J. Hydrol. Hydromech., 69, 2021, 1, 41–48
DOI: 10.2478/johh-2020-0040
ISSN 1338-4333

©2021. This is an open access article distributed under the Creative Commons Attribution NonCommercial-NoDerivatives 4.0 License

sciendo Journal of Hydrology and Hydromechanics

Dynamical bias correction procedure to improve global gridded daily streamflow data for local application in the Upper Blue Nile basin

Haileyesus Belay Lakew[1*], Semu Ayalew Moges[2]

[1] Ethiopian Space Science and Technology Institute, Addis Ababa, Ethiopia.
[2] Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA. E-mail: semu.moges.2000@gmail.com
[*] Corresponding author. Tel.: +251-912-17-94-83. E-mail: haileyesusbelay@gmail.com

Abstract: Recently water resources reanalysis (WRR) global streamflow products are emerging from high- resolution global models as a means to provide long and consistent global streamflow products for assessment of global challenge such as climate change. Like any other products, the newly developed global streamflow products have limitations accurately represent the dynamics of local streamflow hydrographs. There is a need to locally evaluate and apply correction factors for better representation and make use of the data. This research focuses on the evaluation and correction of the bias embedded in the global streamflow product (WRR, 0.25°) developed by WaterGAP3 hydrological model in the upper Blue Nile basin part of Ethiopia. Three spatiotemporal dynamical bias correction schemes (temporal-spatial variable, temporal-spatial constant and spatial variable) tested in twelve watersheds of the basin. The temporal-spatial variable dynamical bias correction scheme significantly improves the streamflow estimation. The Nash-Sutcliffe coefficient (NSCE) improves by 30% and bias decreases by 19% for the twelve streamflow gauging stations applying leave one out cross-validation approach in turn. Therefore, the temporal-spatial variable scheme is applicable and can use as one method for the bias correction to use the global data for local applications in the upper Blue Nile basin.

Keywords: Blue Nile; WaterGAP3; Bias Correction; Water Resource Reanalysis.

INTRODUCTION

Development of accurate hydrologic model predictions for current and potential future climatic conditions requires reliable spatially and temporally distributed climatic information (Hay and Clark, 2003; Jasper et al., 2002). However, the availability and quality of climate data vary from region to region, and observations are sparse and irregularly distributed in many areas (El-Sadek et al., 2011; Haberlandt and Kite, 1998; Kanamaru and Kanamitsu, 2007). Especially in Africa in which the gauging stations are sparse and irregularly distributed, acquiring accurate and consistent hydro-meteorological data is a challenging task. Given this high-resolution satellite and reanalysis precipitation estimation are investigated for more than two decades (Guetter et al., 1996; Hossain and Anagnostou, 2004; Lakew et al., 2017, 2020; Mei et al., 2016; Nikolopoulos et al., 2010).

Along with precipitation, to improve the spatiotemporal availability of streamflow data and to avail a consistent set of data for various global and regional water resources analysis, eartH2Observe (http://www.earth2observe.eu) developed global streamflow WRR products (Schellekens et al., 2017). The products with the preceding version 1 (WRR1, 0.5°) and the recent version 2 (WRR2, 0.25°) using the global hydrological models of WaterGAP, ORCHIDEE, AWRA-L, TESSEL, LISFLOOD, SURFEX and PCR-GLOBWB (Lakew et al., 2019). Such products have a significant role in the assessment of global water balance and water management applications, especially in data-scarce regions of the world.

As such global products usually not correctly represent the local dynamics of the hydrological cycle and also have the limitations to estimate the gauged streamflow data. These global products should be evaluated and corrected before they use for water resources applications for local case studies (Lakew et al., 2019). The first evaluation of the global WRR products is carried out in our previous paper in the same basin of the upper Blue Nile for three case studies of small, medium and large spatial scales (Lakew et al., 2019). The evaluation was for various WRR1 (7 products) and WRR2 (4 products) for daily, monthly as well as wet season temporal scales. The evaluation result highlighted that the recently released WaterGAP3 product exhibits consistently better performance statistics than the earlier coarser-resolution WRR1 and the rest of the WRR2 products at all ranges of temporal and spatial scales (Lakew et al., 2019). Even though the WaterGAP3 product exhibits consistently better performance, it still needs bias correction to capture the gauged streamflow in the upper Blue Nile basin. Therefore, there is a gap in adjusting the mean annual streamflow on the newly released WRR products to correct the bias embedded using bias correction approaches for improvements for various local applications.

Based on this, it is vital to narrow the gap, and the main aim of this study is to evaluate and propose a bias correction procedure that would enable to downscale the global streamflow product WaterGAP3 to a regional scale. The bias correction schemes were carried out at daily temporal scales using twelve observed streamflow gauging stations in the upper Blue Nile basin. The spatiotemporal dynamical bias correction procedures (temporal-spatial variable, temporal-spatial constant and spatial variable) were applied and evaluated for the dynamical bias correction scheme. The leave-one-out cross-validation (LOOCV) approach implemented for each streamflow stations in turn.

