Journal of Hydrology: Regional Studies 8 (2016) 112–129

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions

B. Bisselink[a,*], M. Zambrano-Bigiarini[b], P. Burek[c], A. de Roo[a]

[a] European Commission, Joint Research Centre (JRC), Directorate for Sustainable Resources, Italy
[b] Faculty of Engineering and Sciences, Universidad de La Frontera, Temuco, Chile
[c] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 1 September 2016
Accepted 2 September 2016

Keywords:
Satellite-based rainfall estimates
Highly variable climate conditions
Differential split-sample
Calibration
Model parameter robustness
Hydrological modelling
Southern Africa

ABSTRACT

Study region: Four headwaters in Southern Africa.
Study focus: The streamflow regimes in Southern Africa are amongst the most variable in the world. The corresponding differences in streamflow bias and variability allowed us to analyze the behavior and robustness of the LISFLOOD hydrological model parameters. A differential split-sample test is used for calibration using seven satellite-based rainfall estimates, in order to assess the robustness of model parameters. Robust model parameters are of high importance when they have to be transferred both in time and space. For calibration, the modified Kling-Gupta statistic was used, which allowed us to differentiate the contribution of the correlation, bias and variability between the simulated and observed streamflow.
New hydrological insights: Results indicate large discrepancies in terms of the linear correlation (r), bias (β) and variability (γ) between the observed and simulated streamflows when using different precipitation estimates as model input. The best model performance was obtained with products which ingest gauge data for bias correction. However, catchment behavior was difficult to be captured using a single parameter set and to obtain a single robust parameter set for each catchment, which indicate that transposing model parameters should be carried out with caution. Model parameters depend on the precipitation characteristics of the calibration period and should therefore only be used in target periods with similar precipitation characteristics (wet/dry).

1. Introduction

Hydrological models are widely used for water resources modelling, both drought and flood forecasting, and climate change impact assessment studies, among others. Before applying these models their robustness needs to be tested vis-à-vis with the specific modelling objective to build model credibility and ensure model applicability (Klemeš, 1986). Operational models often need to be calibrated to obtain numerical values of model parameters. The aim of a calibration process is to obtain parameters which allow an acceptable representation of the hydrological behavior of the selected catchment, and moreover to obtain parameters which are robust and, therefore, be transposable towards other time periods as well. This

* Corresponding author.
E-mail address: bernard.bisselink@jrc.ec.europa.eu (B. Bisselink).

http://dx.doi.org/10.1016/j.ejrh.2016.09.003

assumption might only be valid if the uncertainty in the obtained model parameters is low and/or the conditions between the calibration and validation period are similar (stationary conditions). However, there are multiple reasons that might lead to changes in model parameters in time and, therefore, raise a lack of model robustness. The most obvious cause might be an inappropriate model structure (Butts et al., 2004; Bulygina and Gupta, 2009; Reusser and Zehe, 2011; Lin and Beck, 2012; Seiller et al., 2012). Recently, Coron et al. (2014) showed the inability of three models of increasing complexity in reproducing the water balance on different sub-periods. Another explanation for the lack of model robustness can be miscalibration (i.e. poor optimization algorithm) or overcalibration (i.e. insufficient calibration period, too many parameters, wrong objective function) of model parameters, as shown by Wagener et al. (2003), Hartmann and Bárdossy (2005), Son and Sivapalan (2007), Gupta et al. (2009), Ebtehaj et al. (2010), Efstratiadis and Koutsoyiannis (2010), Andréassian et al. (2012), Gharari et al. (2013) and Zhan et al. (2013). In addition, changes in time of some catchment features (e.g., land use change and management, operational rules of reservoirs, changes in groundwater level) are reflected in the model input data, and might also lead to lack of model robustness. For example, Fenicia et al. (2009) showed the major role of changes in land use management and forest age on the catchments' behavior.

To assess the model's robustness under highly variable climate conditions the standard split-sample test, used to calibrate the model in one period and test the model in another period, is not sufficient enough. Klemeš (1986) proposed a more powerful test, the so called *differential split-sample test*, where calibration and validation periods are chosen to represent markedly different hydro-meteorological conditions of the catchment. This differential split-sample test should be applied whenever a model is to be used to simulate flows in a basin under conditions different from those corresponding to the available flow record (Klemeš, 1986). A robust model should demonstrate its ability to perform equally well in the selected calibration and validation periods. Studies that performed a differential split-sample test are relatively scarce, because most models fail this test (Seibert, 2003). The studies of Refsgaard and Knudsen (1996), Donnelly-Makowecki and Moore (1999), Xu (1999), Seibert (2003), Wilby (2005) and Chiew et al. (2009) all applied a differential split-sample test. Most of these studies found a decrease in model performance due to the sensitivity of the model parameters in relation to different climate conditions. More recently, Merz et al. (2011) found in a test for 273 Austrian catchments that the parameters controlling snow and soil moisture were strongly influenced by climatic conditions. Vaze et al. (2010) and Coron et al. (2012) conducted studies with four and three hydrological models, respectively, on southeastern Australian catchments. They also found a strong climate influence in their models. According to Li et al. (2012) dry periods contain more information for model calibration compared to wet periods, when they investigated the transposability of model parameters for dry and wet conditions.

For successful streamflow predictions the model should be forced with accurate precipitation data (Beven, 2004). The impact of precipitation input on model performance is well documented in error analyses (Kavetski et al., 2003, 2006), as a function of catchment size (Moulin et al., 2009), raingauge density (Bárdossy and Das, 2008) or using various geostatistical methods (Sun et al., 2000). However, model robustness problems due to incorrect estimations of precipitation amounts are rarely reported in hydrological modelling, while it is well known that such errors might have a significant effect on the final values of model parameters (Oudin et al., 2006).

Considering the importance of the precipitation input on the reliability of model predictions, it is extremely challenging to perform reliable applications of hydrological models in ungauged or data-scarse areas. For Africa, "ground truth" precipitation is very sparse and, therefore remote sensing can be an ideal technique for obtaining time series of precipitation to be used as input data for hydrological modelling studies. Applications of satellite-based rainfall estimates (SRFE) for hydrological modeling are well documented (for e.g., Thiemig et al., 2013; Artan et al., 2007; Behrangi et al., 2011; Gourley et al., 2011; Stisen and Sandholt, 2010; Cohen Liechti et al., 2012), observing large differences in parameter values obtained from different rainfall inputs (Bitew and Gebremichael, 2011). However, most of these studies perform the standard split-sample test and do not discuss the robustness of the obtained model parameters and how the model structure compensate for the precipitation inaccuracy, and moreover if they are transposable to time periods other than the single validation period.

The aim of this study is to determine the robustness of the fully distributed LISFLOOD hydrological model by using different precipitation estimates as model input. To achieve this aim, this research focuses on five main research questions: (i) How accurate are the different precipitation data sets for streamflow simulations? (ii) What is the effect of uncertain input data (precipitation) on the estimates of model parameters? (iii) How will the model parameters obtained by calibration compensate for precipitation inaccuracy? (iv) Can a different source of precipitation overcome robustness problems? (v) Is a single calibration parameter set sufficient for hydrological forecasting or climate scenario modelling? These research questions are answered performing a differential split-sample test to calibrate the LISFLOOD hydrological model using different precipitation sources, to show differences in model parameters and to ensure a minimum standard for operational validation of this simulation model. Southern Africa is selected as a case study because of its highly inter and intra-annual hydrological variability, mainly due to rainfall patterns characterized by events of short duration and high intensities. Therefore, the precipitation estimates might present large differences with ground observations, i.e., they can be highly inaccurate. The corresponding differences in streamflow bias and variability will allow us to assess differences in model's behavior and robustness of model parameters.

The paper is organized as follows. Section 2 presents the precipitation estimates and other hydrological model data used in this research, providing a description of the sensitivity analysis, calibration procedure and climate characteristics during the hydrological simulations. Section 3 contains a description of the calibration and validation results of the differential split-

sample test, while Section 4 gives a description of the model parameters, their robustness and uncertainty. The conclusions are finally presented in Section 5.

2. Methodology

2.1. Hydrological model

LISFLOOD is a GIS-based spatially-distributed hydrological rainfall-runoff model (De Roo et al., 2000; Van Der Knijff et al., 2010; Burek et al., 2013). Driven by meteorological forcing data, LISFLOOD calculates a complete water balance at a daily time step and every grid-cell defined in the model domain. Processes simulated for each grid cell include snowmelt, soil freezing, surface runoff, infiltration into the soil, preferential flow, redistribution of soil moisture within the soil profile, drainage of water to the groundwater system, groundwater storage, and groundwater base flow. Runoff produced for every grid cell is routed through the river network using a kinematic wave approach. Lakes, reservoirs and retention areas or polders can be simulated by giving their location, size and in- and outflow boundary conditions. LISFLOOD is currently used for studies dealing with water resources (Mubareka et al., 2013; Sepulcre-Canto et al., 2012), climate change impact assessment (Dankers and Feyen, 2008, 2009; Feyen et al., 2009; Rojas et al., 2012), flash flood forecasting (Alfieri et al., 2012) and flood forecasting for Europe (Bartholmes et al., 2009; Pappenberger et al., 2011; Ramos et al., 2007; De Roo et al., 2011; Thielen et al., 2009) and recently for Africa (Thiemig et al., 2015).

Spatially variable input parameters and variables were obtained from different databases. Soil properties were derived from the Harmonized World Soil Database (HWSD). Forest fraction and land use cover were obtained from the Global Land Cover 2000 (GLC2000) dataset (Bartholomé et al., 2003). In addition to land cover, the vegetative properties were obtained from the VGT4AFRICA Project. A more detailed description of the static input maps for Africa is given by (Bódis, 2009). Water use information from the Global Crop Water Model (GCWM – Siebert and Döll, 2008, 2010) is dynamically coupled with LISFLOOD. It is assumed that water is subtracted solely from the streamflow and not from internal storages.

The drainage network of the African river basins were obtained using a sequence of upscaling operations performed on the river network derived from a Shuttle Radar Topography Mission (SRTM; Jarvis et al., 2008) elevation model with spatial resolution of 90 m. With the upscaling from a fine towards a coarser scale the accuracy of the drainage network data can be lost and manual corrections should be applied. However, in the current pan-African setup we applied a new algorithm developed by Wu et al. (2011) for automatic upscaling of river networks, which successfully addresses most of the upscaling issues.

Meteorological forcing data for the pan-African domain of the LISFLOOD model were obtained from the ERA-Interim reanalysis dataset (Simmons et al., 2007). Gridded data products of ERA-Interim include a large variety of surface and upper-air parameters. Here we retrieved 3-hourly or daily estimates of wind speed, minimum and maximum temperature, dewpoint temperature, and solar and thermal radiation at a grid of 0.25° from the original Gaussian reduced grid (about 0.7°). Afterwards, the Penman-Monteith formula was used to compute potential evapotranspiration, evaporation rates for open water and bare soil surfaces, to be used as input data for the hydrological model. The current pan-African setup of LISFLOOD uses a 0.1° grid, which means that all the datasets were re-sampled to 0.1° of spatial resolution.

