Remote Sensing of Environment 171 (2015) 118–131

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent

Beatriz Revilla-Romero [a,b,*], Hylke E. Beck [a], Peter Burek [a,c], Peter Salamon [a], Ad de Roo [a,b], Jutta Thielen [a]

[a] European Commission, Joint Research Centre, Ispra, Italy
[b] Utrecht University, Faculty of Geosciences, Utrecht, the Netherlands
[c] International Institute of Applied Systems Analysis, Laxenburg, Austria

ARTICLE INFO

Article history:
Received 2 April 2015
Received in revised form 13 September 2015
Accepted 23 October 2015
Available online xxxx

Keywords:
Remote sensing
Streamflow
Global hydrology
Model calibration
Global Flood Detection system (GFDS)
LISFLOOD
Global Flood Awareness System (GloFAS)

ABSTRACT

Calibration is a crucial step in the application of hydrological models and is typically performed using in situ streamflow data. However, many rivers on the globe are ungauged or poorly gauged, or the gauged data are not readily available. In this study, we used remotely-sensed surface water extent from the Global Flood Detection System (GFDS) as a proxy for streamflow, and tested its value for calibration of the distributed rainfall-runoff routing model LISFLOOD. In a first step, we identified 30 streamflow gauging sites with a high likelihood of reliable GFDS data. Next, for each of these 30 sites, the model parameters related to groundwater and routing were independently calibrated against in situ and GFDS-derived streamflow time series, and against the raw GFDS surface water extent time series. We compared the performance of the three calibrated and the uncalibrated model simulations in terms of reproducing the in situ streamflow time series. Furthermore, we calculated the gain achieved by each scenario that used satellite-derived information relative to the reference uncalibrated scenario and the one that used in situ data.

Results show that using the raw GFDS data as a proxy for streamflow for calibration improved the skill of the simulated streamflow (in particular the high flows) for 21 of the 30 sites using correlation as a metric. Furthermore, we discuss a calibration strategy using a combination of in situ and satellite data for global hydrological models.

1. Introduction

Hydrological models are indispensable tools for increasing our understanding of the hydrological cycle, for assessing the hydrological implications of climate and land-use change, and for flood and drought forecasting. Model calibration and validation are essential prerequisites to obtain reliable streamflow estimates from hydrological models (Minville et al., 2014; Werner, Blazkova, & Petr, 2005) and are typically performed using in situ streamflow data. However, in situ streamflow data are unavailable for the majority of the land surface, particularly for the most flood vulnerable countries, and the number of operational stations is rapidly decreasing (Hannah et al., 2011; Sivapalan, 2003; Wohl et al., 2012).

Satellite remote sensing has the ability to provide information on hydrological fluxes and state variables at (near-)global coverage and (near-)real time, and at frequent temporal intervals, and as such provides unique opportunities for enhancing model simulations in remote areas (van Dijk & Renzullo, 2011). In recent decades, there have been increasing efforts to improve models by incorporating remotely-sensed data on hydrologic variables such as evaporation (Zhang, Chiew, Zhang, & Li, 2009), surface soil moisture (Beck, De Jeu, Schellekens, Van Dijk, & Bruijnzeel, 2009; Hirpa, Gebremichael, Hopson, Wojick, & Lee, 2014; Wanders, Bierkens, de Jong, de Roo, & Karssenberg, 2014), total water storage (Jiang et al., 2014; Ramillien, Famiglietti, & Wahr, 2008), and snow and ice cover (Bergeron, Royer, Turcotte, & Roy, 2014; Dietz, Kuenzer, Gessner, & Dech, 2012), as well as vegetation-related variables such as leaf area index (LAI; Zhang, Vaze, Chiew, & Liu, 2011) and normalized difference vegetation index (NDVI; Donohue, Roderick, & McVicar, 2007). In addition, various studies have examined the value of remotely-sensed variables related to surface water, including inundation extent, river width, and water levels (see overview in Table 1). Some of these studies used satellite-derived information to calibrate model parameters or to derive empirical rating curves, while others focused on the direct use of changes in water level or width for calibration, using a variety of hydrodynamic or hydrological models. Although these studies reported promising results, a major drawback from the point of view of global hydrological modelling is that they often focused on a small region. Most studies focused on single river reaches and the validation was performed using a relatively small number of gauges ($n < 6$) (Di Baldassarre, Schumann, & Bates, 2009; Domeneghetti et al., 2014; Hostache et al., 2009; Mason, Bates, & Dall' Amico, 2009; Milzow, Krogh, & Bauer-Gottwein, 2011; Montanari et al., 2009; Sun, Ishidaira, & Bastola, 2010, 2012a, 2012b;

* Corresponding author at: Via Enrico Fermi 2749 TP 122, I-21027 Ispra, Italy.
E-mail address: b.revillaromero@gmail.com (B. Revilla-Romero).

http://dx.doi.org/10.1016/j.rse.2015.10.022

Table 1
Summary of relevant studies where satellite-derived inundation extent, river width, and/or water levels were used to calibrate a hydrodynamic and/or hydrological model. Studies are listed in alphabetical order of author.

	Study	Satellite and sensor/Acquisition frequency	Model	Study area and no. of in situ river gauges used	Study period	Objective/Approach	Key findings
1	Di Baldassarre et al. (2009)	ENVISAT ASAR and ERS-2 SAR (1 image each used)	LISFLOOD-FP	10-km of the Dee and 8-km of the Alyn River (UK) (3 gauges)	Dec. 2006 flood event	Use inundation map derived from satellite imagery	Need to move from deterministic binary wet/dry maps to probabilistic fractional flood extent maps
2	Domeneghetti et al. (2014)	ENVISAT and ERS-2/35-day	HEC-RAS	~140 km reach of the Po River (northern Italy) (2 gauges)	16 years	Identification of Manning's roughness coefficient	Combine satellite time series with hydrometric data to increase the reliability of the hydraulic model
3	Getirana (2010)	ENVISAT altimetric data/35 days	MGB-IPH model	Branco River (northern Amazon Basin, Brazil) (11 gauges)	10 years	Use empirical equations to estimate river depth from modelled streamflow	Pioneering use of spatial altimetry data in the automatic calibration of hydrological models
4	Getirana et al. (2013)	ENVISAT altimetric data/35 days	HyMAP (based on CaMa-Flood + ISBA)	Amazon Basin (Brazil) (4 gauges)	10 years	Use of altimetric data to calibrate four parameters: the subsurface runoff time delay, Manning's roughness coefficient, river width, and bankfull height	Demonstrated use of altimetric data in the automatic calibration of model parameters
5	Hostache et al. (2009)	ENVISAT SAR/A few days to 30 days (1 image used)	HEC-RAS	18-km reach of the Alzette River (Luxembourg) (6 gauges)	Jan 2003 flood event	Estimation of spatially distributed water levels from remote-sensing observations and integration of satellite-data information in a hydraulic model in order to reduce model uncertainties	It provides distributed water levels with a high spatial density and provides more reliable hydraulic models thanks to these water levels that allow a spatial evaluation of model performances
6	Mason et al. (2009)	ERS-1 SAR (1 image used)	LISFLOOD-FP	12 km of the Thames River (UK) (No gauges, comparison with LiDAR)	1992 flood event	Comparison of water level probabilities based on flood waterline estimates derived from both SAR and LiDAR data to that derived from only SAR data	The use of water levels unable to restrict the parameter range of acceptable model runs and hence reduce the number of runs necessary to generate a flood inundation uncertainty map
7	Milzow et al. (2011)	ERS-2, and ENVISAT SAR altimetric data/35 days	SWAT	Okavango catchment (Southern Africa) (3 gauges)	11 years	Satellite altimetry was used to derive water level fluctuations at three locations in the catchment for channels approximately 150 m wide	The combination of multiple independent observational datasets improves the parametrization of the hydrological model
8	Montanari et al. (2009)	ERS-2 and ENVISAT SAR altimetric data/35 days (1 image each used)	Nash IUH + HEC-RAS	Alzette River basin (Luxembourg) (2 gauges)	Jan. 2003 flood event	Assessment of the value of remote-sensed water levels in an aggregated modelling system	A first step toward a systematic remote sensing-based surface water monitoring system that may quasi-continuously provide valuable information for sequentially updating coupled H–H models
9	Sun et al. (2012a)	JERS-1 SAR images (1 image used)	HYdrological MODel (HYMOD)	Mekong River (Asia) at Pakse (1 gauge)	4 years	Develop a conceptual framework for rainfall-runoff models in ungauged basins using satellite-derived river width or water surface elevation	Illustrates that under both the average and low designed satellite observational frequencies, the simulated streamflow had an acceptable degree of accuracy
10	Sun et al. (2012b)	TOPEX/Poseidon (T/P)/~10 days	HYMOD	Upper Mississippi Basin at Clinton (US) (1 gauge)	5 years	Attempt to use water surface elevation data obtained from satellite radar altimeters	Comparison with calibration using streamflow data shows that for the new calibration method, the uncertainty in the modelling process is higher, and the parameter space is less constrained
11	Sun et al. (2010)	JERS-1 SAR (16 images used)	HYMOD	Mekong River (Asia) at Pakse (1 gauge)	4 years	Minimize the difference between river widths observed from space and simulated widths by tuning parameters of rainfall-runoff model and at-a-station hydraulic geometry relation simultaneously	Wide applicability for reproducing river streamflow time series on the daily scale in ungauged basin if satellite data is available
12	Tarpanelli et al. (2013)	ENVISAT ASAR (1 image used)	Modello Idrologico Semi-Distribuito (MISD)	20-km reach of Genna stream (Italy) (1 gauge)	6 flood studies (1 day long each)	Calibration of Manning's roughness coefficient by comparing the flooded areas derived from ASAR imagery and hydrologic- hydraulic modelling.	The assessment of the Manning's roughness for the main channel and the floodplain is obtained from hydraulic simulations and satellite data for a flood event on a small basin (90km^2).
14	This study	Passive microwave (AMSR-E, TRMM)/daily	LISFLOOD (GloFAS System)	Africa, Europe, North America, and South America (30 gauges)	13 years	Assess the value of satellite-derived surface water extent for calibration of a hydrological model	Improvement in the timing of the simulated flow peak, and of the model skill in term of volume accuracy using satellite-derived river streamflow, when available

Tarpanelli, Brocca, Melone, & Moramarco, 2013), while others applied the calibration to more locations ($n < 12$) but all within the same catchment (Getirana, 2010; Getirana, Boone, Yamazaki, & Mognard, 2013). In addition, the calibration periods used in these studies are not long enough, if the aim is to improve rainfall-runoff simulations of models at multi-decadal time scales, since in many cases the focus has been on specific flood events of short duration. In these studies, the satellite data were derived from multiple sensors, and ranges of the electromagnetic spectrum, and at different spatial and temporal resolutions. The limitations described are in most cases due to the infrequent overpass

cycle of satellites and/or the high cost of purchasing data at a sufficiently high resolution.

In this research, we used data products derived from the Global Flood Detection System (GFDS; Kugler & De Groeve, 2007) due to its capability for detecting and mapping major river floods worldwide in near-real time. GFDS uses a range of passive microwaves sensors to retrieve surface water extent changes. Such data products have the potential to be linked to a flood forecasting system like the Global Flood Awareness System (GloFAS; Alfieri et al., 2013). The mission of both systems is to be a supporting tool in the context of large flood events for decision makers, including national and regional water authorities, water resource managers, civil protection and first line responders, and international humanitarian aid organizations. Furthermore, both GFDS and GloFAS have the same daily temporal resolution and approximately the same spatial resolution (0.09° and 0.1°, respectively). GloFAS has been pre-operational since 2011, delivering global probabilistic streamflow forecasts. After running the system for three years, and acquiring sufficient data to compare against observations, the model has proven its value in forecasting large-scale floods events around the world. However, like the majority of global hydrological models, GloFAS has not been calibrated yet, and a calibration can potentially further improve its forecast skill.

