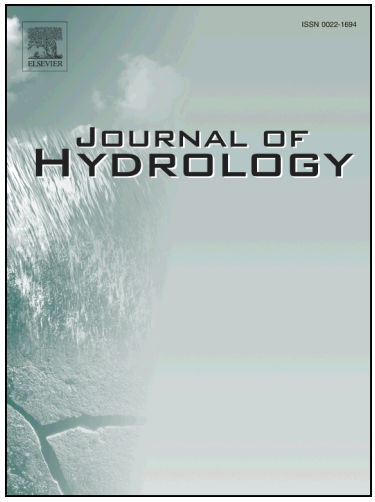



Research papers

# Upper Blue Nile Basin Water Budget from a Multi-Model Perspective

Hahn Chul Jung, Augusto Getirana, Frederick Policelli, Amy McNally, Kristi R. Arsenault, Sujay Kumar, Tsegaye Tadesse, Christa D. Peters-Lidard

PII: S0022-1694(17)30715-1
DOI: https://doi.org/10.1016/j.jhydrol.2017.10.040
Reference: HYDROL 22320

To appear in: *Journal of Hydrology*

Received Date: 1 August 2017
Revised Date: 28 September 2017
Accepted Date: 19 October 2017

Please cite this article as: Jung, H.C., Getirana, A., Policelli, F., McNally, A., Arsenault, K.R., Kumar, S., Tadesse, T., Peters-Lidard, C.D., Upper Blue Nile Basin Water Budget from a Multi-Model Perspective, *Journal of Hydrology* (2017), doi: https://doi.org/10.1016/j.jhydrol.2017.10.040

# Upper Blue Nile Basin Water Budget from a Multi-Model Perspective

Hahn Chul Jung[1,2], Augusto Getirana[1,3], Frederick Policelli[1], Amy McNally[1,3], Kristi R. Arsenault[1,4], Sujay Kumar[1], Tsegaye Tadesse[5], Christa D. Peters-Lidard[1]

[1]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[2]Science Systems and Applications, Inc., Lanham, MD, USA
[3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[4]Science Applications International Corporation, Inc., McLean, VA, USA
[5]National Drought Mitigation Center, University of Nebraska-Lincoln, Lincoln, NE, USA

Corresponding Author
Hahn Chul Jung
8800 Greenbelt Rd., Greenbelt, MD 20771, USA
hahnchul.jung@nasa.gov
+1-301-614-5563

*Submitted to*
*Journal of Hydrology*

## ABSTRACT

Improved understanding of the water balance in the Blue Nile is of critical importance because of increasingly frequent hydroclimatic extremes under a changing climate. The intercomparison and evaluation of multiple land surface models (LSMs) associated with different meteorological forcing and precipitation datasets can offer a moderate range of water budget variable estimates. In this context, two LSMs, Noah version 3.3 (Noah3.3) and Catchment LSM version Fortuna 2.5 (CLSMF2.5) coupled with the Hydrological Modeling and Analysis Platform (HyMAP) river routing scheme are used to produce hydrological estimates over the region. The two LSMs were forced with different combinations of two reanalysis-based meteorological datasets from the Modern-Era Retrospective analysis for Research and Applications datasets (i.e., MERRA-Land and MERRA-2) and three observation-based precipitation datasets, generating a total of 16 experiments. Modeled evapotranspiration (ET), streamflow, and terrestrial water storage estimates were evaluated against the Atmosphere-Land Exchange Inverse (ALEXI) ET, in-situ streamflow observations, and NASA Gravity Recovery and Climate Experiment (GRACE) products, respectively. Results show that CLSMF2.5 provided better representation of the water budget variables than Noah3.3 in terms of Nash-Sutcliffe coefficient when considering all meteorological forcing datasets and precipitation datasets. The model experiments forced with observation-based products, the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) and the Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis (TMPA), outperform those run with MERRA-Land and MERRA-2 precipitation. The results presented in this paper would suggest that the Famine Early Warning Systems Network (FEWS NET)

Land Data Assimilation System incorporate CLSMF2.5 and HyMAP routing scheme to better represent the water balance in this region.

## 1. Introduction

Water balance has been a critical issue throughout the Nile river basin, especially associated with intense hydroclimatic extremes and the impacts on natural and human systems (Senay et al., 2009; FAO, 2011; Bastiaanssen et al., 2014). Studying water budget variables, such as precipitation (P), evapotranspiration (ET), streamflow and terrestrial water storage (TWS), provides improved understanding of water resources under a changing climate system (Simane et al., 2012; Tekleab et al., 2013; Berhane et al., 2014). In the Ethiopian portion of the Blue Nile basin (i.e. the upper Blue Nile), hydrological variability has had major implications for transboundary water supply (e.g. Mellander et al., 2013), periodic drought (e.g. Tadesse et al., 2014; Taye et al., 2015), regional food security (e.g. Shukla et al., 2014; Tadesse et al., 2015; McNally et al., 2016) and land use management (e.g. Gebrehiwot et al., 2011).

Several studies have shown that the water balance in East Africa is likely to shift under a changing climate (Lyon and DeWitt, 2012; Williams and Funk, 2011). The upper Blue Nile basin is vulnerable to negative climate change impacts along with significant interannual climate variability, complex topography, land cover modification, and continued population growth (Taye et al., 2015; Zaitchik et al., 2012). Over the past century, Ethiopia has become warmer with an increasing temperature over time (0.37 °C/decade) and experienced periodic droughts (Simane et al., 2012; Funk et al., 2014). Thus, the application of enhanced land surface models (LSMs) has been motivated by the need to provide enhanced seasonal prediction of hydro-climatic extremes and support

adaptation strategies under evolving climate conditions (e.g. Simane et al., 2014; Tadesse et al., 2014).

Previously, more streamflow and water balance studies in this region have been conducted relative to data sparse regions of Africa. The upper Blue Nile basin has two favorable conditions for hydrological modeling. First, the Ministry of Water and Energy of Ethiopia has been monitoring more than two dozen flow gauging stations and four dozen rainfall stations (Awulachew et al., 2008; Gebrehiwot et al., 2011; Taye et al., 2015). Second, this basin's local hydrology is less influenced by reservoirs when compared to the lower Blue Nile basin which includes Roseires Dam in Sudan, the White Nile basin which includes Lake Victoria, and the main Nile basin which includes Merowe Dam in Sudan and Aswan Dam in Egypt (FAO, 2011). Also, rain-fed agriculture is more dominant than irrigated cropland in the upper Blue Nile basin (Gebremichael et al., 2013). Therefore, incorporation of reservoir operation rules and irrigation impacts are less important for this regional hydrological modeling system.

Intercomparing models and datasets is an effective way to identify strengths and weaknesses of LSMs and meteorological forcings in different regions of the globe and at different spatio-temporal scales. In that sense, numerous model intercomparison initiatives have been created in the past (e.g. Henderson-Sellers et al., 1995; Boone et al., 2004, 2009a, 2009b; van den Hurk et al., 2011; Dirmeyer et al., 2006; Drobinski et al., 2014), providing guidance to the development of future generations of LSMs and improvements for Earth observing systems. In particular, coupling LSMs with river

routing schemes allows us to use in situ streamflow observations to evaluate the water budget at the basin scale (e.g. Oki et al., 1999; Getirana et al., 2014a, 2014b, 2017; Li et al., 2015; Zubieta et al., 2015).

However, little is known from comprehensive model intercomparisons and evaluations in the upper Blue Nile basin. Tekleab et al. (2011) calibrated a simple water balance model against observed streamflow time series and provided the water balances of twenty catchments in the upper Blue Nile basin. The catchment water balance was analyzed using an empirical relationship between the ratio of mean annual actual evaporation to mean annual rainfall and dryness index of the catchment. Bastiaanssen et al. (2014) calculated the annual water balance of 15 catchments in the Nile basin using rainfall data from Tropical Rainfall Measurement Mission (TRMM) and the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) RainFall Estimates (RFE) products in conjunction with actual evapotranspiration from the Operational Simplified Surface Energy Balance (SSEBop) and ETLook models. Also, Senay et al. (2009) estimated water balance components using annual satellite-derived variables such as runoff and evapotranspiration as a percent of rainfall albeit without model validation. Some studies focused on the small-scale water balance of the Lake Tana sub-basin in this region (e.g. Kebede et al., 2006; Setegne et al., 2008; 2010; Wale et al., 2009; Dessie et al., 2015). Other studies have provided parameter estimations for water balance models (e.g. Kim et al., 2008) and the analysis of runoff and sediment fluxes (e.g. Gebremicael et al., 2013; Steenhuis et al., 2009) in the upper Blue Nile.

The objectives of this study are to 1) identify the most suitable combination of LSMs, reanalysis and precipitation data for a water balance study in the upper Blue Nile basin; and 2) support the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS; McNally et al., 2017) by evaluating water budget components to ensure high quality of drought monitoring products. This study provides comparisons of multi-model inputs (i.e. precipitation) and output estimates (i.e. streamflow, terrestrial water storage anomaly, evapotranspiration) and evaluation of the water budget variables using in-situ measurements and satellite-based products. We run retrospective simulations to offer baseline knowledge for an improved modeling framework such as data assimilation or ensemble streamflow prediction. This study relies on remotely sensed data and the NASA Land Information System (LIS; Kumar et al., 2006).

This paper is organized as follows. Section 2 briefly introduces the design of this study, including the study area, land surface models, meteorological forcing data sets, observation-based precipitation data, the river routing scheme, and model setup. Section 3 describes how we evaluate our experiments using in-situ observations, satellite based measurements, and statistical indices. Section 4 presents results and discussion about the model intercomparison and evaluation. The last section summarizes our conclusions from this study.

## 2. Models and datasets

### *2.1. Study area*

The Blue Nile River is the largest tributary of the main Nile River, and its upper part is fully located in Ethiopia (**Fig. 1**). The upper Blue Nile River basin contributes about 60% of the annual streamflow to the main Nile River (UNESCO, 2004; Conway, 2005; Senay et al., 2014). The size of the basin (i.e. a drainage area upstream of the El Diem gaging station) is about 175,000 $km^2$ as compared to the whole Blue Nile basin area, which is about 325,000 $km^2$. Over 95% of the land cover in the basin consists of rain-fed cropland, grassland, wooded grassland, wood land, shrubs and bushes (Gebremicael et al., 2013). The headwater starts at Lake Tana in the Ethiopian highlands, and most of the flow originates from a large number of downstream tributaries. The river has cut a deep canyon through the highlands and drains a large portion of western Ethiopia (Elshamy et al., 2009). Elevations range from ~4000 m in the Ethiopian highlands to ~500 m at the Ethiopia-Sudan border. The precipitation in the upper Blue Nile basin is highly seasonal and subject to a tropical highland monsoon. Its main rainy season (i.e. Kiremt) occurs from June to September, a short rainy season (i.e. Belg) from March to May, and a dry season from October to May (Taye et al., 2015; Mellander et al., 2013; Zaitchik et al., 2012).

