Water Resources Management (2021) 35:1587–1602
https://doi.org/10.1007/s11269-021-02804-8

Evaluation of Reanalysis Precipitation Data and Potential Bias Correction Methods for Use in Data-Scarce Areas

Victoria M. Garibay[1] · Margaret W. Gitau[1] · Nicholas Kiggundu[2] · Daniel Moriasi[3] · Fulgence Mishili[4]

Received: 14 December 2020 / Accepted: 1 March 2021
Published online: 31 March 2021

Abstract

Data availability and accessibility often present challenges to resolving regional water management issues. One primary input essential to models and other tools used to inform policy decisions is daily precipitation. Since observed datasets are not always present or accessible, data from the Climate Forecast System Reanalysis (CFSR) have become a potential alternative. A comparison of CFSR precipitation data to available observed data from stations in the East African countries Kenya, Uganda, and Tanzania showed notable differences between the two datasets, particularly with respect to precipitation totals and number of days receiving rainfall. A sliding window bias correction approach evaluated using 3 methods with 8 different window length and timestep variations showed that empirical quantile mapping with a 30-day sliding window length and 1-day timestep achieved the best performance. A comparison of bias corrected CFSR precipitation data against observed data showed marked improvement in the similarity of the number of wet days and maximum daily rainfall between the two datasets. For precipitation totals, bias correction reduced underprediction errors by 32% and overprediction errors by 81%. Results indicate that bias-corrected CFSR precipitation data provides an improved basis for water resources applications in the study region. Methodologies and approaches are extendable to other data-scarce regions or areas where complete and consistent data are not easily accessible.

Keywords East Africa · Daily rainfall · CFSR precipitation · Bias correction · DownscaleR · Essential characteristics

✉ Margaret W. Gitau
mgitau@purdue.edu

[1] Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, USA

[2] Department of Agricultural and Biosystems Engineering, Makerere University, Kampala, Uganda

[3] USDA-ARS Grazinglands Research Laboratory, El Reno, OK, USA

[4] Department of Agricultural Economics and Agribusiness, Sokoine University of Agriculture, Morogoro, Tanzania

1 Introduction

Lack of reliable and accessible data hampers efforts to form a strong scientific basis for water management policy decisions. Daily precipitation is a crucial input for water resources modelling and an important basis for storm-related planning. In East Africa, for example, the scarcity of good quality, continuous rainfall and other climate data is a challenge that climate and hydrology researchers have frequently faced (Githui 2008; Dinku et al. 2011; Egeru et al. 2014; Alemayehu et al. 2016; Schmocker et al. 2016; Gebrechorkos et al. 2018). A commonly used alternative to observed rainfall data is mathematically generated weather data, such as the data provided by the global, grid-calculated Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010). Other simulated climate data sources that serve a similar purpose include Climate Hazards Group InfraRed Precipitation (CHIRP), Observational-Reanalysis Hybrid (ORH), and Modern-Era Retrospective Analysis for Research and Applications (MERRA). CFSR was chosen for further research because it is freely available and readily accessible, it has performed well in regional comparison studies of similar datasets (Tesfaye et al. 2017; Zhan et al. 2016), and it is already widely used in East Africa (Alemayehu et al. 2016; Schmocker et al. 2016; Anaba et al. 2017; Mainya 2017; Muthuwatta et al. 2018; Lugoi et al. 2019).

The CFSR provides air temperature, surface and upper-level wind speeds, and precipitation totals among other weather parameters. The CFSR dataset covers the time period 1979–2010. The CFSR version 2 (V2) dataset is a similar product covering the period from 2011 to present day, with the effective cutoff being 2019 (Saha et al. 2011). These CFSR data have been used for research in the East African region including evaluation of precipitation trends, hydrologic modelling, study of ecosystem and agricultural productivity, and climate change analysis (Alemayehu et al. 2016; Schmocker et al. 2016; Anaba et al. 2017; Mainya 2017; Muthuwatta et al. 2018; Lugoi et al. 2019). However, previous research on the use of CFSR precipitation data and its effect on model performance for East Africa has produced mixed results. Lakew et al. (2020) and López et al. (2020) recognized the great potential of reanalysis datasets, but noted the importance of evaluating their inherent uncertainties when used at a local scale. In some studies, climate data generated by the CFSR compared well with those from other climate data sources (Tesfaye et al. 2017; Worqlul et al. 2017) while it overestimated precipitation in others (Duan et al. 2019; Zhan et al. 2016), hence the considerations for bias correction in this study.

Bias correction techniques generally fall under two categories: 1) scaling, and 2) distribution adjustment. Examples of scaling techniques include linear scaling (LS), local intensity scaling (LOCI), and power transformation (PTR), while methods such as empirical quantile mapping (EQM) and daily translation (DT), involve modifications to existing distributions (Teutschbein and Seibert 2012; Chen et al. 2013; Smitha et al. 2018). Linear scaling uses a factor derived from mean monthly values of control data in relation to the corresponding values from long-term observed data (Graham et al. 2007). The LOCI method targets the correction of the mean as well as the frequency of wet days in a dataset (Schmidli et al. 2006). Power transformation is a straightforward method which has historically been used in many contexts including the correction of climate data (e.g. Mehan et al. 2019). The method involves a coefficient and power variable, allowing for changes in both the mean and the variance of the dataset (Leander and Buishand 2007). The EQM method works if the simulated data provides an accurate prediction of the relative change in quantiles but not necessarily the actual quantile values (Feigenwinter et al. 2018). EQM makes adjustments to a simulated cumulative distribution function of the modeled data so that it matches that of the observed data, and DT is a similar correction performed by mapping and adjusting the frequency distribution (Chen et al. 2013). For this study, PTR, LOCI, and EQM were determined to be the best prospects based on the completeness of the raw datasets and the data needs and distribution-related limitations of the different methods.

The aim of this study was to evaluate CFSR data as a substitute for on-site measurements in data-scarce areas and identify bias correction methods that would be appropriate to improve the accuracy of the substitute dataset. Specifically, to: 1) determine the extent of discrepancies between the observed data and corresponding CFSR data; 2) identify the most suitable bias correction method using available data in the study region; and, 3) develop bias-corrected datasets for select stations within the study region. The study focused on the East African countries Kenya, Uganda, and Tanzania (Fig. 1). Limited locations in the study region have reasonably detailed weather records. However, even where the data are available, they are difficult to track down and challenging to obtain permission to use. As a result, and consistent with the concept of open data, this study generally targeted observed datasets that were readily available and easily accessible to anyone with internet access.

2 Materials and Methods

As a first step, CFSR precipitation data were compared against available observed daily precipitation data from weather stations in the study area to identify differences between observed and calculated rainfall. Based on results of the analysis, an investigation into potential methods for improving the correspondence between the CFSR and observed data was conducted. New datasets comprising corrected data were then developed based on the most suitable bias correction method as determined. All statistical calculations, bias corrections, and comparisons were conducted using the R environment (R Core Team 2018), selected because it is a reliable software which provides free access to powerful tools.

