Retrospective Analysis and Bayesian Model Averaging of CMIP6 Precipitation in the Nile River Basin

Mohammed Ombadi, Phu Nguyen, Soroosh Sorooshian, and Kuo-Lin Hsu

Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, California

Department of Earth System Science, University of California Irvine, Irvine, California

(Manuscript received 25 June 2020, in final form 2 October 2020)

ABSTRACT: The Nile River basin is one of the global hotspots vulnerable to climate change impacts because of a fast-growing population and geopolitical tensions. Previous studies demonstrated that general circulation models (GCMs) frequently show disagreement in the sign of change in annual precipitation projections. Here, we first evaluate the performance of 20 GCMs from phase six of the Coupled Model Intercomparison Project (CMIP6) benchmarked against a high-spatial-resolution precipitation dataset dating back to 1983 from Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR). Next, a Bayesian model averaging (BMA) approach is adopted to derive probability distributions of precipitation projections in the Nile basin. Retrospective analysis reveals that most GCMs exhibit considerable (up to 64% of mean annual precipitation) and spatially heterogeneous bias in simulating annual precipitation. Moreover, it is shown that all GCMs underestimate interannual variability; thus, the ensemble range is under-dispersive and is a poor indicator of uncertainty. The projected changes from the BMA model show that the value and sign of change vary considerably across the Nile basin. Specifically, it is found that projected changes in the two headwaters basins, namely, the Blue Nile and Upper White Nile, are 0.03% and −1.65%, respectively; both are statistically insignificant at α = 0.05. The uncertainty range estimated from the BMA model shows that the probability of a precipitation decrease is much higher in the Upper White Nile basin whereas projected change in the Blue Nile is highly uncertain both in magnitude and sign of change.

KEYWORDS: Africa; Bayesian methods; Bias; Climate models; General circulation models; Model evaluation/performance

1. Introduction

The Nile River basin constitutes approximately 10% of the African continent (Swain 2008) extending across eleven countries. A total population of 462 million in these countries is growing at an annual growth rate of 2.5%, faster than the average global growth rate estimated at 1.1%. Consequently, the population of these countries is projected to reach 836 million (81% increase) by the year 2050 (The World Bank 2018, 2020). A key challenge, therefore, that face these countries is to sustain the burgeoning food and energy demand of this growing population. Water lies at the heart of natural resources that play a pivotal role in securing this demand. Therefore, assessment of climate change impacts on precipitation is important due to its direct effect on water availability in headwaters countries as well as its impact on the Nile streamflow yield that is the main source of water for riparian countries, namely, Sudan and Egypt.

Several studies have been devoted to the assessment of climate change impacts on precipitation in the Nile River basin and its headwaters basins (e.g., Conway 1996; Yates and Strzepek 1996, 1998; Kim and Kaltuus 2009; Elshamy et al. 2009; Bhattacharjee and Zaitchik 2015; Fenta Mekonnen and Disse 2018). Earlier studies found that general circulation models (GCMs) frequently show disagreement in the sign of change of annual precipitation projections. For instance, Conway (1996) used three GCMs to assess climate change impact on precipitation in the Blue Nile and Lake Victoria subbasins; results showed that percentage change in precipitation ranges from −1.9% to 7.4% in the two subbasins.

More recently, Kim and Kaltuus (2009) showed that mean annual precipitation in the upper Blue Nile subbasin is projected to increase by 11% based on a weighted average of six GCMs’ outcomes. On the contrary, Elshamy et al. (2009) reported the outcomes of 17 GCMs and showed that projected change in mean annual precipitation in the upper Blue Nile subbasin ranges from −15% to +14% with more models reporting a decrease in precipitation. These results, among others, emphasize that there is a wide uncertainty and intermodel differences in precipitation projections, and they indicate that a consensus on how climate change will impact water resources in the Nile basin is yet to be reached.

Two different approaches are commonly adopted to treat uncertainty of GCMs. At one end of the spectrum is the ensemble mean that overlooks historical performance of the models and assigns equal weights to all models. At the other end, there is an approach that selects a number of best-performing models and discards the remaining ones. The former is less accurate at regional scales and in cases in which there is a spread in model projections (Schaller et al. 2011) whereas the latter is highly dependent on the specific metrics chosen for performance evaluation (Schaller et al. 2011; Bhattacharjee and Zaitchik 2015). Between these two extremes lies the approach of model averaging in which models are neither weighted equally nor some of them are discarded entirely. Specifically, model averaging methods take advantage of retrospective analysis of GCMs simulations benchmarked against observations, and they assign weights to models according to their performance. A major issue, however, that lessens the effectiveness of such methods is the dearth of quality controlled, dense gauge precipitation observations in the Nile basin. Here, we surmount this issue by resorting to high

DOI: 10.1175/JHM-D-20-0157.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
spatial resolution and long record of historical observations provided from Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR; Ashouri et al. 2015). PERSIANN-CDR is a high-spatial-resolution satellite-based dataset that is bias adjusted using gauge observations at the monthly scale; thus, providing a unique dataset for retrospective analysis of GCMs.

To this end, the focus of the present study is to first evaluate the performance of 20 GCMs from phase 6 of the Coupled Model Intercomparison Project (CMIP6) against PERSIANN-CDR over the Nile basin. Next, a model averaging approach, namely, Bayesian model averaging (BMA; Raftery et al. 2005) is implemented to derive probability distributions of precipitation projections in the Nile basin for the future period (2015–2100). The remainder of this paper is organized as follows. Section 2 provides a brief description of the data used in this study. Section 3 describes the bias adjustment and Bayesian model averaging approaches used to postprocess precipitation projections of CMIP6 GCMs. Section 4 presents the results of retrospective analysis as well as future projections of precipitation in the Nile basin. Section 5 sums up the findings of the study and draws conclusions.