### *2.2. Land surface models*

The NASA Land Information System (LIS; Kumar et al., 2006) was used as the modeling platform in order to simulate land surface processes in the upper Blue Nile basin. LIS employs the use of high performance terrestrial hydrologic modelling with development led by the Hydrological Sciences Laboratory at NASA Goddard Space Flight Center

(GSFC). The LIS framework includes several community land surface models (LSMs) and supports their application at varying spatial and temporal scales over regional, continental and global domains. Two LSMs, the Noah land surface model, version 3.3, (Noah3.3) and Catchment Land Surface Model, version Fortuna 2.5 (CLSMF2.5), were used to drive the Hydrological Modeling and Analysis Platform (HyMAP; Getirana et al., 2012) river routing scheme and to simulate the hydrological processes in this study area.

The Noah LSM is maintained and released by the National Center for Atmospheric Research (NCAR) and applies finite difference spatial discretization methods and numerically solves the governing equations of the soil-vegetation-snowpack medium to simulate surface energy and water fluxes [more details can be found in Chen et al. (1996) and Ek et al. (2003)]. Noah is operationally used as the land model at the National Centers for Environmental Prediction (NCEP) for weather forecasts and the FLDAS simulation for use in hydro-climate studies and early warning applications.

The Catchment land surface model (CLSM) has been developed by the NASA Global Modeling and Assimilation Office (GMAO) and is the land-surface component of the Goddard Earth Observing System Model, Version 5 (GEOS-5) General Circulation Model (GCM). In contrast to the traditional gridded delineation, CLSM divides areas into irregularly shaped topographic catchments, which each contain a saturated fraction, a sub-saturated fraction, and a wilting fraction. These fractions evolve over time, and are used to determine fluxes and soil states within the catchment [more details can be found

in Koster et al., 2000; Reichle et al., 2011; Houborg et al., 2012]. The soil profile depths were compiled from the FAO/UNESCO Soil Map of the World (Webb et al., 1991).

While default values were used for most model parameters, some parameters were commonly used in both LSMs such as the NCEP monthly albedo, the NCEP monthly greenness, the combined Pennsylvania State University STATSGO and Food and Agriculture Organization of the United Nations (FAO) 16-category soil texture (Reynolds et al., 2000; Tegen et al., 2002), the Shuttle Radar Topography Mission (SRTM) topography elevation, and the Advanced Very High Resolution Radiometer (AVHRR) global land cover classification from the University of Maryland (Hansen et al., 2000).

### *2.3. Meteorological and precipitation datasets*

LSMs require meteorological forcing datasets (e.g. temperature, humidity, downward shortwave and longwave radiation, wind, and surface pressure). The Modern-Era Retrospective analysis for Research and Applications (MERRA) data product is a NASA atmospheric reanalysis for the satellite era using the Goddard Earth Observing System model (GEOS-5) and its associated data assimilation system (Reichle et al., 2011; Reichle, 2012). Our experiments were forced with two different MERRA versions: MERRA-Land and MERRA-2. MERRA-Land is a supplemental land surface data product of MERRA. MERRA-2 is the second version of MERRA with several major upgrades, including observation-based precipitation corrections over Africa (Reichle et al., 2017). Both products are available globally at the hourly time step and horizontal

resolution of 2/3° longitude by 1/2° latitude. These products begin in 1980, but only MERRA-2 system continues to be produced after January, 2017. MERRA-2 is updated operationally with a latency of about one week.

To examine the impact of observation-based precipitation products, three different satellite precipitation datasets were used as inputs to our simulations to compare with the MERRA-based precipitation reanalysis, generating several simulations. These observed products include: 1) the Climate Hazards group InfraRed Precipitation with Stations (CHIRPS), 2) the research-grade Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis version 7 (TMPA; referred to here as TMPA3B42), and 3) the near real-time version of the TMPA3B42 product (TMPA3B42RT).

CHIRPS is a quasi-global (50° S to 50° N) rainfall dataset based on infrared Cold Cloud Duration (CCD) observations, used for seasonal drought monitoring (Funk et al., 2015). The data is temporally disaggregated from daily to 6-hourly at 0.05° spatial resolution and from 1981 to present. CHIRPS incorporates station data with about a three week latency (Funk et al., 2015).

TMPA3B42 is a quasi-global (50° S to 50° N) precipitation product at 0.25° spatial resolution and 3-hourly temporal resolution, using the multi-channel microwave and infrared observations obtained from satellites (Huffman et al., 2007; Wanders et al., 2015). The TMPA3B42 algorithm uses infrared (IR) and passive microwave (PM) sensors and rescales the data based on gauge observations. TMPA3B42 is available from

1998 to present, usually with two months latency. TMPA3B42RT has spatial coverage (60° S to 60° N) and is available from 2000 to present. TMPA3B42RT is available with a latency of about 6 hours and without the gauge based adjustment. Dinku et al. (2007) showed that TMPA3B42RT in Ethiopia performs as well as TMPA3B42 in terms of correlation coefficient and mean error against gauge rainfall data. TMPA3B42RT even performs better than the NOAA Climate Prediction Center (CPC) African Rainfall Estimation Algorithm (RFE), version 2, which includes use of in-situ rainfall data from the Global Telecommunication System (GTS) stations.

### *2.4. River routing scheme*

In this study, HyMAP is driven with both Noah3.3 and CLSMF2.5. Surface runoff and baseflow are converted into streamflow along the river network using a kinematic wave formulation, allowing the comparison against in-situ observations [more details can be found in Getirana et al., 2012]. In river routing schemes, channel geometry, floodplain topography, and roughness coefficient are an acknowledged source of uncertainty (Decharme et al., 2012; Yamazaki et al., 2011; Getirana et al., 2013; Luo et al., 2017). We implemented the HyMAP's global standard parameters except for river width in this study.

Generally, HyMAP defines river width for each grid cell based on an empirical relationship between river width and the mean annual discharge from the global runoff database (Cogley 2003). In this study, we modified the HyMAP-based river width with the Global Width Database for Large Rivers (GWD-LR) (Yamazaki et al., 2014). GWD-

LR calculates effective bank-to-bank river width using the SRTM Water Body Database and the HydroSHEDS flow direction map (Lehner et al., 2008). While GWD-LR shows discontinuities of river width data in small tributaries, this data provides better estimates on the mainstem (greater than ~150 meter in river width) of our study area. Thus, we computed the ratio of GWD-LR to HyMAP defined river width on the mainstem, multiplied the HyMAP defined river width by the ratio, and updated river width input in the river routing scheme.

### *2.5. Model setup*

Both Noah3.3 and CLSMF2.5 were each driven with the two MERRA-based meteorological forcing datasets. Additional model experiments were conducted with each of the three additional precipitation datasets (i.e. CHIRPS, TMPA3B42, and TMPA3B42RT), replacing the MERRA-Land and MERRA-2 precipitation fields but still retaining the other meteorological forcing fields (e.g. air temperature, wind fields, etc.). **Table 1** lists the details of the 16 model experiments. The simulations were performed from 1981 to 2010. All experiments used a constant model time step of 15 minutes and produced on a 0.1° spatial resolution domain with daily-averaged output fields. Bilinear interpolation was applied to the coarser scale meteorological datasets to match the 0.1° resolution required by the LSMs. The first 25 years of simulation were used for model spinup for LSM variables to reach equilibrium (Rodell et al., 2005) and were not considered in the evaluation. The water budget evaluation was carried out for the 2006-2010 period when evaluation datasets (i.e. ET, Q, TWS; see the details in section 3) are most available. The selected LSMs are physically-based models, meaning that the

accuracy of results will depend on the quality of input data. Model calibration is beyond the scope of this study.

## 3. Evaluation approach

### *3.1. Evapotranspiration*

Monthly model-based evapotranspiration estimates were compared and evaluated against the Atmosphere-Land Exchange Inverse (ALEXI) (Yilmaz et al., 2014) and the Global Land Evaporation Amsterdam Model (GLEAM) v3.0a (Martens et al., 2016).

ALEXI was developed as a robust regional framework for a two-source (soil+vegetation) time-integrated model to evaluate the surface energy balance (Anderson et al., 1997). Yilmaz et al. (2014) demonstrated that ALEXI outperformed Noah LSM and the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD16 ET product to estimate ET over the Nile river basin. In this study, ALEXI ET rates are used as the reference for our evaluation. ALEXI is a daily product at 3 km spatial resolution and available from 2007 to 2012 in the Middle East and North Africa (MENA) region.

GLEAM is currently the only global evaporation model driven by microwave remote sensing observations (Miralles et al., 2011). The GLEAM v3.0a used in this study is a global dataset available from 1980 to 2014, dedicated to providing terrestrial evaporation and soil moisture from reanalysis net radiation and air temperature, a combination of gauge-based reanalysis, satellite-based precipitation, and satellite-based vegetation

optical depth (Martens et al., 2016). GLEAM datasets are provided on a 0.25° spatial resolution grid and with a daily temporal resolution.

### *3.2. Streamflow*

We evaluated monthly streamflow estimates from our LIS-HyMAP coupled modeling runs using available daily streamflow observations at the El Diem site (**Fig. 1)**, monitored by the Ministry of Water and Energy of Ethiopia (Uhlenbrook et al., 2012; Gebrehiwot et al., 2011), where the station is located at the basin outlet of this study domain. More than three and half years of observations within 2006-2009 at this station are referenced to evaluate our model streamflow estimates. This study focuses on basin scale water budget variables, but the other streamflow gauges in the study area are mostly located at tributaries with small drainage areas ($< 10^3$ $km^2$) or do not have valid observations during the evaluation period.

### *3.3. Terrestrial water storage anomaly*

The GRACE mission provides measurements of the spatiotemporal changes in Earth's gravity field. We examined monthly basin-averaged TWS anomalies using three GRACE spherical harmonic products from the Center for Space Research at the University of Texas (CSR), the NASA Jet Propulsion Laboratory (JPL), and the German Research Centre for Geosciences (GFZ). We used the latest dataset release, RL05 from GFZ and CSR and RL05.1 from JPL. These products are available from April 2002 to end of 2016. In this study, the land grid scaling coefficients were applied to the GRACE data when generating the time series of TWS anomalies (Landerer and Swenson, 2012).

When analyzing GRACE data, there is a trade-off between spatial resolution and accuracy, such that the fundamental temporal and spatial resolution of the GRACE data is 10 days and 400 km (Rowlands et al., 2005; Swenson et al., 2006). Our study area (~175,000 km$^2$) is slightly greater than the approximate minimum area (~160,000 km$^2$) that can be resolved before errors overwhelm the signal. TWS estimates were calculated using the continuity equation adapted for watersheds:

$$TWS(t) = \int (P(t) - ET(t) - Q(t))dt \quad (1)$$

where $P$ is the input precipitation and $ET$ is LSM-based evapotranspiration. $Q$ is derived from HyMAP at the basin outlet. We calculated both the GRACE and model-based TWS anomalies with a linear trend removal of years 2006-2010. The model performance statistics were computed against the mean of three GRACE spherical harmonic products. The mean absolute error of each of the GRACE products against the mean of the three is equivalent to 1 cm.