2.1 Site Description

East Africa is a uniquely diverse region in terms of climate and geography, both of which vary greatly across the three countries considered in this study (Fig. 1). The region's geomorphology has a strong influence on the climate conditions experienced at a local level, in particular the mountain and valley region (Great Rift Valley) cutting through Kenya and Tanzania, the Lake Victoria region, and distinct plateaus (Berakhi et al. 2015), all of which comprise roughly half of the study area. Kenya is fairly temperate in the southeast where plateaus and mountain formations keep weather mild, but to the northeast where there are mainly plains, the climate is more arid. Bordering the Indian Ocean, coastal areas in Kenya and Tanzania are characterized by heat and humidity. The remaining majority of Tanzania can be characterized as a tropical or subtropical plateau with mild weather, with altitude as the main driver of temperature variation. Uganda also consists mainly of temperate tropical plateaus but is warmer, particularly during dry periods. Average annual precipitation in Kenya ranges from less than 300 mm in arid regions to approximately 2,000 mm in the mountains, however the majority of the country experiences less than 500 mm/yr (Mogaka et al. 2006). Tanzania typically receives 600–800 mm/yr (Rowhani et al. 2011), but in the southern highlands, precipitation exceeds 1,500 mm/yr. Ugandan precipitation can vary from less than 1,000 mm/yr along the northeastern border to over 2,000 mm/yr near Lake Victoria (Jury 2018). All three countries

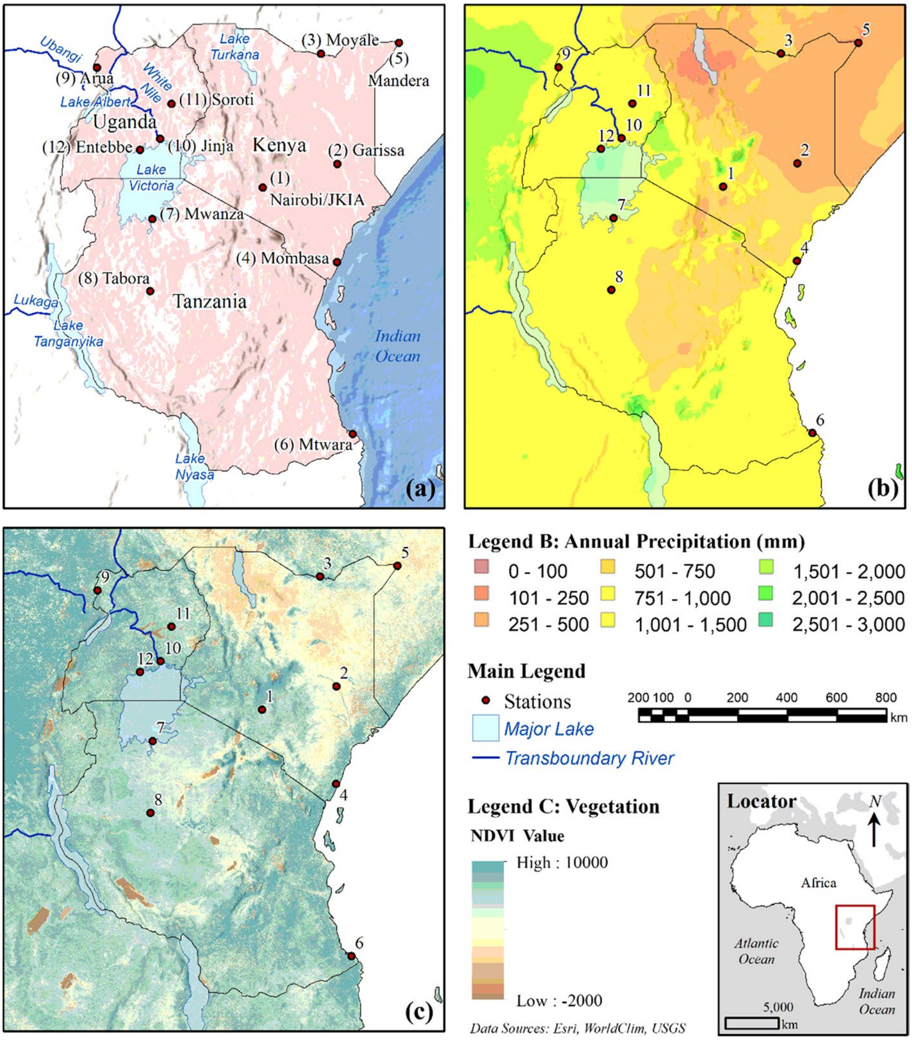

Fig. 1 Map of East Africa showing the position of observed data station locations relative to **a** political map with hillshading, **b** average annual precipitation map, and **c** vegetation map

have distinct rainy seasons, with two (long and short rains) being experienced in Kenya and northeastern Tanzania. In general, future climate projections are in agreement that East Africa will see a rise in average temperature (Allen et al. 2018). However, there is uncertainty about the effects of this change on rainfall patterns, which are highly dependent upon the Indian Ocean and El Niño Southern Oscillation (Indeje et al. 2000; Carabine and Lemma 2014; Cattani et al. 2018).

2.2 Observed Precipitation Data Description and Processing

As a basis of comparison for the CFSR data, available daily precipitation datasets for locations within Kenya and Tanzania were acquired from the NOAA Climate Data Online collections (https://www.ncdc.noaa.gov/cdo-web/). Available data obtained from NOAA for Uganda locations were insufficient for the planned analysis. Thus, the data were sourced directly from the Uganda National Meteorological Authority. The period of concern for observed data was 1979–2010, consistent with the original CFSR dataset. Only stations with 10 or more years or 90% or more days within the period were included due to concerns that fewer data would be insufficient for the planned analysis. After assessing data availability and completeness, 12 stations were selected—five in Kenya, four in Uganda, and three in Tanzania (Fig. 1). Data completeness ranged from 34% for Tanzanian stations to 91% for Entebbe, Uganda. The observed data were screened for observer biases, using methods from Viney and Bates (2004) and Daly et al. (2007), and found to be free of such biases. Only years with complete data were used in annual performance evaluations, and similarly for monthly analyses (Online Resource 1, Table S1). While the CFSR V2 was included in the production of the final product, the V2 dataset was not used in primary analysis as it was more recent than the observed datasets.

