved representation of convection for the East Africa water budget using a convection-permitting model. *J. Climate*, **32**, 2109–2129, https://doi.org/10.1175/JCLI-D-18-0387.1.

——, J. H. Marsham, D. P. Rowell, E. J. Kendon, S. O. Tucker, R. A. Stratton, and L. S. Jackson, 2020: Effects of explicit convection on future projections of mesoscale circulations, rainfall, and rainfall extremes over eastern Africa. *J. Climate*, **33**, 2701–2718, https://doi.org/10.1175/JCLI-D-19-0328.1.

Fontolan, M., A. C. F. Xavier, H. R. Pereira, and G. C. Blain, 2019: Using climate change models to assess the probability of weather extremes events: A local scale study based on the generalized extreme value distribution. *Bragantia*, **78**, 146–157, https://doi.org/10.1590/1678-4499.2018144.

Fowler, H. J., and M. Ekstrom, 2009: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *Int. J. Climatol.*, **29**, 385–416, https://doi.org/10.1002/joc.1827.

——, D. Cooley, S. R. Sain, and M. Thurston, 2010: Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. *Extremes*, **13**, 241–267, https://doi.org/10.1007/s10687-010-0101-y.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale, 2006: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *J. Geophys. Res.*, **111**, D06105, https://doi.org/10.1029/2005JD005965.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10.1038/sdata.2015.66.

Garcia-Aristizabal, A., E. Bucchignani, E. Palazzi, D. D'Onofrio, P. Gasparini, and W. Marzocchi, 2015: Analysis of non-stationary climate-related extreme events considering climate change scenarios: An application for multi-hazard assessment in the Dar es Salaam region, Tanzania. *Nat. Hazards*, **75**, 289–320, https://doi.org/10.1007/s11069-014-1324-z.

Gilleland, E., and R. W. Katz, 2006: Analyzing seasonal to interannual extreme weather and climate variability with the extremes toolkit. *18th Conf. on Climate Variability and Change*, Atlanta, GA, Amer. Meteor. Soc., P2.15, https://ams.confex.com/ams/pdfpapers/101830.pdf.

Golroudbary, V. R., Y. Zeng, C. M. Mannaerts, and Z. Su, 2016: Attributing seasonal variation of daily extreme precipitation

events across the Netherlands. *Wea. Climate Extremes*, **14**, 56–66, https://doi.org/10.1016/j.wace.2016.11.003.

Groupe de Recherche en Hydrologie Statistique, 1996: Intercomparison of regional flood frequency procedures for Canadian rivers. *J. Hydrol.*, **186**, 85–103, https://doi.org/10.1016/S0022-1694(96)03043-0.

Han, F., K. H. Cook, and E. K. Vizy, 2019: Changes in intense rainfall events and dry periods across Africa in the twenty-first century. *Climate Dyn.*, **53**, 2757–2777, https://doi.org/10.1007/s00382-019-04653-z.

Hanel, M., T. A. Buishand, and C. A. T. Ferro, 2009: A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *J. Geophys. Res.*, **114**, D15107, https://doi.org/10.1029/2009JD011712.

Harrison, L., C. Funk, and P. Peterson, 2019: Identifying changing precipitation extremes in sub-Saharan Africa with gauge and satellite products. *Environ. Res. Lett.*, **14**, 085007, https://doi.org/10.1088/1748-9326/ab2cae.

Hosking, J. R. M., and J. R. Wallis, 1997: *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, 224 pp.

Huffman, G. J., 2019: The Transition in Multi-Satellite Products from TRMM to GPM (TMPA to IMERG). NASA Algorithm Theoretical Basis Doc., 5 pp., https://gpm.nasa.gov/sites/default/files/2020-10/TMPA-to-IMERG_transition_201002.pdf.

——, and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

Islam, M. A., 2018: Statistical comparison of satellite-retrieved precipitation products with rain gauge observations over Bangladesh. *Int. J. Remote Sens.*, **39**, 2906–2936, https://doi.org/10.1080/01431161.2018.1433890.

Jackson, L. S., D. L. Finney, E. J. Kendon, J. H. Marsham, D. J. Parker, R. A. Stratton, L. Tomassini, and S. Tucker, 2020: The effect of explicit convection on couplings between rainfall, humidity, and ascent over Africa under climate change. *J. Climate*, **33**, 8315–8337, https://doi.org/10.1175/JCLI-D-19-0322.1.

Jalbert, J., A. C. Favre, C. Bélisle, and J. F. Angers, 2017: A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *J. Roy. Stat. Soc.*, **66**, 941–962, https://doi.org/10.1111/rssc.12212.