DATA AND METHOD

Study area

The Blue Nile River (known as the Abbay River in Ethiopia) rises in the Ethiopian highlands in the region of West Gojam

and flows northward into Lake Tana, located at an elevation of just under 1,800 m. The basin stretches from 34°24'E–39°48'E and 7°42'N–12°30'N and covers a total area of 199,812 km^2 (Figure 1). The basin annual mean streamflow is about 247mm measured at Sudan border. It is the most vital basin in the country by most criteria as it contributes about 45% of the country surface water resources, accommodates 25% of the population, 20% of the landmass, 40% of the nation's agricultural production and most of the hydropower and a significant portion of irrigation potential of the country (Erkossa et al., 2009). The basin characterized by a highly seasonal rainfall pattern with most of the rainfall falling in four months (June to September, JJAS) with a peak in July or August. The mean annual rainfall for the 1961–1990 period amounts to a little over 1,200 mm, of which more than 70% fall during those four months. More than 80% of the annual flow in the Blue Nile results from the summer monsoon and concentrated between June and September (Elshamy et al., 2009; Setegn et al., 2010).

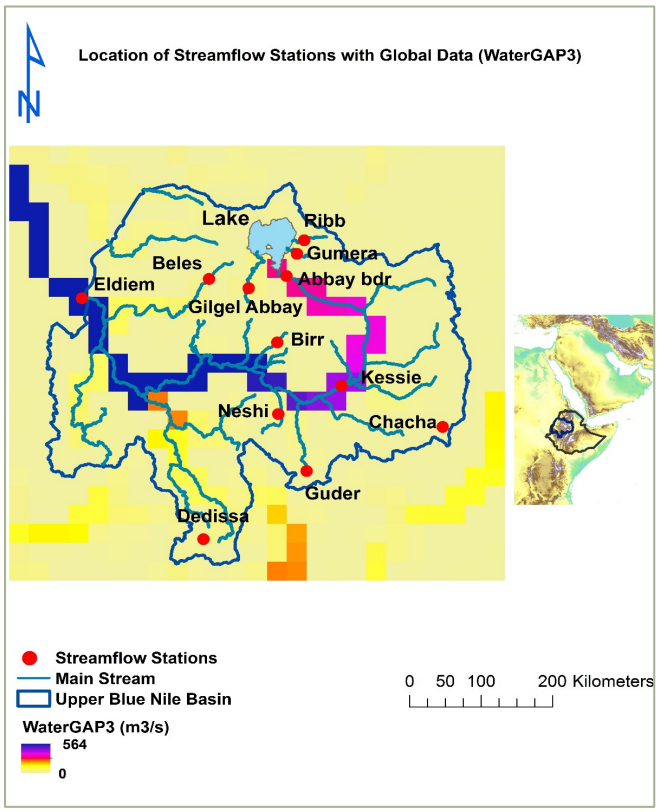


Fig. 1. Location of streamflow stations in the upper Blue Nile basin with the global streamflow data of WaterGAP3 (0.25°).

Local gauge streamflow data

The observed streamflow datasets obtained from the Ethiopian Ministry of Water, Irrigation and Energy used to evaluate the global grid-based reanalysis streamflow data and serve as a reference to estimate the bias factor of the corresponding pixel of the gridded global dataset. The gauging station network in the upper Blue Nile basin is sparse as well as the stations distributed unevenly over the basin area. From the whole set of streamflow gauging stations, twelve gauging stations that would cover the spatial coverage of the basin with a better quality of long records from 1980 to 2010 selected for bias correction and validation procedures. The streamflow gauging stations Eldiem, Kessie, Gilgel Abbay, Beles, Chacha, Guder, Neshi, Dedissa, Ribb, Abbay (Bahir Dar), Gumera and Birr were used and shown in Figure 1. The updated streamflow gauged data is available for a period between 1980 and 2010 for the twelve gauging stations and the mean streamflow with the corresponding basin area shown in Table 1.

Table 1. The upstream basin area and the mean streamflow of the 12 streamflow gauging stations.

No.	Station	Area (km^2)	Mean flow (m^3/s)
1	Eldiem	199,812	1,505.7
2	Kessie	65,784	529.7
3	Gilgel Abbay	1,656	54.6
4	Main Beles	3,431	58.6
5	Chacha	418	4.1
6	Guder	524	12.3
7	Neshi	322	7.6
8	Dedissa (Dembi)	1,806	40.5
9	Ribb	1,592	14.2
10	Abbay (Bahir Dar)	15,321	116.1
11	Gumera	1,394	34.3
12	Birr	978	16

WaterGAP3 reanalysis global streamflow product

The WaterGAP3 (Water Global Assessment and Prognosis) is a global model of water availability as well as water use has been developed to assess the current water resources situation and to estimate the impact of global change on the problem of water scarcity (Müller Schmied et al., 2014; Schellekens et al., 2017). The WaterGAP3 model was developed at the Centre for Environmental Systems Research of the University of Kassel, Germany, in cooperation with the National Institute of Public Health and the Environment of the Netherlands (RIVM). The goals of the model are:

- to enable a comparison of the "freshwater situation" in different parts of the world, i.e. the uses and availability of freshwater to meet various objectives related to the requirements of society and aquatic ecosystems;
- to provide a long term perspective (at least a few decades) on changes in global water resources.

The WaterGAP3 model consists of two main components a Global Water Use model and a Global Hydrology model. The Water Use model takes into account basic socio-economic factors that lead to domestic, industrial and agricultural water use, while the Hydrology model incorporates physical and climate factors that lead to runoff and groundwater recharge (Eisner, 2016). Two multi-model WRR streamflow product of WaterGAP2 is available at 0.5° (WRR1) and WaterGAP3 0.25° (WRR2) grid resolutions, recently produced in the framework of a European Union project (eartH2Observe). In the current study, the high-resolution (0.25°) of WaterGAP3 considered for the bias correction analysis due to its consistent and better performance statistics than from the WRR products at all ranges of temporal and spatial scales evaluated (Lakew et al., 2019).