### *3.4. Evaluation indices*

The performance of LIS-HyMAP simulations is assessed using commonly used statistical indices. For statistical goodness of fit between observed and simulated values, the Nash-Sutcliffe (NS), the Pearson Correlation (r) coefficients, and the root-mean-square error (RMSE) are calculated as:

$$NS = 1 - \frac{\sum(\mathrm{OBS_t} - \mathrm{SIM_t})^2}{\sum(\mathrm{OBS_t} - \overline{\mathrm{OBS}})^2} \quad (2)$$

$$r = \frac{\sum(\mathrm{OBS_t} - \overline{\mathrm{OBS}})(\mathrm{SIM_t} - \overline{\mathrm{SIM}})}{\sqrt{\sum(\mathrm{OBS_t} - \overline{\mathrm{OBS}})^2}\sqrt{\sum(\mathrm{SIM_t} - \overline{\mathrm{SIM}})^2}} \quad (3)$$

$$RMSE = \sqrt{(\sum(\mathrm{OBS_t} - \mathrm{SIM_t})^2)/\mathrm{N}} \quad (4)$$

where $\mathrm{OBS_t}$ and $\mathrm{SIM_t}$ are, respectively, observed and simulated values at time step t, and $\overline{\mathrm{OBS}}$ and $\overline{\mathrm{SIM}}$ are averages of their respective time series. NS ranges from -∞ to 1, where 1 is the optimal case and zero is when simulations are as accurate as the long-term average of the observed values. If NS is lower than zero, then the model provides less skill than using the observed mean as a predictor. NS is more sensitive to large differences between $\mathrm{OBS_t}$ and $\mathrm{SIM_t}$ rather than r, thus NS can be a better evaluation index to assess the timing of peak values in ET and Q hydrological variables. RMSE is calculated only for TWS and not for ET and Q estimates.

For error assessment, relative error ($RE$) and the ratio of standard deviations (RSD) are calculated.

$$RE = \frac{\sum \mathrm{SIM} - \sum \mathrm{OBS}}{\sum \mathrm{OBS}} \quad (5)$$

$$RSD = \frac{\sigma_{\mathrm{SIM}}}{\sigma_{\mathrm{OBS}}} \quad (6)$$

where σ is standard deviation. $RE$ determines how $SIM_t$ is under- or overestimated in comparison to $OBS_t$ for the period studied. When used to evaluate simulated TWS, $RSD$ of simulated and GRACE-based TWS is calculated. $RSD$ compares the amplitudes of the simulated TWS time series against GRACE-based estimates, where values above 1 indicate that $SIM_t$ overestimates the amplitude.

## 4. Results and discussion

### *4.1. Comparison of model precipitation estimates*

The spatial and temporal distributions of precipitation fields have important roles in the water budget. **Fig. 2** shows total annual and mean monthly rates of basin-averaged precipitation from MERRA-Land (ML), MERRA-2 (M2), CHIRPS (CH), TMPA3B42 (TM), and TMPA3B42RT (TR) datasets for years 2006-2010. In **Fig. 2a**, the mean annual rates vary from 1071 mm/yr for ML to 1426 mm/yr for M2. The standard deviations vary from 101 mm/yr for CH to 187 mm/yr for ML. While M2 appears to follow a similar annual trend in the time series as ML, M2 is 33% higher than ML during the years 2006-2010. It implies that when compared to ML, a merged satellite-gauge precipitation product M2 was improved by its precipitation correction algorithm (Reichle et al., 2017). The other three datasets CH, TM, and TR agree well with annual rates within the range of the two MERRA precipitation products. In **Fig. 2b,** mean monthly precipitation values are plotted for the years, 2006-2010. The mean monthly rates vary from 2.93 mm/day for ML to 3.91 mm/day for M2. A larger deviation among the different precipitation data sets is shown for the rainy months from May to August versus

the non-rainy months. The peak month of the precipitation datasets is July except TR in which August is slightly higher than July.

**Fig. 3** shows the spatial distribution of precipitation datasets, temporally averaged at each grid cell for years, 2006-2010. The order of mean daily rates from highest to lowest is M2, CH, TM, TR, and ML, which is similar to the order shown with **Fig. 2**. Though ML shows the highest standard deviation in the time series of annual rates among five precipitation datasets in **Fig. 2a**, ML has the lowest standard deviation in its spatial distribution. M2 seems to have a local effect at longitudes between 36° and 38°, showing higher rates than the other areas. Overall, the average precipitation rates decrease towards the northeast, which is consistent with isohyet lines that Mellander et al. (2013) estimated for the same region. The highest mean daily precipitation rates are found in the southern region of the study area. The maximum rates vary from 4.19 mm/day for ML to 5.95 mm/day for M2.

It is noteworthy in **Figs. 2 and 3** that M2 and TM precipitation values increase when incorporating station-based observations compared to ML and TR, respectively. Modeled ET and total runoff present similar spatial distribution patterns (not shown) as those observed in the precipitation datasets. Thus, higher ET and total runoff rates usually occur in the southern region, coinciding with the higher precipitation rates described above.

### *4.2. Evaluation of model evapotranspiration estimates*

The 16 model-based monthly ET rates in the upper Blue Nile basin were evaluated from January 2007 to December 2010. This period is limited by the availability of the ALEXI ET data. **Table 2** provides the summary of evaluation indices of NS, RE, and r for the model monthly ET analyses against the ALEXI ET rates. A comparison of just the ML and M2 forced experiments (without replacing their precipitation data with the observation-based datasets) showed better ET estimates, with lower RE and higher r values, for M2 (compare ML-ML and M2-M2 in **Table 2**). However, model experiments forced with M2 and the observation-based precipitation datasets (M2-CH, M2-TM, M2-TR) did not make a significant improvement from those with ML (ML-CH, ML-TM, ML-TR) as much as an improvement between ML-ML and M2-M2. This indicates that the other meteorological variables (e.g. air temperature) of ML and M2, except for precipitation, had little influence on the performance of ET estimates (more detailed results and discussion are provided in section 4.5).

**Fig. 4** shows monthly modeled ET estimates from 8 simulations forced with the M2 meteorological forcing dataset (i.e. see simulation numbers 5-8 for Noah3.3 and 13-16 for CLSMF2.5 in **Table 1**). Overall, CLSMF2.5 is more consistent with ALEXI than Noah3.3, whereas GLEAM shows a distinct difference from the other datasets with overall lower values. All datasets show the highest ET values in September-October and the lowest ET values in February-March (**Figs. 4b,e**). The major difference between the two LSMs occurs during lowest period from December-March. CLSMF2.5 agrees well with ALEXI during both high and low periods, whereas Noah3.3 has decreased ET values, close to GLEAM values during those low periods. Yilmaz et al. (2014) also

demonstrated that Noah3.2 captures ET well during the rainy season, but it underestimated ET during the winter months over agricultural areas or regions with shallow ground water tables in the Nile Basin. In this sense, the Noah-MP LSM (Niu et al., 2011) may better represent ET values in this region due to being coupled with an underlying water table scheme, which is not supported in Noah3.3. On the other hand, ET estimates from CLSMF2.5 are close to ALEXI with high NS and r values. ET estimates from Noah3.3 show similarly high r values to CLSMF2.5, but low NS values (see **Table 2**). This supports that CLSMF2.5 better captures the timing of the low season values in monthly ET variability than Noah3.3.

Along with different precipitation datasets, CLSMF2.5 generates a wider range of ET estimates than Noah3.3. The rate of modeled ET from highest to lowest follows the same order of the precipitation rate as M2, CH, TM, and TR. Generally, Noah3.3 overestimates high values and underestimates low values with all precipitation datasets (**Figs. 4b, 4c**). CLSMF2.5 simulations when forced with M2 and CH generate higher ET estimates with positive RE values, whereas those with TM and TR generate lower ET estimates with negative RE values (**Figs. 4e, 4f**).

The mean annual rates of Noah3.3-M2 averaged (**Fig. 4a**), CLSMF2.5-M2 averaged (**Fig. 4d**), ALEXI, and GLEAM ET data are 909, 1010, 979, and 675 mm/yr, respectively, in the upper Blue Nile basin for 2007-2010. Previous ET studies in the Blue Nile basin have provided mean ET estimates at different periods. The Food and Agricultural Organization of the United Nations (FAO)-Nile program estimated mean ET as 863 mm/yr for 1960-

1990 (Hilhorst et al., 2011). Bastiaanssen et al., (2014) estimated mean ET as 737 mm/yr for years, 2005-2010, using the adjusted Operational Simplified Surface Energy Balance (SSEBop) model. Senay et al. (2009) calculated mean ET as 500 mm/yr for 2001-2007, based on standard water balance principles in the upper Blue Nile basin.

### *4.3. Evaluation of model streamflow estimates*

Monthly streamflow time series of the 16 model experiments were evaluated against gauge observations at the El Diem site (**Fig. 5**). NS, RE, and r have been calculated from January 2006 to September 2009, when observations are available (see **Table 2**). Overall, all modeled streamflow estimates are consistent with observations, resulting in high r values (>0.82). Similar to the ET evaluation, the choice of the other meteorological variables between ML and M2, except for precipitation, made little impact on the model Q estimates.

The choice between Noah3.3 and CLSMF2.5, along with different precipitation data, has a significant influence on the accuracy of model streamflow estimates. Generally, Noah3.3 produced higher streamflow estimates with positive RE values, whereas CLSMF2.5 produced lower streamflow with negative RE values, except those when forced with TM. This is mostly explained by the fact that Noah3.3 generates lower ET rates and higher total runoff, resulting in higher streamflows than CLSMF.2.5. More specifically, though both LSMs underestimate low streamflow periods (see **Figs. 5c, 5f**), Noah3.3 simulations provide much higher peak flows, whereas those from CLSM2.5 provide lower peak flows. Simulations from Noah3.3 show higher r values, whereas

simulations from CLSMF2.5 show higher NS values except those forced with TR. The high peaks of all 16 model experiments' streamflow time series occur in August (**Figs. 5b, 5e**), lagging by one month behind the high peaks of the precipitation rates, which occur in July (**Fig. 2b)**.

Interestingly, simulations from CLSMF2.5 show streamflow values rising earlier than those from Noah3.3 and the gauge observations (**Fig. 5e**). This is likely due to the fact that CLSMF2.5 produces little baseflow and makes the routing scheme convert most of the surface runoff to streamflow (Getirana et al., 2017). Baseflow corresponds to longer infiltration processes and slower runoff generation than surface runoff.

In terms of amplitude (see RE values in **Table 2**), the order of modeled Q from highest to lowest is different between Noah3.3 (M2-TM-CH-TR) and CLSMF2.5 (TM-TR-CH-M2). M2 produces the highest Q with Noah3.3, but the lowest Q in CLSMF2.5. This can be explained by noting that the different spatial distributions of the precipitation datasets produced different ratios of total runoff. These differences are also LSM-dependent.