2.2. Data

2.2.1. Precipitation sources

The precipitation products used in this work are the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) 3B42 version 6 (hereafter 3B42V6) and the latest version 7 (3B42V7), the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center morphing technique version 1.0 (CMORPHV1.0), ERA-Interim precipitation corrected using the Global Precipitation Climatology Project (GPCP) dataset (ERAIGPCP), the Global Satellite Mapping of Precipitation moving vector with Kalman filter (GSMaP), The Precipitation Estimation from Remotely Sensed Information Using Neural Networks product (PERSIANN) and The NOAA African Precipitation Estimation Algorithm (RFE 2.0). A brief description of each product is given below.

The TRMM 3B42V6 product has been produced since 1998 and the estimates are produced in four stages (Huffman et al., 2007): (1) the microwave estimates (TMI, AMSR-E, SSM/I and AMSU-B) precipitation are calibrated and combined, (2) infrared (IR) precipitation estimates are created using the calibrated microwave precipitation, (3) the microwave and IR estimates are combined, and (4) rescaling to monthly data is applied using monthly gauge data. The latest version, the 3B42V7, is a reprocessed version of 3B42V6 with changes in the algorithm and includes additional datasets (Huffman et al., 2010; Huffman and Bolvin, 2012). Both 3B42V6 and 3B42V7 product estimates are available approximately 2 months after observation, but are more accurate and suitable for research compared to the near-real-time products (Huffman et al., 2007, 2010). They are released on a 0.25° by 0.25° grid at a 3-hourly temporal resolution and cover all latitudes between 50°N and 50°S.

The main inputs for the CMORPHV1.0 are the combined IR and microwave estimates. The passive microwave estimates are interpolated using atmospheric motion vectors from two successive IR images at 30-min intervals (Joyce et al., 2004). The final product includes the raw, satellite only precipitation estimates as well as bias corrected and gauge-satellite blended precipitation products. The original product has a very high spatial resolution: 8 km grid and half-hourly time step. However,

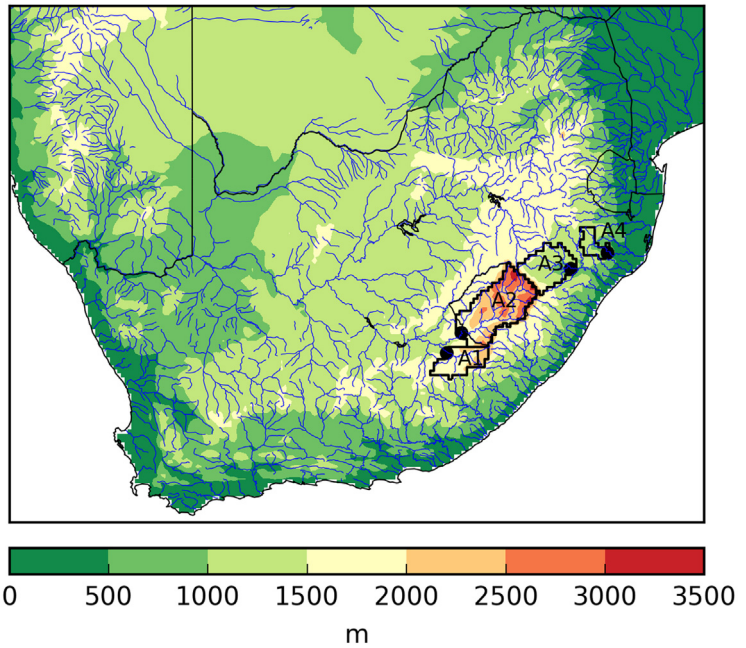

Fig. 1. Elevation map of Southern Africa including the four headwaters with the outlets marked as black dots.

both radar and satellite precipitation datasets are gridded to the 0.25° grid and aggregated to a 3-h time step to ensure common spatial and temporal scales in a global belt extending from 60°N to 60°S.

The ERAIGPCP is estimated by the numerical model based on temperature and humidity information derived from assimilated observations originating from passive microwave and in situ measurements (Dee et al., 2011). Balsamo et al. (2010) reported on systematic biases in the ERA-Interim precipitation data. To correct for these biases, the ERA-Interim precipitation is corrected using the GPCP dataset. Details of the rescaling method can be found in Balsamo et al. (2010).

The main data sources for the GSMaP algorithm are TRMM/TMI, Aqua/AMSR-E, DMSP/SSMI, and IR data. In addition to these, AMSU-B's are included in the GSMaP product. The algorithm uses the Kalman filter to retrieve the precipitation rate (Ushio et al., 2009) and refines the precipitation at each 0.1° pixel using the relation between the IR brightness temperature and surface precipitation rates.

PERSIANN is developed with artificial neural networks estimating precipitation rates using IR. The accuracy is improved with adjustments in the network parameters using precipitation estimates from microwave and ground rain rates where available (Hsu et al., 1997).

The RFE2.0 product (Herman et al., 1997) is based on a combination of passive microwave and IR precipitation estimates. Daily rain gauge station data from Global Telecommunication System (GTS) records is used for bias correction. The spatial resolution corresponds to a 0.1° grid which extends from 40°N to 40°S and 20°W to 55°E.

For this study, each precipitation source is resampled (nearest neighbor algorithm) onto a 0.1° common grid, from July 2001 until June 2010 at a daily time step to drive the LISFLOOD hydrological model.

2.2.2. Observed streamflow

The climate of the Southern African region is characterized as highly variable, both temporally and spatially. Many rivers pass through different climate zones from tropical to extremely arid, and the streamflow regimes are amongst the most variable in the world (Gorgens and Hughes, 1982). Water resource management is of high importance for South Africa to maintain reliable water supplies at times of water stress. Therefore, all major rivers are being regulated to some degree (Walmsley et al., 1999).

Four headwaters were selected (Fig. 1) for the most important rivers in Southern Africa, trying to exclude the unnatural flow regimes downstream as much as possible. Daily streamflow records were collected from 4 gauging stations (Fig. 1), provided by the Department of Water Affairs in South Africa (www.dwaf.gov.za), for the period July 2001 until June 2010.

2.3. Selection of study area

As described in Section 2.2.2, four headwaters in Southern Africa with near-undisturbed hydrological regime were selected for analysis. However, man-made reservoirs exist in both regions A2 and A3, but their small regulation capacity have little impact on streamflow amount, as observed in the flow duration curves (not shown here). The locations of

Table 1
Physiographic characteristics for the study area. The characteristics are calculated based on the ERA-Interim for evapotranspiration (PET) and temperature (TA), and ERA-interim corrected with GPCP for the precipitation (P).

Type	Variable	A1	A2	A3	A4
Station	Name	Roodewal	Oranjedraai	Tugela Ferry	Overvloed
	River	Kraai	Orange	Tugela	White Mfolozi
	Catchment	Orange	Orange	Tugela	Mfolozi
Topography	Area (km^2)	8667	24960	12987	3917
	Elevation (m)	1848	2297	1291	1047
Climate	HI (−)	0.56	0.70	0.67	0.69
	P ($mm\ year^{-1}$)	748	832	875	876
	PET ($mm\ year^{-1}$)	1342	1186	1310	1267
	CORR (−)	0.29	0.47	0.49	0.53
	TA (°C)	12.6	10.5	16.8	17.0
	Q ($m^3 s^{-1}$)	23.4	105.6	43.4	5.8

the headwaters are shown in Fig. 1, while Table 1 lists the climatic and topographical attributes of the selected catchments. The four headwaters are located in or on the Eastern plains of the Drakensberg mountain range where the most important rivers originate. The size of the headwaters range from about 3917 km^2 for A4 to about 24960 km^2 for A2, and the average elevation ranges from 1047 (A4) to 2297 (A1) meters a.s.l. (see Table 1).

According to the calculated humidity index ($HI = P/PET$), the climate in the selected headwaters is sub-humid with an average annual precipitation (P) ranging from 748 (A1) to 876 (A4) $mm\ yr^{-1}$ (according to ERAIGPCP). The average annual evapotranspiration (PET) ranges from 1186 (A1) to 1342 (A4) $mm\ yr^{-1}$. The Pearson product-moment correlation coefficient (CORR) between monthly moisture (P) and energy availability (PET) ranges from 0.29 to 0.53 showing a good positive correlation in A2, A3 and A4 (> 0.4) which, according to Petersen et al. (2012), represents areas with a strong seasonal precipitation regime. Precipitation occurs mainly during the summer months (i.e. October to March) and the major sources of rainfall during summer are thunderstorms, orographically induced storms (Tyson et al., 1976), and occasionally tropical cyclones (Dyson and Van Heerden, 2001; Reason and Keibel, 2004). Although these systems may result in devastating floods, they are an important source of water to supply the increasing water demand in this area. The average annual temperature (TA) ranges from about 10.5 °C in the mountainous areas (A2) to 17 °C in the headwater with the lowest elevation (A4). The average monthly streamflow (Q) ranges from 5.8 (A4) up to 105.6 $m^3\ s^{-1}$ (A2), but the high rainfall variability results in large fluctuations of the streamflow regime.

2.4. Sensitivity analysis

Before using a calibration procedure to obtain parameter values for the hydrological model, we identified parameters with the highest influence in simulated streamflow. In the current LISFLOOD version, experience of the model development team suggests the following parameters to be calibrated. The Upper Zone Time Constant (UZTC) and Lower Zone Time Constant (LZTC) reflect the residence time of water in the upper and lower groundwater zone, respectively. As such, they control the amount and timing of outflow from the groundwater reservoir. The Groundwater Percolation Value (GwPV) controls the flow from the upper and lower groundwater zone. The Ground Water Loss Fraction (GwLoss) is the rate of flow out of the lower groundwater zone, expressed as a fraction of the inflow, GwPV. The Xinanjiang parameter (b_Xinan) is an empirical shape parameter in the Xinanjiang model (Zhao and Liu, 1995) that is used to simulate infiltration. It controls the fraction of saturated area within a grid cell that is contributing to runoff; hence it is inversely related to infiltration. The Power Preferential Flow Parameter (PPrefFlow) is also an empirical shape parameter of the function describing the flow that bypasses the soil-matrix and drains directly to the groundwater. The Channel Manning parameter (CCM) is a multiplier that is applied to the Manning's roughness maps of the channel system. The higher the value the more friction the water will experience and resulting in a lower specific streamflow. The evaporation input does not have a large spatial variation and is subject to errors in Southern Africa (Hughes, 2006; Hughes et al., 2010). Moreover, there are large differences between actual evaporation estimates partially dependent on precipitation (Trambauer et al., 2014). To partially consider uncertainties on the magnitude of the evaporation, related to the different precipitation sources, a multiplier to the evaporation input (CalEvap) is also calibrated. In addition, for catchments with upstream reservoirs (A2 and A3, with 2 reservoirs each) a set of four reservoir parameters are included to describe the inflow and outflow streamflows. Finally, the twelve parameters selected for the global sensitivity analyses are listed in Table 2.

The variance-based method of Sobol (Sobol', 2007) was selected as a global sensitivity analysis technique to identify the relevant model parameters. Here, the Sobol's algorithm proposed by Saltelli et al. (2010) is used to quantify the first- and total-order indices of each model parameter described in the previous section. The first-order sensitivity index (Si) measures the direct contribution of each model parameter to the total model output variance. By definition, the sum of the first-order indices of each single calibration parameter is smaller or equal to 1. The first-order index provides a measure of the direct importance of each parameter, and the larger the first-order sensitivity index the more important the parameter. On the other hand, the total-order sensitivity index (Sti) is a measure of the total effect of each parameter, i.e., its direct effect and all the interactions with other parameters. If the first-order sensitivity index is equal to the total-order sensitivity index for

Table 2
LISFLOOD calibration parameters with both upper and lower bound.

