Hydrology and
Earth System
Sciences

Improving SWAT model performance in the upper Blue Nile Basin using meteorological data integration and subcatchment discretization

Erwin Isaac Polanco[1], **Amr Fleifle**[1,2], **Ralf Ludwig**[3], **and Markus Disse**[1]

[1]Chair of Hydrology and River Basin Management, Faculty of Civil, Geo and Environmental Engineering,
Technische Universität München, Arcisstrasse 21, 80333 Munich, Germany
[2]Irrigation Engineering and Hydraulics Department, Faculty of Engineering, Alexandria University, El-Horia St.,
21544 Alexandria, Egypt
[3]Chair of Geography and Geographic Remote Sensing, Faculty of Geography, Ludwig-Maximilians-Universität München,
Luisenstrasse 37, 80333 Munich, Germany

Correspondence to: Erwin Isaac Polanco (erwinisaac.polanco@tum.de)

Abstract. The Blue Nile Basin is confronted by land degradation problems, insufficient agricultural production, and a limited number of developed energy sources. Hydrological models provide useful tools to better understand such complex systems and improve water resources and land management practices. In this study, SWAT was used to model the hydrological processes in the upper Blue Nile Basin. Comparisons between a Climate Forecast System Reanalysis (CFSR) and a conventional ground weather dataset were done under two sub-basin discretization levels (30 and 87 sub-basins) to create an integrated dataset to improve the spatial and temporal limitations of both datasets. A SWAT error index (SEI) was also proposed to compare the reliability of the models under different discretization levels and weather datasets. This index offers an assessment of the model quality based on precipitation and evapotranspiration. SEI demonstrates to be a reliable additional and useful method to measure the level of error of SWAT. The results showed the discrepancies of using different weather datasets with different sub-basin discretization levels. Datasets under 30 sub-basins achieved Nash–Sutcliffe coefficient (NS) values of -0.51, 0.74, and 0.84; p factors of 0.53, 0.66, and 0.70; and r factors of 1.11, 0.83, and 0.67 for the CFSR, ground, and integrated datasets, respectively. Meanwhile, models under 87 sub-basins achieved NS values of -1.54, 0.43, and 0.80; p factors of 0.36, 0.67, and 0.77; r factors

of 0.93, 0.68, and 0.54 for the CFSR, ground, and integrated datasets, respectively. Based on the obtained statistical results, the integrated dataset provides a better model of the upper Blue Nile Basin.

1 Introduction

Water resources in the upper Blue Nile Basin are not being managed adequately; land use changes, fast population growth, land erosion, and deforestation are some of the causes currently affecting the watershed. Therefore, in order to improve and provide better land use management practices and mitigate the alarming erosion problems researchers need to understand the hydrological conditions of the basin. Physically based, distributed models have provided a very efficient alternative for watershed researchers for analyzing the impact of land management practices on soil degradation, agriculture, water allocation, and chemical yields (Setegn et al., 2008). Due to its versatility and applicability to complex watersheds, researchers have identified the Soil and Water Assessment Tool (SWAT) as one of the most intricate, consistent, and computationally efficient models (Neitsch et al., 2009; Gassman et al., 2007). Recent studies prove that SWAT has received international and interdisciplinary acceptance for modeling large and small watersheds (Malunjkar et al.,

2015; Me et al., 2015; Rafiei Emam et al., 2016; Wang et al., 2017). SWAT provides a wide range of parameters to work with, allowing users to analyze several hydrological processes. It also has the advantage to have been developed to analyze the interaction of several hydrological parameters and the impact of land management practices specifically for large and complex basins; thus, it is a good model to be applied in the upper Blue Nile Basin. However, due to the lack of a unifying theory to accurately model the interaction of the hydrological processes, complex hydrological models suffer from overparameterization and high predictive uncertainty (Sivapalan, 2006). Therefore, it is difficult to simulate the complex interactions of hydrological processes and weather conditions of watersheds without uncertainties.

Among all the input parameters, the meteorological data have the most significant impact on the water balance of a watershed. However, a common problem to set up hydrological models of the upper Blue Nile Basin is related to data limitations. In developing countries, the distribution of meteorological stations is irregular and dispersed (Worqlul et al., 2014). Other weather data problems are related to measuring gauges; many weather data parameters contain missing data periods, and in several cases erroneous measurements are also possible. Thus, many models are often set up based on limited and incomplete data, which may lead to less reliable models. This lack of hydrological and climatic data has impeded in-depth studies of the hydrology of the upper Blue Nile Basin (Tekleab et al., 2011). Several previous studies have modeled the entirety and also small catchments of the Nile Basin, providing good and meaningful results (Tibebe and Bewket, 2011; Setegn et al., 2008, 2010; Swallow et al., 2009; Mulungu and Munishi, 2007). However, most of the hydrological models are built for the Lake Tana basin and its sub-basins: Gumara, Ribb, Gilgel Abay, and Koga (Chebud and Melesse, 2009; Setegn et al., 2008, 2010a, b; Wale, 2008). Dessie et al. (2015) and Kebede et al. (2006) performed a very detailed daily water balance analysis and annual water budget for the Lake Tana basin where the runoff and outflows of ungauged catchments were estimated. Uhlenbrook et al. (2010) performed an analysis of the hydrological processes and responses of Gilgel Abay and Koga catchments by applying the HBV model. Other studies have modeled the entire upper Blue Nile Basin; for instance, Abera et al. (2017) performed a water budget analysis in the upper Blue Nile Basin where precipitation, outflow, and evapotranspiration analyses were done. Betrie et al. (2011) and Easton et al. (2010) also modeled and calibrated the upper Blue Nile Basin using discharge data to estimate sediment yield and erodible areas of the basin; values of the calibrated parameters for flow and sediment were also shown. Dessie et al. (2014) also performed a runoff and sediment yield analysis in the upper Blue Nile Basin, although the main analysis was done at the Lake Tana region. Tekleab et al. (2011) also modeled the upper Blue Nile Basin, where an interesting water balance analysis was done and monthly

streamflow for several subcatchments was modeled. However, most of the studies at large scale in the upper Blue Nile Basin do not provide detailed values for the each of the water balance components of the basin. Another important issue when setting up SWAT models concerns the right number of sub-basins, because the number of meteorological stations to be used by SWAT will depend on the number of sub-basins. For instance, if two stations are located within one sub-basin, SWAT will choose the station nearest to the center of the sub-basin; the other station will be disregarded. However, if more sub-basins are created in a model, and these two stations lie in different sub-basins, then both stations will be considered by SWAT, which provides different water balance results.

Therefore, the first objective of this study has been the comparison of different weather datasets at large scale and under different sub-basin discretization levels. Two models were created using different subcatchment discretization (30 and 87 sub-basins), hereafter named SWAT30 and SWAT87, respectively (Fig. 3). The time frame of the models was from 1990 to 2004, using a 4-year warm-up period (1990–1993), a 6-year calibration period (1994–1999), and a 5-year validation period (2000–2004). This comparison provided a better understanding of the effects of different sub-basin discretization levels on the total water balance of a watershed. It also helped to identify the temporal and spatial constraints of both datasets. Roth and Lemann (2016) performed a comparison between CFSR and conventional data in small catchments in the Ethiopian highlands, where they showed that the CFSR data provided unreliable results. However, Roth and Lemann (2016) made it clear that the CFSR data were tested only in very small catchments ranging from 112 to 477 ha and not at large scale, also suggesting that CFSR data should be carefully checked and compared with conventionally measured data of similar climatic stations. Furthermore, this study proposes an integration of CFSR and conventional weather data to be used at large scale in the upper Blue Nile Basin with an area of approximately 199 812 km^2. Additionally, the CFSR stations used in the study were compared with conventionally measured data. Based on the obtained statistical results, the integration of these two datasets provides better models and a better representation of the magnitudes and distribution of the different weather variables in the upper Blue Nile Basin.

After a hydrological model has been set up, a critical point to determine its quality is the water balance. Therefore, in addition to graphical assessments, other statistical indicators as the Nash–Sutcliffe coefficient (NS), percent bias (PBIAS), and ratio of the root mean square error (RMSE) to the standard deviation of measured data were proposed by Moriasi et al. (2007). Based on these commonly used statistical indicators, most of the SWAT models provide very good results for discharge values at the outlet of a basin (van Griensven et al., 2012). However, the evaluation of the models based on both evapotranspiration and water balance is not discussed in detail, and the evapotranspiration behavior of a catchment
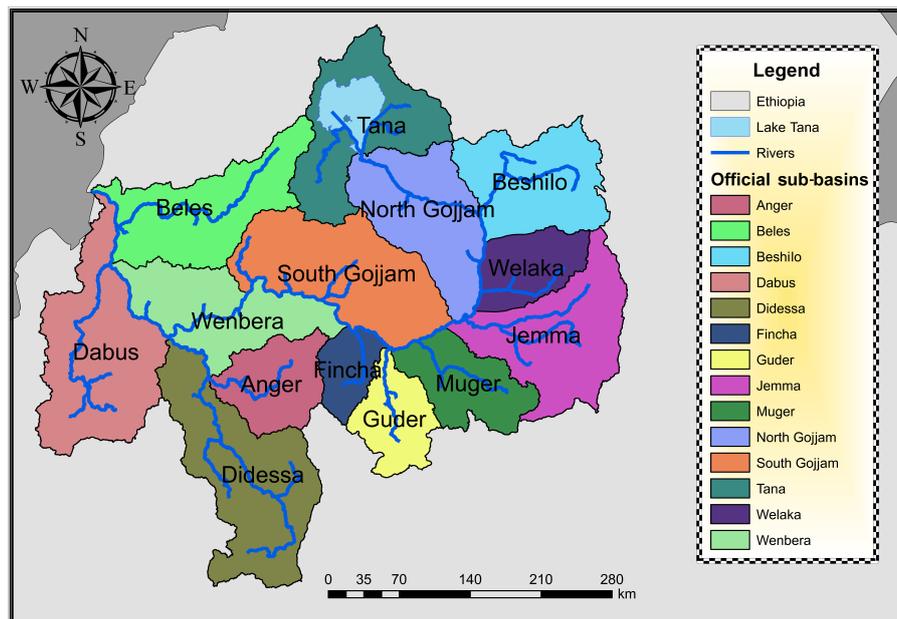
Figure 1. Official sub-basin distribution of the upper Blue Nile Basin.

is usually not presented. Several published documents could even report unrealistic parameter values (van Griensven et al., 2012). Therefore, the second objective of this study has been to propose an index, the SWAT error index (SEI), to quantify the level of error of a hydrological model. The SEI uses flexible weighting values for the relative root mean square error (rRMSE) obtained from measured flow discharge data and satellite evapotranspiration data. SEI proved to be a useful additional method to develop models that can provide a better representation of the water balance of a watershed.