### *4.4. Evaluation of model terrestrial water storage anomaly estimates*

The 16 different model TWS estimates were compared to GRACE data for January 2006 to December 2010. **Fig. 6** shows the 16 model basin-averaged TWS time series, and Table 2 highlights the evaluation indices when compared with the mean of the three GRACE spherical harmonic products. GRACE-based TWS time series show that the equivalent water height anomalies range ~±200 mm. If we convert this to storage change

by multiplying it by the basin area (~175,000 $km^2$), we get approximately ~±35 $km^3$ which is stored and then leaves the upper Blue Nile basin every year. Generally, the model outputs agree with GRACE estimates with high NS (>0.71) and r (>0.85) as well as RSD being close to 1 (0.80 to 1.35). Similar to the ET and Q evaluations, the choice of the other meteorological variables between ML and M2, except for precipitation, made little impact on the model TWS estimates.

For a comparison of the two LSMs, simulations from Noah3.3 show that the TWS anomalies peak in August one month earlier than those from CLSMF2.5 and GRACE. This is related to the fact that Noah3.3 underestimates ET during winter season and produces much high Q peak values. This leads to CLSMF2.5 having higher NS and r values than Noah3.3. Also, CLSMF2.5 shows higher RSD than Noah3.3 in the TWS anomaly evaluation. This implies that Noah3.3 produces higher streamflow values and results in lower amplitudes of TWS anomalies, despite lower ET rates than CLSMF2.5 (see **Eq. 1**).

The RMSE for TWS for all experiments, except for Noah3.3-ML-TR, are less than 61 mm, which is the sum of the leakage error and the residual error in the scaled GRACE data for this region (Swenson and Wahr, 2002).

### *4.5. A comparative analysis of water budget variables*

A comparative analysis of the three water budget variables (i.e. ET, Q, and TWS) was performed with the 16 LSM experiments and with the HyMAP routing scheme, and the

results are shown in **Fig. 7**. The indicated values for each shaded bar in the graph represents the average of the three water budget variables and for each derived evaluation index and for each experiment. The NS and correlation (r) coefficients were calculated for all model water budget variables. While relative errors (RE) were calculated for both Q and ET, the ratio of standard deviations (RSD) was calculated for TWS anomaly. When averaging the RE and RSD evaluation indices together, we subtracted 1 from RSD and calculated the average of the rescaled index with RE.

In terms of NS, CLSMF2.5 shows better performance than Noah3.3 in each evaluation index of the three water budget variables. CLSMF2.5 produces Q and TWS estimates with slightly higher NS values than Noah3.3. For ET estimates, CLSMF2.5 outperforms Noah3.3 because Noah3.3 underestimates ET rates during the winter months. When CLSMF2.5 was forced with the ML or M2 meteorological forcing dataset and CH, TM, TR precipitation datasets, the average of the three evaluation indices is about 0.8. CLSMF2.5 when forced with M2-M2 produces high NS values for Q and TWS, but shows lower NS than one for ET with the other precipitation datasets.

In terms of RE or RSD-1, all experiments show the average values within +/- 0.06 except those when forced with ML-ML. In ET, all experiments show negative RE values except CLSMF2.5 when forced with CH and M2 precipitation datasets. Overall, Noah3.3 provides lower RE in ET, higher RE in Q, and lower RSD values in TWS than CLSMF2.5.

In terms of r, all experiments show high average r (> 0.88) values. CLSMF2.5 when forced with TR precipitation shows lower r values than simulations with the other precipitation datasets. Overall, Noah3.3 shows higher correlation values in Q and lower correlation values in TWS than CLSMF2.5.

Simulations forced with CH, TM or TR precipitation datasets outperform those with either the ML or M2 precipitation fields, as shown in the average evaluation indices of these water budget variables. For r, CH has the highest correlation values in both LSMs, whereas TM performs well with Noah3.3. For the evaluation of RE or RSD, CH has lower errors in both LSMs' set of experiments. For the evaluation of NS, TM is the best with both LSMs, whereas CH performs well with CLSMF2.5.

## 5. Conclusions

This study focuses on the evaluation of the water budget over the upper Blue Nile basin from a modeling perspective. We generated 16 different model experiments including two LSMs (Noah3.3 and CLSMF2.5), two reanalysis forcing datasets (MERRA-Land and MERRA2), and three additional observation-based precipitation datasets (CHIRPS, TMPA3B42, and TMPA3B42RT). The HyMAP model was used to route surface runoff and baseflow in the river network to generate streamflow. These particular models and forcing datasets were chosen in the existing framework of NASA's LIS software. The spatial and temporal distributions of precipitation (P) fields were investigated. Three water budget variables (i.e. evapotranspiration (ET), streamflow (Q), and terrestrial water

storage (TWS) anomaly) were compared and evaluated using in-situ and satellite observations. The intercomparision and evaluation of these models and datasets offered improved understanding and modeling of basin scale water budget variables in the upper Blue Nile basin.

Among the five different precipitation datasets, the gauge-adjusted products such as MERRA-2, CHIRPS, and TMPA3B42 were expected to provide better spatial and temporal distributions of precipitation over the basin. While precipitation from MERRA-2 showed improved precipitation relative to precipitation from MERRA-Land, little improvement in the other meteorological variables was seen in the evaluation of modeled ET, Q, and TWS anomaly estimates. The spatial distribution of precipitation from MERRA-2 appears to include a local bias and needs to be modified. In the upper Blue Nile basin, a real-time version of TMPA3B42 (i.e. TMPA3B42RT) outperformed MERRA2 and MERRA-Land with the evaluation of NS and RE.

Significant uncertainty in evapotranspiration analyses has been a known issue in the region (Yilmaz et al., 2014). In this study, we evaluated the 16 modeled ET estimates against ALEXI data, which have also been used to validate actual ET estimates in this region (Allam et al., 2016). The results showed that Noah3.3 produces lower ET and higher total runoff (or streamflow) than CLSMF2.5 when using the same precipitation data. This is likely caused by low baseflow produced by CLSMF2.5. The evaluation of Q was performed against in-situ observations at the Ethiopia-Sudan border. All experiments forced with the CH, TM and TR observation-based precipitation datasets showed high

model streamflow performances. In further studies, more streamflow gauge stations along the upstream tributaries can be analyzed to explore smaller scale hydrological and hydraulic processes related to forcing error, limited representation of physical processes, and inaccurate parameterization of the routing scheme. Simulated TWS anomalies were compared against the mean of three GRACE spherical harmonic products. All experiments, except for Noah3.3-ML-TR, provide TWS errors that are less than the published leakage error for this region (Swenson and Wahr, 2002).

The comparative analysis shows that CLSMF2.5 provided better representation of the water budget variables in terms of Nash-Sutcliffe coefficient, though CLSMF2.5 produces little baseflow in the runoff generation process. In the upper Blue Nile basin, simulations forced with CHIRPS or TMPA3B42 precipitation data show better modeled ET, Q, and TWS anomaly estimates than the other precipitation datasets. Currently, FLDAS operationally runs Noah3.3 forced with MERRA2 and CHIRPS and provides these model outputs to a public archive server. However, FLDAS does not currently include CLSMF2.5 and HyMAP routing scheme outputs in its routine production. In regards to this, the results from this study could suggest that by including CLSM and HyMAP in the FLDAS production suite in the future, additional information would be provided for enhancing the drought monitoring for FEWS NET applications.

The evaluation can be sensitive to different temporal and spatial model resolutions. In this study, monthly modeled ET, Q, TWS estimates at 0.1° grid scale are evaluated. The temporal resolution is constrained by the availability of GRACE products. The spatial

resolution is matched with the FLDAS outputs. Further studies are needed to carry out the evaluation at different resolutions and reveal the impact of the scales onto the comparative analysis of water budget variables in this region.

In this study, we didn't attempt to evaluate different precipitation datasets as compared to the other three water budget variables. Previous studies (Funk et al., 2014, Huffman et al., 2007, Reichle et al., 2017) have already demonstrated that by incorporating in situ observation, precipitation datasets CHIRPS, MERRA-2, TMPA3B42 are better than CHIRP, MERRA-Land, TMPA3B42RT, respectively. Besides, ground observations are generally point data whereas these satellite- and/or model-based precipitation datasets represents averages in large scale. Therefore, we focus on the evaluating the impact of precipitation datasets on modeled ET, Q and TWS estimates rather than directly evaluating each of the five different precipitation datasets.

These results are essential to continue to make potential improvements in the parametrizations, physics, calibration of the LSMs being used. Also, these analyses are useful for future applications such as seasonal forecast modeling, agricultural production estimates, water resource management, and model algorithm development (e.g. data assimilation, irrigation, etc.) within the NASA LIS framework.

***Acknowledgments***. This research was supported by NASA Interdisciplinary Research in Earth Science Program (NNH12ZDA001N-IDS) and NASA Earth Science Applications (NNH13ZDA001N-WATER) and FEWS NET's Participating Agency

Program Agreement Water Availability Monitoring Activity. Computing was supported by the resources at the NASA Center for Climate Simulation (NCCS). We thank Martha Anderson for the ALEXI ET data. Streamflow data were provided by the Ministry of Water and Energy of Ethiopia. The GRACE data are provided by the GRCTellus (http://grace.jpl.nasa.gov), which is supported by the NASA MEaSUREs Program.

## REFERNCES

Allam, M.M., Figueroa, A.J., McLaughlin, D.B., Eltahir, E.A.B., 2016. Estimation of evaporation over the upper Blue Nile basin by combining observations from satellites and river flow gauges. *Water Resour. Res*., 52, 644–659, doi:10.1002/2015WR017251.

Anderson, M.C., Norman, J.M., Diak, G.R., Kutas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ*., 60, 195–216.

Awulachew, S.B., Mccartney , M., Steenhuis, T.S., Ahmed, A.A., 2008. A Review of Hydrology, Sediment and Water Resource Use in the Blue Nile Basin. International Water Management Institute, Colombo, Sri Lanka (IWMI Working Paper 131).

Bastiaanssen, W.G.M., Karimi, P., Rebelo, L., Duan, Z., Senay, G., Muthuwatte, L., Smakhtin, V., 2014. Earth Observation Based Assessment of the Water Production and Water Consumption of Nile Basin Agro-Ecosystems. *Remote Sens.*, 2014, *6*, 10306-10334; doi:10.3390/rs61110306

Berhane, F., Zaitchik , B., Dezfuli, A., 2014. Subseasonal Analysis of Precipitation Variability in the Blue Nile River Basin. J. Hydrometeor., 27, 325-344, doi: 10.1175/JCLI-D-13-00094.1.

Boone, A. Getirana, A.C.V., Demarty, J., Cappelaere, B., Galle, S., Grippa, M., Lebel, T., Mougin, E., Peugeot, C., Vischel, T., 2009a. AMMA Land Surface Model Intercomparison Project Phase 2, (ALMIP-2). Gewex News, 9(4), 9-10.

Boone, A., de Rosnay, P., 24 co-authors, 2009b. The AMMA Land Surface Model Intercomparison Project. *Bull. Amer. Meteor.* Soc., 90(12), 1865- 1880. doi:10.1175/2009BAMS2786.1

Boone, A., Habets, F., Noilhan , J., 20 co-authors, 2004. The Rhone-Aggregation Land Surface Scheme Intercomparison Project: An Overview. J. Climate, 17, 187-208.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H., Koren, V., Duan, Y., Ek, M., Betts, A., 1996. Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res*., 101, 7251-7268.