2.3 CFSR Precipitation Data Description and Processing

CFSR 6-h precipitation totals were downloaded from the National Center for Atmospheric Research (NCAR, https://rda.ucar.edu/) Research Data Archive. The available spatial resolution varies from 0.3×0.3 to 2.5×2.5 degrees depending on the parameter. For the total precipitation dataset, the finest resolution available is 0.5×0.5 degrees. Daily totals were calculated by summing 24-h periods beginning at 18:00 UTC the day prior to the day of record to accommodate the time zone difference. Generally, the closest CFSR grid point to each station was selected for evaluation (Online Resource 1, Table S1). This was due to the high variation in climate and topography across a half-degree square in this study region, and hence, concern regarding the influence that surrounding grid points would have on the CFSR values used in the study. Jinja and Soroti were positioned between two proximate stations. In these cases, the grid point dataset which resulted in the highest Spearman's rank correlation coefficient when compared against the available daily data was selected.

2.4 Comparisons Between Observed and CFSR Precipitation Data

Descriptive properties (mean and median) and essential characteristics (Mehan et al. 2017) for rainfall data calculated on a monthly basis for each station were compared with those from the corresponding CFSR grid point data. Essential characteristics considered in this study included: daily maximum precipitation (Max); number of days receiving rainfall (NWet); number of days receiving rainfall in excess of the 95th and 99th percentiles of observed precipitation for each station (N95 and N99, respectively); and, the sum of precipitation received on those days in which precipitation exceeded the 95th and 99th percentiles (S95 and S99, respectively). These characteristics were selected based on their importance in previous work (Gitau 2016; Mehan et al. 2017) and their relevance to climate change and extreme weather events (Zhang et al. 2011). Additionally, plots were created of

CFSR daily precipitation quantiles against quantiles for all daily precipitation observations in months with complete data, with CFSR datasets being trimmed to match observed data periods.

Preliminary analysis showed several discrepancies between the CFSR and observed data. In general, the precipitation obtained by CFSR was more frequent and lower in magnitude, indicating the tendency for the re-analysis to spread rainfall over multiple days rather than simulating larger rainfall events. This tendency has previously been reported for CFSR precipitation (Nkiaka et al. 2017) and has also been observed with other model-based datasets (Mehan et al. 2019). For most stations, there was similarity in month-to-month patterns in median (Med) and NWet, but monthly values for the two datasets being compared were substantially different. Q-Q plots for daily rainfall revealed substantial differences between the distributions of observed and CFSR datasets. This preliminary analysis indicated the need to correct or transform the CFSR data to improve their representation of observed data. More details on the results of the comparisons are presented in Section 3.1 and Online Resource 2, Figure S1.

2.5 Evaluation and Selection of Bias Correction Methods

Based on the success and applicability of different approaches to correcting bias in climate datasets (Schmidli et al. 2006; Leander and Buishand 2007; Chen et al. 2013; Smitha et al. 2018), the sliding window approach in conjunction with one of the aforementioned bias correction methods (LOCI, PTR, EQM) was selected for further investigation as a potential solution for improving the representation from the CFSR dataset. The sliding window technique accounts for the seasonal characteristic of precipitation data as bias correction is performed over a window length of a given duration once every time step. Sliding window bias correction requires some initial precipitation data to train the correction. Bias correction was performed using the well-developed and tested downscaleR (Bedia et al. 2017), part of the climate4R bundle (Iturbide et al. 2019). DownscaleR supports several different bias correction methods and has a history of success and flexibility in application (Casanueva et al. 2019; Araya-Osses et al. 2020; Bedia et al. 2020).

Base testing of the bias correction methods was performed using the Nairobi Jomo Kenyatta International Airport (JKIA) dataset, selected because it had been used to conduct a variety of preliminary evaluations. Bias correction approaches for all available data were assessed using combinations comprising different: bias correction methods (LOCI, PTR, and EQM); window lengths (30, 45, 60, and 90 days); and, time steps (1 and 15 days)—resulting in a total of 24 combinations (naming convention: "Method Window Length–Time Step", e.g. EQM 90–15). Two subsets were formed from the observed data: a training period comprising the first 10 complete years; and, a testing period comprising the final year with complete data. These subsets were then used as inputs to the bias correction function from the downscaleR package.

The performance of the different approaches was evaluated based on the root mean squared error (RMSE), Nash–Sutcliffe efficiency coefficient (NSE), and mean absolute error (MAE) calculated for the different characteristics. The different approaches were ranked best (1) to worst (24) for each performance metric. Metric rankings for each approach were then summed to form a score. Lower scores corresponded to a better rank and better performance overall. To avoid giving undue influence to the first set of 10 years that were used for testing, the scoring was performed again using all 11 years that had complete data. Cross validation (Jolliffe and Stephenson 2012) was used to verify the robustness of the highest performing approach. This technique was chosen as it avoided the complete exclusion of any portion of the already limited observed dataset. For the cross validation, the chosen bias correction method was applied to all 11 configurations of 10 years used for training with one reserved for testing. The mean and 95% confidence interval for each month's error values were established for each statistic, with a small confidence interval considered to be an indication that there was little variation in the errors between configurations. The bias correction approach that performed best was applied to the observed data from the 12 stations. The mean and 95% confidence interval for the error for the essential characteristics from the final bias correction were compared against the error range produced in the validation of the method for JKIA to confirm that the bias correction technique performed on a similar level for all stations.

3 Results

3.1 Comparison of CFSR and Observed Data

The comparison of the CFSR dataset to the available observed data revealed discrepancies in the overall distribution of rainfall quantities, as determined from Q-Q plots (Fig. 2-I, blue; Online Resource 2, S3-S4). For Garissa, Mombasa, and Mwanza, the CFSR exhibited behavior very similar to that of Nairobi/JKIA, severely underestimating the quantities of days with higher precipitation. The Q-Q plots for several of the remaining stations showed poor fit in the moderate precipitation days, while the distribution of CFSR data for Entebbe was substantially different from that of observed data.

The total CFSR precipitation by station varied between 41.3% and 196.4% of the corresponding sum of daily precipitation observations for all complete months. Mean and range values of the absolute monthly errors (Table 1) indicated that although N95 and N99 were on average within a day or two of the observed number, there were substantial errors in the corresponding totals, S95 and S99. This was attributable to the tendency for the CFSR to predict monthly patterns correctly but incorrectly estimate the magnitude of the change or the baseline amount (Online Resource 2, Fig. S1). Nonetheless, the absolute error for the N95 and N99 could go as high as 18 and 8 days, respectively, with maximum monthly absolute errors for the corresponding extreme rainfall summations being 750 mm and 589 mm. Errors in the number of wet days were substantial, with average values ranging between 4 and 19 days and maximum monthly error up to 30 days, or essentially an entire month. Most of this is fairly consistent with overall observations on the CFSR and climate models in general (Nkiaka et al. 2017; Mehan et al. 2019). Discrepancies in maximum daily rainfall were also high with average absolute errors ranging between 10 and 28 mm, and maximum values ranging between 70 and 291 mm. In general, the highest errors in Med and NWet were observed at JKIA, while Mandera and Mwanza had the lowest errors. For the majority of metrics and stations, however, the minimum absolute error observed was zero, indicating that there were months in which the two datasets were perfectly matched. From these observations, it was determined that the CFSR precipitation dataset could potentially be refined and improved through bias correction.

