Hydrological Evaluation of Satellite-Based Rainfall Estimates over the Volta and Baro-Akobo Basin

Vera Thiemig[1]

Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy

Utrecht University, Faculty of Geosciences, Utrecht, The Netherlands

Rodrigo Rojas

Climate Risk Management Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy

Mauricio Zambrano-Bigiarini

Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy

Ad De Roo

Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy

Utrecht University, Faculty of Geosciences, Utrecht, The Netherlands

[1] *Corresponding author address*: Vera Thiemig, Via E. Fermi 2749, TP 261, 21027 Ispra (VA), Italy. E-mail: vera.thiemig@gmail.com

Abstract

How useful are satellite-based rainfall estimates (SRFE) as forcing data for hydrological applications? *Which SRFE should be favoured for hydrological modelling? What could researchers do to increase the performance of SRFE-driven hydrological simulations*? To address these three research questions, four SRFE (CMORPH, RFE 2.0, TRMM-3B42 and PERSIANN) and one reanalysis product (ERA-Interim) are evaluated within a hydrological application for the time period 2003-2008, over two river basins (Volta and Baro-Akobo) which hold distinct physiographic, climatologic and hydrologic conditions. The focus was on the assessment of: a) the individual and combined effect of SRFE-specific calibration and bias correction on the hydrological performance, b) the level of complexity required regarding bias correction and interpolation to achieve a good hydrological performance, and c) the hydrological performance of SRFE during high- and low-flow conditions. Results show that 1) the hydrological performance is always higher if the model is calibrated to the respective SRFE rather than to interpolated ground observations; 2) for SRFE that are afflicted with bias, a bias-correction step prior to SRFE-specific calibration is essential, while for SRFE with good intrinsic data quality applying only a SRFE-specific model calibration is sufficient; 3) the more sophisticated bias-correction method used in this work (histogram equalization) results generally in a superior hydrological performance, while a more sophisticated spatial interpolation method (Kriging with External Drift) seems to be of added value only over mountainous regions; 4) the bias correction is not over-proportionally important over mountainous catchments, as it solely depends on where the SRFE show high biases (e.g. for PERSIANN and CMORPH over lowland areas); and 5) the hydrological performance during high-flow conditions is superior thus promoting the use of SRFE for applications focusing on the high-end flow spectrum. These results complement a preliminary "ground truthing" phase and provide insight on the usefulness of SRFE for

hydrological modelling and under which conditions they can be used with a given level of reliability.

Keywords: satellite-based rainfall estimates; SRFE; hydrological modeling; hydrological evaluation; bias correction; Upper Nile

1. Introduction

Hydrological models facilitate worldwide the efficient management of one of the most valuable natural resources: water. A plethora of hydrological applications have been developed aiming at quantifying each (terrestrial) component of the water cycle for past, present and future conditions (see, e.g. (Döll et al., 2003; Silberstein, 2006). Results from these models are used to, for example, issue flood warnings (e.g. (Cloke and Pappenberger, 2009), estimate drinking water availability (e.g. (Soboll et al., 2011), determine ecological flows required to maintain a healthy environment (e.g. (Dyson et al., 2008)), or to optimise water allocation schemes (e.g. (de Condappa et al., 2009)). The reliability and accuracy of these applications is therefore essential for decision-making and usually entails some sort of economic, social and environmental benefits and costs.

Precipitation data is the most crucial atmospheric driver for hydrological modelling as it influences the accuracy of these applications to a large extent. In this context, the global decline of rain gauge networks proves to be disadvantageous (Hughes, 2006). This has led researchers to consider the use of satellite-derived rainfall estimates (SRFE) instead. With a suitable spatio-temporal resolution (e.g. 0.25° and 24 h), and being released uninterrupted and in near real-time, publically available, and easily accessible, most SRFE hold a large potential as forcing data for medium- to large-scale hydrological modelling, especially for data-sparse and ungauged basins.

However, SRFE are subjected to a variety of potential errors, which originate from e.g. discontinuous revisit time of observing sensors and weak relationships between remotely sensed signal and rainfall rate (Bitew and Gebremichael, 2011). In this regard, a commonly experienced flaw of SRFE is the bias. The presence of bias in precipitation estimates is unfavourable for water balance calculations as the total water quantity is preserved within the hydrological model. Therefore, the questions at stake are: 1) *How useful are these SRFE as forcing data for hydrological modelling*? 2) *Which SRFE should be favoured for hydrological modelling? 3) What could researchers do to increase the performance of SRFE-driven hydrological simulations*? Answering these questions would allow us to provide insight about the appropriateness of using SRFE for hydrological applications. To ensure a justified usage of SRFE as input to hydrological models, however, a thorough validation is required.

There are two methods for validating SRFE: either through ground truthing, or through model-based applications. The first method refers to the traditional approach comparing SRFE against ground observed precipitation. This approach has been applied extensively, resulting into a comprehensive literature (here relevant for Africa only: (Adler et al., 2003; Ali et al., 2005; Asadullah et al., 2008; Dinku et al., 2010; Dinku et al., 2007; Diro et al., 2009; Hughes, 2006; Laurent et al., 1998; McCollum et al., 2000; Nicholson et al., 2003; Symeonakis et al., 2009; Thorne et al., 2001; Xie and Arkin, 1995)). The second approach refers to the evaluation of SRFE by assessing their performance within a target application. An example of this approach is the evaluation of SRFE based on their capabilities to reproduce the observed streamflow, also referred to as "hydrological evaluation". This method is rather recent but continues to gain popularity amongst researchers (see e.g. (Artan et al., 2007; Behrangi et al., 2011; Bitew and Gebremichael, 2011; Gourley et al., 2011; Jiang et al., 2012)). Even though both methods can be independently applied, they can be considered as complementary: the first one provides insight into the intrinsic data quality of the SRFE, whereas the second one assesses the usefulness of the SRFE within a certain application.

However, the abovementioned studies on the hydrological evaluation of SRFE, a) validated either a single SRFE over a wider area or multiple SRFE over a single target area; b) used traditional performance indicators such as the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1974), bias (absolute, relative, normalised or fractional), Root Mean Square Error (RMSE, standard or normalised), Mean Absolute Error (MAE) or coefficient of determination (R^2); c) examined the improvement in hydrological performance by calibrating the model with the respective SRFE rather than with rain gauge data; and d) mostly obviated a step to correct for biases in the precipitation estimates or applied a rather simple bias-correction technique.

This study provides an innovative perspective on the hydrological evaluation of SRFE for five reasons. First, we evaluate multiple SRFE over multiple physiographic and climatic conditions. Second, we assess the individual and combined effect of SRFE-specific model calibration and bias correction on the hydrological performance. Third, we make use of state-of-the-art calibration algorithms and a novel model performance indicator. Fourth, we test two different bias-correction methods to find the optimal way of compensating the bias of SRFE in data-sparse regions. Fifth, by combining detailed knowledge on the intrinsic data quality obtained during the ground truthing phase (Thiemig et al., 2012) with the results of this current study, we gain the unique opportunity to differentiate among potential impacts arising from the input data, the hydrological model and from the physiographic and climatic conditions on hydrological simulations in Africa.

In this study, we focus on the hydrological evaluation of four SRFE, namely, CMORPH, RFE 2.0, TRMM-3B42 and PERSIANN and one reanalysis product called ERA-Interim. These products are validated over two African basins (Volta and Baro-Akobo), which hold distinct physiographic and climatic conditions. For the hydrological assessment we use LISFLOOD (Van Der Knijff et al., 2010), a physically-based hydrological model, which has been calibrated using the Particle Swarm Optimisation (PSO) algorithm (Kennedy and Eberhart, 1995) for the time period 2003-2006. Additionally, we implement two different bias-correction methods to correct the bias in the SRFE: factor correction (FC) and histogram equalization (HE), in combination with two spatial interpolation methods, Inverse Distance Weighted (IDW) and Kriging with External Drift (KED) to define the observed targets for bias correction.

This study intends to answer the three aforementioned questions by focussing on: a) the impact of SRFE-specific model calibration and bias correction on the hydrological performance; b) regarding bias correction and spatial interpolation, the level of complexity of the method required to achieve an acceptable hydrological performance, and c) the usefulness of SRFE for specific flow conditions (high-flow and low-flow). Our results will help to elucidate the limits of predictability when using SRFE as input for hydrological modelling. The ultimate goal of this study is to provide insight on the usefulness of SRFE for hydrological modelling and to select the "best" way of increasing the hydrological performance given the limitations of each SRFE.

The remainder of the article is organised as follows: Section 2 describes the study areas and precipitation data. Section 3 presents the workflow, the hydrological modelling framework including details on LISFLOOD, the calibration algorithm, bias-correction methods and the performance indicator. Results are presented in Section 4, while discussion and concluding remarks are rounded off in Section 5 including among other things the answers to the research questions as well as recommendations for SRFE end-users.

2. Data

2.1. Study Areas

The hydrological evaluation of SRFE was done over the three upper catchments of the Volta River Basin, namely, Black Volta, White Volta and Oti, and the Upper Baro-Akobo catchment, which is part of the Nile River Basin. The study area including the delineation of sub-catchments and the location of meteorological and hydrological stations is shown in Figure 1.

The two basins differ from each other with respect to physiographic and climatic conditions as well as the hydrological responses. While the Volta is a medium- to large-size lowland basin, located in the tropical wet and dry zone, with a rather short but pronounced flood period from mid-July to the end of October with inter-annual variable flood peaks exceeding 2500 m^3/s, the Baro-Akobo is a small- to medium-size mountainous basin, with a typical highland climate and a prolonged flood period from June to November with flood peaks of only around 1200 m^3/s. Further details on topography and climate are presented in Table 1, while hydrological information is depicted in Figure 2.

(insert Figure 1 here)

(insert Table 1 here)

(insert Figure 2 here)

2.2. Precipitation data

2.2.1. Ground observations

Information regarding the number of meteorological ground stations, station density, data coverage and data provider can be obtained for each river basin from Table 1 (see Figure 1 for location of the stations). We consider this data set as representative since it is the most complete, accurate and independent information at hand, taking into consideration the general data availability, the quality checks done by the data provider and the fact that 79% of these observations are not part of the publically available GTS-station data, but data from national institution without public domain and hence are not input to any SRFE (explicitly TRMM-3B42 or RFE 2.0), respectively.

To be able to use point precipitation as forcing data to the hydrological model as well as to assess the relevance of the spatial interpolation method for bias correction (see Sections 3.2.3 and 4.3), point precipitation was interpolated to areal (raster) precipitation using two methods: Kriging with External Drift, and Inverse Distance Weighted (Burrough and McDonnell, 1998; Goovaerts, 2000). For KED interpolations we used high-resolution terrain elevation data (SRTM; Shuttle Radar Topography Mission) provided by NASA as a secondary variable (referred to as external drift) to define a trend to guide the estimation of the primary variable at each grid cell to improve the performance of the spatial interpolation of point precipitation. The interpolation was executed on a daily time step for the 6-year time period (2001-2006). The whole process was automated by using the hydroTSM R package (Zambrano-Bigiarini, 2011). The spatial resolution was set to 0.1° x 0.1° for both approaches.