2 Materials and methods

2.1 Study site

The upper Blue Nile Basin, also known as Abay Basin, is located in the northwestern highlands of Ethiopia, approximately between latitude 7°40′ and 12°51′ N, and longitude 34°25′ and 39°49′ E, with elevations raging between 483 and 4248 m a.s.l. The total area of the upper Blue Nile Basin is approximately 199 812 km^2, including two sub-basins shared with the Republic of the Sudan in the northern region. The climate in the upper Blue Nile Basin fluctuates from humid to semi-arid and it is mainly dominated by latitude and altitude, with average temperatures ranging from 13 °C in the southeastern to 26 °C in the southwestern regions. The lowest rainfall data detected during the current research period (1990–2004) correspond to the eastern region for the sub-basins of Beshilo, North Gojjam, South Gojjam, Welaka, Jemma, Muger, Guder, and Fincha, where the precipitation drops below 1000 mm yr^{-1} (Figs. 1 and 4); meanwhile, the highest precipitation ranges belong to the western region (Didessa, Wenbera, Anger, Dabus, and Beles), with precipitation above 1900 mm yr^{-1} (Figs. 1 and 4). The topographic disparity and variations in altitude of the upper Blue Nile Basin have a great impact on the weather, soil, and vegetation conditions. Consequently, rainy seasons are very variable in this watershed; for instance, the total discharge peaks at the El Diem gauging station can reach 7000 m^3 s^{-1} and dry seasons can go as low as 100 m^3 s^{-1} (Figs. 7 and 8). Soils in the upper Blue Nile Basin are mainly dominated by 10 types (Fig. 2): Eutric Nitosols, Eutric Cambisols, Humic Fluvisols, Cambic Arenosols, Chromic Vertisols, Dystric Cambisols, Eutric Fluvisols, Eutric Regosols, Orthic Acrisols, and Pellic Vertisols (FAO, 2015).

2.2 Datasets

A Shuttle Radar Topographic Mission digital elevation model (SRTM DEM) from the Consultative Group on International Agricultural Research – Consortium for Spatial Information (CGIAR-CSI) was used to set up the model. This DEM has a resolution of 90 m and was used to perform an automatic watershed delineation of the upper Blue Nile Basin, where the flow direction, flow accumulation, and stream network were automatically determined by SWAT.

The second input dataset was a land use map, which was obtained from the GIS portal of the International Livestock Research Institute (ILRI) and corresponds to the year 2004 (http://data.ilri.org/geoportal/catalog/main/home.page).

The soil map used for these models was developed by the Food and Agriculture Organization of the United
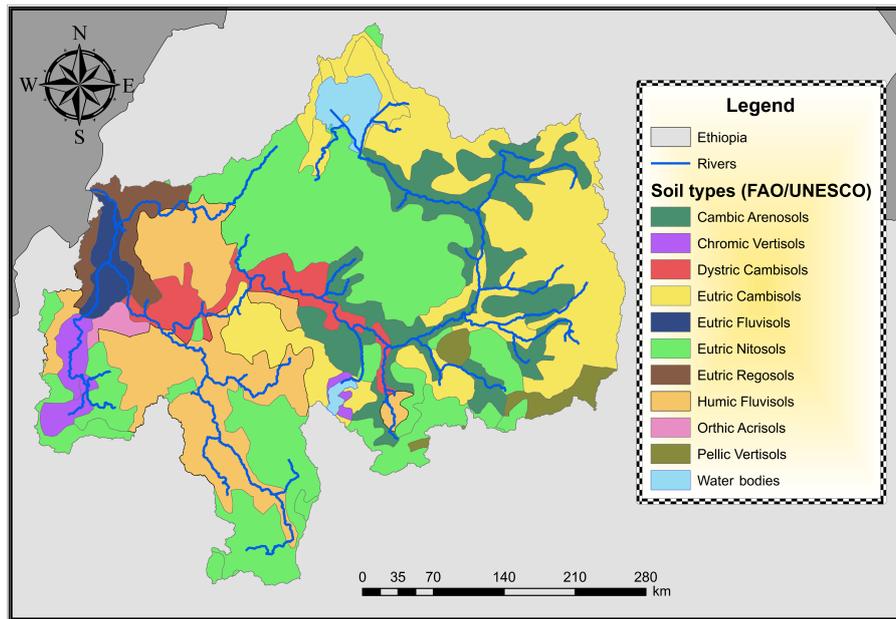
Figure 2. FAO/UNESCO soil map of the upper Blue Nile Basin.
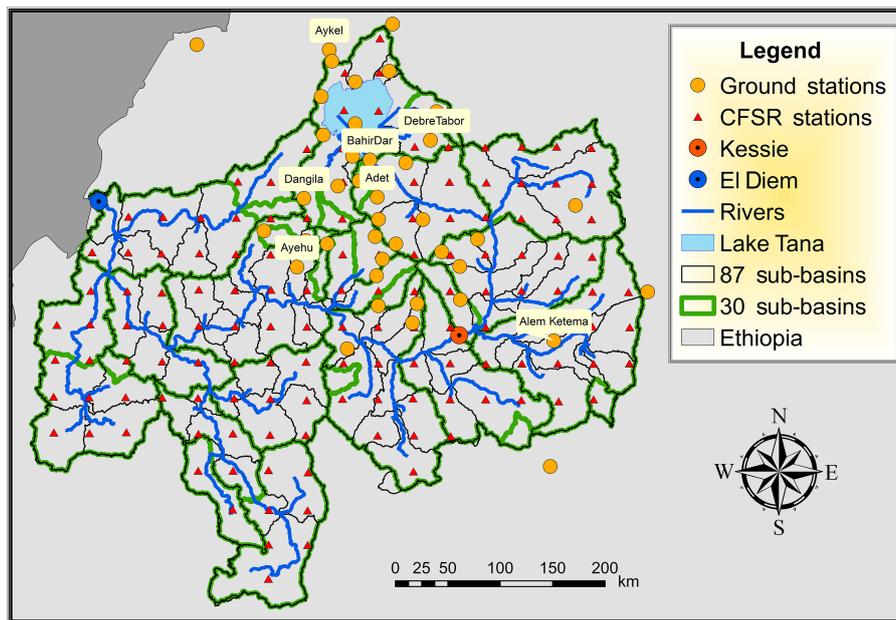
Figure 3. Weather and hydrometric gauging stations in the upper Blue Nile Basin under two discretization levels of 30 and 87 sub-basins (SWAT30 and SWAT87).

Nations (FAO-UNESCO). This world soil map was prepared by FAO and UNESCO at 1 : 5 000 000 scale (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/). The information provided by this map was used in combination with the Harmonized World Soil Database v1.2, a database that combines existing regional and national soil information (http://www.

fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/).

The last input dataset was the meteorological information. Two weather datasets from different sources were used to set up the models. The first weather dataset was collected from the National Meteorology Agency of Ethiopia (NMA). The data used for these models correspond to 42 stations distributed in the upper Blue Nile Basin (Fig. 3). However, only
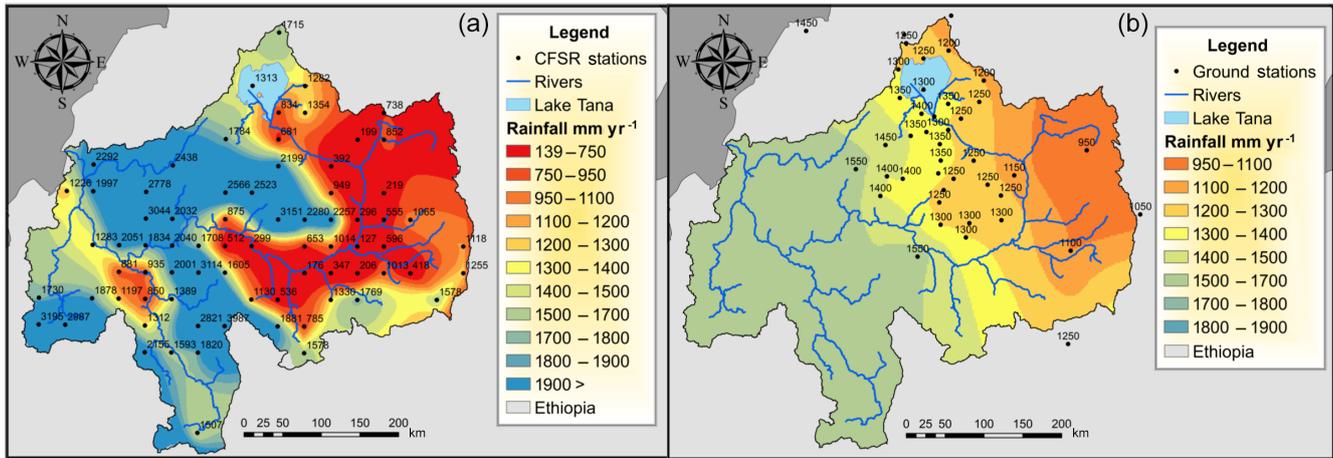
Figure 4. Spatial annual rainfall variation in the upper Blue Nile Basin using two different data sources: the CFSR dataset **(a)** and the ground dataset **(b)**.

15 of these stations are capable of measuring all five parameters needed to set up SWAT: rainfall, temperature, relative humidity, solar radiation, and wind speed. Moreover, few of these 15 station have complete and continuous data available for the entire period under study (1990–2004). For instance, the collected data for solar radiation were limited to 2 stations only, wind speed was available for 4 stations, only maximum temperature was available for 4 stations, relative humidity was available for 3 stations, and precipitation was available for all 42 stations. Additionally, the quality of these observed data is somehow questionable. Many meteorological stations are more than 10 years old, and their constant technical failure due to the lack of continuous expert maintenance also questions the quality of the data. A large part of the available ground data has been collected from old stations that could have in many cases malfunctioning, defected, and outdated devices. The second weather dataset was the Climate Forecast System Reanalysis (Fig. 3), a dataset that has been produced by the National Centers for Environmental Prediction (NCEP) (http://globalweather.tamu.edu/). CFSR data bring several uncertainties due to their multiple spatial and temporal interpolations (Dile and Srinivasan, 2014). These data were generated using different assimilation techniques that include satellite radiances, advanced coupled atmospheric, oceanic, and land surface modeling components. The global atmospheric resolution of CFSR data is approximately 38 km. These atmospheric, oceanic, and land surface output products are available at a $0.5° \times 0.5°$ latitude and longitude resolution. Both weather datasets used for these models correspond to the period 1990–2004.

For the analysis of the quality of the SWAT models, monthly flow discharge data and evapotranspiration data were used. The flow discharge data were obtained from the Ministry of Water, Irrigation and Electricity of Ethiopia and correspond to the gauging stations at Kessie and El Diem at the main stream of the upper Blue Nile Basin (Fig. 3). For the

evapotranspiration analysis, data from the MOD16 Global Terrestrial Evapotranspiration Project (http://www.ntsg.umt.edu/project/mod16) were obtained. The global evapotranspiration data from MOD16 are regular $1\,km^2$ land surface datasets for the 109.03 million km^2 of vegetated area in the whole world at different time intervals (8 days, monthly, and annual) from which monthly data generated specifically for the Nile Basin were used.