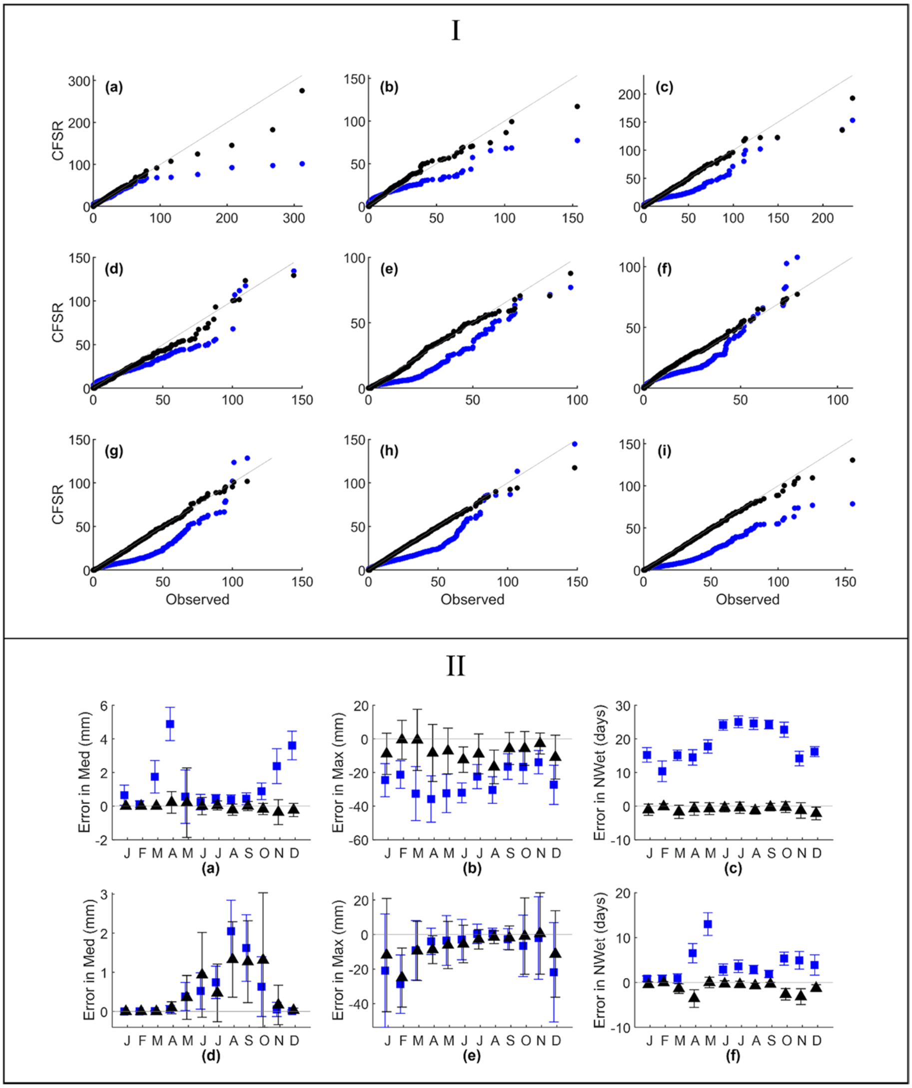

Fig. 2 (I) Q-Q plots of daily rainfall for **a** Nairobi/JKIA, **b** Garissa, **c** Mombasa, **d** Mtwara, **e** Mwanza, **f** Tabora, **g** Arua, **h** Soroti, and **i** Entebbe. (II) Mean and 95% confidence interval for monthly error in: median daily precipitation (Med) for **a** Mombasa and **d** Arua; maximum daily precipitation (Max) for **b** Entebbe and **e** Mtwara; and, number of days receiving rainfall (NWet) for **c** Nairobi /JKIA and **f** Mandera. Blue markers represent original CFSR; Black markers represent bias corrected CFSR

3.2 Bias Correction Method and Window Length Determination

Of the 24 bias correction method-window length–time step combinations evaluated (Online Resource 1, Tables S2, S3), the EQM 45–1, EQM 30–1, and EQM 30–15 approaches received the top three scores for the first decade ranking. The ranking with all 11 years resulted in EQM 30–1, EQM 45–15, and LOCI 30–15 as the top scoring approaches. In

Table 1 Range and mean of absolute errors in raw (top) and corrected (bottom) CFSR data computed on a monthly basis