2. Data and study area
a. CMIP6

Many climate models participating in CMIP6 have reported their simulations for the different CMIP6 experiments. In the present study, two experiments are of concern: historical and the future scenario that corresponds to high greenhouse gas emissions, and it is the equivalent to RCP8.5 “business as usual” scenario in CMIP5. Currently, a set of 20 models have reported their simulations for both historical and SSP5-8.5 experiments. These models have been used in this study to examine climate change impact on precipitation in the Nile basin, and they are shown in Table 1. For each model, we only consider the first ensemble member for future projections under SSP5-8.5. Also, we consider the dataset at monthly temporal resolution for both historical and SSP5-8.5.

b. PERSIANN-CDR

PERSIANN-CDR (Ashouri et al. 2015; see also Nguyen et al. 2018) is a satellite-based precipitation dataset based on infrared (IR) imagery. It has near-global coverage (60°S–60°N) with a spatial resolution of 0.25° × 0.25° and a daily temporal resolution. PERSIANN-CDR is suitable for climatic studies because of its long record of 37+ years (from 1983 to delayed present). It is particularly advantageous because it is bias adjusted using Global Precipitation Climatology Project (GPCP) monthly 2.5° × 2.5° precipitation data (Adler et al. 2018). Therefore, it maintains monthly precipitation at 2.5° × 2.5° that is consistent with GPCP while capturing spatial rainfall variability at higher latitudes.

---

Table 1. CMIP6 models used in the present study and their spatial resolution. An asterisk indicates approximate resolution since the native resolution is not in regular grids. Expansions of many model acronyms can be found online (https://www.ametsoc.org/PubsAcronymList).

<table>
<thead>
<tr>
<th>Model</th>
<th>Institute</th>
<th>Resolution (Lat° × Lon°*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth3</td>
<td>EC-Earth Consortium, Europe</td>
<td>0.702 × 0.703*</td>
</tr>
<tr>
<td>Earth3-Veg</td>
<td>EC-Earth Consortium, Europe</td>
<td>0.702 × 0.703*</td>
</tr>
<tr>
<td>MPI-ESM1.2-HR</td>
<td>Max Planck Institute for Meteorology, Germany</td>
<td>0.935 × 0.937*</td>
</tr>
<tr>
<td>CESM2(WACCM)</td>
<td>National Center for Atmospheric Research, United States</td>
<td>0.942 × 1.25</td>
</tr>
<tr>
<td>FIO-ESM-2.0</td>
<td>First Institute of Oceanography-Qingdao National Laboratory for Marine Science and Technology (FIO-QLNM), China</td>
<td>0.942 × 1.25</td>
</tr>
<tr>
<td>NorESM2-MM</td>
<td>NorESM Climate Modeling Consortium (NCC), Norway</td>
<td>0.942 × 1.25</td>
</tr>
<tr>
<td>FGOALS-f3-L</td>
<td>Chinese Academy of Sciences, China</td>
<td>1 × 1.25</td>
</tr>
<tr>
<td>BCC_CSM2(m)</td>
<td>Beijing Climate Center, China</td>
<td>1.121 × 1.125*</td>
</tr>
<tr>
<td>MIROC6</td>
<td>Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, The University of Tokyo, National Institute for Environmental Studies, and RIKEN Center for Computational Science (MIROC), Japan</td>
<td>1.4 × 1.406</td>
</tr>
<tr>
<td>ACCESS-CM2</td>
<td>Commonwealth Scientific and Industrial Research Organisation–Australian Research Council Centre of Excellence for Climate System Science (CSIRO-ARCCSS), Australia</td>
<td>1.25 × 1.875</td>
</tr>
<tr>
<td>ACCESS-ESM1.5</td>
<td>CSIRO, Australia</td>
<td>1.25 × 1.875</td>
</tr>
<tr>
<td>KAGE-1.0-G</td>
<td>National Institute of Meteorological Sciences/Korea Meteorological Administration (NIMS-KMA), Republic of Korea</td>
<td>1.25 × 1.875</td>
</tr>
<tr>
<td>INM-CM4.8</td>
<td>Institute for Numerical Mathematics, Russia</td>
<td>1.5 × 2</td>
</tr>
<tr>
<td>INM-CM5.0</td>
<td>Institute for Numerical Mathematics, Russia</td>
<td>1.5 × 2</td>
</tr>
<tr>
<td>IPSL-CM6a-LR</td>
<td>Institut Pierre Simon Laplace, France</td>
<td>2.128 × 2.5</td>
</tr>
<tr>
<td>MPI-ESM1.2-LR</td>
<td>Max Planck Institute for Meteorology, Germany</td>
<td>1.865 × 1.875*</td>
</tr>
<tr>
<td>NESM3</td>
<td>Nanjing University of Information Science and Technology, China</td>
<td>1.865 × 1.875*</td>
</tr>
<tr>
<td>FGOALS-g3</td>
<td>Chinese Academy of Sciences, China</td>
<td>2.279 × 2*</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>NCC, Norway</td>
<td>2.895 × 2.5</td>
</tr>
<tr>
<td>CanESM5</td>
<td>Canadian Centre for Climate Modeling and Analysis, Canada</td>
<td>2.789 × 2.813*</td>
</tr>
</tbody>
</table>
spatial resolution. This last point emphasizes that PERSIANN-CDR has sufficient credibility to be used as a baseline dataset for evaluation of CMIP6 GCMs. PERSIANN-CDR is widely used for a range of hydrologic and hydroclimatic studies (e.g., Ombadi et al. 2018; Nguyen et al. 2020), and it has previously been used for evaluation of GCMs (Nguyen et al. 2017). Here, we use PERSIANN-CDR at monthly temporal accumulations.

c. Study area

In this study, we consider the entire Nile basin for our analysis (shown in Fig. 1; gray lines). The analysis is performed at the grid scale 1° × 1° because of the wide variability of climate and precipitation regimes in the Nile basin. This variability is clearly shown in Fig. 1 with the south-to-north gradient in precipitation that represents the variability in climate from tropical humid in the south to hyperarid in the north. Throughout this study, we carry out the analysis at the grid scale and then aggregate the results at the entire Nile basin as well as its headwaters basins, namely, the Blue Nile and Upper White Nile basins (gray lines in Fig. 1). We focus on these two subbasins because of their significant contribution to the Nile streamflow yield.