Jones, C., F. Giorgi, and G. Asrar, 2011: The Coordinated Regional Downscaling Experiment: CORDEX, an international downscaling link to CMIP5. *CLIVAR Exchanges*, No. 56, International CLIVAR Project Office, Southampton, United Kingdom, 34–39, https://www.clivar.org/sites/default/files/documents/Exchanges56.pdf.

Kendon, E. J., and Coauthors, 2017: Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Amer. Meteor. Soc.*, **98**, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1.

——, R. A. Stratton, S. Tucker, J. H. Marsham, S. Berthou, D. P. Rowell, and C. A. Senior, 2019: Enhanced future changes in wet and dry extremes over Africa at convection-permitting scale. *Nat. Commun.*, **10**, 1794, https://doi.org/10.1038/s41467-019-09776-9.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, https://doi.org/10.1175/JCLI4066.1.

——, G. M. Flato, X. Zhang, N. P. Gillett, F. Zwiers, and K. J. Anderson, 2018: Risks from climate extremes change differently from 1.5°C to 2.0°C depending on rarity. *Earth's Future*, **6**, 704–715, https://doi.org/10.1002/2018EF000813.

Kijazi, A. L., and C. J. C. Reason, 2005: Relationships between intraseasonal rainfall variability of coastal Tanzania and ENSO. *Theor. Appl. Climatol.*, **82**, 153–176, https://doi.org/10.1007/s00704-005-0129-0.

Kimani, M. W., J. C. B. Hoedjes, and Z. Su, 2017: An assessment of satellite-derived rainfall products relative to ground observations over East Africa. *Remote Sens.*, **9**, 430, https://doi.org/10.3390/rs9050430.

Klein, C., and Coauthors, 2021: Combining CMIP data with a regional convection-permitting model and observations to project extreme rainfall under climate change. *Environ. Res. Lett.*, **16**, 104023, https://doi.org/10.1088/1748-9326/ac26f1.

Knote, C., G. Heinemann, and B. Rockel, 2010: Changes in weather extremes: Assessment of return values using high resolution climate simulations at convection-resolving scale. *Meteor. Z.*, **19**, 11–23, https://doi.org/10.1127/0941-2948/2010/0424.

Mannshardt-Shamseldin, E. C., R. L. Smith, S. R. Sain, L. O. Mearns, and D. Cooley, 2010: Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *Ann. Appl. Stat.*, **4**, 484–502, https://doi.org/10.1214/09-AOAS287.

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

Martins, E. S., and J. R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, **36**, 737–744, https://doi.org/10.1029/1999WR900330.

Misiani, H. O., D. L. Finney, Z. T. Segele, J. H. Marsham, A. Tadege, G. Artan, and Z. Atheru, 2020: Circulation patterns associated with current and future rainfall over Ethiopia and South Sudan from a convection-permitting model. *Atmosphere*, **11**, 1352, https://doi.org/10.3390/atmos11121352.

Mittal, N., and Coauthors, 2021: Tailored climate projections to assess site-specific vulnerability of tea production. *Climate Risk Manage.*, **34**, 100367, https://doi.org/10.1016/j.crm.2021.100367.

Monsieurs, E., O. Dewitte, and A. Demoulin, 2019: A susceptibility-based rainfall threshold approach for landslide occurrence. *Nat. Hazards Earth Syst. Sci.*, **19**, 775–789, https://doi.org/10.5194/nhess-19-775-2019.

Muthoni, F. K., V. O. Odongo, J. Ochieng, E. M. Mugalavai, S. K. Mourice, I. Hoesche-Zeledon, M. Mwila, and M. Bekunda, 2019: Long-term spatial-temporal trends and variability of rainfall over eastern and southern Africa. *Theor. Appl. Climatol.*, **137**, 1869–1882, https://doi.org/10.1007/s00704-018-2712-1.

Onyutha, C., 2020: Analyses of rainfall extremes in East Africa based on observations from rain gauges and climate change simulations by CORDEX RCMs. *Climate Dyn.*, **54**, 4841–4864, https://doi.org/10.1007/s00382-020-05264-9.

Overeem, A., T. A. Buishand, and I. Holleman, 2009: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.*, **45**, W10424, https://doi.org/10.1029/2009WR007869.

Papalexiou, S. M., and D. Koutsoyiannis, 2013: Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.*, **49**, 187–201, https://doi.org/10.1029/2012WR012557.

Pinto, I., C. Lennard, M. Tadross, B. Hewitson, A. Dosio, G. Nikulin, H. J. Panitz, and M. E. Shongwe, 2016: Evaluation and projections of extreme precipitation over southern Africa from two CORDEX models. *Climatic Change*, **135**, 655–668, https://doi.org/10.1007/s10584-015-1573-1.

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.

Ragulina, G., and T. Reitan, 2017: Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. *Hydrol. Sci. J.*, **62**, 863–879, https://doi.org/10.1080/02626667.2016.1260134.