Schemes for bias correction

Both global precipitation products of satellite and reanalysis rainfall estimates exhibit large systematic and random errors (Bitew and Gebremichael, 2011; Habib et al., 2014; Lakew et al., 2020). Hence, WRR global products consist of errors cascaded from precipitation forcing, global hydrological model formulation and processing simplification. Thus, the bias in global streamflow products should be evaluated and corrected before the global streamflow products use for various applications for local case studies.

We used a multiplicative daily bias reduction estimator (Bias Factor, BF) to correct the bias in the WaterGAP3 model product. The correction for the WaterGAP3 bias, the pixel-based daily bias factors were calculated for some gauging stations and spatially interpolated with the same WaterGAP3 spatial 0.25° grid resolution. For the case of streamflow estimates, we used the nearest neighbor method of spatial interpolation as well as it is applicable for the insertion of the bias factor values at each pixel for the entire basin with 0.25° spatial resolution. For the evaluation and analysis, three dynamical bias correction schemes were applied and evaluated. Temporal-spatial variable, temporal-spatial constant and spatial variable bias factors were estimated to correct WaterGAP3 estimates. The new set of spatiotemporal bias-corrected of WaterGAP3 streamflow gridded data produced by dividing the uncorrected global gridded data of WaterGAP3 by the spatially interpolated gridded bias factor for the respective time window that is computed by the three methods discussed below.

I. Temporal Spatial Variable, BF_{TSV}

In this study, we estimated and corrected the bias in WaterGAP3 estimates as follows. The multiplicative daily bias factor (BF) of a given streamflow gauging station with the WaterGAP3 grid global streamflow data formulated as follows. For each streamflow observation station, 365 bias factors determined. The same with the daily temporal scale, the monthly bias factor was computed using Equation (1). For monthly temporal scale, 12 bias factors generated for each streamflow station.

$$BF_{TSV} = \frac{WaterGAP3(i,t)}{G(i,t)} \tag{1}$$

where G and WaterGAP3 represent daily (monthly) gauge and global streamflow product, respectively, i refers to gauge location, and t refers to a Julian day or month number. The subscript "TSV" stands for "Temporal Spatial Variable" since the bias factor in this formulation estimated for a specific pixel with 0.25° spatial resolution and a particular day or month. Each pixel has the corresponding bias factor value for a specific day (month). The bias factor is dynamic spatially and temporally.

This scheme helps in adjusting the bias at a pixel-based at 0.25° spatial and at a daily or monthly temporal scales (i.e., time and space varying), and based on using the BF_{TSV} factor estimated from Equation (1). The interpolation yields a spatial and temporally varying field of BFs over the entire study area and over the whole time window. The spatially gridded WaterGAP3 daily or monthly streamflow fields were then divided by the spatially gridded and interpolated BF_{TSV} bias fields for the respective time windows. This result a new set of bias-corrected WaterGAP3 estimates that as such are bias-corrected in a temporally and spatially varying scheme.

II. Temporal Spatial Constant, BF_{TSC}

Temporal and spatial constant (TSC) bias correction: in this formulation, the bias factor estimated using Equation (2). The average daily WaterGAP3 and the average observed streamflow over 31 years (1980–2010) for the twelve stations computed and obtained by dividing the average WaterGAP3 product with the average observed streamflow values. Then the bias factor of the eleven stations computed and interpolated for the 0.25° grid resolution to get temporally and spatially lumped value.

$$BF_{TSC} = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N} WaterGAP3(i,t)}{\sum_{t=1}^{T}\sum_{n=1}^{N} G(i,t)} \tag{2}$$

where n is the total number of gauges within the spatial coverage of the study, and T is the full duration of the study period. The bias correction, in this case, is applied by dividing the WaterGAP3 estimates by the bias factor, BF_{TSC}, to result in a new set of WaterGAP3 estimates that are bias-corrected in a spatially and temporally-lumped scheme.

III. Spatial Variable, BF_{SV}

In this formulation, the BF is temporally lumped over the entire domain but, it is still estimated for each streamflow gauge stations Equation (3):

$$BF_{SV} = \frac{\sum_{t=1}^{T} WaterGAP3(i,t)}{\sum_{t=1}^{T} G(i,t)} \tag{3}$$

The bias correction, in this case, is applied by dividing each WaterGAP3 field by the bias factor, BF_{SV}, to result in a new set of WaterGAP3 estimates that are bias-corrected in a temporally lumped but spatially varying scheme.

The bias factor at a given pixel for the three schemes was computed by dividing the uncorrected WaterGAP3 to the observed (gauged) streamflow. This computation might lead to unexpected extremely large or small (near to zero) bias factor values. To avoid the propagation of large and small magnitudes of the bias factor to the neighbor pixels during interpolation, we restricted the bias factor value between 0.5 and 2. The preliminary evaluation of the global WaterGAP3 product for the most streamflow gauging stations bias value was between 0.5 and 2. If the BF value of the streamflow station of a given pixel is greater than 2 or less than 0.5, the BF will be replaced automatically by 2 or 0.5 respectively before the interpolation.

The bias factor of the streamflow stations was interpolated using the nearest neighbor method as well the interpolated map of the bias factor values shown in Figure 2. The map shows the interpolated map of the BF_{TSV} scheme for temporally and spatially varied schemes for randomly selected date of the first July of 2000 (peak runoff season).

