# On the contribution of remote sensing-based calibration to model multiple hydrological variables

**A. M. Oliveira[1], A. S. Fleischmann[1], and R. C. D. Paiva[1]**

[1] Instituto de Pesquisas Hidráulicas (IPH), Universidade Federal do Rio Grande do Sul – UFRGS, Av. Bento Gonçalves, 9500, Porto Alegre 90050-260, RS, Brazil.

Corresponding author: Aline Meyer Oliveira (alinemey@gmail.com)

**Key Points:**

- Calibration/evaluation of a hydrological-hydrodynamic model with five remote sensing-based water cycle variables in a tropical region
- Different calibration strategies with current remotely-sensed observations were able to improve water cycle representation
- Model calibration with multiple remotely sensed variables highlighted deficiencies in model structure and parameterization, and observations.

**Abstract**

The accuracy of hydrological model predictions is limited by uncertainties in model structure and parameterization, and observations used for calibration, validation and model forcing. Conventionally, calibration is performed with discharge estimates. However, the internal processes in the model might be misrepresented, i.e., the model might be getting the “right results for the wrong reasons”, which compromises model reliability. An alternative is to calibrate the model parameters with remote sensing (RS) observations of the water cycle. Previous studies highlighted its potential to improve discharge estimates, but put much less effort on investigating other variables of the water cycle. In this study, we analyzed in detail the contribution of five different RS-based variables (water level (h) from Jason-2, flood extent (A) from ALOS-PALSAR, terrestrial water storage (TWS) anomalies from GRACE, evapotranspiration (ET) from MOD16 and soil moisture (W) from SMOS) to calibrate a hydrological-hydrodynamic model for a tropical study region with floodplains in the Amazon basin. Calibration with TWS, ET, W, and h+W were able to improve discharge estimates by around 16% to 48%. Water cycle representation was also improved (e.g., calibration with h improved not only h estimates but also A, TWS and ET). By analyzing differing calibration setups, a consistent selection of complementary variables for model calibration resulted in better performances than incorporating all RS variables into the calibration. By looking at multiple RS observations of the water cycle, we were able to found inconsistencies in model structure and parameterization, which would remain unknown if only discharge observations were considered.

## Plain Language Summary

Hydrological models are important tools for many applications in water resources, such as natural hazards management, quantification of impacts of climate change or anthropogenic effects on the water cycle. However, there are uncertainties in these models, which might lead to inaccurate predictions. In many cases, they are related to calibrating parameters of the model by comparing in-situ streamflow observations with streamflow modeled estimates. Nonetheless, internal processes in the model might be misrepresented, i.e., the model might be getting the "right results for the wrong reasons", which compromises model reliability and its estimates. An alternative is to calibrate the model parameters with remote sensing (RS) observations of the water cycle. In this study, we analyzed the contribution of five different RS-derived variables (water level, flood extent, anomalies in total terrestrial water storage, evapotranspiration, and soil moisture) to calibrate model parameters. We found that RS-based calibration was able to improve water cycle representation (e.g., calibration with water level was able to improve estimates of water level itself, but also flood extent, terrestrial water storage and evapotranspiration). Moreover, by looking at multiple RS observations of the water cycle, we were able to found inconsistencies in model structure and parameterization, which would remain unknown if only discharge observations were considered.

## 1 Introduction

The accurate representation of hydrologic processes in mathematical models remains a key challenge in water resources research and applications (Baroni et al., 2019; Clark et al., 2015; Kirchner, 2006; Nearing et al., 2016; Semenova & Beven, 2015) due to uncertainties in model structure (Wagener et al., 2003), parameterization (Gharari et al., 2014; Shafii & Tolson, 2015), and observations (Di Baldassarre & Montanari, 2009). These uncertainties might lead to inaccurate predictions of hydrological variables for water resources and natural hazards management (Grimaldi et al., 2019; Montanari & Koutsoyiannis, 2014), and for quantification of impacts of climate change and anthropogenic effects on the water cycle (Haddeland et al., 2006; Teutschbein & Seibert, 2012; C. Y. Xu et al., 2005). This problem has led for instance to initiatives to better constrain the terrestrial water budget by fusing models and Earth Observation datasets (M. Pan & Wood, 2006; Pellet et al., 2019).

Traditionally, hydrological models are calibrated against gauged streamflow data, which might hamper predictions in ungauged sites, and it does not provide reliability of an accurate representation of internal model processes, leading to uncertainty in hydrologic predictions (Hrachowitz et al., 2013). Moreover, there are many parameter sets that provide equally acceptable performances on streamflow evaluation (equifinality), but they might be "right for the wrong reasons" (Beven, 2006; Kirchner, 2006). Several solutions have been proposed to improve process representation and reduce uncertainty in model predictions, such as the generalized likelihood uncertainty estimation (Beven & Binley, 1992), dynamic identifiability analysis (Wagener et al., 2003), multiscale parameter regionalization (Samaniego et al., 2010), and multi-objective calibration

(Yapo et al., 1998). Another alternative is the use of complementary datasets besides streamflow observations for model validation (e.g., Alkama et al., 2010; Motovilov et al., 1999; Neal et al., 2012; Siqueira et al., 2018), calibration (e.g., Crow et al., 2003; Franks et al., 1998; Lo et al., 2010; López et al., 2017; Rajib et al., 2016), or data assimilation (e.g., Brêda et al., 2019; Houser et al., 1998; Mitchell et al., 2004; Paiva et al., 2013; Pathiraja et al., 2016; Reichle et al., 2002; Vrugt et al., 2005).

The use of complementary datasets (i.e., observations of hydrological variables besides discharge) for model calibration has been proved as a promising approach to improve representation of processes in hydrological models (Clark et al., 2015), to reduce uncertainty in hydrological predictions (Gharari et al., 2014), to address equifinality issues (Beven, 2006) and to make predictions in ungauged or poorly-gauged sites (Sivapalan et al., 2003). However, distributed data on complementary hydrological variables (e.g., evapotranspiration, soil moisture) are scarce, and in-situ measurements present poor spatial and temporal representativeness. As a consequence, calibration of hydrological models based on other hydrological variables did not become a common practice.

In this context, remote sensing (RS) observations have stood out in the last decade because of their increasing spatial and temporal resolutions, free availability in many cases, and capability to record less monitored hydrological variables such as soil moisture, evapotranspiration, and terrestrial water storage (Lettenmaier et al., 2015). For instance, GRACE mission provided monthly estimates of changes in water storage on a global coverage with an accuracy of 2 cm when estimated uniformly over the land and ocean regions (Tapley et al., 2004). Missions such as SMOS, SMAP, AMSR-E and ASCAT were estimated to provide soil moisture data with a median RMSE of 0.06-0.10 $m^3/m^3$ for the CONUS (Karthikeyan et al., 2017). Altimeters such as Envisat, Jason-2 and ICESat-1 and ICESat-2 can yield water level data with an accuracy ranging from 0.04 m to 0.42 m, involving trade-offs between temporal resolution from 10 to 91 days, and cross-track separation from 15 to 315 km (Jarihani et al., 2013), and the future SWOT mission focuses on surface waters (Biancamaria et al., 2016).

Previous studies have analyzed the value of integrating RS data into hydrological modeling through calibration or data assimilation (see review in Xu et al., 2014 and Jiang & Wang, 2019). In special, RS-based calibration of hydrological models is a promising approach, but it is novel and it has not been fully explored to its potential yet. Therefore, in the next section we present a literature review to identify what are directions and questions that would help us move forward in understating the contributions of RS-based calibration of hydrological models.