The aim of this study is to assess the potential of using remotely-sensed GFDS surface water extent to improve hydrological model simulations through calibration of large-scale catchments having a broad range of morphologic, river, climatic, physical, and vegetation characteristics located in Africa, Europe, North America, and South America. The specific objectives of this study are: (1) to assess whether calibration using only satellite-derived passive microwave data improves the skill of the simulated streamflow; (2) to quantify the gain in skill using various calibration scenarios; and (3) to summarize the benefits and limitations of calibration using satellite-derived data.

2. Data and study regions

2.1. Satellite-derived data

The GFDS system (http://www.gdacs.org/flooddetection/) measures daily water surface extent changes using a range of satellite-based passive microwave sensors. The method uses the difference in brightness temperature at a frequency of 36.5 GHz between water and land surface to detect the proportion of within-pixel water and land (Kugler & De Groeve, 2007). The retrieved changes in brightness temperature are first gridded into a product with a pixel size of 0.09°×0.09°, and the system then provides a daily output. Satellites never pass over the same track at exactly the same time and therefore the operational GFDS applies a four-day forward-running mean to compute the final value, which also fills any gaps present in the record (Kugler & De Groeve, 2007). The GFDS flood-merged product currently (i.e. since January 2015) uses data from both the Advanced Microwave Scanning Radiometer 2 (AMSR-2) and the Global Precipitation Mission (GPM) sensors. However, GFDS used data from the Tropical Rainfall Measuring Mission (TRMM) for the period 1998–2014 and from the AMSR — Earth Observing System (AMSR-E) for the period June 2002–October 2011. GFDS data outputs are available since 1998 for latitudes 38°S–38°N. From June 2002 after the inclusion of AMSR-E data the coverage was expanded to latitudes 90°S–90°N. Brakenridge, Nghiem, Anderson, and Mic (2007); Brakenridge, Nghiem, Anderson, and Chien (2005) and Revilla-Romero, Thielen, Salamon, De Groeve, and Brakenridge (2014) demonstrated that the GFDS data can be used to monitor streamflow, for multiple catchments around the globe. For this study, we employed both the daily raw GFDS signal and GFDS estimated streamflow time series. To obtain GFDS estimated streamflow, we used in situ streamflow data to translate the dimensionless signal into volumetric streamflow units for each studied site. Regression equations were obtained using monthly means from daily values, and GFDS-measured streamflow was derived as follows:

$$Q_{\text{GFDS measured of X month}} = a_{\text{month}} + b_{\text{month}} \cdot \text{signal}. \quad (1)$$

A full description of the methodology can be found in Revilla-Romero et al. (2014). Data from in situ gauges are needed for this and therefore this approach cannot be used at locations that are completely ungauged. However, these datasets can fill data gaps or extend the records of discontinued gauges.

2.2. In situ streamflow data

Daily in situ streamflow time series were obtained from the Global Runoff Data Centre (GRDC, 2010) for gauging stations fulfilling the criteria presented in Section 3.1.

2.3. Reference climatology and input runoff forcing

To apply the rainfall-runoff model we used offline simulations from ERA-Interim/Land (Balsamo et al., 2013), a global reanalysis of land-surface meteorological variables from 1980 to 2013 at a spatial resolution of 80-km. ERA-Interim/Land is the result of a land-surface model simulation using the HTESSEL land surface scheme (Balsamo et al., 2009), with meteorological forcing from ERA-Interim and precipitation adjustments based on the Global Precipitation Climatology Project (GPCP) v2.2. (Adler et al., 2003; Huffman, Adler, Bolvin, & Gu, 2009). HTESSEL is the land-surface scheme of the operational ECMWF weather model Integrated Forecast System (IFS). Output datasets from HTESSEL are used within the current GloFAS model setup (further details in Section 3.2.1).

3. Methodology

3.1. Selection of testing sites

We selected test sites based on Revilla-Romero et al. (2014), which examined how site morphologic, hydrologic, climatic, physical, and vegetation characteristics influence the quality of the GFDS signal which can be influenced by ground conditions such as presence of snow/ice, land cover type, and topography (Brakenridge et al., 2012). For example, the canopy and crown water content from closed forests type can influence the microwave emission properties (Chukhlantsev, 2006). Furthermore, the capacity of the GFDS to detect changes in streamflow volume also depends on local characteristics of the river such as the mean daily streamflow, floodplain size, and channel cross-section (Revilla-Romero et al., 2014). Additionally, the presence of hydraulic control measures such as regulated dams can cause a disagreement between the simulated streamflow that measured in situ, if the impacts of the dams are not accounted for by the hydrological model. In such cases, further research is needed to quantify the added value of GFDS since the flow regimes are influenced by the non-natural factors.

From our initial set of 322 gauging stations, we used the following criteria for the GFDS grid cell containing the gauging station, in order to exclude or include sites with features that can affect the quality of the GFDS signal or complicate the simulation of the streamflow:

1. Exclude sites that have, closely upstream dams/hydraulic control measures (GRanD Database; Lehner, Verdin, & Jarvis, 2008), land cover classified as “artificial” or “closed forest” (Globcover; Bontemps et al., 2010), or “Intermediate wetland/lakes” wetland type (Global Lakes and Wetlands Database; Lehner & Döll, 2004). This reduced the dataset from 322 to 240 stations.
2. Include all sites that are located within grid cells classified as inundated area by the Global Inundation Extent from Multi-Satellites

(GIEMS, Prigent et al., 2012), which has a resolution of 0.25°× 0.25°, based on the calculated maximum extent from 1998 to 2007, and that meet either of the following two criteria:

2.1. At least 50% of the GFDS grid cell is classified as "floodplain and freshwater marsh" or "swamp forest and flooded forest" wetland cover type by the Global Lakes and Wetlands Database (Lehner & Döll, 2004). Out of the 240 remaining stations, 30 fulfilled this criterion.

2.2. A mean daily streamflow greater than 500 $m^3 s^{-1}$. Revilla-Romero et al. (2014) showed that during normal flow conditions (i.e., not floods) the quality of the GFDS signal was lower (i.e., higher noise) on those rivers with mean daily streamflow less than 500 $m^3 s^{-1}$. However, during flood conditions, the GFDS signal is expected to perform well at those locations. The mean flow was calculated from in situ observation time series. However, if ungauged, this can be obtained from datasets such as the global composite of mean annual surface runoff (Fekete, Vörösmarty, & Grabs, 2002; Müller Schmied et al., 2014) or from hydrological modelling simulations. Out of the 240 remaining stations, 61 fulfilled this criterion, of which 21 stations also fulfil criterion 2.1.

Thus, 21 stations out the 240 locations fulfilled both criteria 2.1 and 2.2, a further 9 stations fulfil criterion 2.1 only, and a further 40 stations fulfil criterion 2.2 only, yielding 70 eligible sites in total. We randomly selected 30 out of 70 eligible sites in different climatic regions to test the usefulness of the GFDS signal for model calibration. Only 30 sites were considered as a proof of concept study, to optimize computational time. The 30 stations are distributed throughout South America (18), North America (5), Europe (4) and Africa (3) (Fig. 1). See Appendix A Table A1 for additional details for each site.

Additionally, we calculated the Topographic Index (TI, Beven & Kirkby, 1979) within our study sites using the global high-resolution (15 arc sec) map of Marthews, Dadson, Lehner, Abele, and Gedney (2015), which is based on the HydroSHEDS dataset (Lehner et al., 2008). TI is a measure of the relative propensity for the soil, at a given point, to become saturated at the surface, given the area that drains into it and its local outflow slope (Beven & Kirkby, 1979). Its values range from about 0 to 25, and are low at ridge tops and high in valleys along drainage paths and in zones of water concentration on the landscape (Wilson & Gallant, 2000). In Marthews et al. (2015), TI mean values were spatially calculated by wetlands type (Lehner & Döll, 2004). "Floodplains and freshwater marsh" mean value is 7.38 and for river pixels is 8.81. For this study, we carried out a spatial analysis based on both the performance of the GFDS obtained for the locations from Revilla-Romero et al. (2014), and the mean TI value within each location. We found that those sites with higher performance in terms of r have at least 16% of each satellite grid location covered by a TI value around 8 or higher.

The high resolution global inundation map (GIEMS-D15, Fluet-Chouinard, Lehner, Rebelo, Papa, & Hamilton, 2015) could also be used as a guidance for the selection of locations according to the GFDS signal quality. However, it was not available for distribution at the moment this study was carried out.

3.2. Hydrological model

3.2.1. LISFLOOD

LISFLOOD is a GIS-based spatially distributed hydrological model, which includes a one-dimensional channel routing model (Van Der Knijff, Younis, & De Roo, 2010). The LISFLOOD model has been used for flood forecasting (Forzieri et al., 2014; Thielen, Bartholmes, Ramos, & de Roo, 2009; Thiemig, Bisselink, Pappenberger, & Thielen, 2014), flood inundation modelling (Bates & De Roo, 2000), climate change impact assessments (Alfieri, Burek, Feyen, & Forzieri, 2015; Rodrigo Rojas, 2013), land use change impact assessments (De Roo, Odijk, Schmuck, Koster, & Lucieer, 2001; De Roo, Schmuck, Perdigao, & Thielen, 2003), and parameter uncertainty assessments (Feyen, Vrugt, Nualláin, van der Knijff, & De Roo, 2007), among many other purposes. For example, LISFLOOD is currently running within the European Flood Awareness System (EFAS) on an operational basis (Pappenberger, Thielen, & Del Medico, 2011; Thielen et al., 2009) for the entire Europe on a grid of 5 × 5 km spatial resolution. LISFLOOD is also used for the African Flood Forecasting System (AFFS; Thiemig, Bisselink, Pappenberger, & Thielen, 2015) at 0.1° spatial resolution. Within GloFAS, a combination of HTESSEL and a simplified setup of LISFLOOD is used at the global scale at a resolution of 0.1° (Fig. 2). This means that LISFLOOD uses surface runoff and sub-surface runoff fluxes from HTESSEL as model input, instead of using the basic meteorological data (precipitation, temperature, etc.) to force the model. LISFLOOD subsequently simulates groundwater storage, lakes and reservoirs, and the river routing processes. For this study, a 34-year (1980–2013) daily streamflow time series has been calculated using ERA-Interim/Land surface runoff.

To run the model, we used the improved river network obtained from Kimball, Li, Huang, Leung, and Adler (2012) and new global river width estimates from the Global River Width Database for Large Rivers (GRWD-LR, Yamazaki et al., 2014). As a result of this update of river widths, the "bank full" channel depth was recalculated using the same method as in the LISFLOOD setup used in Burek et al. (2013), while the Manning's roughness coefficient and channel slope remained the same. Other efforts such as the inclusion of lakes and reservoirs (Zajac, Salamon, Burek, De Roo, & Revilla-Romero, in preparation), and tests on the effect of reference climatology on the skill of the flood forecast (Hirpa et al., submitted for publication) are underway. Further details of the LISFLOOD model and the description of the equations can be found in Burek et al. (2013).