The average annual KED and IDW precipitation fields are shown for both basins in the first and second column of Figure 3 respectively. The KED precipitation field shows for the Volta Basin an increasing gradient from the dry north (600 mm) to the wet south (1600 mm), ending in an abrupt reduction in precipitation (1000 mm) at the coastal zone, while the IDW precipitation field shows a rather homogeneous distribution ranging over the whole basin between 800 and 1000 mm. For the Baro-Akobo, both precipitation fields show an increasing precipitation gradient from the river mouth in the west to the highlands in the north and southeast; for the KED field this gradient ranges from 1400 mm to up to 2600 mm, while it ranges between 1500 mm and 1950 mm for the IDW field. The KED precipitation patterns are for both basins in full agreement with values reported by (Shahin, 2002) and (Romilly and Gebremichael, 2011), respectively.

2.2.2. Satellite-Based Rainfall Estimates

There are three main data sources for SRFE: geostationary thermal infrared (TIR), passive microwave (PMW), and rain gauges. Each of these data sources holds its particular strengths and limitations. For example, TIR data have a unique temporal and spatial coverage, sensing almost the whole globe every one hour or less. TIR information, i.e. the cloud top brightness temperature, is particularly valuable for the distinction between raining and non-raining, however they are rather poor in the estimation of the actual precipitation amount since the sensor signal does not penetrate through the clouds. PMW, on the contrary, proves better in estimating the precipitation amount due to the more direct physical relationship between sensor signal and precipitation, but runs on a much lower temporal frequency and on a coarser spatial resolution. Lastly, rain gauge data provide the most direct information about precipitation at surface level, but only for certain point locations and are not spatially inclusive and comprehensive. The concept of SRFE is to combine the favourable characteristics of the different data sources using various merging strategies, to achieve accurate precipitation estimates at surface level with a high spatial and temporal resolution.

The SRFE evaluated in this study were selected based on a number of characteristics such as: a) whole coverage over Africa, b) good temporal and spatial resolution (min. ≤ 24h and ≤ 0.25°), c) preferentially near real-time availability, and d) public domain availability. Additionally, we excluded SRFE that showed a poor intrinsic data quality during the ground truthing phase (Thiemig et al., 2012) and included the re-analysis product ERA-Interim for the sake of its underlying numerical weather prediction model, which is very similar to the one used to calculate the ECMWF-EPS. ERA-Interim and ECMWF-EPS will become of great importance in our future analysis of flood predictions and therefore we included ERA-Interim into the current analysis.

Table 2 provides details on the selected SRFE and re-analysis product (ERA-Interim), including spatio-temporal coverage and resolution, product data sources, merging techniques as well as the main outcome of the ground truthing phase over the respective study area. The average annual precipitation between 2003 and 2006 is shown in Figure 3, which gives a good indication on the accuracy of the SRFE to reproduce the spatial precipitation pattern. For more information regarding the nature of the SRFE the reader is referred to the additional references indicated in Table 2.

(insert Table 2 here)

(insert Figure 3 here)

3. Method: Hydrological evaluation

3.1. Workflow

To answer the research questions we focused on the three focal points as stated in Section 1 by following the workflow depicted in Figure 4.

In order to assess the effect of SRFE-specific calibration and/or bias-correction, a reference performance for each SRFE needed to be defined. Therefore, LISFLOOD was calibrated for each catchment using the interpolated observed precipitation fields obtained using KED (Step 1). It has been decided to use the KED fields as these resemble the observed precipitation pattern reported by (Shahin, 2002) and (Romilly and Gebremichael, 2011) the closest. The resulting calibrated (optimised) parameter set is referred to as the "base-line parameterisation" (BLP). In Step 2, each SRFE is run with the model set-up of Step 1 (BLP) resulting into the reference performance for each SRFE presented in Section 4.1. Once the reference performance has been calculated, the influence of SRFE-specific calibration and bias correction is assessed in Steps 3 and 4 respectively. In Step 3, LISFLOOD is re-calibrated for each SRFE and each sub-catchment, resulting into 20 model calibrations (5 SRFE x 4 sub-catchments), to investigate the impact of calibrating the hydrological model for each SRFE, without any bias correction. In Step 4, LISFLOOD is run for each sub-catchment, with each version of bias-corrected SRFE individually using the parameter set of Step 1 (60 model runs; 5 SRFE x 3 BC methods x 4 sub-catchments). Finally, the combined effect of SRFE-specific calibration and bias correction is assessed in Step 5, by recalibrating each model setting of Step 4 (60 model calibrations).

The results of Steps 3 to 5 (see Section 4.2) show the individual and combined effect of SRFE-specific calibration and bias correction on the hydrological performance. The most convenient bias-correction approach is discussed in Section 4.3 through a detailed analysis of Step 4. The usefulness of SRFE for different flow conditions is investigated in Section 4.4 based on a separate consideration of low- and high-flow conditions of the hydrological simulations of Step 5 (only HE-KED). Finally, the validation of the hydrological performance is done for each individual SRFE using the same model settings as of Step 5 (only HE-KED) in Section 4.5.

(insert Figure 4 here)

3.2. Hydrological Modelling Framework

3.2.1. LISFLOOD model

LISFLOOD is a fully-distributed and physically-based hydrological model developed for flood forecasting and impact assessment studies. This model simulates the spatial and temporal patterns of catchment responses in large river basins as a function of spatial information on topography, soils and land cover. LISFLOOD is a versatile GIS-based hydrological model used for large-scale assessment of water resources (floods or droughts), flood warnings (European Flood Awareness System, www.efas.eu) and climate change impacts. Within this frame the model has been tested exhaustively all over Europe (the whole list of publications is found on http://floods.jrc.ec.europa.eu/publications/floods-a-climate-change) and recently also over various parts in Africa (Thiemig et al., 2010). A complete description of the model structure and equations is available in (Van Der Knijff et al., 2010).

The hydrological model was set up for the two study areas with a spatial resolution of 0.1°. GIS-based model parameters were either extracted or derived from multiple data sources such as the Harmonized World Soil Database 1.0, the VGT4AFRICA project or the SRTM. Meteorological variables (except precipitation) were obtained from the ERA-Interim fields, while parameters related to the groundwater response, infiltration, groundwater losses and channel routing were determined through model calibration.

3.2.2. Model Calibration

LISFLOOD has been calibrated based on a 4-year period (2003-2006, using 2002 as warm-up) for each individual sub-catchment using raw and bias-corrected SRFE, respectively. Calibration was done using the hydroPSO R package (Zambrano-Bigiarini and Rojas, 2012), which implements a state-of-the-art Particle Swarm Optimisation (PSO) algorithm to carry out a global parameter optimisation.

PSO is an evolutionary optimisation algorithm originally developed by (Kennedy and Eberhart, 1995). In PSO each individual of the population (referred to as a particle) searches the global optimum in a multidimensional search-space considering the personal and collective past experiences. The algorithm is highly efficient and has been applied to a vast collection of case studies (see, e.g., (Poli, 2008). (Zambrano-Bigiarini and Rojas, 2013) validated hydroPSO against standard global optimisation algorithms such as the Shuffled Complex Evolution Algorithm (SCE-UA) (Duan et al., 1993), DiffeRential Evolution Adaptive Metropolis (DREAM) (Vrugt et al., 2009), and Standard PSO 2011 (SPSO-2011) (Clerc, 2012), finding an outstanding performance of hydroPSO in terms of efficiency, effectiveness and scalability for a set of benchmarking functions. On the basis of these results, the hydroPSO was selected as the calibration engine for this study. For a detailed description of hydroPSO the reader is referred to (Zambrano-Bigiarini and Rojas, 2013).

The selection of model parameters to be calibrated is listed in Table 3, including their respective physically-reasonable ranges. The performance of each particle was assessed using a modified version of the Kling-Gupta Efficiency (Gupta et al., 2009) (see Section 3.2.4.).

Comparing the hydrological performance of each SRFE obtained with the BLP (Step 2) against the one obtained after SRFE-specific calibration (Step 3), will provide insight on how important a SRFE-specific calibration is.

(insert Table 3 here)

3.2.3. Bias-correction

To assess the influence of bias correction on the hydrological performance two different bias-correction methods, namely factor correction (FC) and histogram equalization (HE), were tested.

The FC method refers to a rescaling of precipitation based on a multiplier. This multiplier is calculated for each calendar month during the wet season (here: April - October) only if a tendency of either over- or underestimation is prevailing for that respective calendar month. A tendency is prevailing if, when comparing the monthly accumulations of the ground observation with those of the respective SRFE pixel, at least 75% of these values show the same sign of over-or underestimation. In this case, a monthly correction factor ($F_{SRFE,mon}$) is computed for the whole study area as:

$$F_{SRFE,mon} = \frac{\sum_{i=1}^{n} P_{SRFE_i}}{\sum_{i=1}^{n} P_{obs_i}} \qquad \text{eq. (1)}$$

where P_{obs} and P_{SRFE} are, respectively, the monthly precipitation of the ground observation and the corresponding pixel of the SRFE at the location of the ground observation station i, and n is the number of ground stations being considered (only stations with complete data coverage for that respective calendar month are considered). The resulting multiplier ($F_{SRFE,mon}$) is then applied on each daily SRFE map of the particular month.

The HE is a recent bias correction method used to correct precipitation estimates from climate models (Krajewski and Smith, 1991; Piani et al., 2010). The idea behind this method is the derivation of a "transfer function" (TF) that maps the histogram of the SRFE to match the histogram of the observations. This transfer function is calculated for each raster cell of the SRFE and for each calendar month, hence for each raster cell of the SRFE a corresponding observation is required. Two different methods were applied to interpolate the observations: Inverse Distance Weighted (Burrough and McDonnell, 1998) and Kriging with External Drift (Goovaerts, 2000). Depending on which spatial interpolation method was used to grid the observations, the bias-correction method is referred to as HE-IDW or HE-KED.

Comparing FC against HE gives the impact of the bias-correction method on the hydrological performance, while comparing HE-IDW against HE-KED gives the impact of the spatial interpolation method. The aim of this particular analysis is to assess whether using more sophisticated approaches (HE in general, but also specifically HE-KED) provides an improvement of hydrological performance that compensates the computational and human effort required, or if using simpler approaches (FC or HE-IDW) can result in comparable or even better hydrological performances. This analysis will provide insight on what bias-correction method can be pursued, as well as what level of complexity is required to perform the spatial interpolation of the reference field for obtaining acceptable hydrological performances.

3.2.4. Performance indicator

The model performance during calibration was assessed using the modified Kling-Gupta Efficiency (KGE') (Kling et al., 2012). The KGE' is a recent performance indicator based on the equal weighting of three sub-components: linear correlation (r), bias ratio (β) and variability (γ), between simulated (s) and observed (o) discharge. KGE' is defined as follows:

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \qquad \text{eq. (2a)}$$

$$\beta = \frac{\mu_s}{\mu_o} \qquad \text{eq. (2b)}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \qquad \text{eq. (2c)}$$

where r is the Pearson product-moment correlation coefficient, μ is the mean discharge [m³/s], CV is the coefficient of variation and σ is the standard deviation of the discharge [m³/s]. KGE', r, β and γ are dimensionless and their optimum is at unity. The actual value of KGE' gives the lower limit of any of the three sub-components.