R Core Team, 2013: A language and environment for statistical computing. R Foundation for Statistical Computing, https://www.r-project.org/.

Rowell, D. P., and R. Chadwick, 2018: Causes of the uncertainty in projections of tropical terrestrial rainfall change: East Africa. *J. Climate*, **31**, 5977–5995, https://doi.org/10.1175/JCLI-D-17-0830.1.

Scarrott, C., and A. MacDonald, 2012: A review of extreme value threshold estimation and uncertainty quantification. *Revstat Stat. J.*, **10**, 33–60, http://dx.doi.org/10.57805/revstat.v10i1.110.

Schliep, E. M., D. Cooley, S. R. Sain, and J. A. Hoeting, 2010: A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, **13**, 219–239, https://doi.org/10.1007/s10687-009-0098-2.

Semmler, T., and D. Jacob, 2004: Modeling extreme precipitation events—A climate change simulation for Europe. *Global Planet. Change*, **44**, 119–127, https://doi.org/10.1016/j.gloplacha.2004.06.008.

Senior, C. A., and Coauthors, 2021: Convection-permitting regional climate change simulations for understanding future climate and informing decision-making in Africa. *Bull. Amer. Meteor. Soc.*, **102**, E1206–E1223, https://doi.org/10.1175/BAMS-D-20-0020.1.

Shiferaw, A., T. Tadesse, C. Rowe, and R. Oglesby, 2018: Precipitation extremes in dynamically downscaled climate scenarios over the greater horn of Africa. *Atmosphere*, **9**, 112, https://doi.org/10.3390/atmos9030112.

Shongwe, M. E., G. J. Van Oldenborgh, B. J. J. M. Van Den Hurk, B. De Boer, C. A. S. Coelho, and M. K. Van Aalst, 2009: Projected changes in mean and extreme precipitation in Africa under global warming. Part I: Southern Africa. *J. Climate*, **22**, 3819–3837, https://doi.org/10.1175/2009JCLI2317.1.

Sieck, L. C., S. J. Burges, and M. Steiner, 2007: Challenges in obtaining reliable measurements of point rainfall. *Water Resour. Res.*, **43**, W01420, https://doi.org/10.1029/2005WR004519.

Stephenson, A., and J. Tawn, 2004: Bayesian inference for extremes: Accounting for the three extremal types. *Extremes*, **7**, 291–307, https://doi.org/10.1007/s10687-004-3479-6.

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.

Sylla, M. B., F. Giorgi, E. Coppola, and L. Mariotti, 2013: Uncertainties in daily rainfall over Africa: Assessment of gridded observation products and evaluation of a regional climate model simulation. *Int. J. Climatol.*, **33**, 1805–1817, https://doi.org/10.1002/joc.3551.

Tarhule, A., 2005: Damaging rainfall and flooding: The other Sahel hazards. *Climatic Change*, **72**, 355–377, https://doi.org/10.1007/s10584-005-6792-4.

Thomassen, E. D., E. J. Kendon, H. J. D. Sørup, S. C. Chan, P. L. Langen, O. B. Christensen, and K. Arnbjerg-Nielsen, 2021: Differences in representation of extreme precipitation events in two high resolution models. *Climate Dyn.*, **57**, 3029–3043, https://doi.org/10.1007/s00382-021-05854-1.

Timmermans, B., M. Wehner, D. Cooley, T. O'Brien, and H. Krishnan, 2019: An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dyn.*, **52**, 6651–6670, https://doi.org/10.1007/s00382-018-4537-0.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205.

Wainwright, C. M., D. L. Finney, M. Kilavi, E. Black, and J. H. Marsham, 2020: Extreme rainfall in East Africa, October 2019–January 2020 and context under future climate change. *Weather*, **76**, 26–31, https://doi.org/10.1002/wea.3824.

Wang, X. L., B. Trewin, Y. Feng, and D. Jones, 2013: Historical changes in Australian temperature extremes as inferred from extreme value distribution analysis. *Geophys. Res. Lett.*, **40**, 573–578, https://doi.org/10.1002/grl.50132.

Wilson, P. S., and R. Toumi, 2005: A fundamental probability distribution for heavy rainfall. *Geophys. Res. Lett.*, **32**, L14812, https://doi.org/10.1029/2005GL022465.

World Bank, 2019: Disaster risk profile: Malawi. Global Facility for Disaster Reduction and Recovery Doc., 13 pp., https://www.gfdrr.org/sites/default/files/publication/malawi_low.pdf.

Wright, D. B., C. D. Bosma, and T. Lopez-Cantu, 2019: U.S. hydrologic design standards insufficient due to large increases in frequency of rainfall extremes. *Geophys. Res. Lett.*, **46**, 8144–8153, https://doi.org/10.1029/2019GL083235.