*Revised Manuscript with no changes marked
Click here to view linked References

Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia

Zheng Duan[1], Ye Tuo[1], Junzhi Liu[2,3*], Hongkai Gao[4,5*], Xianfeng Song[6], Zengxin Zhang[7], Lei Yang[2,3], Dagnenet Fenta Mekonnen[1,8]

[1]Chair of Hydrology and River Basin Management, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

[2]Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, 210023 Nanjing, China

[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, School of Geography, Nanjing Normal University, 210023 Nanjing, China

[4] Key Laboratory of Geographic Information Science (Ministry of Education of China), East China Normal University, Shanghai, China.

[5] School of Geographical Sciences, East China Normal University, 200241 Shanghai, China.

[6] College of Resources and Environment, University of Chinese Academy of Sciences, 100049 Beijing, China

[7] Joint Innovation Center for Modern Forestry Studies, College of Biology and the Environment, Nanjing Forestry University, 210037 Nanjing, China

[8] Amhara Regional State Water, Irrigation and Energy Resource Development Bureau, Bahirdar, Ethiopia

[*]Corresponding authors: Junzhi Liu (liujunzhi@njnu.edu.cn),

Hongkai Gao (hkgao@geo.ecnu.edu.cn)

Abstract

Precipitation and air temperature are key drivers of watershed models. Currently there are many open-access gridded precipitation and air temperature datasets at different spatial and temporal resolutions over global or quasi-global scale. Motivated by the scarcity and substantial temporal and spatial gaps in ground measurements in Africa, this study evaluated the performance of three open-access precipitation datasets (i.e. CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data), TRMM (Tropical Rainfall Measuring Mission) and CFSR (Climate Forecast System Reanalysis)) and one air temperature dataset (CFSR) in driving Soil and Water Assessment Tool (SWAT) model in simulation of daily and monthly streamflow in the upper Gilgel Abay Basin, Ethiopia. The "best" available measurements of precipitation and air temperature from sparse gauge stations were also used to drive SWAT model and the results were compared with those using open-access datasets. After a comprehensive comparison of a total of eight model scenarios with different combinations of precipitation and air temperature inputs, we draw the following conclusions: (1) using measured precipitation from even sparse available stations consistently yielded better performance in streamflow simulation than using all three open-access precipitation datasets; (2) using CFSR air temperature yielded almost identical performance in streamflow simulation to using measured air temperature from gauge stations; (3) among the three open-access precipitation, overall CHIRPS yielded best performance. These results suggested that the CHIRPS precipitation available at high spatial resolution (0.05°) together with CFSR air temperature can be a promising alternative open-access data source for streamflow simulation in this data-scarce area in the case of limited access to desirable gauge data.

Keywords: Blue Nile; Climate Hazards Group InfraRed Precipitation with Station data; Tropical Rainfall Measuring Mission; Climate Forecast System Reanalysis; SWAT; satellite precipitation

1 Introduction

Hydrological models or rainfall-runoff models are essential for understanding the hydrological processes of river basins and supporting operational management of water resources characterized with large spatial and temporal variability (Uhlenbrook et al., 2010; Tuo et al., 2016). Precipitation and air temperature are two necessary weather variables required as inputs to hydrological models. An accurate representation of the temporal and spatial variability of precipitation and air temperature is essential for achieving good simulation and prediction of hydrological processes from models (Wagner et al., 2012; Tuo et al., 2016; Laiti et al., 2018).

Ideally a reasonably dense network of gauge stations are needed to obtain the reliable measured precipitation and air temperature data that are adequate to effectively represent the weather at the basin scale. In reality, the network of gauge stations is often sparse and the point-based measurements with limited coverage are insufficient to capture the spatial and temporal variability of weather variables. Unfortunately, at global scale the number of gauge stations has been significantly declined. This data availability situation is even worse in developing countries and remote areas where measurements are not available or even not existent. Sometimes even data are available, strict data sharing policy could constraint the free access to the public, or the data quality is very poor. For example, despite the importance of Nile River as vital water resource for local population, the understanding of hydrology is still quite limited which is mainly due to the data scarcity and unfavorable data quality (Uhlenbrook et al., 2010; Dile & Srinivasan, 2014; Roth & Lemann, 2016). Very often we are facing limited availability of in-situ measurement, which hinders us to do hydrological Prediction in Ungauged Basins (PUB) (Hrachowitz et al., 2013). Therefore, there is a clear need for improving data collection (if human and financial resources allow) and/or exploring alternative data sources which are more feasible.

Many studies have been conducted to explore the accuracy of using open-access weather data (most focused on only precipitation data) in driving hydrological models in streamflow simulation by using available gauge precipitation data as reference. Our current study focuses on the widely-used Soil and Water Assessment Tool (SWAT) model (Arnold et al., 1998; Arnold & Fohrer, 2005; Gassman et al., 2007; Song et al., 2011, and more in the SWAT Literature Database at https://www.card.iastate.edu/swat_articles/). SWAT is also a popular model for many studies of Nile basin where is overall poorly gauged (see a review by Griensven et al., 2012). For SWAT community, a common source of weather data (precipitation, air temperature and other variables) is the Climate Forecast System Reanalysis (CFSR) data. The CFSR data are promoted and popularized by the SWAT official website through providing ready-to-use weather data in desired format with the data portal at http://globalweather.tamu.edu/. The CFSR is an interpolated dataset on a 38-km grid using climate forecast system with most available in-situ data and satellite data (Radcliffe Z & Mukundan, 2017). The readily availability of weather data in the required format attracted many studies to use CFSR data to drive hydrological models.

Several studies evaluated the performance of using CFSR precipitation to drive SWAT in streamflow simulation. However, contrasting findings were reported from different studies. For example, using CFSR precipitation was found to yield satisfactory streamflow simulation in Lake Tana Basin, Ethiopia (Dile & Srinivasan, 2014), in four small watersheds in USA and the Gumera watershed in Ethiopia (Fuka et al., 2014). But CFSR was found to generate unsatisfactory streamflow simulation in two upstream watersheds of the Three Gorges Reservoir in China (Yang et al., 2014) and in two watersheds in USA (Radcliffe & Mukundan, 2017). The latter found that using the PRISM (Parameter-elevation Relationships on Independent Slopes Model) precipitation data as input yielded satisfactory to even very good streamflow simulation in the same watersheds. All aforementioned studies only explicitly evaluated the performance of CFSR precipitation data but did not comprehensively evaluate the other weather variables (e.g. air temperature) from CFSR. It should be noted that the minimum requirements in weather data input for SWAT model include daily precipitation and daily air temperature (maximum and minimum temperature). Then one research question arises: what is the performance of using CFSR air temperature data together with other better precipitation data to drive SWAT in streamflow simulation? This is particularly relevant for data-scarce or ungauged basins where reliable air temperature data from gauge stations are not available or even nonexistent, thereby hindering the application of SWAT model and other models in such regions. Therefore, this study aims to answer this research question.

Besides the CFSR precipitation data, currently there are many open-access gridded precipitation datasets at different spatial and temporal resolutions over the global or quasi-global scale (Duan et al., 2016). A detailed summary of available precipitation datasets can be found in Tapiador et al., (2012). Overall, the accuracy of different open-access gridded precipitation datasets vary from region to region and thus evaluation of certain precipitation products in a range of regions with different characteristics is important for both product developers and users. Such importance attracted a vast amount of studies that have been carried out to evaluate a single or multiple precipitation products at scales varying from the quasi-global to basin scales (Awange et al., 2016; Bitew & Gebremichael, 2011; Duan & Bastiaanssen, 2013a; Duan et al., 2012; Jiang et al., 2017; Liu et al., 2015; Tan & Duan, 2017; Tang et al., 2016; Yong et al., 2010).

Most gridded precipitation datasets are at the spatial resolution of 0.25° with one grid representing mean precipitation over an area of about 625 km^2, thus such datasets cannot sufficiently reflect the spatial variability of precipitation for relatively small areas. Among them, the TRMM (Tropical Rainfall Measuring Mission) multi-satellite precipitation analysis (TMPA) product (Huffman et al., 2007) is one of the most widely used products at 0.25° and has been used in many applications. It is worth noting that the recently (in 2015) released CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) precipitation dataset (Funk et al., 2015) stands out by providing daily precipitation at the finest spatial resolution of 0.05° (one grid representing around 25 km^2) from 1981 to present. This high spatial resolution enables it to better describe the spatial variability of precipitation and favors its application in hydrological studies at wider scales including the small basins. In addition, CHIRPS was found to be as accurate as or even better than other seven commonly used precipitation products in Adige Basin in Italy after comprehensive evaluation at multiple temporal (daily to annual) and spatial scales (Duan et al., 2016). The follow-up study further demonstrated that using the CHIRPS product as input to the SWAT model resulted in satisfactory performance in simulating monthly streamflow in the same basin (Tuo et al., 2016). A recent evaluation showed that the CHIRPS precipitation data have higher accuracy than other four gridded precipitation datasets in the Upper Blue Nile Basin (Bayissa et al., 2017). The evaluation was carried out by comparing gridded dataset with gauge-based measurements at daily, monthly, and seasonal time scales. Given its aforementioned special feature and good performance, CHIRPS can be a good alternative open-access data source in various applications. To our best knowledge, no study has been conducted to evaluate the performance of using CHIRPS precipitation in driving SWAT to simulate streamflow at the daily scale.

In this study, we focused on a basin upper Gilgel Abay within Lake Tana Basin in Ethiopia where the data scarcity has been mentioned in many previous studies. The data scarcity motives us to explore the alternative data source particularly the relatively new CHIRPS precipitation data. The main objective of this study is to determine the suitable weather data inputs for SWAT in this data-scarce basin. We evaluated the performance of using different combinations of four precipitation datasets (gauge and three open-access datasets, CHIRPS, TRMM, CFSR) and two air temperature datasets (gauge and CFSR) in driving SWAT for daily and monthly streamflow simulation.

The remainder of this paper is organized as follows: Section 2 introduces the study area. Section 3 provides a brief description of data and methods. Section 4 presents the detailed results and discussion. Finally, Section 5 summarizes main findings and additional suggestion for future studies.

2 Study area

The upper Gilgel Abay Basin is located in northwestern highlands of Ethiopia (Fig. 1) It belongs to the Lake Tana Basin. Lake Tana is the largest lake in Ethiopia and the third largest in the Nile River Basin (Setegn et al., 2010). Lake Tana is a vast circular-shaped and shallow lake with water level fluctuations of approximately 1.6 m among seasons. The surface water area of Lake Tana ranges from 2966 to around 3100 km^2 depending on the seasonal fluctuation of lake level (Duan & Bastiaanssen, 2013b). Lake Tana is the source of the Blue Nile River and the Blue Nile River contributes more than 60% of total flow into the Nile River at Aswan in Egypt (Uhlenbrook et al., 2010). Therefore, water resources of Lake Tana are of great importance for Ethiopia and other Nile Basin riparian countries. Despite of such importance, Lake Tana Basin is a poorly gauged basin with ungauged areas accounting for more than 50% of the total area (Wale et al., 2009). Previous studies showed that more than 93% of lake inflow is from four main tributary rivers and the Gilgel Abay is the main tributary by contributing about 60% of the inflow to the lake (Uhlenbrook et al., 2010).