3.2.2. Calibration parameters, simulation scenarios and studied period

Similar to the calibration of LISFLOOD for Europe (Feyen et al., 2007; Zajac et al., 2013), the global setup needs calibration of the parameters to control percolation to the lower groundwater zone, the residence time of the upper and lower groundwater zone, and the routing parameters. The calibrated parameters are listed in Table 2. The parameter space was defined by physically reasonable lower and upper limits for each parameter. Initially, when the LISFLOOD model was setup for GloFAS in 2011 for the pre-operational launch, these parameters were carefully selected using our previous experience with developing and calibrating the model. In this study, however, we carried out parameter value estimation through model calibration for a selection of test sites. Note that this setup of LISFLOOD uses the surface and sub-surface runoff dataset from HTESSEL. Therefore, the input fluxes into the system remain equal, and the parameters that control the infiltration capacity, preferential bypass flow, and snow-melt coefficient will not be calibrated. We recognize that this is a drawback and a potential limitation of the calibration methodology. There are three different calibration scenarios (Table 3). The uncalibrated scenario is Qsim0, and all calibrated scenarios that use different calibrated data such as in situ observed streamflow, GFDS satellite-derived streamflow, and the raw GFDS surface extent signal are named Qsim1, Qsim2, and Qsim3, respectively. The performance of the calibration runs were compared with the uncalibrated (Qsim0) simulated streamflow of the LISFLOOD model embedded within GloFAS.

Note that due to the dimensionless character of the raw GFDS signal, and to avoid substantial changes in water volume while optimizing the flow peak, when using scenario Qsim3 the parameters that control the quantity of loss rate out of the lower zone (*GwLoss*) and the transmission loss, including simulation of evaporation, water extraction, and leaching (*TransSub*), are not calibrated, but instead these are set to the default values. For this scenario Qsim3, the inclusion of the two aforementioned parameters provides too many degrees of freedom because there is not enough information in the GFDS signal alone to constrain

Table 2
LISFLOOD calibration parameters, including upper and lower bound, and default value for the uncalibrated simulation setup. *Note that *GwLoss* and *TransSub* are not calibrated when using the raw GFDS signal (scenario Qsim3).

Parameter name	Description	Unit	Min	Max	Default
Tuz	Time constant for water in upper zone	D	3	40	20
Tlz	Time constant for water in lower zone	d	50	2500	1000
GwPerc	Maximum rate of percolation going from the Upper to the Lower response box	mm d^{-1}	0.01	2	0.5
GwLoss*	Maximum loss rate out of the Lower response box	mm	0	0.35	0
TransSub*	Linear transmission loss parameter (Rao & Maurer, 1996)	–	0	0.6	Estimated based on relationship with mean annual potential evaporation
CalChanMan	Multiplier applied to Channel Manning's n (bankfull routing)	–	0.1	15	4.0

them. As a consequence, it might take very high values in order to optimize the linear correlation between simulated and in situ streamflow and result in an important loss of flow volume. During our initial tests, this error was observed and was substantial, giving unrealistic streamflow simulations for some catchments such as the Amazon and Mississippi.

To define the calibration and validation periods for each catchment, we used the split-sample procedure (Refsgaard & Storm, 1990) for the period of simultaneous in situ and satellite-derived data, where the first half was used for validation and the second half for calibration. The first year of the calibration period was used as warm-up, and thus disregarded in computing the performance statistics. This means that for stations located below 38°N with sufficient in situ data, the calibration period is 2004–2010 with 2004 used as a warm-up, and the validation period is 1998–2003. For stations located above 38°N with sufficient in situ data, the calibration period was 2006–2010 with 2006 used as a warm-up, while the validation period is mid-2002-2005. The aforementioned period was modified when a particular site had less than 10 years in total of in situ data available during these periods, as was the case for the Zambezi River at Senanga, the Niger River at Lokoja, and the Parnaiba River at Luzilandia.

To initialise the groundwater storage of the model, we performed a pre-run for each simulation and parameter set in order to calculate the average recharge. Even though computational time increased, this was done because the time needed to initialise any storage component of the model is dependent on the average residence time of the water contained in it, and the warm-up period might not be sufficiently long for a correct initialisation of the storage.

3.3. Calibration procedure

The aim of model calibration is to obtain a set of parameters for which the model simulation outputs are as close as possible to the ground measurements. We test the use of both the raw remotely-sensed surface water extent signal ($GFDS_{raw}$) and the re-scaled signal (Q_{GFDS}) as a proxy for in situ measured streamflow. For all scenarios, we calibrated the model using an automatic process with the same model setup, input dataset (Section 2), range of parameter values, and calibration/validation periods (Section 3.2.2). We performed a single-objective calibration for the in situ observed streamflow (Qsim1) and GFDS estimated streamflow (Qsim2) to assess the skill of the model simulations using this optimisation method. In this case, we used the modified Kling–Gupta Efficiency (KGE′, Kling, Fuchs, & Paulin, 2012) as an objective function.

KGE′ is a performance indicator based on the equal weighting of linear correlation (r), bias ratio (β) and variability (γ), between simulated (s) and observed (o) streamflow values:

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

$$\beta = \frac{\mu_s}{\mu_o} \tag{2.b}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{2.c}$$

where CV is the coefficient of variation, μ is the mean streamflow [$m^3\ s^{-1}$], and σ is the standard deviation of the streamflow [$m^3\ s^{-1}$]. For KGE′ r, β, and γ have their optimum at unity. The KGE′ measures the Euclidean distance from the ideal point (unity) of the Pareto front, and is therefore able to provide an optimal solution which is simultaneously good for bias, flow variability, and correlation.

Our purpose was to implement the best calibration framework for the setup used for the LISFLOOD model as the intrinsic characteristics of the data permitted. Since the $GFDS_{raw}$ data contain only information on variability and timing, but not on magnitude of the flows in volumetric streamflow units, for Qsim3 we carried out a calibration using the Pearson linear correlation coefficient (r):

$$r = \frac{\sum_{i=1}^{n}(s_i-\mu_s)(o_i-\mu_o)}{\sqrt{\sum_{i=1}^{n}(s_i-\mu_s)^2}\sqrt{\sum_{i=1}^{n}(o_i-\mu_o)^2}} \tag{3}$$

where s is the simulated and o the observed streamflow values, which range from −1 to 1, with 1 being the optimum value.

Numerous studies have successfully applied evolutionary algorithms for calibrating hydrological models (Duan, Sorooshian, & Gupta, 1992; Maier et al., 2014; Nicklow et al., 2010). For the calibration of LISFLOOD we used the evolutionary algorithm implemented using the Distributed Evolutionary Algorithm in Python (DEAP) module (Fortin, De Rainville, Gardner, Parizeau, & Gagńe, 2012) using a population size of 12, a recombination pool size of 24, and crossover and mutation probabilities set to 0.9 and 0.1, respectively. The number of generations was set by trial-and-error to 30, which we found was sufficient to achieve convergence for all testing sites.

3.4. Performance comparison and skill quantification

The calibrated streamflow time series of each calibration scenario was compared with the uncalibrated streamflow. In addition, an

Table 3
Summary of simulation scenarios.

Scenario	Description	Objective function
Qsim0	Uncalibrated	–
Qsim1	Calibrated using Q_{ground}	KGE′
Qsim2	Calibrated using Q_{GFDS}	KGE′
Qsim3	Calibrated using $GFDS_{raw}$	Linear correlation (r)

evaluation against the in situ streamflow observations was also done. For this, the r, KGE′, Nash–Sutcliffe efficiency (NSE, Nash & Sutcliffe, 1970), and percentage of bias (PBIAS) were used. A description of the two latter validation metrics can be found in the Appendix A at the end of this paper. In addition, hydrographs, scatterplots, and obtained skill scores by the calibrated simulated time series were carefully analysed and compared with the uncalibrated simulated time series (Qsim0) for the period 1998–2010, and also with the in situ streamflow time series. A better performance in terms of reproducing in situ streamflow observations for Qsim3 than Qsim0 for the validation period would mean that the GFDS signal has value for that particular catchment.

Furthermore, to understand the added value of using the satellite-derived data (Q_{GFDS} and GFDS) for model calibration, we calculated the gain as follows:

$$Gain\ \ Q_{ground} = B - A \tag{4}$$

$$Gain\ \ Q_{GFDS} = C - A \tag{5}$$

$$Gain\ \ GFDS = D - A \tag{6}$$

$$Gain\ \ Q_{GFDS}\ \ vs\ \ Q_{ground} = Gain\ \ Q_{GFDS} - Gain\ \ Q_{ground} \tag{7}$$

where A, B, C, and D represents the skill scores obtained when comparing the scenario Qsim0, Qsim1, Qsim2, and Qsim3, respectively, to in situ measurements. This was carried out for the r and KGE′ score. At each location, a positive value means that either Qim1 (Eq. (4)), Qsim2 (Eq. (5)), or Qsim3 (Eq. (6)) scenario obtained a higher skill score than Qsim0, whereas a negative value means that the score decreased after calibration. Furthermore, we calculated the overall difference in gain between the scenarios that use Q_{GFDS} and Q_{ground} (Eq. (7)).

4. Results

4.1. Assessing the use of satellite derived surface water data for calibration

For each scenario, including the uncalibrated model, simulated streamflow time series were compared with the in situ measurements. Furthermore, skill scores (r, KGE, NSE, and PBIAS) obtained for each calibrated run (Qsim1, Qsim2, and Qsim3) were compared with the uncalibrated run (Qsim0). An example of the calibration optimisation technique that use the GFDS signal as input, compared with in situ measurement (Q_{ground}), and the simulated uncalibrated time series is presented in Fig. 3, for station G1156 (Amazonas River at Obidos-Linigrafo). We decided to illustrate the evaluation through this station for the richness of information that it provides. However, the Supplementary material (Table S1), includes tables with all the skill scores obtained for each station, for the uncalibrated and calibrated runs, during both the calibration and validation periods.

Looking at the figure, the time series for this location (G1156) reveal that the model exhibits a too-early bias in the streamflow seasonality, in particular for the uncalibrated scenario. After calibration, the timing of the streamflow at the peak improves whereas the timing at the low flows worsens. This response can be partially explained by the fact that, in this study we used the full time series of the GFDS signal, and therefore we also included the signal in low-flow periods. During the periods with lesser streamflow volume, the signal-to-noise ratio is lower, and therefore the signal might not be as accurate as during high flows or flood conditions. We also tested the possibility of applying a cut-off value of the signal at each site for calibration, in other words, to use only the signal above a certain threshold where the noise is minimal. However, for the current studied locations we could not find an optimal solution between the cut-off value and the remaining length of data available for calibration. In order to implement this, further research is needed. Therefore, as our main interest is on the performance at the peak flows, we opted to include the full time series from the GFDS signal. Furthermore, r is influenced by few large values, and the low flow might have changed as a side effect of the calibration. In addition, it is possible that when the streamflow volume starts to decrease, the signal retrieved from the satellite is still capturing some areas that remain wet for a longer time within the pixel under observation. This might occur when the flow has exceeded the bank-full channel capacity, but is also common in meandering and braided rivers. Additionally, this delay could be further influenced partly by the operational 4-day mean applied to the GFDS signal.