2.3 Water balance and evapotranspiration processes in SWAT

Water balance in watersheds is one of the most important factors used to determine if a model is good enough for any particular application. Hence, analyses of the processes involved in the estimation of the water balance of a watershed (evapotranspiration, runoff, and groundwater) can provide more details about the hydrological behavior of a watershed and can be used to understand the interaction of main hydrological processes (Zhang et al., 1999). For the input data processing and hydrological estimation, SWAT uses two levels of discretization: sub-basins and hydrologic response units (HRUs). HRUs are contained in the sub-basins and are defined based on the land use map, soil map, and slope classes. HRUs allow the model to reflect differences in evapotranspiration and other hydrologic conditions for each crop and soil type. The water balance in SWAT is calculated for each HRU using the following formula (Neitsch et al., 2009):

$$\mathrm{SW}_t = \mathrm{SW}_0 + \sum_{i=1}^{t} \left(R_{\mathrm{day}} - Q_{\mathrm{surf}} - E_{\mathrm{a}} - W_{\mathrm{seep}} - Q_{\mathrm{gw}} \right), \quad (1)$$

where SW_t is the final soil water content (mm), SW_0 is the initial soil water content on day i (mm), R_{day} is the amount of rainfall on day i (mm), Q_{surf} is the amount of surface runoff on day i (mm), E_{a} is the amount of evapotranspiration on

day i (mm), W_{seep} is the amount of water entering the vadose zone from the soil profile on day i (mm), and Q_{gw} is the amount of return flow on day i (mm).

SWAT can estimate the evapotranspiration using several methods, from which the Hargreaves and Penman–Monteith methods were compared in this study (Figs. 11 and 12). The Hargreaves method calculates the potential evapotranspiration using minimum and maximum daily temperatures as input data (Hargreaves and Samani, 1982). This method was chosen as a better option for the upper Blue Nile Basin due to the data scarcity of the meteorological stations in the basin. The Hargreaves equation can be used with the sole input of temperature data, while the Penman–Monteith requires more data, for instance, wind speed, solar radiation, and relative humidity. The Hargreaves method has been recommended for computing potential evaporation in cases when only the maximum and minimum temperatures are available (Allen et al., 1998). A study from Tekleab et al. (2011) was also able to successfully use the Hargreaves equation to calculate the potential evaporation in the upper Blue Nile Basin. Several improvements were made to the original equation since 1975 (Hargreaves and Samani, 1982). The final form of the Hargreaves equation used in SWAT and published in 1985 (Hargreaves and Samani, 1985) is as follows (Neitsch et al., 2009):

$$\lambda E_0 = 0.0023 \cdot H_0 \cdot (T_{max} - T_{min})^{0.5} \cdot \left(\overline{T}_{av} + 17.8\right), \qquad (2)$$

where λ is the latent heat of vaporization (MJ kg^{-1}), E_0 is the potential evapotranspiration (mm day^{-1}), H_0 is the extraterrestrial radiation (MJ m^{-2} day^{-1}), T_{max} and T_{min} are the maximum and minimum air temperature for a given day (°C), respectively, and T_{av} is the mean air temperature for a given day.

Following the potential evapotranspiration, the actual evapotranspiration must be calculated. Initially, SWAT calculates the evaporated water intercepted by the canopy; then, maximum transpiration and soil evaporation are calculated. Evaporation from canopy is very significant in forested areas and in several cases can be higher than transpiration. Transpiration for the Hargreaves equation is calculated as (Neitsch et al., 2009)

$$E_t = \frac{E_0' \cdot \text{LAI}}{3.0}, \qquad (3)$$

where E_t is the maximum transpiration on a given day (mm H$_2$O), E_0' is the potential evapotranspiration adjusted for evaporation of free water in the canopy (mm H$_2$O), and LAI is the leaf area index.

Evaporation from the soil on a given day is calculated with following equation (Neitsch et al., 2009):

$$E_s = E_0' \cdot \text{cov}_{sol}, \qquad (4)$$

where E_s is the maximum soil evaporation on a given day (mm H$_2$O), E_0' is the potential evapotranspiration adjusted for evaporation of free water in the canopy (mm H$_2$O), and cov$_{sol}$ is the soil cover index.

2.4 Weather data processing and integration

If input data are used without the respective analyses, models will provide less reliable results. Also, even small errors in temperature or precipitation can result in considerable inaccuracies and impacts on the model results (Maraun et al., 2010). Tekleab et al. (2011) and Uhlenbrook et al. (2010) checked the data quality of streamflow data in the upper Blue Nile Basin based on comparison graphs and additionally a double mass analysis. In this study, the data quality and consistency of the time series on monthly basis in terms of magnitude and spatial distribution of the five input variables required by SWAT were also analyzed through comparison graphs (Figs. 4–6) to determine the deficiencies of the two datasets (CFSR and ground datasets) and to form an integrated dataset.

In the first case, the ground dataset was used without alterations to create the SWAT models. This ground dataset obtained from the NMA corresponds to 42 stations in the upper Blue Nile Basin, where most of the meteorological stations were located in the eastern part of the watershed (Fig. 3). Additionally, the data obtained from these stations had several months of missing data, leading to temporal uncertainties.

For the second case, the SWAT models were set up using the CFSR dataset, also without alterations. This dataset is evenly distributed at 38 km resolution, with over 100 stations available for the upper Blue Nile Basin, and is temporally continuous.

However, after performing a quality check through a comparison of maps and graphs between the ground and CFSR datasets (Figs. 4–6), it was noticed that not all the weather variables from CFSR are reliable. The precipitation distribution appeared to be underestimated in the eastern region of the upper Blue Nile Basin and overestimated in the western region (Fig. 4). The map created from the ground stations (Fig. 4b) showed a precipitation distribution in the western region that was the result of SWAT using the precipitation values from the nearest stations. Two stations in the eastern part, Alem Ketema and Adet (Figs. 5a, b and 6a, b), showed the underestimation of the CFSR rainfall at the eastern region, and Ayehu (Figs. 5c and 6c) showed the overestimation of the CFSR rainfall in the western region. For this reason, additional CFSR rainfall stations were not used in the integrated dataset. However, the graphical and statistical comparisons of the few available stations for relative humidity, temperature, and solar radiation showed an acceptable level of agreement between the ground and CFSR datasets. The seasonal behavior and magnitudes of the values for these variables are similar; additionally, the 1 : 1 graphs showed an acceptable degree of matching (Fig. 6). For instance, the values for relative humidity for Debre Tabor and Aykel with both datasets show very similar values (Figs. 5d, e and 6d, e). The comparisons of maximum temperature for Aykel also showed a good degree of matching (Figs. 5g and 6g), although for Bahir Dar the results were not very good, showing
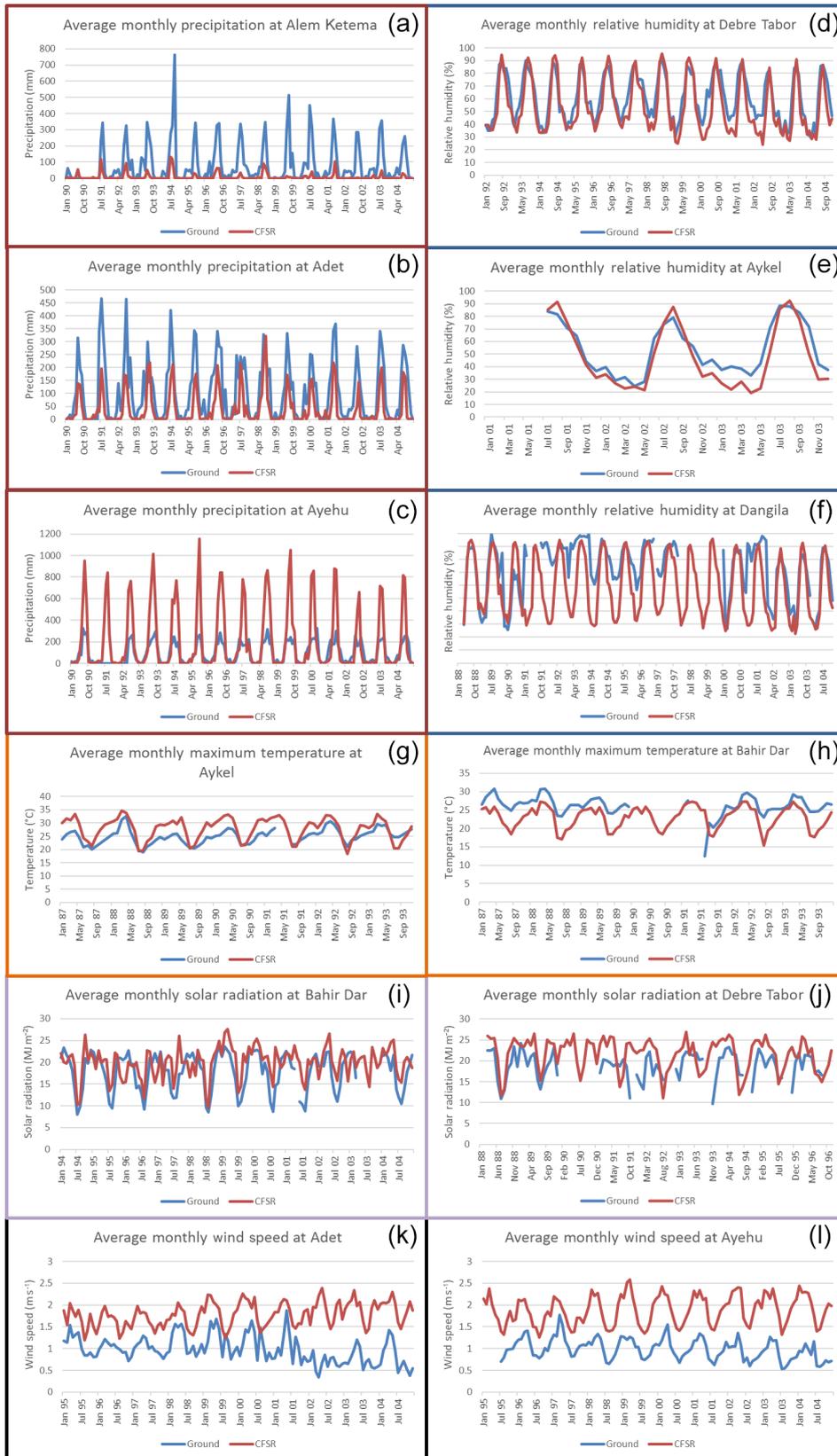
Figure 5. Comparisons between the ground and CFSR weather datasets. Panels **(a–c)** are average monthly precipitation; **(d–f)** are average monthly relative humidity; **(g, h)** are average monthly maximum temperatures; **(i, j)** are average monthly solar radiation; and **(k, l)** are average monthly wind speed.
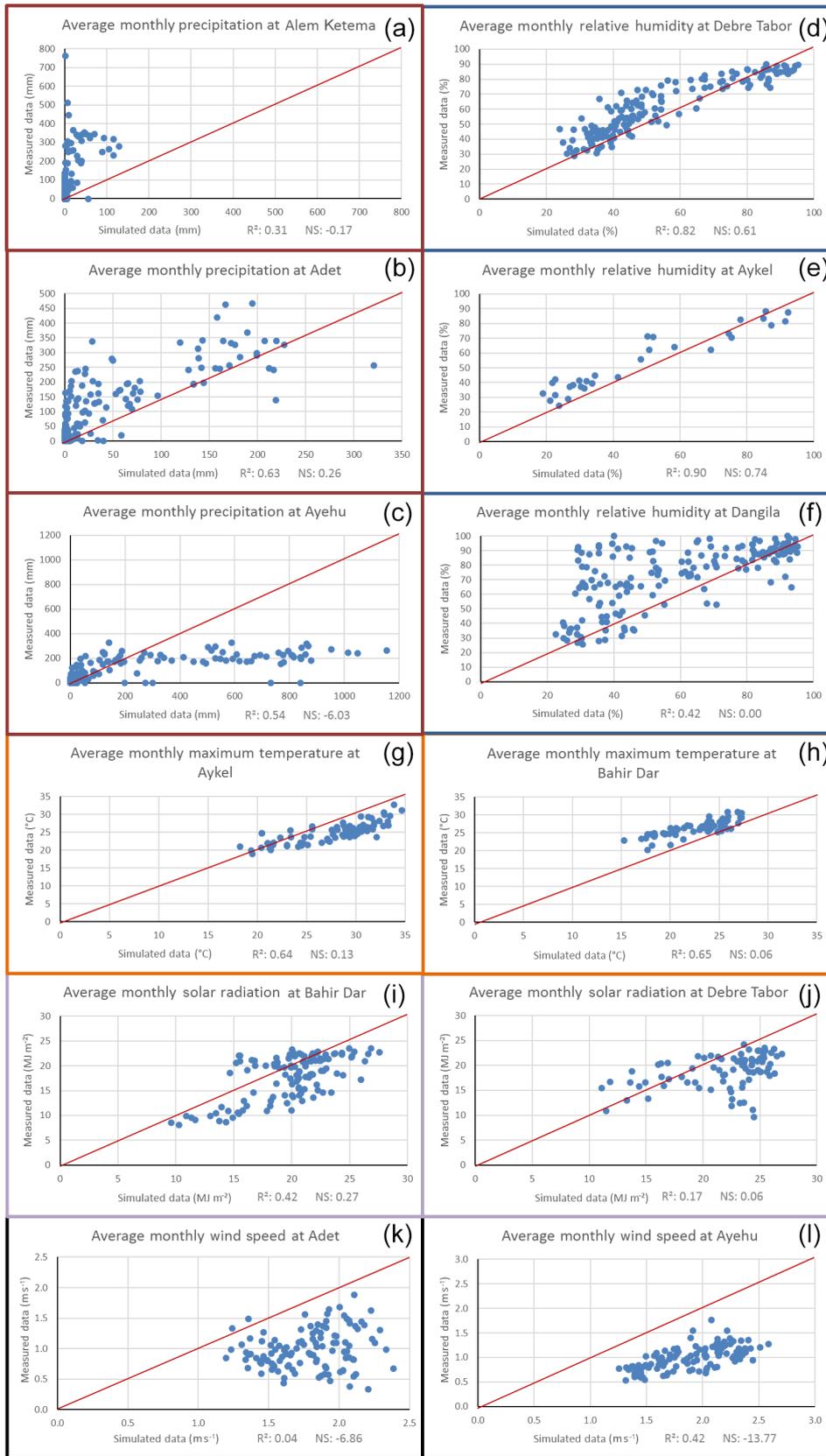
Figure 6. Significance of matching between the ground and CFSR weather datasets. Panels **(a–c)** are average monthly precipitation; **(d–f)** are average monthly relative humidity; **(g, h)** are average monthly maximum temperatures; **(i, j)** are average monthly solar radiation; and **(k, l)** are average monthly wind speed.