According to Kling (personal communication, 2012) the hydrological performance can be classified using KGE' as following:

- good (KGE' ≥ 0.75),
- intermediate (0.75 > KGE' ≥ 0.5),
- poor (0.5 > KGE' > 0.0) and

- very poor (KGE’ ≤ 0.0).

For the actual analysis of the hydrological performance not only the KGE’ value is taken into account, but also its three sub-components (r, β and γ) as they provide an excellent opportunity to elucidate the causes behind a non-optimal model performance. A mismatch of timing and shape of the hydrograph, for example, is reflected by a low value of the linear correlation coefficient (r), while a poor mass balance and a poor variability of daily discharge are expressed by bias (β) and variability (γ) ratios very different from unity, respectively.

Besides the benefit of explicitly discriminating between these three sub-components (r, β and γ), using KGE’ as objective function during calibration has been demonstrated to improve the bias and variability ratio considerably, while the correlation coefficient is only slightly decreased, compared to the often-used Nash Sutcliffe Efficiency (NSE). Moreover, NSE has shown the tendency to underestimate the variability of flows and exhibit less efficiency in constraining the bias ratio (Gupta et al., 2009). For a full discussion of the advantages of using KGE’ over NSE we refer the reader to (Gupta et al., 2009).

3.3. High- and low-flow conditions

While for some hydrological applications, SRFE that result in a moderate to good hydrological performance throughout the year are sufficient, there exist some applications that require a particular high performance for a certain flow condition. Flood forecasting or inundation modelling, for example, require a high accuracy during high-flow period, while other applications such as drought or environmental flow modelling require a high accuracy during low-flow seasons. Therefore, we analyse the hydrological performance during low- and high-flow conditions separately.

The distinction between high- and low-flow season is done through baseflow separation on the observed time series following the automated digital filter approach by (Arnold and Allen, 1999), in which baseflow (b_t) is calculated as:

$$b_t = Q_t - \frac{0.925 \cdot q_{t-1} + 1.925}{2 \cdot (Q_t - Q_{t-1})} \qquad \text{eq. (3)}$$

Where Q_t is the original streamflow, q is the filtered surface runoff and t the time step. In this approach, time periods in which the discharge consists almost exclusively of baseflow correspond to the low-flow season, whereas the remaining time periods correspond to the high-flow season. Once the time periods of the high- and low-flow seasons were identified, they were used to separate the hydrological simulations of Steps 1 and 5 into high-flow and low-flow periods and then analysed individually. The objective of this particular analysis is to get a better understanding of the hydrological performance during different flow conditions as well as to identify which SRFE shows a superior performance during a particular flow condition and topographic feature.

4. Results

4.1. Reference performance

Figure 5 shows the hydrological performance of each SRFE when LISFLOOD is run with its base-line parameters (BLP). These results serve as benchmark in order to estimate the impact of SRFE-specific calibration and bias correction at a later stage.

Considering the classification of hydrological performance described in section 3.2.4, different tendencies are observed for lowland (B, N & S) and mountainous (G) catchments. Over the lowland areas the hydrological performance is quite diverse depending on the SRFE: it is good to intermediate using RFE 2.0 and TRMM-3B42, poor using ERA-Interim, and very poor for

CMORPH and PERSIANN. Over the mountainous catchment, however, almost all SRFE show a poor performance, with only CMORPH being slightly better.

The three components of KGE' (r, β and γ) are useful to identify the source of the performance flaws. Most of the poor and very poor performances observed are due to large deviations of β and γ from their optimum, which indicate a poor agreement in the mass balance and distributional shape, respectively. At the same time, r (representing the temporal dynamic and shape of the hydrograph) is in almost all cases the component comparatively closest to unity. From Figure 5, it appears that the very poor performances of CMORPH and PERSIANN have a large bias ratio (β>2) and most of them a small variability (γ<1), indicating large overestimation of mass balance and less flow variability, respectively. Poor performances, on the contrary, show mostly an underestimation of discharges (β close to 0). These results are in full agreement with the findings about the bias ratio of the raw SRFE shown in (Thiemig et al., 2012).

(insert Figure 5 here)

4.2. The influence of calibration, bias correction and both combined on the hydrological performance

Figure 6 presents the influence on the hydrological performance considering SRFE-specific calibration, bias correction and both combined. We should note that the middle and right-hand columns only show the best performance, irrespectively of the bias-correction method employed, and that the dot size quantifies the change in performance compared to the reference runs (BLP) shown in Figure 5. In other words, the smaller the dot size, the smaller the effect of the particular process (SRFE-specific calibration or bias correction). Hence, in this case, the values (colors) will be similar to the ones of the reference performance.

Results after SRFE-specific calibration (left-hand column) show that the hydrological performance was improved in all the cases, with the best KGE' values for SRFE that initially showed a good performance (RFE 2.0 and TRMM-3B42, Figure 5), and the lowest performance for products with an initially poor or very poor performance. Nevertheless, the absolute improvement is larger for SRFE with an initially poor or very poor performance than for those with an initially good performance. Considering the three sub-components, SRFE-specific calibration produced the largest improvement for the bias ratio (β), while the improvement of the variability of flow (γ) and of the timing and shape of the hydrograph (r) were negligible. However, even though the SRFE-specific calibration reduced the bias ratio largely, it is not capable of correcting the mass balance perfectly (β not approximating unity) for SRFE that are afflicted with large biases in their intrinsic data quality (see β in Figure 5). Products with an initially good or intermediate performance such as RFE 2.0 and TRMM-3B42 (over lowlands), approximate after SRFE-specific calibration their feasible hydrological optimal performance (i.e. KGE' close to 1).

The bias correction (middle column) has, as the SRFE-specific calibration, a large impact on the bias ratio (β) and a rather negligible impact on the variability of flow (γ) as well as on the timing and shape of the hydrograph (r). Consequently, the bias correction improved the hydrological performance in almost all the cases, with the largest improvement for SRFE with an initially poor or very poor hydrological performance, due to their large bias ratio (β diverging largely from 1), namely CMORPH, PERSIANN and ERA-Interim over the lowlands and all SRFE (except CMORPH) over the mountains. On the contrary, the influence of bias correction is negligible for an initially good or intermediate-performing product (e.g. RFE 2.0 and TRMM-3B42 over lowlands), which can clearly be seen by the size and color of the dots of the β component. For some of these products, even though the effect of bias correction is small, the variability (γ) of the hydrological performance is slightly worsened (underestimated) over Nawuni (TRMM-3B42) and

Saboba (RFE 2.0, TRMM-3B42 and ERA-Interim) after bias correction. An in-depth analysis has shown that in all of these cases the HE was used as bias-correction method. Previous research on the HE has shown two issues that might explain the worsening of the variability: first, the underestimation of the lower and higher end of the bias-corrected PDFs of precipitation estimates (Rojas et al., 2011) and (Dosio et al., 2012), and second, the presence of numerical artifacts coming from the derivation of the TF by using Ordinary Least Squares (OLS) fitting (Piani et al., 2010).

Overall, the best hydrological performances are obtained if the hydrological model is calibrated using the bias-corrected SRFE (right-hand column), with all three sub-components (r, β and γ) showing an average close to 1 for all catchments and an average KGE' of 0.87, 0.84, 0.9 and 0.88 for Bui, Nawuni, Saboba and Gambella, respectively. Considering the classification in section 3.2.4, most of the products result into a good hydrological performance after applying the workflow described in Section 3.1.

Considering the impact of SRFE-specific calibration and bias correction of SRFE on the hydrological performance, the evaluation has shown that both improved the initial performance obtained with BLP, mostly by reducing the bias ratio component of the KGE'. However, the impact of bias correction is larger than that of SRFE-specific calibration for initially (Step 2) poor and very poor performing products, and vice versa for initially good and intermediate performing products.

(insert Figure 6 here)

4.3. The most convenient bias-correction approach

Figure 7 shows the hydrological performance for the three different versions of bias-corrected SRFE (FC, HE-IWD, HE-KED) for each sub-catchment. It is worth noting two aspects: 1) the

“NA” indication for RFE 2.0 and TRMM-3B42 are due to the fact that both products are quite close to the observations over the lowland catchments and thus they do not fulfill the prerequisites for the calculation of the correction factor as stated in Section 3.2.3; and 2) the general effect of bias correction on r, β and γ as already discussed in Section 4.2 will not be repeated, unless it contributes to the distinction among the different bias-correction approaches.

Results show that the choice of bias-correction method has a substantial effect for all catchments. Considering the classification of the hydrological performance, it is clear that using a more sophisticated bias-correction method (HE) results into a better hydrological performance in all sub-catchments. This is mostly due to the reduction of the bias ratio (β), which is considerably better reproduced using HE rather than FC. The fact that the variability of flow (γ) shows also a different pattern for FC and HE over the lowlands (B, N & S), with FC overestimating (γ>1) and HE underestimating (γ<1), is not decisive, as none of the methods outperforms the other. The timing and shape of the hydrograph (r) plays a negligible role as it is the same for both bias-correction methods.

The choice of the spatial interpolation field, on the contrary, appears rather subsidiary since the differences in hydrological performance are small. For lowland catchments (B, N & S) this might even be applicable as both interpolation fields lead to similar performances considering KGE’, with the only difference being that HE-IDW shows a tendency to underestimation and HE-KED to overestimation. However, over the mountainous catchment (G) those small differences move within a range that is hydrologically highly relevant, i.e. within the transition from intermediate to good hydrological performance. In this catchment, all four statistical measures (KGE’, r, β, and γ) show higher scores for HE-KED than for HE-IDW. For example, KGE’ is on average 0.85 and 0.77 for HE-KED and HE-IDW, respectively, while β ranges between 0.9-1.07 and 0.84-0.92. This might be partially explained by the fact that using high-resolution terrain elevation information as auxiliary data might improve the spatial interpolation of precipitation in mountainous areas. Hence, we could hypothesize that the more sophisticated the spatial interpolation method, the better the hydrological performance.

In summary and with regard to the bias-correction method, the HE results generally in a superior hydrological performance, while the more sophisticated interpolation algorithm (KED) seems to be of added value only over mountainous regions.

(insert Figure 7 here)

4.4. The performance of SRFE during high-flow and low-flow seasons

Figure 8 shows the hydrological performance during high-flow and low-flow season when both SRFE-specific calibration and bias correction (HE-KED) are used. In general, the hydrological performance is better during the high-flow season than during the low-flow season. During high-flow season almost all SRFE achieve a good hydrological performance over all the sub-catchments, with exception of TRMM-3B42 over the mountainous area (G) and PERSIANN over most of the lowlands (B, N & S), which hold an intermediate performance. The limiting factor of the slightly poorer performances is due to a weak correlation, meaning that the timing and shape of the hydrograph are not properly reproduced, which was clearly visible by inspecting the corresponding hydrographs (not included here). The hydrographs showed diverse tendencies such as over- or underestimation as well as delayed or early onset of the high-flow season depending on the individual year being considered. Regarding TRMM-3B42 this is in full agreement with the findings of (Bitew and Gebremichael, 2011), who observed likewise an inconsistent model performance of TRMM-3B42 over mountainous areas.