Due to the dimensionless character of the GFDS signal in this calibration scenario (Qsim3) we only use linear correlation as an optimisation function, but measurements of the goodness-of-fit such as KGE′ and NSE might also improve consequently. Meanwhile, the percentage of the bias (PBIAS) may keep the same value. This is because the parameters that control the quantity of loss rate out of the lower zone (*GwLoss*) and the transmission loss (*TransSub*), including simulation of evaporation, water extraction, and leaching, were not calibrated for scenario Qsim3, and therefore have the same values as Qsim0. However, for scenarios Qsim1 and Qsim2, the inclusion of the GwLoss and TransSub parameters in the calibration resulted in reductions in the bias for stations where the uncalibrated simulated streamflow was either substantially underestimated (e.g., G0729), or overestimated (e.g., G1090, G1177, G1197). As a consequence, if we also aim to improve the skill of the simulated volumes, and there is availability of historical in situ measurements, we could calibrate the model using the satellite-derived streamflow measurements (Q_{GFDS}) to extent the calibration period. An example of this can be found in the Supplementary material, for station G1177 (River Mearim at Bacabal). Both scenarios Qsim0 and Qsim3 obtained poor KGE, NSE, and PBIAS scores, whereas the skill of the simulated streamflow improved on the calibration for Qsim1 and Qsim2.

In order to determine how well the simulated streamflow that uses GFDS estimated streamflow (Qsim2) and GFDS raw signal (Qsim3) correspond with the simulation that uses in situ observed streamflow (Qsim1) and the uncalibrated simulation (Qsim0), we used box-plots (Fig. 4). Here the results of the correlation coefficient (r), the KGE′, and the NSE obtained by comparing all simulated time series against in situ streamflow time series, are presented. All the obtained individual values, also for PBIAS, are displayed in the Supplementary material. Looking at the linear correlation (r) scores (Fig. 4, panel a), similar distributions were obtained. Furthermore, for scenario Qsim1, Qsim2, and Qsim3 the inter-quartile range got smaller in absolute terms and their correlation values improved. Among these, however, scenario Qsim3 obtained a higher score for certain stations probably because this scenario specifically used linear correlation as the metric to be optimized during the calibration procedure. Looking at the KGE (Fig. 4, panel b) similar distributions were also obtained, although a clear overall improvement of the simulations in Qsim1 and Qsim2 is shown in comparison with Qsim0. This was expected because these two scenarios used KGE for calibration, providing an optimal solution which is equally good for bias, flow variability, and correlation. As a consequence of improving the correlation score on scenario Qsim3 during the calibration period, it resulted in higher inter-quartile KGE′ values in comparison with Qsim0, although this is not valid for the validation period, even though the median value increased. Finally, the NSE values (Fig. 4, panel c), show a similar pattern as observed with KGE′ scores. Qsim1 and Qsim2 obtained higher values than Qsim0, with Qsim1 the best scenario. Moreover, Qsim3 shows some improvement for the calibration period in comparison with Qsim0, but not very relevant for the validation period. For clarity, in the KGE′ and NSE′ graphs the station G1177 is not shown but its KGE′ values are: calibration [−3.63, −0.60, −0.60, −1.50] and validation [−2.55, −1.11, −1.11, −3.40], and NSE: calibration [−23.91, −7.90, −7.91, −31.50] and validation [−43.37, −15.82, −15.84, −56.81] for Qsim0, Qsim1, Qsim2 and Qsim3, respectively.

4.2. Quantifying the gain in skill for each calibration scenario

We calculated the improvement or deterioration of the skill score obtained during the calibration and validation of the calibration scenarios (Qsim1, Qsim2, and Qsim3), in relation to the uncalibrated scenario (Qsim0), and the differences in gain between Qsim1 and Qsim2. This was done firstly by comparing the simulated streamflow of each scenario against in situ measurements, as previously shown, and secondly, by calculating the gain based on the r and KGE′ score of scenarios for Qsim1 (Eq. (4)), Qsim2 (Eq. (5)), and Qsim3 (Eq. (6)) with respect to the uncalibrated scenario (Qsim0), and comparing the results of Eqs. (4) and (5) (Eq. (7)).

Fig. 5 shows in panel a) that for the calibration period, Qsim3 obtained overall the highest r gain, and in this respect the gain in scenarios Qsim1 and Qsim3 are very similar. On the other hand, for the validation period there are no important differences between the different scenarios. Note that during calibration of Qsim2, the skill (KGE′) is calculated based on the Q_{GFDS}, so the skill always improved compared with Qsim0. For calibration of Qsim3, the skill (r) is calculated based on the GFDS, so the r skill always improved in comparison with Qsim0. However, in this figure and analysis, the comparison is made of Qsim2 or Qsim3 against in situ streamflow measurements. So although these data are regarded as "ground truth", previous research showed that the analysis of rating curve uncertainty leads to errors from 1.8 to 38.4% of the total volume studied (Di Baldassarre & Montanari, 2009; Pappenberger et al., 2006). Note that in the cases where the gain in Qsim3 is negative, this is because although there was an improvement in the r score during the calibration in comparison with the raw GFDS signal, when comparing the simulated streamflow with in situ observations, it is possible that that score slightly decreased. In panel b) of the same figure, we see how the KGE gains on the Qsim1 and Qsim2 are both positive in comparison to the uncalibrated run, and very similar during the calibration and validation periods, whereas Qsim3 obtained overall smaller gains and the median gain value in KGE is close to 0 as expected because its calibration was not optimized in this score.

Furthermore, Fig. 6 shows for scenario Qsim2 (panel a) and Qsim3 (panel b) the KGE gain values geographically distributed. For scenario Qsim2 and for most of the stations, there was an important improvement of the KGE values in both the calibration and validation period. However, for the stations located in Missouri (G0729) river in North America, the simulated streamflow did not substantially improve, looking at the validation period, whereas for the other stations with negative KGE Gain values, G1083, G1121, G1129, and G1286, their values was less than −0.03. Therefore, there was no improvement after calibration in the skill to reproduce in situ observations. For example, for the Yukon River, the uncalibrated model already underestimates flow using the lowest value of the TranSub parameter, and therefore calibration was unsuccessful in reducing the flow underestimation.

For scenario Qsim3, the overall gain is smaller, as expected, because KGE′ was not used as an optimisation objective during calibration. For the calibration period, four locations obtained lower gain with the range [−0.02 to −0.69], and 11 locations on the validation [−0.02 to −0.99]. However, the increase of the r score leads to an improvement of KGE′ for most of the stations, whereas for others, the KGE′ is smaller than before calibration. For this scenario, the decrease is especially notable also for Missouri river (G0729) in North America, and for Uraricoera River (G1121) in South America in the calibration. On the other hand for the validation period there are more stations with lower scores. This is the case for example for the Parnaiba (G1168) and Mearim (G1177) Rivers, where the model highly overestimates the streamflow both before and after calibration. Therefore, these results, indicate that once the locations are selected based on the GFDS criteria for a reliable quality signal, the performance of the streamflow simulations for scenario Qsim3, in terms of volumetric accuracy, is more dependent on the a priori performance of the hydrological model rather than on the geographical dependences on the quality of the GFDS signal.

4.3. Assessing the variability of the calibrated model parameters

Fig. 7 shows that the median values and spread for each of the parameters (as described in Table 2) are similar between calibration scenarios that used in situ streamflow time series (Qsim1 and Qsim2). However, using GFDS signal (Qsim3) we obtained different median and inter-quartile parameter values for the time constant for water in the upper zone (*Tuz*), the groundwater percolation rate (*GwPerc*), and the multiplier of the Manning's roughness maps of the channel system (*CalChanMan*). The *GwPerc* parameter determines the amount of water that percolates from the upper to the lower zone. In scenario Qsim3, we found that a median lower percolate rate was applied, therefore a larger amount of water is either stored, or added to the channel. In line with this, for this calibration scenario, the time constant *Tuz* is smaller, which means that there is an increase in volume and velocity of inflow to the river channel. As a result, the *CalChanMan* parameter that is used to fine-tune the timing of the channel routing on open channels, obtained higher values. This means that the higher the Manning's roughness, the slower the velocity of the streamflow through the channel. Therefore, these three parameters were found to be more dependent on the calibration methodologies and/or the input dataset used.

However, for the scenarios that used streamflow for calibration (Qsim1 and Qsim2), we also calibrated the amounts of inflow to the deeper groundwater systems (*GwLoss*) and losses due to evaporation, leaching or water use (*TransSub*). In order to optimize the simulated streamflow, this resulted for both scenarios in a higher median amount of percolation from the upper to the lower zone, and also a slighter higher median value of loss to groundwater compared with both the uncalibrated and Qsim3 scenarios. Again, it should be noted that the raw GFDS signal is dimensionless and does not contain information about the quantity of streamflow unless a (linear) relation between the satellite signal and historical in situ streamflow measurements is established (as has been done using Q_{GFDS} in Qsim2).

We also tested the exclusion of the groundwater percolation (*GwPerc*) parameter for calibration of the scenario Qsim3, and calibrated for the remaining *Tuz*, *Tlz*, and *CalChanMan*. In comparison to the results from Qsim3 illustrated in Fig. 7, we obtained a similar median value for *Tuz*, but lower median values for *Tlz* and *CalChanMan*. This means that for this calibration method applied to the LISFLOOD model, in order to optimize the streamflow simulation, the movement of the flow from the lower zone to the river channel is faster than previously, but the streamflow moves through the channel more slowly.

5. Discussion

5.1. Assessing the use of satellite derived surface water data for calibration

In this study, the use of satellite-derived passive microwave data from GFDS has been tested for calibration of the rainfall-runoff LISFLOOD setup within the GloFAS flood forecasting system, in order to enhance the skill of the simulated streamflow. As in previous studies, use of satellite-derived data does not aim at replacing in situ measurements for model calibration, but rather to use them as an alternative for ungauged locations. Another possibility, as demonstrated here, is to use the satellite data as an extension of historical time series at locations where the gauges are no longer active, or for periods when in situ data are not available. The results of our study correspond with other research carried out in the area of calibrating hydrological or hydraulic models using satellite-derived data, such as inundation extent, river width, and water levels from several satellite platforms (see Table 1). In these studies, the usefulness of satellite-derived information to