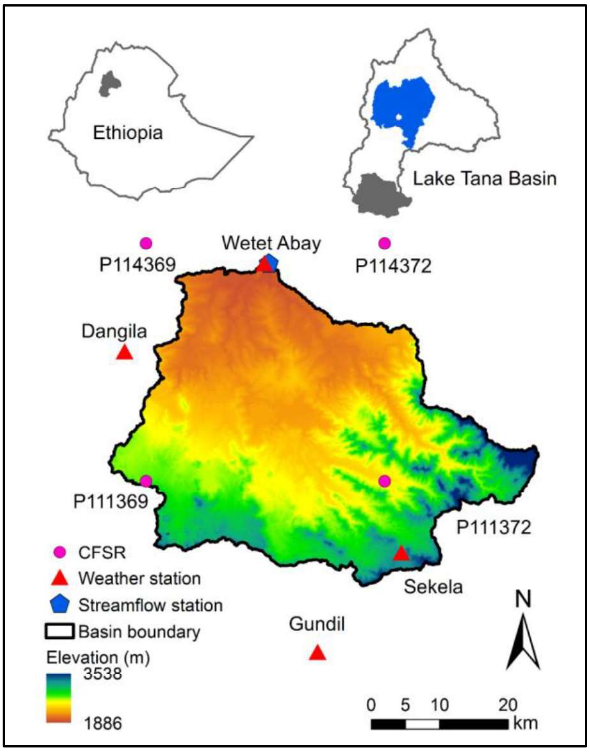


Fig. 1. Locations of the upper Gilgel Abay Basin, one streamflow gauging station and four weather stations, and CFSR stations.

The upper Gilgel Abay Basin has a total area of 1656 km^2. The elevation ranges from 1886 to 3538 m above the mean sea level. The high elevation is located in the southern, west and southeast part. The geology is composed of quaternary basalts and alluviums and the dominant land use types are agricultural and agro-pastoral land with rainfed agriculture accounting for 74% (Uhlenbrook et al., 2010). The dominant soil type is clay. The mean annual precipitation is 1811 mm/year based on the analysis of available rain gauge data between 2000 and 2007. The climate of this region is tropical highland monsoon with a rainy season (June–September) and a dry season (October–March). The seasonal distribution of rainfall is mainly controlled by the north–south movement of the Inter Tropical Convergence Zone (ITCZ) (Taye & Willems, 2012). The air temperature shows a large diurnal but small seasonal variability. Based on measured air temperature from gauge stations for the period 2000-2007, the annual mean daily maximum air temperature is 25.4°C and minimum air temperature is 9.8°C, and the daily average air temperature is 17.6°C.

3 Datasets and methods

3.1 In-situ measurements from gauge stations

In-situ measurements of weather data from four gauge stations were obtained from Ethiopian National Meteorological Agency. Measured daily streamflow from a single station at the outlet of upper Gilgel Abay Basin were obtained from the Hydrology Department of the Ministry of Water Resources of Ethiopia. The locations of these stations are shown in Fig. 1. For weather data, two stations (Wetet Abay and Sekela) are within the basin and the other two (Dangila and Gundil) are around with Dangila station being much closer to the basin. After intensive and rigorous analyses of measured data, finally the available data constrained us to focus on the period 1998-2007 for which data are relatively more complete. For this period, all four stations had daily precipitation data, while three stations excluding Sekela had daily maximum and minimum air temperature, but there were still temporal gaps with more substantial for air temperature data than precipitation. The data gaps and scarcity in this region have been commonly mentioned in many previous studies (Dile & Srinivasan, 2014; Roth & Lemann, 2016), which is indeed the motivation of this study for exploring the performance of alternative open-access weather data. Fig. 2 shows the summary of data gaps for precipitation and air temperature. It is worth noting that in some period, data were available in only one station. For example, from October to December in 2002, daily maximum temperature was only available in the Wetet Abay station. Considerable uncertainty existed in such situations. The SWAT model can automatically fill missing weather data by using weather generator which needs more efforts and more historical data to prepare. In this study, we filled the data gaps before using them as inputs to SWAT. The data gaps were filled as follows: for the dates of data gaps, the data from the closest station were used if possible. In the case of all stations have data gaps for certain dates, then the data gaps were filled by taking available data from the same dates in the closest years for the same station. In this study, we did not interpolate weather stations data as there were only four stations that are insufficient for a reasonable interpolation based on geostatistical methods. We used the weather stations in the normal/standard way to SWAT. The SWAT model (ArcSWAT interface) will automatically distribute the weather data to the subbasins by using data from only one gauge station that is nearest to the centroid of each subbasin (Tuo et al., 2016).

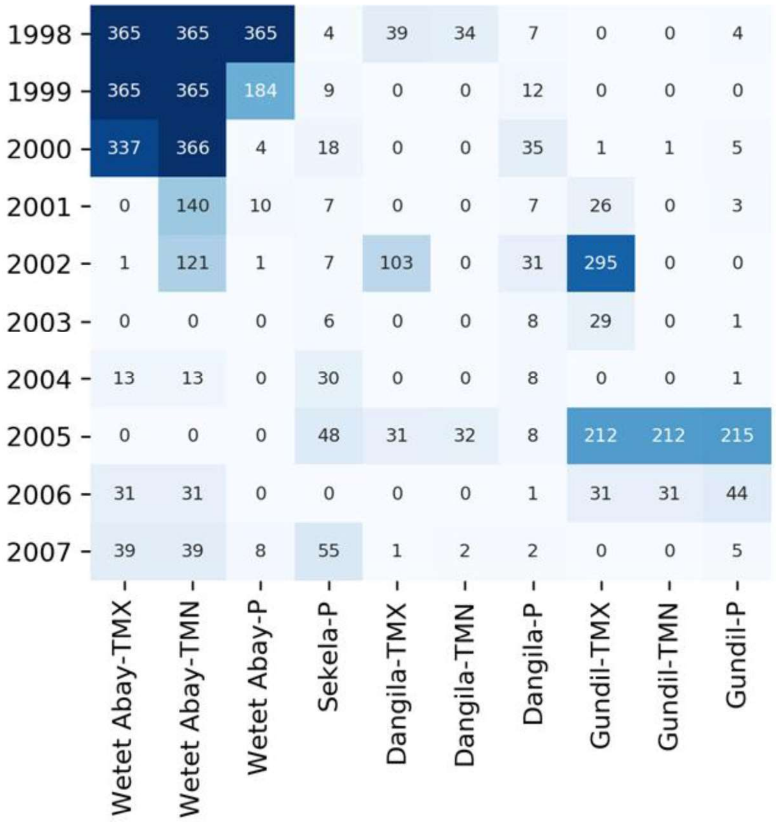


Fig. 2. Data gaps for air temperature and precipitation gauge data. The number in each grid means the number of days with missing data in each year. TMX, TMN, P means daily maximum temperature, minimum temperature and precipitation, respectively.

For streamflow data, the station had more complete data with only 19 values missing (October 8-26, 2006) during the entire period. Streamflow data were used for calibration and validation of the SWAT model in streamflow simulation. The 19 missing data were within the validation period, in this study they were not filled and instead these dates with missing data (October 2006) were simply discarded for validation to avoid additional uncertainty caused by gap-filling.

3.2 CHIRPS precipitation data

CHIRPS stands for the Climate Hazards Group InfraRed Precipitation with Station data. The CHIRPS data provides daily precipitation data at the spatial resolution of 0.05° for the quasi-global coverage of 50°N-50°S from 1981 to present. The latest product is the Version 2.0 product that was released in February 2015. The CHIRPS product and its supporting data are available at: http://chg.geog.ucsb.edu/data/chirps/. The main used datasets for the construction of CHIRPS product include the monthly precipitation climatology (CHPclim) that is created using rain gauge stations collected from FAO and GHCN, the Cold Cloud Duration (CCD) information based on thermal infrared data archived from CPC and NOAA National Climate Data Center (NCDC), the Version 7 TRMM 3B42 data, the Version 2 atmospheric model rainfall field from the NOAA Climate Forecast System (CFS), and the rain gauge stations data from multiple sources. First, the CCD data are calibrated with TRMM 3B42 to generate the 5-daily CCD-based precipitation estimates which are further converted to the fractions of the long-term mean precipitation estimates. The fractions are then multiplied with CHPclim data to remove the systematic bias and the derived product is called CHIRP product. Finally, the CHIRP product is blended with rain gauge stations data using a modified inverse distance weighting algorithm to produce the CHIRPS. All the processing mentioned above are performed at the 5-daily timescales. The daily CCD data and daily CFS data are finally used to disaggregate the 5-daily products to daily precipitation estimates using a simple redistribution method. More detailed information on CHIRPS can be found in Funk et al. (2015). Daily CHIRPS products at the spatial resolution of 0.05° the period 1998-2007 were used and evaluated in this study. SWAT does not allow to directly use gridded precipitation as input as it is not a fully distributed model. Thus we computed the area-weighted average daily CHIRPS data from all grids within the subbasin to represent the effective daily precipitation for each subbasin and then further using them as input to the SWAT model following Tuo et al. (2016). To avoid the edge effect during averaging, the CHIRPS grid cells were firstly disaggregated by 10 times (0.005°) but maintaining original grid locations and values before performing area-weighted averaging.

3.3 TRMM 3B42 precipitation data

The TRMM 3B42 product is one type of the TMPA (TRMM Multi-satellite Precipitation Analysis) products (Huffman et al., 2007). TRMM 3B42 product provides 3-hourly and daily precipitation at the spatial resolution of 0.25° for the quasi-global coverage of 50° N–50° S from 1998 to present. The applied algorithm is the TMPA algorithm that combines precipitation estimates from microwave and infrared satellites, as well as the gauge-interpolated monthly gridded product from GPCC (Global Precipitation Climatology Centre). More details about TMPA algorithms can be found in (Huffman et al., 2007) and Huffman and Bolvin (2015). All TRMM products including 3B42 can be freely downloaded from Goddard Earth Sciences Data and Information Services Center at http://mirador.gsfc.nasa.gov and other sources. The latest version (Version 7) daily accumulated TRMM 3B42 product for the common period 1998-2007 were used in this study, and the data are simply referred to TRMM for conciseness hereafter. Similarly, we firstly disaggregated the TRMM grids by 50 times (to 0.005°) to reduce the edge effect during averaging. Then area-weighted average daily TRMM data from all grids within the subbasin were computed to represent the effective daily precipitation for each subbasin, which were then used as inputs of the SWAT model.

3.4 CFSR precipitation and air temperature data

The CFSR, as the product of the National Centers for Environmental Prediction (NCEP), was designed and executed as a global coupled atmosphere–ocean–land surface–sea ice system to provide the best estimate of the state of these coupled domains (Saha et al., 2010). This system uses most available in situ and satellite observations and provides a range of atmospheric, oceanic, and land surface output products at an hourly time resolution for any geographic location around the globe. The CFSR global atmosphere products are at the spatial resolution of ~38 km with 64 levels extending from the surface to 0.26 hPa. More details about CFSR can be found in Saha et al. (2010). The available CFSR data spans from 1979 to 2014 with planed update to present. The online Global Weather Data for SWAT data portal https://globalweather.tamu.edu/ popularizes the application of CFSR in SWAT modelling community because it provides readily weather data (precipitation, air temperature, relative humidity, wind speed and solar radiation) required by SWAT in the ready-to-use format. Specially, this data portal provides CFSR data like a normal weather station using the centroid of the CFSR grid as the coordinate of each CFSR weather point/station (Dile and Srinivasan, 2014). Users just need to enter the coordinates of the bounding box covering the area of interest and then the data portal would generate the required weather data from the CFSR weather stations within the box. We followed the norm to request the precipitation and air temperature data covering the upper Gilgel Abay Basin for the period 1998-2007 and they were directly used as inputs to the SWAT model. The locations of CFSR weather stations are shown in Fig. 1. Finally only three CFSR weather stations (P114369, P111369 and P111372) located in or closer to the Gilgel Abay Basin were actually used in the SWAT model as the SWAT model automatically selects only one gauge station that is nearest to the centroid of each subbasin (Tuo et al., 2016).