The performance during low-flow season is very different over lowland and mountainous catchments; with a mostly good performance over the mountains (G) and poor performance over the lowlands (B, N & S). The poor to very poor performances show deficits in all three sub-components, with the mismatch in variability (γ) between observed and simulated discharge being the most pronounced. Recalling that we are looking here at hydrological simulations that result from using the hydrological model with bias-corrected SRFE, it can be reasonably assumed that the mismatch originates mainly from the standard deviation (σ), as the differences in mean discharge (μ) between simulated and observed discharge are presumably less pronounced after bias-correction. Hence, the mismatch in variability (γ) arises mainly from differences in the deviation from the mean behavior of the hydrograph.

The superior hydrological performance during high-flow compared to low-flow conditions (over lowland catchments), combined with the fact that the hydrological performance obtained using SRFE and interpolated observations are similar to each other, indicates that the hydrological model is better suited to flood forecasting or other applications focused on high-flow conditions for the studied areas. Poor performance during low-flow conditions might be attributed to model deficiencies or an improper performance indicator used during calibration. Analysis of these issues, however, is beyond the scope of this article.

(insert Figure 8 here)

4.5. Validation

Figure 9 shows the hydrological performance during an independent 2-year validation period (2007-2008; 2006 as warm-up) when both SRFE-specific calibration and bias correction (HE-KED) are used. Note that the hydrological performance using KED interpolated observed precipitation is not included due to a lack of observed precipitation data during this time period. The classification of hydrological performance shows a general decline of performance compared to the results achieved during calibration. Product-wise, RFE 2.0 maintains a good performance, CMORPH and PERSIANN decrease slightly to an intermediate performance, while TRMM-

3B42 and ERA-Interim show heterogeneous performances over the different sub-catchments. Considering the sub-catchments individually, no particular tendency can be seen.

The decline in performance is mainly due to deviations of β from its optimum. The reason for this might be related to the length of the time period used for deriving the bias-corrected fields. The bias-correction approach used here (HE) is based on the derivation of a transfer function (TF), which is then assumed to be valid for a different target period. In our case, the TF has been derived for the time period 2003-2006 and then applied to the period 2003-2008. Hence, the time period for the derivation of TF may be too short for computing a reliable transfer function and the validation period has not been used for the construction of the TF and thus might not be perfectly fitting. Hence, our assumption of stationarity, meaning that the TF and its associated parameters are also valid during the application period might not hold, and consequently it would be advisable to use longer time periods for obtaining more robust TFs. However, considering the data availability, this was the best possible approach. Considering TRMM-3B42, another reason might explain the deviation of β from its optimum. Version 6 uses two different gauge analyses namely GPCC and CAMS (at different times) to correct the monthly bias. Using CAMS (since May 2005) has shown to be deficient in some regions and hence might explain the heterogeneous performance.

Periodic updating of the calibration and bias correction when new data become available may compensate the decline in hydrological performance when the hydrological model is used in a time period different from the one used for the calibration and derivation of the transfer function.

(insert Figure 9 here)

5. Discussion and Conclusion

The usefulness of satellite-derived rainfall estimates (SRFE) as forcing data for hydrological applications was investigated here. Four SRFE (CMORPH, RFE 2.0, TRMM-3B42 and PERSIANN) and one re-analysis product (ERA-Interim) were evaluated over two African river basins (Volta and Baro-Akobo), both holding distinct climatic, physiographic and hydrologic characteristics. We aimed at addressing three research questions: *How useful are these SRFE as forcing data for hydrological modelling*? *Which SRFE should be favoured for hydrological modelling? What could researchers do to increase the performance of SRFE-driven hydrological simulations*? Within this context we assessed a) the individual and combined effect of SRFE-specific calibration and bias correction on the hydrological performance, b) the level of complexity required regarding bias-correction and spatial interpolation methods to achieve a good hydrological performance, and c) the performance of SRFE during high- and low-flow conditions.

5.1. Answers to key research questions

How useful are these SRFE as forcing data for hydrological modelling?

Results from the hydrological evaluation make clear that the selected SRFE have a good potential to be used as input data source for hydrological modeling. This is mainly due to two facts: a) most of the SRFE achieve a good hydrological performance over most of the climatic and geomorphologic conditions analysed, when SRFE-specific calibration and/or bias correction are used to (partially) compensate for intrinsic data quality flaws of the SRFE; and b) that the hydrological performance obtained using SRFE and interpolated observations as forcing data for the hydrological model are after calibration similar to each other. This outcome is highly desirable, especially for data-sparse and ungauged basins.

Which SRFE should be favoured for hydrological modelling?

Generally, the SRFE that requires the least effort for a good hydrological performance is the desirable one. Hence, the one that has a good intrinsic data quality and does not need to be bias-corrected prior to model application. However, as the quality of the SRFE is not homogeneous over different climatologic and geomorphologic conditions, there is no straightforward recommendation for a specific SRFE to use. The user has to consider the intrinsic data quality of the SRFE for the specific target area, either through ground truthing or by running the hydrological model with some parameterisation based on expert knowledge (here: BLP), and, finally, select the most accurate from the start. In our case, the selection would be RFE 2.0 and TRMM-3B42 for Volta (lowlands) and CMORPH for Baro-Akobo (mountains). After selection of the SRFE the user can consider further measures to increase the hydrological performance (see next point).

What could researchers do to increase the performance of SRFE-driven hydrological simulations?

From the evaluation of the results, a number of recommendations can be given to increase the hydrological performance. Figure 10, which shows the observed and simulated hydrographs (of Step 1 to Step 5) for CMORPH and RFE 2.0 for the two study areas, provides the context for the following recommendations:

(insert Figure 10 here)

1) Prior to any further measures to improve the hydrological performance, the intrinsic data quality of the selected SRFE needs to be assessed, either through ground truthing or by running the hydrological model with a parameterization based on expert knowledge (here: BLP). In the latter case, a good to intermediate hydrological performance indicates a good intrinsic data quality of the SRFE, while a poor to very poor performance indicates the presence of quality flaws within the SRFE. Figure 10 suggests quality flaws of CMORPH over lowlands and RFE 2.0

over mountainous areas while the data quality seems to be high for CMORPH over mountainous areas and RFE 2.0 over lowlands.

2) If a certain SRFE has a good intrinsic data quality, then only SRFE-specific calibration is recommended. Additional bias correction does not produce a significant improvement to the performance achieved after calibration (Figure 10, panel c).

3) If, on the contrary, a given SRFE shows a quality flaw during ground truthing (usually bias), then applying a bias correction to the SRFE prior to SRFE-specific calibration is essential to obtaining a good hydrological performance (Figure 10, panels a and d).

4) Regarding the selection of the bias-correction method, the more sophisticated approach (histogram equalization) results generally in a superior hydrological performance than when using a simpler method (factor correction). Whereas for the spatial interpolation algorithm, the more sophisticated interpolation (Kriging with External Drift) seems to be of added value only over mountainous regions, as the improvement is within a range that is hydrologically highly relevant and hence justifies the larger workload during the interpolation phase (see Figure 11).

Bias correction should be applied to SRFE that are afflicted with biases for two reasons: first, it appears more sensible to correct the forcing data that produces, due to its data quality flaw, a systematic over- or underestimation of discharge, rather than distorting the calibration parameters beyond commensurability to force the model to reproduce the observed hydrological pattern. Secondly, the hydrological evaluation shows that bias correction reduces the bias more effectively than SRFE-specific calibration.

Applying SRFE-specific calibration is in any case a general recommendation, as it always leads to an improved hydrological performance compared to a hydrological model that has been calibrated to interpolated observations and then forced with SRFE. This has also been shown in previous studies by e.g. (Bitew and Gebremichael, 2011) and (Stisen and Sandholt, 2010).

5.2. Further issues

Previous studies suggest a relationship between the nature of the SRFE (i.e. the main type of data source (IR, PMW) as well as the presence of ground observations) and the quality of the hydrological performance. This study has shown a weaker hydrological performance over the lowland catchments (B, N & S) for those SRFE that do not ingest any ground observations (CMORPH and PERSIANN), which is in agreement with the findings of (Behrangi et al., 2011). Knowing that only a minor percentage of the ground observations used for the hydrological evaluation are publically available (ca. 21 %; see Section 2.2.1), and hence used by RFE 2.0 and TRMM-3B42 to adjust their estimates quantitatively, we could argue that a small number of ground observations might have the potential to improve the intrinsic data quality of the SRFE substantially, and as a result also improve the hydrological performance. The fact that the same tendency is not shown over the mountainous catchment (G) might suggest that the available data density might not be sufficient, considering the complex topography, and hence do not favour SRFE that incorporate ground observations. For those areas, SRFE that ingest primarily PMW (here CMORPH) show a consistent and better performance, which was also observed by (Bitew and Gebremichael, 2011), and suggest that over complex topographies the higher accuracy of the PMW is more important than the high spatio-temporal resolution of the IR.

Lastly, the choice of the performance indicator (KGE') might be questioned due to the fact that it has not yet been widely applied in hydrology, hence it might be unfavorable to the audience that cannot draw upon past experiences. However, due to its powerful nature, originating from an equal, independent and simultaneous consideration of bias ratio, variability ratio and linear correlation, KGE' has an enormous potential over more conventional indicators (e.g. Nash-Sutcliffe Efficiency or RMSE). Calibrating upon KGE' resembles a multi-objective calibration, optimising simultaneously several attributes of the hydrological performance. Furthermore, using KGE' during the evaluation phase provides valuable insight into the hydrological performance. Knowing the origin of the performance flaws gives the opportunity to address those separately to further increase the performance.

5.3. Final implications

As a result of the hydrological evaluation, SRFE showed significant potential as forcing data to hydrological applications focusing on high-flow conditions (such as dam storage capacity calculations or flood management) for the areas under study. SRFE are also useful for general water budget calculations and similar applications, as the general performance was good. However, for our study areas, the use of SRFE are not advisable for hydrological applications focusing solely on the reproduction of low-flow conditions, as the hydrological performance for these conditions was poor. This, however, does not preclude that SRFE could be used for meteorological drought monitoring.

The cause for the poor hydrological performance during low-flow conditions is not entirely clear, as the sub-components of KGE' indicate various flaws. However, the fact that the hydrological simulations driven by interpolated ground observations show similar flaws suggests that the hydrological model might not be capable of reproducing low-flow conditions. The latter could be explained by a poor model structure or by the performance indicator chosen during calibration, which optimises the hydrological simulation from a number of perspectives, but with no particular emphasis on the low-flow spectrum. Given this uncertainty, it is presently not possible to indicate a potential applicability of SRFE for applications focusing on low-flow conditions, although this is open for future research. One approach might be to repeat the calibration with a performance indicator that concentrates predominantly on the low-flow conditions such as the Heteroscedastic Maximum Likelihood Error (HMLE) (Sorooshian and Dracup, 1980).