	Range of errors (mean error)						
Station	Med (mm)	Max (mm)	NWet (Days)	N95 (Days)	S95 (mm)	N99 (Days)	S99 (mm)
JKIA	(0.0, 12.6) 2.8	(0.1, 291.5) 15.7	(0.0, 30.0) 18.7	(0.0, 15.0) 1.3	(0.0, 750.4) 40.7	(0.0, 8.0) 0.4	(0.0, 589.3) 43.8
	(0.0, 15.5) 0.3	(0.0, 279.5) 16.9	(0.0, 13.0) 2.6	(0.0, 15.0) 1.1	(0.0, 694.8) 39.7	(0.0, 8.0) 0.4	(0.0, 533.7) 44.3
Garissa	(0.0, 8.6) 0.7	(0.0, 131.0) 11.6	(0.0, 24.0) 7.2	(0.0, 18.0) 2.7	(0.0, 218.7) 27.0	(0.0, 3.0) 0.3	(0.0, 275.9) 33.2
	(0.0, 6.3) 0.3	(0.0, 83.9) 11.2	(0.0, 13.0) 3.7	(0.0, 7.0) 1.0	(0.0, 175.4) 25.7	(0.0, 3.0) 0.3	(0.0, 285.3) 28.2
Moyale	(0.0, 7.0) 0.4	(0.0, 103.2) 12.9	(0.0, 19.0) 5.5	(0.0, 8.0) 1.1	(0.0, 341.2) 29.7	(0.0, 5.0) 0.3	(0.0, 194.9) 12.5
	(0.0, 2.3) 0.0	(0.0, 120.7) 12.8	(0.0, 11.0) 2.2	(0.0, 8.0) 1.1	(0.0, 282.7) 22.9	(0.0, 6.0) 0.3	(0.0, 343.6) 24.5
Mombasa	(0.0, 8.2) 1.6	(0.0, 185.5) 18.8	(0.0, 24.0) 10.6	(0.0, 9.0) 1.2	(0.0, 563.7) 47.3	(0.0, 5.0) 0.3	(0.0, 400.5) 26.2
	(0.0, 10.9) 0.5	(0.0, 96.5) 16.6	(0.0, 15.0) 3.6	(0.0, 7.0) 1.0	(0.0, 408.4) 41.4	(0.0, 4.0) 0.3	(0.0, 487.9) 47.3
Mandera	(0.0, 6.0) 0.2	(0.0, 102.0) 10.4	(0.0, 22.0) 4.2	(0.0, 18.0) 1.6	(0.0, 209.6) 20.2	(0.0, 4.0) 0.3	(0.0, 276.6) 19.8
	(0.0, 3.7) 0.0	(0.0, 96.9) 10.1	(0.0, 12.0) 1.7	(0.0, 7.0) 0.9	(0.0, 232.7) 17.7	(0.0, 4.0) 0.3	(0.0, 210.0) 16.1
Mtwara	(0.0, 8.1) 1.7	(0.4, 118.0) 17.8	(1.0, 19.0) 10.9	(0.0, 7.0) 1.0	(0.0, 301.4) 41.6	(0.0, 2.0) 0.3	(0.0, 281.9) 38.6
	(0.0, 4.3) 0.3	(0.0, 108.3) 17.0	(0.0, 13.0) 2.9	(0.0, 8.0) 0.9	(0.0, 378.5) 40.6	(0.0, 2.0) 0.3	(0.0, 437.0) 43.0
Mwanza	(0.0, 2.3) 0.3	(0.0, 70.4) 20.2	(0.0, 17.0) 4.4	(0.0, 7.0) 1.2	(0.0, 232.8) 44.4	(0.0, 2.0) 0.4	(0.0, 198.2) 21.8
	(0.0, 5.5) 0.4	(0.0, 81.1) 15.3	(0.0, 13.0) 3.5	(0.0, 7.0) 0.9	(0.0, 285.5) 36.9	(0.0, 2.0) 0.4	(0.0, 350.6) 39.4
Tabora	(0.0, 7.7) 1.1	(0.0, 78.3) 14.0	(0.0, 20.0) 4.7	(0.0, 8.0) 1.0	(0.0, 264.7) 31.2	(0.0, 2.0) 0.3	(0.0, 127.0) 20.0
	(0.0, 11.4) 0.8	(0.0, 77.5) 11.9	(0.0, 15.0) 2.9	(0.0, 13.0) 1.3	(0.0, 294.6) 39.2	(0.0, 2.0) 0.3	(0.0, 383.0) 48.9
Arua	(0.0, 7.0) 0.6	(0.0, 101.6) 22.7	(0.0, 21.0) 4.8	(0.0, 8.0) 1.3	(0.0, 431.0) 53.2	(0.0, 4.0) 0.3	(0.0, 202.4) 22.5
	(0.0, 18.2) 0.7	(0.0, 87.7) 18.4	(0.0, 19.0) 3.9	(0.0, 9.0) 1.4	(0.0, 393.7) 56.8	(0.0, 3.0) 0.4	(0.0, 506.6) 54.6
Jinja	(0.0, 6.5) 1.4	(0.0, 96.0) 21.0	(0.0, 26.0) 10.7	(0.0, 8.0) 1.3	(0.0, 433.8) 58.8	(0.0, 3.0) 0.3	(0.0, 376.5) 35.1
	(0.0, 8.3) 0.5	(0.0, 86.0) 19.1	(0.0, 19.0) 4.1	(0.0, 10.0) 1.5	(0.0, 329.5) 64.2	(0.0, 3.0) 0.4	(0.0, 410.5) 68.0
Soroti	(0.0, 12.7) 1.3	(0.0, 129.0) 22.9	(0.0, 22.0) 6.9	(0.0, 7.0) 1.3	(0.0, 377.8) 49.5	(0.0, 4.0) 0.3	(0.0, 318.5) 24.2
	(0.0, 15.6) 0.8	(0.0, 108.4) 18.7	(0.0, 17.0) 4.2	(0.0, 10.0) 1.4	(0.0, 426.8) 53.8	(0.0, 4.0) 0.4	(0.0, 514.2) 55.3
Entebbe	(0.0, 7.2) 0.7	(0.6, 145.8) 28.0	(0.0, 19.0) 6.3	(0.0, 7.0) 1.4	(0.0, 376.2) 58.8	(0.0, 3.0) 0.3	(0.0, 236.3) 24.8
	(0.0, 9.3) 0.8	(0.0, 126.2) 23.6	(0.0, 17.0) 5.6	(0.0, 8.0) 1.5	(0.0, 444.5) 66.0	(0.0, 4.0) 0.5	(0.0, 500.1) 57.0

Abbreviations: *Med* Median daily precipitation, *Max* Maximum daily precipitation, *NWet* Number of wet days, *N95* Number of days above the 95th percentile, *S95* Sum of precipitation on days above the 95th percentile, *N99* Number of days above the 99th percentile, *S99* Sum of precipitation on days above the 99th percentile general, EQM 30–1 performed very well in correcting NWet, N95, and Max and performed moderately well for remaining metrics; both EQM 45–1 and 45–15 achieved slightly better results with respect to NWet but performance was sacrificed in N99 and N95, and while LOCI 30–15 achieved excellent results in mean correction, it performed comparatively poorly with NWet. From the two rankings, EQM 30–1 was found to be more robust and was selected for further use.

3.3 Evaluation of Corrected Data

Ranges and means of absolute errors in corrected data are presented in Table 1. A selection of graphs for Med, Max, and NWet showing the mean and 95% confidence interval on the error between CFSR and observed data as well as between bias corrected and observed data are presented in Fig. 2-II. Reduction in the distance between the mean error point and zero error and reductions in the span of the confidence interval were indicators of improvement. Any instances in which the mean error diverged further from zero following correction were considered inadequate with respect to the performance of the correction. For each characteristic, two graphs were included to exhibit examples of cases where the bias correction had a beneficial effect and where corrected data performed the same or worse than the uncorrected data. Corresponding graphs for the remaining characteristics are presented in Online Resource 2, Fig. S2. Heatmaps summarizing the efficacy of the bias correction based on performance metrics are presented as supplementary information (Online Resource 2, Fig. S5-S7).

Marked improvements were seen in NWet and Max for all stations (Table 1), although a slightly higher average error in Max was seen at JKIA while the range of errors increased at Moyale and Mwanza. The bias correction was particularly effective for NWet at JKIA, with its reduction in magnitude of average monthly error from approximately 19 days to less than 3 days. Results for N95 and S95 were mixed, with improvements primarily seen in Kenyan stations. For Med at Mombasa (Fig. 2-II), improvements pertained to March/April and November/December periods, capturing values corresponding to long and short rains respectively. For Max at Entebbe the underestimation for all months was improved while bias correction did not make a substantial impact for Med at Arua and Max at Mtwara (Fig. 2-II); errors in N99 were generally small and mostly unchanged for raw and corrected CFSR data. Results for monthly mean errors in S99 were, however, either not as favorable or produced mixed results (Table 1; Online Resource 2, Fig. S3-S5). Overall, the mean error of Med, Max, NWet, and N95 were brought closer to zero, even though the span of the confidence intervals did not always change substantially. Additionally, the range of total precipitation after bias correction as a percentage of the total precipitation observed ranged between 73.4%-115.4%, representing reductions of 32% and 81% in under-prediction and over-prediction errors, respectively. In general, bias correction resulted in an improvement in the data over the uncorrected dataset.