3.5 SWAT model and model setup

SWAT stands for Soil and Water Assessment Tool. It is a semi-distributed, process-based and time-continuous river basin model, which was developed by the Agricultural Research Service of the United States Department of Agriculture-Agricultural Research Service (Arnold et al., 1998). SWAT can be used to model hydrological processes, soil erosion, and water quality in river basins and evaluate the impact of land use change/land management practices on water, sediment and nutrients yields (Neitsch et al., 2011; Song et al., 2011; Tuo et al., 2016). In SWAT, the river basin is first divided to subbasins and further to the Hydrologic Response Units (HRUs) which is the smallest spatial unit. The HRU is generated by a unique combination of land use, soil type and slope. Simulation of hydrology consists of two major phases: the land phase and routing phase. For the land phase, the hydrological cycle simulated by SWAT is based on the soil water balance, and this phase calculates the quantity of water, sediment and nutrients loads from land to the main channel. SWAT offers two methods for estimating surface runoff: the SCS curve number method (USDA-SCS, 1972) that requires daily precipitation as input and the Green and Ampt infiltration method (Green & Ampt, 1911) that requires sub-daily precipitation. The routing phase controls the movement of these loads through the channel network to the outlet of a river basin. The Manning's equation is used to define the rate and velocity of channel flow and flow/water is routed through channels using either variable storage routing or Muskingum routing. More details about the SWAT model can be found in the official theoretical documentation (Neitsch et al., 2011) and review paper (Gassman et al., 2007) as well as SWAT literature database available at https://www.card.iastate.edu/swat_articles/. The SWAT model has been embedded as easy-to-use toolbar as ArcSWAT in ArcGIS interface. The ArcSWAT (Version 2012.10_3.18) was used for setting up SWAT model in this study.

Besides the weather data, which the detailed procedures are mentioned above, the SWAT model requires elevation data, land use map and soil map with information on soil properties. Below describes the data source and processing for setting up the SWAT model in our study. The Digital Elevation Model (DEM) data at the spatial resolution of about 30 m from the Shuttle Radar Topographic Mission 1 arc-second global product were downloaded from USGS EarthExplorer at https://earthexplorer.usgs.gov/. The DEM was used to perform the automatic watershed delineation and used to compute topographic parameters for the SWAT model. The land use map representing the year of 2004 was obtained from the International Livestock Research Institute (ILRI) at http://data.ilri.org/geoportal/catalog/main/home.page. The world soil map developed by the Food and Agriculture Organization (FAO) at 1:5000000 scale was obtained at http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/. Similar to Mekonnen et al. (2018), the Harmonized World Soil Database v1.2 together with this FAO soil map and associated information was used to prepare the required soil properties in SWAT.

SWAT also provides several options for calculating certain hydrological components such as potential evapotranspiration. The default setting for potential evapotranspiration is the Penman–Monteith method which requires more weather data (i.e., wind speed and solar radiation and relative humidity) than the simple Hargreaves method (Hargreaves & Samani, 1982) which requires only air temperature data. Given the common data scarcity, like most previous studies in this study area or nearby regions (Dile & Srinivasan, 2014; Setegn et al., 2010; Tekleab et al., 2011), we used the Hargreaves method for calculating potential evapotranspiration in this study, and all other default settings (e.g. the SCS curve number method for surface runoff and the variable storage routing method for water routing) in SWAT were used.

By only changing different weather data (precipitation and air temperature) as inputs, we were able to set up a number of eight model scenarios, a combination of two air temperature data (guage and CFSR) and four precipitation data (guage, CHIRPS, TRMM, and CFSR), to investigate the effects of different weather data on streamflow simulation.

3.6 Model calibration using SWAT-CUP and model evaluation

In this study, for all eight model scenarios, the SWAT model was run at daily timescale, and the first two years (1998-1999) were considered as warm-up period to mitigate the effect of initial conditions of hydrological modelling. The period 2000-2003 was considered as calibration period, in which sensitivity parameters were calibrated to fit the observed daily streamflow. The remaining period 2004-2007 was used for validation.

The automatic calibration was performed for daily streamflow simulation by using the Sequential Uncertainty Fitting algorithm version 2 (SUFI-2) (Abbaspour et al., 2004; Abbaspour et al., 2007) in the SWAT-CUP tool (Abbaspour, 2015). The sensitivity analysis was firstly performed with SWAT-CUP using one-at-a-time procedure (Abbaspour, 2015) and a number of eight parameters were finally identified as highly sensitive parameters (Table 1). The selection of sensitive parameters is consistent with previous studies (Mekonnen et al., 2018; Setegn et al., 2010). In this study, the same eight parameters were considered for calibration for each SWAT model. The same initial range (Table 1) was used for the eight parameters among all SWAT models to enable a fair starting point and comparison. Following Abbaspour (2015), the calibration procedures were performed with three iterations with 1000 simulations (so a total 3000 simulations during the calibration) being run for each iteration using the Nash-Sutcliffe Efficiency (NSE, Nash & Sutcliffe, 1970) as the objective function. After each iteration, the range of each parameter was updated (normally narrowed down) based on both the new parameters suggested by the SWAT-CUP tool (Abbaspour et al., 2004; Abbaspour et al., 2007) and their reasonable physical boundaries. More details about the calibration procedures can be found in Abbaspour (2015) and Abbaspour et al. (2015). For evaluating model performance in streamflow simulation using different precipitation and air temperature as inputs, the best one among the 3000 simulations from each SWAT model was compared.

For model evaluation and comparison purpose, we used three indicators, i.e. NSE and the coefficient of determination (R^2) and the percent bias (PBIAS, %). Calculations of these indicators were performed using R package hydroGOF (Zambrano-Bigiarini, M., 2014). The NSE measures the quantity difference between the simulated streamflow and the measured streamflow, a value of 1 is the optimal value for NSE, a negative value of NSE means that the model has no skill in the simulation compared to simply using the mean as a predictor (Bitew & Gebremichael, 2011). The R^2 ranges from 0 to 1 and represents the trend similarity between the simulated streamflow and measured. The closer the R^2 value to the optimal value of 1, the better model performance is. The PBIAS measures the average tendency of the simulated values to be larger or smaller than the corresponding observed values. The optimal PBIAS value is 0, and positive (negative) values indicate overestimation (underestimation) bias in the simulation. We followed the criteria proposed by Moriasi et al. (2007) to classify the performance of model to the respective categories: unsatisfactory ($NSE \leq 0.50$, $PBIAS \geq \pm 25\%$), satisfactory ($0.50 < NSE \leq 0.65$, $\pm 15\% \leq PBIAS < \pm 25\%$), good ($0.65 < NSE \leq 0.75$; $\pm 10\% \leq PBIAS < \pm 15\%$) and very good ($NSE > 0.75$, $PBIAS < \pm 10\%$).

Table 1

List of eight parameters considered for calibration and their default values, calibrated ranges and physical ranges. In the SWAT-CUP ‘‘a__’’, ‘‘v__’’ and ‘‘r__’’ means to modify the default value by adding a specified value, to replace the default value by the specified value, and to make a relative change to the initial parameter values, respectively (Abbaspour, 2015). More details on parameter calibration with SWAT-CUP can be found in Yang et al. (2008) and Tuo et al., (2016).

Parameters	Description	Default	Calibrated range	Physical range
r_CN2.mgt	SCS runoff curve number	HRU specific	-0.3/0.1	35/98
r_SOL_AWC.sol	Available water capacity of the soil layer[mm H_2O/mm soil]	Soil layer specific	-0.5/0.5	0/1
v_ESCO.hru	Soil evaporation compensation factor	0.95	0/1	0/1
v_GW_DELAY.gw	Groundwater delay [days]	31	0/500	0/500
v_GW_REVAP.gw	Groundwater "revap" coefficient	0.02	0.02/0.2	0.02/1
v_GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur [mm]	1000	0/5000	0/5000
a_REVAPMN.gw	Threshold depth of water in the shallow aquifer for "revap" to occur [mm]	750	-500/250	0/1000
v_CH_K2.rte	Effective hydraulic conductivity [mm/hr]	0	0/150	-0.01/500

4 Results and discussion

4.1 Comparison of precipitation and temperature inputs

Fig. 3 shows the cumulative fraction of the daily precipitation averaged over the studied basin from four sources during the calibration and validation period (2000-2007). Four products display different probability of occurrence of dry day (rain=0 mm/day), which are 44%, 47%, 30%

and 38% for gauge, CHIRPS, TRMM and CFSR, respectively. Overall all, CHIRPS and TRMM had very similar distribution for all precipitation intensity expect for the dry days. The difference in dry days between CHIRPS and TRMM could be partly due to the spatial resolution issue; TRMM spreads the rain out over the 0.25° pixels that may contain 0.05° pixels with no rain as indicated by CHIRPS. For precipitation intensity with 0-10 mm/day, the distribution of CFSR is very close to that of the gauge measurements. Four products showed larger difference for precipitation within 10-50 mm /day within largest being at the threshold of 20 mm/day. The CFSR data set had the highest frequency (10%) of precipitation beyond 20 mm/day, while the other three data sets had less than 5%. The average annual precipitation from 2000 to 2007 were 1811 mm, 1491 mm, 1471 mm and 2173 mm for gauge, CHIRPS, TRMM and CFSR, respectively.

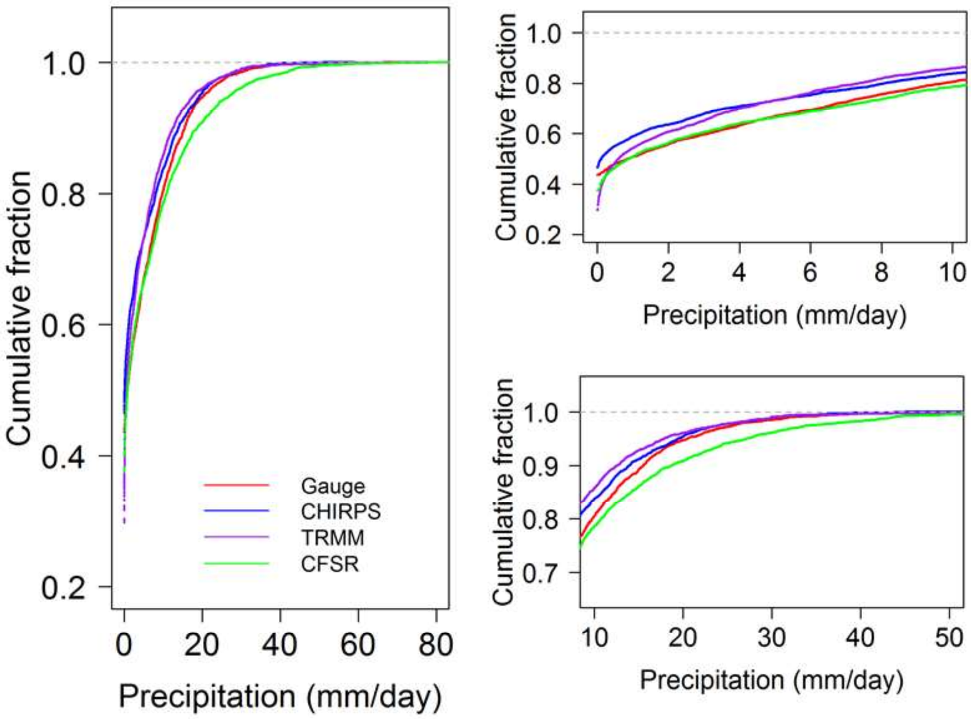


Fig. 3. The cumulative fraction of daily precipitation from four data sets (Gauge, CHIRPS, TRMM and CFSR) at the basin scale during 2000-2007.

Fig. 4 shows the comparison of monthly precipitation over the basin from four datasets. All four datasets showed the same seasonal pattern with rainy months centered in June-September. However, clearly CFSR consistently had more precipitation than the other three data sets for the rainy months through the entire period, especially during the validation period (2004-2007). The pattern and magnitude of monthly precipitation from CHIRPS and TRMM data set were much similar. Both data sets were in much better agreement with gauged precipitation through the entire period, but their peaks were usually lower than gauge data.