## 1.1 Literature review on calibration of hydrological models with RS data

A comprehensive, yet non-exhaustive literature review of studies that used RS datasets for parameter estimation in hydrological models is presented in this section and summarized in Figure 1. A total of 62 research articles was found, which are listed in the Supporting Information (Table S1). Most previous publications about calibrating hydrological models with RS products involved large study areas (> 1000 km²), what is expected because of the coarse resolution of RS products. Most studies used RS-derived evapotranspiration for model calibration, followed by soil moisture (Figure 1b), but there have been attempts for calibration of up to eight different RS-derived variables (Nijzink et al., 2018). This indicates a still existent knowledge gap regarding which RS-derived variables are more useful for model calibration. Indeed, many recent studies have been investigating the added value of RS-derived information to calibrate hydrological models (Figure 1d).

Most of the studies used only one RS product for model calibration (Figure 1e, in black), while thirteen (three) studies used two (three) products, e.g., Kittel et al. (2018) calibrated the parameters of a hydrological model with water level observations from Envisat and Jason-2, and TWS from GRACE. Only two study used more than three RS products for model calibration (Nijzink et al., 2018 and Schattan et al., 2020). Therefore, we identified a knowledge gap on the use of multiple RS products for hydrological model calibration, which would allow a better understanding of the redundancy and complementarity between variables observed by RS.

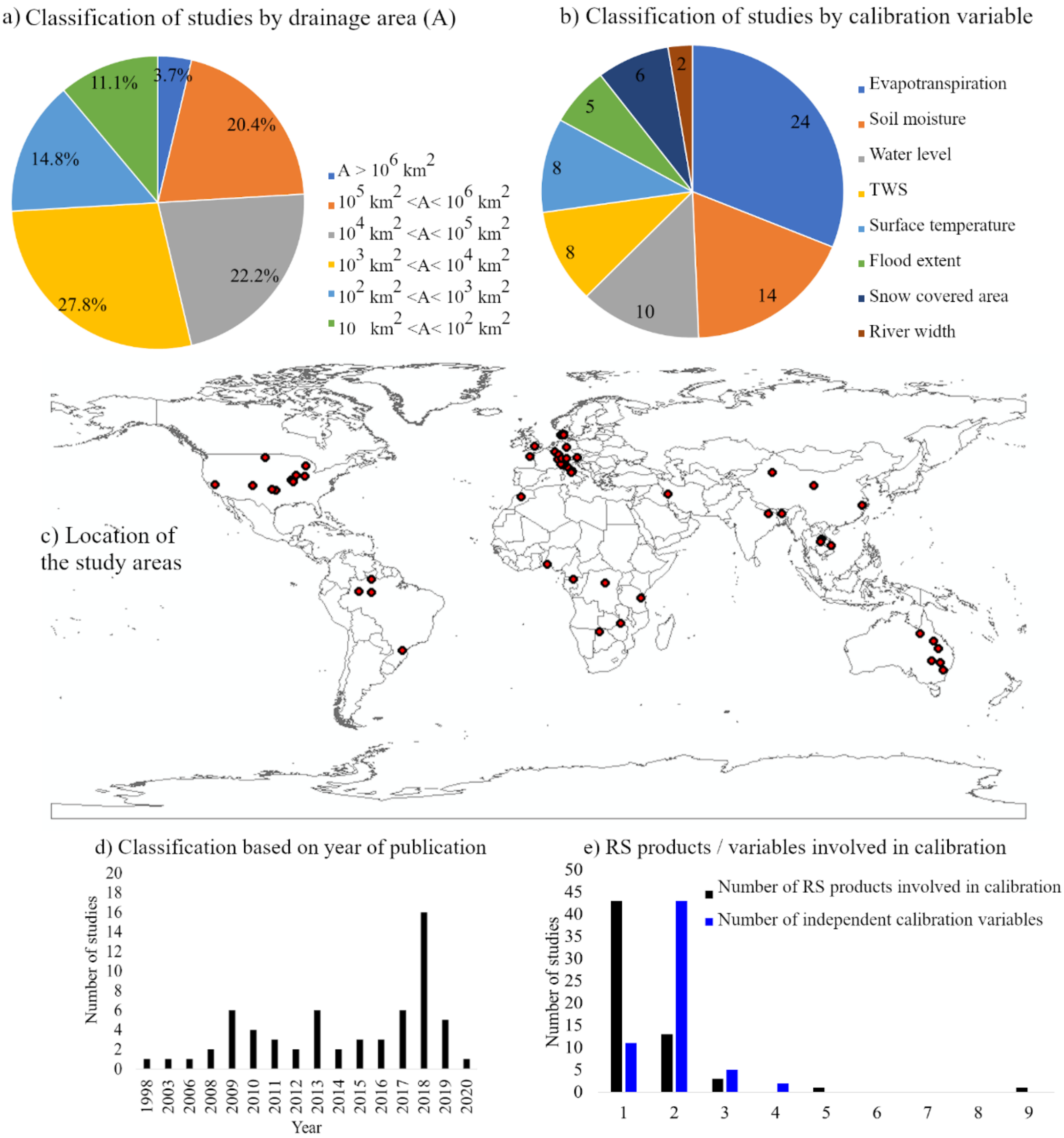


**Figure 1**. Summary of the literature review on 62 studies that incorporated RS datasets for parameter estimation in hydrological models (see Table S1 in Supporting Information). (a) Classification of publications based on the drainage area of study sites (a total of 51 studies informed the drainage area; an average value was considered for publications that used multiple study areas); (b) distribution of studies based on the calibration variable; (c) geographical distribution of study areas from 58 publications (the four remaining publications cover large domains: Nijzink et al., 2018; Poméon et al., 2018; Rakovec et al., 2016; Werth & Güntner, 2010); (d) number of publications per year; and (e) number of RS products involved in calibration (in black) and number of independent calibration variables (in blue).

Some studies addressed the used of RS data to estimate discharge in ungauged basins, following the Prediction in Ungauged Basins (PUB) initiative (e.g., Kittel et al., 2018; Sun et al., 2010), while others focused on narrowing the parameter search space, and thus equifinality reduction, by combining multiple observations for calibration (e.g., Nijzink et al., 2018; Pan et al., 2018). This is confirmed by Figure 1e (in blue), which demonstrates that the vast majority of researches used two variables for calibration (in general, discharge and a RS-derived variable). Within these studies, some analyzed model performance in terms of discharge only, while others considered different variables, providing a more comprehensive discussion on inconsistencies of hydrological models (e.g., Koch et al., 2018; Li et al., 2018). In this study, we attempt to address the latter approach, by analyzing model performance based on multiple variables.

Previous studies can also be classified based on how RS data are incorporated into the model calibration procedure: 40 previous articles used RS-based spatially distributed information, thus calibrating the model with distributed objective functions (e.g., pixel-by-pixel or by sub-basin), while 18 previous publications incorporated RS data as an average for the whole basin.

There is still a need for more studies in tropical regions (especially South America) (Figure 1c), which have particular hydro-climatic characteristics, thus leading to different requirements on model process representation(e.g., snow-related processes might not be so relevant in some tropical areas, whereas an accurate representation of floodplains might be). In this study, a tropical region with extensive floodplains in the Amazon is adopted as a case study.