3. Methods

a. Bias adjustment

CMIP6 model simulations and PERSIANN-CDR data were first regridded to a common spatial resolution of 1° × 1° using bilinear interpolation. Bias adjustment coefficients were then calculated for each grid from the historical simulations (1983–2014) according to the following linear model:

\[ y^H = a + b f^H, \]  

where \( y^H \) is PERSIANN-CDR annual (or monthly) precipitation time series at a given grid for the period (1983–2014), and \( f^H \) is the corresponding annual (or monthly) precipitation from the \( k \)th GCM model; the superscript \( H \) refers to “historical,” and \( a \) and \( b \) are the bias adjustment coefficients.

b. Bayesian model averaging

1) General description

Bayesian model averaging (Raftery et al. 2005; see also Duan et al. 2007; Ajami et al. 2007) aims to reduce multimodel uncertainty by assigning weights to all models, with the weights representing posterior probabilities of the models given historical observations. BMA has previously been used to derive probability distributions of continental precipitation and temperature projections from a CMIP3 multimodel ensemble (Duan and Phillips 2010). The BMA predictive distribution is a weighted sum of conditional probability distributions of individual models. Let us consider the same notations used earlier and denote by \( f_k \) annual (or monthly) precipitation projections of the \( k \)th model. BMA yields the following predictive model:

\[ p(y|f_1, f_2, \ldots, f_K) = \sum_{k=1}^{K} w_k p_k(y|f_k), \]
where \( y \) is the sought-after precipitation projections. The left-hand side represents the probability density function (pdf) of the BMA model that is equal to a weighted sum of the individual conditional pdfs of models 1, 2, \ldots, \( K \). As noted earlier, the weights \( w_k \) represent posterior probabilities of models conditioned on historical observations; thus, they sum to 1. The pdfs \( p_k \) for \( k = 1, 2, \ldots, K \) are commonly assumed to be normal distributions, which is the case in the present study. The weights \( w_k \) are estimated by maximizing the log-likelihood function of the pdf in the left-hand side using historical observations. Put simply, \( y^{H} \) and \( f_k^{H} \) are substituted for \( y \) and \( f_k \), respectively, in Eq. (2) in order to estimate \( w_k \). Several techniques such as the expectation-maximization algorithm (Dempster et al. 1977) can be used to converge to a local maximum of the log-likelihood function. Here, we use a differential evolution–Markov chain (DE-MC) algorithm (Ter Braak 2006) to find the optimum values of \( w_k \).

2) DE-MC ALGORITHM AND ESTIMATION OF BMA WEIGHTS

The DE-MC algorithm (Ter Braak 2006) combines the genetic algorithm variant of differential evolution (Storn and Price 1997) and the sampling techniques of Markov chain Monte Carlo for optimization over real parameter space. The essential idea of DE-MC is that a large number of \( N \) chains, each consists of a randomly sampled vector of parameters, are run in parallel. The \( N \) chains exchange information with each other according to the simple rules of DE, and they are updated sequentially to provide a new candidate solution at each step. This process continues for a number of \( T \) steps that is often referred to as the number of generations, in connection with the natural selection basis of genetic algorithms. Both \( N \) and \( T \) are hyperparameters specified for the algorithm, and they are chosen to be 500 and 10,000, respectively in the present study. The choice of hyperparameters can be judged tentatively by tracking the convergence rate of the algorithm across generations, and the chosen values of hyperparameters were found to be optimal in this study.

At each generation of \( T \), the algorithm evaluates the fitness of the \( N \) population members using an objective function. In the present study, the objective function is obtained from Eq. (2) modified by substituting \( y^{H} \) and \( f_k^{H} \) for \( y \) and \( f_k \), respectively. Specifically, the conditional probability distribution \( p_k(y^{H} | f_k^{H}) \) of the \( k \)th model is assumed to be a normal distribution of the following form:

\[
p_k(y^{H}|f_k^{H}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y^{H}(t) - f_k^{H}(t))^2 \right\},
\]

where \( \sigma^2 \) is the variance of the normal distribution and is considered as a parameter to be optimized by the DE-MC, whereas \( y^{H}(t) \) and \( f_k^{H}(t) \) are the annual historical PERSIANN-CDR observations and simulations of the \( k \)th model, respectively, for the year \( t \). For numerical stability, the objective function is considered as a log likelihood of the following form:

\[
\mathcal{L}(y^{H} | f_1^{H}, f_2^{H}, \ldots, f_K^{H}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} w_k p_k(y^{H} | f_k^{H}),
\]

where \( \mathcal{L} \) refers to the log likelihood of the distribution and \( n \) is the number of years in the historical period (1983–2014). The log likelihood in Eq. (4) is used as an objective function by the DE-MC algorithm to evaluate the fitness of the \( N \) population members at each generation. Each candidate solution consists of the vector of parameters \( w_k \) in addition to the variance \( \sigma^2 \). Note that all of the foregoing steps are implemented for each spatial grid in the study area separately.