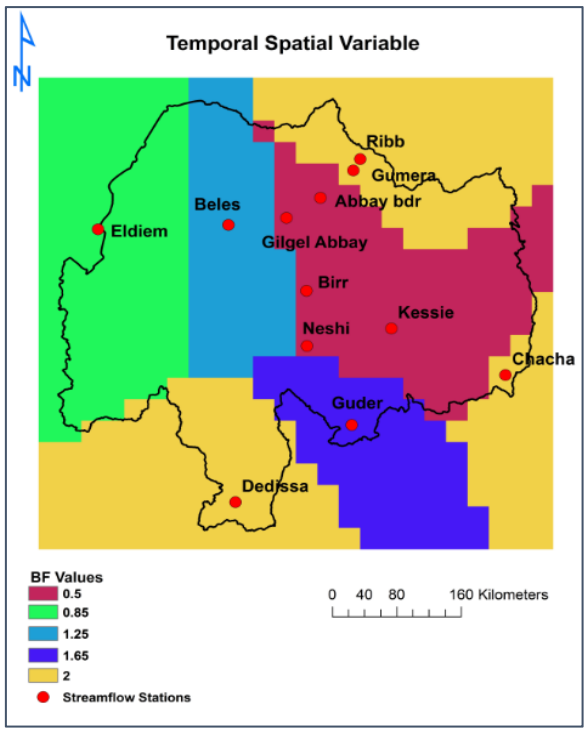


Fig. 2. The interpolated bias factor values of the upper Blue Nile basin using the nearest neighbor method.

From the total stations, three stations of Chacha, Guder and Neshi have less drainage area from the grid cell of the global streamflow data (Table 1). This might have an impact on the rest stations that have large drainage area during the BF interpolation. However, the BF values restricted between the minimum and maximum bounds of 0.5 and 2. Furthermore, the three stations of Chacha, Guder and Neshi have independent BF values shown in Figure 2 for temporally and spatially varied BF values. The three stations considered in this study due to the shortage of streamflow stations.

Performance of bias corrected reanalysis streamflow (WaterGAP3)

The new set of bias-corrected streamflow product computed by the three schemes evaluated using statistical metrics between the bias-corrected and observed streamflow at a daily and monthly temporal scales. The three statistical metrics are described below.

First, for statistical goodness of fit of simulated streamflow, we utilized the Nash-Sutcliffe coefficient of efficiency (Nash and Sutcliffe, 1970).

$$NSCE = 1 - \frac{\sum\left(Q_{i,o} - Q_{i,c}\right)^2}{\sum\left(Q_{i,o} - \overline{Q_o}\right)^2} \quad (4)$$

where $Q_{i,o}$ is the observed streamflow of the ith day; $Q_{i,c}$ is the global streamflow (WaterGAP3) of ith day; and $\overline{Q_o}$ is the average of all the daily observed streamflow values. If NSCE ≤ 0, then the model provides no skill in relation to using the observed mean as a predictor and values greater than zero indicating better agreement.

Second, the Pearson correlation coefficient (CC) is used to assess the agreement between simulated and observed streamflow as follows:

$$CC = \frac{\sum\left(Q_{i,o} - \overline{Q_o}\right)\left(Q_{i,c} - \overline{Q_c}\right)}{\sqrt{\sum\left(Q_{i,o} - \overline{Q_o}\right)^2\left(Q_{i,c} - \overline{Q_c}\right)^2}} \quad (5)$$

where $\overline{Q_c}$ is the average of all daily simulated streamflow (WaterGAP3) values.

Third, relative bias ratio assesses the systematic bias of the simulated discharge:

$$Bias = \frac{\sum Q_{i,c} - \sum Q_{i,o}}{\sum Q_{i,o}} * 100\% \quad (6)$$

The best skill occurs with NSCE = 1, CC = 1, and Bias = 0%.

Cross-validation

In cross-validation (CV), some parts of the datasets are held back when the model is fitted, and used for evaluating the error of the fitted model. The hold-out is called the test, the other part of the training data set. Cross-validation approaches differ in how many of the data are allocated to test/training, and how often the CV itself is repeated (Lachenbruch and Mickey, 1968; Stone, 1977). In our case, we used the leave-one-out cross-validation (LOOCV) method for the analysis. In LOOCV, a single observation from the sample used as test data and the remaining observations as training data. It leaves out only 1 data point and does that for each data point in turn (Badr et al., 2014). This is done repeatedly N times, with N has chosen equal to the sample size (12). The advantage is that the procedure delivers the same results every time because all possible options evaluated.

RESULT

Evaluation of bias correction schemes

To evaluate the three bias correction schemes that were applied for the gridded based global WaterGAP3 streamflow product. For the demonstration, we selected three streamflow gauging stations from a total of twelve stations. The three stations represent a wide range of different spatial scales are small (Gilgel Abbay, 1,656 km^2), medium (Kessie, 65,784 km^2) and large scale (Eldiem, 199,812 km^2). The recent and the available updated observed streamflow data from 2000 to 2010 taken for the comparison of the three schemes.