Acronyms

AMSR-E	Advanced Microwave Scanning Radiometer
AMSU	Advanced Microwave Sounding Unit
BLP	"base-line parameterisation"
β	bias ratio
CAMS	Climate Anomaly Monitoring System
FC	factor correction
GPCP	Global Precipitation Climatology Project
GTS	Global Telecommunication System
HE	histogram equalization
HMLE	Heteroscedastic Maximum Likelihood Error
IDW	inverse distance weighted
γ	variability ratio
KED	Kriging with External Drift
KGE'	modified Kling-Gupta Efficiency
MAE	Mean Absolute Error
NSE	Nash Sutcliffe Efficiency
OLS	Ordinary Least Squares
PDF	probability density function
PMW	passive microwave
PSO	Particle Swarm Optimisation
R	linear correlation

R^2	coefficient of determination
RMSE	Root Mean Square Error
SCE-UA	Shuffled Complex Evolution Algorithm
SRFE	satellite-based rainfall estimates
SRTM	Shuttle Radar Topography Mission
SSM/I	Spatial Sensor Microwave/ Imager on board
TCI	TRMM Combined Instrument
TIR	thermal infrared
TF	“transfer function”
TMI	Advanced Microwave Sounding Radiometer on board the TRMM spacecraft
TRMM	Tropical Rainfall Measuring Mission
VarBC	variational bias correction

Acknowledgements

First of all, we would like to express our sincere gratitude to the meteorological and hydrological institutions for providing us with the required hydrological and meteorological ground measurements that enabled this study. These include the national members of the GLOWA Volta project, the Ethiopian National Meteorological Department as well as the Ethiopian Ministry of Water and Energy. Special acknowledgment also goes to NASA, ECMWF, the University of Arizona, and NOAA/CPC for providing easy access to the SRFEs and their assistance in particular cases. We would also like to thank William Becker for proofreading this article, as well as Alessandra Bianchi for her technical support.

References

Adler, R.F. et al., 2003. The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). Journal of Hydrometeorology, 4: 1147-1167.

Ali, A., Amani, A., Diedhiou, A. and Lebel, T., 2005. Rainfall estimation in the Sahel. Part II: Evaluation of rain gauge networks in the CILSS countries and objective intercomparison of rainfall products. Journal of Applied Meteorology, 44(11): 1707-1722.

Arnold, J.G. and Allen, P.M., 1999. AUTOMATED METHODS FOR ESTIMATING BASEFLOW AND GROUND WATER RECHARGE FROM STREAMFLOW RECORDS1. JAWRA Journal of the American Water Resources Association, 35(2): 411-424.

Artan, G. et al., 2007. Adequacy of satellite derived rainfall data for stream flow modelling. *Natural Hazards and Earth System Sciences*, 43(2): 167-185.

Asadullah, A., McIntyre, N. and Kigobe, M., 2008. Evaluation of five satellite products for estimation of rainfall over Uganda. Hydrological Sciences Journal, 53(6): 1137-1150.

Behrangi, A. et al., 2011. Hydrologic evaluation of satellite precipitation products over a mid-size basin. Journal of Hydrology, 397(3-4): 225-237.

Bitew, M.M. and Gebremichael, M., 2011. Assessment of high-resolution satellite rainfall for streamflow simulation in medium watersheds of the East African highlands. Hydrology and Earth System Sciences Discussions, 7(5): 8213-8232.

Burrough, P.A. and McDonnell, R.A., 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford, 333 pp.

Clerc, M., 2012. Standard Particle Swarm Optimisation.

Cloke, H.L. and Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology, 375(3-4): 613-626.

de Condappa, D., ChaponniÃ¨re, A. and Lemoalle, J., 2009. A decision-support tool for water allocation in the Volta Basin. Water International, 34(1): 71-87.

Dee, D.P. et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656): 553-597.

Dinku, T., Ceccato, P., Cressman, K. and Connor, S.J., 2010. Evaluating detection skills of satellite rainfall estimates over desert locust recession regions. Journal of Applied Meteorology and Climatology, 49: 1322-1332.

Dinku, T. et al., 2007. Validation of satellite rainfall products over East Africa's complex topography. International Journal of Remote Sensing, 28(7): 1503-1526.

Diro, G.T., Grimes, D.I.F., Black, E., O'Neill, A. and Pardo-Iguzquiza, E., 2009. Evaluation of reanalysis rainfall estimates over Ethiopia. International Journal of Climatology, 29(1): 67-78.

Döll, P., Kaspar, F. and Lehner, B., 2003. A global hydrological model for deriving water availability indicators: Model tuning and validation. Journal of Hydrology, 270(1-2): 105-134.

Dosio, A., Paruolo, P. and Rojas, R., 2012. Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: Analysis of the

climate change signal. Journal of Geophysical Research D: Atmospheres, 117(17).
Duan, Q.Y., Gupta, V.K. and Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. Journal of Optimization Theory and Applications, 76(3): 501-521.
Dyson, M., Bergkamp, G. and Scanlon, J., (eds), 2008. Flow - The essentials of environmental flows, International Union for Conservation of Nature and Natural Resources, Gland, Switzerland.
Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology, 228(1-2): 113-129.
Gourley, J.J. et al., 2011. Hydrologic evaluation of rainfall estimates from radar, satellite, gauge, and combinations on Ft. Cobb basin, Oklahoma. Journal of Hydrometeorology, 12(5): 973-988.
Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2): 80-91.
Hsu, K.L., Gao, X., Sorooshian, S. and Gupta, H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. Journal of Applied Meteorology, 36(9): 1176-1190.
Huffman, G.J. et al., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of Hydrometeorology, 8(1): 38-55.
Huffman, G.J., Adler, R.F., Bolvin, D.T. and Nelkin, E., 2010. The TRMM Multi-Satellite Precipitation Analysis. In: F. Hossian and M. Gebremichael (Editors), Satellite Rainfall Applications for Surface Hydrology. Springer Publishing Company, London, pp. 3-22.
Hughes, D.A., 2006. Comparison of satellite rainfall data with observations from gauging station networks. Journal of Hydrology, 327(3-4): 399-410.
Jiang, S. et al., 2012. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. Journal of Hydrology, 452-453: 213-225.
Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology, 5(3): 487-503.
Kennedy, J. and Eberhart, R., 1995. Particle swarm optimization, Neural Networks, 1995. Proceedings., IEEE International Conference on, pp. 1942-1948 vol.4.
Kling, H., Fuchs, M. and Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 424-425: 264-277.
Krajewski, W.F. and Smith, J.A., 1991. On the estimation of climatological Z-R relationships. Journal of Applied Meteorology, 30(10): 1436-1445.
Laurent, H., Jobard, I. and Toma, A., 1998. Validation of satellite and ground-based estimates of precipitation over the Sahel. Atmospheric Research, 47-48: 651-670.

McCollum, J.R., Gruber, A. and Ba, M.B., 2000. Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa. Journal of Applied Meteorology, 39(5): 666-679.

Nicholson, S.E. et al., 2003. Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part I: Validation of GPCC rainfall product and Pre-TRMM satellite and blended products. Journal of Applied Meteorology, 42(10): 1337-1354.

Piani, C. et al., 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. Journal of Hydrology, 395(3-4): 199-215.

Poli, R., 2008. Analysis of the Publications on the Applications of Particle Swarm Optimisation.

Rojas, R., Feyen, L., Dosio, A. and Bavera, D., 2011. Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations. Hydrology and Earth System Sciences, 15(8): 2599-2620.

Romilly, T.G. and Gebremichael, M., 2011. Evaluation of satellite rainfall estimates over Ethiopian river basins. Hydrol. Earth Syst. Sci. (HESS), 15(5): 1607-7938.

Shahin, D.M., 2002. Hydrology and Water Resources of Africa, 41. Springer/Book Series: Water Science and Technology Library, Kluwer Academic Publishers, 686 pp.

Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environmental Modelling and Software, 21(9): 1340-1352.

Soboll, A. et al., 2011. Integrated regional modelling and scenario development to evaluate future water demand under global change conditions. Mitigation and Adaptation Strategies for Global Change, 16(4): 477-498.

Sorooshian, S. and Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resources Research, 16(2): 430-442.

Stisen, S. and Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. Hydrological Processes, 24(7): 879-891.

Symeonakis, E., Bonifaçio, R. and Drake, N., 2009. A comparison of rainfall estimation techniques for sub-Saharan Africa. International Journal of Applied Earth Observation and Geoinformation, 11(1): 15-26.

The NOAA Climate Prediction Center, N., 2002. African Rainfall Estimation Algorithm - Version 2.0.

Thiemig, V. et al., 2010. Ensemble flood forecasting in Africa: a feasibility study in the Juba-Shabelle river basin. Atmospheric Science Letters, 11(2): 123-131.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V. and De Roo, A., 2012. Validation of Satellite-Based Precipitation Products Over Sparsely-Gauged African River Basins. Journal of Hydrometeorology, 13(6): 1760-1783.

Thorne, V., Coakeley, P., Grimes, D. and Dugdale, G., 2001. Comparison of TAMSAT and CPC rainfall estimates with raingauges, for Southern Africa. International Journal of Remote Sensing, 22(10): 1951-1974.

Van Der Knijff, J.M., Younis, J. and De Roo, A.P.J., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. International Journal of Geographical Information Science, 24(2): 189-212.
Vrugt, J.A. et al., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation, 10(3): 273-290.
Xie, P. and Arkin, P.A., 1995. An Intercomparison of Gauge Observations and Satellite Estimates of Monthly Precipitation. Journal of Applied Meteorology, 34(5): 1143-1160.
Zambrano-Bigiarini, M., 2011. hydroTSM: Time series management, analysis and interpolation for hydrological modelling. , R package, Ispra.
Zambrano-Bigiarini, M. and Rojas, R., 2012. hydroPSO: Model-Independent Particle Swarm Optimisation for Environmental Models, R package, Ispra.
Zambrano-Bigiarini, M. and Rojas, R., 2013. hydroPSO: A model-independent Particle Swarm Optimisation software for model calibration. Environmental Modelling and Software, [accepted for publication].