Parameter	Unit	Min	Max
UZTC	days	5	40
LZTC	days	50	2500
GwPV	mm day^{-1}	0.5	2
GwLoss	–	0.01	0.70
b_Xinan	–	0.01	1
PPrefFlow	–	1	4
CCM	–	0.1	15
CalEvap	–	0.8	1.2
rnlim	–	0.10*	0.80*
rflim	–	0.81*	1.0*
rnormq	$m^3 s^{-1}$	0.1*	20*
rndq	$m^3 s^{-1}$	12*	1200*

*Ranges are reservoir dependent and only used for region A2 and A3 in the global sensitivity analysis.

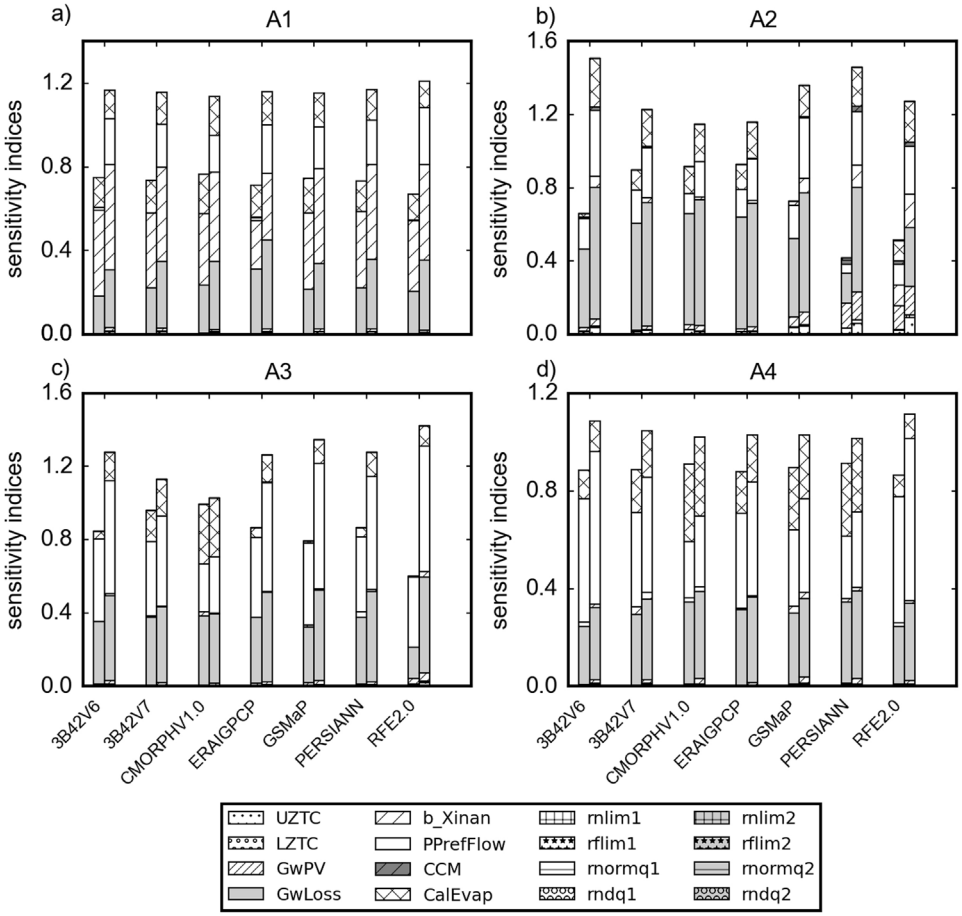

Fig. 2. Barplots of the sensitivity indices with respect to the calibration parameters for each precipitation source based on the period 2002–2006. For each precipitation source, two bars are displayed: the left one is for first-order sensitivity indices and the right one is for the total-order sensitivity indices.

a given parameter, it means that this parameter does not interact with the other parameters. Conversely, if the first-order sensitivity index is lower than the total-order sensitivity index, it indicates strong interactions between this parameter and other parameters (high complexity model). Note that the sum of the total-order sensitivity indices may be larger than 1 (because some effects are counted more than once in the sum).

The Sobol's method was used to identify the most important parameters when the LISFLOOD model was forced with each one of the seven SRFE described in Section 2.2.1. The evaluation of both the first- and total-order sensitivity indices is shown in Fig. 2. For region A1 (Fig. 2a), the most sensitive parameters are related to infiltration (PPrefFlow, b_Xinan), groundwater processes (GwLoss) and evaporation (CalEvap). Both GwLoss and b_Xinan are clearly the most important variables for every precipitation source. The parameter PPrefFlow is found to have a very weak direct influence, but a high

interaction with the other parameters. For region A2 (Fig. 2b), the most influential parameter is GwLoss for most of the SRFEs, except for the RFE2.0 precipitation product. GwLoss explains by itself more than 50% of the model output variance. Several precipitation products led to model parameters highly related amongst them (i.e., they are not independent), as shown by the height difference of the vertical bars representing the sum of all first-order sensitivity indices: 50% or more. For region A3 (Fig. 2c), only three parameters (GwLoss, PPrefFlow and CalEvap) are influential for the model output variance. The interaction between the parameters is variable, with weak interaction for CMORPHV1.0 and strong interaction for RFE2.0. Fig. 2d shows similar results for region A4, but with a stronger direct influence and a weaker interaction amongst those three parameters. Note that the parameters related to the reservoirs (regions A2 and A3) are of no importance at all for the model output variance. The aforementioned results show that only parameters related to evaporation (CalEvap), infiltration (PPrefFlow) and groundwater processes (GwLoss) are the most important for the next calibration step. The strong interactions observed amongst them clearly indicate that trying to optimize one parameter at a time would be inefficient. Although an independent calibration of only the most important parameters could be attempted, we calibrated jointly the 8 parameters (UZTC, LZTC, GwPV, GwLoss, b_Xinan, PPrefFlow, CCM and CalEvap). The parameters to be calibrated and their respective physically-reasonable ranges are listed in Table 2. The parameters related to the reservoirs are neglected in the calibration procedure as they were not identified as important during the global sensitivity step.

2.5. Calibration procedure

The open source *hydroPSO* R package v0.3-3 (Zambrano-Bigiarini, 2013; Zambrano-Bigiarini and Rojas, 2013) was used for calibrating the LISFLOOD model in the four selected catchments. *hydroPSO* is a new global optimisation/calibration tool based on the Particle Swarm Optimisation (PSO) algorithm. *hydroPSO* is an effective and efficient calibration tool, as shown in recent applications with different hydrological and environmental models (Thiemig et al., 2013; Brauer et al., 2014a,b; Abdelaziz and Zambrano-Bigiarini, 2014; Silal et al., 2015). More details about *hydroPSO* can be found in Zambrano-Bigiarini and Rojas (2013) and Zambrano-Bigiarini et al. (2013).

The differential split-sample test was performed over two time periods of four years (Jul 2002–Jun 2006 and Jul 2006–Jun 2010), running the LISFLOOD model with a daily time step. For each model simulation a pre-run is performed to initialize the groundwater storage and subsequently an additional warm-up period of six months is used to initialize the water storage components of the model.

The objective function chosen to drive the optimization was the modified Kling-Gupta efficiency (KGE') between simulated and observed streamflows (Kling et al., 2012), which was maximized towards an optimal value of 1.

$$KGE' = 1 - \sqrt{K + L + M}, \tag{1}$$

where K is the correlation term

$$K = (r - 1)^2 \tag{2}$$

L the bias term

$$L = \left(\beta - 1\right)^2 \tag{3}$$

and M the variability term

$$M = (\gamma - 1)^2 \tag{4}$$

with r is the Pearson product-moment correlation coefficient, β is the bias ratio:

$$\beta = \frac{\mu_s}{\mu_o} \tag{5}$$

and γ is the variability ratio:

$$\gamma = \frac{CV_s}{CV_o} = \frac{\frac{\sigma_s}{\mu_s}}{\frac{\sigma_o}{\mu_o}} \tag{6}$$

where μ is the mean streamflow ($m^3\ s^{-1}$), CV is the coefficient of variation and σ is the standard deviation of the streamflow ($m^3\ s^{-1}$), and the indices s and o represent simulated and observed streamflow, respectively. KGE', r, β, and γ have their optimum at unity. The hydrological performance is classified as mentioned in Thiemig et al. (2013) and according Kling's review comment (2016): excellent ($KGE' \geq 0.9$), good ($0.9 > KGE' \geq 0.75$), intermediate ($0.75 > KGE' \geq 0.5$), poor ($0.5 > KGE' > 0.0$) and very poor ($KGE' \leq 0.0$). More information about KGE' in relation to other objective functions can be found in Gupta et al. (2009). Although the model was run with daily time steps, the results of the calibration are analyzed on a monthly basis.

2.6. Climate characteristics of 2002–2006 and 2006–2010

The calibration of LISFLOOD is carried out for each of the seven precipitation products described in Section 2.2.1, using two calibration periods with highly different climate characteristics: 2002–2006 and 2006–2010. Fig. 3 summarizes the

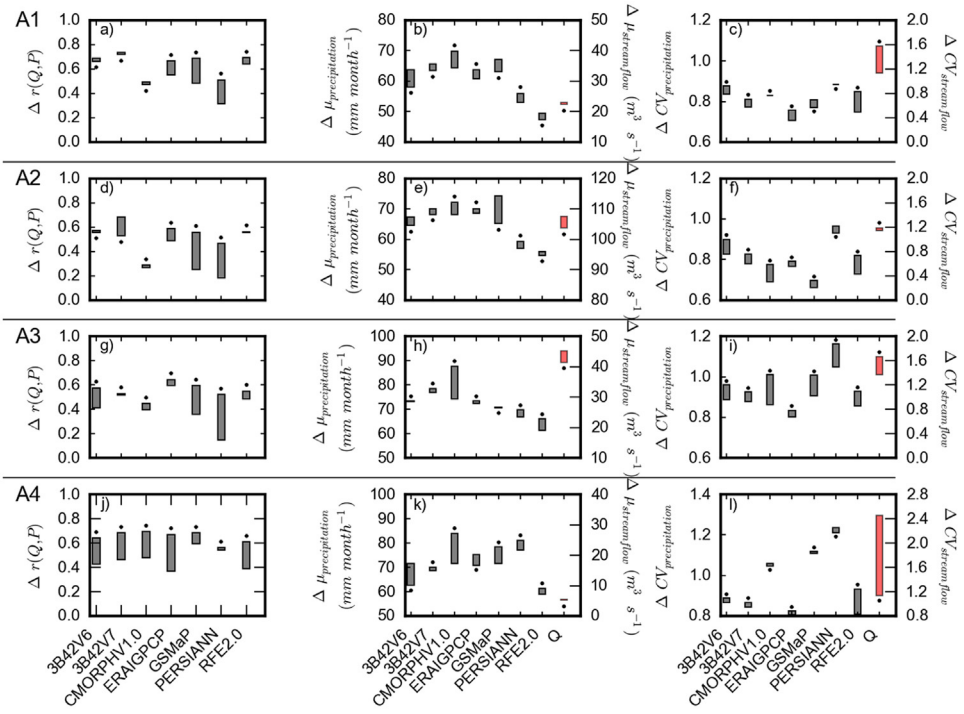

Fig. 3. Barplots with the change in correlation (panels a, d, g, j), mean (panels b, e, h, k) and variation (panels e, f, i, l) between 2002–2006 and 2006–2010 for each precipitation source and observed streamflow in each study area A1 (panels a, b, c), A2 (panels d, e, f), A3 (panels g, h, i), and A4 (panels j, k, l). The value of 2002–2006 is the upper or lower value of the vertical bar accompanied with the black dot.

climate characteristics of each selected catchment, by showing the relation between monthly precipitation and observed streamflow during the two calibration periods. Fig. 3 will be used as a guideline to explain the results which are presented in the following section.