Most studies used simple flood wave routing schemes such as those based on the kinematic wave, usually adopted in rainfall-runoff models. Only ten researches attempted to couple hydrologic and hydrodynamic models, which is especially relevant for representing flat regions with wetlands (Hodges, 2013; Neal et al., 2012; Pontes et al., 2017). Here, we used a tightly coupled hydrological-hydrodynamic model, being the first study to analyze impacts of calibration of hydraulic parameters (i.e., Manning's coefficient, river width and depth) on hydrological variables (e.g., evapotranspiration and soil moisture).

In general, we identified a lack of researches that use multiple RS variables for model calibration, assessing its impacts on the water cycle representation. In this study, we evaluate the use of multiple RS products to calibrate model parameters, and analyze the redundancy and complementarity between different variables and processes. Therefore, we provide contributions to the literature on what can we learn from model limitations and inconsistencies by looking at multiple RS observations of the water cycle. We also provide insights on how can RS-based calibration improve discharge estimates, and on what is the added value of multi-variable calibration with RS observations.

## 2 Methods

### 2.1 Experimental design

A hydrological-hydrodynamic model (MGB; (Collischonn et al., 2007)) is set up for a case study in the Amazon (Purus River Basin) with a priori parameter sets based on their variability as reported in literature. The study is then divided into two steps.

Firstly, a sensitivity analysis is performed to understand parameter uncertainty and the correlation between model state variables.

Then, a calibration step is performed in which the model is calibrated with the well-known MOCOM-UA optimization algorithm considering six variables: (1) streamflow in-situ observations (one gauge at the basin outlet), and RS observations of (2) water level (one satellite altimetry virtual station), (3) flood extent (sum of flooded areas over the Lower Purus River Basin), (4) terrestrial water storage (TWS), (5) evapotranspiration, and (6) soil moisture. Variables (4), (5) and (6) are considered as an average for the whole basin. The calibration of each variable is performed individually (one-at-a-time), and evaluated for all variables. All experiments are performed in triplicate, and we use state-of-the-art RS products that are freely available. The model is calibrated and evaluated for the same period (2008-2011), given limitations on the availability of simultaneous RS time coverage. A final test is performed in which two multi-variable calibration experiments are conducted: (i) calibration with all analyzed variables, except discharge; and (ii) calibration with two complementary variables which are selected for simultaneous calibration.

## 2.2 Hydrological-hydrodynamic model: MGB

The MGB ("Modelo de Grandes Bacias", a Portuguese acronym for "Large Basin Model") is a semi-distributed, hydrological-hydrodynamic model (Collischonn et al., 2007; Pontes et al., 2017). It was chosen for this study because (1) it has been wide and successfully applied in several South American basins (e.g., Paiva et al., 2013; Siqueira et al., 2018); (2) it is representative and similar to other conceptual hydrological models as VIC (Liang et al., 1994), SWAT (Arnold et al., 2012), and mHM (Samaniego et al., 2010); and (3) the hydrological component is tightly coupled to a hydrodynamic routing scheme, allowing the simulation of complex flat, tropical basins. Moreover, the source code of MGB is freely available at www.ufrgs.br/lsh.

Within the model structure, basins are discretized into unit-catchments, which are further divided into Hydrological Response Units (HRU's) based on soil type and land use. A vertical water balance is performed for each HRU, considering canopy interception, soil infiltration, evapotranspiration, and generation of surface, subsurface and groundwater flows. Flow generated in each HRU is routed to the outlet of the unit-catchment with linear reservoirs. Outflow from each unit-catchment is then propagated through the stream network by using a 1D hydrodynamic model based on the inertial approximation proposed by Bates et al. (2010). The stream network is derived from Digital Elevation Model (DEM) processing. Other model inputs are precipitation and climate data, and soil type and land use maps.

## 2.3 A priori uncertainty of model parameters

Within MGB model, there are parameters related to vegetation cover (leaf area index, vegetation height and Penman-Monteith surface resistance), river hydraulics (Manning's roughness, and width and depth parameters related to geomorphological relationships), and conceptual parameters related to soil water budget (Wm, b, Kbas, Kint, CI, CS, CB), which are further detailed in Supporting Information (Table S2).

The a priori uncertainty of MGB model parameters is estimated based on their variability as reported in literature. Supporting Information (Table S2) presents the calibration parameters, their initial values, range, and the references that support these assumptions.

## 2.4 Sensitivity analysis

In other to understand parameter uncertainty in the MGB model, multiple model runs were conducted considering four uncalibrated model setups: (1) varying only soil parameters; (2) varying only vegetation parameters; (3) varying only hydraulic parameters; (4) varying all parameters together. One hundred runs were conducted, in triplicate, resulting in three hundred runs for each setup.

Parameters were varied considering a uniform distribution, and results were analyzed in terms of RMSD (root mean square deviation) of each variable, by comparing each run with a reference one (i.e., the initial run with the initial parameter set as defined in Table S2 of the Supporting Information). This was performed in order to understand the sources of model uncertainties related to different sets of parameters (e.g., are flood extent estimates sensitive to vegetation parameters, or are ET estimates sensitive to hydraulic parameters?). The uncertainty of the model was also compared to uncertainty in the observations, as derived from literature.

In order to understand which variables are related to each other, another analysis was performed in which for each run the Kling-Gupta Efficiency (KGE; Gupta et al., (2009)) was also computed by comparing each run with the reference one. The correlation between the KGE of all variables was computed with the Pearson coefficient (r), with the aim to understand the correlation between the multiple variables in the model. In this step, neither RS data nor discharge observations are incorporated into the model yet.

## 2.5 Model calibration

The adopted calibration algorithm is MOCOM-UA (Yapo et al., 1998; Multi-objective global optimization for hydrologic models) due to its satisfactory performance when coupled with hydrological models (e.g., Collischonn et al., 2008; Maurer et al., 2009; Naz et al., 2014). MOCOM-UA is an evolutionary algorithm, based on SCE-UA (Duan et al., 1992), that simultaneously optimizes a model population with respect to different objective functions. Here, the population size was set to 100 individuals. Varying model parameters and their ranges are described in Supporting Information (Table S2).

In the one-at-a-time calibration, for each variable, three objective functions that summarize the agreement between simulated and observed (RS) time-series are simultaneously optimized: Pearson correlation (r), ratio of averages ( $\mu_{sim} / \mu_{obs}$ ), and ratio of standard deviations ( $\sigma_{sim} / \sigma_{obs}$ ), which is associated to the individual terms of Kling-Gupta Efficiency (KGE, Gupta et al., 2009).

Then, for the multi-variable calibration, the objective functions are the KGE of each variable considered: firstly, five objective functions were considered (KGE of all variables except discharge); secondly, two objective functions were adopted (KGE of selected variable 1, and KGE of selected variable 2).