Cogley, J.G., 2003. GGHYDRO—Global hydrographic data, release 2.3. Trent Tech. Note 2003-1, 11.

Conway, D., 2005. From head water tributaries to international river: Observing and adapting to climate variability and change in the Nile Basin. *Global Environ. Change*, 15, 99–114.

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., Prigent, C., 2012. Global off-line evaluation of the ISBA-TRIP flood model. *Climate Dyn.,* 38, 1389–1412, doi:10.1007/s00382-011-1054-9.

Dessie, M., Verhoest, N.E.C., Pauwels, V.R.N., Adgo, E., Deckers, J., Poesen, J., Nyssen, J., 2015. Water balance of a lake with floodplain buffering: Lake Tana, Blue Nile Basin, Ethiopia. *J. Hydrol.*, 522, 174–186.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.J., Ropelewski, C.F., 2007. Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, 28, 1503-1526, doi: 10.1080/01431160600954688.

Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., 2006. GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, 87, 1381–1397, doi:10.1175/BAMS-87-10-1381.

Drobinski, P., coauthors, 2014. HyMeX: A 10-year multi- disciplinary program on the Mediterranean water cycle. *Bull. Amer. Meteor. Soc.*, 95, 1063–1082, doi:10.1175/ BAMS-D-12-00242.1.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grummann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational Mesoscale Eta Model. *J. Geophys. Res.*, 108, 8851, doi:10.1029/2002JD003296.

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.

FAO – Food and Agriculture Organization of the United Nations, 2011. Information products for Nile Basin Water Resources Management, Synthesis Report, FAO-Nile Basin Project GCP/INT/945/ITA 2004 to 2009.

Funk, C., Hoell, A., Shukla, S., Blade, I., Liebmann, B., Roverts, J.B., Roberson, F.R., Husak, G., 2014. Predicting East African spring droughts using Pacific and Indian Ocean sea surface temperature indices. *Hydrol. Earth Syst. Sci*., 18, 4965–4978, doi:10.5194/hess-18-4965-2014.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards

infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data*, 2:150066.

Gebrehiwot, S.G., Ilstedt, U., Gardenas, A.I., Bishop, K., 2011. Hydrological characterization of watersheds in the Blue Nile Basin, Ethiopia, *Hydrol. Earth Syst. Sci.*, 15, 11-20, doi: 10.5194/hess-15-11-2011.

Gebremicael, T.G., Mohamed, Y.A., Betrie, G.D., van der Zaag, P., Teferi, E., 2013. Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps. *J. Hydrol*., 482, 57-68.

Getirana, A., Boone, A., Peugeot, C., 2017. Streamflows over a West African basin from the ALMIP-2 model ensemble. *J. Hydrometeor*., doi: http://dx.doi.org/10.1175/JHM-D-16-0233.1.

Getirana, A., McNally, A., Roningen, J., Zaitchik, B., Arsenault, K., Jung, H.C., Peters-Lidard, C., 2015. Forecasting Water Availability in Data Sparse and Heavily Managed Catchments in Africa and the Middle East. *GEWEX NEWS*, 27(4), 8-11.

Getirana, A., Boone, A., Peugeot, C., 2014a. Evaluating LSM-based water budgets over a West African basin assisted with a river routing scheme. *J. Hydrometeor*., doi: 10.1175/JHM-D-14-0012.1

Getirana, A., Dutra, E., Guimberteau, M., Kam, J., Li, H., Decharme, B., Zhang, Z., Ducharne, A., Boone, A., Balsamo, G., Rodell, M., Toure, A.M., Xue, Y., Drapeau, G., Arsenault, K., Kumar, S.V., Leung, L. R., Peters-Lidard, C., Ronchail, J., Sheffield, J., 2014b. Water balance in the Amazon basin from a land surface model ensemble. *J. Hydrometeor*., doi: 10.1175/JHM-D-14-0068.1

Getirana, A.C.V., Boone, A., Yamazaki, D., Mognard, N., 2013. Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon basin. *Water Resources Research*. DOI: 10.1002/wrcr.20077.

Getirana, A.C.V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., Mognard, N., 2012. The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon basin. *J. Hydrometeor*., 13, 1641–1665, doi:10.1175/JHM-D-12-021.1.

Hansen, M., DeFries, R., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover classification at 1km resolution using a decision tree classifier. *Int. J. Remote Sens.*, 21: 1331-1365.

Henderson-Sellers, A., Pitman, A.J., Love, P.K., Irannejad, P., Chen, T., 1995. The Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, 76, 489–503.

Hilhorst, B., Burke, J.,, Hoogeveen, J., Fremken, K., Faures, J.M., Gross, D., 2011. Information Products for Nile Basin Water Resources Management; FAO: Rome, Italy.

Houborg, R., Rodell, M., Li, B., Reichle, R., Zaitchik, B.F., 2012. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.*, 48, W07525, doi:10.1029/2011WR011291.

Kebede, S., Travi, Y., Alemayehu, T., Marc, V., 2006. Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia, *J. Hydrol*., 316, 233–247.

Kim, U., Kaluarachchi, J.J., 2008. Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia. *J. Hydrol*., 362, 39-56.

Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., Kumar, P., 2000. A catchment based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.-Atmos*., 105, 24809–24822, doi:10.1029/2000JD900327.

Kumar, S.V., Peters-Lidard, C.D., Tian, Y., Geiger, J., Houser, P.R., Olden, S., Lighty, L., Eastman, J.L., Dirmeyer, P., Doty, B., Adams, J., Wood, E.F., Sheffield, J., 2006. LIS – An interoperable framework for high resolution land surface modeling, *Environ. Modell. Software*, 21, 1402-1415.

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, 48, W04531, doi:10.1029/2011WR011453.

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. *Eos Trans. AGU*, 89, doi:10.1029/2008EO100001.

Li, H., Leung, L.R., Getirana, A., Huang, M., Wu, H., Xu, Y., Guo, J., Voisin, N., 2015. Evaluating Global Streamflow Simulations by a Physically-based Routing Model Coupled with the Community Land Model. *J. Hydrometeor*., DOI: 10.1175/JHM-D-14-0079.1.

Luo, X., Li, H., Leung, L.R., Tesfa, T.K., Getirana, A., Papa, F., Hess, L.L., 2017. Modeling surface water dynamics in the Amazon Basin using MOSART-Inundation-v1.0: Impacts of geomorphological parameters and river flow

representation. *Geoscientific Model Development*, 10, 1233–1259. DOI: 10.5194/gmd-10-1233-2017.

Lyon, B., DeWitt, D., 2012. A recent and abrupt decline in the East African long rains. *Geophys. Res. Lett.*, 39, L02702, doi:10.1029/2011GL050337.

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2016. GLEAM v3.0: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev. Discuss.*, doi: 10.5194/gmd-2016-162.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wnag, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data*, doi: 10.1038/sdata.2017.12.

McNally, A., Shukla, S., Arsenault, K.R., Wang, S., Peters-Lidard, C.D., Verdin, J.P., 2016. Evaluating ESA CCI soil moisture in East Africa. *Int. J. Appl. Earth Obs.*, 48, 96–109.

Mellander, P.E, Gebrehiwot, S.G., Gardenas, A.I., Bewket, W., Bishop, K., 2013. Summer rains and dry seasons in the Upper Blue Nile Basin: the predictability of half a century of past and future spatiotemporal patterns. *PLoS ONE*, 8, e68461, doi:10.1371/journal.pone.0068461.

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453–469, doi:10.5194/hess-15-453-2011.

Niu, G.Y., Yang, Z.L., Mitchell, K. E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): Model description and evaluation with local-scale measurements. *J. Geophys. Res*., 116, D12109, doi:10.1029/2010JD015139.

Oki, T., Nishimura, T., Dirmeyer, P., 1999. Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *J. Meteor. Soc*. Japan, 77, 235–255.

Reichle, R.H., 2012. The MERRA-Land Data Product. GMAO Office Note No. 3, Version 1.2.

Reichle, R.H., Liu, Q., Koster, R.D., Draper, C.S., Mahanama, S.P.P., Partyka, G.S., 2017. Land Surface Precipitation in MERRA-2. *Journal of Climate*, 30, 1643-1664, doi: 10.1175/jcli-d-16-0570.1.

Reichle, R.H., Koster, R.D., De Lannoy, G.J.M., Forman, B.A., Liu, A., Mahanama, S.P.P., Toure, A., 2011. Assessment and enhancement of MERRA land surface hydrology estimates, *J. Climate*, 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.

Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resour. Res*., 36, 3653–3662.

Rodell, M., Houser, P.R., Berg, A.A., Famiglietti, J.S., 2005. Evaluation of 10 methods for initializing a land surface model. *J. Hydrometeor*., 6, 146–155, doi:10.1175/JHM414.1.

Rowlands, D.D., Luthcke, S.B., Klosko, S.M., Lemoine, F.G.R., Chinn, D.S., McCarthy, J.J., Cox, C.M., Anderson, O.B., 2005. Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. Geophys. Res. Lett., 32, L04310, doi:10.1029/2004GL021908.

Senay, G.B., Velpuri, N.M., Bohms, S., Demissie, Y., Gebremichael, M., 2014. Understanding the hydrologic sources and sinks in the Nile Basin using multisource climate and remote sensing data sets. *Water Resour. Res.*, 50, 8625–8650, doi:10.1002/2013WR015231.

Senay, G.B., Assante, K., Artan, G., 2009. Water balance dynamics in the Nile Basin. *Hydrol. Process.*, 23, 3675–3681, doi: 10.1002/hyp.7364.

Setegne, 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.

Setegne, S.G., Srinivasan, R., Dargahi, B., 2008. Hydrological Modelling in the Lake Tana Basin, Ethiopia using SWAT model. *Open Hydrol. J.*, 2, 24–40.

Shukla, S., McNally, A., Husak, G., Funk, C., 2014. A seasonal agricultural drought forecast system for food-insecure regions of East Africa. *Hydrol. Earth Syst. Sci.*, 18, 3907–3921, doi:10.5194/hess-18-3907-2014.

Simane, B., Zaitchik, B.F., 2014. The Sustainability of Community-Based Adaptation Projects in the Blue Nile Highlands of Ethiopia. *Sustainability*, *6*, 4308-4325; doi:10.3390/su6074308.

Simane, B., Zaitchik, B.F., Mesfin, D., 2012. Building climate resilience in the Blue Nile/Abay Highlands: a framework for action. *Int J Environ Res Public Health*, 9(2), 610-631, doi: 10.3390/ijerph9020610.

Steenhuis, T.S., Collick, A.S., Easton, Z.M., Leggesse, E.S., Bayabil, H.K., White, E.D., Awulachew, S.B., Adgo, E., Ahmed, A.A., 2009. Predicting discharge and sediment for the Abay (Blue Nile) with a simple model. *Hydrol. Process.* 23, 3728–3737, doi: 10.1002/hyp.7513.