For region A1, the lowest correlation between the observed streamflow and precipitation is observed for CMORPHV1.0 in 2002–2006 and PERSIANN in 2006–2010. The highest correlation between observed streamflow and precipitation is found using the 3B42V7 product for both calibration periods (Fig. 3a). The difference in the amount of observed streamflow (Q, Fig. 3b) between 2002–2006 and 2006–2010 is negligible whereas the streamflow in 2002–2006 is slightly lower compared to 2006–2010. The difference in precipitation amount from different products is dispersed, with the CMORHPV1.0 as the wettest and the RFE2.0 as the driest product for 2002–2006, while for 2006–2010 GSMaP is the wettest and RFE2.0 is the driest. Note that the precipitation mean for the CMORPHV1.0, ERAIGPCP and PERSIANN products in 2002–2006 are larger compared to 2006–2010, which is in contrast with the observed streamflow which is larger in 2006–2010 compared to 2002–2006. The variability of the observed streamflow is larger in 2002–2006 compared to 2006–2010 (Fig. 3c), which is also, but to a lesser extent, observed from the 3B42V6, CMORPHV1.0, 3B42V7, ERAIGPCP and RFE2.0 products. In contrast, the variability from the GSMaP and the PERSIANN product is larger in 2006–2010 compared to 2002–2006.

For region A2, the distribution of the correlation, mean and variability between the precipitation products and observed streamflow in 2002–2006 and 2006–2010 is quite similar as for region A1. The lowest correlation between the observed streamflow and precipitation is again observed for CMORPHV1.0 in 2002–2006 and PERSIANN in 2006–2010. The highest correlation between observed streamflow and precipitation is found with the ERAIGPCP product for the 2002–2006 period and 3B42V7 for the 2006–2010 period (Fig. 3d). The amount of observed streamflow is higher for the 2006–2010 period compared to the 2002–2006 period (Fig. 3e). For the precipitation, the CMORPHV1.0 product is the wettest and the RFE2.0 product the driest out of all products for the 2002–2006 period. For 2006–2010, the GSMaP product is the wettest and the RFE2.0 is the driest product. The variability of the observed streamflow is in 2002–2006 slightly larger compared to 2006–2010 (Fig. 3f). The PERSIANN precipitation product has the largest variability in both periods but a contradicting distribution compared to the observed streamflow between 2002–2006 and 2006–2010. The rest of the products have a similar precipitation variability distribution but a much lower value compared to the observed streamflow.

For region A3, the lowest correlation between the observed streamflow and precipitation is observed for CMORPHV1.0 in 2002–2006 and PERSIANN in 2006–2010. The highest correlation between observed streamflow and precipitation is found with the ERAIGPCP product for 2002–2006 and 3B42V7 for 2006–2010 (Fig. 3g). The amount of observed streamflow is higher for the 2006–2010 period compared to the 2002–2006 period (Fig. 3h). For the precipitation, the CMORPHV1.0 product is the wettest and the RFE2.0 product the driest out of all products for the 2002–2006 period. For 2006–2010, the 3B42V7 product is the wettest and the RFE2.0 is the driest product. The variability of the observed streamflow is larger in

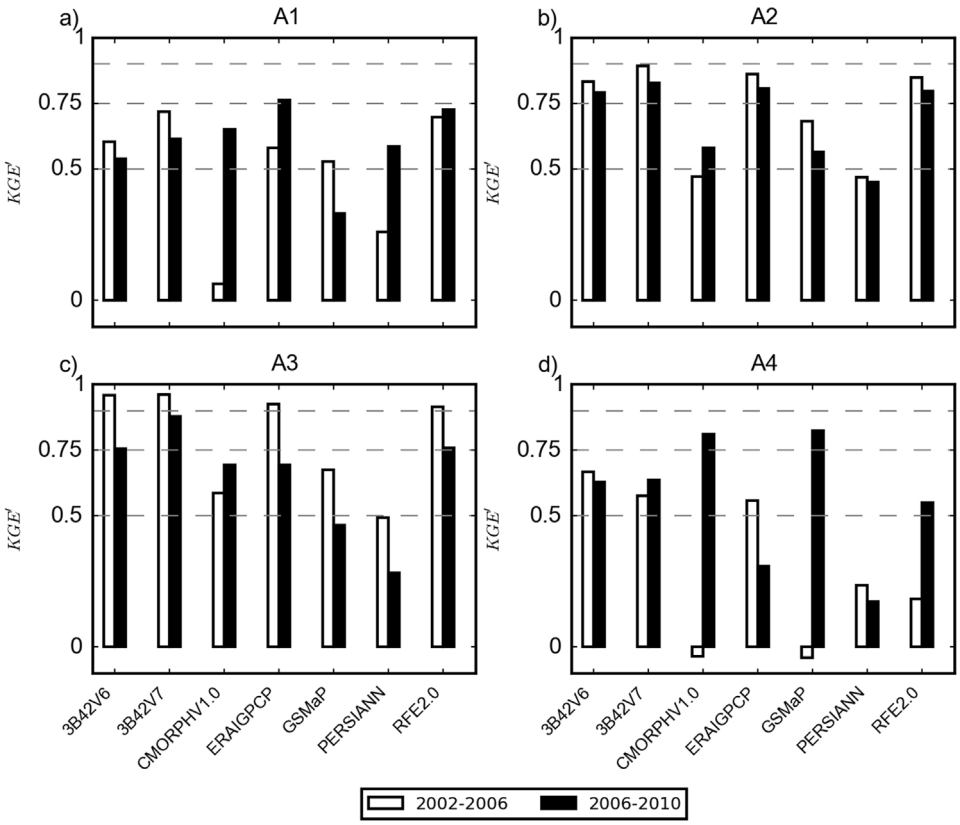

Fig. 4. Barplots of calibration results in terms of *KGE′* in the two periods of the differential split sample test: 2002–2006 and 2006–2010, for each precipitation product. Panels a), b), c) and d) correspond to region A1, A2, A3 and A4, respectively.

2002–2006 compared to 2006–2010 (Fig. 3i). The PERSIANN precipitation product has the largest precipitation variability in both periods, while ERAIGPCP has the lowest.

For region A4, the correlation between the observed streamflow and precipitation is quite similar for all precipitation products for both calibration periods with a higher correlation for the 2002–2006 period (Fig. 3j). The difference in the amount of observed streamflow between 2002–2006 and 2006–2010 is negligible (Fig. 3k). For precipitation, the CMORPHV1.0 product is the wettest and RFE2.0 is the driest for the 2002–2006 period. For 2006–2010, PERSIANN is the wettest and RFE2.0 is the driest product. The variability of the observed streamflow is in 2002–2006 much lower compared to 2006–2010 (Fig. 3l). The PERSIANN precipitation product has the largest variability in both periods.

3. Results

3.1. Calibration

Fig. 4 shows the best values of *KGE′* obtained in the two calibration periods of the differential split sample test: 2002–2006 and 2006–2010, when the LISFLOOD model is calibrated with every precipitation product for each study area. In addition, taking into account that high values of K, L and M (Eqs. (2)–(4)) lead to low values of *KGE′* (Eq. (1)), we computed the relative contributions of K, L, and M to the *KGE′* values obtained in each calibration ($100 \cdot K$, $LorMK + L + M$), in order to identify the *limiting factor* that prevents *KGE′* to achieve an ideal value of 1.0. Fig. 5 shows the relative contribution of each limiting factor to the *KGE′* value: K, L, M (Eqs. (1)–(4)), which represent the impact of the linear correlation (r), bias (β) and variability (γ) between simulated and observed streamflow on the calibration results.

The calibration results for A1 (Fig. 4a) were deemed to be satisfactory (intermediate or good) for the simulations with products 3B42V6, 3B42V7, ERAIGPCP and RFE2.0 in both calibration periods. On the other hand, calibration results for the CMORPHV1.0 and PERSIANN were poor for the period 2002–2006 calibration but intermediate for 2006–2010. The opposite was true for simulations forced by the GSMaP product. Fig. 5a shows with vertical bars the relative contribution of K, L and M to the *KGE′*, making clear that the poor performance of model simulations forced by CMORPHV1.0 and PERSIANN during 2002–2006 was mainly due to a poor correlation between simulated and observed streamflow (high K value). These findings are in agreement with the results of Fig. 3a, where both CMORPHV1.0 and PERSIANN presented the lowest correlation with

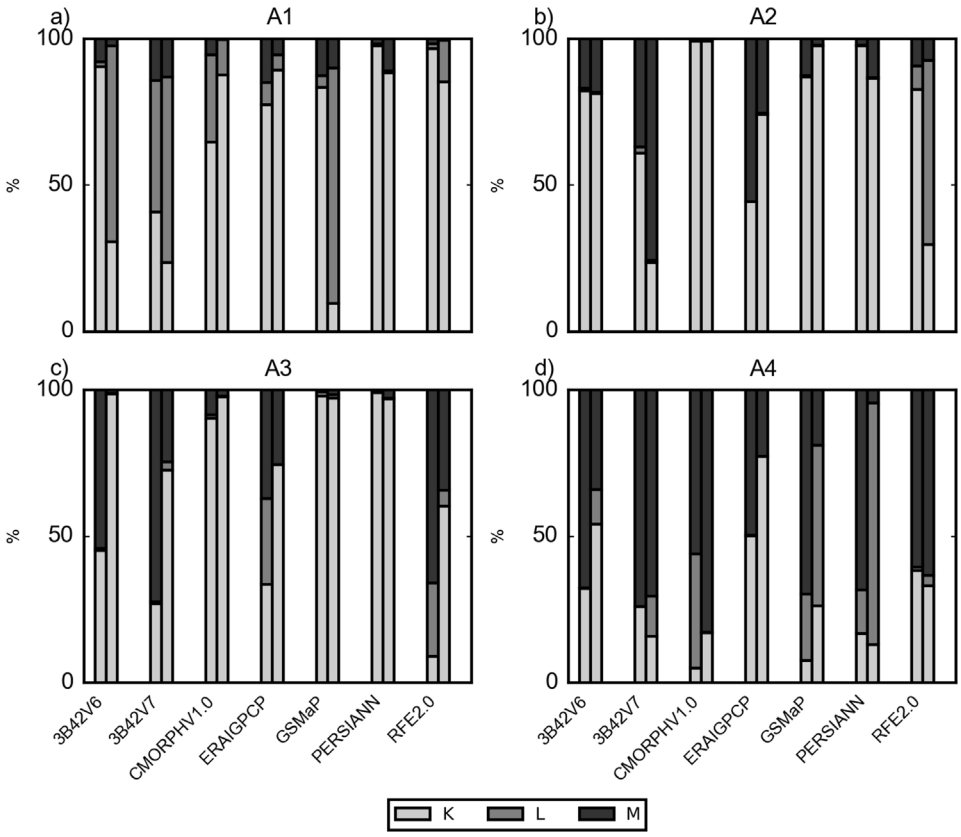

Fig. 5. Barplots with the relative contribution (in %) of the correlation (K), bias (L) and variability (M) which prevent the *KGE′* to achieve an optimum value of 1.0. For each precipitation product, K, L and M were computed in the two calibration periods of the differential split sample test: 2002–2006 and 2006–2010. Panels a), b), c) and d) correspond to region A1, A2, A3 and A4, respectively.