Table 1: Topographic and climatic characteristics of the study areas

river system	Volta			Nile
topographic information				
sub-catchment	Black Volta	White Volta	Oti	Baro-Akobo
reference gauging station	Bui	Nawuni	Saboba	Gambella
data provider	Geoportal of Volta Basin Authority			Ethiopian Ministry of Water and Energy
drainage area [km²]	130 000	100 000	53 000	76 000
altitude (min / max / average) [m ASL]	60 / 762 / 287	60 / 530 / 270	40 / 920 / 245	400 / 3100 / 1677
average slope [°], [m/km]	0.7, 12	0.6, 11	0.9, 16	4.4, 77
climatic information				
mean annual precipitation [mm]	1033	964	1155	2009
no. of meteorological stations	68			14
station density [km²/ station]	8800			12 700
data provider	Geoportal of Volta Basin Authority			Ethiopian National Meteorology Agency

1 **Table 2:** Description of the selected SRFE including outcome of the ground truthing phase (Thiemig et al., 2012)

	CMORPH	RFE 2.0	TRMM 3B42 v6	PERSIANN	ERA-Interim
Provider	NOAA-CPC	NOAA-CPC	NASA	University of California, Irvine	ECMWF
Spatial coverage	60°N to 60° S, globally	40° N - 40° S, 20° W - 55°E	50°N to 50° S, globally	60°N to 60° S, globally	Global
Temporal coverage	Since 06.12.2002	Since 01.01.2001	Since 01.01.1998	Since 01.03.2000	Since 01.01.1989
Spatial resolution	0.25°	0.1°	0.25°	0.25°	~ 79 km
Temporal resolution	3 h	24 h	3 h	6 h	6 h
main product data sources	Geostationary IR, SSM/I, AMSU, AMSR-E, TMI	Geostationary IR, SSM-I, AMSU-B and GTS stations	Geostationary IR, TCI, SSM/I, AMSU, CAMS and GPCP	Geostationary IR, TRMM 2A12, SSM/I and AMSU	4D-Var, VarBC
merging approach	Precipitation estimates are solely based on MW data. IR data are only used to derive a cloud motion field to propagate precipitation in higher spatial and temporal resolution.	Precipitation is firstly approximated from each individual satellite source using the ML method, decreasing data gaps, random errors and systematic bias. The quantity of this approximation is then adjusted using GTS interpolated rainfall fields.	MW-based estimations are merged and calibrated, and subsequently combined with IR-based estimates. The combined approximation is then rescaled using monthly CAMS and GPCP data.	A relationship between IR and precipitation rate is established using a neural network. The network is additionally trained with MV data. The actual precipitation estimates are solely based on instantaneous IR observations.	Precipitation is estimated by a numerical model based on temperature and humidity information derived from assimilated observations originating from PMV data and in-situ measurements.
reference	(Joyce et al., 2004)	(The NOAA Climate Prediction Center, 2002)	(Huffman et al., 2007; Huffman et al., 2010)	(Hsu et al., 1997)	(Dee et al., 2011)
main outcome of ground truthing phase for the selected study areas	• Overestimate the amount of precipitation during wet periods	• Capture the intraseasonal variability, the spatial distribution	• Capture the intraseasonal variability, the spatial distribution	• Large quantitative deviations of monthly and annual values	• Persistent overestimation of light rainfall events and

(Thiemig et al., 2012) (see also Figure 3)	as well as the number of rainy days per year (Volta) • Superior ability to reproduce daily, monthly and annual precipitation over mountainous areas (Baro-Akobo)	pattern, the average annual precipitation and the timing of the highest annual precipitation event well (Volta) • Underestimation of precipitation over mountainous areas (Baro-Akobo)	pattern, the average annual precipitation and the timing of the highest annual precipitation event well (Volta) • Underestimation of precipitation over mountainous areas (Baro-Akobo)	• Large overestimations of precipitation amount and number of rainy days; mostly during wet season (Volta) • Underestimation of precipitation over mountainous areas (Baro-Akobo)	underestimation of heavy rainfall events • Capture intraseasonal variability and spatial distribution pattern well (Volta) • Clear overestimation of precipitation over mountainous areas (Baro-Akobo)

AMSR-E: advanced microwave scanning radiometer; AMSU: advanced microwave sounding unit; CAMS: climate anomaly monitoring system; GPCP: global precipitation climatology project; GTS: Global Telecommunication System; SSM/I: spatial sensor microwave/ imager on board; TCI: TRMM combined instrument; TMI: advanced microwave sounding radiometer on board the TRMM spacecraft; TRMM: tropical rainfall measuring mission; VarBC: variational bias-correction

Table 3: Calibrated parameters in the LISFLOOD hydrological model

Parameter	Description	Unit	Min	Max
UZTC	Time constant for water in upper zone	days	3	40
LZTC	Time constant for water in lower zone	days	50	2500
GwPV	Groundwater percolation value	mm/day	0.5	2
GWLoss	Maximum loss rate out of Lower response box, expressed as a fraction of lower zone outflow.	-	0.01	0.35
b_Xinan	Power in Xinanjiang distribution function	-	0.01	1
PPrefFlow	Power that controls increase of proportion of preferential flow with increased soil moisture storage	-	0.5	8
CCM	Multiplier applied to Channel Manning's n	-	0.1	15
CCM2	Multiplier applied to Channel Manning's n for second line of routing	-	0.1	15
CalEvap	Multiplier applied to potential evapo(transpi)ration rates	-	0.5	2

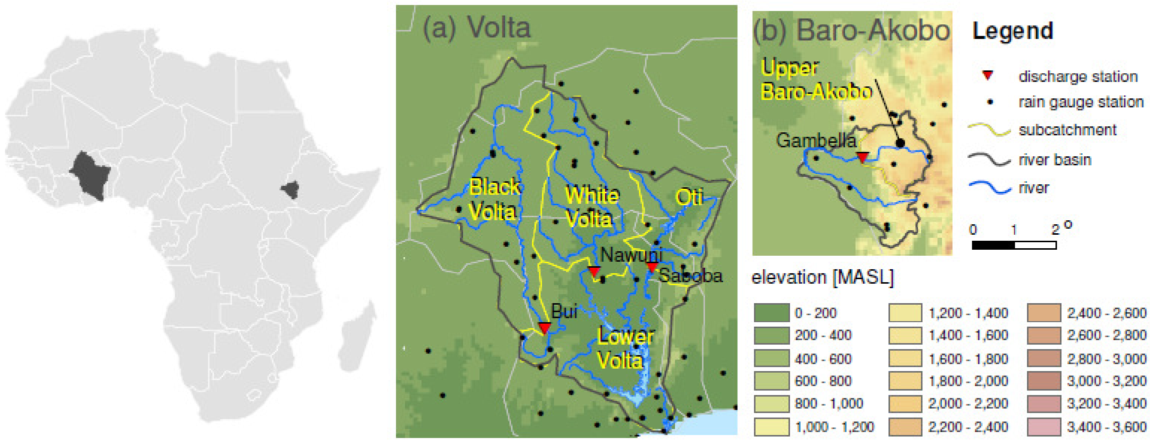

Figure 1: Overview of the geographical location, including the terrain elevation, rain gauge and discharge stations as well as subcatchment delineation for the study areas.

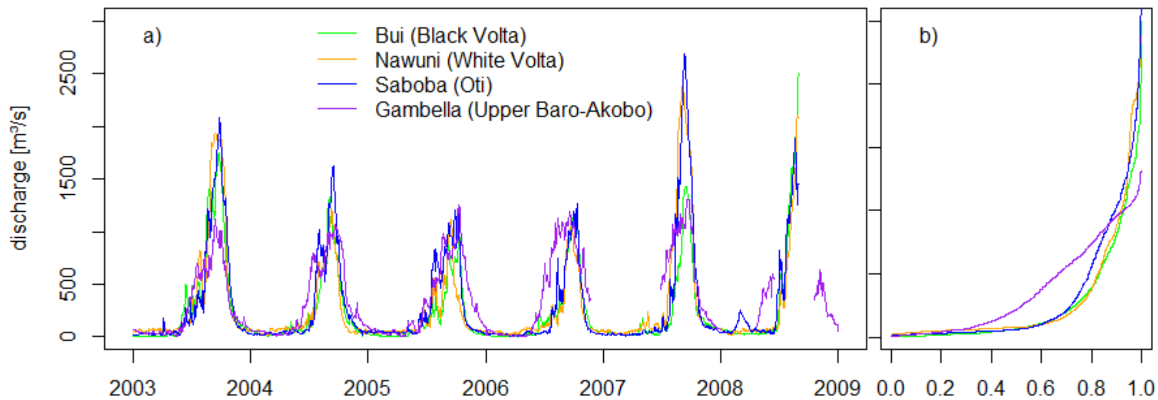

Figure 2: a) observed streamflow and b) quantiles at the reference gauging stations (for location see Figure 1)

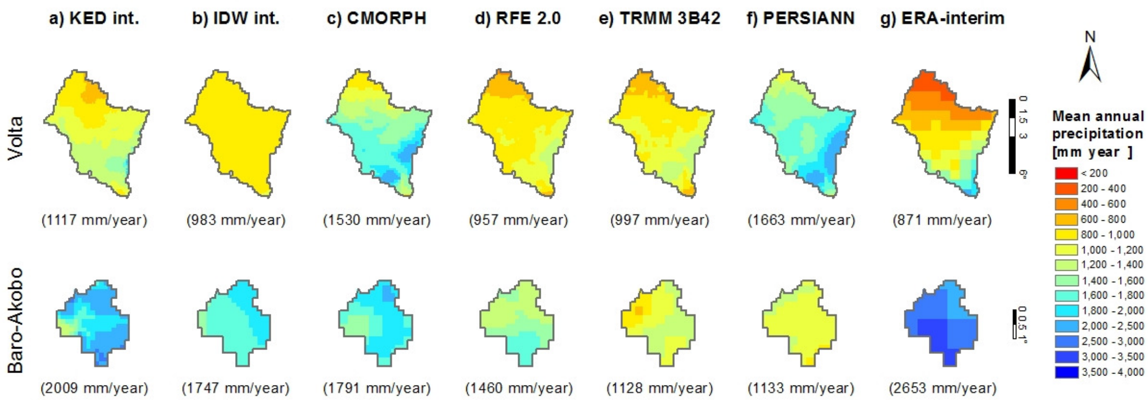

Figure 3: Observed and SRFE-based mean annual precipitation for the reference period 2003-2006. (note both basins are shown on a different spatial scale)