Results are expressed in terms of a Skill Score (Equation 1; Zajac et al., 2017).

$$S = \frac{KGE_{calibrated} - KGE_{initial}}{1 - KGE_{initial}} \quad (1)$$

## 2.6 Model setup

The Bare Earth Digital Elevation Model (O'Loughlin et al., 2016) was used for stream network computation and basin discretization with the IPH-HydroTools GIS package (Siqueira et al., 2016). Unit-catchments were discretized by dividing the stream network into fixed length reaches of 10 km, resulting in 2957 unit-catchments for the whole basin. Soil type and land cover maps were extracted from the HRU discretization developed by Fan et al. (2015): (1) deep and (2) shallow forested areas, (3) deep and (4) shallow agricultural areas, (5) deep and (6) shallow pasture, (7) wetlands, (8) semi-impervious areas, and (9) open water. In the Purus River Basin, 57.4% of the region is covered by forest with deep soils, 26.9% by forest with shallow soils, and 13.7% by wetlands (i.e., river floodplains). Daily precipitation data were derived from TMPA 3B42 (version 7), with spatial resolution of 0.25º x 0.25º (Huffman et al., 2007), which were extracted and interpolated by the inverse distance weighting method for the centroid of each unit-catchment. Long term climate averages for mean surface air temperature, relative humidity, insolation, wind speed and atmospheric pressure are from the Climatic Research Unit database (New et al., 2000), available at a spatial resolution of 10', and interpolated with the nearest neighbor method.

## 2.7 Calibration/Evaluation Data

The following data were used for model calibration and evaluation:

-*In-situ discharge measurements* were obtained from the Brazilian Water Agency Hidroweb database (available at <

http://www.snirh.gov.br/hidroweb/publico/apresentacao.jsf>), at the gauge “Canutama” (code: 13880000; location: S ° 32' 20.04"; W 64° 23' 8.88"; drainage area: 236,000 $km^2$, period of available data: 1973 to 2016). Uncertainty in discharge observations can be estimated as ranging from 6.2% to 42.8% at the 95% confidence level, with an average of 25.6% (Di Baldassarre & Montanari, 2009).

- *Remotely sensed water level data* were obtained from Jason-2 mission, which presents an orbit cycle of approximately 10 days, and tracks separated by approximately 300 km at the equator (Lambin et al., 2010). It presents an accuracy of approximately 0.28 m (Jarihani et al., 2013), and data are available since 2008. The virtual station presented in Figure 1 corresponds to Track 165. Processed data for this study were downloaded from the Hydroweb/Theia database (http://hydroweb.theia-land.fr). Simulated and RS water level data were compared in terms of anomaly (values subtracted from long term average).

- *Satellite flood extent data* were derived from ALOS-PALSAR imagery, which presents a recurrence cycle of 46 days (from 2006 to 2011) and a ground resolution of 100 m (Rosenqvist et al., 2007). Images were downloaded from Alaska Satellite Facility (available at <https://www.asf.alaska.edu/>) in processing level 1.5, which already presents geometric and radiometric corrections. A 3 x 3 median filter was used to remove speckle noise (Lee et al., 2014). Images were classified into water (backscattering coefficient less than -14 dB), non-flooded forest (between -14 dB and -6.5 dB), and flooded forest (higher than -6.5 dB) classes, according to Hess et al. (2003) and Lee et al. (2014). The accuracy of flood extent estimates was estimated based on the RMSE between the resulting classification of this study, and the dual-season mapping developed by Hess et al. (2003). Simulated and RS flood extent data were compared for the pink area depicted in Figure 1, in order to avoid spurious flood extent data in regions that are known to be not subject to flooding.

- *Satellite-based terrestrial water storage (TWS) anomalies* were extracted from GRACE mission, launched in March 2002. It provides monthly TWS estimates~~,~~ based on anomalies in gravitational potential, at a resolution of 300-400km, with an uniform accuracy of 2 cm over the land and ocean regions (Tapley et al., 2004). TWS anomalies were retrieved from three processing centers - GFZ (Geoforschungs Zentrum Potsdam, Germany), CSR (Center for Space Research at University of Texas, USE), and JPL (Jet Propulsion Laboratory, USA), available at <https://grace.jpl.nasa.gov/>, and then averaged for the whole basin. Simulated and RS TWS were compared in terms of anomaly (values subtracted from long term average).

- *Satellite-based evapotranspiration* estimates were retrieved from MOD16 product, derived by an algorithm presented by Mu et al. (2011) based on Penman-Monteith equation. The dataset covers the period from 2000-2010 with a spatial resolution of 1 $km^2$ for global vegetated land areas. Because of that, even though MGB evapotranspiration is calculated for flooded areas (main channel and floodplains) and vegetation for water balance purposes, only the vegetation-ET output was compared to MOD16. MOD16 products are provided in 8-days, monthly and annual intervals. Monthly intervals were used here and averaged for the whole basin (mm/month).

Accuracy of MOD16 along the Amazon basin is estimated as 0.76 mm/day (Gomis-Cebolla et al., 2019).

- *Satellite-based soil moisture* is derived from SMOS mission (Kerr et al., 2001), processed by CATDS, and downloaded in processing level 4, which combines lower level products with data from other sensors and modeling/data assimilation techniques. Daily L4 root zone soil moisture at 0-1m (Al Bitar et al., 2013) were used, and data from ascending and descending orbits were averaged for the whole basin. Since MGB model represents the soil as a bucket (i.e., one only soil layer), SMOS values were rescaled for the range 0 - 100% for comparison with the model based saturation degree, according to the Min/Max Correction method described by Tarpanelli et al. (2013), and applied by some studies (e.g., Rajib et al., 2016; Silvestro et al., 2015).

## 2.8 Study area: Purus River Basin

The Purus River Basin (Figure 2) in Amazon presents a drainage area of approximately 236,000 km², and discharge values range from around 1,000 (June-December) to 12,000 m³/s (January-July) at Canutama gauge. Because of its large scale, it is compatible with the spatial resolution of RS products (e.g., a pixel of GRACE presents spatial resolution of roughly 300-400 km). Purus river presents minor anthropogenic influence, which simplifies the modeling process. Besides, the climate is tropical, and mean annual rainfall is 2147 mm (according to in-situ gauges). Purus was also selected because of its representativeness of tropical regions as the Amazon basin, which is the largest river in the world (Holeman, 1968), and it is characterized by extensive floodplains (Junk, 1997). For instance, on the lower Purus, the floodplain width is in the order of 30 km, which corresponds to approximately 30 times the main channel width (Paiva et al., 2011). These floodplains allow a satisfactory flood extent monitoring by RS image classification, which contributes to the suitability of Purus River Basin for this study.