3.4 Bias-Corrected Datasets

The daily Q-Q plots for each bias corrected dataset against the available measured data (Fig. 2-I, black) showed marked improvements in the distributions of the CFSR data in relation to the distributions of the observed data, particularly in middle quantiles. For some of the stations the bias corrections resulted in a small number of missing values, impacting less than 0.3% of the data during the 1979–2010 time period. These missing values were thought to be attributable to data overflow errors occurring during the correction. As the proportion of the dataset impacted was very small, the value for any day for which the output was missing was replaced with the corresponding uncorrected CFSR value for that date. A performance comparison with and without the replaced values showed no substantial differences. Based on the performance results, bias corrected datasets were developed for the period of 1 January 1979 to 31 December 2019 for each of the 12 station locations. These datasets will be made available for public use through the Purdue University Research Repository (https://purr.lib.purdue.edu).

4 Discussion

The use of simulated rainfall data provides a solution for areas with inconsistent datasets. However, errors in the generated rainfall data are propagated through subsequent applications such as hydrologic models, water balance approximations, or characterizations of storm events and patterns that rely solely on these simulated datasets. In this study, the potential for improvements to the precipitation portion of a well-used, openly available climate dataset—CFSR—was explored. Approaches for bias correction of the dataset were evaluated through the application of 24 different combinations of bias correction methods, sliding window dimensions, and time steps. The highest performing approach was applied to CFSR data for the 12 selected stations, and results were compared against corresponding observed data to determine the extent of improvements to the preexisting dataset. Overall, the bias correction provided improved datasets based on the comparisons.

The discrepancy in the way the rainfall was distributed on a daily basis was the most important and yet particularly challenging to address. The CFSR typically underestimated higher rainfall amounts while over-estimating the number of wet days, suggesting the tendency to spread rainfall over a number of days rather than simulating large events. This was especially important due to the potential impact on the ability to capture extreme events for use with applications such as flood prediction or agricultural needs. While the three bias correction methods evaluated have a history of performing well with precipitation data (Fang et al. 2015; Luo et al. 2018), the EQM approaches performed at the highest level, overall. The EQM method used was nonparametric (Amengual et al. 2012), which was one of the contributing factors to its better performance. Since it did not rely on a particular distribution, it was able to capture more of the irregularities in the individual distributions of each station.

One reason LOCI may have lagged behind in performance is that many of the characteristics that were the focus for this study were based on precipitation extremes which LOCI does not typically capture well (Schmidli et al. 2006). It was surprising that the LOCI method did not show a better performance for the wet day counts since the method does specifically make adjustments to wet day frequency (Chen et al. 2013). A suggestion to address distribution of extrema across months within a season and disconnection between monthly corrections is through iterative bias corrections using multiple window lengths (Pierce et. al 2015). While computationally taxing, such a method may be an excellent option to consider if bias correcting for a single location. The 30–1 sliding window adopted for this study achieved the best results based on the employed metrics and is consistent with the widely accepted window length of approximately one month (Wilcke et al. 2013; Pierce et al. 2015; Ma et al. 2019).

Another approach to the development of a more robust bias correction methodology that might be considered in the future is a power transformation based on one or more climate factors. In this power transformation, parameters would be scaled as needed to reduce the difference between the observed and simulated coefficients of variation, similar to existing methods (Leander and Buishand 2007; Teutschbein and Seibert 2012) but with the added nuance of the parameters being a function of the aforementioned climate factors. Initially, it was hoped that such a methodology could be adopted in this study. However, distances between the observed data stations paired with the dramatic changes in climate and topography over relatively short distances in the study region (Mogaka et al. 2006; Rowhani et al. 2011; Foerch et al. 2015; Berakhi et al. 2015), were a serious impairment to the identification of common attributes among stations in relation to the corresponding CFSR performance.

The study methodology was strategically designed to use readily accessible tools and gridded data so it can be applied to any station with sufficient data available. Perspectives differ on what constitutes a sufficient period of observed climate data for downscaling or as a basis of manipulation, with recommended lengths ranging from as few as four years (Maurer et al. 2013) to 30 years (Mekonnen et al. 2019). In this study, a 10-year threshold was used with a comfortable exception made for Entebbe as its dataset for the 30-year period was nearly 91% complete. In data-scarce locations, it may be difficult or impossible to amass 30—or even just 10—years of data. In this study, bias correction did not necessarily perform better at stations with more data and, in some cases, the performance was worse. These results provide evidence that a strict minimum period of continuous data might not indicate the suitability of datasets for use in bias correction. Hofer et al. (2015) suggested that even small amounts of data are sufficient provided that the bias correction makes a skilled improvement. Thus, the potential and value of a bias correction is best determined on a case by case basis.

5 Conclusions

Lack of reliable and accessible precipitation data is a perpetual problem hindering effective water resources management in the study region and other data scarce regions in general. This study was conducted to determine the suitability of CFSR rainfall data as a substitute for on-site measurements in data-scarce regions and identify bias correction methods that would be appropriate to improve the accuracy of the dataset, with specific focus on the East African countries—Kenya, Uganda, and Tanzania. Results showed that the CFSR provides an easily accessible, reliable option for data-scarce regions and can be improved with bias correction to make it more representative of observed data. The sliding window technique with a 30-day window length and single-day timestep applied to Empirical Quantile Mapping (EQM 30–1) provided the best method for this application. Corrected CFSR datasets overall provided an improved representation of the observed data, particularly with respect to the number of wet days and also for the median, maximum, and 95th percentile rainfall. The resulting bias corrected datasets and the methodology outlined for optimizing a sliding window bias correction can be used to further water research and other water resource management endeavors in the study area. Additionally, the approaches used in this study provide solutions applicable to other regions where precipitation data records are scarce, incomplete, or inaccessible, and, potentially, to similar substitute datasets.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11269-021-02804-8.

Author's Contributions All authors contributed to the conceptualization of the study. V. M. Garibay and M. W. Gitau designed the framework for the analyses. V. M. Garibay conducted the analysis. V. M. Garibay and M. W. Gitau wrote the manuscript in consultation with N. Kiggundu, D. Moriasi, and F. Mishili. All authors reviewed the results and contributed to the final version of the manuscript.

Funding This work was funded in part by USDA National Institute of Food and Agriculture, Hatch Project IND00000752.

Data Availability Datasets developed through this work will be made available through the Purdue University Research Repository (PURR).

Code Availability Not applicable.

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Conflict of Interest The authors declare no potential conflicts of interest.