observed streamflow during 2002–2006. However, this relation is not always valid as the model performance for 2006–2010 with PERSIANN is reasonably good in spite of PERSIANN having the lowest correlation with the observed streamflow out of all precipitation products. On the other hand, Fig. 4a shows that simulations forced by GSMaP were the worst for the 2006–2010 period, which was mainly due to the high relative contribution of the bias term L (Fig. 5a). The previous result is in agreement with the fact that GSMaP was the wettest precipitation product during that period (see Fig. 3b). Calibration results for A2 (Fig. 4b) are intermediate or good for the simulations driven by products 3B42V6, 3B42V7, ERAIGPCP, GSMaP and RFE2.0 for both calibration periods. The poor performance of the CMORPHV1.0 and PERSIANN simulations in the 2002–2006 period was caused by a low correlation between observed and simulated streamflow (Fig. 5b). Fig. 3d shows also low correlations between precipitation and observed streamflow in 2002–2006, for both CMORPHV1.0 and PERSIANN. Fig. 5b shows that the poor performance of the PERSIANN simulations in 2006–2010 was the low correlation between observed and simulated streamflow, possibly related to the poorest correlation between precipitation and observed streamflow obtained by PERSIANN out of all products in 2006–2010 (Fig. 3e).

The calibration results for A3 (Fig. 4c) are satisfactory (excellent, intermediate or good) for the 3B42V6, 3B42V7, CMORPHV1.0, ERAIGPCP and RFE2.0 simulations for both calibration periods. The results of the GSMaP and PERSIANN simulations were intermediate in 2002–2006 and poor for the 2006–2010 period, mainly due to the weak correlation between simulated and observed streamflow (Fig. 5c). The previous result is possibly due to the low correlation between GSMaP and PERSIANN precipitation and the observed streamflow, which were the poorest out of all products in 2006–2010 (Fig. 3g).

In region A4 (Fig. 4d), the calibration results for the 2002–2006 period were intermediate or good only for the 3B42V6, 3B42V7 and ERAIGPCP products. The poor results obtained with CMORPHV1.0, GSMaP, PERSIANN and RFE2.0 products were mainly caused by the difference in the variability term (M in Eqs. (1) and (4)) between simulated and observed streamflow, which is apparently related to high variability of those precipitation products (Fig. 3l). The CMORPHV1.0, GSMaP, PERSIANN and RFE2.0 are matching the variability of the streamflow more closely compared to the three remaining products. Controversially, high precipitation variability of these four precipitation products results in streamflow overestimation due to overestimation of both rainfall amount and duration. The calibration results of the simulations forced by ERAIGPCP and PERSIANN in 2006–2010 were poor due to the contribution of correlation (K term in Eqs. (1) and (2)) and bias (L term in Eqs. (1) and (3)) between simulated and observed streamflow, respectively. This is in agreement with Fig. 3j, where the corre-

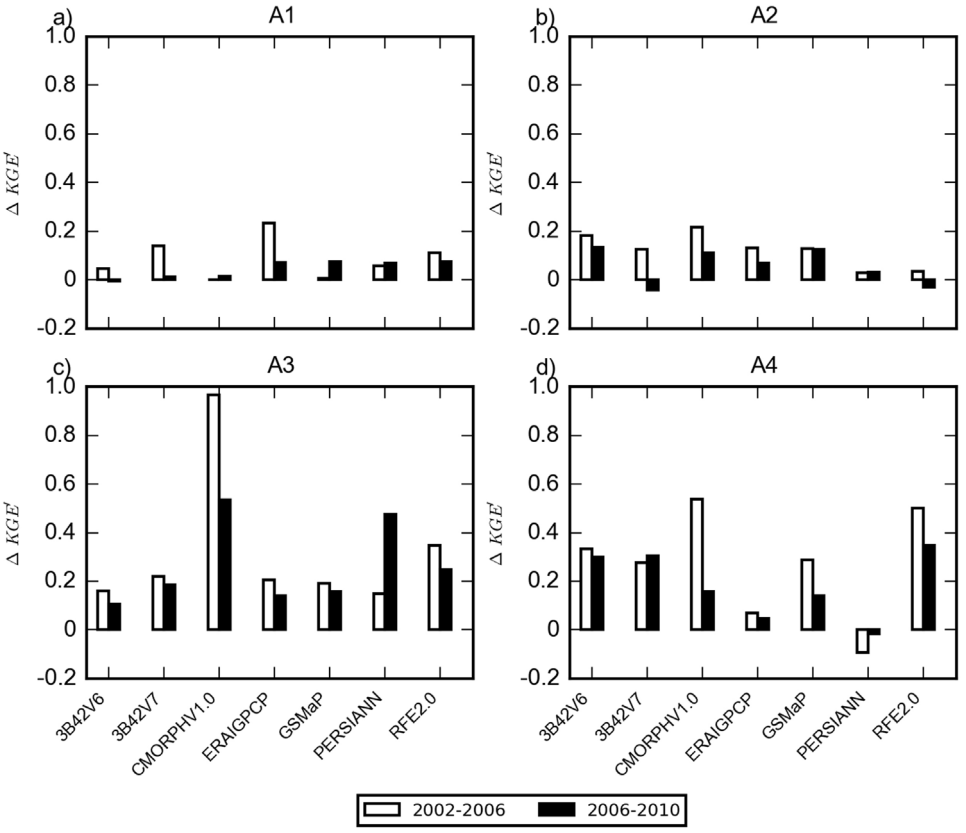

Fig. 6. Barplots of the validation results in terms of the variation of KGE' ($\Delta KGE' = KGE'$ calibration run 2002–2006 (2006–2010) minus KGE' validation run 2002–2006 (2006–2010) with optimum parameter values obtained from 2006–2010 (2002–2006)) in the two periods of the differential split sample test for each precipitation product. Panels a), b), c) and d) corresponds to region A1, A2, A3 and A4, respectively.

lation between ERAIGPCP precipitation and observed streamflow is the lowest, and with Fig. 3k where the mean monthly precipitation of PERSIANN is highest out of all products.

3.2. Robustness of model parameters

In order to test the robustness of the model parameters obtained in each calibration period, all the "optimum" parameter sets were validated on an independent time period with a different climatological regime, as explained in Section 2.6, for each headwater and precipitation product. In particular, when using the differential split-sample approach, the performance of model parameters obtained during the calibration in 2002–2006 (2006–2010) was compared against the performance of a validation run obtained with model parameters calibrated in the other time period 2006–2010 (2002–2006). Fig. 6 shows the KGE' values of the validation period run in 2002–2006 (2006–2010) using the optimum parameter set obtained during the calibration run in 2006–2010 (2002–2006) with the latter reference ($\Delta KGE' = KGE'_{CAL} - KGE'_{VAL}$). On the other hand, Fig. 7 shows for every precipitation product, study area and calibration period, the relative contribution of K, L and M to the obtained KGE' of the validation run. Model parameters obtained during calibration are considered *robust* if the model performs equally well in the validation run using model parameters derived in a different precipitation regime, i.e., the $\Delta KGE'$ should be small. Note that in a small number of cases the KGE' values obtained during the validation run are higher than the KGE' values obtained during the calibration run in the same period (e.g., PERSIANN in A4). This is due to the aggregation from daily to monthly values to compute the KGE'.

Fig. 6a shows that "optimum" parameter sets obtained for almost all the products in the study area A1 were deemed robust, as shown by the small value of $\Delta KGE'$. The only exception was the run driven by the ERAIGPCP precipitation, which showed an important reduction in model performance when running the model in 2002–2006 with parameters obtained for the 2006–2010 time period. The vertical bars in Fig. 7a show that the KGE' in the 2002–2006 validation period for the ERAIGPCP products is mainly limited by the bias (L) between simulated and observed streamflow. The comparison of observed and simulated hydrographs (not shown here) reveals that the observed streamflow in the validation period 2002–2006 is slightly lower compared to the calibration period 2006–2010. Controversially, the average ERAIGPCP precipitation in 2002–2006 is wetter compared to 2006–2010, which results in an overestimation of the observed streamflow.

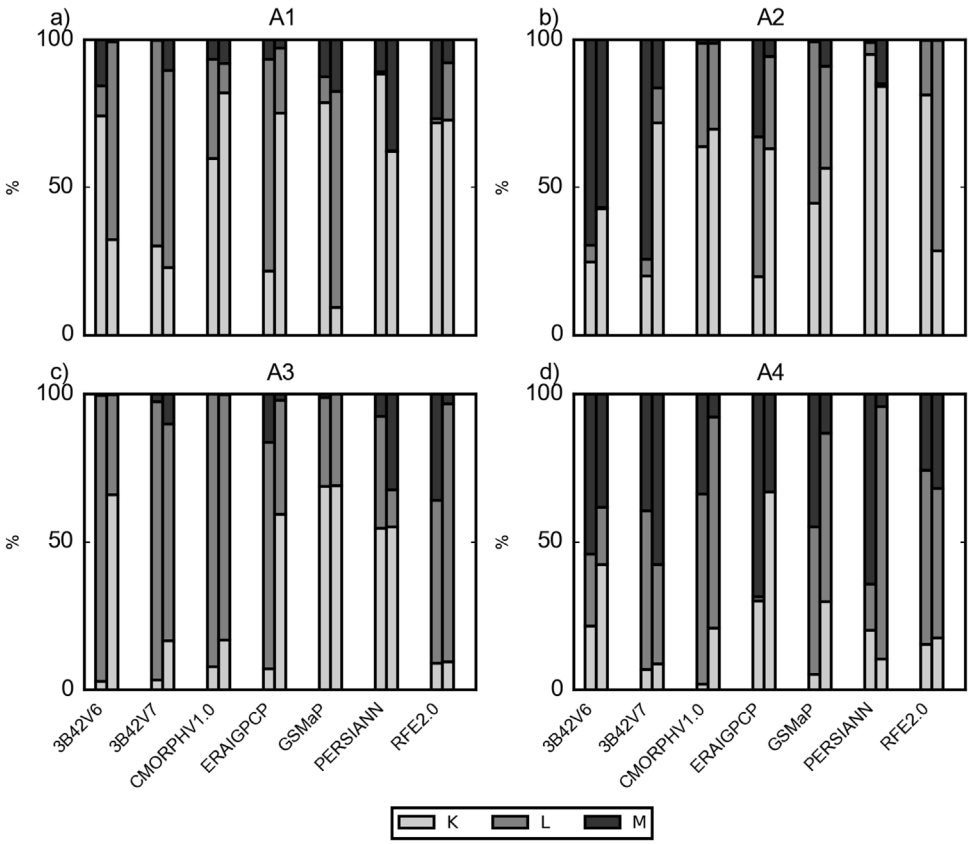

Fig. 7. Barplots with the relative contribution (in %) of the correlation (K), bias (L) and variability (M) which prevent *KGE′* to achieve an optimum value of 1.0. For each precipitation product, K, L and M were computed in the two validation periods of the differential split sample test: 2002–2006 and 2006–2010. Panels a), b), c) and d) correspond to region A1, A2, A3 and A4, respectively.