3) COMPUTATION OF BMA UNCERTAINTY RANGE

One of the main advantages of the BMA approach is offering a posterior distribution instead of a single expected value (i.e., mean). Consequently, this distribution can be sampled to obtain uncertainty ranges with regard to precipitation projections. Following the estimation of the parameters \( w_k \) and \( \sigma^2 \) as illustrated in the previous section, the posterior distribution of precipitation projections is given by Eq. (2) after substituting the obtained values of parameters. To sample this posterior distribution, we follow two simple steps: first, a number in the range 1, 2, \ldots, \( K \) is drawn randomly with probabilities \( w_1, w_2, \ldots, w_K \); second, a random sample is drawn from the distribution \( p_k(y|f_k) \). These two steps are repeated 1000 times to approximate the posterior distribution at each spatial grid.

c. Evaluation metrics

In the present study, we first evaluate the performance of CMIP6 GCMs in simulating annual precipitation for the historical period (1983–2014). This is carried out using metrics of bias, relative bias, seasonality and spatial correlations, coefficient of variability and rank histogram. Bias in this study is calculated as the difference given by mean annual precipitation of CMIP6 GCM minus mean annual precipitation of PERSIANN-CDR. Thus, positive and negative values indicate overestimation and underestimation, respectively. Relative bias is obtained by normalizing the bias; specifically, dividing by mean annual precipitation of PERSIANN-CDR.

Seasonality correlation is computed as Pearson correlation coefficient between the climatology of monthly CMIP6 GCMs and PERSIANN-CDR precipitation whereas the spatial correlation is computed as Pearson correlation coefficient between mean annual precipitation of CMIP6 GCMs and PERSIANN-CDR for all spatial grids within a basin. The coefficient of variation for CMIP6 GCMs and PERSIANN-CDR is computed as the ratio of the standard deviation of annual precipitation to its mean. Finally, the rank histogram (Hamill 2001) is an efficient and convenient method to assess the reliability of ensemble forecasts. Its fundamental idea is to keep track of the rank of observed precipitation with respect to forecasts of ensemble members at each time step of forecast; these ranks are then used to construct the histogram. If the ensemble range effectively captures uncertainty, then the histogram is expected to be uniform. On the other hand, if the ensemble range is too narrow (wide), the rank histogram will be concentrated near the ends (center). Note that, in the evaluation of historical model performance as well as in future projections, we also examine the performance of ensemble mean of the 20 CMIP6 GCMs. This ensemble mean is an arithmetic average of the 20 GCMs, and it...
should not be confused with the BMA mean. The latter is the mean of the posterior distribution obtained from Bayesian model averaging and is only used in future projections.

4. Results and discussion


1) BIAS IN ANNUAL PRECIPITATION

We first examine the performance of the different GCMs in simulating the mean value of annual precipitation for the baseline period (1983–2014). Figure 2 shows the bias in spatially averaged annual precipitation over the Nile, Blue Nile, and Upper White Nile basins for each GCM as well as the ensemble mean with respect to PERSIANN-CDR. There is a clear spread between the models with a bias range from $-2430$ to $389$ mm, from $-2619$ to $661$ mm, and from $-2738$ to $791$ mm in the Nile, Blue Nile, and Upper White Nile basins, respectively (see Table 2). These biases are significant since they represent up to 64%, 61%, and 64% of mean annual precipitation in the three basins, respectively. Although the ensemble mean reduces the biases, it fails to outperform the best performing model in the three basins.

Figure 3 shows the biases proportional to mean annual precipitation (i.e., relative bias) of the 20 GCMs and the ensemble mean for each grid (1° × 1°) in the Nile basin. Apart from intermodel differences in bias, Fig. 3 shows that there is a considerable spatial variability in bias within individual models. The values of relative bias over large areas of the basin exceed ±0.3 (stippled grids in Fig. 3), which underscores the importance of bias adjustment of GCM outputs prior to evaluation of future projections. In addition to examining the ability of GCMs in simulating the amount of total precipitation in the basins, it is important to investigate their accuracy in simulating the spatial patterns of precipitation. Table 2 shows the spatial correlation coefficient of the 20 GCMs and the ensemble mean against PERSIANN-CDR. This reflects how well each model represents the spatial variability of annual precipitation within the Nile basin and its two headwaters basins. Clearly, all the models fairly represent the spatial variability of annual precipitation within the Nile basin as evidenced by correlation coefficients greater than 0.8. Furthermore, the ability of the models to represent spatial variability within the Blue Nile basin is quite reasonable with a minimum correlation coefficient of 0.58. However, the correlation of spatial variability within the Upper White Nile basin is drastically lower, with many models showing a negative correlation, and a maximum correlation coefficient of only 0.49. This highlights that while the GCMs performance in terms of bias is comparable in the Nile basin and its headwaters basins, the GCMs specifically underperform in the Upper White Nile basin with regard to representation of spatial variability. We speculate that the lower performance of GCMs in simulating the spatial variability of precipitation in the Upper White Nile basin is due to the complex rainfall regime in this region. In addition to mechanisms such as monsoonal winds that modulate rainfall in East Africa (i.e., Blue Nile basin), this region is also affected by the interplay of several factors including the intertropical convergence zone (ITCZ), El Niño–Southern Oscillation, and the quasi-biennial oscillation, among others (Nicholson and Yin 2002; Dezfuli 2017).

2) INTERANNUAL VARIABILITY AND UNCERTAINTY

Figure 4a shows the annual precipitation coefficient of variation (ratio of standard deviation to mean) for the 20 GCMs
Table 2. Evaluation of CMIP6 GCMs precipitation against PERSIANN-CDR over the entire Nile, Blue Nile (B Nile), and Upper White Nile (W Nile) basins. Boldface type indicates the best-performing model according to the metric under consideration. An em dash indicates that the correlation coefficient is negative.