a slight underestimation (Figs. 5h and 6h). The solar radiation comparison at Bahir Dar (Figs. 5i and 6i) also showed a good agreement between both datasets, although results at Debre Tabor (Figs. 5j and 6j) showed slightly different results. The exception was the wind speed data, which in both cases at Adet and Ayehu (Figs. 5k, l and 6k, l) were overestimated by the CFSR dataset.

Therefore, these two datasets were integrated to form a third input dataset for SWAT with the objective of overcoming their spatial and temporal limitations. Tekleab et al. (2011) and Uhlenbrook et al. (2010) filled in missing streamflow data of the upper Blue Nile Basin using regression analysis, which is also a good approach to fill in missing meteorological values. However, in this study, the missing values of the ground dataset refer to complete time series of a specific station and variable. Thus, to create the integrated dataset, the 42 rainfall stations of the ground dataset were taken as the basis; this means that the locations of the weather stations of the final integrated dataset correspond to the locations of the 42 rainfall stations of the ground dataset. From there, the missing variables (relative humidity, temperature, and solar radiation values) of those 42 rainfall stations were completed by using the variables of their nearest CFSR stations. The integrated dataset has 42 stations where the data for each variable were combined as follows: the precipitation is formed by 42 rainfall stations taken entirely from the ground dataset; the relative humidity is formed by 3 stations from the ground dataset and 39 stations from the CFSR dataset; the maximum temperature is formed by 4 stations from the ground dataset and 38 stations from the CFSR dataset; the values for the minimum temperature were taken totally from the CFSR dataset; the solar radiation was formed by 2 stations from the ground dataset and 40 stations from the CFSR dataset; no wind speed data were used in the models. However, missing daily values within a variable were completed by the built-in SWAT weather generator. This integrated dataset contained more data than the ground dataset and also provided more reliable precipitation values and distribution than those provided by the CFSR dataset.

2.5 Parameterization for the calibration and validation of the models

One of the major constrains of hydrological modeling is the difficulty of the parameterization of different variables (Hauhs and Lange, 2008). The correct combination of the values of the parameters influencing the ground water, runoff, and evapotranspiration processes is a key point in a model calibration. The characterization of watersheds considering their most influential variables is a good approach to determine the predictive capabilities of a model (McDonnell et al., 2007). Initially, it is recommended to perform calibrations for annual discharge values once acceptable results are acquired; a calibration based on monthly values can be performed to achieve more detailed results (Neitsch et al., 2009).

During a model calibration, a potential value can be assigned for each parameter and for each HRU, which would generate a large number of parameters. However, these values can also be applied as a global modification to estimate parameters by multiplying or adding values. Table 2 shows the parameterization applied to the respective regions in the watershed to calibrate streamflow at Kessie and El Diem, where r stands for relative values and v for values to be replaced. The same parameterization was applied to all the models with different subcatchment delineations and data sources. Land coverage, soil types, and slope have a great impact on the total water balance, and a calibration with wrong parameters values will only produce models with good statistical results but with less realistic representation of the actual properties of the watershed. Therefore, the values of the parameters were modified within the ranges specified by the SWAT input/output documentation 2012 (Arnold et al., 2012). For instance, the available water content of the soils was calibrated in such a way that it did not change the physical properties of the soils. The curve number 2 (CN2) values were defined within different ranges based on the type of land cover.

2.6 Statistical indices and SWAT quality analyses

2.6.1 Calibration and validation with flow discharge

In the case of hydrological modeling, the limitation of the data quality and capabilities of the model to represent the complexity of the hydrological process often constitutes obstacles. Therefore, models must be calibrated, and a statistical analysis is also required to determine how reliable the results of the model are prior to their applications (Bastidas et al., 2002). Since sediment data for the upper Blue Nile Basin are very limited, the calibration and validation of the models were done using flow discharge data only. The calibrated stations were Kessie and El Diem at the mainstream of the Blue Nile River (Fig. 3). For the automatic calibration, the Sequential Uncertainty Fitting version 2 (SUFI-2) was used to efficiently calculate the coefficient of determination (R^2) and NS as likelihood measures, trying to catch the seasonal dynamics and magnitudes of the measured discharge data.

SUFI-2 is a sequential parameter estimation method that operates within parameter uncertainty domains (Tanveer et al., 2016). SUFI-2 performs several iterations, where each iteration provides better results than the previous iteration and reduces the parameters ranges. In SUFI-2, the objective is to capture most of the observed values within the 95PPU (95 % prediction uncertainty) range at the same time that thinner 95PPU range is preferable. The 95PPU represents the uncertainty in the model outputs. Therefore, the simulation starts assuming large and physically meaningful parameter ranges, so that the measured data fall within the 95PPU and continuously decrease the ranges of the 95PPU and produce better results. The final 95PPU is the 95 % of the observed data captured within the final 95PPU band, which are defined by

the final parameter intervals. Therefore, the best simulation is the best iteration within the 95PPU, and considering that is difficult to claim a specific parameter range for a certain watershed, any solution within the 95PPU should be an acceptable solution. The fit of simulated results within the 95PPU is quantified through the p factor and r factor. The p factor is the percentage of observed data that fall within the 95PPU and ranges from 0 to 1, while the r factor is the thickness of the 95PPU band and ranges from 0 to infinity. The quality of a calibration and the prediction uncertainty are judged based on how close the p factor is to 1 and how close the r factor is to 0 (Yang et al., 2007). A p factor of 1 and r factor of 0 represent the measured data. As the number of iterations increases, SUFI-2 continues to reduce the 95PPU thickness and produces smaller values for the p factor and r factor, trying to find a better combination of the parameter values. The uncertainty in SUFI-2 is expressed as a uniform distribution of parameter ranges, and parameter uncertainties are considered for any possible source in variables, for instance, model inputs, model structure, model parameters, and also measured data (Abbaspour et al., 2015). The uncertainties in the outputs are expressed as the 95PPU. The uncertainty analysis in SUFI-2 is based on the concept that a single parameter value generates a single model response, while a parameter range or propagation of the parameter uncertainty leads to the 95PPU.

The coefficient of determination (R^2) is a measure of how well the regression line represents the data and gives a measure of the proportion of the fluctuation of a variable that is predictable from another variable. The values for this coefficient denote the strength of the linear relation between Q_m and Q_s, representing the percentage of the data closest to the line of best fit. The R^2 objective function provided in SWAT-CUP is as follows:

$$R^2 = \frac{\left[\sum\limits_{i=1}^{n}\left(Q_{m,i} - \overline{Q}_m\right)\left(Q_{s,i} - \overline{Q}_s\right)\right]^2}{\sum\limits_{i=1}^{n}\left(Q_{m,i} - \overline{Q}_m\right)^2 \sum\limits_{i=1}^{n}\left(Q_{s,i} - \overline{Q}_s\right)^2}, \tag{5}$$

where Q indicates discharge values, "m" and "s" stand for observed and simulated values, respectively, and i is the ith measured or simulated data.

NS is widely used as goodness-of-fit indicator that expresses the potential predictive ability of a hydrological model (Nash and Sutcliffe, 1970). The Nash–Sutcliffe objective function provided in SWAT-CUP is as follows:

$$NS = 1 - \frac{\sum\limits_{i=1}^{n}\left(Q_m - Q_s\right)_i^2}{\sum\limits_{i=1}^{n}\left(Q_{m,i} - \overline{Q}_m\right)^2}, \tag{6}$$

where Q indicates discharge values, "m" and "s" stand for observed and simulated data, respectively, and the bar stands for the average values.