The results of the three statistical metrics show that the bias-corrected data, applying the bias factor of the temporal-spatial variable (BF_{TSV}) outperforms the other method of bias correction for the three different spatial scales as shown in Table 2. In the case of the Eldiem (large basin), the NSCE performance shows significant improvement from 0.80 to 0.95 for the uncorrected and corrected WaterGAP3 product respectively applying BF_{TSV} bias correction scheme. The bias also reveals a high reduction from 13.9% to 1.2% as the WaterGAP3 is corrected using BF_{TSV}. Similarly, in the case of the medium-scale of Kessie shows a high improvement of NSCE and reduction of bias applying BF_{TSV} scheme. The NSCE improves from 0.66 to 0.93 as well as the bias reduces from –12% to 7.2%. The NSCE performance result from Gilgel Abbay of small basin scale makes sure that the scheme of bias correction using BF_{TSV} gives the best performance from among the methods used for the three case studies of small, medium and large basin scales. The bias efficiency metric shows that the result from BFsv scores minimum bias value for medium and large basin scales. However, the bias value of Gilgel Abbay shows highest from the other schemes. The result of bias of Gilgel Abbay and the performance of NSCE of the three case studies shows that BF_{TSV} is applicable for the streamflow bias correction in the upper Blue Nile basin. The BF_{TSC} and BF_{SV} manifest poor performance even from the uncorrected WaterGAP3 products, that is due to the bias values are highly varied temporally and spatially in the upper Blue Nile basin and should not be lumped in both aspects. Lumping the bias factors temporally or spatially might decrease the performance even from the uncorrected WaterGAP3 product.

Table 2. Different schemes of bias correction for the three-different sized basin scales from 2000 to 2010.

Station		Uncorrected WaterGAP3	Correction Schemes for WaterGAP3		
			BF_{TSV}	BF_{TSC}	BF_{SV}
Eldiem	NSCE	0.80	**0.95**	0.76	0.83
	Bias (100%)	13.9	**1.2**	21.3	0.1
	CC	0.91	**0.98**	0.91	0.91
Kessie	NSCE	0.66	**0.93**	0.60	0.58
	Bias (100%)	–12	**–7.2**	–3.9	0.38
	CC	0.83	**0.97**	0.81	0.81
Gilgel Abbay	NSCE	0.66	**0.76**	0.67	0.61
	Bias (100%)	–21.7	**–5.3**	–16.7	–31.3
	CC	0.83	**0.87**	0.83	0.83

Evaluation of the selected bias correction scheme and streamflow

The result from the evaluation of the three bias correction schemes of the recent period from 2000 to 2010 reveals the BF_{TSV} is applicable for the upper Blue Nile basin. Considering this, we implemented this bias correction scheme applying twelve gauging stations of Eldiem, Kessie, Gilgel Abbay, Beles, Chacha, Guder, Neshi, Dedissa, Ribb, Abbay (Bahir Dar), Gumera and Birr in the upper Blue Nile basin The stations are with a long period of record from 1980 to 2010 at daily temporal scale. There are 365 BF_{TSV} values for each of the eleven stations values interpolated using the nearest neighbor method and the remaining one station used for validation applying leave-one-out cross-validation approach. This procedure carried out for each twelve streamflow stations that were used for bias correction in the upper Blue Nile basin, leaving one observation out of the calibration data set, recalibrating the model, and doing that for each data point in turn.

The result reveals that the bias correction applying BF_{TSV} for the gridded streamflow data WaterGAP3 shows significant improvement for long-term data for the case studies. Nash-Sutcliffe coefficient (NSCE) improves by 30% and bias decreases by 19% for the twelve streamflow gauging stations (Table 3).

The NSCE performance of the uncorrected WaterGAP3 product from 1980 to 2010 shows that only Eldiem scores higher than 0.70. Eldiem is the most downstream station, where due to the large upstream area errors average out more as compared to streamflow from smaller basins. Typical, large-scale hydrological models perform better for downstream stations than for upstream stations and the rest show relatively weak performance below 0.6 (except Gilgel Abbay). Especially, Abbay (Bair Dar) shows the lowest NSCE performance of –1.56 that is due to the effect of lake Tana. This result indicates that the uncorrected product of WaterGAP3 manifests poor performance, especially it has less efficiency for the stations that have a reservoir effect at the upstream of the streamflow gauging stations.

The new set of bias-corrected WaterGAP3 streamflow product shows high NSCE efficiency improvement. From the total 12 stations, 33% of the stations score above 0.8, and 58% of the stations score above 0.7 of NSCE evaluation metric in the corrected WaterGAP3 new set of product.

Based on bias performance evaluation of the corrected gridded WaterGAP3 streamflow product, more than half of the stations score less than 10% of bias, such as Gumera and Ribb score the minimum bias value of –1.9% and 3.8% respectively. More than 80% of the stations score below 25% bias value for the new set of bias-corrected streamflow gridded data.

The overall NSCE results manifest that except Guder (0.87) the stations with small watershed area show relatively weak performance compared to the large watersheds for the corrected WaterGAP3 product. The small watersheds with the weak performance of NSCE are Chacha (0.41), Neshi (0.52) and

Table 3. Performance of daily and monthly WaterGAP3 streamflow with and without bias-correction using BF_{TSV} (1980–2010).