After the validation in region A2 (Fig. 6b) the model performance with the parameters forced with the 3B42V6, 3B42V7, ERAIGPCP and RFE2.0 products are deemed robust for both periods. Note that, in comparison with the calibration, the contribution of the bias (L) becomes a more pronounced variable in reducing the *KGE′* in most of the products. This is even more pronounced in region A3 (Fig. 6c). In both calibration periods (Fig. 4c) the limiting factor is mostly dominated by the correlation (K) or variability (M), whereas in the validation the limiting factor is the bias (L) between simulated and observed streamflow (Fig. 7c) sometimes resulting in large reductions of the *KGE′*. In Fig. 3h is found that the average precipitation of the CMORPHV1.0, ERAIGPCP, PERSIANN and RFE2.0 products is higher in 2002–2006 compared to 2006–2010. However, the opposite is true for the observed streamflow and consequently this result in bias issues and therefore a lack of robustness between calibration and validation. Calibrating these precipitation products in 2006–2010 (dry) and validating them in 2002–2006 (wet) result in an overestimation of the observed streamflow as the average observed streamflow is lower and the average precipitation is higher in the validation period compared to the calibration period. The opposite is also valid. Calibrating these precipitation products in 2002–2006 (wet) and validating them in 2006–2010 (dry) result in an underestimation of the observed streamflow as the average streamflow is higher and the average precipitation is lower in the validation period compared to the calibration period.

The limiting factor for poor calibration results in region A4 was mostly dominated by the variability term (M). However, poor validation results are observed in Fig. 6d due to both the variability (M) and bias (L) term as limiting factors (Fig. 7d). Only the reduction in *KGE′* of the validation of the ERAIGPCP simulations is low in both periods, but the calibration results in 2006–2010 were already poor due to poor correlation between observed and simulated streamflow. For this reason, no robust model parameters are found in region A4 no matter which precipitation product.

4. “Optimum” parameter sets under highly variable climate conditions

As seen in Fig. 7 the bias (L) between observed and simulated streamflow is an important factor in the loss of model performance in the validation period. Apparently, the model parameters are compensating the bias between simulated and observed streamflow during the calibration but these model parameters fail to capture the observed streamflow during the

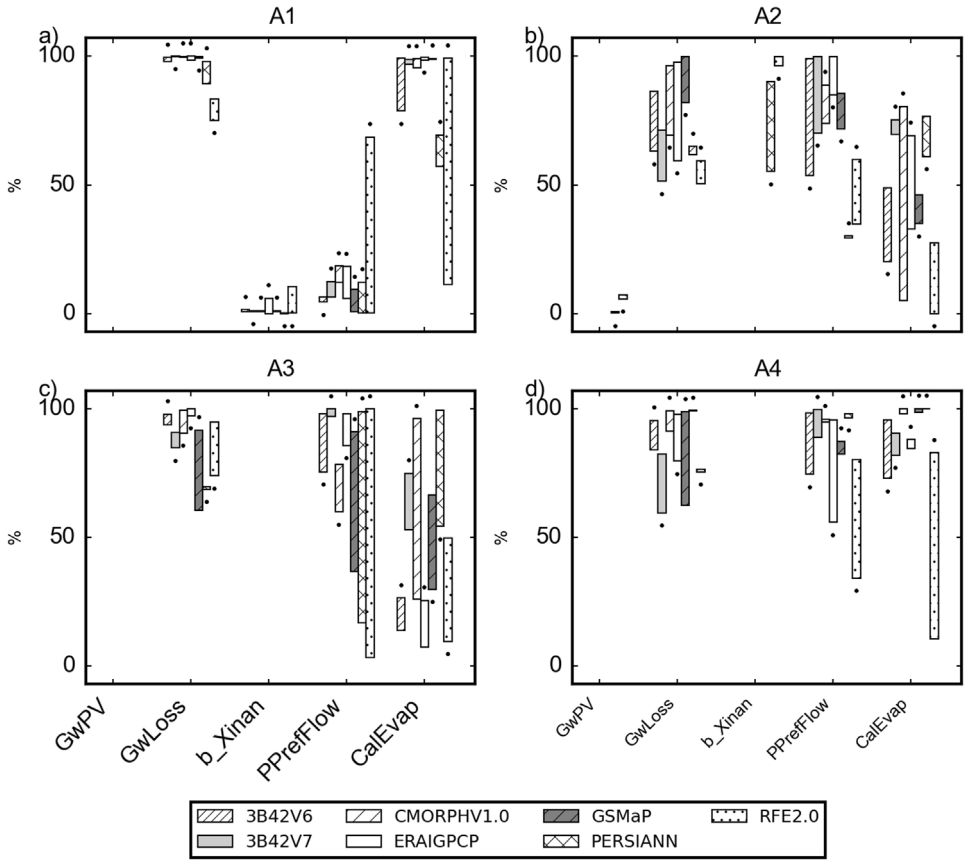

Fig. 8. Barplots with the relative change (in %) in the magnitude of the best parameter set found in the 2002–2006 and 2006–2010 calibration periods, whereas the value of 2002–2006 is the upper or lower value of the vertical bar accompanied with the black dot. The lower and upper bound values for the parameters (Table 2) represent 0 and 100% respectively and only the parameters with a total-order sensitivity index larger than 0.1 are included (Fig. 2).

validation period. Compensation for correlation and variability failure between observed and simulated streamflow is less pronounced as they are the dominant limiting factors in the calibration periods.

In Fig. 8 the length of the vertical bars show the relative change in the magnitude of the best parameter set found in the two calibration periods. In general it can be observed that the "optimum" parameter sets differ significantly between the calibration periods. Different values found for the optimum parameter sets in different time periods are a clear indication of parametric uncertainty in the obtained model parameters (Beven, 1993).

For region A1 (Fig. 8a), the difference in the model parameters between the calibration periods is negligible for most precipitation products except for RFE2.0. The model parameters GwLoss and CalEvap are reaching the upper values of the calibration range for almost all precipitation products in both calibration periods. This means that the model parameters are compensating the excess of rainfall by eliminating water both to deeper groundwater and by increasing the evaporation rate to match the observed streamflow. Typically, these parameter values result in bias problems between simulated and observed streamflow depending on whether one calibration period is drier or wetter compared to the other period. For example, Fig. 8a shows that the optimum value for the GwLoss and CalEvap parameters were found in the upper boundary of their calibration range, when LISFLOOD was calibrated with ERAIGPCP in 2006–2010. However, validating the previous optimum parameter set during a wetter period (Fig. 3b) results consequently in an overestimation of the observed streamflow (Fig. 5a). Although the results for both the calibration and validation in both periods for the RFE2.0 precipitation product are satisfactory (Figs. 4 a and 6 a), the obtained model parameters are quite different. However, the range of the two most sensitive parameters, GwLoss and b_Xinan (Fig. 3a) is small.

For region A2 (Fig. 8b), large differences in the best parameters obtained in the two calibration periods are observed. As seen in Fig. 2b the model performance is the most sensitive to the GwLoss parameter for most of the precipitation products. A large variation in the GwLoss value between the two calibration periods is needed to compensate for the precipitation difference between the two periods, resulting mostly in a reduction in model performance in the independent validation periods (Fig. 6b). The model parameter values from the PPrefflow and CalEvap are considerably different for the RFE2.0 product between the calibration periods while the model performance and robustness was judged to be good. Apparently,

the high interaction between the model parameters forced with RFE2.0 produced good results even with a different set of parameters.

The small reduction in *KGE′* in the validation periods for the 3B42V6, 3B42V7 and ERAIGPCP precipitation products in region A3 is probably related to the low model parameter ranges of the most sensitive model parameters for these products (Fig. 8c). Moreover, these are also the products with small differences between the mean precipitation in both calibration periods and therefore compensation is hardly needed.

For region A4, the model parameters reach the upper boundaries of the parameter ranges to eliminate water to match the observed streamflow for the wetter CMORPHV1.0, GSMaP and PERSIANN products (Fig. 8d). Although the model parameter ranges for the 3B42V6, 3B42V7 and ERAIGPCP are relatively small, the model parameters were not capable to compensate for the high variability of the streamflow and match the observed streamflow.

5. Conclusion

In this study, we applied the differential split-sample test to investigate the robustness of model parameters of a fully distributed hydrological model, by calibrating the model forced by seven satellite-based precipitation estimates. For model calibration we used the modified Kling-Gupta efficiency (*KGE′*) as objective function, which allowed us to differentiate the relative contribution of linear correlation (r), bias (β) and variability (γ) between the simulated and observed streamflow to the obtained *KGE′* value. We selected four regions in Southern Africa with highly variable climate conditions, which resulted in large precipitation differences between the seven precipitation products. Large discrepancies in terms of the r, β and γ between the simulated and observed streamflow were observed when using SRFEs as model input. Following a differential split-sample approach to test the robustness of the best parameter sets of the LISFLOOD model, the observed streamflow time series were split in two periods with highly different climatic characteristic: 2002–2006 and 2006–2010, and then model parameters obtained during calibration in one period were transferred to the other period to assess its performance. Thus, two optimized parameter sets were validated on two independent time periods, which is essential in highly variable climate and non-stationary conditions (Xu, 1999). From the results obtained in this work it can be concluded that a SRFE product which is accurate enough (in comparison to ground observations) can help to avoid future model robustness problems. Therefore, accurate precipitation estimates are essential for obtaining robust model parameters before extrapolating them in time and space. However, this statement is catchment and time period specific, as the catchment behavior in highly variable climate is difficult to capture and on the limit of model's capability. Therefore, hydrological analyses or predictions which are based on transferred model parameters in highly variable climate or non-stationary conditions should be taken with care. Note that reasons for the lack of model robustness can be multiple (see Introduction) and we only take uncertain precipitation estimates into account.