References

Alemayehu T, van Griensven A, Bauwens W (2016) Evaluating CFSR and WATCH data as input to SWAT for the estimation of the potential evapotranspiration in a data-scarce Eastern-African catchment. J HydrolEng 21(3):05015028. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001305

Allen MR, Dube OP, Solecki W, Aragón-Durand F, Cramer W, Humphreys S, Kainuma M, Kala J, Mahowald N, Mulugetta Y, Perez R, Wairiu M, Zickfeld K (2018) Framing and Context. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T (eds) Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. https://www.ipcc.ch/sr15/. Accessed 23 Aug 2020

Amengual A, Homar V, Romero R, Alonso S, Ramis C (2012) A statistical adjustment of regional climate model outputs to local scales: application to Platja de Palma. Spain J Clim 25(3):939–957. https://doi.org/10.1175/JCLI-D-10-05024.1

Anaba LA, Banadda N, Kiggundu N, Wanyama J, Engel B, Moriasi D (2017) Application of SWAT to assess the effects of land use change in the Murchison bay catchment in Uganda. Comp Water Energ Environ Eng 6:24–40. https://doi.org/10.4236/cweee.2017.61003

Araya-Osses D, Casanueva A, Román-Figueroa C, Uribe JM, Paneque M (2020) Climate change projections of temperature and precipitation in Chile based on statistical scaling. Clim Dyn 54(9–10):4309–4330. https://doi.org/10.1007/s00382-020-05231-4

Bedia J, Gutiérrez JM, Herrera S, Iturbide M, Manzanas R (2017) downscaleR: An R package for bias correction and statistical downscaling. https://github.com/SantanderMetGroup/downscaleR. Accessed 18 June 2019

Bedia J, Baño-Medina J, Legasa MN, Iturbide M, Manzanas R, Herrera S, Casanueva A, San-Martín D, Cofiño AS, Gutiérrez JM (2020) Statistical downscaling with the downscaleR package (v3.10): contribution to the VALUE intercomparison experiment. Geosci Model Dev 13(3):1711–1735. https://doi.org/10.5194/gmd-13-1711-2020

Berakhi RO, Oyana TJ, Adu-Prah S (2015) Land use and land cover change and its implications in Kagera river basin, East Africa. AfrGeogr Rev 34(3):209–231. https://doi.org/10.1080/19376812.2014.912140

Carabine E, Lemma A (2014) The IPCC's Fifth Assessment Report: What's in it for Africa?. Overseas Development Institute and Climate and Development Knowledge Network. https://cdkn.org/wp-content/uploads/2014/04/AR5_IPCC_Whats_in_it_for_Africa.pdf. Accessed 23 Aug 2020

Casanueva A, Kotlarski S, Herrera S, Fischer AM, Kjellstrom T, Schwierz C (2019) Climate projections of a multivariate heat stress index: the role of downscaling and bias correction. Geosci Mod Dev. https://doi.org/10.5194/gmd-12-3419-2019

Cattani E, Merino A, Guijarro JA, Levizzani A (2018) East Africa rainfall trends and variability 1983–2015 using three long-term satellite products. Remote Sens 10(6):1–26. https://doi.org/10.3390/rs10060931

Chen J, Brissette FP, Chaumont D, Braun M (2013) Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour Res 49:4187–4205. https://doi.org/10.1002/wrcr.20331

Daly C, Gibson WP, Taylor GH, Doggett MK, Smith JI (2007) Observer bias in daily precipitation measurements at United States cooperative network stations. Bull Am MeteorolSoc 88(6):899–912. https://doi.org/10.1175/BAMS-88-6-899

Dinku T, Asefa K, Hilemariam K, Grimes D, Connor S (2011) Improving availability, access and use of climate information. WMO Bull 60(2):80–86. https://library.wmo.int/doc_num.php?explnum_id=7013. Accessed 23 Apr 2020

Duan Z, Tuo Y, Liu J, Gao H, Song X, Zhang Z, Yang L, Mekonnen DF (2019) Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia. J Hydrol 569:612–626. https://doi.org/10.1016/j.jhydrol.2018.12.026

Egeru A, Osaliya R, MacOpiyo L, Mburu J, Wasonga O, Barasa B, Said M, Aleper D, Mwanjalolo GJM (2014) Assessing the spatio-temporal climate variability in semi-arid Karamoja sub-region in north-eastern Uganda. Int J Environ Stud 71(4):490–509. https://doi.org/10.1080/00207233.2014.919729

Fang GH, Yang J, Chen YN, Zammit C (2015) Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. Hydrol Earth Syst Sci 19:2547–2559. https://doi.org/10.5194/hess-19-2547-2015

Feigenwinter I, Kotlarski S, Casanueva A, Fischer AM, Schwierz C, Liniger MA (2018) Exploring quantile mapping as a tool to produce user-tailored climate scenarios for Switzerland. Technical Report MeteoSwiss No. 270. https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/service-und-publikationen/publikationen/doc/MeteoSchweiz_Fachbericht_270_final.pdf. Accessed 25 Oct 2020

Foerch G, Tenywa M, Zizinga A, Luwesi C, Mekuriaw A, Munyaneza O, Foerch N (2015) Integrated water resources management in Eastern Africa: Coping with 'complex' hydrology. Glob Water Partnership. https://doi.org/10.13140/RG.2.1.5054.2164

Gebrechorkos SH, Hülsmann S, Bernhofer C (2018) Evaluation of multiple climate data sources for managing environmental resources in East Africa. Hydrol Earth Syst Sci 22:4547–4564. https://doi.org/10.5194/hess-22-4547-2018

Gitau M (2016) Long-term seasonality of rainfall in the southwest Florida Gulf coastal zone. Clim Res 69:93–105. https://doi.org/10.3354/cr01399

Githui FW (2008) Assessing the impacts of environmental change on the hydrology of the Nzoia catchment, in the Lake Victoria Basin. PhD Thesis, Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Belgium. https://www.oceandocs.org/bitstream/handle/1834/6857/ktf0244.pdf?sequence=1. Accessed 8 Aug 2020

Graham LP, Andréasson J, Carlsson B (2007) Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin. Clim Chang 81:293–307. https://doi.org/10.1007/s10584-006-9215-2

Hofer M, Marzeion B, Mölg T (2015) A statistical downscaling method for daily air temperature in data-sparse, glaciated mountain environments. Geosci Model Dev 8:579–593. https://doi.org/10.5194/gmd-8-579-2015

Indeje M, Semazzi FHM, Ogallo LJ (2000) Enso signals in East African rainfall seasons. Int J Climatol 20:19–46. https://doi.org/10.1002/(SICI)1097-0088(200001)20:1%3c19::AID-JOC449%3e3.0.CO;2-0

Iturbide M, Bedia J, Herrera S, Baño-Medina J, Fernández JJ, Frías MD, Manzanas R, San-Martín D, Cimadevilla E, Cofiño AS, Gutiérrez JM (2019) The R-based climate4R open framework for reproducible climate data access and post-processing. Environ Model Softw 111:42–54. https://doi.org/10.1016/j.envsoft.2018.09.009