			Daily		Monthly	
No.	Station		WaterGAP3	Bias corrected WaterGAP3	WaterGAP3	Bias corrected WaterGAP3
1	Eldiem (199,812 km^2)	NSCE	0.73	0.85	0.79	0.85
		Bias (100%)	24.2	6.1	24.2	6.1
		CC	0.90	0.94	0.93	0.95
2	Kessie (65,784 km^2)	NSCE	0.37	0.73	0.43	0.74
		Bias (100%)	7	–5.8	6.6	–0.58
		CC	0.66	0.91	0.71	0.92
3	Gilgel Abbay (1,656 km^2)	NSCE	0.67	0.86	0.84	0.87
		Bias (100%)	–18	–8.6	–17	–8.4
		CC	0.83	0.93	0.93	0.95
4	Main Beles (3,431 km^2)	NSCE	0.44	0.78	0.71	0.79
		Bias (100%)	–17.3	–12.5	–10.9	–6.7
		CC	0.67	0.91	0.85	0.97
5	Chacha (418 km^2)	NSCE	0.12	0.41	0.45	0.82
		Bias (100%)	63.4	25.1	63.4	25.3
		CC	0.62	0.73	0.85	0.91
6	Guder (524 km^2)	NSCE	0.48	0.87	0.79	0.91
		Bias (100%)	21.7	7.1	21.7	7.1
		CC	0.77	0.92	0.91	0.94
7	Neshi (322 km^2)	NSCE	0.19	0.52	0.22	0.55
		Bias (100%)	–53.2	–38.1	–53	–38
		CC	0.64	0.83	0.77	0.90
8	Dedissa (1,806 km^2)	NSCE	0.30	0.68	0.48	0.76
		Bias (100%)	–38.3	–23.2	–38	–20
		CC	0.62	0.86	0.87	0.94
9	Ribb (1,592 km^2)	NSCE	0.54	0.84	0.75	0.88
		Bias (100%)	8.6	–3.8	7.9	–29
		CC	0.80	0.92	0.89	0.92
10	Abbay (Bahir Dar) (15,321 km^2)	NSCE	–1.56	0.64	–1.5	0.67
		Bias (100%)	107.7	41	106	40.5
		CC	0.61	0.84	0.63	0.87
11	Gumera (1,394 km^2)	NSCE	0.47	0.70	0.71	0.87
		Bias (100%)	15.4	1.9	15.3	1.9
		CC	0.78	0.84	0.88	0.93
12	Birr (978 km^2)	NSCE	0.34	0.46	0.34	0.68
		Bias (100%)	–36.4	–25.6	–36.3	–25.2
		CC	0.63	0.71	0.85	0.88

Birr (0.46) as shown in Table 3. This indicates that large watersheds with a large magnitude of streamflow have high performance to estimate the observed streamflow for the corrected global product of WaterGAP3. This is because of that the WaterGAP3 product has better performance of minimum bias for the larger basin than smaller basin to estimate the streamflow magnitude.

The daily WaterGAP3 and observed streamflow were aggregated to the monthly mean flow to investigate the effectiveness of the temporal-spatial variable scheme for the monthly temporal scale and shown in Table 3. The bias correction of the monthly BF_{TSV} was carried out as the same procedure as the daily temporal scale. The monthly temporal scale bias-corrected new product of WaterGAP3 improves by much and performs better even from the daily temporal scale for the whole streamflow stations shown in Table 3. These all results show that the bias-adjusted new set of high resolution (0.25°) gridded streamflow data shows high performance to estimate the observed streamflow applying BF_{TSV} scheme. This new set of bias-corrected data can be in use for the data-scarce region of the upper Blue Nile basin, Ethiopia.

For the randomly selected the year of 2000, the corrected and uncorrected WaterGAP3 daily products plotted using hydrograph for Gilgel Abbay and Gumera stations. From the twelve case studies that used for bias correction scheme (Gilgel Abbay and Gumera) stations were selected randomly to plot the hydrograph (Figure 3). The hydrograph of both Gilgel Abbay and Gumera shows that the corrected WaterGAP3 manifests significant improvement from the uncorrected WaterGAP3 to capture the shape and the magnitude of the hydrograph for both Gilgel Abbay and Gumera stations. This result indicates that the corrected WaterGAP3 product using BF_{TSV} bias correction scheme is applicable for local applications for the data-scares region of the upper Blue Nile.

As a final step of this dynamical bias correction, the analysis extended to find out how many of the twelve observational time series are required to improve performance applying BF_{TSV} bias correction scheme at a daily time scale. The analysis was carried out by taking a different number of stations using leave-one-out cross-validation (LOOCV) technique for each trial and each station and shown in Table 4. The average values of the evaluation metrics of NSCE, Bias and CC for each analysis were computed and tabulated below. The result shows that for the upper Blue Nile basin that has a drainage area of 199,812 km^2, the BF_{TSV} bias correction procedure is applicable for more than nine streamflow stations to improve the performance. Taking a lesser number of stations for the analysis might make the performance worse even from the uncorrected WaterGAP3. Mainly, if the stations are less than six, it would deteriorate the original global streamflow data shown in Table 4. This is due to the stations' bias factor (BF) are spatially varied and taking an insignificant number of stations of less than six would make the performance worse even from the original global streamflow data (uncorrected) of WaterGAP3 model.