2.6.2 Actual evapotranspiration analysis

In addition to the calibration and validation of the SWAT models with flow discharge, comparisons with evapotranspiration data could also provide more details to quantify the reliability of hydrological models. Therefore, actual evapotranspiration data for the upper Blue Nile Basin were obtained from the MODIS Global Terrestrial Evapotranspiration Project (MOD16). These are global estimated data from land surface by using satellite remote sensing data. These data are intended to be used to calculate regional water balances; hence, they are a very important source of data for watershed management and hydrological models analyses. The original MOD16 evapotranspiration (ET) algorithm (Mu et al., 2007) was based on the Penman–Monteith equation (Monteith, 1965), while the current MOD16 ET has used the improved evapotranspiration algorithm (Mu et al., 2011). In this improved algorithm, the sum of the evaporation from the wet canopy surface, transpiration from the dry canopy surface, and evapotranspiration from the soil surface constitute the total daily ET (Mu et al., 2011). The formulae for the total daily ET(λE) and potential ET (λE_{POT}) are

$$\lambda E = \lambda E_{wet\_C} + \lambda E_{trans} + \lambda E_{SOIL}$$

$$\lambda E_{POT} = \lambda E_{wet\_C} + \lambda E_{POT\_trans} + \lambda E_{wet\_SOIL} + \lambda E_{SOIL_{POT}}, \tag{7}$$

where λE_{wet_C} is the evaporation from the wet canopy surface, λE_{trans} is the transpiration from the dry canopy surface (plant transpiration), λE_{SOIL} is the evaporation from the soil surface, λE_{POT_trans} is the potential plant transpiration, and $\lambda E_{SOIL_{POT}}$ is the potential soil evapotranspiration.

Previous studies have already shown that the annual ET values derived from the MOD16 algorithm are lower than those provided by hydrological models, principally when using the Hargreaves method. For instance, Ruhoff et al. (2013) detected an underestimation of 21 % in the evapotranspiration provided by MOD16 in the Rio Grande Basin, Brazil, where the underestimation was mainly caused by the misclassification of the land use. Sun et al. (2007) also identified certain disadvantages in the MOD16 evapotranspiration. Nevertheless, in this study, the evapotranspiration estimations from SWAT were compared with satellite evapotranspiration data. This was done only as comparison and not with the objective of calibrating the models, and also as a test to understand the performance of the proposed SEI.

Evapotranspiration estimations shown as percentage of the average annual precipitation are frequently given for the upper Blue Nile Basin. However, these percentages would yield totally different amounts depending on the average annual precipitation provided by different weather data sources and under different sub-basin discretization. Therefore, a comparison of the actual evapotranspiration data provided by MOD16 with the values calculated by SWAT under the Hargreaves and Penman–Monteith equations was done to show the level of discrepancy between datasets (Figs. 11,

Table 1. Average annual water balance components in the upper Blue Nile Basin based on different literature.

Cherie (2013)		
Hydrologic parameters	Calibration period 1976–1982 (mm yr^{-1})	Validation period 1992–1995 (mm yr^{-1})
Precipitation	1338	1348
Evapotranspiration	962	960
Revap/shallow aquifer	59	58
Surface runoff	143	151
Return flow	70	38
Transmission losses	9	9
Mengistu and Sorteberg (2012)		
Hydrologic parameters	Calibration period 1991–1996 (mm yr^{-1})	Validation period 1997–2000 (mm yr^{-1})
Precipitation	1422	1547
Evapotranspiration	820.9	816
Groundwater in the shallow aquifer	264.8	302
Surface runoff	314.4	410
Transmission losses	11	12
Groundwater recharge	286	327

12, and 14). MOD16 ET data are available only for the period 2000–2010; hence, the comparison was done only for 5 years (2000–2004).

2.6.3 SWAT error index

A common problem of hydrological models is the wrong combination of the values of the calibrated parameters, which can also lead to good graphical results, and consequently good statistical values, but wrong water balance values. Consequently, good R^2 and NS values do not always denote the reliability of a model. R^2 and NS are common statistical parameters used to evaluate and compare time series in hydrological models (Abbaspour, 2015; De Almeida Bressiani et al., 2015; Dile and Srinivasan, 2014; Gebremicael et al., 2013). Additionally, rainfall distribution, parameterization, and evapotranspiration are also crucial points to be considered in any hydrological model. Therefore, in this study, after good calibration and validation values for R^2 and NS were achieved, and after a comparison between the SWAT ET and MOD16 ET values was done, an index (the SEI) to quantify the model quality has been introduced. This index is intended to be used only as an additional indicator to assess the reliability of the SWAT model, where the rRMSE was chosen as the fitting function.

Several reliable measured flow discharge datasets are available for rivers, but that is not the case for evapotranspiration data. However, satellite evapotranspiration data are available for most watersheds in the world. Furthermore, the measured discharge dataset and the satellite estimated evapotranspiration dataset do not have the same level of reliability. Therefore, SEI uses different weighting values (W_1 and W_2) to define differences in the level of reliability of the datasets: 0.7 for flow discharge and 0.3 for evapotranspiration. The proposed equation for SEI is as follows:

$$\text{SEI} = W_1 \left(\frac{\left(\sqrt{\text{rac} \sum_{i=1}^{n} (Q_{oi} - Q_{si})^2 N} \right)}{(Q_{o\,max} - Q_{o\,min})} \right) + W_2 \left(\frac{\left(\sqrt{\frac{\sum_{i=1}^{n} (\text{ET}_{oi} - \text{ET}_{si})^2}{N}} \right)}{(\text{ET}_{o\,max} - \text{ET}_{o\,min})} \right). \quad (8)$$

The first part of the equation corresponds to the rRMSE of the values obtained from the discharge data, where Q_{oi} is the observed discharge data (m^3 s^{-1}), Q_{si} is the simulated discharge data (m^3 s^{-1}), $Q_{o\,max}$ is the maximum value of the observed discharge data, and $Q_{o\,min}$ is the minimum value of the observed discharge dataset. The second part of the formula corresponds to the rRMSE achieved from the evapotranspiration data that were obtained from MOD16, where ET_{oi} is the MOD16 evapotranspiration values, ET_{si} is the SWAT-simulated evapotranspiration data, and $\text{ET}_{o\,max}$ and $\text{ET}_{o\,min}$ are the maximum and minimum values of the MOD16 evapotranspiration data, respectively. W_1 and W_2 are the assigned weighted values for discharge and evapotranspiration, respectively.

SEI ranges from 0 to $+\infty$, with 0 corresponding to the ideal value. When the SEI value of the model is closer to 0, the model will have a better match with the flow discharge and the evapotranspiration data. Since SEI includes the rRMSE values for discharge and evapotranspiration data, a model with good SEI results represents a model with a good agreement between these two hydrological processes, which are two important processes influencing the water balance of a watershed. By analyzing the SEI results, the quality of the combination of the parameter used for the calibration could also be evaluated and is less expectable to have a wrong parameterization. SEI was tested for two cases: the first one in whole upper Blue Nile Basin and the second in the Ribb sub-catchment in the Lake Tana region.

3 Results and discussions

3.1 Impact of different subcatchment discretization levels and rain gauge combinations

After analyzing the different datasets under different discretization levels, it was detected that the input data and the parameterization have a critical impact not only on the water balance but also on the sub-basins' distribution. The water balance analysis was done for two calibrated stations, three datasets, and two different sub-basin distributions. Water balance results for the upper Blue Nile Basin and also the values
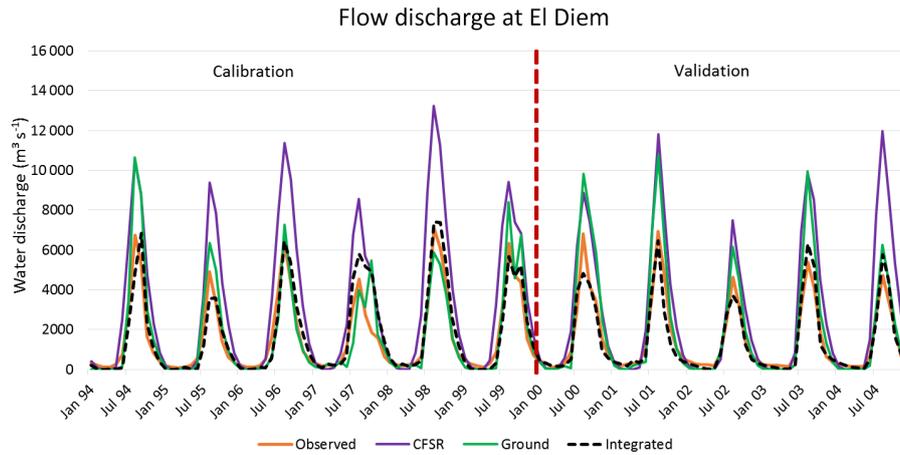
Figure 7. Calibration and validation of SWAT30 at El Diem. Calibration results achieved R^2 and NS values of integrated data: 0.88, 0.84; ground data: 0.86, 0.74; and CFSR data: 0.94, -0.51, respectively. Validation results achieved R^2 and NS values of integrated data: 0.92, 0.91; ground data: 0.96, 0.45; and CFSR data: 0.92, -0.48, respectively.
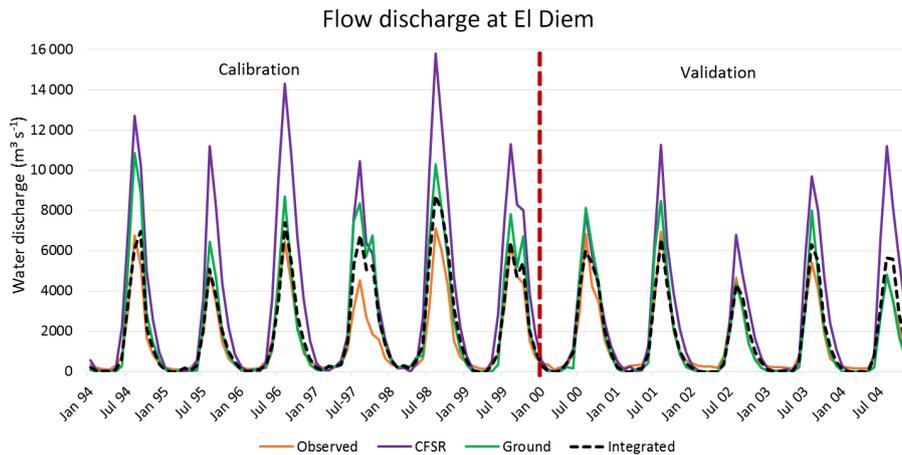
Figure 8. Calibration and validation of SWAT87 at El Diem. Calibration results achieved R^2 and NS values of integrated data: 0.92, 0.80; ground data: 0.92, 0.43; and CFSR data: 0.96, -1.54, respectively. Validation results achieved R^2 and NS of integrated data: 0.94, 0.91; ground data: 0.95, 0.85; and CFSR data: 0.89, -0.05, respectively.

for the different hydrological processes and models are given in Table 3; values for these hydrological processes from the literature are also given in Table 1 (Cherie, 2013; Mengistu and Sorteberg, 2012). The average annual precipitation in the upper Blue Nile Basin differs between the literature (Table 1) and also between dataset sources (Table 3). The uncertainty of the rainfall in the upper Blue Nile Basin is also noticeable when models with different sub-basin delineations are compared and show different values (Table 3, Figs. 7 and 8 for El Diem; Figs. 9 and 10 for Kessie; with SWAT30 and SWAT87, respectively). With the values provided in Table 2, it was possible to obtain good statistical values for the calibrated models (Table 4).