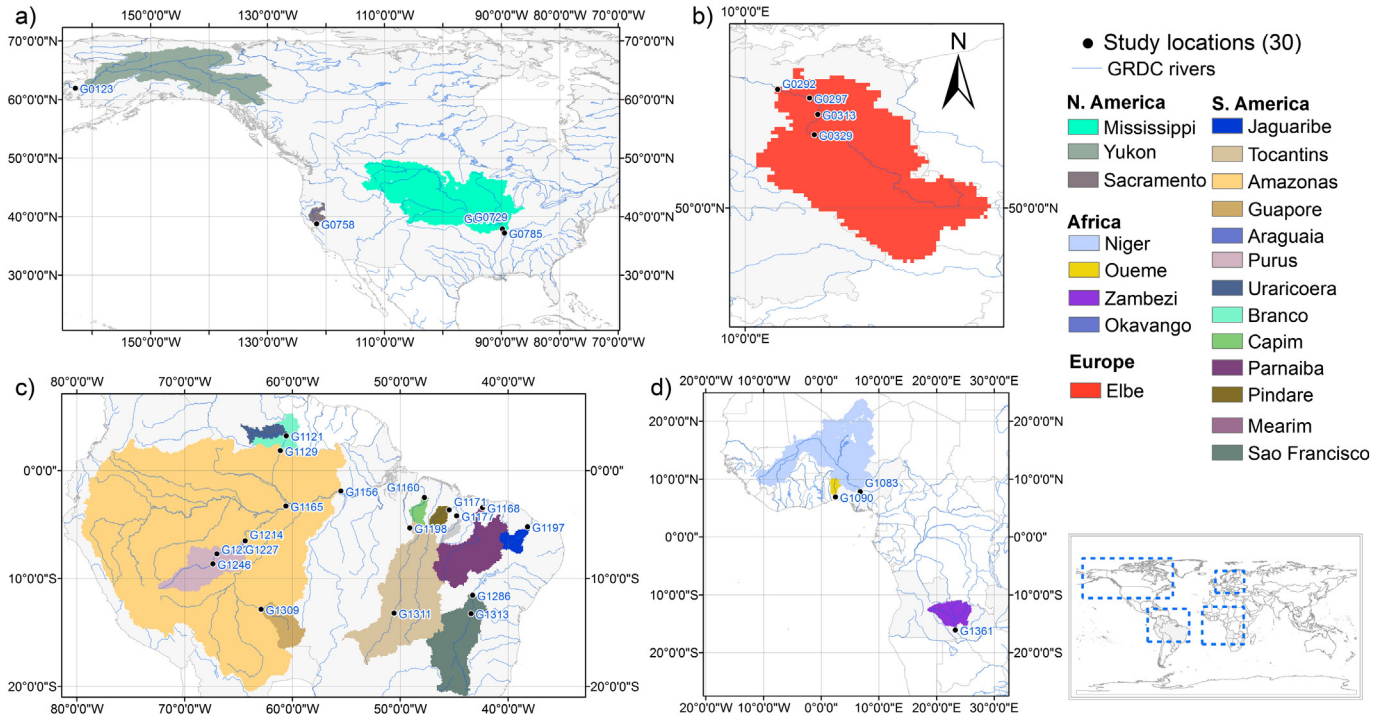

Fig. 1. Gauging stations used for testing the value of the GFDS signal for hydrological model calibration including their associated catchments in (a) North America, (b) Europe, (c) South America, and (d) Africa. Shaded polygons represent the drainage areas at the outlets of the studied river basins.

increase the performance and reliability of the models was highlighted. Some of these studies focus on the model calibration for specific flood events, others (like this study) focus on longer periods, greater than 10 years (e.g. Domeneghetti et al., 2014; Getirana, 2010; Getirana et al., 2013; Milzow et al., 2011). However, these studies use, for example, altimetric data from ERS and/or ENVISAT missions, with an image acquisition frequency of 35 days, while the data used in our study are obtained from several passive microwave sensors in the GFDS system, which have a daily visit frequency. Furthermore, this study is the first to test the use of surface water changes data from GFDS for calibration of the LISFLOOD-GloFAS model setup.

5.2. Outlying the overall benefits and limitations of using GFDS satellite-derived data for calibration

The results have shown the overall benefits of using raw GFDS data for calibration in (virtually) ungauged basins. Therefore, even though the gain in the calibration when using raw satellite data can be limited,

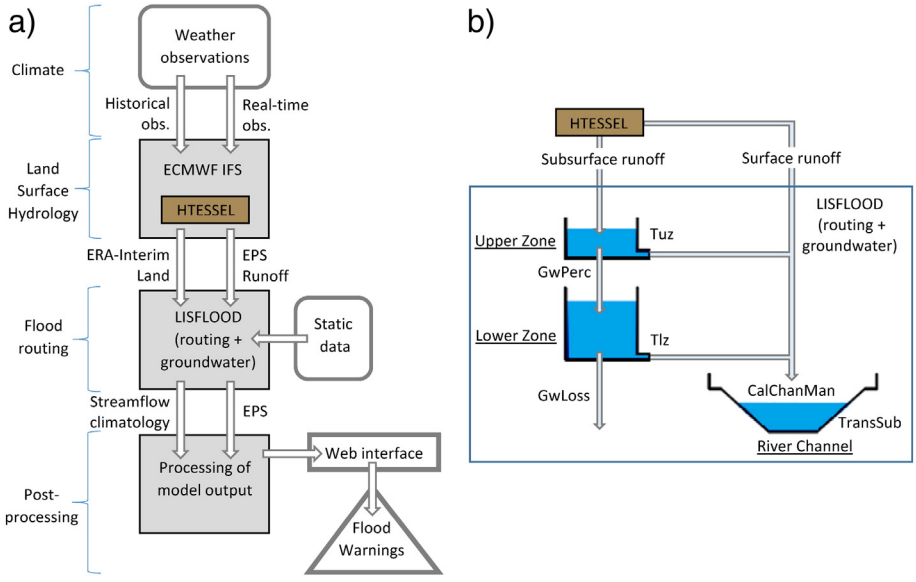

Fig. 2. a) Overview of the GloFAS setup. The blue-contoured polygon indicates the input and output datasets and model used within this study (adapted from Alfieri et al., 2013); b) Schematic of the LISFLOOD model used in this study (adapted from Burek et al., 2013). LISFLOOD uses the surface and sub-surface runoff outputs from HTESSEL as input. Light blue arrows in panel b) represent water fluxes. The parameter names are explained in Table 2.

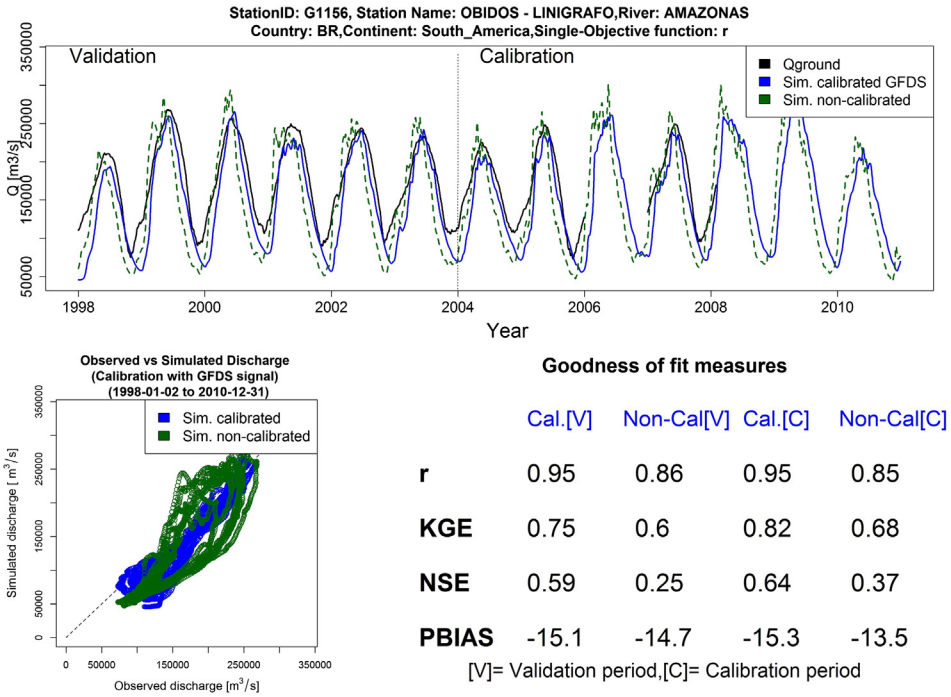

Fig. 3. Results of LISFLOOD calibration using GFDS signal for the Qsim3 optimization (r) method for the Obidos-Linigrafo station (Amazonas River, Brazil). Skill scores show the results of comparing the model simulations (calibrated and uncalibrated) with the in situ observed streamflow. In the time series panel, the dashed line marks the end of the validation period, and beginning of the calibration period.

mainly to the timing of the peaks, it is nonetheless an attractive option for calibration due to its complete independence from in situ measurements and, consequently, its suitability for completely ungauged catchments. Furthermore, the gain in the correlation using the GFDS signal is in line with the amount of improvement in that same metric by using Q_{ground}, dataset which marks, within the presented calibration framework, the limits of how well the simulated streamflow can be calibrated at those locations.

It is expected that the application of this calibration concept to a global flood forecasting system such as GloFAS, should also improve the forecast performance. However, quantifying the improvement in forecast skill is beyond the scope of this study. Furthermore, the results also show that for a global hydrologic model calibration, the use of all three types of data – namely in situ streamflow, GFDS-derived streamflow, and raw GFDS signal – would be likely to result in the best possible calibrated model setup. Nevertheless, use of the GFDS signal as a proxy for measured streamflow is not possible for every site with an in situ gauge (Revilla-Romero et al., 2014), due to the influence on the satellite signal of local factors such as river geometry, the amount of average daily streamflow, and the floodplain dimension. Therefore, one limitation of an approach using all three datasets is confidence in the quality or signal/noise ratio of the GFDS satellite signal at locations where its validation is challenging due to the lack of in situ observations. Furthermore, uncertainties in both the raw GFDS signal to capture surface water changes and GFDS-derived streamflow to reproduce in situ streamflow, have not yet been quantified.

5.3. Future research direction

An additional potential value of using the GFDS signal within a hydrological model would be to include the signal at specific sites for post-processing (Bogner & Kalas, 2008) through data assimilation and error correction of the streamflow forecast for the GloFAS model, or spatial data assimilation analogous to the use of remotely sensed soil moisture in the LISFLOOD model (Revilla-Romero et al., in preparation; Wanders, Karssenberg, de Roo, de Jong, & Bierkens, 2014), thereby providing more reliable information of the model simulations.

Furthermore, we acknowledge that a calibration scheme where the input fluxes remain equal for the surface and sub-surface runoff fractions has some limitations. Therefore, including this element will also be considered for calibration of the full global hydrological model. This could be done either by calibrating the HTESSEL land surface model, or by using the full LISFLOOD global setup to test the calibration. It is also planned to use this full version of LISFLOOD for a multi-model framework setup of GloFAS. Once the full calibration of the model is carried out, the optimal set of parameters defined for each catchment, will be applied within the pre-operational GloFAS flood forecasting system. However, further research is required on the effect of setting a single parameter value for large catchments, and how to increase the parameter variability within a single catchment.