<table>
<thead>
<tr>
<th>Model</th>
<th>Nile Bias (mm)</th>
<th>B Nile Bias (mm)</th>
<th>W Nile Bias (mm)</th>
<th>Nile Spatial correlation</th>
<th>B Nile Spatial correlation</th>
<th>W Nile Spatial correlation</th>
<th>Nile Seasonality correlation</th>
<th>B Nile Seasonality correlation</th>
<th>W Nile Seasonality correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS-CM2</td>
<td>−239</td>
<td>−615</td>
<td>−238</td>
<td>0.81</td>
<td>0.82</td>
<td>—</td>
<td>0.94</td>
<td>0.97</td>
<td>0.55</td>
</tr>
<tr>
<td>ACCESS-ESM1.5</td>
<td>389</td>
<td>497</td>
<td>696</td>
<td>0.89</td>
<td>0.9</td>
<td>—</td>
<td>0.91</td>
<td>0.94</td>
<td>0.55</td>
</tr>
<tr>
<td>BCC_CSM2(m)</td>
<td>157</td>
<td>−6</td>
<td>517</td>
<td>0.87</td>
<td>0.88</td>
<td>—</td>
<td>0.96</td>
<td>0.99</td>
<td>0.75</td>
</tr>
<tr>
<td>CanESM5</td>
<td>63</td>
<td>−56</td>
<td>190</td>
<td>0.83</td>
<td>0.91</td>
<td>—</td>
<td>0.93</td>
<td>0.96</td>
<td>0.55</td>
</tr>
<tr>
<td>CESM2(WACCM)</td>
<td>61</td>
<td>205</td>
<td>163</td>
<td>0.91</td>
<td>0.9</td>
<td>0.11</td>
<td>0.91</td>
<td>0.96</td>
<td>0.68</td>
</tr>
<tr>
<td>Earth3</td>
<td>−31</td>
<td>−65</td>
<td>61</td>
<td>0.9</td>
<td>0.79</td>
<td>0.49</td>
<td>0.88</td>
<td>0.99</td>
<td>0.5</td>
</tr>
<tr>
<td>Earth3-Veg</td>
<td>−68</td>
<td>−138</td>
<td>18</td>
<td>0.89</td>
<td>0.78</td>
<td>0.47</td>
<td>0.89</td>
<td>0.99</td>
<td>0.54</td>
</tr>
<tr>
<td>FGOALS-f3-L</td>
<td>−430</td>
<td>−619</td>
<td>−738</td>
<td>0.85</td>
<td>0.87</td>
<td>0.09</td>
<td>0.87</td>
<td>0.97</td>
<td>0.5</td>
</tr>
<tr>
<td>FGOALS-g3</td>
<td>−261</td>
<td>−242</td>
<td>−446</td>
<td>0.68</td>
<td>0.58</td>
<td>—</td>
<td>0.81</td>
<td>0.95</td>
<td>0.54</td>
</tr>
<tr>
<td>FIO-ESM-2.0</td>
<td>249</td>
<td>179</td>
<td>377</td>
<td>0.9</td>
<td>0.93</td>
<td>0</td>
<td>0.81</td>
<td>0.72</td>
<td>0.59</td>
</tr>
<tr>
<td>INM-CM4.8</td>
<td>131</td>
<td>−98</td>
<td>688</td>
<td>0.86</td>
<td>0.83</td>
<td>0.27</td>
<td>0.88</td>
<td>0.95</td>
<td>0.72</td>
</tr>
<tr>
<td>INM-CM5.0</td>
<td>90</td>
<td>−184</td>
<td>525</td>
<td>0.87</td>
<td>0.86</td>
<td>0.12</td>
<td>0.87</td>
<td>0.9</td>
<td>0.77</td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>87</td>
<td>−30</td>
<td>315</td>
<td>0.86</td>
<td>0.85</td>
<td>—</td>
<td>0.9</td>
<td>0.94</td>
<td>0.63</td>
</tr>
<tr>
<td>KAGE-1.0-G</td>
<td>1</td>
<td>−348</td>
<td>374</td>
<td>0.85</td>
<td>0.8</td>
<td>0.08</td>
<td>0.95</td>
<td>0.99</td>
<td>0.52</td>
</tr>
<tr>
<td>MIROC6</td>
<td>326</td>
<td>661</td>
<td>791</td>
<td>0.87</td>
<td>0.73</td>
<td>0.18</td>
<td>0.98</td>
<td>0.98</td>
<td>0.71</td>
</tr>
<tr>
<td>MPI-ESM1.2-HR</td>
<td>−198</td>
<td>−405</td>
<td>−193</td>
<td>0.85</td>
<td>0.84</td>
<td>0.22</td>
<td>0.9</td>
<td>0.93</td>
<td>0.76</td>
</tr>
<tr>
<td>MPI-ESM1.2-LR</td>
<td>−210</td>
<td>−391</td>
<td>−349</td>
<td>0.91</td>
<td>0.95</td>
<td>0.19</td>
<td>0.95</td>
<td>0.96</td>
<td>0.7</td>
</tr>
<tr>
<td>NESM3</td>
<td>−121</td>
<td>−235</td>
<td>−328</td>
<td>0.89</td>
<td>0.95</td>
<td>—</td>
<td>0.96</td>
<td>0.97</td>
<td>0.76</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>−94</td>
<td>−140</td>
<td>−266</td>
<td>0.84</td>
<td>0.9</td>
<td>—</td>
<td>0.95</td>
<td>0.97</td>
<td>0.68</td>
</tr>
<tr>
<td>NorESM2-MM</td>
<td>5</td>
<td>118</td>
<td>21</td>
<td>0.92</td>
<td>0.9</td>
<td>0.27</td>
<td>0.98</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>Ensemble Mean</td>
<td>−5</td>
<td>−96</td>
<td>109</td>
<td>0.92</td>
<td>0.89</td>
<td>0.08</td>
<td>0.96</td>
<td>0.99</td>
<td>0.71</td>
</tr>
</tbody>
</table>