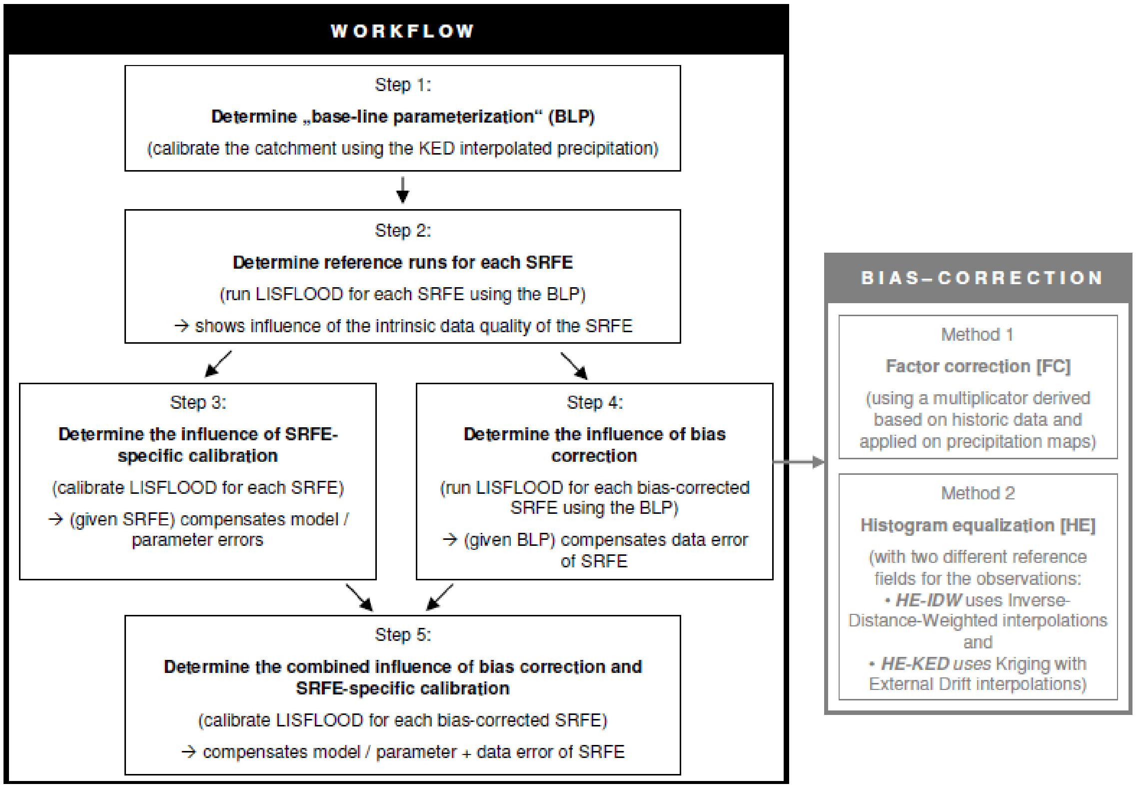

Figure 4: Methodological framework

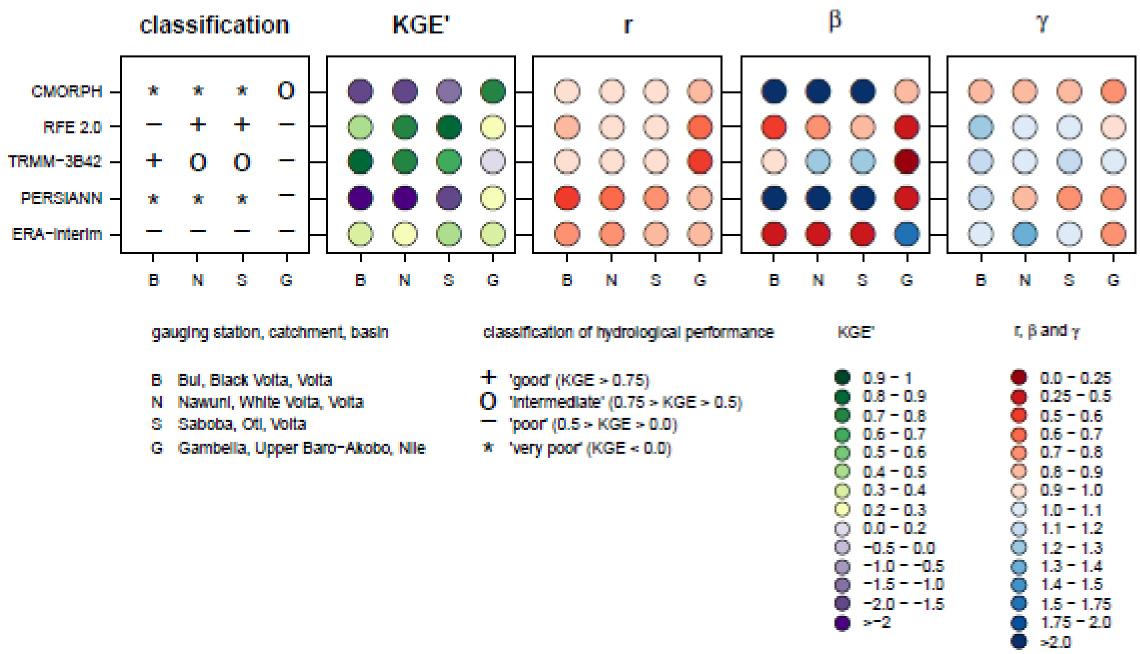

Figure 5: Reference hydrological performance for each SRFE retrieved by running LISFLOOD with BLP for different catchments

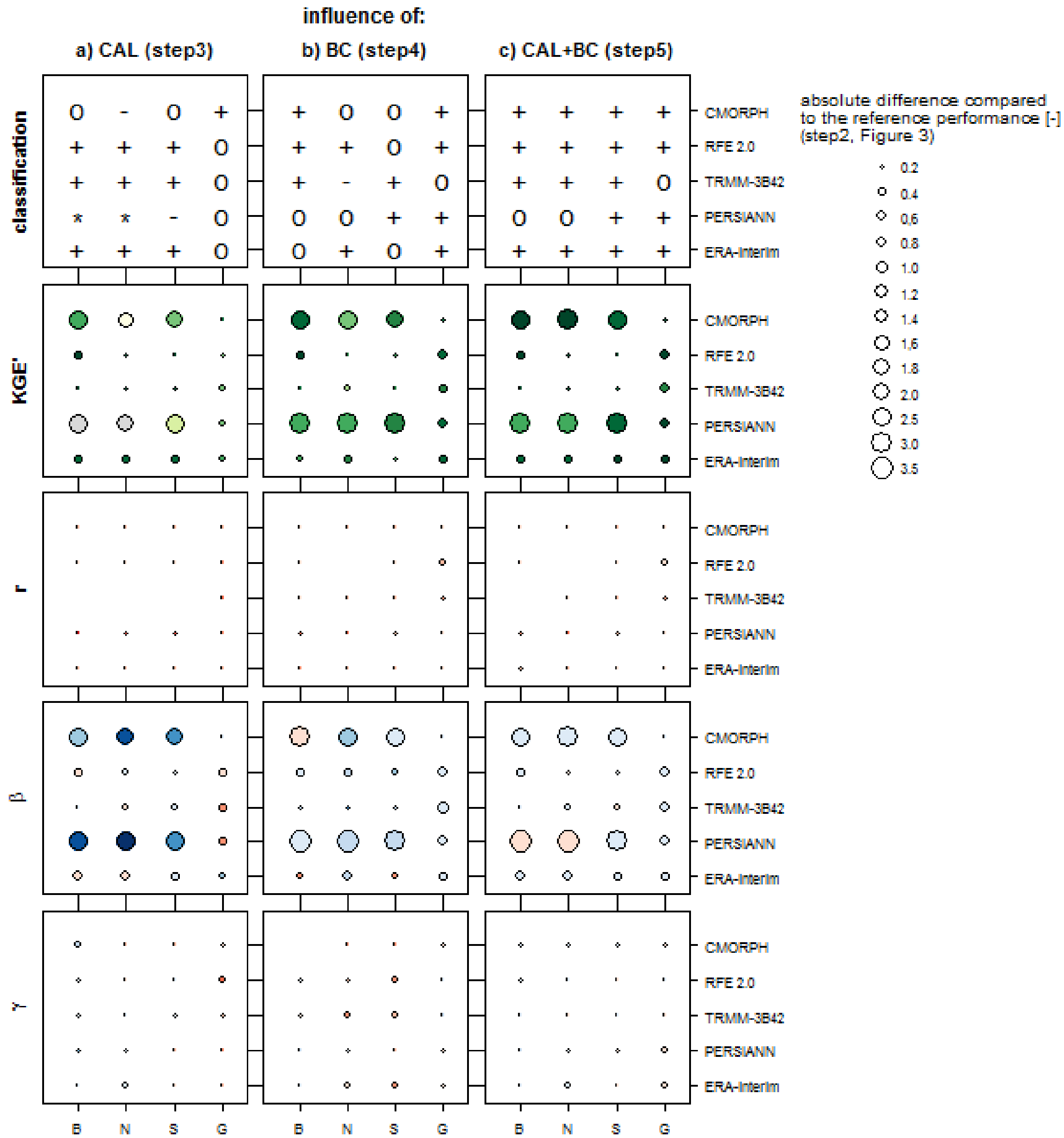

Figure 6: Impact of SRFE-specific calibration (left column), bias correction (middle column) and both combined (right column) on the hydrological performance [same legend as in **Figure 5**].

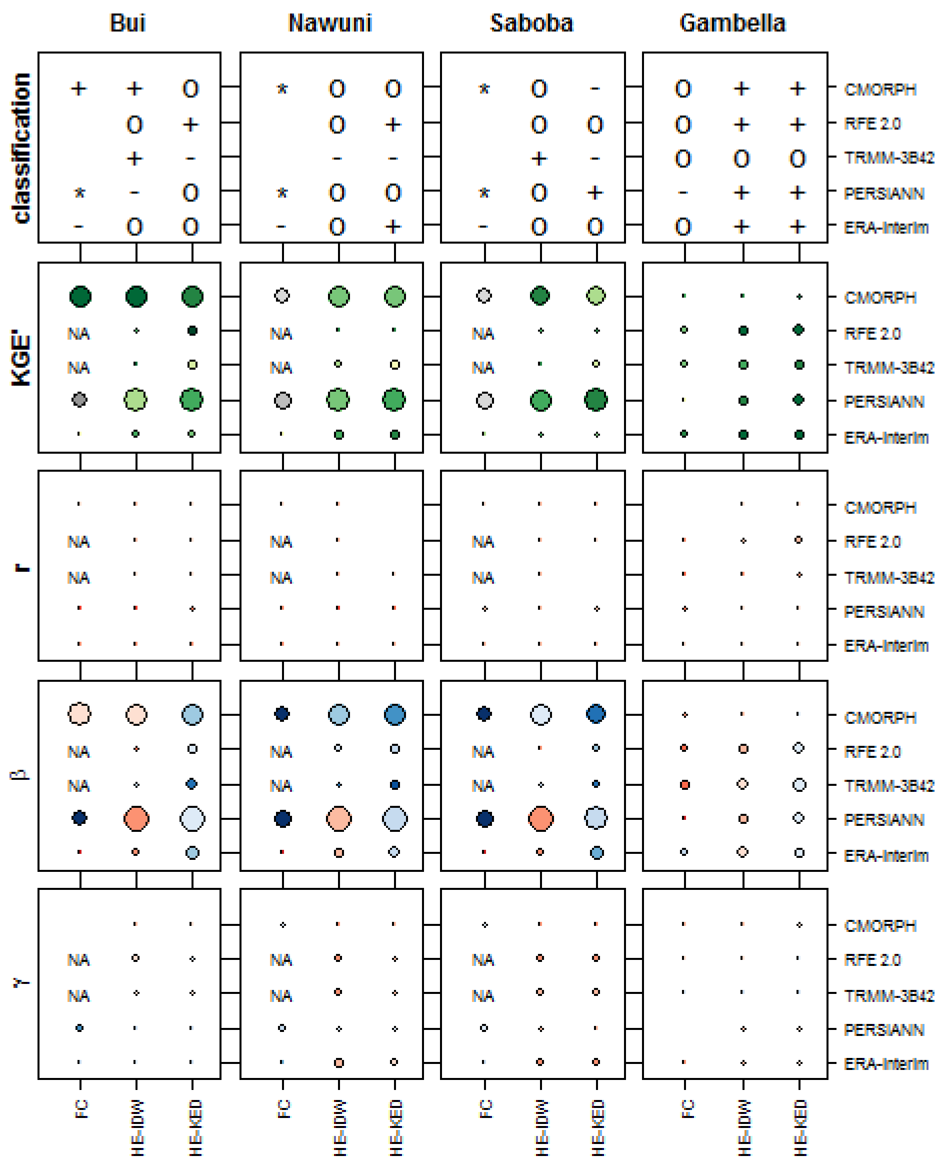

Figure 7: Impact of different bias-correction methods (FC and HE) and different interpolation methods (IDW and KED) ingested by HE on the hydrological performance [same legend as in **Figure 5** and **Figure 6**].