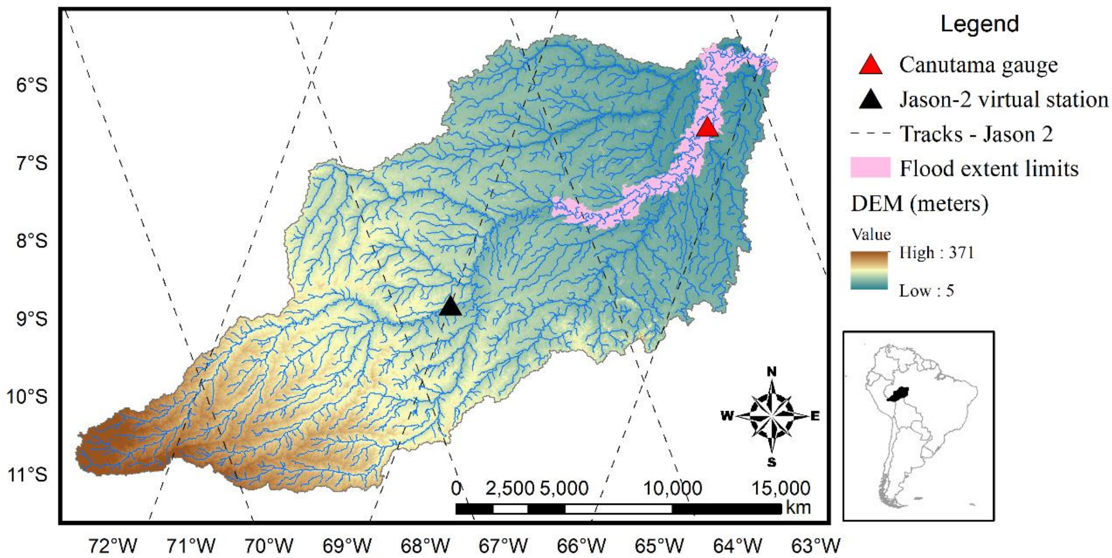

**Figure 2**. Study area: Purus River Basin. Bare Earth Digital Elevation Model (O'Loughlin et al., 2016) and drainage network are presented on the back. It also presents locations of the discharge gauge (Canutama, triangle in red), tracks of the spatial altimetry mission Jason 2 (dashed black lines) and location of the altimetry virtual station (triangle, in black), and the area used for extraction of flood extent (Lower Purus, pink polygons).

# 3 Results and discussion

Results are structured as follows. Firstly, the sensitivity analysis is presented with discussions on model uncertainties (Section 3.1). Then, results for model calibration are presented, with discussions on how RS-based model calibration can improve discharge and water cycle representation (Section 3.2).

## 3.1 Sensitivity analysis

A sensitivity analysis was carried out to understand the a priori uncertainty of the model (Figure 3), by considering six output variables (discharge, water level, flood extent, TWS anomalies, vegetation ET, and soil moisture), and regarding the dispersion provided by varying different parameter sets (hydraulic, soil, vegetation, all). These uncertainties are also compared with an estimate of the observations' uncertainties.

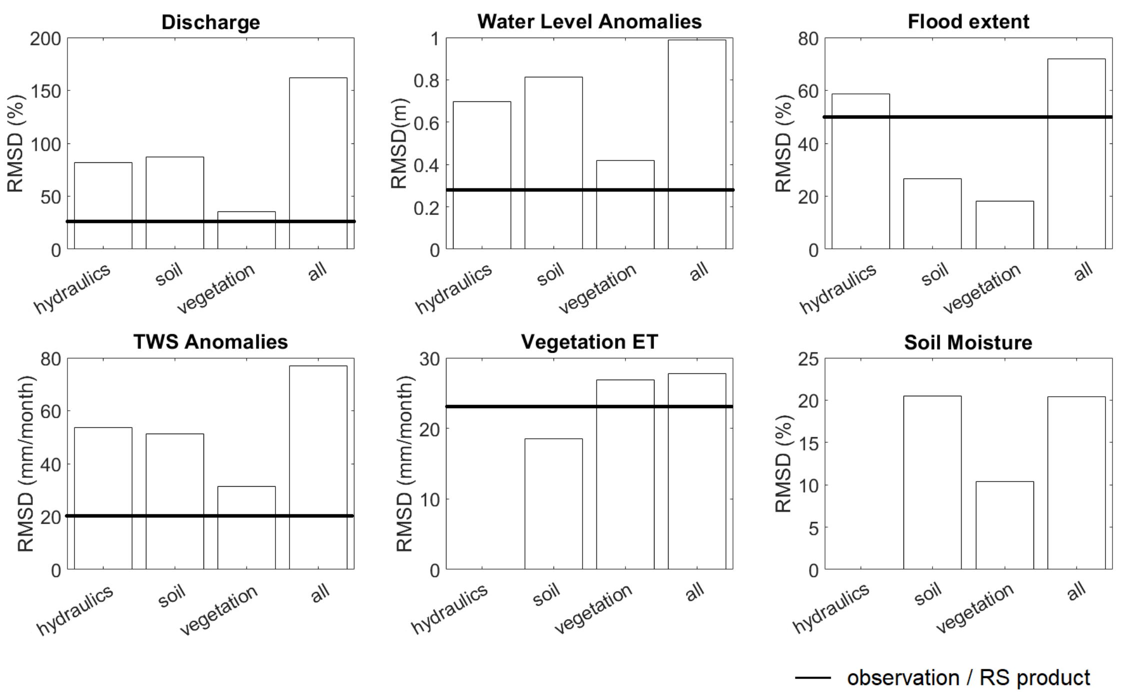


**Figure 3**. Sensitivity analysis of multiple model output variables to varying sets of parameters (hydraulics, soil, vegetation, overall). The a priori uncertainty of the model parameters, for each output variable, is compared to the reported uncertainty for the observation / RS product,

previously described in the Cal/Eval data section (no uncertainty estimation is provided for the soil moisture root zone product given absence of this estimate for the Amazon region).

### 3.1.1 How do model uncertainties relate to uncertainties in observations?

Some variables present observations/RS products that have uncertainties significantly lower than the overall uncertainties of the model, e.g., 25 % for discharge observations, while model overall parameter uncertainty is ~160%. This pattern is also found for water level and TWS estimates, and implies that these observations might be useful to constrain the model. On the other hand, uncertainties in RS products of flood extent (~50%) and vegetation ET (~23%) are in the same order of magnitude of model overall parameter uncertainty, which might hamper their contribution for model calibration, due to their high uncertainties.

### 3.1.2 Which sets of parameters are related to which variables?

The overall uncertainties in the model are related to differing sets of parameters: discharge, water level, and TWS are more strongly related to hydraulics and soil parameters, and to a lesser extent to vegetation parameters. Flood extent estimates are strongly related to hydraulic parameters, and less to soil and vegetation. As expected, soil moisture and vegetation ET estimates relate to vertical water balance processes, therefore they are insensitive to hydraulic parameters. Soil moisture (SM) is more sensitive to soil parameters, while vegetation ET is more sensitive to vegetation parameters. Therefore, if model calibration with either of these variables (ET or SM) is achieved through optimization of hydraulic parameters, it would highlight that the model would have “gotten the right results for the wrong reasons”.

### 3.1.3 Which variables are related between each other?

By varying all parameters, there is a high correlation (greater or equal to 0.4) between discharge and flood extent, water level and flood extent, flood extent and TWS, and ET and TWS (Figure 4). High correlations between discharge, water level and flood extent were expected because these variables are strongly associated through river transport processes. However, it is surprising that correlation between discharge and water level is not too high (0.30), and this is probably due to high uncertainties in hydraulic parameters. Furthermore, high correlations between TWS and flood extent might be related to surface water storage dynamics which are specific for regions with floodplains.

In general, a high correlation between variables in Figure 4 should be reflected in positive results when calibrating with a given variable and evaluating with the other highly correlated variable (one-at-a-time calibration). This may also indicate that observations of these variables are redundant. On the other hand, high correlations in

Figure 4 followed by deterioration after the one-at-a-time calibration process might indicate structural errors in the model, or in the observations. Conversely, low correlations in Figure 4, followed by improvement in performances with the calibration with multiple variables, might indicate complementarity between variables.