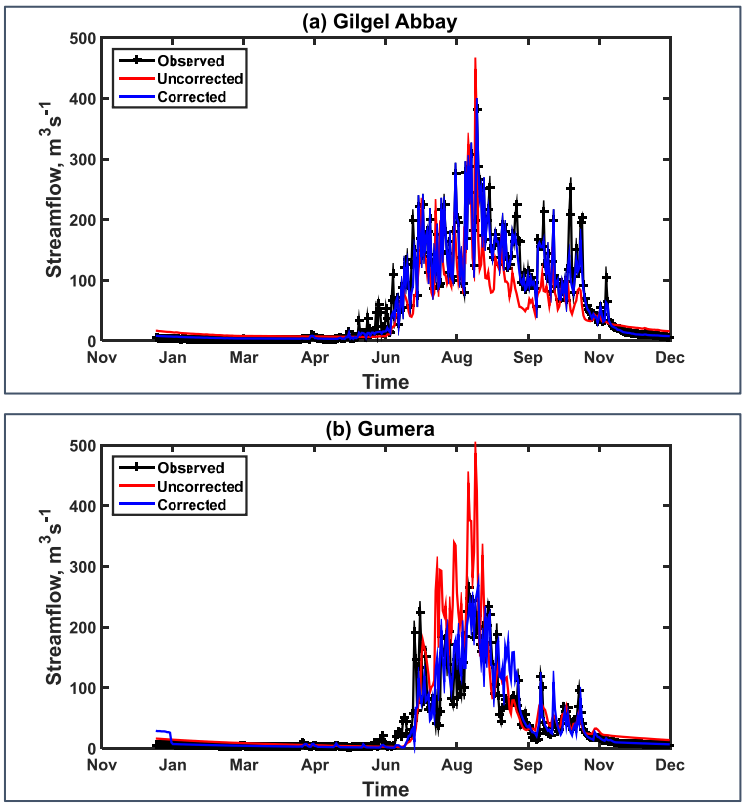


Fig. 3. Hydrograph of the corrected, uncorrected WaterGAP3 and observed streamflow for (a) Gilgel Abbay and (b) Gumera watersheds for the year of 2000.

Table 4. Performance for different number of stations for the bias correction procedure in the upper Blue Nile basin.

	Number of stations taken for the bias correction								Un corrected
	12	11	10	9	6	3	2	1	
NSCE	0.70	0.67	0.63	0.56	0.34	0.28	0.27	0.21	0.36
Bias (100%)	16.6	18.2	19.8	22.1	34.9	38.9	40.1	43.6	34.3
CC	0.86	0.85	0.83	0.81	0.69	0.64	0.64	0.54	0.71

DISCUSSION AND CONCLUSIONS

This study contributes to the improvement of the performance of global WaterGAP3 product by implement bias correction schemes to develop a new set of dynamically bias-corrected gridded streamflow product for the data-scarce region of the upper Blue Nile basin. The new bias-corrected WaterGAP3 daily data is compared with the uncorrected and also with previous study findings. Previously, (Lakew et al., 2019) the first evaluation of the global WRR products (including the WaterGAP3) was carried out at a range of spatial and temporal scales in the nested watersheds (Gilgel Abbay, Kessie and Eldiem) for the upper Blue Nile basin. The results revealed that the WaterGAP3 product manifests better and consistent performance from the other products. However, the findings have highlighted that the accuracy and reliability of global hydrological products vary greatly as a function of the model selected and spatial/temporal requirements and thus uncertainty evaluation and bias correction of those products should precede their use in practice. Therefore, this study aims to implement bias correction to improve the performance of the WaterGAP3 product using dynamical bias correction schemes.

In the previous studies and findings, the results revealed that the reanalysis streamflow cannot capture the gauged streamflow for local applications. The daily reanalysis streamflow product of the uncorrected WaterGAP3 has high bias range from 7% to 100.7%, and the NSCE performance is poor and between –1.56 to 0.7 for the selected streamflow gauging stations in the upper Blue Nile basin. This indicates that the uncorrected global WaterGAP3 streamflow product has an error in streamflow generation mechanisms to estimate the observed streamflow for the local basin river network and needs bias correction or adjustment for various applications at the regional scale. However, not a lot has been done to demonstrate how these products can be ready for various local application by reducing their errors.

This research focuses on improving the quality of global streamflow data by bias-correction to make the product ready for diverse applications in the upper Blue Nile basin, Ethiopia. Three spatiotemporal dynamical bias correction schemes (temporal-spatial variable, temporal-spatial constant and spatial variable) tested in the twelve watersheds (streamflow gauging stations) of the basin. For the bias correction scheme and analysis, from the whole observed streamflow stations in the basin, twelve stations that have a better quality of records selected for bias correction procedure using long term set of observed data from 1980 to 2010 at a daily temporal scale.

The bias correction applying BF_{TSV} manifests significant improvement compared to the other schemes for the global streamflow dataset. The NSCE improves by 30% and bias decreases by 19% for streamflow gauging stations that used for the bias correction in the upper Blue Nile basin. From the total stations, 33% of the stations score above 0.8, and 58% of the stations score above 0.7 of NSCE. Based on bias evaluation metrics, more than half of the stations score less than 5% of bias in the corrected WaterGAP3 new set of daily gridded streamflow product. The leave-one-out cross-validation (LOOCV) approach applied for the bias correction and validation procedures.

These results reveal that the temporal-spatial variable scheme is applicable and can use as one method for the bias correction to use the global data for local applications. The bias-adjusted new set of high resolution (0.25°) gridded streamflow data applying BF_{TSV} can be in use as one source of data for various applications in the upper Blue Nile basin, Ethiopia. The bias correction was applied only using twelve streamflow stations in the upper Blue Nile basin. We recommend further investigation based on the analysis of a reasonably large number of stations with high-quality and recently updated streamflow data.

Taking more streamflow stations (>9) for the procedure makes the bias correction BF_{TSV} scheme applicable and improves the performance. On the other side, taking a lesser number of stations (<6), will make the performance worse even from the uncorrected WaterGAP3 model and not recommended to apply for the bias correction procedure.

Acknowledgements. The authors would like to thank the Ethiopian Ministry of Water, Irrigation and Energy for the updated observed streamflow data.