and PERSIANN-CDR. Clearly, all models severely underestimate the interannual variability in the Nile basin and its headwaters basins. Specifically, the average coefficient of variation for the 20 GCMs is less than that of PERSIANN-CDR by a factor of 4–7. Consequently, the bias adjusted ensemble of GCMs is underdispersive, which entails that the ensemble does not represent the true uncertainty in annual precipitation. This is demonstrated in Fig. 4b, which shows the rank histogram of PERSIANN-CDR with respect to the bias-adjusted GCMs ensemble for the period 1983–2014. If the ensemble truly captures the variability of annual precipitation, the rank histogram in bins 2–19 should contain 19/21, or 90.5%, of the observations showing in Table 2 ensemble mean providing adequate performance, it does not capture the variability of annual precipitation, the rank histogram in bins m-1 should contain m/21, or 90.5%, of the observations. Consequently, the bias adjusted ensemble of GCMs is underdispersive, which entails that the ensemble does not represent the true uncertainty in annual precipitation.

3) Seasonal cycle

Here we evaluate the performance of the GCMs in capturing the seasonal cycle of precipitation. This is particularly important from the standpoint of assessing the hydrological impacts of climate projections such as variability in the Nile River flow and reservoir operations. Figure 5 shows the observed climatology monthly precipitation (red line) as well as simulations of the 20 GCMs (black dashed lines) and their ensemble mean (solid black line). The two headwaters basins are characterized by distinct precipitation regimes; see Figs. 5b and 5c. Specifically, precipitation in the Blue Nile subbasin is monsoonal with pronounced seasonality (July–September) whereas the Upper White Nile subbasin experiences two rainy seasons (March–May, October–December) (Conway 2005). Specifically, the seasonal rainfall pattern in the Upper White Nile basin follows the seasonal migration of the ITCZ, which leads to a bimodal seasonal cycle (Kizza et al. 2012). The seasonal cycle over the entire Nile basin, thereby, is a reflection of the cycles at the two headwaters basins; specifically, there is a major peak in (July–September) and a less pronounced one around (April–May). Despite overestimation and underestimation bias, the GCMs adequately capture the seasonal variability in precipitation. This is particularly apparent in the Nile and Blue Nile basins with the ensemble mean showing a correlation coefficient of 0.96 and 0.99, respectively, in capturing the seasonal cycle (see Table 2). On the contrary, the GCMs are less capable of capturing the seasonal cycle over the Upper White Nile basin with a correlation coefficient of 0.71 for ensemble mean; in addition, the ensemble mean overestimates the second rainy season (October–December).

Overall, there are numerous observations to be drawn from the retrospective analysis of GCMs simulations; however, two key findings are particularly worthy of consideration. First, the notion of a best-performing model is very sensitive to the specific metric used for evaluation as well as the spatial domain of analysis. Table 2 shows the best-performing model with respect to each metric (in boldface font). Clearly, a different “best performing” model can be selected according to each metric and spatial domain. For instance, KAGE-1.0-G is the best-performing model in terms of bias in annual precipitation over the entire Nile basin (bias = 1 mm), whereas NorESM2-MM is the best-performing model in capturing the seasonal cycle of precipitation in the three basins. Second, although the ensemble mean provides adequate performance, it does not
outperform all individual models as clearly shown in Table 2. This pinpoints that the ensemble mean is sensitive to ensemble members at the end of the performance spectrum. It also underlines that analysis of future projections can benefit from advanced model averaging schemes that consider retrospective model performance to provide a superior estimate to that of individual models.

4) A CLOSER LOOK INTO THE PERFORMANCE OF GCMS

The results in Table 2 show that the performance of GCMS differs substantially depending on the specific metric chosen as a criterion. An extreme example is the MPI-ESM1.2-LR model that is the best-performing model in capturing the seasonal variability of precipitation in the Blue Nile basin while it is one of the six lowest performing models in simulating the amount of annual precipitation in the same basin. These results emphasize that performance of GCMS must be interpreted within the context of the selected metric, and that the metric should be chosen in consistency with the objective of the study. Because the present study is mostly concerned with future projections of annual precipitation, we will focus in the reminder of this section on drawing conclusions on the performance of GCMS in simulating annual precipitation (i.e., the bias results in Fig. 2 and Table 2). Half of the 20 CMIP6 GCMS used in this study consists of pairs of models each of which is provided from a single modeling institute; thus, it is interesting to examine whether models that share the same lineage performs in a similar manner or not. EC-Earth3 and EC-Earth3-Veg are both provided by the European EC-Earth Consortium, and they have exactly similar spatial resolution. Both models perform well in simulating annual precipitation in all basins of the Nile River and consistently show similar sign of bias; precisely, underestimation in Nile and Blue Nile basins and overestimation in Upper White Nile basin. It thus appears that the differences between the two models, mainly the interactive vegetation module added in EC-Earth3-Veg, have insignificant impact on the performance of models in simulating annual precipitation amounts in this region. Similar conclusions with
regard to other pairs of models can be drawn for the Russian Institute for Numerical Mathematics models (INM-CM4.8 and INM-CM5.0) and the Chinese Academy of Sciences models (FGOALS-g3-L and FGOALS-g3).