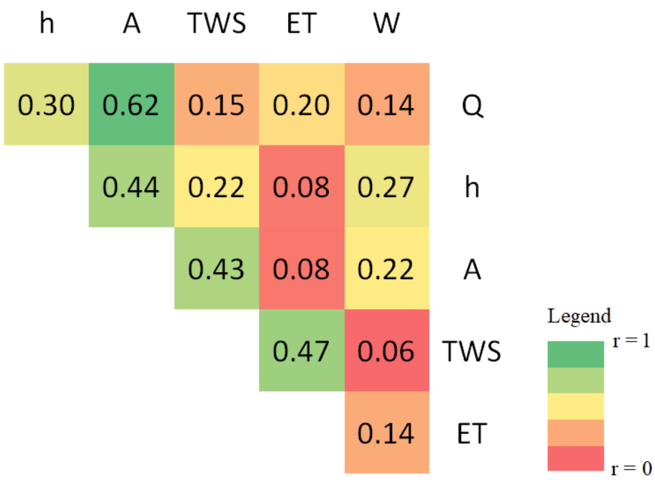


**Figure 4**. Correlation matrix (Pearson coefficient) between performance metrics (KGE based on a reference simulation) for the six analyzed variables, by varying all parameters together.

## 3.2 Model calibration

### 3.2.1 How RS-based model calibration improves discharge estimates?

Calibration with RS products of TWS, vegetation ET and soil moisture led to improvements in discharge estimates (Figure 5a). Nonetheless, RS products of water level and flood extent led to overestimation of discharge estimates in wet periods (Figure 5a).

This could be due to high uncertainties in the observations (Figure 3), but if this was the case, it would also be reflected in a poor performance for water level and flood extent when discharge is the target variable for calibration (Figure 5b), which does not occur. Therefore, calibration with discharge leads to reasonable parameter sets for the performance of discharge itself, and also water level and flood extent. However, it does not lead to the best hydraulic arrangement, which might be achieved more successfully when calibrating with water level or flood extent.

On the other hand, both water level and flood extent observations are representative of a specific location in the basin (Figure 2), and calibration with these variables might lead to the best parameter arrangement for these locations, but not for the whole watershed. A more spatially-consistent use of these observations should improve their usability to constrain models and improve discharge estimates, such as the studies of Kittel et al. (2018), that used radar altimetry measurements at 12 locations in the basin, Schneider et al. (2017), that used data from 13 virtual stations, or Liu et al. (2015), that used water level measurements at 4 virtual stations, and flood extent for stream segments at different locations in the basin.

In spite of the limitations with water level and flood extent variables for discharge prediction in this study, other RS variables, such as TWS, ET, and soil moisture were able to improve discharge estimates by 16.1%, 48.4%, and 26.3% (Figure 6), which is especially relevant in the context of the Prediction in Ungauged Basins initiative (Hrachowitz et al., 2013; Sivapalan et al., 2003). These results agree with previous studies, such as López et al. (2017) that found good performances in discharge estimates by model calibration with GLEAM ET and ESA CCI soil moisture, or Nijzink et al. (2018), that found improvements in discharge by using soil moisture products (AMSR-E, ASCAT) and TWS from GRACE.

The multi-variable calibration experiment considering all variables except discharge (Figure 6b) resulted in a Skill Score of 3.1% for discharge, which is low, and might reflect the inability of retrieving discharge measurements based on the calibration of RS-derived variables. This is probably because of the calibration scheme setup, which combines too many constraints (five objective functions). This limits the degrees of freedom in the calibration procedure, and leads to a fast convergence because the parameter search space is not appropriately explored. For instance, it had an average of 28 iterations while variables as discharge or flood extent took 309 and 173 iterations, on average, respectively, to converge. Moreover, all uncertainty from RS observations are incorporated into the calibration. An alternative to deal with uncertainties from RS observations in the calibration procedure would be to explicitly include them into the objective functions (Aires, 2014; Croke, 2009; Foglia et al., 2009; Peña-Arancibia et al., 2015).

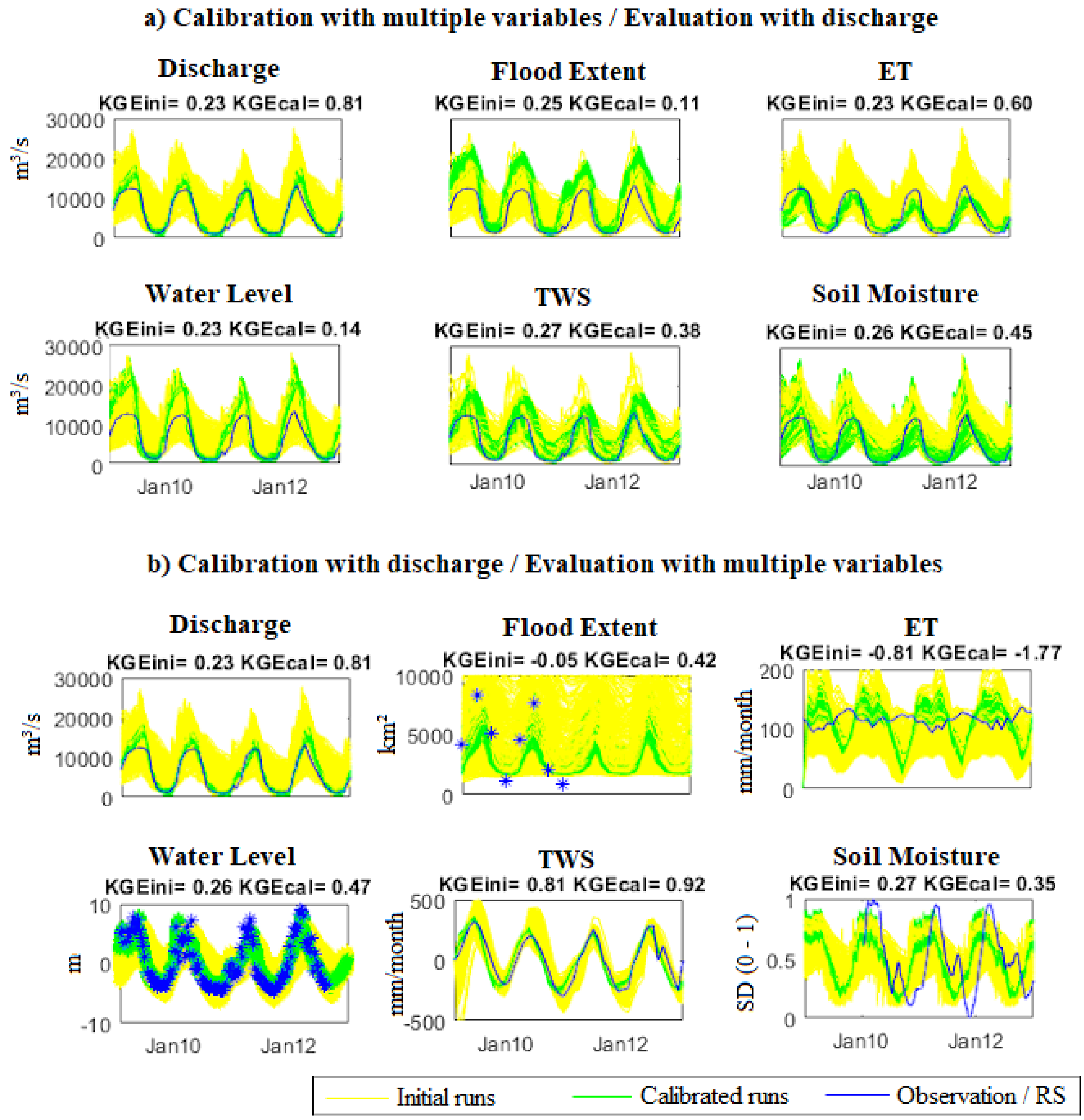


**Figure 5**. (a) Time series of discharge, when calibrating the model with six different variables. (b) Time series of the six variables when calibrating the model with discharge observations only. *KGEini* is the mean KGE of initial runs, and *KGEcal* the mean KGE of calibrated runs. Time series for all variables by calibrating the model with all setups is presented in supporting information (Figure S1).