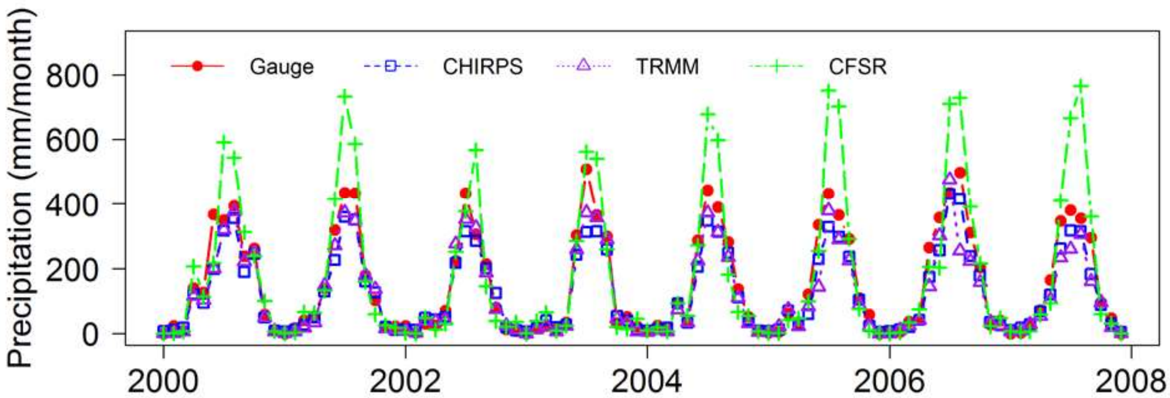


Fig. 4. Comparison of monthly precipitation totals from four data sets (gauge, CHIRPS, TRMM and CFSR) at the basin scale during 2000-2007.

Fig. 5 displays the cumulative fraction of daily maximum and minimum air temperature at the basin scale from the two data sets (gauge and CFSR) during 2000-2007. Fig. 6 presents the monthly mean of daily maximum and minimum air temperature. Overall, CFSR data set agreed better with gauge measurements for the daily maximum air temperature than for the daily minimum air temperature. Similar finding was reported in two basins in Malaysia (Tan et al., 2017). For the daily maximum air temperature, two data sets showed good similarity in the seasonal pattern and magnitude during the entire period except for four months in 2002 when daily maximum temperature was only available in the Wetet Abay station. Analysis of historical air temperature data showed that Wetet Abay station had higher daily maximum air temperature than other stations. Therefore, using gauge data from only Wetet Abay station biased toward higher average daily maximum air temperature at the basin scale in 2002 (Figure 6). There is no snowfall in this study area and thus the air temperature input would be mainly used to compute the potential evapotranspiration (PET) in SWAT. The resulting PET would further affect the computation of water balance in SWAT. To explore the impact of using air temperature input from the two data sets on SWAT modelling, we further compared the PET estimates. The time-series of monthly PET totals from the two data sets are shown in Fig. 6. The seasonal pattern of PET is very similar to that of daily maximum air temperature. Similarly, the PET estimates from two data sets were in good agreement except for the same periods when larger discrepancy occurred in daily maximum air temperature. Therefore, given such good agreement in PET estimates, we expected the impacts of using air temperature input from the two data sets would have very limited influence on the SWAT modelling results, which was ascertained in our following analysis.

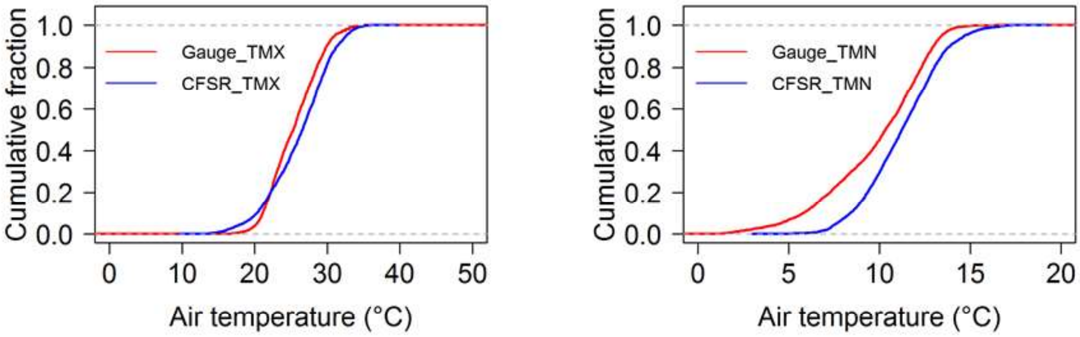


Fig. 5. The cumulative fraction of daily maximum (TMX) and minimum (TMN) air temperature from gauge and CFSR data set at the basin scale during 2000-2007.

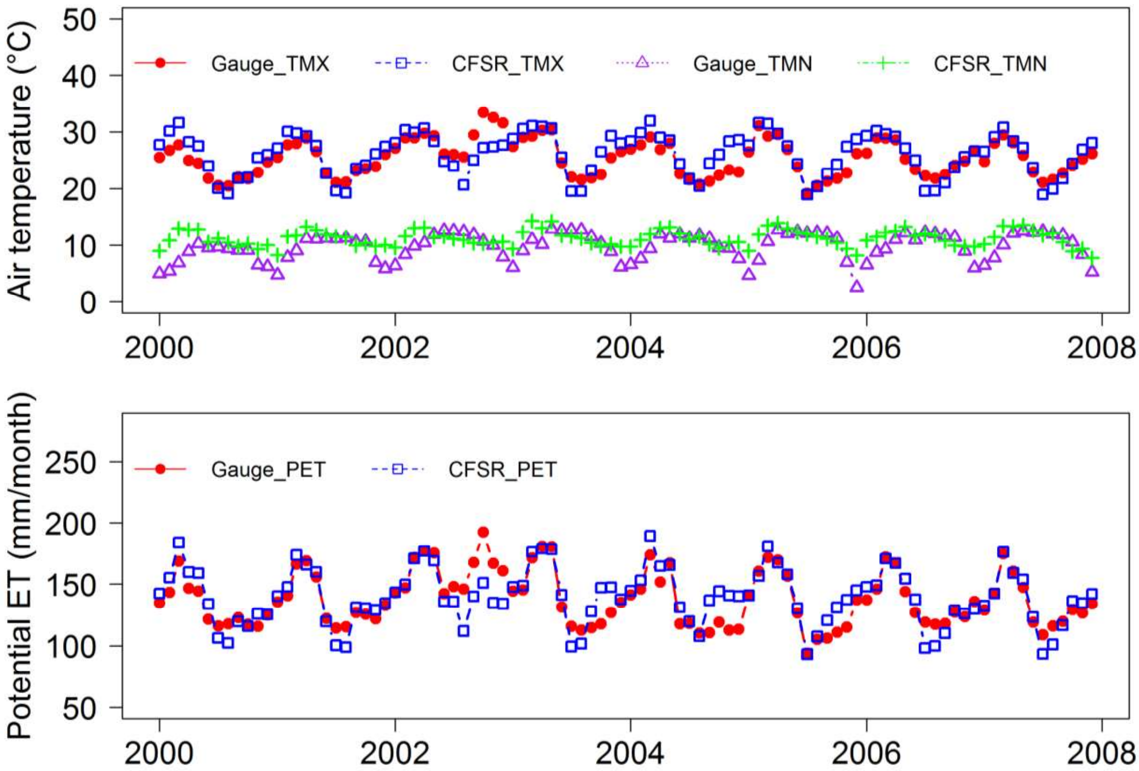


Fig. 6. Comparison of monthly mean daily maximum (TMX) and minimum (TMN) air temperature from gauge and CFSR data set, and their resulting potential evapotranspiration (PET) estimates at the basin scale during 2000-2007.

4.2 Results of streamflow simulation using different precipitation and temperature inputs

4.2.1 Simulation results without calibration

We first evaluated the performance of all eight models without calibration. For conciseness, hydrographs of results without calibration are not shown here, but the model evaluation statistics at the daily and monthly timescales are presented in Table 2 and 3, respectively. According to the guidelines by Moriasi et al. (2007), all eight model scenarios yielded unsatisfactory daily streamflow simulation with NSE values of less than 0.5 for both two periods 2000-2003 and 2004-2007. Only the two models with gauge precipitation input had PBIAS less than 10%, indicating the very good performance on average. The models using the same precipitation but different air temperature inputs had almost the same performance. Using CFSR precipitation as input resulted in the worst performance with lowest NSE values of 0.05 and -1.1 and high positive PBAIS values of around 19% and over 46%for the two considered periods, respectively. This is mainly due to the high overestimation in precipitation by CFSR (Fig. 3). As far as the performance at the monthly scale is concerned, almost all eight models yielded quite good monthly streamflow simulation with NSE > 0.64 and R^2 > 0.82 (except for the period 2004-2007 using CFSR precipitation data). All models except ones with gauge precipitation input had high PBIAS values showing the average tendency of considerable underestimation in simulations by using CHIRPS and TRMM as inputs or overestimation in simulations by using CFSR as input. Using gauge precipitation as input performed best with both NSE and R^2 values larger than 0.90 and small PBIAS. Models using precipitation from CHIRPS and TRMM performed comparably with TRMM slightly better for 2000-2003 while CHIRPS better for 2004-2007. During 2000-2003, using CFSR precipitation as input even outperformed CHIRPS and TRMM, but it yield unsatisfactory simulation (significant overestimation in streamflow) for 2004-2007 with NSE of 0.07.

4.2.2 Simulation results after calibration

Fig. 7 shows comparison of daily measured and simulated streamflow from the four models using gauge air temperature and four different precipitation data sets for the calibration (2000-2003) and validation (2004-2007) periods after calibration. Fig. 8 shows the same as Fig. 7 except using CFSF air temperature as input instead of gauge data. Fig. 9 and 10 shows simulation results at the monthly scale for all eight models. Table 2 and 3 summarizes model evaluation statistics for all eight models at the daily and monthly timescales, respectively. It can be found that if the same precipitation dataset was used, using gauge and CFSR air temperature datasets had almost identical performance.

Overall most models can well captured seasonal patterns. Hydrograph at the daily timescale (Fig. 7 and Fig. 8) showed reasonable agreement between the observed and simulated streamflow using gauge and CHIRPS precipitation as inputs except overestimation and underestimation in a very few events. Using gauge precipitation performed best in daily streamflow simulation with NSE of 0.69 to 0.78. This translates into very good and good performance according to the guideline by Moriasi et al. (2007). The noticeable overestimation in 25 July 2005 during the validation period was caused by the recorded extremely high precipitation from Sekela station (103.5 mm/day). Using CHIRPS precipitation yielded satisfactory performance with NSE of 0.52 to 0.57 and very good performance in terms of PBIAS within 10%. Using TRMM precipitation yielded unsatisfactory performance in terms of NSE, but the NSE were very close to the threshold 0.50 of being satisfactory, and PBIAS within 25% shows satisfactory performance on average simulations. Using CFSR precipitation resulted in satisfactory performance which was even slightly better than that using gauge precipitation for the calibration period, but the performance was very poor for the independent validation period with NSE of 0.01/0.04. This suggests that the calibrated parameters cannot be used for prediction, which is mainly due to the inconsistent behavior of CFSR precipitation in the two periods. As shown in Fig. 7 and Fig. 8, the overestimation of precipitation by CFSR appears to be more severe during the validation period than during the calibration period, and thus the calibrated parameters cannot compensate for such overestimation.