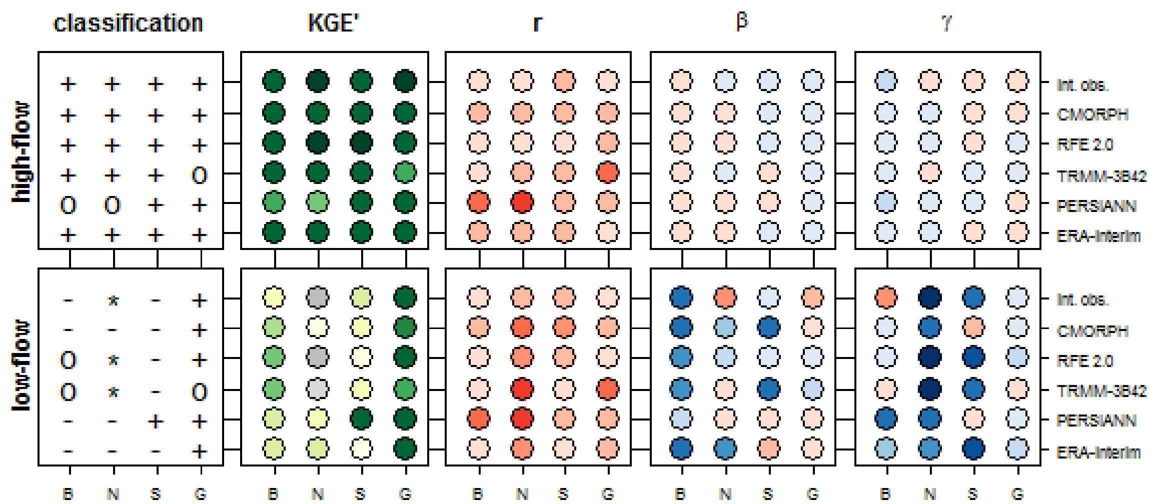

Figure 8: Hydrological performance during high-flow and low-flow season [same legend as in **Figure 5**]

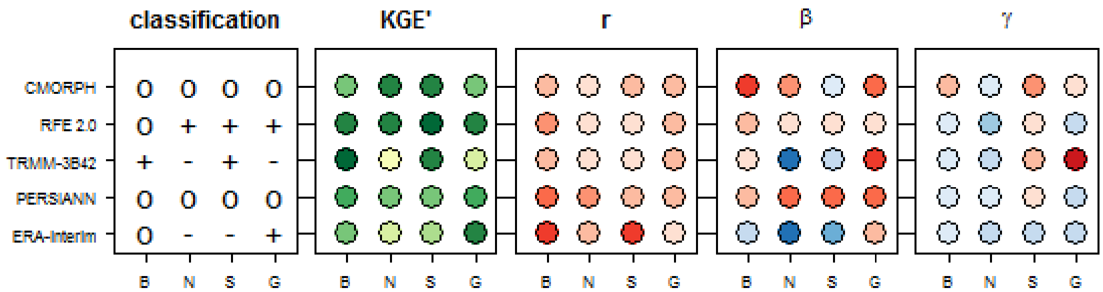

Figure 9: Hydrological performance during the validation period 2007-2008 [same legend as in **Figure 5**]

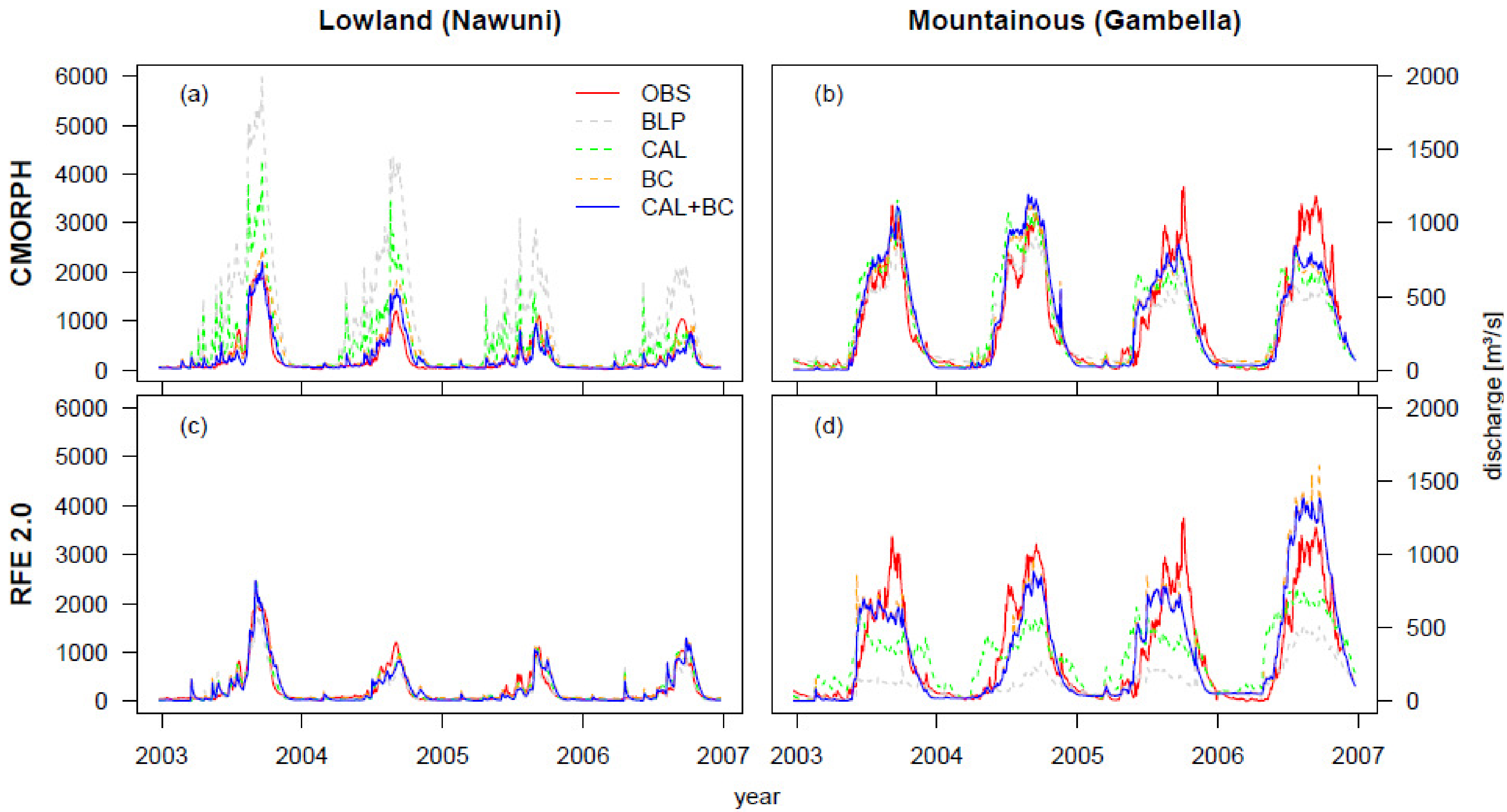

Figure 10: Observed and simulated hydrographs (BLP, CAL, BC, CAL + BC) of CMORPH and RFE 2.0 for a lowland and mountainous catchment during the calibration period 2003-2006.

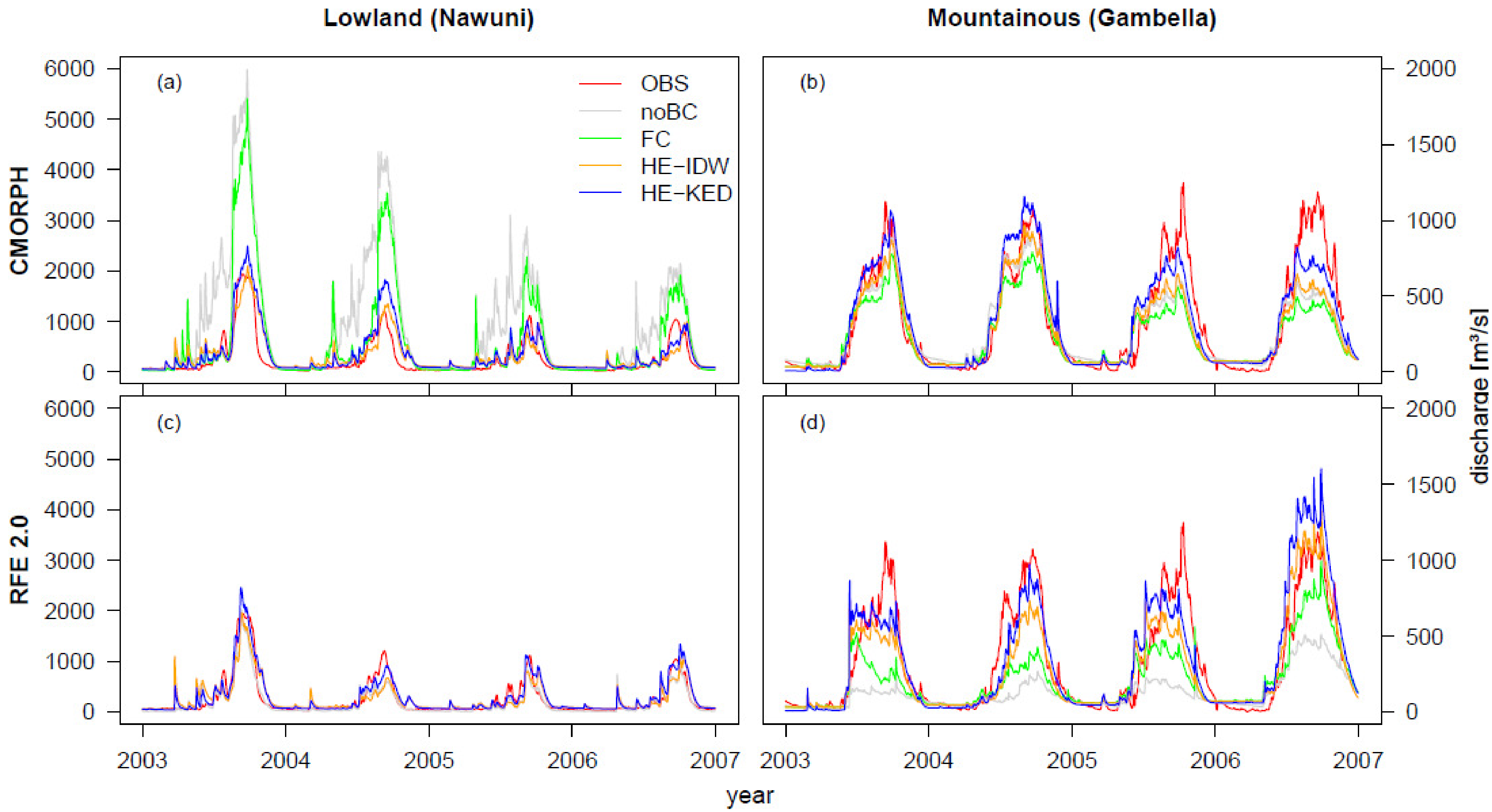

Figure 11: Observed and simulated hydrographs (noBC, FC, HE-IDW and HE-KED) of CMORPH and RFE 2.0 for a lowland and mountainous catchment during the calibration period 2003-2006.