Swenson, S.C., Wahr, J., 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *J. Geophys. Res.*, 107, 2193, doi:10.1029/2001JB000576.

Swenson, S., Yeh, P.J.F., Wahr, J., Famiglietti, J., 2006. A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophys. Res. Lett.*, 33, L16401, doi:10.1029/2006GL026962.

Tadesse, T., Bathke, D., Wall, N., Peter, J., Haigh, T., 2015. Participatory research workshop on seasonal prediction of hydro-climatic extremes in the Greater Horn of Africa. *B. Am. Meteorol. Soc.*, 96, doi: 10.1175/BAMS-D-14-00280.1

Tadesse, T., Demisse, G.B., Zaitchik, B., Dinku, T., 2014. Satellite-based hybrid drought monitoring tool for prediction of vegetation condition in Eastern Africa: A case study for Ethiopia. *Water Resour. Res.*, 50, 2176–2190, doi:10.1002/2013WR014281.

Taye, M.T., Willems, P., Block, P., 2015. Implications of climate change on hydrological extremes inthe Blue Nile basin: A review. *J. Hydrol. Reg. Stud.* 4, 280-293.

Tegen, I., Harrison, S.P., Kohfeld, K., Prentice, I.C., Coe, M., Heimann, M., 2002. Impact of vegetation and preferential source areas on global dust aerosol: Results from a model study. *J. Geophys. Res*., 107,14-27, doi:10.1029/2001JD000963.

Tekleab, S., Uhlenbrook, S., Mohamed, Y., Savenije, H.H.G., Temesgen, M., Wenninger, J., 2011. Water balance modeling of Upper Blue Nile catchments using a top-down approach. *Hydrol. Earth Syst. Sci*., 15, 2179–2193, doi:10.5194/hess-15-2179-2011

Tekleab, S., Mohamned, Y., Uhlenbrook, S., 2013. Hydro-climatic trends in the Abay/Upper Blue Nile basin, Ethiopia. *Phys. Chem. Earth*, 61–62, 32–42.

Uhlenbrook, S., Mohamed, Y., Gragne, A.S., 2010. Analyzing catchment behavior through catchment modeling in the Gilgel Abay, Upper Blue Nile River Basin, Ethiopia. *Hydrol. Earth Syst. Sci.*, 14, 2153–2165, doi:10.5194/hess-14-2153-2010.

UNESCO – United Nations, Educational, Scientific and Cultural Organization, 2004. National Water Development Report for Ethiopia, UN-WATER/WWAP/2006/7,WorldWater Assessment program, Report, MOWR, Addis Ababa, Ethiopia.

van den Hurk, B., Best, M., Dirmeyer, P., Pitman, A., Polcher, J., Santanello, J., 2011. Acceleration of land surface model de- velopment over a decade of GLASS. *Bull. Amer. Meteor. Soc*., 92, 1593–1600, doi:10.1175/BAMS-D-11-00007.1.

Wale, A., Rientjes, T.H.M., Gieske, A.S.M., Getachew, H.A., 2009. Ungauged catchment contributions to Lake Tana's water balance. *Hydrol. Process*., 23, 3682–3693.

Webb, R., Rosenzweig, C., Levine, E., 1991. A global data set of soil particle size properties, NASA Tech. Memo., 4286, 1–29.

Williams, A.P., Funk, F., 2011. A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. *Clim. Dyn*., 37, 2417-2435.

Yamazaki, D., O'Loughlin, F., Trigg, M.A., Miller, Z.F., Pavelsky, T.M., Bates, P.D., 2014. Development of the Global Width Database for Large Rivers, *Water Resour. Res.*, 50, 3467–3480, doi:10.1002/2013WR014664.

Yamazaki, D., Kanae, S., Kim, H., Oki, O., 2011. A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, 47, W04501, doi:10.1029/2010WR009726.

Yilmaz, M.T., Anderson, M.C., Zaitchik, B., Hain, C.R., Crow, W.T., Ozdogan, M., Chun, J.A., Evans, J., 2014. Comparison of prognostic and diagnositic surface flux modeling approaches over the Nile River basin. *Water Resour. Res.*, 50, 386–408, doi:10.1002/2013WR014194.

Zaitchik, B.F., Simane, B., Habib, S., Anderson, M.C., Ozdogan, M., Foltz, J.D., 2012. Building Climate Resilience in the Blue Nile/Abay Highlands: A Role for Earth System Sciences. *Int. J. Environ. Res. Public Health*, *9*, 435-461; doi:10.3390/ijerph9020435.

Zubieta, R., Getirana, A., Espinoza, J.C., Lavado, W., Saavedra, M., 2015. Impacts of satellite-based precipitation datasets on the rainfall-runoff modeling of the Western Amazon basin (Peru and Ecuador). *J. Hydrometeor.*, 528, 599–612. doi: 10.1016/j.jhydrol.2015.06.064.

## Figure Captions

**Fig. 1.** Study area of the upper Blue Nile river basin with river networks, country borders, and location of the streamflow station at the El Diem site.

**Fig. 2.** (a) Annual and (b) mean monthly precipitation rates from MERRA-Land (ML), MERRA-2 (M2), CHIRPS (CH), TMPA3B42 (TM), and TMPA3B42RT (TR) datasets in the upper Blue Nile basin for 2006-2010.

**Fig. 3.** Spatial distribution of mean daily precipitation rates (mm/day) in the upper Blue Nile basin for 2006-2010. Gray lines represent the mainstem and tributaries.

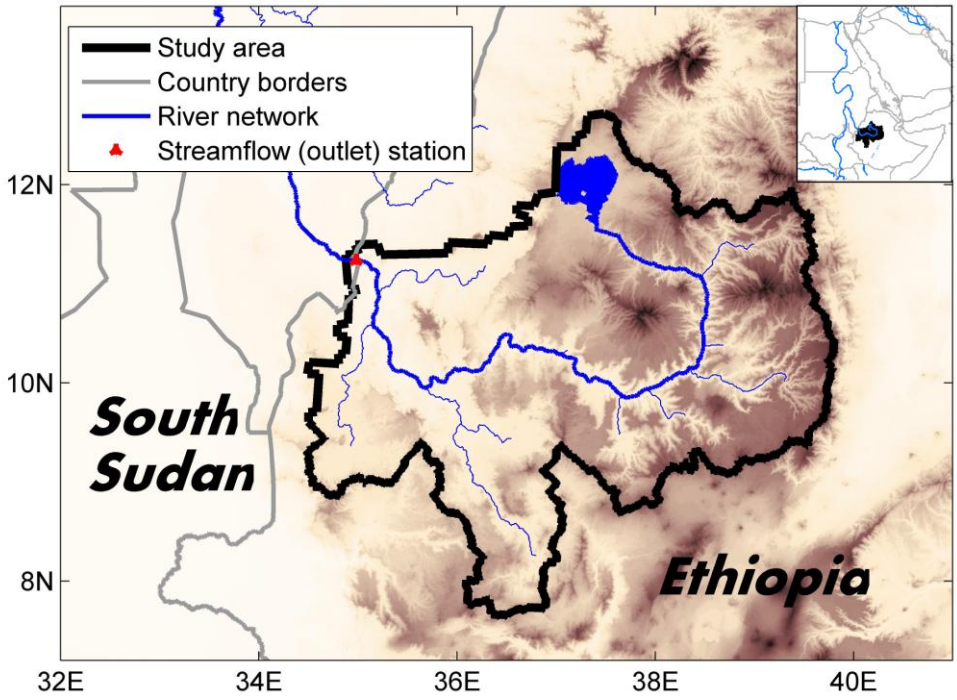
**Fig. 4.** Monthly evapotranspiration (ET) rates from Noah3.3-MERRA2 (a, b, c) and CLSMF2.5-MERRA2 (d, e, f) simulations for years, 2006-2010. The seasonality (b, e) and scatterplot (c, f) of the model monthly ET rates were evaluated against ALEXI ET rates from January 2007 to December 2010.

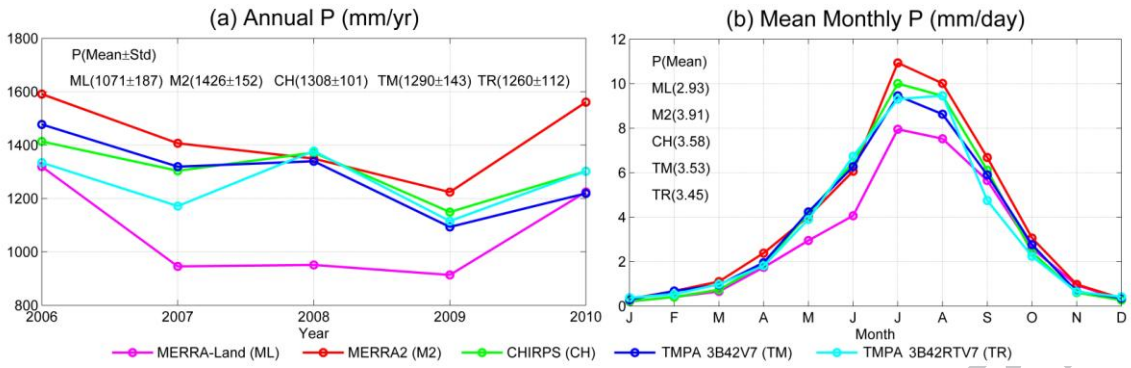
**Fig. 5.** Monthly streamflow rates (Q) at the EL Diem site from Noah3.3-MERRA2 (a, b, c) and CLSMF2.5-MERRA2 (d, e, f) simulations for years, 2006-2010. The seasonality (b, e) and scatterplot (c, f) of the model monthly streamflow rates were evaluated against gauge observations at the El Diem site from January 2006 to September 2009.