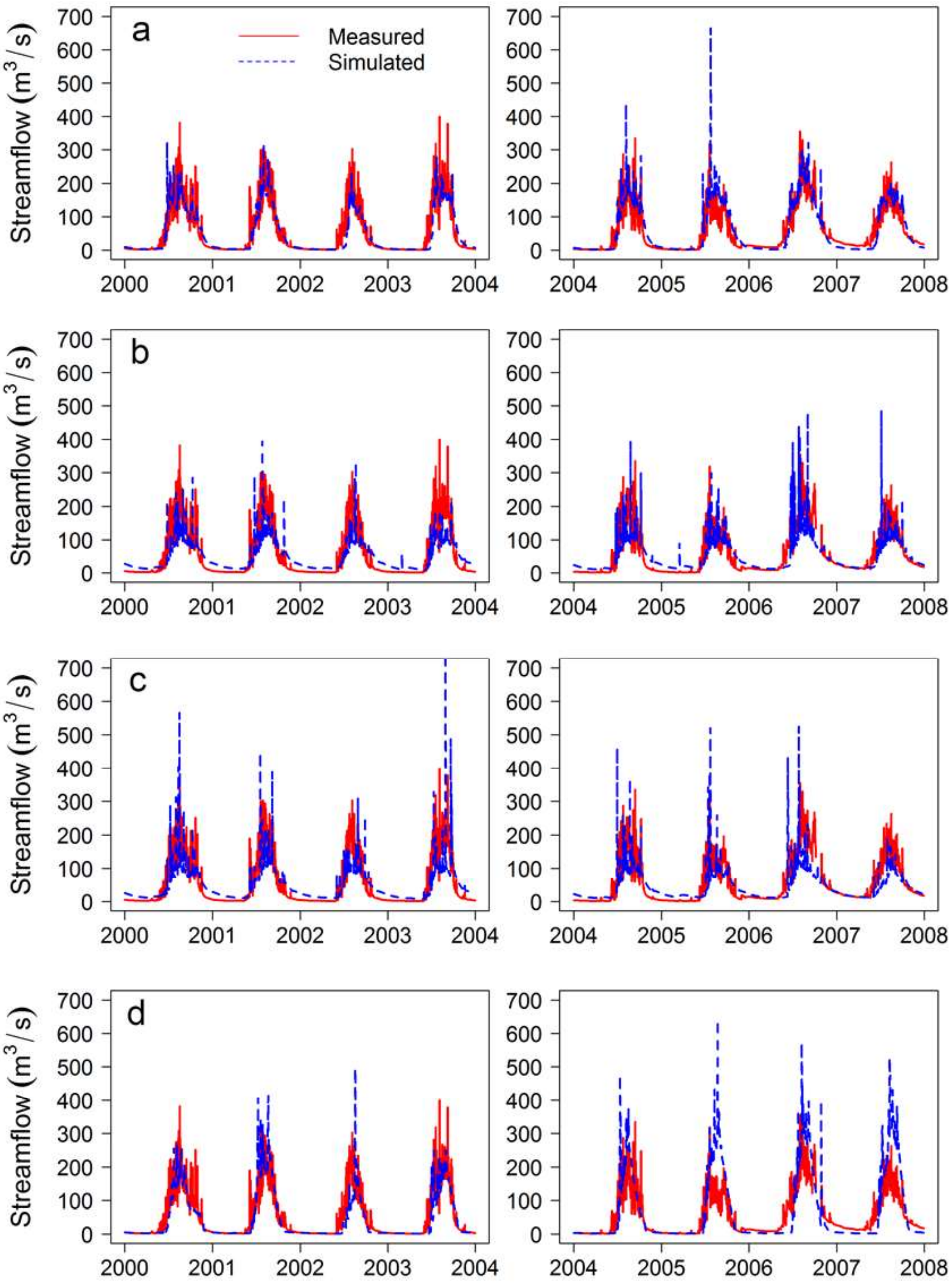


Fig. 7. Comparison of daily measured and simulated streamflow from models using gauge air temperature and four different precipitation datasets (a: Gauge, b: CHIRPS, c: TRMM, d: CFSR) for the calibration period 2000-2003 and validation period 2004-2007.

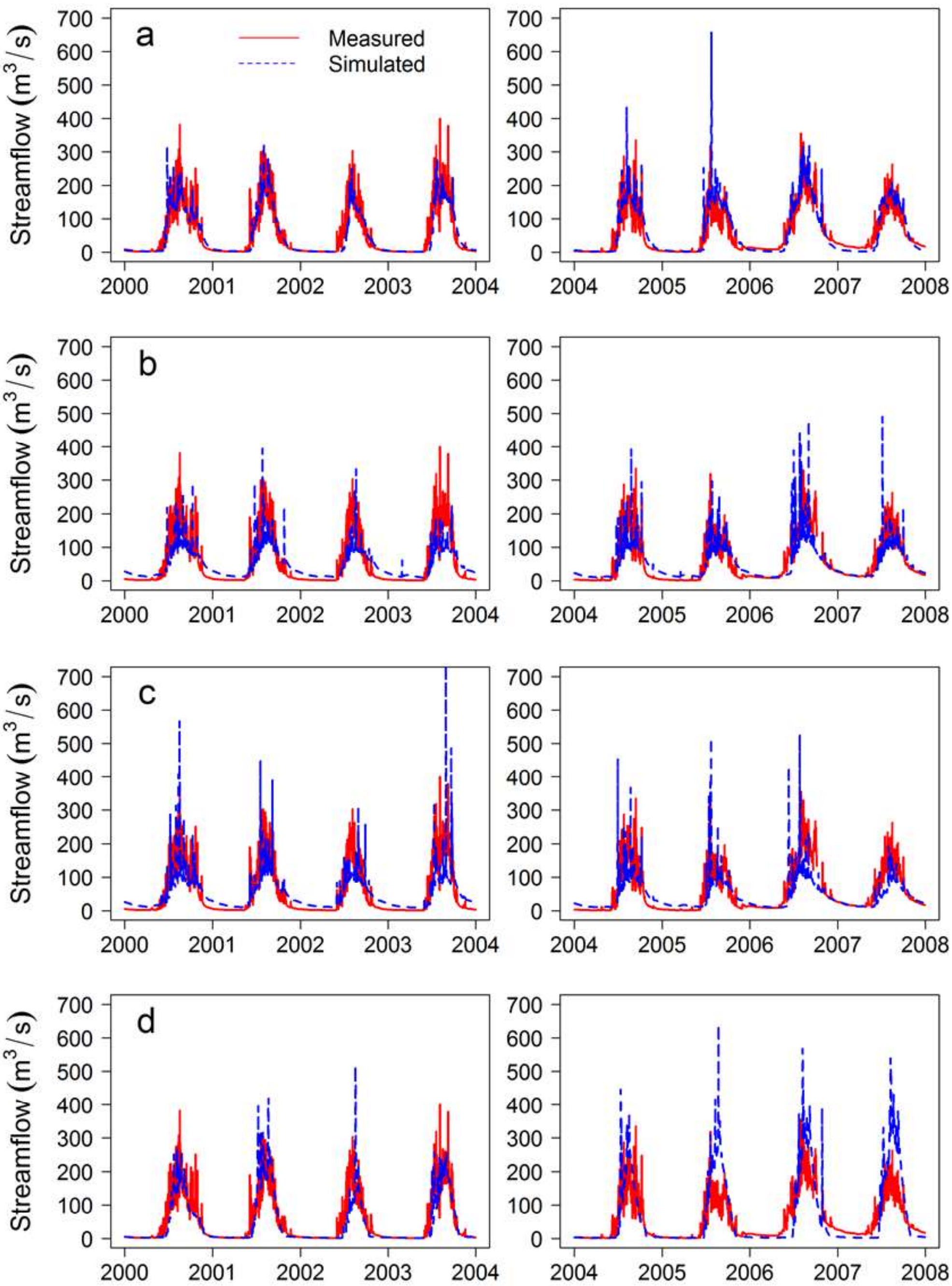


Fig. 8. Comparison of daily measured and simulated streamflow from models using CFSR air temperature and four different precipitation datasets (a: Gauge, b: CHIRPS, c: TRMM, d: CFSR) for the calibration period 2000-2003 and validation period 2004-2007.

When daily results were aggregated to monthly timescale, all models showed better performance with good agreement with measured streamflow in hydrography (Fig. 9 and Fig. 10) and better evaluation statistics (Table 3). Using gauge precipitation still yielded better performance than the other three precipitation datasets with NSE of 0.92 for both calibration and validation period. CHIRPS and TRMM performed comparably well with each being better for a certain period, but hydrography showed both overestimated low flow and underestimate high flow through the entire period 2000-2007. Using CFSR precipitation still cannot satisfactorily simulate monthly streamflow during the validation period with NSE of 0.26 and substantial overestimation as shown in hydrograph.

Table 2

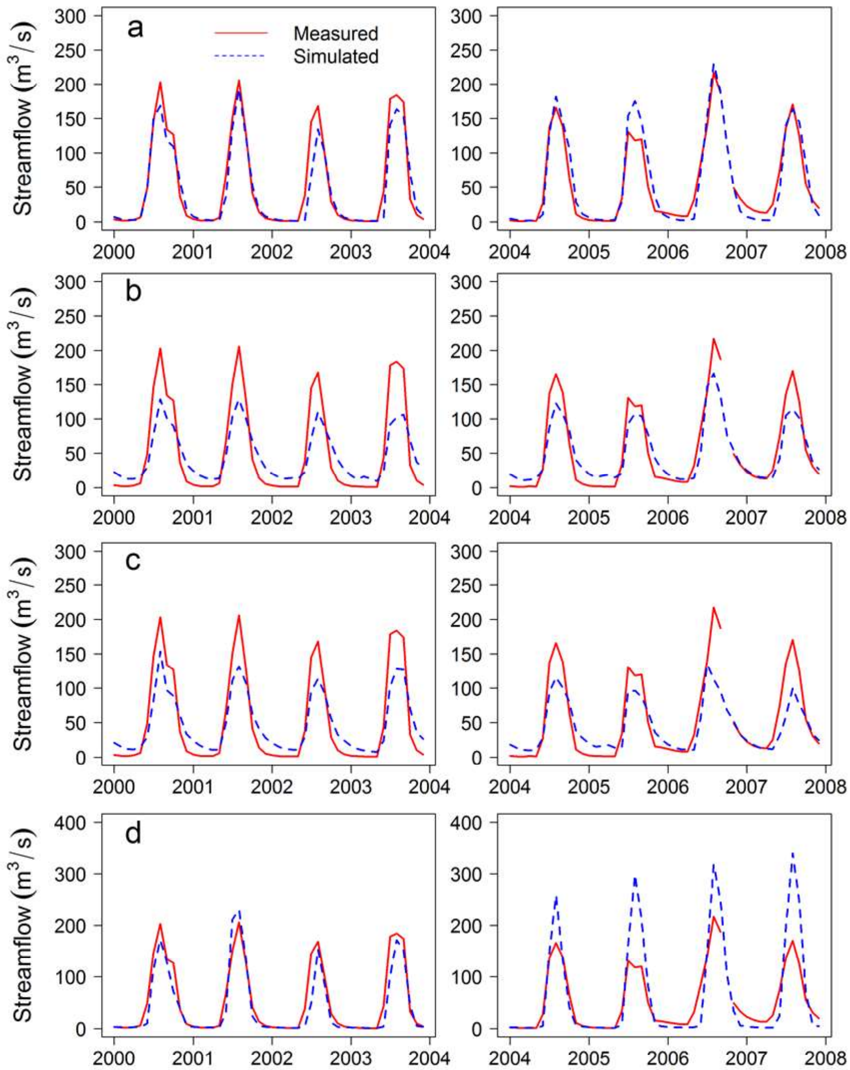
Evaluation statistics for the performance of eight models in daily streamflow simulation

Precipitation data	Temperature data	Without calibration						After calibration					
		2000-2003			2004-2007			2000-2003 (calibration)			2004-2007 (Validation)		
		NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS
Gauge	Gauge	0.49	0.54	-9.80	0.14	0.46	4.40	0.76	0.77	-10.50	0.69	0.75	4.00
	CSFR	0.50	0.54	-9.30	0.16	0.47	2.30	0.78	0.79	-8.00	0.70	0.75	3.60
CHIRPS	Gauge	0.44	0.50	-34.30	0.31	0.44	-26.70	0.56	0.61	-10.50	0.52	0.53	-7.40
	CSFR	0.44	0.49	-33.80	0.31	0.45	-28.70	0.57	0.61	-9.80	0.52	0.54	-8.60
TRMM	Gauge	0.33	0.42	-29.70	0.23	0.38	-38.80	0.49	0.49	-7.40	0.41	0.44	-19.00
	CSFR	0.34	0.42	-29.20	0.24	0.39	-40.40	0.49	0.50	-6.20	0.41	0.44	-20.00
CFSR	Gauge	0.05	0.48	18.90	-1.16	0.51	47.30	0.64	0.67	-18.8	0.01	0.67	20.60
	CSFR	0.05	0.48	19.50	-1.15	0.51	46.20	0.64	0.67	-18.00	0.04	0.67	17.70

Table 3

Evaluation statistics for the performance of eight models in monthly streamflow simulation

Precipitation data	Temperature data	Without calibration						After calibration					
		2000-2003			2004-2007			2000-2003 (calibration)			2004-2007 (Validation)		
		NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS
Gauge	Gauge	0.90	0.93	-9.80	0.93	0.94	4.40	0.92	0.94	-10.50	0.92	0.94	4.10
	CSFR	0.90	0.93	-9.30	0.94	0.94	2.20	0.94	0.95	-8.00	0.92	0.95	3.70
CHIRPS	Gauge	0.68	0.89	-34.30	0.82	0.91	-26.50	0.71	0.88	-10.20	0.85	0.91	-6.60
	CSFR	0.69	0.89	-33.80	0.82	0.92	-28.50	0.72	0.88	-9.50	0.85	0.91	-7.80
TRMM	Gauge	0.77	0.93	-29.70	0.64	0.83	-38.70	0.80	0.92	-7.10	0.72	0.85	-18.30
	CSFR	0.77	0.93	-29.10	0.64	0.85	-40.40	0.80	0.92	-6.00	0.72	0.86	-19.30
CFSR	Gauge	0.81	0.88	18.70	0.07	0.87	46.30	0.86	0.88	-18.80	0.26	0.85	20.60
	CSFR	0.81	0.89	19.40	0.07	0.87	45.10	0.87	0.89	-18.00	0.29	0.85	17.80



516

Fig. 9. Comparison of monthly measured and simulated streamflow from models using gauge air temperature and four different precipitation datasets (a: Gauge, b: CHIRPS, c: TRMM, d: CFSR) for the calibration period 2000-2003 and validation period 2004-2007.