Conflicts of interest. The authors declare no conflict of interest.

REFERENCES

Badr, H.S., Zaitchik, B.F., Guikema, S.D., 2014. Application of Statistical Models to the Prediction of Seasonal Rainfall Anomalies over the Sahel. J. Appl. Meteorol. Climatol., 53, 614–636. https://doi.org/10.1175/JAMC-D-13-0181.1

Bitew, M.M., Gebremichael, M., 2011. Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. Hydrol. Earth Syst. Sci., 15, 1147–1155. https://doi.org/10.5194/hess-15-1147-2011

Eisner, S., 2016. Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients. PhD thesis. University of Kassel, Kassel, Germany. Available at http://nbn-resolving.de/urn:nbn:de:hebis:34-2016031450014

El-Sadek, A., Bleiweiss, M., Shukla, M., Guldan, S., Fernald, A., 2011. Alternative climate data sources for distributed hydrological modelling on a daily time step. Hydrol. Process., 25, 1542–1557. https://doi.org/10.1002/hyp.7917

Elshamy, M.E., Seierstad, I.A., Sorteberg, A., 2009. Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. Hydrol. Earth Syst. Sci., 13, 551–565. https://doi.org/10.5194/hess-13-551-2009

Erkossa, T., Awlachew, S., Haileslassie, A., Denekew Yilma, A., 2009. Impacts of improving water management of smallholder agriculture in the Upper Blue Nile Basin. https://doi.org/DOI: 10.22004/ag.econ.212433

Guetter, A.K., Georgakakos, K.P., Tsonis, A.A., 1996. Hydrologic applications of satellite data: 2. Flow simulation and soil water estimates. J. Geophys. Res. Atmos., 101, 26527–26538. https://doi.org/10.1029/96JD01655

Haberlandt, U., Kite, G.W., 1998. Estimation of daily space–time precipitation series for macroscale hydrological modelling. Hydrol. Process., 12, 1419–1432. https://doi.org/10.1002/(SICI)1099-1085(199807)12:9<1419::AID-HYP645>3.0.CO;2-A

Habib, E., Haile, A., Sazib, N., Zhang, Y., Rientjes, T., 2014. Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. Remote Sens., 6, Article Number: 6688. https://doi.org/doi.org/10.3390/rs6076688

Hay, L.E., Clark, M.P., 2003. Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. J. Hydrol., 282, 56–75. https://doi.org/10.1016/s0022-1694(03)00252-x

Hossain, F., Anagnostou, E.N., 2004. Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. J. Geophys. Res. Atmos., 109, Article Number: D07102. https://doi.org/10.1029/2003JD003986

Jasper, K., Gurtz, J., Lang, H., 2002. Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. J. Hydrol., 267, 40–52. https://doi.org/https://doi.org/10.1016/S0022-1694(02)00138-5

Kanamaru, H., Kanamitsu, M., 2007. Fifty-seven-year California reanalysis downscaling at 10 km (CaRD10). Part II: Comparison with North American regional reanalysis. J. Clim., 20, 5572–5592. https://doi.org/10.1175/2007jcli1522.1

Lachenbruch, P.A., Mickey, M.R., 1968. Estimation of error rates in discriminant analysis. Technometrics, 10, 1–11. https://doi.org/10.1080/00401706.1968.10490530

Lakew, H.B., Moges, S.A., Asfaw, D.H., 2017. Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile basin: A case study of Gilgel Abbay. Hydrology, 4, Article Number: 39. https://doi.org/https://doi.org/10.3390/hydrology4030039

Lakew, H.B., Moges, S.A., Anagnostou, E.N., Nikolopoulos, E.I., Asfaw, D.H., 2019. Evaluation of global water resources reanalysis runoff products for local water resources applications: Case study - Upper Blue Nile basin of Ethiopia. Water Resour. Manag., https://doi.org/10.1007/s11269-019-2190-y

Lakew, H.B., Moges, S.A., Asfaw, D.H., 2020. Hydrological performance evaluation of multiple satellite precipitation products in the upper Blue Nile basin, Ethiopia. J. Hydrol. Reg. Stud., 27, Article Number: 100664. https://doi.org/https://doi.org/10.1016/j.ejrh.2020.100664

Mei, Y., Nikolopoulos, E., Anagnostou, E., Zoccatelli, D., Borga, M., 2016. Error analysis of satellite precipitation-driven modeling of flood events in complex Alpine terrain. Remote Sens., 8, Article Number: 293. https://doi.org/https://doi.org/10.3390/rs8040293

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F.T., Flörke, M., Döll, P., 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol. Earth Syst. Sci., 18, 3511–3538. https://doi.org/10.5194/hess-18-3511-2014

Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol., 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nikolopoulos, E.I., Anagnostou, E.N., Hossain, F., Gebremichael, M., Borga, M., 2010. Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. J. Hydrometeorol., 11, 520–532. https://doi.org/10.1175/2009jhm1169.1

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., Weedon, G.P., 2017. A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset. Earth Syst. Sci. Data, 9, 389–413. https://doi.org/10.5194/essd-9-389-2017

Setegn, S.G., Srinivasan, R., Melesse, A.M., Dargahi, B., 2010. SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia. Hydrol. Process., 24, 357–367. https://doi.org/10.1002/hyp.7457

Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J. R. Stat. Soc. Ser. B, 39, 44–47. https://doi.org/10.1111/j.2517-6161.1977.tb01603.x

Received 31 January 2020
Accepted 25 September 2020