6. Conclusion

This study has evaluated the potential of remotely-sensed estimates of surface water extent, from the Global Flood Detection System (GFDS), to improve hydro-meteorological model simulations of the LISFLOOD model for large-scale catchments. Main conclusions of the study, based on the 30 sites located in Africa, Europe, North America, and South America, and the different calibration scenarios, are summarized as follows:

1) Using satellite-derived surface water extent as a proxy for streamflow (scenario Qsim3) for calibration generally leads to more a skilled streamflow simulation than the uncalibrated streamflow simulation, based on a comparison of the simulated time series with the in situ streamflow data. The calibration framework used linear correlation

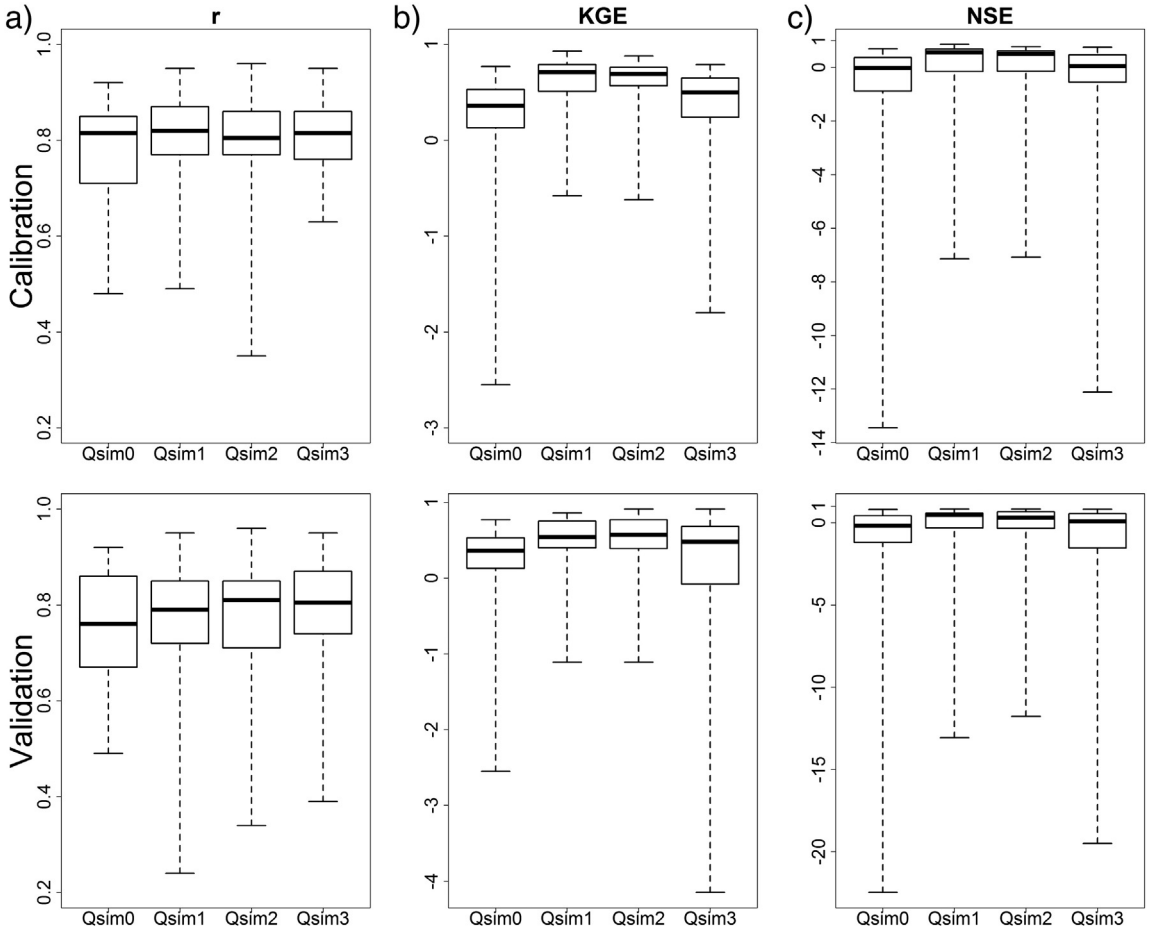

Fig. 4. Boxplots of the (a) correlation coefficient (r), (b) Kling-Gupta Efficiency (KGE′), and (c) Nash–Sutcliffe (NSE) for each of the simulations (Qsim0-Qsim4), obtained for the calibration and validation periods versus in situ observation time series.

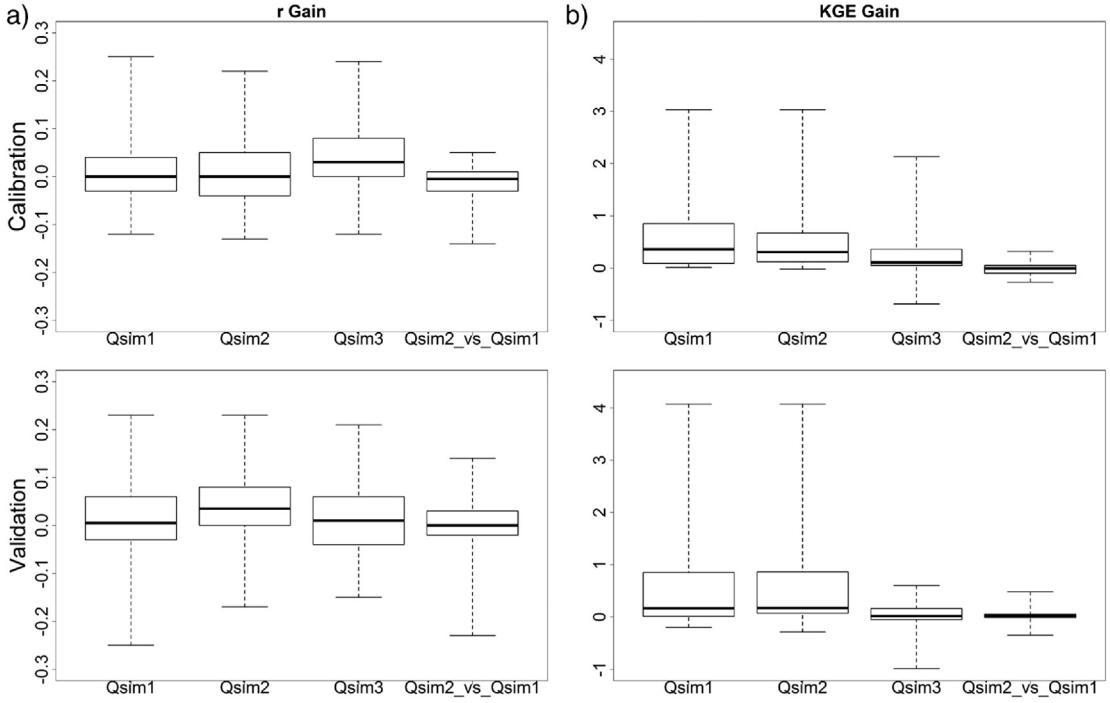

Fig. 5. Summary of the gains for each scenario (Qsim1, Qsim2, and Qsim3) compared with Qsim0, and between the calibrated scenarios with streamflow time series (Qsim1 and Qsim2).

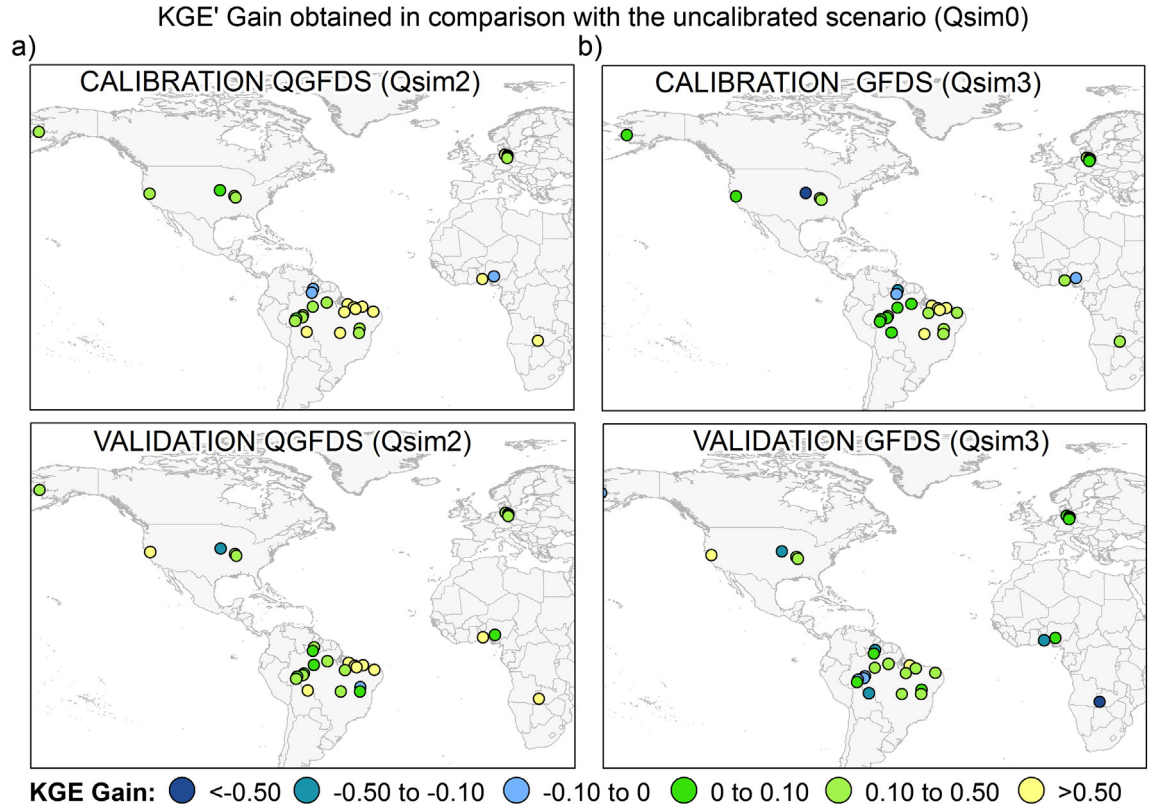

Fig. 6. Spatial distribution of the gain or loss of the Kling–Gupta Efficiency (KGE′) coefficient obtained by each of the simulation scenarios calibrated with satellite-derived information (Qsim2 and Qsim3) in comparison with the uncalibrated scenario (Qsim0), a) during calibration and, b) validation periods. Skill scores are obtained by comparing each simulated streamflow set with in situ observed streamflow.

(r) for optimisation, due to the dimensionless of the raw GFDS signal. However, the independence of GFDS from in situ measurement is a key strength of this dataset, in terms of its application for completely ungauged catchments.

2) Using observed streamflow gauge data to calibrate the model produced marginally better results than using satellite-derived streamflow estimated data. This is especially relevant for the KGE metric, which, also takes into account flow variability and bias,

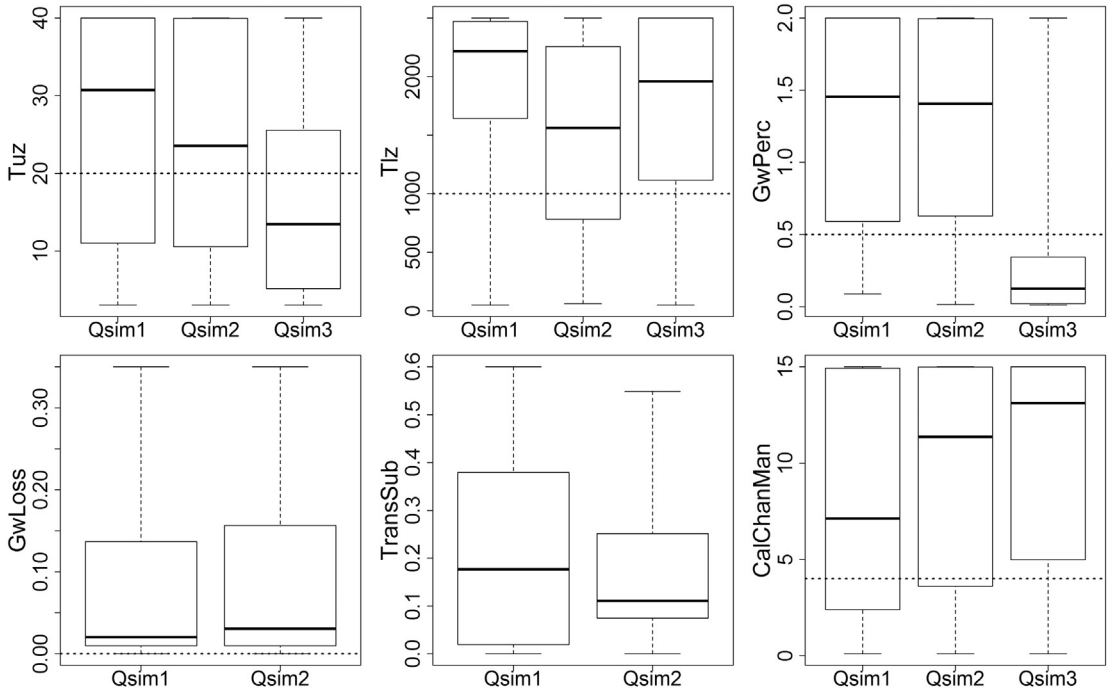

Fig. 7. The six calibrated parameters distribution for the three calibration scenarios. Refer to Table 3 for the description of the scenarios. The default parameter values used for the uncalibrated run (Qsim0) are marked with the dashed line, except for the TransSub parameter, where an input global map was used for the uncalibrated run with values [0–0.45]. GwLoss and TransSub were not calibrated for Qsim3 scenario.

compared with linear correlation. However, at multiple sites, we needed to use shorter potential evaluation periods due to the lack of (recent) in situ data, whereas satellite-derived data are available almost continuously on a daily basis since 1998 or 2002 (depending of the latitudes) until present.