On the contrary, the Norwegian models of the NorESM Consortium (NorESM2-MM and NorESM2-LM) show opposite sign to each other across all basins of the Nile River. The same behavior can be observed for the Australian models (ACCESS-ESM1.5 and ACCESS-CM2) with opposite sign of bias across all basins of the Nile River. Although each pair of these models share the same lineage, they provide substantially different performance in simulating the amount of annual precipitation. We also note that there appears to be no coherent relationship between the spatial resolution of GCMs and their performance in the present study. For instance, the two models with the lowest resolution (CanESM5 and NorESM2-LM) are consistently ranked among the top performing models in terms of bias in annual precipitation (see Fig. 2).

b. Precipitation projections for the period 2015–2100

1) BMA MEAN PRECIPITATION PROJECTIONS

Here, we analyze mean precipitation projections obtained from the BMA model for the period 2015–2100 with respect to PERSIANN-CDR for the baseline period (1983–2014). Annual precipitation series of the 20 GCMs for the period 2015–2100 were first bias adjusted using the coefficients estimated from Eq. (1). Next, the BMA weights and their corresponding BMA precipitation projections were computed. These calculations were performed at the grid scale as opposed to the entire spatial domain due to the wide climatic variability and the different precipitation regimes in the Nile basin.

Figure 6a shows the projected changes in mean annual precipitation spatially averaged over the entire Nile basin from the 20 GCMs, ensemble mean, and BMA. There is a spread in model projections with 14 models indicating an increase in mean annual precipitation and 6 models showing a decrease. Overall, percentage change in mean annual precipitation ranges from −1.7% to 3.2%. The BMA shows a statistically insignificant increase of 1.34% (p value = 0.2) (see Fig. 6a and Table 3) as compared with 0.82% from the ensemble mean. Figure 6b shows the percentage change in mean annual precipitation projected from BMA for the period 2015–2100 with respect to the baseline period (1983–2014). Clearly, there is spatial variability both in the sign and magnitude of change. A slight decrease in precipitation is observed in southern regions (the Upper White Nile subbasin) whereas the eastern regions (Blue Nile subbasin) show both an increase and a decrease in precipitation. The statistically significant changes in precipitation, at a significance level of 0.05, are observed over the riparian arid regions (stippled grids in Fig. 6b) that have almost no impact on Nile streamflow. Specifically, there is a significant increase in precipitation in northern Sudan (15°–20°N), and a precipitation decrease to the northward. Furthermore, Fig. 6c shows the number of models that agree in the sign of change with BMA projections out of the 20 GCMs used in this study. It appears that spatial grids in which the projected BMA change is significant (stippled grids in Fig. 6b) are grids in which a large number of the 20 GCMs agree in the sign of change.

It is important to narrow the analysis down to regional scales of unified precipitation regimes. Here we focus on headwaters basins, namely, the Blue Nile and Upper White Nile subbasins (see Fig. 1). These basins are characterized by distinct precipitation regimes as shown in Fig. 5. Figure 7a shows the decadal moving average of percentage change in projected annual precipitation at the Blue Nile subbasin. Intermodel differences are clearly present, with a range from −5% to 5% (dashed thin black lines). BMA and ensemble mean are nearly equivalent, and they show no noticeable change in precipitation. To be precise, BMA shows a change of 0.03%, which is not statistically significant with a p value of 0.49 (see Table 3). At the Upper White Nile subbasin (Fig. 7b), BMA deviates from the ensemble mean, and it indicates a decrease of −1.65% in mean annual precipitation, with a p value of 0.09 (see Table 3).
In addition to precipitation projections of the BMA and ensemble mean, Fig. 7 and Table 3 show the projected change in precipitation in each basin from a selected subset of 3 models. The selection criterion is to identify the three models with the least bias in the historical period (1983–2014); see Fig. 1 and Table 2. In each basin, a subset of three models is selected, and its mean is calculated. Table 3 shows that the estimate of the best three models is consistently opposite in sign to the estimate of BMA and ensemble mean. However, their projected changes are small (<1%) and statistically insignificant at $\alpha = 0.05$. We also examined precipitation projections for the rainy seasons in the Nile headwaters basins due to their impact on the variability of the Nile streamflow. The results are shown in Table 4, and they do not show a statistically significant trend, whether decreasing or increasing. Of particular importance is the June–August rainy season in the Blue Nile basin since it contributes 60% of the annual Nile flow. Table 4 shows that the projected change is statistically insignificant, with a decrease of $-0.09\%$ ($p$ value = 0.49).

2) UNCERTAINTY IN BMA PRECIPITATION PROJECTIONS

As discussed earlier, the bias adjusted GCMs ensemble is underdispersive; thus, it underestimates the uncertainty of
precipitation. The BMA approach provides a remedy to this problem because it accounts for two types of variability. Specifically, the BMA total variability is decomposed into two components: between and within variability (Raftery et al. 2005). The former considers the spread of ensemble members whereas the latter accounts for the variability within the individual members. This is clearly shown in Fig. 7, which shows the BMA 90% confidence interval (shaded pink area). While the spread of models (black dashed lines) is limited to an approximate range from $-5\%$ to $5\%$ in the two basins, the BMA 90% confidence interval extends to approximately $\pm 20\%$. This extended uncertainty is the result of the BMA approach consideration of the within variability that is not accounted for in the multimodel ensemble.