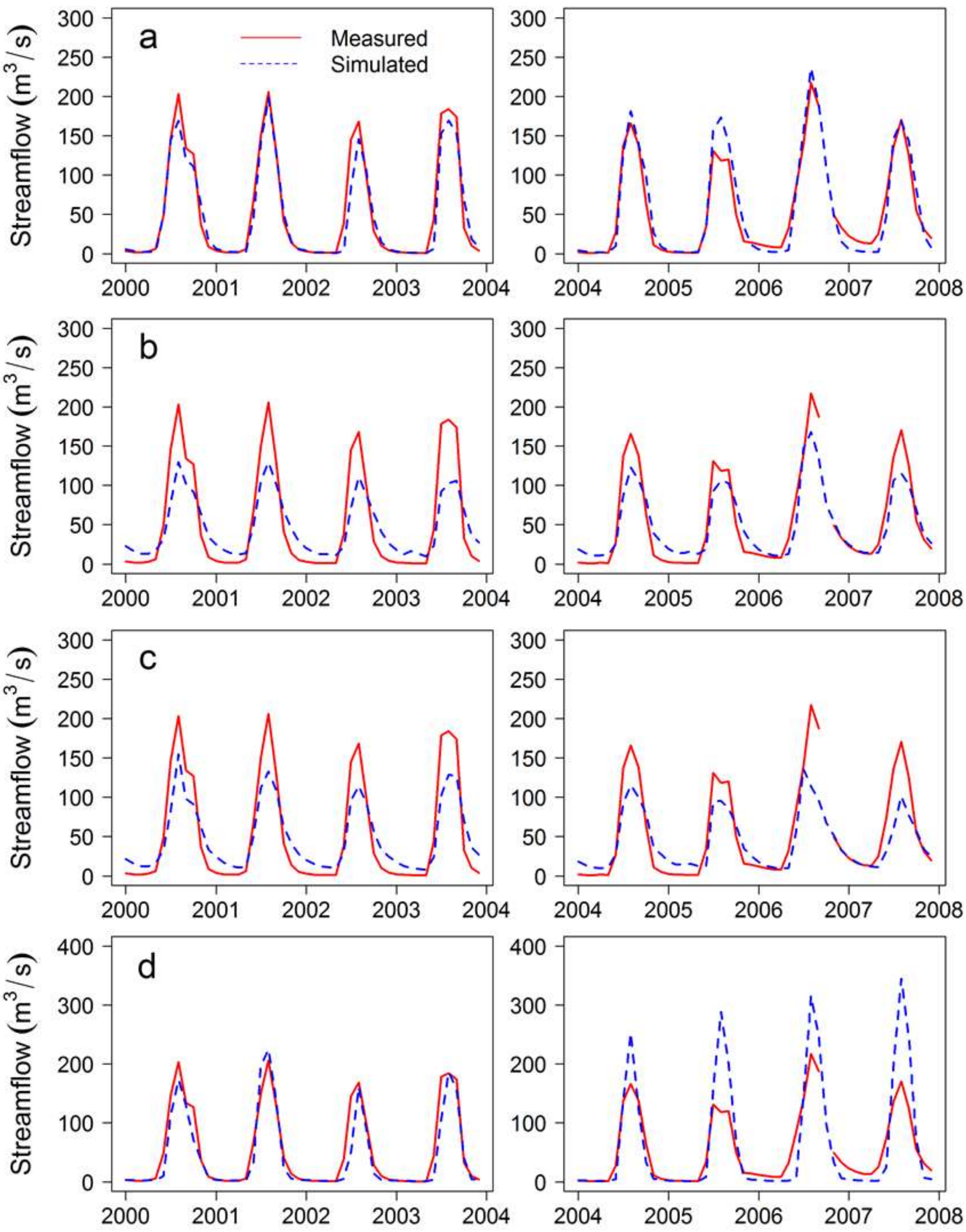


Fig. 10. Comparison of monthly measured and simulated streamflow from models using CFSR air temperature and four different precipitation datasets (a: Gauge, b: CHIRPS, c: TRMM, d: CFSR) for the calibration period 2000-2003 and validation period 2004-2007.

In summary, we can conclude that using different precipitation datasets as inputs to SWAT had much larger influence on streamflow simulation than using different air temperature datasets in this area. Using CFSR air temperature can yield equal performance to using gauge air temperature in driving SWAT model in this study area. This is a good news for researches who are interested in this study area given the limited availability and large amount of gaps in the gauge air temperature data as mentioned in Section 3.1. About the selection of precipitation dataset, this study showed that overall measured precipitation from gauge stations (even though with limited availability and sparse coverage) are still the one that yielded the best simulation result in this study area. This finding is consistent with other studies (Dile & Srinivasan, 2014; Tuo et al., 2016; Worqlul et al., 2015; Yang et al., 2014) which reported better performance using gauge precipitation or interpolation of gauge data than other gridded products. However, the open-access high resolution gridded products CHIRPS was found to yield satisfactory performance in daily and monthly streamflow simulation, and thus it can be a good choice in this study area. In addition, in the case of no access to gauge data at all, the combination of CHIRPS precipitation and CFSR air temperature can be used as an alternative data source to drive hydrological model in streamflow simulation in this data-scarce area.

4.2.3 Comparison of calibrated parameters

Table 4 presents the optimal values of the calibration parameters for all eight models after calibration using SWAT-CUP. Models using air temperature from gauge and CFSR had exactly the same values if they used the same precipitation data except for using the gauge precipitation. The models using gauge precipitation but different air temperature data had different optimal parameter sets, but after careful examination, we found that both parameter sets were ranked as top two parameters sets with very slight difference in the NSE value. In other words, in the case of using gauge precipitation as input, when the best parameter set from model using CFSR temperature was used, the model with gauge temperature could still yield similarly good performance to that using its own best parameter. This reflects the effect of parameter equifinality (Beven & Binley, 1992). Interestingly, models with CHIRPS and TRMM precipitation as input had the same best parameter sets, but using CHIRPS yielded better performance in daily streamflow simulation (NSE=0.56 and NSE=0.52) than using TRMM (NSE=0.49 and NSE=0.41) for both calibration and validation periods.

Table 4

Optimal parameters calibrated for all eight models

Parameters	r__CN2.mgt	r__SOL_AWC.sol	v__ESCO.hru	v__GW_DELAY.gw	v__GW_REVAP.gw	a__GWQMN.gw	a__REVAPMN.gw	v__CH_K2.rte
GaugeP_GaugeT	-0.27	0.27	0.95	1.85	0.07	622.43	223.34	7.39
CHIRPSP_GaugeT	0.09	-0.5	0.98	327.18	0.04	285.43	168.56	145.33
TRMMP_GaugeT	0.09	-0.5	0.98	327.18	0.04	285.43	168.56	145.33
CFSRP_GaugeT	-0.28	0.12	0.67	1.88	0.13	3581.52	-280.59	4.72
GaugeP_CFSRT	-0.27	0.09	0.95	1.87	0.03	607.84	154.37	4.35

CHIRPSP_CFSRT	0.09	-0.5	0.98	327.18	0.04	285.43	168.56	145.33
TRMMP_CFSRT	0.09	-0.5	0.98	327.18	0.04	285.43	168.56	145.33
CFSRP_CFSRT	-0.28	0.12	0.67	1.88	0.13	3581.52	-280.59	4.72

Note. GaugeP_GaugeT means the model using gauge precipitation and gauge air temperature as inputs. CHIRPSP_CFSRT means the model using CHIRPS precipitation dataset and CFSR air temperature data as inputs, and so forth.

Overall, the calibrated parameters using gauge and CFSR precipitation data were similar, and those using TRMM and CHIRPS precipitation data were similar. For example, both gauge and CFSR precipitation leaded to reductions in the parameter CN2 by 27% and 28%, respectively, while both TRMM and CHIRPS leaded to slight increase in CN2 by 9%. Increase in CN2 would result in more runoff by SWAT. For the parameter SOL_AWC that is responsible for available water capacity of the soil layer, both gauge and CFSR precipitation leaded to increase but the increase was less for CFSR. CHIRPS and TRMM precipitation datasets resulted in decrease in SOL_AWC. The decrease in SOL_AWC would generally result in less runoff (Neitsch et al., 2011). For the groundwater delay time (GW_DELAY), gauge and CFSR precipitation had similarly small values, which will resulted in more rapid recharge of the shallow aquifer and discharge to the stream (Radcliffe & Mukundan, 2017). However, CHIRPS and TRMM had very large value for GW_DELAY which translates into slow recharge of the shallow aquifer and discharge to the stream.

In summary, during calibration different parameter values were compensating the difference in precipitation inputs to increase the agreement with measured streamflow at the basin outlet. This might lead to different hydrological components (e.g. surface runoff and groundwater contribution). Therefore, even though all models can fit well the measured streamflow, the partition of water balance components can be different among models (Tuo et al., 2016). This is the inherent limitation of calibrating and validating a model based on only the streamflow at the basin outlet. Unfortunately, this is a common practice in hydrological modelling because measurements for other components are often not available. Many studies have already stressed that simulation of other water balance components from the model that is calibrated with only outlet streamflow should be used with great caution (Bitew & Gebremichael, 2011). Once data allows, the multi-variable and multi-site calibration should be performed to overcome this uncertainty (Tuo et al., 2018). For example, the satellite-based evapotranspiration or soil moisture data could be considered to constrain calibration together with outlet streamflow. In this regard, several studies have been carried out to explore the added values of multi-variable in improving hydrological modelling in other regions (e.g. (Herman et al., 2018)). The same topic (multi-variable and multi-site calibration) is interesting and within our plan for further study in this data-scarce basin in Africa.

4.3 Discussion with existing studies in the same study area

Several studies have been carried out to evaluate the performance of different precipitation datasets in driving hydrological model (particularly SWAT) in streamflow simulation in the same basin or region, e.g. Lake Tana Basin and Blue Nile Basin. We discussed our results with two most relevant previous studies which considered the same precipitation datasets (CFSR and TRMM) with our study.