Figures 7 and 8 show the magnitude and dynamics of the measured and estimated monthly discharge flow at El Diem. The integrated dataset provided good statistical values for R^2

and NS (Table 4) under both discretization levels. The other models using the ground and CFSR datasets also showed good R^2 results but very low NS values, with the exception of SWAT87 with ground data (Table 4, Figs. 7 and 8). Although R^2 is always high in all the models, R^2 is a coefficient that measures only the dynamics of a model, meaning that the models behave with accuracy matching the seasonality of the rainfalls and dry periods in the upper Blue Nile Basin. However, NS is probably a more important factor to be considered, as it can be used to quantitatively describe the accuracy of model outputs. Calibrations and validations at Kessie showed good statistical values for the models using the ground and integrated datasets, achieving good R^2 and NS values (Table 4, Figs. 9 and 10).

SWAT30 under the CFSR dataset provides an average annual precipitation of 1253 mm (Table 3), while SWAT87 av-
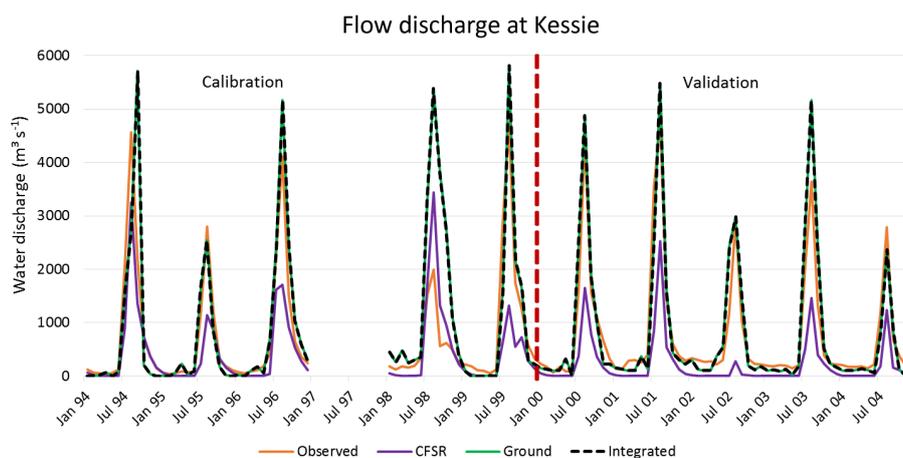
Figure 9. Calibration and validation of SWAT30 at Kessie. Calibration results achieved R^2 and NS values of integrated data: 0.74, 0.74; ground data: 0.74, 0.72; and CFSR data: 0.87, 0.46, respectively. Validations results achieved R^2 and NS values of integrated data: 0.76, 0.74; ground data: 0.78, 0.74; and CFSR data: 0.86, 0.49, respectively.
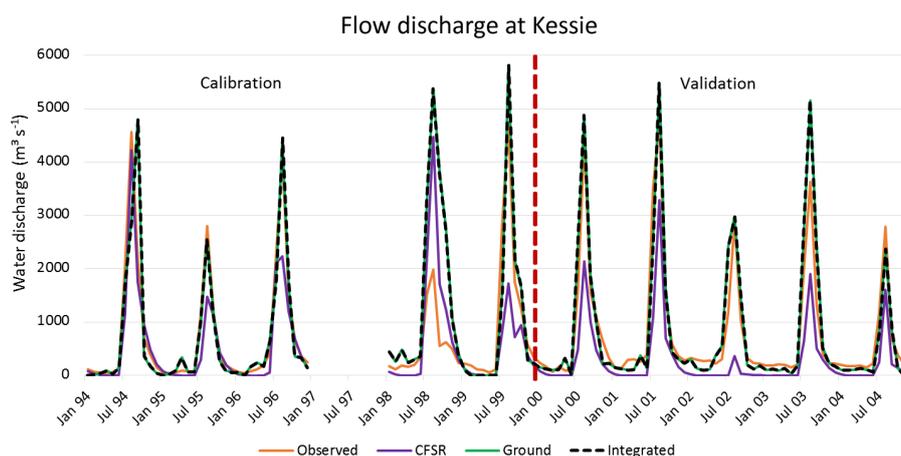
Figure 10. Calibration and validation of SWAT87 at Kessie. Calibration results achieved R^2 and NS values of integrated data: 0.77, 0.72; ground data: 0.77, 0.72; and CFSR data: 0.77, 0.37, respectively. Validations results achieved R^2 and NS values of integrated data: 0.78, 0.78; ground data: 0.80, 0.76; and CFSR data: 0.74, 0.37, respectively.

erage annual precipitation increases to 1481 mm. This rainfall increase provided by the CFSR dataset is caused by the number of sub-basins, SWAT87 considered more stations than SWAT30. However, both average annual precipitation values compared to the other two datasets and to the literature (Table 1) are still within acceptable ranges for the upper Blue Nile Basin, and it is not the main factor affecting the water balance, but its distribution in the watershed (Fig. 4). Figures 9 and 10 showed how CFSR data underestimate the precipitation in the eastern part of the basin (at Kessie) compared to those provided by the ground and integrated datasets. Figures 9 and 10 also showed the effect of the number of sub-basins on the simulated discharge flow. The flow discharge provided by the CFSR data is slightly higher in SWAT87 compare to SWAT30, although in both cases this dataset continues to underestimate the flow discharge

at Kessie. As the precipitation in the watershed changes in magnitude and distribution, the parameterization for the calibration of the models will be different. Therefore, in order to meet good R^2 and NS for the model with an incorrect precipitation distribution (in this case, the CFSR data), the values of the parameters needed to be modified to unrealistic values.

3.2 Average annual evapotranspiration and the impact of different data sources and potential evapotranspiration methods

The evapotranspiration has been another critical factor subject to analysis in this study. Depending on the weather dataset, the evapotranspiration values in the upper Blue Nile Basin varied from 729 mm yr^{-1} in SWAT30 with the CFSR dataset up to 932 mm yr^{-1} in SWAT30 with the in-
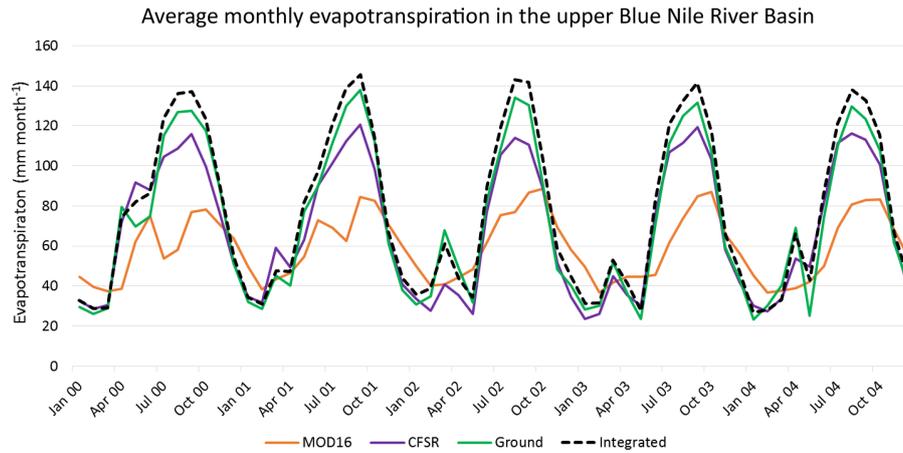
Figure 11. Average monthly evapotranspiration analysis using SWAT87 and the Hargreaves method, with R^2 and NS values of the integrated dataset: 0.63, -2.32; ground dataset: 0.60, -1.32; and CFSR dataset: 0.63, -1.20, respectively, compared to the MOD16 data.
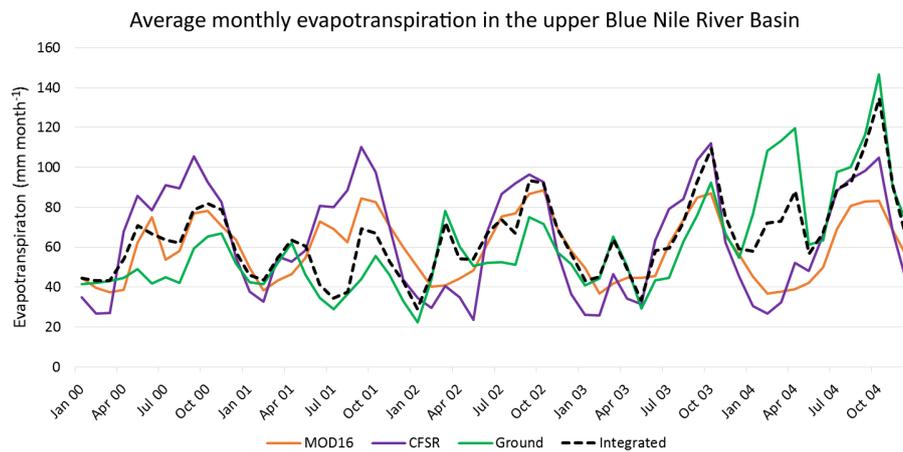
Figure 12. Average monthly evapotranspiration analysis using SWAT87 and the Penman–Monteith method, with R^2 and NS values of the integrated dataset: 0.36, -0.02; ground dataset: 0.34, -0.10; and CFSR dataset: 0.74, 0.03, respectively, compared to the MOD16 data.

tegrated dataset. SWAT models using the ground and integrated datasets and the Hargreaves equation showed acceptable discharge values and trends compared to those of measured discharge data (Figs. 7 and 8). However, the models overestimated the evapotranspiration values compared to those provided by MOD16 (Fig. 11). Nevertheless, when using the Penman–Monteith method, the SWAT models using the ground and integrated datasets provided more similar evapotranspiration values, and better R^2 and NS values compared to the values given by the MOD16 evapotranspiration data (Fig. 12). The best match with the evapotranspiration values provided by MOD16 is obtained using the CFSR dataset; this model provided low evapotranspiration values (Fig. 12) and consequently overestimated the flow discharges (Figs. 7 and 8). For the second test done in the Ribb subcatchment, the evapotranspiration rates provided by the ground and CFSR datasets were much better, having relatively good

statistical values compared to those obtained at large scale in the upper Blue Nile Basin (Figs. 13 and 14).