Jolliffe IT, Stephenson DB (2012) Introduction. In: Jolliffe IT, Stephenson DB (eds) Forecast verification: a practitioner's guide in atmospheric science, 2nd edn. John Wiley and Sons, Ltd., Chichester, pp 1–9

Jury MR (2018) Uganda rainfall variability and prediction. Theor Appl Climatol 132:905–919. https://doi.org/10.1007/s00704-017-2135-4

Lakew HB, Moges SA, Anagnostou EN, Nikolopoulos EI, Asfaw DH (2020) Evaluation of global water resources reanalysis runoff products for local water resources applications: case study-upper blue Nile Basin of Ethiopia. Water Resour Manage 34:2157–2177. https://doi.org/10.1007/s11269-019-2190-y
Leander R, Buishand TA (2007) Resampling of regional climate model output for the simulation of extreme river flows. J Hydrol 332:487–496. https://doi.org/10.1016/j.jhydrol.2006.08.006
López PL, Sultana T, Kafi MAH, Hossain MS, Khan AS, Masud MS (2020) Evaluation of global water resources reanalysis data for estimating flood events in the Brahmaputra River Basin. Water Resour Manage 34:2201–2220. https://doi.org/10.1007/s11269-020-02546-z
Lugoi LP, Bamutaze Y, Martinsen V, Dick ØB, Almås ÅR (2019) Ecosystem productivity response to environmental forcing, prospect for improved rain-fed cropping productivity in lake Kyoga Basin. Appl Geogr 102:1–11. https://doi.org/10.1016/j.apgeog.2018.11.001
Luo M, Liu T, Meng F, Duan Y, Frankl A, Bao A, De Maeyer P (2018) Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the Kaidu River Basin in Western China. Water 10(8):1046. https://doi.org/10.3390/w10081046
Ma Q, Xiong L, Xia J, Xiong B, Yang H, Xu C-Y (2019) A censored shifted mixture distribution mapping method to correct the bias of daily IMERG satellite precipitation estimates. Remote Sens 11(11):1345. https://doi.org/10.3390/rs11111345
Mainya J (2017) Modelling Runoff, Soil Erosion and Sediment Yield in Sosiani Catchment in Kenya Using Arcswat. Thesis, Institute of Water and Energy Sciences, Pan African University, Algeria. http://repository.pauwes-cop.net/bitstream/handle/1/139/MT_Johnstone%20Mainya.pdf?sequence=1. Accessed 14 Aug 2020
Maurer EP, Das T, Cayan DR (2013) Errors in climate model daily precipitation and temperature output: time invariance and implications for bias correction. Hydrol Earth Syst Sci 17:2147–2159. https://doi.org/10.5194/hess-17-2147-2013
Mehan S, Gitau MW, Flanagan DC (2019) Reliable future climatic projections for sustainable hydrometeorological assessments in the Western Lake Erie Basin. Water 11(3):581. https://doi.org/10.3390/w11030581
Mehan S, Guo T, Gitau MW, Flanagan DC (2017) Comparative study of different stochastic weather generators for long-term climate data simulation. Climate 5(2):26. https://doi.org/10.3390/cli5020026
Mekonnen DG, Moges MA, Mulat AG, Shumitter P (2019) The impact of climate change on mean and extreme state of hydrological variables in Megech watershed, Upper Blue Nile Basin, Ethiopia. In: Melesse AF, Abtew W, Senay G (eds) Extreme Hydrology and Climate Variability, First Edition:123–135. Elsevier. https://doi.org/10.1016/B978-0-12-815998-9.00011-7
Mogaka H, Gichere S, Davis R, Hirji R (2006) Climate Variability and Water Resources Degradation in Kenya: Improving Water Resources Development and Management. The World Bank, Working Paper No. 69. https://doi.org/10.1596/978-0-8213-6517-5
Muthuwatta L, Sood A, McCartney M, Silva NS, Opere A (2018) Understanding the impacts of climate change in the Tana River Basin, Kenya. Proc IAHS 379:37–42. https://doi.org/10.5194/piahs-379-37-2018
Nkiaka E, Nawaz NR, Lovett JC (2017) Evaluating global reanalysis precipitation datasets with rain gauge measurements in the Sudano-Sahel region: case study of the Logone catchment, Lake Chad Basin. Meteorol Appl 24:9–18. https://doi.org/10.1002/met.1600
Pierce DW, Cayan DR, Maurer EP, Abatzoglou JT, Hegewisch KC (2015) Improved bias correction techniques for hydrological simulations of climate change. J Hydrometeor 16(6):2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1
R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 22 Aug 2018
Rowhani P, Lobell DB, Linderman M, Ramankutty N (2011) Climate variability and crop production in Tanzania. Agric For Meteorol 151:449–460. https://doi.org/10.1016/j.agrformet.2010.12.002
Saha S et al (2010) NCEP Climate Forecast System Reanalysis (CFSR) 6-hourly Products, January 1979 to December 2010. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.https://doi.org/10.5065/D69K487J
Saha S et al (2011) Updated daily. NCEP Climate Forecast System Version 2 (CFSv2) 6-hourly Products. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D61C1TXF
Schmidli J, Frei C, Vidale PL (2006) Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. Int J Climatol 26:679–689. https://doi.org/10.1002/joc.1287

Schmocker J, Liniger HP, Ngeru JN, Brugnara Y, Auchmann R, Brönnimann S (2016) Trends in mean and extreme precipitation in the Mount Kenya region from observations and reanalyses. Int J Climatol 36:1500v1514. https://doi.org/10.1002/joc.4438

Smitha PS, Narisimhan B, Sudheer KP, Annamalai H (2018) An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. J Hydrol 556:100–118. https://doi.org/10.1016/j.jhydrol.2017.11.010

Tesfaye TW, Dhanya CT, Gosain AK (2017) Evaluation of ERA-Interim, MERRA, NCEP-DOE R2 and CFSR Reanalysis precipitation Data using Gauge Observation over Ethiopia for a period of 33 years. AIMS Environ Sci 4(4):596–620. https://doi.org/10.3934/environsci.2017.4.596

Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J Hydrol 456–457:12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Viney NR, Bates BC (2004) It never rains on sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. Int J Climatol 24:1171–1192. https://doi.org/10.1002/joc.1053

Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of regional climate models. Clim Chang 120:871–887. https://doi.org/10.1007/s10584-013-0845-x

Worqlul AW, Yen H, Collick AS, Tilahun SA, Langan S, Steenhuis TS (2017) Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: The upper Blue Nile Basin, Ethiopia. CATENA 152:242–251. https://doi.org/10.1016/j.catena.2017.01.019

Zhan W, Guan K, Sheffield J, Wood EF (2016) Depiction of drought over sub-Saharan Africa using reanalyses precipitation data sets. J Geophys Res Atmos 121:10,555-10,574

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. Wiley Interdiscip Rev Clim Chang 2:851–870. https://doi.org/10.1002/wcc.147