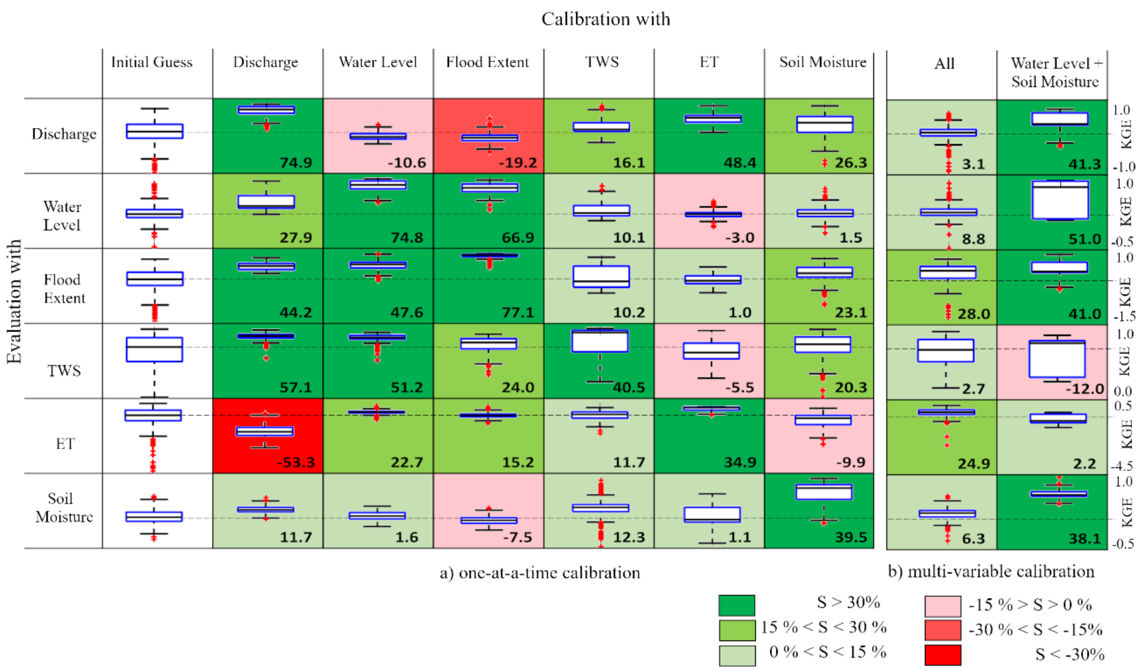


**Figure 6**. Boxplots of skill score for the evaluation of multiple variables with the (a) one-at-a-time (discharge, water level, flood extent, TWS, vegetation ET, soil moisture) and (b) multi-variable calibration (all except discharge, water level + soil moisture). Colors refer to classes of skill score. Please note that the KGE scales are different for each variable.

### 3.2.2 How RS-based model calibration improves representation of the water cycle?

When performing a one-at-a-time calibration, the performance of the variable itself always gets improved, which is evidenced by the green main diagonal (Figure 6a). Calibration with water level was also able to improve estimates of flood extent, TWS and ET; calibration with flood extent improved water level, TWS and ET; calibration with TWS slightly improved all variables, but to a lesser extent; calibration with ET was able to improve discharge estimates; and calibration with soil moisture improved discharge, flood extent and TWS.

In a perfect modeling framework, calibration with any variable should improve the performance of all other variables. However, we have identified that this did not happen in our experiments. This can be due to uncertainties in model structure, in parameterization, or in the observations. Previous studies have also found significant advantages in using RS-based model calibration in order to identify structural model issues (e.g., Werth et al., 2009; Willem Vervoort et al., 2014; Winsemius et al., 2008), detect uncertainties in input data (e.g., Milzow et al., 2011), identify deficiencies in model parameterization (e.g., Franks et al., 1998; Koppa et al., 2019), or increase model reliability (e.g., Koch et al., 2018; Manfreda et al., 2018).

According to what has already been presented in Figure 5b and supporting information (Figure S1), calibration with discharge improved estimates of discharge itself, water level, flood extent, TWS, and soil moisture, to a lesser extent. However, calibration with discharge surprisingly deteriorated the performance for vegetation ET time series. Vegetation ET estimated by MOD16 varies at maximum 30mm/month, while MGB calibration with discharge observations led to variations of 100 mm/month in vegetation ET, reaching around 30 mm/month in the driest periods, while MOD16 estimates are limited to a minimum of 100 mm/month in these periods (time series in Figure 5b). However, one can notice that not even the seasonality between MGB and MOD16 time series agree. This could be due to relatively high uncertainties in vegetation ET estimates from MOD16 for the Amazon basin (around 23 mm/month, according to Gomis-Cebolla et al., 2019). Nonetheless, it could also be related to model structural and/or parameter deficiencies, in which case the model might be “right for the wrong reasons”. In order to identify the source of this ET inconsistency, we have compared MOD16 and MGB results to in-situ measurements of ET in Purus River Basin, provided by Gomis-Cebolla et al. (2019) and Maeda et al. (2017). We found a much stronger agreement both in seasonality and in amplitude of in-situ observations with MOD16 observations than with MGB model output. Hasler & Avissar (2007) have already warned about the overestimation of dry season water stress in hydrological models, probably related to the misrepresentation of soil water availability for plants. This was also found by Maeda et al. (2017), which highlighted that ET did not necessarily reach the lowest values during the driest periods, because of the plants’ access to deep soil water, which has also been previously documented by Nepstad et al. (1994). They found that, in the Southern Amazon ecotone, deep root water intake plays a key role in maintaining ecosystem productivity during dry season. MGB model is probably misrepresenting these processes, which would remain unknown if only discharge time series were observed.

Even though calibration with discharge observations was not able to retrieve ET estimates, calibration with the remaining variables (except for soil moisture) was able to improve ET estimates. For instance, in Figure 4, ET and water level presented low correlation (r=0.08), but calibration with water level improved ET estimates by 22.7%. On the other hand, in Figure 4, ET and TWS presented high correlation (r=0.47), but calibration with TWS improved ET estimates by only 11.7%.

In general, calibration with TWS did not present much influence on any of the variables. Consistently, TWS estimates got relatively easily improved by calibration with any variable (except ET). These results for TWS contrast with previous work from Lo et al., 2010; Nijzink et al., 2018; Rakovec et al., 2016; Schumacher et al., 2018; and Werth & Güntner, 2010, which highlighted the valuable nature of GRACE data when incorporated into hydrological modeling. This can be due to the high seasonality of Purus River Basin, in which TWS does not aggregate much information, biasing the calibration with high correlation values. Even for an uncalibrated setup TWS performances were very good: KGE values were around 0.8, while for all other variables, except for ET (for which KGE values were negative), KGE values were around 0.3 for the uncalibrated setup. A future development could involve a deseasonalized TWS into the calibration scheme.