3.3 SEI evaluation

In the first case, SEI results for the El Diem station (Table 5) showed that the behavior and capability of SEI to quantify the level of error of a model through an evaluation of both flow discharge and evapotranspiration estimations is good. For instance, values in Table 5 showed that the lower the value of the discharge data is, the more the value for evapotranspiration tends to increase. This is because the flow discharge data are being matched; however, the evapotranspiration increases and tends to overestimate those values provided by MOD16 ET. If MOD16 ET had a good representation of the evapotranspiration data of a watershed, then the rRMSE values for both discharge and evapotranspiration values should be closer to 0, which could provide better SEI val-
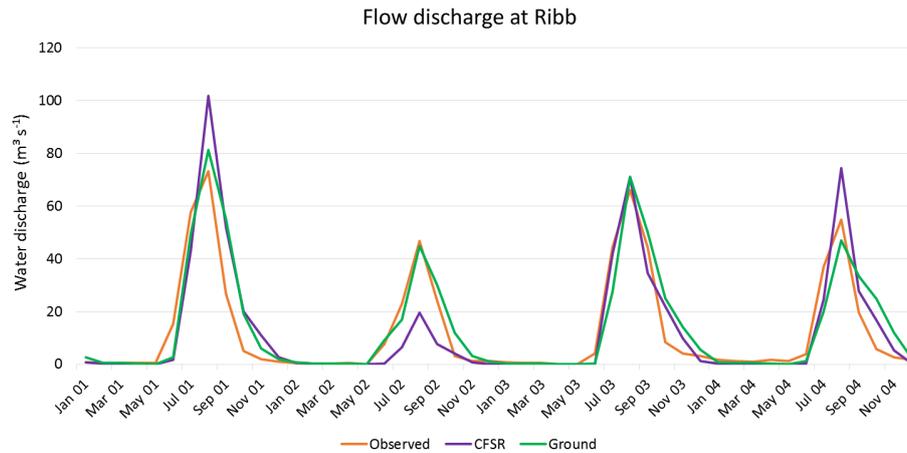
Figure 13. Flow discharge in the Ribb subcatchment. Calibration with outflow data achieved R^2 and NS values of the CFSR dataset: 0.81, 0.75; and ground dataset: 0.85, 0.83, respectively.
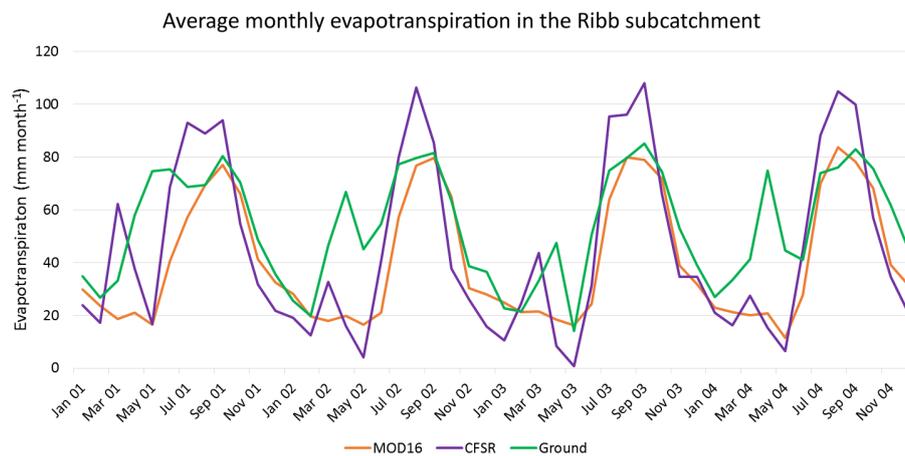
Figure 14. Average monthly evapotranspiration in the Ribb subcatchment. Statistical results achieved R^2 and NS values of the CFSR dataset: 0.78, 0.47; and ground dataset: 0.59, 0.24, respectively, compared to the MOD16 data.

ues (in the second test done at Ribb subcatchment). However, SEI showed that the models using the integrated datasets are more reliable than the other two datasets, achieving SEI values of 0.29 and 0.27 for SWAT30 and SWAT87, respectively. It also demonstrated that the CFSR dataset is less accurate, with an SEI value of 0.4 for both SWAT30 and SWAT87. In the second test done at the Ribb subcatchment, the calibration with flow discharge data provided good statistical results, where the CFSR dataset achieved R^2 and NS values of 0.81 and 0.75, respectively, and the ground dataset achieved R^2 and NS values of 0.85 and 0.83, respectively (Fig. 13 and Table 6). Unlike the SEI test performed for the entire upper Blue Nile Basin, the statistical results obtained from the comparison of the evapotranspiration data in the Ribb subcatchment are significantly better. The CFSR dataset achieved R^2 and NS values of 0.78 and 0.47, respectively, while the ground dataset achieved R^2 and NS values of 0.59 and 0.24, respectively (Fig. 14 and Table 6). SEI

showed better values than those obtained from the first test done in the whole Blue Nile Basin. The CFSR dataset provided better R^2 and NS values than the ground dataset for the evapotranspiration analysis; however, the ground dataset performed better during the calibration with outflow data (Table 6). The SEI value for both datasets was 0.16, a much better value that those obtained in the first test (Table 5). This second test provides a better understanding of how SEI works; it also proved how using reliable evapotranspiration data can improve the SEI values.

4   Conclusions

The CFSR dataset and a conventional observed ground dataset were analyzed in terms of statistical results, water balance, and precipitation distribution in the upper Blue Nile Basin. After detecting their limitations and disadvantages, an integration of both datasets was proposed with the purpose of

Table 2. Parameterization of the SWAT models using the SUFI-2 algorithm for the period 1990–2004. BSN means applied to the entire basin.

Parameter	Description	Type of change	Threshold Min	Max	Fitted value	Ranges of fitted absolute values for the Blue Nile Basin calibration
CN2	Curve number for moisture condition 2	r	−0.1	0.1	−0.05	60-87
SOL_AWC	Available water capacity of the soil	r	−2	2	1.7	0.095–0.49
ESCO	Soil evaporation compensation factor HRU	v	0.01	1	0.01	0.01
EPCO	Plant uptake compensation factor HRU	v	0.01	1	0.01	1
ESCO	Soil evaporation compensation factor BSN	v	0.01	1	0.01	0.01
EPCO	Plant uptake compensation factor BSN	v	0.01	1	0.01	1
CANMX	Maximum canopy storage	v	0	100	100	57
RCHRG_DP	Deep aquifer percolation fraction	v	0.01	1	0.01	0.01

Table 3. Water balance analysis in the upper Blue Nile Basin (1990–2004).

Hydrological component	Water balance in the Blue Nile Basin (all values in mm yr^{-1})					
	SWAT30			SWAT87		
	CFSR data	Ground data	Integrated data	CFSR data	Ground data	Integrated data
Precipitation	1253	1301	1270	1481	1209	1243
Evapotranspiration	729	887	932	848	798	860
Revap/shal. aquifer	27	31	31	27	27	28
Surface runoff	172	167	114	228	166	125
Return flow	274	107	139	307	136	147
Lateral flow	40	50	50	80	73	74
Perc. to deep aquifer	313	199	175	349	168	181
Rechg. deep aquifer	16	10	9	17	8	9

overcoming their uncertainties and limitations. This data integration method was effectively used in the upper Blue Nile Basin to create a better SWAT model and can also be applied in other watersheds where observed data are limited and incomplete. However, data analyses and tests should always be performed before performing an integration for other watersheds. Despite its limitations, the CFSR dataset continues

to be an important source that can be very useful in regions where conventional measured data are not available.

A comparison of the three datasets under different discretization levels was also performed. This comparison was important for obtaining a better understanding of how crucial the sub-basin discretization process is during a SWAT model setup. The comparisons showed that the three input datasets, under models with a different number of sub-basins, yield

Table 4. Statistical results for the calibrations and validations with outflow data at the El Diem and Kessie gauging stations. Bold values mean "models with low performance".

Sub-basins		CFSR dataset		Ground dataset		Integrated dataset	
		30	87	30	87	30	87
El Diem							
Calibration	R^2	0.94	0.96	0.86	0.92	0.88	0.92
	NS	**−0.51**	**−1.54**	0.74	**0.43**	0.84	0.80
	p factor	0.53	0.36	0.66	0.67	0.70	0.77
	r factor	1.11	0.93	0.83	0.68	0.67	0.54
Validation	R^2	0.92	0.89	0.96	0.95	0.92	0.94
	NS	**−0.48**	**−0.05**	**0.45**	0.85	0.91	0.91
Kessie							
Calibration	R^2	0.87	0.77	0.74	0.77	0.74	0.77
	NS	**0.46**	**0.37**	0.72	0.72	0.74	0.72
	p factor	0.49	0.57	0.60	0.63	0.60	0.63
	r factor	0.61	0.71	0.72	0.59	0.72	0.59
Validation	R^2	0.86	0.74	0.78	0.80	0.76	0.78
	NS	**0.49**	**0.37**	0.74	0.76	0.74	0.78

Table 5. SEI results for the upper Blue Nile Basin.

Process	Weighting	CFSR dataset		Ground dataset		Integrated dataset	
		rRMSE	Weighted rRMSE	rRMSE	Weighted rRMSE	rRMSE	Weighted rRMSE
SWAT30							
Water discharge	0.7	0.33	0.231	0.17	0.119	0.098	0.068
Evapotranspiration	0.3	0.58	0.174	0.70	0.21	0.75	0.225
SWAT error index		0.4		0.33		0.29	
SWAT87							
Water discharge	0.7	0.37	0.259	0.17	0.119	0.1	0.07
Evapotranspiration	0.3	0.46	0.138	0.58	0.174	0.66	0.198
SWAT error index		0.4		0.29		0.27	

Table 6. Statistical results for the Ribb subcatchment in the Lake Tana region of the upper Blue Nile Basin.

Statistical results for the Ribb subcatchment								
Process	Weighting	CFSR dataset				Ground dataset		
		R^2	NS	rRMSE	Weighted rRMSE	R^2	NS	rRMSE	Weighted rRMSE
Water discharge	0.7	0.81	0.75	0.13	0.091	0.85	0.83	0.11	0.077
Evapotranspiration	0.3	0.78	0.47	0.23	0.069	0.59	0.24	0.28	0.084
SWAT error index				0.16				0.16	

different results. The number of sub-basins in a SWAT model will affect the magnitude of the flow discharge and hence the total water balance of a watershed.

The comparison of the results of SWAT30 demonstrates that the values for the total annual average precipitation at El Diem are similar for the three datasets. Nevertheless, only

the model using the CFSR dataset was not able to achieve good water balance results under similar parameterization. The quality of the CFSR rainfall data is not reliable for the upper Blue Nile Basin, although this case cannot be generalized for other watersheds in the world. However, this dataset needs to be equally verified in other watersheds before it is used. For the second case, the three datasets were analyzed in more detail using SWAT87, and although an exact number of the correct precipitation amounts in the upper Blue Nile Basin cannot be given, CFSR data showed an overestimation of the rainfall and also a wrong precipitation distribution compared to the other datasets. Additionally, the model under 87 sub-basins was the model that provided more details in terms of the number of HRUs and also achieved better statistical values. Therefore, this study proposes that 87 is a suitable number of sub-basins for the upper Blue Nile Basin. SWAT87 is more suitable to perform several types of hydrological analyses and propose watershed management practices in the Blue Nile Basin.