3) The continuous availability of the Global Flood Detection System data is a key strength, in terms of their use as a proxy for streamflow, and this is especially useful for large scale hydrological models where the data can be used to extend of historical observations from a discontinued gauge, or for periods when in situ data are not available.

Acknowledgments

We wish to thank the Global Runoff Data Centre (GRDC) for providing the in situ observed streamflow data. We would also like to thank the Joint Research Centre (JRC) Crisis Monitoring and Response Technologies (CRITECH) team for providing the GFDS data, and Catherine Prigent, Filipe Aires, and Fabrice Papa of the CNRS — Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique (LERMA) for providing the GIEMS dataset. Furthermore, we thank Niall McCormick for proof-reading the manuscript.

Appendix A

Table A1
List of study stations. *Station location adapted to be located on the river network (center of the 0.1° cell).

ID	Station Name	River Name	Country	Continent	Drainage area provided	Lat*	Lon*
G0123	Pilot Station, AK	Yukon River	US	North America	831,390	61.85	−162.85
G0292	Neu-Darchau	Elbe River	DE	Europe	131,950	53.25	10.85
G0297	Wittenberge	Elbe River	DE	Europe	123,532	52.95	11.85
G0313	Tangermuende	Elbe River	DE	Europe	97,780	52.55	11.95
G0329	Barby	Elbe River	DE	Europe	94,060	52.05	11.75
G0729	Rulo, NE	Missouri River	US	North America	1,084,845	40.05	−95.35
G0758	Verona, CA	Sacramento River	US	North America	55,040.1	38.75	−121.65
G0777	Chester, IL	Mississippi River	US	North America	1,835,274	37.95	−89.95
G0785	Thebes, IL	Mississippi river	US	North America	1,847,188	37.35	−89.45
G1083	Lokoja	Niger	NG	Africa	-	7.85	6.75
G1090	Bonou	Oueme	BJ	Africa	46,990	6.95	2.45
G1121	Fazenda Passarao	Rio Uraricoera	BR	South America	50,985	3.25	−60.55
G1129	Caracarai	Rio Branco	BR	South America	124,980	1.85	−61.05
G1156	Obidos-Linigrafo	Amazonas	BR	South America	4,680,000	−1.95	−55.55
G1160	Badajos	Rio Capim	BR	South America	38,178	−2.55	−47.75
G1165	Manacapuru	Amazonas	BR	South America	2,147,736	−3.35	−60.65
G1168	Luzilandia	Rio Parnaiba	BR	South America	322,823	−3.45	−42.25
G1171	Pindare-Mirim	Rio Pindare	BR	South America	34,030	−3.65	−45.45
G1177	Bacabal	Rio Mearim	BR	South America	27,650	−4.25	−44.75
G1197	Peixe Gordo	Rio Jaguaribe	BR	South America	48,200	−5.25	−38.15
G1198	Maraba	Tocantins	BR	South America	690,920	−5.35	−49.05
G1214	Canutama	Rio Purus	BR	South America	230,012	−6.55	−64.45
G1227	Labrea	Rio Purus	BR	South America	226,351	−7.25	−64.75
G1237	Seringal Fortaleza	Rio Purus	BR	South America	153,016	−7.75	−66.95
G1246	Valparaiso-Montante	Rio Purus	BR	South America	103,285	−8.65	−67.35
G1286	Morpara	Sao Francisco	BR	South America	345,000	−11.55	−43.25
G1309	Pedras Negras	Rio Guapore	BR	South America	110,000	−12.85	−62.95
G1311	Luiz Alves	Araguaia	BR	South America	117,580	−13.25	−50.55
G1313	Bom Jesus Da Lapa	Sao Francisco	BR	South America	271,000	−13.25	−43.45
G1361	Senanga (60,370,001)	Zambezi	ZM	Africa	284,538	−16.05	23.25

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2015.10.022.

References

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P. -P., Janowiak, J., ... Nelkin, E. (2003). The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*, 1147–1167. http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS — Global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*, 1161–1175. http://dx.doi.org/10.5194/hess-17-1161-2013.

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, *19*, 2247–2260. http://dx.doi.org/10.5194/hess-19-2247-2015.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Cloke, H., Dee, D., ... Vitart, F. (2013). ERA-interim/land: A global land water resources dataset. *Hydrology and Earth System Sciences Discussions*, *10*, 14705–14745. http://dx.doi.org/10.5194/hessd-10-14705-2013.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A.K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*, 623–643. http://dx.doi.org/10.1175/2008JHM1068.1.

Bates, P.D., & De Roo, A.P.J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*, 54–77. http://dx.doi.org/10.1016/S0022-1694(00)00278-X.

Beck, H.E., De Jeu, R.A.M., Schellekens, J., Van Dijk, A.I.J.M., & Bruijnzeel, L.A. (2009). Improving curve number based storm runoff estimates using soil moisture proxies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*, 250–259. http://dx.doi.org/10.1109/JSTARS.2009.2031227.

Bergeron, J., Royer, A., Turcotte, R., & Roy, A. (2014). Snow cover estimation using blended MODIS and AMSR-E data for improved watershed-scale spring streamflow simulation in Quebec. *Hydrological Processes*, *28*, 4626–4639. http://dx.doi.org/10.1002/hyp.10123.

Beven, K.J., & Kirkby, M.J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie

du bassin versant. *Hydrological Sciences Bulletin*, *24*, 43–69. http://dx.doi.org/10.1080/02626667909491834.

Bogner, K., & Kalas, M. (2008). Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the upper Danube catchment. *Atmospheric Science Letters*, *9*, 95–102. http://dx.doi.org/10.1002/asl.180.

Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., & Perez, J.R. (2010). *GLOBCOVER 2009 — Products description and validation report.*

Brakenridge, G.R., Nghiem, S.V., Anderson, E., & Chien, S. (2005). Space-based measurement of river runoff. *EOS. Transactions of the American Geophysical Union*, *86*, 185–188. http://dx.doi.org/10.1029/2005EO190001.

Brakenridge, G.R., Nghiem, S.V., Anderson, E., & Mic, R. (2007). Orbital microwave measurement of river discharge and ice status. *Water Resources Research*, 43. http://dx.doi.org/10.1029/2006WR005238 (n/a–n/a).

Brakenridge, R.G., Cohen, S., Kettner, A.J., De Groeve, T., Nghiem, S.V., Syvitski, J.P.M., & Fekete, B.M. (2012). Calibration of satellite measurements of river discharge using a global hydrology model. *Journal of Hydrology*, *475*, 123–136. http://dx.doi.org/10.1016/j.jhydrol.2012.09.035.

Burek, P., van der Knijff, J., & de Roo, A. (2013). *LISFLOOD — distributed water balance and flood simulation model — Revised user manualhttp://dx.doi.org/10.2788/24719.*

Chukhlantsev, A. (2006). *Microwave radiometry of vegetation canopies.* Dordrecht: Kluwer Academic Publishers.

De Roo, A., Odijk, M., Schmuck, G., Koster, E., & Lucieer, A. (2001). Assessing the effects of land use changes on floods in the Meuse and Oder catchment. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, *26*, 593–599. http://dx.doi.org/10.1016/S1464-1909(01)00054-5.

De Roo, A., Schmuck, G., Perdigao, V., & Thielen, J. (2003). The influence of historic land use changes and future planned land use scenarios on floods in the Oder catchment. *Recent Development in River Basin Research and Management. Phys. Chem. Earth Parts ABC*, *28*. (pp. 1291–1300). http://dx.doi.org/10.1016/j.pce.2003.09.005.

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*, 913–921. http://dx.doi.org/10.5194/hess-13-913-2009.

Di Baldassarre, G., Schumann, G., & Bates, P.D. (2009). A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. *Journal of Hydrology*, *367*, 276–282. http://dx.doi.org/10.1016/j.jhydrol.2009.01.020.

Dietz, A.J., Kuenzer, C., Gessner, U., & Dech, S. (2012). Remote sensing of snow —A review of available methods. *International Journal of Remote Sensing*, *33*, 4094–4134. http://dx.doi.org/10.1080/01431161.2011.640964.

van Dijk, A.I.J.M., & Renzullo, L.J. (2011). Water resource monitoring systems and the role of satellite observations. *Hydrology and Earth System Sciences*, *15*, 39–55. http://dx.doi.org/10.5194/hess-15-39-2011.

Domeneghetti, A., Tarpanelli, A., Brocca, L., Barbetta, S., Moramarco, T., Castellarin, A., & Brath, A. (2014). The use of remote sensing-derived water surface data for hydraulic model calibration. *Remote Sensing of Environment*, *149*, 130–141. http://dx.doi.org/10.1016/j.rse.2014.04.007.

Donohue, R.J., Roderick, M.L., & McVicar, T.R. (2007). On the importance of including vegetation dynamics in Budyko's hydrological model. *Hydrology and Earth System Sciences*, *11*, 983–995.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*, 1015–1031. http://dx.doi.org/10.1029/91WR02985.

Fekete, B.M., Vörösmarty, C.J., & Grabs, W. (2002). High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, *16*, 15-1–15-10. http://dx.doi.org/10.1029/1999GB001254.

Feyen, L., Vrugt, J.A., Nualláin, B.Ó., van der Knijff, J., & De Roo, A. (2007). Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology*, *332*, 276–289. http://dx.doi.org/10.1016/j.jhydrol.2006.07.004.

Fluet-Chouinard, E., Lehner, B., Rebelo, L. -M., Papa, F., & Hamilton, S.K. (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sensing of Environment*, *158*, 348–361. http://dx.doi.org/10.1016/j.rse.2014.10.015.

Fortin, F. -A., De Rainville, F. -M., Gardner, M. -A., Parizeau, M., & Gagńe, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*, 2171–2175.

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., & Bianchi, A. (2014). Ensemble projections of future streamflow droughts in Europe. *Hydrology and Earth System Sciences*, *18*, 85–108. http://dx.doi.org/10.5194/hess-18-85-2014.

Getirana, A.C.V. (2010). Integrating spatial altimetry data into the automatic calibration of hydrological models. *Journal of Hydrology*, *387*, 244–255. http://dx.doi.org/10.1016/j.jhydrol.2010.04.013.

Getirana, A.C.V., Boone, A., Yamazaki, D., & Mognard, N. (2013). Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon Basin. *Water Resources Research*, *49*, 614–629. http://dx.doi.org/10.1002/wrcr.20077.