We summarize our findings as follows:

- The (lack of) model performance in terms of *KGE′* differs substantially between model runs driven by the seven different precipitation products and is partially reflected by their precipitation characteristics. A high positive correlation between the upstream monthly precipitation estimate and the observed streamflow proved to be a good *a priori* indicator for obtaining good model performance in a posterior calibration process, as long as the precipitation variability (*CV*) is not too high. High precipitation variability often results in overestimation of the duration of the rainfall events and, therefore, an overestimation of the total amounts as well (e.g., the r(P,Q) between PERSIANN and observed streamflow in region A4 for 2002–2006 is 0.56 (significant), but the *CV* of the precipitation has a value of 1.21 which result in a poor *KGE′* of 0.24).
- Model parameters related to infiltration (PPrefFlow and b_Xinan) and water losses in terms of groundwater (GwLoss) and evaporation (CalEvap) are the most sensitive parameters for the LISFLOOD model in Southern Africa. These parameters are mostly important to compensate the biased precipitation inputs. Most likely, for wet precipitation products (e.g., CMORPHV1.0) the interaction between the model parameters is less important compared to dry products (e.g., RFE2.0). To compensate the precipitation bias this parameters find their optimum value in the upper or lower bounds of their calibration range. The sensitive model parameters of dry precipitation products interact much more with each other, as the timing and shape of the hydrograph becomes important if compensating the precipitation bias is less dominant.
- Parameter sets obtained during calibration compensate, up to a certain extent, the bias in the precipitation estimate to match the observed steamflow, as the bias is hardly a limiting factor for the obtained *KGE′* in the calibration process. Therefore, to minimize the lack of model performance in the validation periods, the dry/wet trend of the precipitation estimate should be in general agreement with the dry/wet trend of the observed streamflow.
- The compensation of the precipitation characteristics between the calibration periods leads to different optimum parameter sets unless the model parameters reach the upper limits, which means that the catchments behavior is not captured within the parameter ranges. This often happened if the simulated streamflow overestimates the observed streamflow and the parameters GwLoss and CalEvap reach their upper limit to match the observed streamflow.
- The CMORPHV1.0, GSMaP and PERSIANN products are – in general – not suitable to drive hydrological modeling studies in the Southern African region with the LISFLOOD model as these products often have a poor correlation with the observed streamflow and generate much more rainfall compared with the other products, which results in large differences between simulated and observed streamflows. The best model performance is obtained with products which ingest gauge data for

bias correction (3B42V6, 3B42V7, ERAIGPCP and RFE2.0 products) resulting in the best conditions in terms of the r, β and γ between simulated and observed streamflow.

- According to Vaze et al. (2010) and Li et al. (2012), models calibrated over dry periods perform better in wet periods than the other way round. These findings might be true (not exhaustively verified here due to low sample size) because of the common overestimation of observed streamflow in (semi-) arid areas in hydrological modelling (Vaze et al., 2010; Gallart et al., 2007; Perrin et al., 2007; Lidén and Harlin, 2000; Gan et al., 1997; Hughes, 1997). However, these findings do not rule out robust model parameters. According to our findings, model parameters obtained for dry catchments with occasional rainfall events and therefore a high intra- and interannual variability in streamflow (e.g., A4) are not able to capture the catchments' behavior. This may be due to the fact that the model parameters are highly influenced by a few events and therefore have a higher risk for errors when transferred to a different period. Moreover, the spatial resolution of the precipitation products might be too coarse (0.1 or 0.25°) for an accurate detection of local heavy precipitation events and, therefore, they do not capture well the volume and timing of the peak flows especially in a small catchment like region A4.
- Different calibration periods related to precipitation or streamflow variability (wet/dry) result in different optimum parameters obtained in the calibration phase, which reflects the effect of non-stationarity conditions on the parameter estimates (Beven, 1993). However, the parameter distribution tends to show more spread with "insensitive" model parameters compared to "sensitive" parameters in the time periods used for calibration (Fig. 8). For this reason along with the interaction between parameters, the model is able to produce good results with a different set of parameters, i.e. equifinality (Beven, 1993). Therefore, an estimation of parameter uncertainty and their lack of robustness would only make sense on "sensitive" model parameters.
- The behavior of catchments in Southern Africa, but probably also in other areas with highly variable climate conditions (e.g., Australia or the Iberian Peninsula) is difficult to capture using a single model parameter set for the whole simulation period. Transposing these parameters to other time periods (scenarios in climate impact studies) or areas (in regionalization studies) should be carried out with caution. We agree with the conclusion of Vaze et al. (2010) that model parameters obtained from a wet (dry) calibration period should be used for periods where a wetter (drier) future is predicted, which does not mean that the parameters are robust.

Conflict of interest

The authors declare that there are no conflict of interest.

Acknowledgments

This work was carried out within the framework of the FP7 EU project GLOWASIS. Observed streamflow data in this study have been obtained from the Department of Water Affairs in South Africa.

We thank Harald Kling and an anonymous reviewer for providing valuable comments to improve the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejrh.2016.09.003.

References

Abdelaziz, R., Zambrano-Bigiarini, M., 2014. Particle swarm optimization for inverse modeling of solute transport in fractured gneiss aquifer. J. Contam. Hydrol. 164, 285–298, http://dx.doi.org/10.1016/j.jconhyd.2014.06.003.

Alfieri, L., Thielen, J., Pappenberger, F., 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. J. Hydrol. 424–425, 143–153, http://dx.doi.org/10.1016/j.jhydrol.2011.12.038.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., 2012. All that glitters is not gold: the case of calibrating hydrological models. Hydrol. Process. 26, 2206–2210, http://dx.doi.org/10.1002/hyp.9264.

Artan, G., Gadain, H., Smith, J.L., Asante, K., Bandaragoda, C.J., Verdin, J.P., 2007. Adequacy of satellite derived rainfall data for stream flow modeling. Nat. Hazards 43, 167–185, http://dx.doi.org/10.1007/s11069-007-9121-6.

Bárdossy, A., Das, T., 2008. Influence of rainfall observation network on model calibration and application. Hydrol. Earth Syst. Sci. 12, 77–89.

Bódis, K., 2009. Development of a Data Set for Continental Hydrologic Modelling. European Commission, Joint Research Centre, Ispra, http://dx.doi.org/10.2788/46925.

Balsamo, G., Boussetta, S., Lopez, P., Ferranti, L., 2010. Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A.

Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The european flood alert system EFAS – part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci. 13, 141–153.

Bartholomé, E., Belward, A.S., Achard, F., 2003. Global Land Cover 2000 database, European Commission, Joint Research Centre.

Behrangi, A., Khakbaz, B., Jaw, T.C., AghaKouchak, A., Hsu, K., Sorooshian, S., 2011. Hydrologic evaluation of satellite precipitation products over a mid-size basin. J. Hydrol. 397, 225–237, http://dx.doi.org/10.1016/j.jhydrol.2010.11.043.

Beven, K.J., 1993. Prophecy, reality, and uncertainty in distributed hydrological modeling. Adv. Water Resour. 16, 41–51, http://dx.doi.org/10.1016/0309-1708(93)90028-E.

Beven, K.J., 2004. Rainfall – Runoff Modelling: The Primer. John Wiley & Sons Ltd., New York.