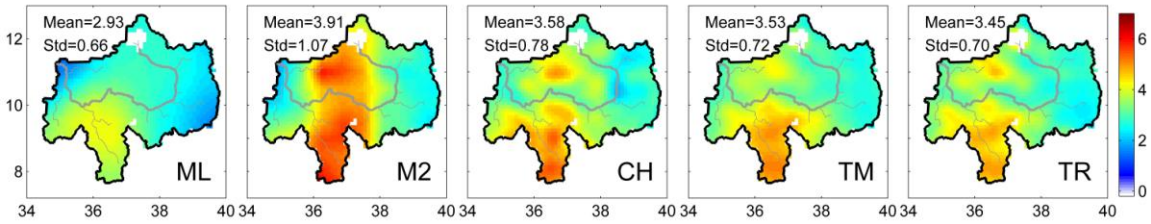
**Fig. 6.** Monthly terrestrial water storage (TWS) anomalies of the upper Blue Nile River Basin from Noah3.3-MERRA2 (a, b, c) and CLSMF2.5-MERRA2 (d, e, f) simulations

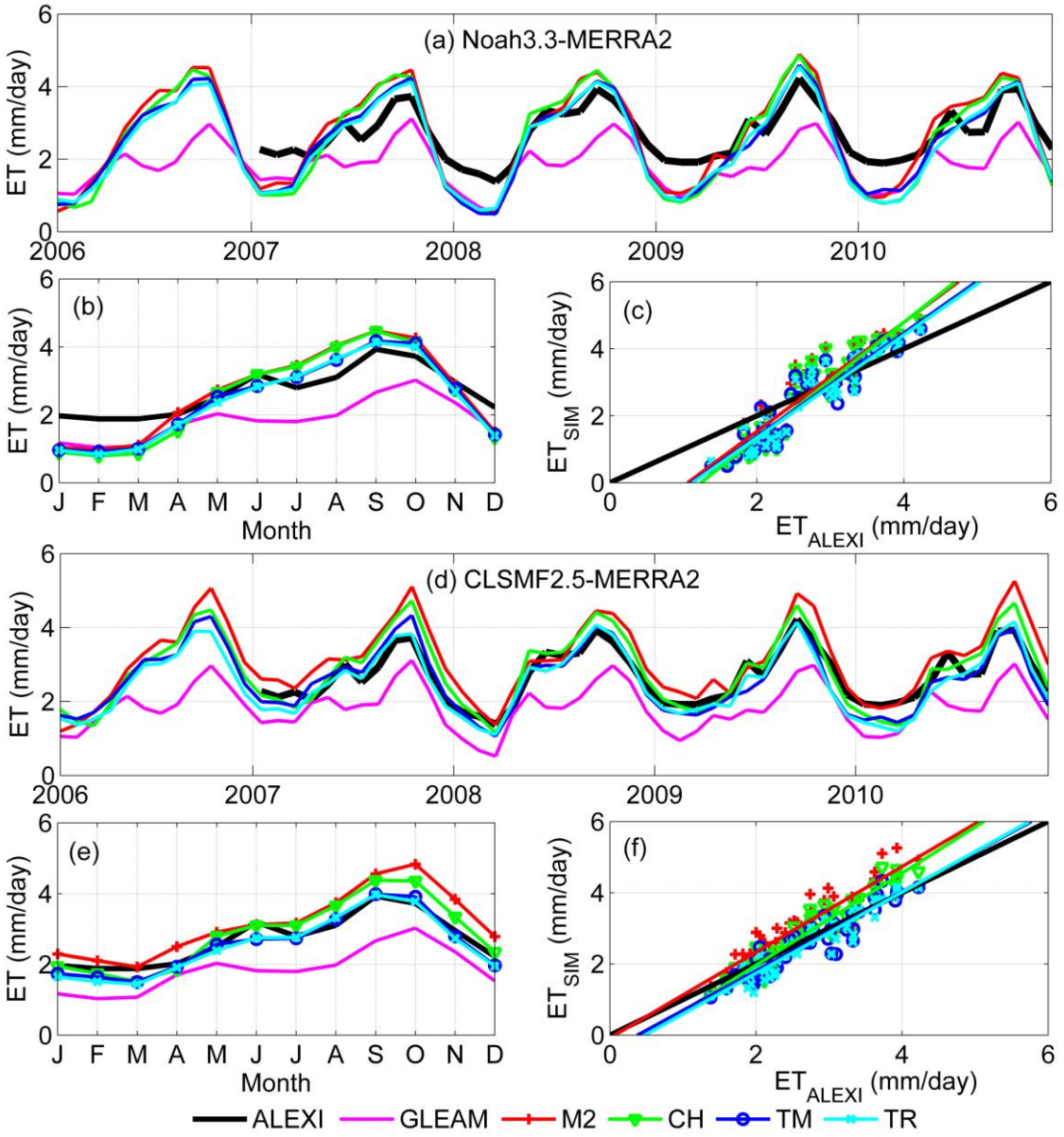
for 2006-2010. The seasonality (b, e) and scatterplot (c, f) of the model monthly TWS anomalies were evaluated against the average of three GRACE spherical harmonic products from CSR, JPL, and GFZ.

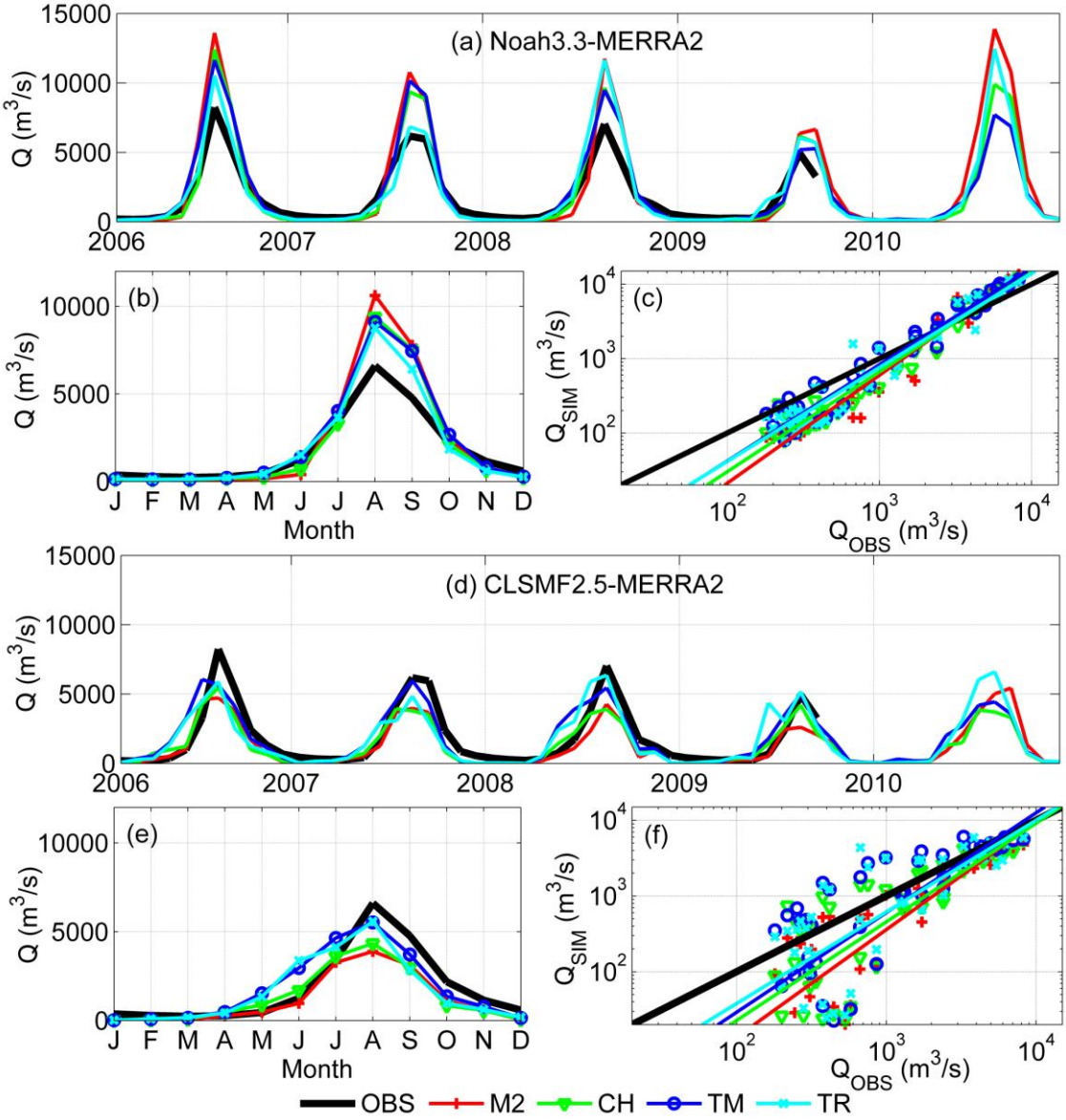
**Fig. 7.** A comparative analysis of three water budget variables (i.e. ET, Q, and TWS) from the 16 LSM simulations and the HyMAP river routing scheme (▾ These Noah3.3-M2-CH model output fields are operationally generated by FLDAS).