Global Runoff Data Centre (GRDC) (2010). *The river discharge time series, Koblenz, Germany*. Federal Institute of Hydrology (BfG) ([WWW Document]. URL http://grdc.bafg.de/(accessed 1.20.14)).

Hannah, D.M., Demuth, S., van Lanen, H.A.J., Looser, U., Prudhomme, C., Rees, G., ... Tallaksen, L.M. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*, 1191–1200. http://dx.doi.org/10.1002/hyp.7794.

Hirpa, F.A., Gebremichael, M., Hopson, T.M., Wojick, R., & Lee, H. (2014). Assimilation of satellite soil moisture retrievals into a hydrologic model for improving river discharge. In V. Lakshmi, D. Alsdorf, r. Anderson, S. Biancamaria, M. Cosh, J. Antin, G. Huffman, W. kustas, P. van oevelen, T. painter, J. parajka, rR. tthew, & C. Rüdiger (Eds.), *Remote sensing of the terrestrial water cycle* (pp. 319–329). John Wiley & Sons, Inc.

Hirpa, F.A., Salamon, P., Alfieri, L., Thielen, J., Zsoter, E., & Pappenberger, F. (2015). The effect of reference climatology on global flood forecasting. *Journal of Hydrometeorology (submitted for publication).*

Hostache, R., Matgen, P., Schumann, G., Puech, C., Hoffmann, L., & Pfister, L. (2009). Water level estimation and reduction of hydraulic model calibration uncertainties using satellite SAR images of floods. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 431–441. http://dx.doi.org/10.1109/TGRS.2008.2008718.

Huffman, G.J., Adler, R.F., Bolvin, D.T., & Gu, G. (2009). Improving the global precipitation record: GPCP version 2.1. *Geophysical Research Letters*, *36*, L17808. http://dx.doi.org/10.1029/2009GL040000.

Jiang, D., Wang, J., Huang, Y., Zhou, K., Ding, X., & Fu, J. (2014). The review of GRACE data applications in terrestrial hydrology monitoring. *Advances in Meteorology*, *2014*http://dx.doi.org/10.1155/2014/725131.

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424–425*, 264–277. http://dx.doi.org/10.1016/j.jhydrol.2012.01.011.

Kugler, Z., & De Groeve, T. (2007). *The global flood detection system.* Luxembourg: Office for Official Publications of the European Communities.

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*, 1–22. http://dx.doi.org/10.1016/j.jhydrol.2004.03.028.

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *EOS. Transactions of the American Geophysical Union*, *89*, 93–94. http://dx.doi.org/10.1029/2008EO100001.

Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., ... Reed, P.M. (2014). Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling and Software*, *62*, 271–299. http://dx.doi.org/10.1016/j.envsoft.2014.09.013.

Marthews, T.R., Dadson, S.J., Lehner, B., Abele, S., & Gedney, N. (2015). High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, *19*, 91–104. http://dx.doi.org/10.5194/hess-19-91-2015.

Mason, D.C., Bates, P.D., & Dall' Amico, J.T. (2009). Calibration of uncertain flood inundation models using remotely sensed water levels. *Journal of Hydrology*, *368*, 224–236. http://dx.doi.org/10.1016/j.jhydrol.2009.02.034.

Milzow, C., Krogh, P.E., & Bauer-Gottwein, P. (2011). Combining satellite radar altimetry, surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. *Hydrology and Earth System Sciences*, *15*, 1729–1743. http://dx.doi.org/10.5194/hess-15-1729-2011.

Minville, M., Cartier, D., Guay, C., Leclaire, L. -A., Audet, C., Le Digabel, S., & Merleau, J. (2014). Improving process representation in conceptual hydrological model calibration using climate simulations. *Water Resources Research*, *50*, 5044–5073. http://dx.doi.org/10.1002/2013WR013857.

Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., & Hoffmann, L. (2009). Calibration and sequential updating of a coupled hydrologic–hydraulic model using remote sensing-derived water stages. *Hydrology and Earth System Sciences*, *13*, 367–380. http://dx.doi.org/10.5194/hess-13-367-2009.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F.T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*, 3511–3538. http://dx.doi.org/10.5194/hess-18-3511-2014.

Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*, 282–290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.

Nicklow, J., Reed, P., Savic, D., Harrell, L., Chan-Hilton, A., Karamouz, M., ... ASCE Task Committee on Evolutionary Computation in Environmental and Water Resources Engineering (2010). State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*, *136*, 412–432. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000053.

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J. -B., Pfister, L., & Fraipont, P. (2006). Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources*, *29*, 1430–1449. http://dx.doi.org/10.1016/j.advwatres.2005.11.012.

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. *Hydrological Processes*, *25*, 1091–1113. http://dx.doi.org/10.1002/hyp.7772.

Prigent, C., Papa, F., Aires, F., Jimenez, C., Rossow, W.B., & Matthews, E. (2012). Changes in land surface water dynamics since the 1990s and relation to population pressure. *Geophysical Research Letters*, *39*, L08403http://dx.doi.org/10.1029/2012GL051276.

Ramillien, G., Famiglietti, J.S., & Wahr, J. (2008). Detection of continental hydrology and glaciology signals from GRACE: A review. *Surveys in Geophysics*, *29*, 361–374. http://dx.doi.org/10.1007/s10712-008-9048-9.

Rao, C.X., & Maurer, E.P. (1996). A simplified model for predicting daily transmission losses in a stream Channel1. *JAWRA Journal of the American Water Resources Association*, *32*, 1139–1146. http://dx.doi.org/10.1111/j.1752-1688.1996.tb03484.x.

Refsgaard, J.C., & Storm, B. (1990). Construction, calibration and validation of hydrological models. In M.B. Abbott, & J.C. Refsgaard (Eds.), *Distributed hydrological modelling. Water Science and Technology Library*. (pp. 41–54). Springer Netherlands.

Revilla -Romero, B., Wanders, N., Burek, P., Salamon, P., De Roo, A., & Thielen, J. (2015). *Integrating remotely sensed surface water extent and a large-scale hydrological model through data assimilation.* (in preparation).

Revilla-Romero, B., Thielen, J., Salamon, P., De Groeve, T., & Brakenridge, G.R. (2014). Evaluation of the satellite-based global flood detection system for measuring river discharge: Influence of local factors. *Hydrology and Earth System Sciences*, *18*, 4467–4484. http://dx.doi.org/10.5194/hess-18-4467-2014.

Rodrigo Rojas, L.F. (2013). Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation. *Global Environmental Change*, *23*http://dx.doi.org/10.1016/j.gloenvcha.2013.08.006.

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*.

Sun, W., Ishidaira, H., & Bastola, S. (2010). Towards improving river discharge estimation in ungauged basins: Calibration of rainfall-runoff models based on satellite observations of river flow width at basin outlet. *Hydrology and Earth System Sciences*, *14*, 2011–2022. http://dx.doi.org/10.5194/hess-14-2011-2010.

Sun, W., Ishidaira, H., & Bastola, S. (2012a). Calibration of hydrological models in ungauged basins based on satellite radar altimetry observations of river water level. *Hydrological Processes*, *26*, 3524–3537. http://dx.doi.org/10.1002/hyp.8429.

Sun, W., Ishidaira, H., & Bastola, S. (2012b). Prospects for calibrating rainfall-runoff models using satellite observations of river hydraulic variables as surrogates for in situ river discharge measurements. *Hydrological Processes*, *26*, 872–882. http://dx.doi.org/10.1002/hyp.8301.

Tarpanelli, A., Brocca, L., Melone, F., & Moramarco, T. (2013). Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery. *Hydrological Processes*, *27*, 1321–1330. http://dx.doi.org/10.1002/hyp.9550.

Thielen, J., Bartholmes, J., Ramos, M. -H., & de Roo, A. (2009). The European flood alert system — Part 1: Concept and development. *Hydrology and Earth System Sciences*, *13*, 125–140. http://dx.doi.org/10.5194/hess-13-125-2009.

Thiemig, V., Bisselink, B., Pappenberger, F., & Thielen, J. (2014). A pan-African flood forecasting system. *Hydrology and Earth System Sciences Discussions*, *11*, 5559–5597. http://dx.doi.org/10.5194/hessd-11-5559-2014.

Thiemig, V., Bisselink, B., Pappenberger, F., & Thielen, J. (2015). A pan-African medium-range ensemble flood forecast system. *Hydrology and Earth System Sciences*, *19*, 3365–3385. http://dx.doi.org/10.5194/hess-19-3365-2015.

Van Der Knijff, J.M., Younis, J., & 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*, 189–212. http://dx.doi.org/10.1080/13658810802549154.

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., & Karssenberg, D. (2014a). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, *50*, 6874–6891. http://dx.doi.org/10.1002/2013WR014639.

Wanders, N., Karssenberg, D., de Roo, A., de Jong, S.M., & Bierkens, M.F.P. (2014b). The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrology and Earth System Sciences*, *18*, 2343–2357. http://dx.doi.org/10.5194/hess-18-2343-2014.

Werner, M., Blazkova, S., & Petr, J. (2005). Spatially distributed observations in constraining inundation modelling uncertainties. *Hydrological Processes*, *19*, 3081–3096. http://dx.doi.org/10.1002/hyp.5833.

Wilson, J.P., & Gallant, J.C. (2000). *Terrain analysis: Principles and applications.* John Wiley & Sons.

Wohl, E., Barros, A., Brunsell, N., Chappell, N.A., Coe, M., Giambelluca, T., ... Ogden, F. (2012). The hydrology of the humid tropics. *Nature Climate Change*, *2*, 655–662. http://dx.doi.org/10.1038/nclimate1556.

Wu, H., Kimball, J.S., Li, H., Huang, M., Leung, L.R., & Adler, R.F. (2012). A new global river network database for macroscale hydrologic modeling. *Water Resources Research*, *48*, W09701. http://dx.doi.org/10.1029/2012WR012313.

Yamazaki, D., O'Loughlin, F., Trigg, M.A., Miller, Z.F., Pavelsky, T.M., & Bates, P.D. (2014). Development of the global width database for large rivers. *Water Resources Research*, *50*, 3467–3480. http://dx.doi.org/10.1002/2013WR014664.

Zajac, Z., Salamon, P., Burek, P., De Roo, A., & Revilla-Romero, B. (2015). *Evaluating the effects of lake and reservoir parameterization in a global river routing model on uncertainty of daily river discharge.* (in preparation).

Zajac, Z., Zambrano-Bigiarini, M., Salamon, P., Burek, P., Gentile, A., & Bianchi, A. (2013). Calibration of the LISFLOOD hydrological model for Europe, Calibration Round 2013. *JRC technical report, European Commission*. Italy: Joint Research Centre. Ispra.

Zhang, Y., Chiew, F.H.S., Zhang, L., & Li, H. (2009). Use of remotely sensed actual evapotranspiration to improve rainfall–runoff modeling in southeast Australia. *Journal of Hydrometeorology*, *10*, 969–980. http://dx.doi.org/10.1175/2009JHM1061.1.

Zhang, Y.Q., Vaze, J., Chiew, F.H.S., & Liu, Y. (2011). Incorporating vegetation time series to improve rainfall-runoff model predictions in gauged and ungauged catchments. *Modelling and Simulation Society of Australian and New Zealand. Presented at the MODSIM 2011 International Congress on Modelling and Simulation, Perth, Australia* (pp. 3455–3461).