Bitew and Gebremichael (2011) evaluated four gridded precipitation products at 0.25° spatial resolution including TRMM3B42 in driving SWAT for daily streamflow simulation in the same upper Gilgel Abay Basin. The model was calibrated for the period 2003-2004 and validated for 2006-2007. The authors reported only analysis of validation period at daily timescale. They found that using TRMM3B42 resulted in unsatisfactory daily streamflow simulation with substantial underestimation. The evaluation statistics showed that R^2 values were 0.50 and less than 0.2 for 2006 and 2007, respectively, while NSE were 0.16 and negative. Our study found the same unsatisfactory performance of TRMM in driving SWAT for daily streamflow, which is in good agreement with Bitew and Gebremichael (2011). However, our evaluation statistics for using TRMM3B42 were much better. This could be mainly due to two reasons: (1) mostly importantly *Bitew and Gebremichael* (2011) used old version of TRMM product, while our study used the latest product. Previous study already showed that latest version performed much better than previous version and had reasonably good agreement with gauge-based measurements in the same region (Duan & Bastiaanssen, 2013a). (2) Besides the difference in precipitation data and other data used for setting up SWAT model, the calibration strategy used by Bitew and Gebremichael (2011) might not be able to find the optimal values for TRMM3B42, although they did not explicitly detailed the calibration procedures rather just simply mentioned the application of automatic and manual calibration. Our study used a more objective calibration with the same starting parameter ranges in a sufficient number of iterations, which increases the possibility of finding optimal parameter values for each precipitation product and allow for a more fair inter-comparison among different precipitation products.

Dile and Srinivasan (2014) was perhaps the first study that evaluated the performance of using CFSR in driving SWAT for streamflow simulation in Lake Tana Basin with the upper Gilgel

Abay Basin included. They evaluated the performance of CFSR precipitation and air temperature in monthly streamflow simulation using SWAT without calibration for the period 1993-2007. They concluded that using CFSR data yielded satisfactory performance (NSE=0.79) in simulating monthly streamflow. Our study found that without calibration using CSFR precipitation and air temperature yielded very good performance (NSE=0.81) in monthly streamflow simulation for the period 2000-2003, but very poor performance (NSE=0.07) for the 2004-2007. Thus, our finding partially contradicts with their findings. After careful comparison, we found that as shown in Fig. 1 of Dile and Srinivasan (2014), they somehow consistently discarded all CFSR data in the western part of the study area, even there are CFSR stations located within the study area. This is because they used a smaller bounding box (particularly a larger west longitude value of 36.89°E) than actually needed for covering the entire study area (the area stretches out to the west longitude value of 36.82°E) when they requested the data from the CFSR data portal at https://globalweather.tamu.edu/. As a result, their study used only two CFSR stations (P111372 and P114372) but missed inclusion of another two CFSR stations (P114369 and P111369) that actually should be considered for the upper Gilgel Abay Basin. We analyzed the precipitation data from all the four CFSR stations and found that the other two stations have substantially higher amount of precipitation. To be specific, the average daily precipitation during the period 2000-2007 is 4.8 mm/day for P111372, 2.2 mm/day for P114372, 8.4 mm/day for P111369 and 6.5 mm/day for P114369. Our study used more CFSR stations that should normally be used, and thus CFSR precipitation resulted in severe overestimation particularly in the validation period 2004-2007. While Dile and Srinivasan (2014) used two CFSR stations with lower amount precipitation, and thus better simulation result can be obtained.

To further test our speculation and make a proper comparison, we did further analysis: we did intentionally used the same two CFSR stations as Dile and Srinivasan (2014) did to run SWAT model, and we further considered results without calibration as well as after calibration. Table 5 shows the evaluation statistics for performance of streamflow simulation using precipitation and air temperature from only two CFSR stations. Without calibration, the monthly streamflow simulation showed very good performance with NS of 0.76 for both calibration and validation period, which is now in agreement with conclusion by Dile and Srinivasan (2014). This confirms our speculation. However, strictly speaking, the evaluation by Dile and Srinivasan (2014) did not reflect the complete accuracy of CFSR because of the unintentionally exclusion of two stations. It should be noted that normally users of CFSR will use a larger box covering entirely the study area to select data like what we did in this study, then the good results reported by Dile and Srinivasan (2014) cannot be reproduced. In addition, without ground measurements as reference, pre-selection of CFSR stations cannot be performed in a favorable manner.

Table 5

Evaluation statistics for the performance of model using air temperature and precipitation from only two CFSR stations as Dile and Srinivasan (2014) did in daily and monthly streamflow simulation

Timescale	Without calibration						After calibration					
	2000-2003			2004-2007			2000-2003 (calibration)			2004-2007 (Validation)		
	NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS	NSE	R^2	PBIAS
Daily	0.13	0.34	-27.30	-0.35	0.41	0.10	0.56	0.60	-25.80	0.36	0.66	2.30
Monthly	0.76	0.82	-27.40	0.76	0.86	-1.50	0.77	0.81	-25.80	0.63	0.85	2.00

Furthermore, our analysis showed that without calibration using the only two stations from CFSR still performed unsatisfactorily for daily streamflow with NSE of 0.13 and -0.35 in the calibration and validation periods, respectively. This suggests that the reported good performance of a certain precipitation at monthly timescale does not necessarily guarantee the equally good performance at finer timescale (e.g. daily). Local community should pay due attention to this issue when selecting precipitation products. Even after performing the same calibration strategy, the data from only two CFSR stations can yield satisfactory performance (NSE=0.56) in daily streamflow simulation for the calibration period but fail to generate satisfactory for the validation period (NSE=0.36). Fig. 11 shows the comparison of simulated and measured streamflow at daily and monthly timescale. Therefore, taken together, considering both calibration and validation periods, CFSR precipitation data is not a good alterative data source in this study area. By contrast, CHIRPS precipitation data yielded more consistent performance and the performance was as good as (if not better than) CFSR in daily and monthly streamflow simulation.

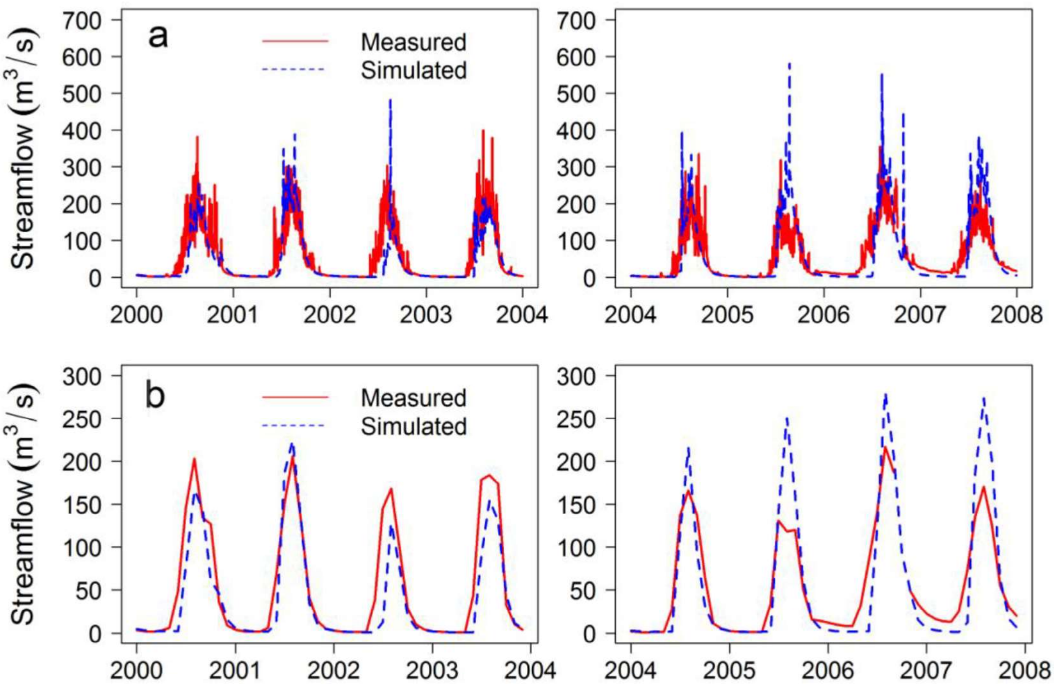


Fig. 11. Comparison of daily (the top panel a) and monthly (the bottom panel b) measured and simulated streamflow from model using air temperature and temperature data from only two CFSR stations as (Dile and Srinivasan, 2014) did for the calibration period 2000-2003 and validation period 2004-2007.

4.4 General discussion and recommendations for future study

Overall the CHIRPS precipitation outperformed TRMM and CFSR precipitation products in driving SWAT model for streamflow simulation in this study. The CFSR product tended to overestimate precipitation and yielded unsatisfactory streamflow simulation using SWAT. Similar significant overestimation of CFSR precipitation data have also been reported in many other regions with different sizes and environmental conditions, e.g. Singapore (Tan et al., 2018), two basins in Malaysia (Tan et al., 2017), several basins in China (Zhu et al., 2016; Gao et al., 2018a), the Mekong River Basin (Chen et al., 2018), and six basins in West Africa (Poméon et al., 2017). It seems that only a limited number of studies reported the reasonable performance of CFSR precipitation, e.g. in four small basins in USA and the Gumera basin in Ethiopia (Fuka et al., 2014). This suggests that the large uncertainty of CFSR precipitation product and it should be used with great cautions. In contrast, literature search showed a very limited number of studies that evaluated the performance of CFSR air temperature. One existing study by Tan et al. (2017) found good correlation of CFSR air temperature product with in-situ measurements in two basins in Malaysia and further using CFSR air temperature can yield good streamflow simulation using SWAT. Their findings are consistent with ours in the current study.

Since the CHIRPS precipitation product (released in 2015) is a relatively new product, thus there are a relatively small number (around 30 journal publications) of studies on the assessment of CHIRPS product and comparison with other widely used products such as TRMM. It is interesting to mention that similar to our study many studies have reported that CHIRPS product has good performance being comparably good or even better than TRMM product, for example, in Mozambique (Toté et al., 2015), Adige basin in Italy (Duan et al., 2016), Upper Blue Nile (Bayissa et al., 2017), East Africa (Kimani et al., 2017, Gebrechorkos et al., 2018), West Africa (Poméon et al., 2017) and Haihe River Basin, China (Gao et al., 2018). After a comprehensive global evaluation of 22 precipitation products, Beck et al. (2017) also concluded that CHIRPS is a viable choice for tropical regions.

It should be noted that this study only evaluated the performance of the CHIRPS, TRMM and CFSR precipitation products and CFSR air temperature at the daily and monthly scales. The CHIRPS stands out in terms of finer spatial resolution (0.05°), but it only provides daily precipitation product. The TRMM and CFSR products with sub-daily temporal resolutions are expected to have beneficial potentials for applications that require precipitation and streamflow simulation at sub-daily scales, e.g. flood simulation. We recommend to evaluate performances of multiple gridded precipitation products at sub-daily scales in future studies. One particular product to evaluate is the Global Precipitation Measurement (GPM) product, the Integrated Multi-satellite Retrievals for GPM (IMERG) became available from March 2014, due to its high temporal (30-minute) and spatial resolution (0.1°) (Yuan et al. 2018).

5 Conclusions

Motivated by the scarcity and substantial temporal and spatial gaps in ground measurements in many basins in Africa, this study evaluated the performance of using three open-access precipitation datasets (CHIRPS, TRMM and CFSR) and one air temperature dataset (CFSR) in driving SWAT model in simulation of daily and monthly streamflow in the upper Gilgel Abay Basin, Ethiopia. The “best” available measurements of precipitation and air temperature from sparse gauge stations were also used to drive SWAT model and the results were compared with those using open-access datasets. After a comprehensive comparison of a total of eight model scenarios, we can draw the following conclusions.

(1) Using measured precipitation from even sparse available stations consistently yielded better performance in streamflow simulation than using all three open-access precipitation datasets, and thus all three open-access precipitation datasets cannot be substitute for ground measurements.

(2) Using CFSR air temperature yielded almost identical performance in streamflow simulation to using measured air temperature from gauge stations. This suggests the favorable accuracy of CFSR air temperature to use for hydrological modelling in this region. This is a good news for the local community as the availability and quality of measured air temperature is often worse than that of precipitation.