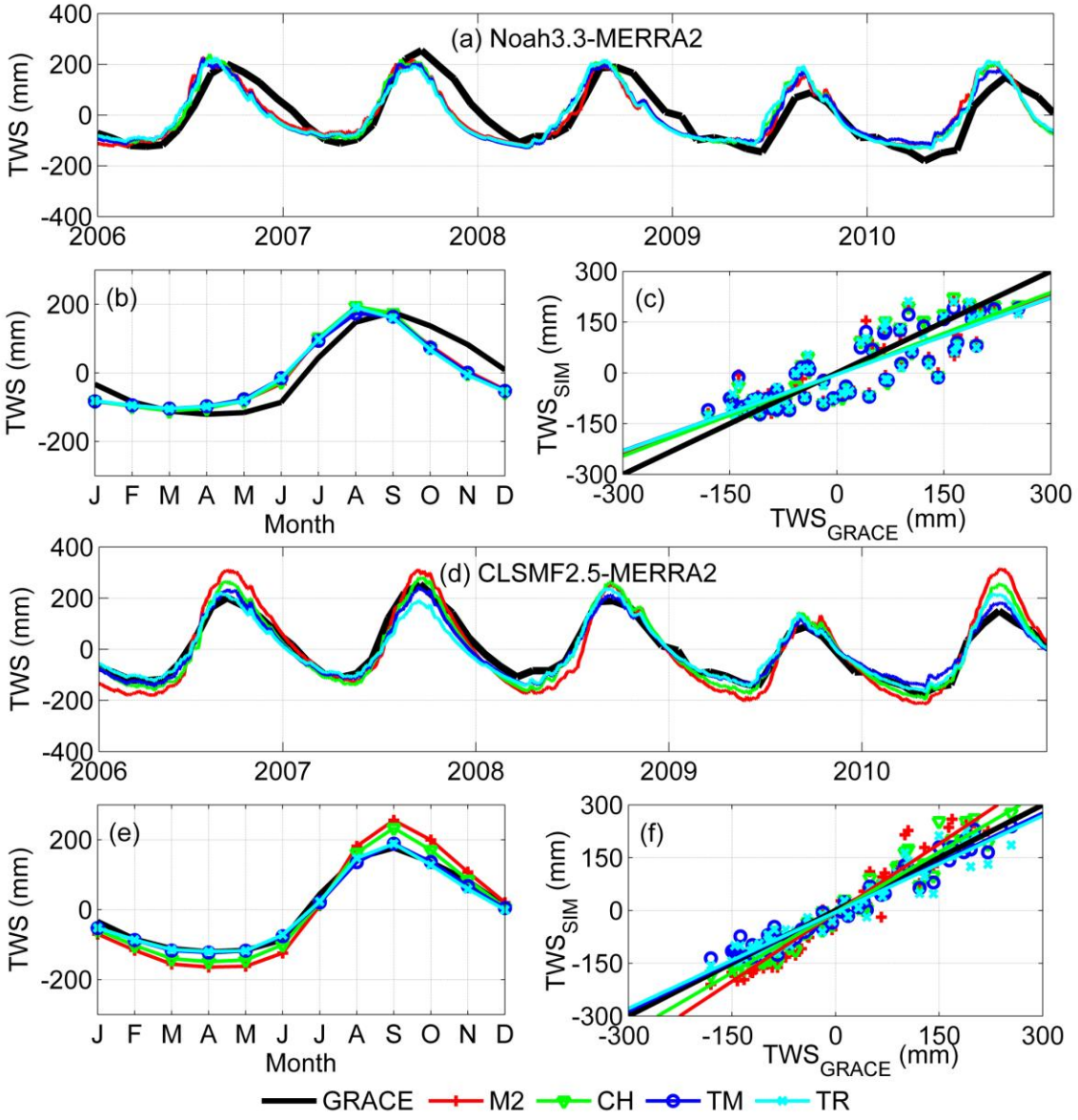


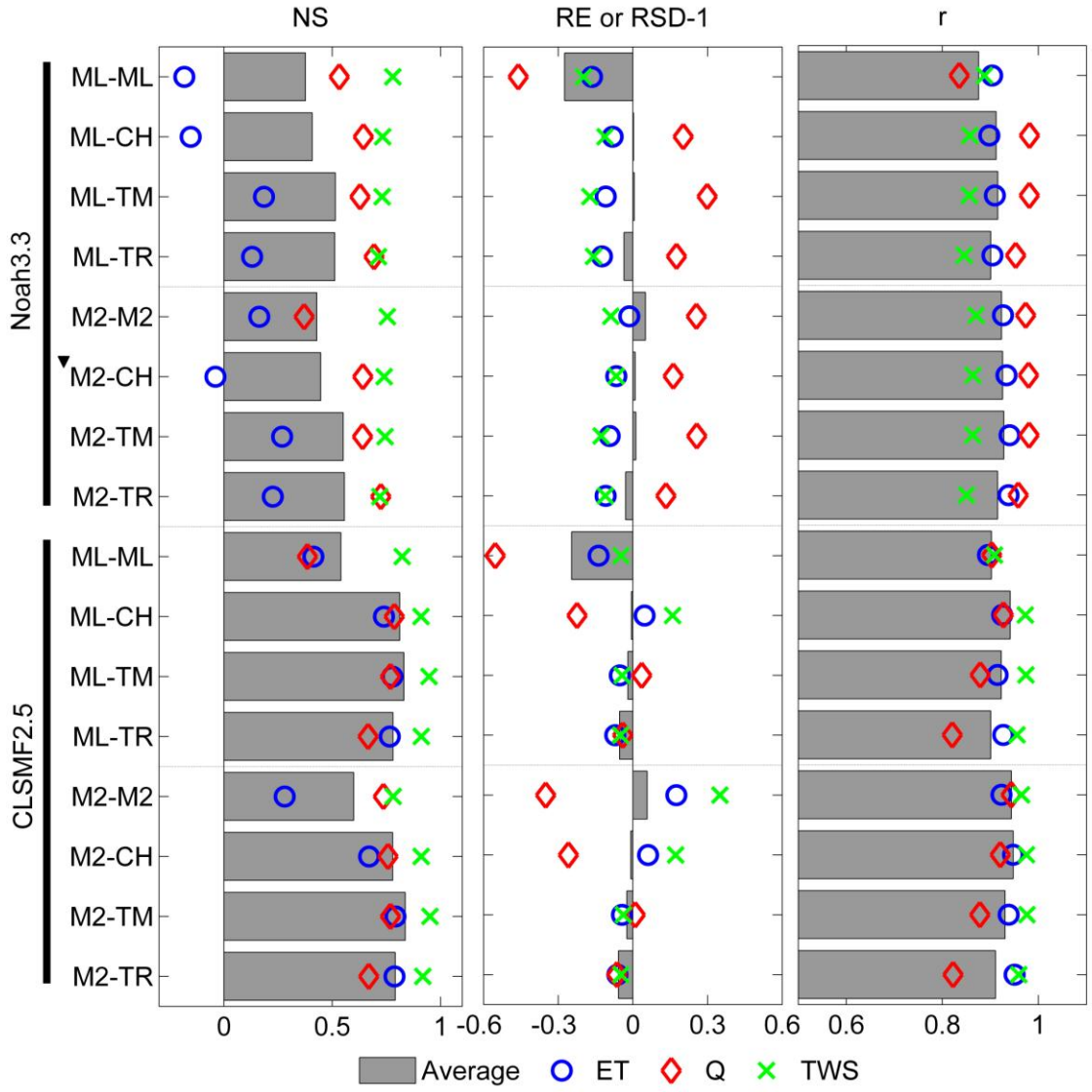







(a) Noah3.3-MERRA2
Q (m3/s)
15000
10000
5000
0
2006
2007
2008
2009
2010
(b)
Q (m3/s)
10000
5000
0
J F M A M J J A S O N D
Month
(c)
QSIM (m3/s)
10^4
10^3
10^2
10^2
10^3
10^4
QOBS (m3/s)
(d) CLSMF2.5-MERRA2
Q (m3/s)
15000
10000
5000
0
2006
2007
2008
2009
2010
(e)
Q (m3/s)
10000
5000
0
J F M A M J J A S O N D
Month
(f)
QSIM (m3/s)
10^4
10^3
10^2
10^2
10^3
10^4
QOBS (m3/s)
OBS
M2
CH
TM
TR


(a) Noah3.3-MERRA2
TWS (mm)
400
200
0
-200
-400
2006
2007
2008
2009
2010
(b)
TWS (mm)
200
0
-200
J F M A M J J A S O N D
Month
(c)
TWS_SIM (mm)
300
150
0
-150
-300
-300 -150 0 150 300
TWS_GRACE (mm)
(d) CLSMF2.5-MERRA2
(e)
(f)
GRACE
M2
CH
TM
TR


NS
RE or RSD-1
r
Noah3.3
CLSMF2.5
ML-ML
ML-CH
ML-TM
ML-TR
M2-M2
M2-CH
M2-TM
M2-TR
0
0.5
1
-0.6
-0.3
0
0.3
0.6
0.6
0.8
1
Average
ET
Q
TWS

## TABLES

**Table 1.** List of land surface models (LSMs), reanalysis-based meteorological forcing, and satellite-based precipitation data sets.

| Simulation (No.) | LSM | Base meteorological forcing data | Precipitation source |
|---|---|---|---|
| 1 | Noah3.3 (Chen et al., 1996; Ek et al., 2003) | MERRA-Land (ML; Reichle et al., 2011) | MERRA-Land (ML) |
| 2 | | | CHIRPS (CH; Funk et al., 2014) |
| 3 | | | TMPA3B42 (TM; Huffman et al., 2007) |
| 4 | | | TMPA3B42RT (TR; Huffman et al., 2007) |
| 5 | | MERRA2 (M2; Reichle et al., 2017) | MERRA2 (M2) |
| ▾6 | | | CHIRPS (CH) |
| 7 | | | TMPA3B42 (TM) |
| 8 | | | TMPA3B42RT (TR) |
| 9 | CLSMF2.5 (Koster et al., 2000; Reichle et al., 2011) | MERRA-Land (ML) | MERRA-Land (ML) |
| 10 | | | CHIRPS (CH) |
| 11 | | | TMPA3B42 (TM) |
| 12 | | | TMPA3B42RT (TR) |
| 13 | | MERRA2 (M2) | MERRA2 (M2) |
| 14 | | | CHIRPS (CH) |
| 15 | | | TMPA3B42 (TM) |
| 16 | | | TMPA3B42RT (TR) |

▾FLDAS operationally runs and provides model outputs from Noah3.3-MERRA2-CHIRPS.

**Table 2.** Summary of evaluation indices for monthly evapotranspiration (ET), streamflow (Q), and total water storage (TWS) anomaly from the 16 LSM simulations. The best statistical results in each column are in boldface.

| LSM | Met. | Precip. | ET | | | Q | | | TWS | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | NS | RE | r | NS | RE | r | NS | RSD | r | RMSE (mm) |
| Noah3.3 | ML | ML | -0.18 | -0.16 | 0.90 | 0.53 | -0.46 | 0.84 | 0.78 | 0.80 | 0.89 | 54 |
| | | CH | -0.15 | -0.08 | 0.90 | 0.64 | 0.20 | **0.98** | 0.73 | 0.89 | 0.86 | 60 |
| | | TM | 0.18 | -0.11 | 0.91 | 0.63 | 0.30 | **0.98** | 0.73 | 0.83 | 0.86 | 60 |
| | | TR | 0.13 | -0.12 | 0.90 | 0.69 | 0.18 | 0.95 | 0.71 | 0.84 | 0.85 | 62 |
| | M2 | M2 | 0.16 | **-0.01** | 0.93 | 0.37 | 0.26 | 0.97 | 0.75 | 0.91 | 0.87 | 57 |
| | | CH | -0.04 | -0.07 | 0.93 | 0.64 | 0.16 | **0.98** | 0.74 | 0.93 | 0.86 | 59 |
| | | TM | 0.27 | -0.09 | 0.94 | 0.64 | 0.26 | **0.98** | 0.74 | 0.87 | 0.86 | 59 |
| | | TR | 0.23 | -0.11 | 0.94 | 0.72 | 0.13 | 0.96 | 0.72 | 0.89 | 0.85 | 61 |
| CLSMF2.5 | ML | ML | 0.41 | -0.14 | 0.89 | 0.38 | -0.55 | 0.90 | 0.82 | 0.95 | 0.91 | 49 |
| | | CH | 0.74 | 0.05 | 0.92 | **0.79** | -0.22 | 0.93 | 0.91 | 1.16 | 0.97 | 35 |
| | | TM | 0.78 | -0.05 | 0.92 | 0.77 | 0.04 | 0.88 | **0.95** | **0.96** | 0.97 | 27 |
| | | TR | 0.76 | -0.07 | 0.93 | 0.67 | -0.04 | 0.82 | 0.91 | 0.95 | 0.96 | 35 |
| | M2 | M2 | 0.28 | 0.17 | 0.92 | 0.74 | -0.35 | 0.94 | 0.78 | 1.35 | 0.96 | 54 |
| | | CH | 0.6 | 0.0 | **0.9** | 0.7 | - | 0.9 | 0.9 | 1.17 | **0.98** | 35 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 7 | 6 | **5** | 6 | 0.26 | 2 | 1 | | | | |
| | | TM | **0.79** | -0.04 | 0.94 | 0.77 | **0.01** | 0.88 | **0.95** | **0.96** | **0.98** | **26** | |
| | | TR | **0.79** | -0.06 | **0.95** | 0.67 | -0.06 | 0.82 | 0.92 | 0.95 | 0.96 | 33 | |

## Highlights:

(1) To compare model outputs (i.e., evapotranspiration, streamflow, total water storage anomaly) from different combinations of land surface models, meteorological forcing datasets, and observation-based precipitation data in the upper Blue Nile basin

(2) To evaluate monthly time series of model water budget variables using in situ measurement and satellite-based products

(3) To identify the most suitable combination of LSMs and meteorological datasets for a water balance study

(4) To suggest that the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) incorporate CLSMF2.5 and HyMAP routing scheme to better represent the water balance in this region