Figure 8 shows the distributions of the BMA precipitation projections for the period 2015–2100 expressed as a percentage change with respect to the baseline period (1983–2014). The distributions show also the mean (black dashed line) and the 90% confidence interval limits (gray dashed lines). The mean values are the same as those shown in Table 3. Figure 8a shows the distribution for the Nile basin; the 90% interval range is from $-3.2\%$ to $5.4\%$ with a width of 8.6%. This shows that the probability of an increase in precipitation is higher than that of a decrease. On the contrary, Fig. 8c shows that the probability of a decrease in rainfall at the Upper White Nile basin is higher with a 90% confidence interval range from $-9\%$ to $5.9\%$ with a width of 14.9%. For the Blue Nile basin, the uncertainty range is wider; specifically, from $-11.2\%$ to $16.3\%$ with a width of 27.5%. Besides the wide range of uncertainty in the Blue Nile basin, Fig. 8c shows that the distribution is more centered around 0%; thus, there is also increased uncertainty in the sign of change in precipitation projections.

5. Conclusions

This study examined the performance of 20 CMIP6 GCMs in simulating precipitation for the period 1983–2014 over the Nile basin and then used a Bayesian model averaging scheme to derive precipitation projections for the period 2015–2100. The main findings of retrospective analysis are as follows:

### Table 3. Projected changes in mean annual precipitation in the Nile, Blue Nile, and Upper White Nile basins. The \(p\) values of the projected changes are in parentheses.

<table>
<thead>
<tr>
<th>Basin</th>
<th>Ensemble mean</th>
<th>Best three models</th>
<th>BMA</th>
<th>BMA 90% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nile</td>
<td>0.82% (0.3)</td>
<td>$-0.19%$ (0.45)</td>
<td>1.34% (0.2)</td>
<td>Lower (%)</td>
</tr>
<tr>
<td>Blue Nile</td>
<td>0.43% (0.42)</td>
<td>$-0.92%$ (0.33)</td>
<td>0.03% (0.49)</td>
<td></td>
</tr>
<tr>
<td>Upper White Nile</td>
<td>$-0.45%$ (0.36)</td>
<td>0.17% (0.44)</td>
<td>$-1.65%$ (0.09)</td>
<td></td>
</tr>
</tbody>
</table>
The bias in most GCMs simulations is significant (up to 64% of mean annual precipitation), which consequently pinpoints the importance of bias adjustment prior to analysis of precipitation projections. In addition, the spatial patterns of bias vary considerably within individual models both in the sign and value.

Although all models fairly represent spatial patterns and seasonal cycle of precipitation over most regions in the Nile basin, the results show that the performance of models is less accurate at the Upper White Nile basin.

Selection of a so-called best-performing model is highly dependent on the specific metric chosen as a criterion. Moreover, the results show that the ensemble mean usually does not outperform all individual models.

All models severely underestimate the interannual variability as represented by the coefficient of variation. As a result, the ensemble range underestimates the uncertainty of precipitation.

Bayesian model averaging shows that projected changes in precipitation vary spatially across the Nile basin with clear regional patterns; in particular, a mild decrease of $-1.65\%$ in the Upper White Nile subbasin, almost no change ($0.03\%$) in the Blue Nile subbasin, and significant changes (both increasing and decreasing) in the arid riparian Nile basin. For the Blue Nile subbasin, our results are similar to those reported by Elshamy et al. (2009), which showed no change in annual precipitation as based on 17 CMIP3 GCMs. However, they are at odds with results in Kim and Kaluarachchi (2009) and Fenta Mekonnen and Disse (2018), which showed an increase of 11% and 2.1%–43.8%, respectively. In general, it is not possible to make a conclusive judgement on which study, among previous studies and including the present one, has more credibility because they differ significantly in the models, climate scenarios, future time period, and geographical regions. Nonetheless, we argue that a strict and more cautious approach relative to previous ones has been adopted in this study.

**Table 4.** BMA projected changes in seasonal precipitation in the Blue Nile and Upper White Nile basins.

<table>
<thead>
<tr>
<th>Basin</th>
<th>June–August</th>
<th></th>
<th>October–December</th>
<th></th>
<th>March–May</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Change in mean (%)</td>
<td>$p$ value</td>
<td>Change in mean (%)</td>
<td>$p$ value</td>
<td>Change in mean (%)</td>
<td>$p$ value</td>
</tr>
<tr>
<td>Blue Nile</td>
<td>$-0.09$</td>
<td>0.49</td>
<td>$-0.83$</td>
<td>0.45</td>
<td>0.18</td>
<td>0.49</td>
</tr>
<tr>
<td>Upper White Nile</td>
<td>$0.37$</td>
<td>0.48</td>
<td>$0.47$</td>
<td>0.45</td>
<td>$-0.09$</td>
<td>0.49</td>
</tr>
</tbody>
</table>

**Fig. 7.** Ten-year moving averages of percentage change in projected annual precipitation for the period 2015–2100 with respect to the baseline period (1983–2014), spatially averaged over the (a) Blue Nile and (b) Upper White Nile subbasins. The horizontal axis shows the year at the end of the 10-yr time window. Dashed thin black lines, thick black lines, red lines, and blue lines indicate projections of the 20 GCMs, ensemble mean, BMA, and “best three models,” respectively. The pink-shaded area represents 90% uncertainty bounds of the BMA model.
Last, the BMA probability distributions show that the probability of a decrease in annual precipitation is more likely in the Upper White Nile basin. Moreover, the uncertainty in annual precipitation projections over the Blue Nile basin is higher both in terms of values and sign of change.

Acknowledgments. This research was partially supported by the U.S. Department of Energy (DoE Prime Award DE-IA0000018), California Energy Commission (CEC Award 300-15-005), and University of California (4600010378 TO 15 Am 22). Author contributions: M. Ombadi designed and implemented the study; M. Ombadi wrote the paper; and P. Nguyen, S. Sorooshian, and K. Hsu provided insightful feedback on the drafts of the paper.

Data availability statement. CMIP6 data used in this study are available online (https://esgf-node.llnl.gov/search/c mip6/).


REFERENCES