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

Brauer, C.C., Teuling, A.J., Torfs, P.J.J.F., Uijlenhoet, R., 2014a. The Wageningen Lowland Runoff Simulator (WALRUS): a lumped rainfall-runoff model for catchments with shallow groundwater. Geosci. Model Dev. 7, 2313–2332, http://dx.doi.org/10.5194/gmd-7-2313-2014.
Brauer, C.C., Torfs, P.J.J.F., Teuling, A.J., Uijlenhoet, R., 2014b. The Wageningen Lowland Runoff Simulator (WALRUS): application to the hupsel brook catchment and the cabauw polder. Hydrol. Earth Syst. Sci. 18, 4007–4028, http://dx.doi.org/10.5194/hess-18-4007-2014.
Bulygina, N., Gupta, H., 2009. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. 45, 1–20, http://dx.doi.org/10.1029/2007WR006749.
Burek, P., Van Der Knijff, J., De Roo, A., 2013. LISFLOOD, distributed water balance and flood simulation model – Revised user manual 2013.
Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266, http://dx.doi.org/10.1016/j.jhydrol.2004.03.042.
Chiew, F.H.S., Teng, J., Vaze, J., Post, D.A., Perraud, J.M., Kirono, D.G.C., Viney, N.R., 2009. Estimating climate change impact on runoff across southeast Australia: method, results, and implications of the modeling method. Water Resour. Res. 45, 1–17, http://dx.doi.org/10.1029/2008WR007338.
Cohen Liechti, T., Matos, J.P., Boillat, J.-L., Schleiss, A.J., 2012. Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin. Hydrol. Earth Syst. Sci. 16, 489–500, http://dx.doi.org/10.5194/hess-16-489-2012.
Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F., 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. Water Resour. Res. 48, 1–17, http://dx.doi.org/10.1029/2011WR011721.
Coron, L., Andréassian, V., Perrin, C., Bourqui, M., Hendrickx, F., 2014. On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrol. Earth Syst. Sci. 18, 727–746, http://dx.doi.org/10.5194/hess-18-727-2014.
Dankers, R., Feyen, L., 2008. Climate change impact on flood hazard in Europe: an assessment based on high-resolution climate simulations. J. Geophys. Res. Atmos. 113, 1–17, http://dx.doi.org/10.1029/2007JD009719.
Dankers, R., Feyen, L., 2009. Flood hazard in Europe in an ensemble of regional climate scenarios. J. Geophys. Res. Atmos. 114, 1–16, http://dx.doi.org/10.1029/2008JD011523.
De Roo, A., Wesseling, C.G., Van Deursen, W.P.A., 2000. Physically based river basin modelling within a GIS: the LISFLOOD model. Hydrol. Process. 14, 1981–1992, http://dx.doi.org/10.1002/1099-1085(20000815/30)14:11/12<1981:AID-HYP49>3.0.CO;2-F.
De Roo, A., Thielen, J., Salamon, P., Bogner, K., Nobert, S., Cloke, H., Demeritt, D., Younis, J., Kalas, M., Bódis, K., Muraro, D., Pappenberger, F., 2011. Quality control, validation and user feedback of the European Flood Alert System (EFAS). Int. J. Digit. Earth 4, 77–90, http://dx.doi.org/10.1080/17538947.2010.510302.
Dee, D.P., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteor. Soc. 137, 553–597 http://dx.doi.org/10.1016/S0022-1694(99)00056-6.
Donnelly-Makowecki, L.M., Moore, R.D., 1999. Hierarchical testing of three rainfall-runoff models in small forested catchments. J. Hydrol. 219, 136–152, http://dx.doi.org/10.1016/S0022-1694(99)00056-6.
Dyson, L.L., Van Heerden, J., 2001. The heavy rainfall and floods over the northeastern interior of South Africa during February 2000. S. Afr. J. Sci. 97, 80–86.
Ebtehaj, M., Moradkhani, H., Gupta, H.V., 2010. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling. Water Resour. Res. 46, 1–14, http://dx.doi.org/10.1029/2009WR007981.
Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrol. Sci. J. 55, 58–78, http://dx.doi.org/10.1080/02626660903526292.
Fenicia, F., Savenije, H.H.G., Avdeeva, Y., 2009. Anomaly in the rainfall-runoff behaviour of the Meuse catchment. Climate, land use, or land use management? Hydrol. Earth Syst. Sci. 13, 1827–1837.
Feyen, J., Shannon, K., Neville, M., 2009. Water and urban development paradigms: towards an integration of engineering, design and management approaches. In: Proceedings of The International Urban Water Conference John Wiley & Sons, Ltd. Heverlee, Proceedings of the International Urban Water Conference, Heverlee, Belgium, 15–19 September, 2008, p. 694.
Gallart, F., Latron, J., Llorens, P., Beven, K., 2007. Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions. Adv. Water Resour. 30, 808–823, http://dx.doi.org/10.1016/j.advwatres.2006.06.005.
Gan, T.Y., Dlamini, E.M., Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. J. Hydrol. 192, 81–103.
Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G., 2013. An approach to identify time consistent model parameters: sub-period calibration. Hydrol. Earth Syst. Sci. 17, 149–161, http://dx.doi.org/10.5194/hess-17-149-2013.
Gorgens, A.H.M., Hughes, D.A., 1982. Synthesis of streamflow information relating to the semi-arid Karoo Biome of South Africa. S. Afr. J. Sci. 78, 58–68.
Gourley, J.J., Hong, Y., Flamig, Z.L., Wang, J., Vergara, H., Anagnostou, E.N., 2011. Hydrologic evaluation of rainfall estimates from radar, satellite, gauge, and combinations on ft. Cobb Basin, Oklahoma. J. Hydrometeorol. 12, 973–988, http://dx.doi.org/10.1175/2011JHM1287.1.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91, http://dx.doi.org/10.1016/j.jhydrol.2009.08.003.
Hartmann, G., Bárdossy, A., 2005. Investigation of the transferability of hydrological models and a method to improve model calibration. Adv. Geosci. 5, 83–87, http://dx.doi.org/10.5194/adgeo-5-83-2005.
Herman, A., Kumar, V.B., Arkin, P.A., Kousky, J.V., 1997. Objectively determined 10-day African rainfall estimates created for famine early warning systems. Int. J. Remote Sens. 18, 2147–2159, http://dx.doi.org/10.1080/014311697217800.
Hsu, K., Gao, X., Sorooshian, S., Gupta, H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteorol. 36, 1176–1190, http://dx.doi.org/10.1175/1520-0450(1997)036<1176:pefrsi>2.0.co;2.
Huffman, G.J., Bolvin, D.T., 2012. TRMM and other Data Precipitation Data set Documentation.
Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38–55, http://dx.doi.org/10.1175/JHM560.1.
Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., 2010. The TRMM multi-satellite precipitation analysis (TMPA). Satell. Rainfall Appl. Surf. Hydrol., 3–22, http://dx.doi.org/10.1007/978-90-481-2915-7_1.
Hughes, D.A., Kapangaziwiri, E., Sawunyama, T., 2010. Hydrological model uncertainty assessment in southern Africa. J. Hydrol. 387, 221–232, http://dx.doi.org/10.1016/j.jhydrol.2010.04.010.
Hughes, D.A., 1997. Southern African FRIEND–The Application of Rainfall-Runoff Models in the SADC Region, Water Research Commission Report. Pretoria, South Africa.
Hughes, D.A., 2006. Water resources estimation in less developed regions—issues of uncertainty associated with a lack of data. Predict. Ungauged Basins Promise Prog. 72–79.
Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the Globe Version 4, available at CGIAR-CSI SRTM 90 m Database: http://srtm.csi.cgiar.org (accessed 14.06.16.).
Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5, 487–503, http://dx.doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.
Kavetski, D., Franks, S.W., Kuczera, G., 2003. Calibration of watershed models. In: Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Water Science and Application. American Geophysical Union, pp. 49–68, http://dx.doi.org/10.1029/WS006.
Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1 Theory. Water Resour. Res. 42, 1–9, http://dx.doi.org/10.1029/2005WR004368.
Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J. 31, 13–24, http://dx.doi.org/10.1080/02626668609491024.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424–425, 264–277, http://dx.doi.org/10.1016/j.jhydrol.2012.01.011.
Li, C.Z., Zhang, L., Wang, H., Zhang, Y.Q., Yu, F.L., Yan, D.H., 2012. The transferability of hydrological models under nonstationary climatic conditions. Hydrol. Earth Syst. Sci. 16, 1239–1254, http://dx.doi.org/10.5194/hess-16-1239-2012.
Lidén, R., Harlin, J., 2000. Analysis of conceptual rainfall-runoff modelling performance in different climates. J. Hydrol. 238, 231–247, http://dx.doi.org/10.1016/S0022-1694(00)00330-9.
Lin, Z., Beck, M.B., 2012. Accounting for structural error and uncertainty in a model: an approach based on model parameters as stochastic processes. Environ. Model. Softw. 27–28, 97–111, http://dx.doi.org/10.1016/j.envsoft.2011.08.015.
Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: implications for climate impact analyses. Water Resour. Res. 47, 1–17, http://dx.doi.org/10.1029/2010WR009505.
Moulin, L., Gaume, E., Obled, C., 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrol. Earth Syst. Sci. 13, 99–114.
Mubareka, S., Maes, J., Lavalle, C., De Roo, A., 2013. Estimation of water requirements by livestock in Europe. Ecosyst. Serv. 4, 139–145, http://dx.doi.org/10.1016/j.ecoser.2013.03.001.
Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J. Hydrol. 320, 62–83, http://dx.doi.org/10.1016/j.jhydrol.2005.07.016.
Pappenberger, F., Thielen, J., Del Medico, M., 2011. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System. Hydrol. Process. 25, 1091–1113, http://dx.doi.org/10.1002/hyp.7772.
Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., Mathevet, T., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. Hydrol. Sci. J. 52, 131–151, http://dx.doi.org/10.1623/hysj.52.1.131.
Petersen, T., Devineni, N., Sankarasubramanian, A., 2012. Seasonality of monthly runoff over the continental United States: causality and relations to mean annual and mean monthly distributions of moisture and energy. J. Hydrol. 468–469, 139–150, http://dx.doi.org/10.1016/j.jhydrol.2012.08.028.
Ramos, M.-H., Thielen, J., Bartholmes, J., 2007. EFAS EPS-based forecasts?: in-depth case studies analyses & statistical evaluation of summer 2005 and spring 2006 flood forecasts, 3rd HEPEX Workshop. Stresa.
Reason, C.J.C., Keibel, A., 2004. Tropical cyclone eline and its unusual penetration and impacts over the southern african mainland. Weather Forecast 19, 789–805, http://dx.doi.org/10.1175/1520-0434(2004)019<0789:TCEAIU>2.0.CO;2.
Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resour. Res. 32, 2189–2202, http://dx.doi.org/10.1029/96WR00896.
Reusser, D.E., Zehe, E., 2011. Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. Water Resour. Res. 47, 1–15, http://dx.doi.org/10.1029/2010WR009946.
Rojas, R., Feyen, L., Bianchi, A., Dosio, A., 2012. Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. J. Geophys. Res. Atmos. 117, http://dx.doi.org/10.1029/2012jd017461.
Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270, http://dx.doi.org/10.1016/j.cpc.2009.09.018.
Seibert, J., 2003. Reliability of model predictions outside calibration conditions. Nord. Hydrology 34, 477–492, http://dx.doi.org/10.2166/nh.2003.028.
Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrol. Earth Syst. Sci. 16, 1171–1189, http://dx.doi.org/10.5194/hess-16-1171-2012.
Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H., Vogt, J., 2012. Development of a combined drought indicator to detect agricultural drought in Europe. Nat. Hazards Earth Syst. Sci. 12, 3519–3531, http://dx.doi.org/10.5194/nhess-12-3519-2012.
Siebert, S., Döll, P., 2008. The Global Crop Water Model (GCWM): Documentation and first results for irrigated crops, Frankfurt Hydrology Paper. Frankfurt.
Siebert, S., Döll, P., 2010. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. J. Hydrol. 384, 198–217, http://dx.doi.org/10.1016/j.jhydrol.2009.07.031.
Silal, S.P., Little, F., Barnes, K.I., White, L.J., 2015. Predicting the impact of border control on malaria transmission: a simulated focal screen and treat campaign. Malar. J. 14, 268, http://dx.doi.org/10.1186/s12936-015-0776-2.
Simmons, A., Uppala, S., Dee, D., Kobayashi, S., 2007. ERA-Interim: New ECMWF reanalysis products from 1989 onwards, ECMWF Newsletter. ECMWF Newsletter n.110.
Sobol', I.M., 2007. Global sensitivity analysis indices for the investigation of non linear mathematical models. Mat. Model. 19 (11), 23–24.
Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. Water Resour. Res. 43, 1–18, http://dx.doi.org/10.1029/2006WR005032.
Stisen, S., Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. Hydrol. Process. 24, 879–891, http://dx.doi.org/10.1002/hyp.7529.
Sun, X., Mein, R.G., Keenan, T.D., Elliott, J.F., 2000. Flood estimation using radar and raingauge data. J. Hydrol. 239, 4–18, http://dx.doi.org/10.1016/S0022-1694(00)00350-4.
Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., 2009. The European flood alert system – part 1: concept and development. Hydrol. Earth Syst. Sci. 13, 125–140.
Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A., 2013. Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. J. Hydrol. 499, 324–338, http://dx.doi.org/10.1016/j.jhydrol.2013.07.012.
Thiemig, V., Bisselink, B., Pappenberger, F., Thielen, J., 2015. A pan-African medium-range ensemble flood forecast system. Hydrol. Earth Syst. Sci. 19, 3365–3385, http://dx.doi.org/10.5194/hess-19-3365-2015.
Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., Van Beek, L.P.H., Uhlenbrook, S., 2014. Comparison of different evaporation estimates over the African continent. Hydrol. Earth Syst. Sci. 18, 193–212, http://dx.doi.org/10.5194/hess-18-193-2014.
Tyson, P.D., Preston-Whyte, R.A., Schulze, R.E., 1976. The Climate of the Drakensberg. Pietermaritzburg.
Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Kachi, M., Oki, R., Morimoto, T., Kawasaki, Z.-I., 2009. A Kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. J. Meteorol. Soc. Jpn. 87A, 137–151, http://dx.doi.org/10.2151/jmsj.87a.137.
Van Der Knijff, J.M., Younis, J., De Roo, A., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. Int. J. Geogr. Inf. Sci. 24, 189–212, http://dx.doi.org/10.1080/13658810802549154.
Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J., 2010. Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies. J. Hydrol. 394, 447–457, http://dx.doi.org/10.1016/j.jhydrol.2010.09.018.
Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol. Process. 17, 455–476, http://dx.doi.org/10.1002/hyp.1135.
Walmsley, R.D., Walmsley, J.J., Consultants, M., Silberbauer, M., D. of affairs and Forestry, 1999. Freshwater systems and resources.
Wilby, R.L., 2005. Uncertainty in water resource model parameters used for climate change impact assessment. Hydrol. Process. 19, 3201–3219, http://dx.doi.org/10.1002/hyp.5819.
Wu, H., Kimball, J.S., Mantua, N., Stanford, J., 2011. Automated upscaling of river networks for macroscale hydrological modeling. Water Resour. Res. 47, 1–18, http://dx.doi.org/10.1029/2009WR008871.
Xu, C., 1999. Operational testing of a water balance model for predicting climate change impacts. Agric. For. Meteorol. 98–99, 295–304, http://dx.doi.org/10.1016/S0168-1923(99)00106-9.

Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent particle swarm optimisation software for model calibration. Environ. Model. Softw. 43, 5–25, http://dx.doi.org/10.1016/j.envsoft.2013.01.004.
Zambrano-Bigiarini, M., Clerc, M., Rojas, R., 2013. Standard particle swarm optimisation 2011 at CEC-2013: a baseline for future PSO improvements. 2013 IEEE Congr. Evol. Comput., 2337–2344, http://dx.doi.org/10.1109/CEC.2013.6557848.
Zambrano-Bigiarini, M., 2013. Particle Swarm Optimisation, with focus on Environmental Models.
Zhan, C., Song, X., Xia, J., Tong, C., 2013. An efficient integrated approach for global sensitivity analysis of hydrological model parameters. Environ. Model. Softw. 41, 39–52, http://dx.doi.org/10.1016/j.envsoft.2012.10.009.
Zhao, R.J., Liu, X.R., 1995. The Xinanjiang model. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, USA, pp. 215–232.