(3) Among the three precipitation datasets, overall CHIRPS yielded the best performance and it was the only one that can achieve satisfactory simulation of daily streamflow. The recommended CFSR precipitation by previous study consistently overestimated

precipitation and using CFSR precipitation resulted in inconsistent and overall unsatisfactory performance in daily and monthly streamflow simulation.

(4) Even without calibration, using CHIRPS and TRMM precipitation datasets comparably resulted in satisfactory and up to very good performance in monthly streamflow simulation. This further demonstrates the applicability of SWAT model in this study area and the reasonable accuracy of the two datasets at the monthly timescale.

(5) Using different precipitation datasets resulted in different best parameters during calibration. Therefore, simulation of other water balance components from the model that is calibrated with only outlet streamflow should be used with great caution, as also stressed by Bitew and Gebremichael (2011). Multi-variable and multi-site calibration is a promising way to overcome this limitation to a certain degree.

(6) Taken together, the CHIRPS precipitation available at high spatial resolution (0.05°) together with CFSR air temperature can be a promising alternative open-access data source for streamflow simulation with SWAT in this data-scarce area in the case of limited access to desirable gauge data.

Due to non-availability of gauged wind speed, solar radiation and relative humidity, this study did not explore the performance of using complete CFSR weather data in driving SWAT. The complete CFSR weather data enable users to use the other two more data-demanding methods for calculating potential evapotranspiration. Previous studies showed that different methods resulted in large different potential evapotranspiration estimates and further had certain effects on streamflow simulation by SWAT (Samadi, 2017). This is an interesting topic for future study. In addition, future studies can also include further testing of CHIRPS data in more different regions, and the added values of currently available satellite products in constraining calibration and spatially evaluation of hydrological models particularly in poorly and even ungauged basins.

Acknowledgments

We thank all providers as specified in Section 3 of this manuscript for providing datasets for this study. This work was supported by National Basic Research Program of China (Project No.: 2015CB954102), Natural Science Foundation of Jiangsu Province of China (Project No.: BK20150975), National Natural Science Foundation of China (Project No.: 41601413, 41431177), PAPD (Project No.: 164320H116), and Program of Innovative Research Team of Jiangsu Higher Education Institutions of China. We would like to thank two anonymous reviewers and the Editor for constructive comments which helped improve the quality of this paper.

References

Abbaspour, K., 2015. SWAT-CUP 2012: SWAT calibration and uncertainty programs: A user manual. Department of Systems Analysis, Integrated Assessment and Modelling (SIAM), Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland.

Abbaspour, K.C., Johnson, C.A., Genuchten, M.T.V., 2004. Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. Vadose Zone Journal, 3(4): 1340-1352.

Abbaspour, K.C. et al., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology, 524: 733-752.

Abbaspour, K.C. et al., 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology, 333(2): 413-430.

Arnold, J.G., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling. Hydrological Processes, 19(3): 563-572.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT. Jawra Journal of the American Water Resources Association, 34(1): 73-89.

Awange, J. et al., 2016. Uncertainties in remotely sensed precipitation data over Africa. International Journal of Climatology, 36(1): 303-323.

Bayissa, Y., Tadesse, T., Demisse, G., Shiferaw, A., 2017. Evaluation of Satellite-Based Rainfall Estimates and Application to Monitor Meteorological Drought for the Upper Blue Nile Basin, Ethiopia. Remote Sensing, 9(7): 669.

Beck, H.E. et al., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21(12): 6201-6217.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3): 279-298.

Bitew, M., Gebremichael, M., 2011. Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. Hydrology and Earth System Sciences, 15(4): 1147.

Chen, A., Chen, D., Azorin-Molina, C., 2018. Assessing reliability of precipitation data over the Mekong River Basin: A comparison of ground-based, satellite, and reanalysis datasets. International Journal of Climatology, 38(11): 4314-4334.

Dile, Y.T., Srinivasan, R., 2014. Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: An application in the blue nile river basin. Journal of the American Water Resources Association, 50(5): 1226-1241. DOI:10.1111/jawr.12182

Duan, Z., Bastiaanssen, W., 2013a. First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure. Remote Sensing of Environment, 131: 1-13.

Duan, Z., Bastiaanssen, W.G.M., 2013b. Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. Remote Sensing of Environment, 134(5): 403-416.

Duan, Z., Bastiaanssen, W., Liu, J., 2012a. Monthly and annual validation of TRMM Mulitisatellite Precipitation Analysis (TMPA) products in the Caspian Sea Region for the period 1999–2003, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. IEEE, pp. 3696-3699.

Duan, Z., Bastiaanssen, W.G.M., Liu, J., 2012b. Monthly and annual validation of TRMM Mulitisatellite Precipitation Analysis (TMPA) products in the Caspian Sea Region for the period 1999–2003, Geoscience and Remote Sensing Symposium, pp. 3696-3699.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., Disse, M., 2016. Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. Science of The Total Environment, 573: 1536-1553.

Fuka, D.R. et al., 2014. Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrological Processes, 28(22): 5613-5623.

Funk, C. et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data, 2(9): 150066.

Gao, X., Zhu, Q., Yang, Z., Wang, H., 2018a. Evaluation and hydrological application of CMADS against TRMM 3B42V7, PERSIANN-CDR, NCEP-CFSR, and gauge-based datasets in Xiang River basin of China. Water (Switzerland), 10(9).

Gao, F. et al., 2018b. Evaluation of CHIRPS and its application for drought monitoring over the Haihe River Basin, China. Natural Hazards, 92(1): 155-172.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Transactions of the ASABE, 50(4): 1211-1250.

Gebrechorkos, S.H., Hülsmann, S., Bernhofer, C., 2018. Evaluation of multiple climate data sources for managing environmental resources in East Africa. Hydrology and Earth System Sciences, 22(8): 4547-4564.

Griensven, A.v., Ndomba, P., Yalew, S., Kilonzo, F., 2012. Critical review of SWAT applications in the upper Nile basin countries. Hydrology and Earth System Sciences, 16(9): 3371-3381.

Green, W.H., Ampt, G. A., 1911. Studies on Soil Phyics. The Journal of Agricultural Science, 4(1), 11-24.

Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. Journal of the Irrigation & Drainage Division, 108(3): 225-230.

Herman, M.R. et al., 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. Journal of Hydrology, 556: 39-49.

Hrachowitz, M. et al., 2013. Decade of Predictions in Ungauged Basins (PUB)-a review. Hydrological Sciences Journal, 58(6), 1198-1255.

Huffman, G.J. et al., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of hydrometeorology, 8(1): 38-55.

Jiang, S. et al., 2017. Hydrologic Evaluation of Six High Resolution Satellite Precipitation Products in Capturing Extreme Precipitation and Streamflow over a Medium-Sized Basin in China. Water, 10(1): 25.

Kimani, M.W., Hoedjes, J.C.B., Su, Z., 2017. An assessment of satellite-derived rainfall products relative to ground observations over East Africa. Remote Sensing, 9(5).

Laiti, L. et al., Testing the Hydrological Coherence of High- Resolution Gridded Precipitation and Temperature Data Sets. Water Resources Research.

Liu, J., Duan, Z., Jiang, J., Zhu, A., 2015. Evaluation of three satellite precipitation products TRMM 3B42, CMORPH, and PERSIANN over a subtropical watershed in China. Advances in Meteorology, 2015.

Mekonnen, D.F., Duan, Z., Rientjes, T., Disse, M., 2018. Analysis of combined and isolated effects of land-use and land-cover changes and climate change on the upper Blue Nile River basin's streamflow. Hydrology and Earth System Sciences, 22: 6187-6207.

Moriasi, D.N. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the Asabe, 50(3): 885-900.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles ☆. Journal of Hydrology, 10(3): 282-290.

Poméon, T., Jackisch, D., Diekkrüger, B., 2017. Evaluating the performance of remotely sensed and reanalysed precipitation data over West Africa using HBV light. Journal of Hydrology, 547: 222-235.

Radcliffe, D.E., Mukundan, R., 2017. PRISM vs. CFSR Precipitation Data Effects on Calibration and Validation of SWAT Models. Journal of the American Water Resources Association, 53(1): 89-100. DOI:10.1111/1752-1688.12484

Roth, V., Lemann, T., 2016. Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands. Hydrology and Earth System Sciences, 20(2): 921-934. DOI:10.5194/hess-20-921-2016

Saha, S. et al., 2010. The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8): 1015-1058.

Samadi, S.Z., 2017. Assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the southeastern United States. Hydrology Research, 48(2): 395-415.

Setegn, S.G., Srinivasan, R., Melesse, A.M., Dargahi, B., 2010. SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia. Hydrological Processes, 24(3): 357-367.

Song, X., Duan, Z., Kono, Y., Wang, M., 2011. Integration of remotely sensed C factor into SWAT for modelling sediment yield. Hydrological Processes, 25(22): 3387-3398.

Tan, M.L., Chua, V.P., Tan, K.C., Brindha, K., 2018. Evaluation of TMPA 3B43 and NCEP-CFSR precipitation products in drought monitoring over Singapore. International Journal of Remote Sensing, 39(8): 2089-2104.

Tan, M.L., Duan, Z., 2017. Assessment of GPM and TRMM precipitation products over singapore. Remote Sensing, 9(7): 720

Tan, M.L., Gassman, P.W., Cracknell, A.P., 2017. Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins. Water, 9(3)..

Tang, G., Ma, Y., Long, D., Zhong, L., Hong, Y., 2016. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. Journal of Hydrology, 533: 152-167.

Tapiador, F.J. et al., 2012. Global precipitation measurement: Methods, datasets and applications. Atmospheric Research, 104-105(1): 70-97.

Taye, M.T., Willems, P., 2012. Temporal variability of hydroclimatic extremes in the Blue Nile basin. Water Resources Research, 48(3): 3513.

Tekleab, S. et al., 2011. Water balance modeling of Upper Blue Nile catchments using a top-down approach. Hydrology and Earth System Sciences, 15(7): 2179.

Toté, C. et al., 2015. Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. Remote Sensing, 7(2): 1758-1776.

Tuo, Y., Duan, Z., Disse, M., Chiogna, G., 2016. Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). Science of the Total Environment, 573: 66-82.

Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. A Multi-Objective Approach to Improve SWAT Model Calibration in Alpine Catchments. Journal of Hydrology.

Uhlenbrook, S., Mohamed, Y., Gragne, A., 2010. Analyzing catchment behavior through catchment modeling in the Gilgel Abay, upper Blue Nile River basin, Ethiopia. Hydrology and Earth System Sciences, 14(10): 2153-2165.

USDA-SCS, 1972. SCS national engineering handbook, section 4: hydrology. US Department of Agriculture, Washington, DC.

Wagner, P.D., Fiener, P., Wilken, F., Kumar, S., Schneider, K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology, 464-465(5): 388-400.

Wale, A., Rientjes, T.H.M., Gieske, A.S.M., Getachew, H.A., 2009. Ungauged catchment contributions to Lake Tana's water balance. Hydrological Processes, 23(26): 3682-3693.

Worqlul, A.W. et al., 2014. Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. Hydrology and Earth System Sciences, 18(12): 4871-4881.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology, 358(1-2): 1-23.

Yang, Y., Wang, G., Wang, L., Yu, J., Xu, Z., 2014. Evaluation of gridded precipitation data for driving SWAT model in area upstream of three gorges reservoir. PloS one, 9(11): e112725.

Ye, T., Marcolini, G., Disse, M., Chiogna, G., 2018. A Multi-Objective Approach to Improve SWAT Model Calibration in Alpine Catchments. Journal of Hydrology.

Yong, B. et al., 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. Water Resources Research, 46(7).

Yuan, F. et al., 2018. Evaluation of hydrological utility of IMERG Final run V05 and TMPA 3B42V7 satellite precipitation products in the Yellow River source region, China. Journal of Hydrology. In Press.

Zambrano-Bigiarini, M., 2014. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-8.

Zhu, Q., Xuan, W., Liu, L., Xu, Y.P., 2016. Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China. Hydrological Processes, 30(17): 3061-3083.