Furthermore, the SEI has proved to be a useful additional tool to express the level of error of SWAT models. This index used the weighted rRMSE of the discharge and evapotranspiration data. SEI was tested in two locations, being the second case done at the Ribb subcatchment more accurate. Nevertheless, further tests and improvements should be done to this index. SEI also showed that the integrated dataset successfully achieved better and more reliable results than the ground and CFSR datasets. The integrated dataset improved the results of the model, obtaining better R^2, NS, and SEI values.

Although further improvements must be made to the methods proposed in this study, the integration of datasets, the sub-basin delineation, and the application of the SEI are important approaches that can be applied in other watersheds and can significantly help to develop better hydrological models.

References

Abbaspour, K. C.: SWAT-CUP: SWAT calibration and uncertainty programs-A user manual, Tech. rep., Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland, 2015.

Abera, W., Formetta, G., Brocca, L., and Rigon, R.: Modeling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data, Hydrol. Earth Syst. Sci., 21, 3145–3165, https://doi.org/10.5194/hess-21-3145-2017, 2017.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop Evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper No. 56, Food and Agriculture Organization, Land and Water, Rome, Italy, 1998.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B., and Neitsch, S. L.: Soil & Water Assessment Tool, Input/Output documentation, version 2012, Texas Water Resources Institute, Texas, 246–248, 2012.

Bastidas, L. A., Gupta, H. V., and Sorooshian, S.: Emerging paradigms in the calibration of hydrologic models, Mathematical Models of Large Watershed Hydrology, Water Resources Publications, LLC, Englewood, CO, USA, 1, 25–56, 2002.

Betrie, G. D., Mohammed, Y. A., Van Griensven, A., and Srinivasan. R.: Sediment management modelling in the Blue Nile Basin using SWAT model, Hydrol. Earth Syst. Sci., 15, 807–818, https://doi.org/10.5194/hess-15-807-2011, 2011.

Chebud, Y. A. and Melesse, A. M.: Modelling lake stage and water balance of Lake Tana, Ethiopia, Hydrol. Process., 23, 3534–3544, 2009.

Cherie, N. Z.: Downscaling and Modeling the Effects of Climate Change on Hydrology and Water Resources in the Blue Nile Basin, Ethiopia, PhD Thesis, University of Kassel, Kassel, Germany, 2013.

De Almeida Bressiani, D., Srinivasan, R., Jones, C. A., and Mendiondo, E. M.: Effects of different spatial and temporal weather data resolutions on the stream flow modeling of a semi-arid basin, Northeast Brazil, Int. J. Agr. Biol. Eng., 8, 1–16, https://doi.org/10.3965/j.ijabe.20150803.970, 2015.

Dessie, M., Verhoest, N. E. C., Pauwels, V. R. N., Admasu, T., Poesen, J., Adgo, E., Deckers, J., and Nysse, J.: Analyzing runoff processes through conceptual hydrological modelling in the Upper Blue Nile Basin, Ethiopia, Hydrol. Earth Syst. Sci., 18, 5149–5167, https://doi.org/10.5194/hess-18-5149-2014, 2014.

Dessie, M., Verhoest, N. E. C., Pauwels, V. R. N., Adgo, E., Deckers, J., Poesen, J., and Nyssen, J.: Water balance of a lake with

floodplain buffering: Lake Tana, Blue Nile Basin, Ethiopia, J. Hydrol., 522, 174–186, 2015.

Dile, Y. T. and Srinivasan, R.: Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin, J. Am. Water Resour. Assoc., 50, 1226–1241, https://doi.org/10.1111/jawr.12182, 2014.

Easton, Z. M., Fuka, D. R., White, E. D., Collick, A. S., Ashagre, B. B., McCartney, M., Awlachew, S. B., Ahmed, A. A., and Steenhuis, T. S.:. A multi basin SWAT model analysis of ruoff and sedimentation in the Blue Nile, Ethiopia, Hydrol. Earth Syst. Sci., 14, 1827–1841, https://doi.org/10.5194/hess-14-1827-2010, 2010.

FAO.: World reference base for soil resources 2014: International soil classification system for naming soils and creating legends for soil maps, World Soil Resources Report, Food and Agriculture Organization of the United Nations, Rome, Italy, 2015.

Gassman, W. P., Reyes, M. R., Green, C. H., and Arnold, J. G.: The soil and water assessment tool: historical development, applications, and future research direction, T. ASABE, 50, 1211–1250, 2007.

Gebremicael, T. G., Mohamed, Y. A., Betrie, G. D., van der Zaag, P., and Teferi, E.: 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, https://doi.org/10.1016/j.jhydrol.2012.12.023, 2013.

Hargreaves, G. H. and Samani, Z. A.: Estimating potential evaporation, J. Irrig. Drain. Eng.-ASCE, 108, 225–230, 1982.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1, 96–99, 1985.

Hauhs, M. and Lange, H.: Classification of runoff in headwater catchments: A physical problem?, Geogr. Compass, 2, 235–254, 2008.

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

Malunjkar, V. S., Shinde, M. G., Ghotekar, S. S., and Atre, A. A.: Estimation of Surface Runoff using SWAT Model, Int. J. Invent. Eng. Sci., 3, 2015.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, https://doi.org/10.1029/2009RG000314, 2010.

McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J., and Weiler, M.: Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology, Water Resour. Res., 43, 1–6, 2007.

Me, W., Abell, J. M., and Hamilton, D. P.: Effects of hydrologic condition on SWAT model performance and parameter sensitivity for a small, mixed land use catchment in New Zealand, Hydrol. Earth Syst. Sci., 19, 4127–4147, https://doi.org/10.5194/hess-19-4127-2015, 2015.

Mengistu, D. T. and Sorteberg, A.: Sensitivity of SWAT simulated streamflow to climatic changes within the Eastern Nile River basin, Hydrol. Earth Syst. Sci., 16, 391–407, https://doi.org/10.5194/hess-16-391-2012, 2012.

Monteith, J. L.: Evapotranspiration and Environment, Symp. Soc. Exp. Biol., 19, 205–234, 1965.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, T. ASABE, 50, 885–900, 2007.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, 2007.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, 2011.

Mulungu, D. M. M. and Munishi, S. E.: Simiyu river catchment parameterization using SWAT model, J. Phys. Chem. Earth A/B/C, 32, 1032–1039, 2007.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I: A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., and Williams, J. R.: Soil and Water Assessment Tool Theoretical Documentation Version 2009, Grassland, Soil and Water Research Laboratory and Black land Research Center, Texas, 2009.

Rafiei Emam, A., Kappas, M., Hoang Khanh Nguyen, L., and Renchin, T.: Hydrological Modeling in an Ungauged Basin of Central Vietnam Using SWAT Model, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2016-44, 2016.

Roth, V. and Lemann, T.: Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands, Hydrol. Earth Syst. Sci., 20, 921–934, https://doi.org/10.5194/hess-20-921-2016, 2016.

Ruhoff, A. L., Paz, A. R., Aragao, L. E. O. C., Mu, Q., Malhi, Y., Collischonn, W., Rocha, H. R., and Running, S. W.: Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin, Hydrological Sciences Journal – Journal des Sciences Hydrologiques, 58, 1658–1676, https://doi.org/10.1080/02626667.2013.837578, 2013.

Setegn, S. G., Srinivasan, R., and Dargahi, B.: Hydrological Modelling in the Lake Tana Basin, Ethiopia Using SWAT Model, Open Hydrol. J., 2, 49–62, 2008.

Setegn, S. G., Darfahi, B., Srinivasan, R., and Melesse, A. M.: Modeling of sediment yield from Anjeni-gauged watershed, Ethiopia using SWAT model, J. Am. Water Resour. Assoc., 46, 514–526, https://doi.org/10.1111/j.1752-1688.2010.00431.x, 2010a.

Setegn, S. G., Srinivasan, R., Melese, A. M., and Dargahi, B.: SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia, Hydrol. Process., 24, 357–367, 2010b.

Sivapalan, M.: Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., John Wiley & Sons, Ltd., https://doi.org/10.1002/0470848944.hsa012, 2006.

Sun, Z., Wang, Q., Ouyang, Z., Watanabe, M., Matsushita, B., and Fukushima, T.: Evaluation of MOD16 algorithm using MODIS

and ground observational data in winter wheat field in North China Plain, Hydrol. Process., 21, 1196–1206, 2007.

Swallow, B. M., Sang, J. K., Nyabenge, M., Bundotich, D. K., Duraiappah, A. K., and Yatich, T. B.: Tradeoffs, synergies and traps among ecosystem services in the Lake Victoria basin of East Africa, Environ. Sci. Poll., 12, 504–519, https://doi.org/10.1016/j.envsci.2008.11.003, 2009.

Tanveer, A., Ghulam, N., Boota, M. W., Fiaz, H., Azam, M. I., Jun, J. H., and Faisal, M.: Uncertainty analysis of runoff and sedimentation in a forested watershed using sequential uncertainty fitting method, Sci. Cold Arid Reg., 8, 297–310, 2016.

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

Tibebe, D. and Bewket, W.: Surface runoff and soil erosion estimation using the SWAT model in the Keleta watershed, Ethiopia, Land Degrad. Dev.,22, 551–564, https://doi.org/10.1002/ldr.1034, 2011.

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

van Griensven, A., Ndomba, P., Yalew, S., and Kilonzo, F.: Critical review of SWAT applications in the upper Nile basin countries, Hydrol. Earth Syst. Sci., 16, 3371–3381, https://doi.org/10.5194/hess-16-3371-2012, 2012.

Wale, A.: Hydrological balance of Lake Tana Upper Blue Nile Basin, Ethiopia, Enschede, the Netherlands, 2008.

Wang, H., Sun, F., Xia, J., and Liu, W.: Impact of LUCC on streamflow based on the SWAT model over the Wei River basin on the Loess Plateau in China, Hydrol. Earth Syst. Sci., 21, 1929–1945, https://doi.org/10.5194/hess-21-1929-2017, 2017.

Worqlul, A. W., Maathuis, B., Adem, A. A., Demissie, S. S., Langan, S., and Steenhuis, T. S.: Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia, Hydrol. Earth Syst. Sci., 18, 4871–4881, https://doi.org/10.5194/hess-18-4871-2014, 2014.

Yang, J., Reichert, P., and Abbaspour, K. C.: Hydrolgocial modeling of the Chaohe Basin in China: Statistical model formulation and bayesian inference, J. Hydrogl., 340, 167–182, https://doi.org/10.1016/j.jhydrol.2007.04.006, 2007.

Zhang, L., Dawes, W. R., and Walker, G. R.: Predicting the effect of vegetation changes on catchment average water balance, Tech. Rep. 99/12, Coop. Res. Cent. Catch. Hydrol., Canbera, 1999.