Flood extent and water level performances were highly improved by calibration of discharge, water level and flood extent, but it did not affect much ET (which actually was degraded with discharge calibration) and soil moisture. This is probably due to the relationship between water level and flood extent with river transport processes (e.g., flood routing and floodplain storage), while ET and soil moisture are more related to vertical hydrological processes (e.g., soil water balance). This highlights the complementarity between variables that relate to different processes.

Soil moisture does not present high correlations with other variables, thus, coherently, calibration with other variables does not present a large influence on soil moisture performance. Calibration with soil moisture, on the other hand, improves performances of all variables (water level to a lesser extent), except for ET.

#### 3.2.3 What is the added value of complementary RS observations?

By calibrating with all variables (Figure 6b), we found improvements for all variables, with the most significant improvements for flood extent (S = 28.0%) and ET (S = 24.9%). However, Skill Score for discharge performance was S = 3.1%, which is low, and might reflect the inability of retrieving discharge measurements based on the calibration of the RS-derived variables (as discussed previously).

Therefore, we chose a specific arrangement of two complementary variables in order to check if this calibration setup might lead to better retrievals for discharge and the other variables. The chosen variables were soil moisture, and water level, because of their complementarity, based on the Skill Score values in Figure 6: calibration with water level improves all variables but discharge (and soil moisture to a lesser extent), while calibration with soil moisture improves all variables, but ET (and water level to a lesser extent).

Results showed an improvement for all variables (ET to a lesser extent), except for TWS. The calibration arrangement of water level and soil moisture led to improvements not only to soil moisture and water level themselves, but also to flood extent (S = 41.0 %; mean KGE = 0.39) and discharge (S = 41.3 %; mean KGE = 0.57), which is extremely relevant in the context of the PUB initiative (Hrachowitz et al., 2013; Sivapalan et al., 2003). These results agree with previous works that found an improvement in model performances by multi-variable calibration of soil moisture and evapotranspiration (e.g., Koppa et al., 2019; López et al., 2017), discharge and evapotranspiration (e.g., Herman et al., 2018; Pan et al., 2018; Poméon et al., 2018), discharge and soil moisture (e.g., Li et al., 2018; Rajib et al., 2016), discharge and TWS (e.g., Rakovec et al., 2016; Schumacher et al., 2018; Werth & Güntner, 2010), and discharge and water level (e.g., Kittel et al., 2018; Schneider et al., 2017; W. Sun et al., 2012). However, it is difficult to compare this study to previous works, because most of them used discharge observations as constraints. In this study, we avoided the use of discharge observations for multi-variable calibration, in order to analyze the applicability of the method for poorly-gauged regions.

Even though TWS presents a negative Skill Score, its mean KGE value is 0.8, which is still relatively high. Calibration with water level and soil moisture did not present much influence on ET performance, because of the specificities regarding ET in this watershed, i.e., given that the model setup that does not represent deep root water intake during dry season, as discussed previously.

By comparing the two frameworks for multi-variable calibration (all versus two selected variables), we found that calibration with all variables is useful to some extent, but consistently selecting complementary variables for model calibration resulted in best overall performance.

### 3.3 Are we getting the right results for the right sets of parameters?

When analyzing the dispersions of parameters before and after calibration with each variable (Figure 7 for selected Supporting Information (Figure S2) for all calibrated parameters), it can be observed that the range of parameters vary largely depending on the calibration variable. For instance, Wm is a soil conceptual parameter that relates to maximum storage of water in the soil. In the calibration based on all variables but ET it converged to low values (300), while in the calibration with ET it reached high values (2000). This probably occurred in order to compensate, by parameterization, a structural error in the model, i.e., the model inability to represent deep root water uptake in dry season. These trade-offs between model parameters during calibration has also been reported and discussed by Koppa et al. (2019).

While all variables are sensitive to soil parameters, soil moisture and ET are insensitive to hydraulic parameters (Figure 3), and thus calibrating with ET or soil moisture should not change hydraulic parameters. However, it does. Therefore, this highlights equifinality issues in model parameterization, and that the model might be getting the right results for the wrong sets of parameters. Previous studies reported reduction in equifinality by using a multi-variable calibration framework (e.g., Pan et al., 2018; Silvestro et al., 2015; Wambura et al., 2018; Zink et al., 2018), but this was not verified in this study.

Many previous studies have also highlighted the use of multi-variable calibration to narrow parameters' search space (Nijzink et al., 2018; W. Sun et al., 2018), but this is not observed in our results: for most parameters (except for Wm), calibration with the combination of water level and soil moisture resulted in a wide range of values, being in some cases similar to the initial range (due to insensitivity of the parameter to a given variable), and in some cases highly dispersed (e.g., vegetation height; Figure S2 in Supporting Information). This can be due to differing convergence sets of parameters between each of the triplicate runs. One interesting result relates to channel Manning's coefficient, which was estimated with median values higher than 0.045 when calibrating with water level and flood extent, while with the other hydrological variables it got smaller values (Figure S2 in supporting information). This highlights the equifinality problem of Manning parameter, which has been studied in details in the literature (Neal a et al., 2015).

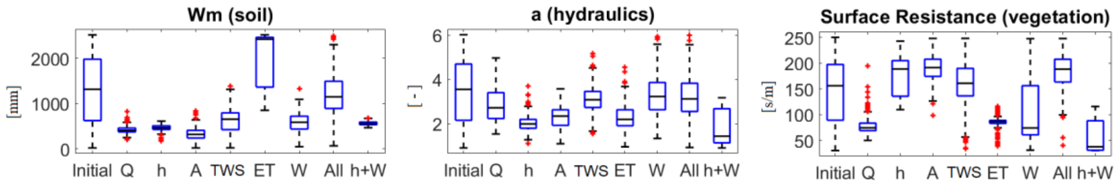


**Figure 7**. Boxplots of dispersion of three model parameters before (Initial) and after the one-at-a-time calibration (Q – discharge; h – water level; A – flood extent; TWS – total water storage anomalies; ET - vegetation ET; W – soil moisture), and multi-variable calibration (All – variables except discharge; h+W – water level and soil moisture). Description of parameters is presented in Supporting Information (Table S2). A complete figure with boxplots for all parameters is presented in Supporting Information (Figure S2).

# 4 Conclusion

We calibrated and evaluated a hydrological-hydrodynamic model with five different RS-based observations of the water cycle: water levels (Jason-2), flood extent (ALOS-PALSAR), TWS (GRACE), vegetation ET (MOD16), and soil moisture (SMOS), for a study basin in a tropical region with floodplains (Purus River Basin in the Amazon), and analyzed the redundancy and complementarity between different variables and processes.

Results showed that calibration with current RS observations was able to improve discharge estimates. For instance, calibration with TWS improved discharge estimates by 16.1% in comparison to an uncalibrated model setup, calibration with ET improved discharge estimates by 48.4% and soil moisture by 26.3%, and a joint scheme of water level + soil moisture was able to improve discharge estimates by 41.3%. We conclude that RS observations are useful to retrieve discharge estimates, but the utility of each RS variable might depend on the study area characteristics.

Our results also showed that RS-based calibration led to an overall improvement of the water cycle representation. For instance, calibration with water level was able to improve estimates of water level itself, but also flood extent, TWS and ET; calibration with soil moisture was able to improve estimates of soil moisture itself, but also discharge, flood extent and TWS. This is especially relevant in the context of the PUB initiative (Hrachowitz et al., 2013; Sivapalan et al., 2003).

Moreover, calibration with multiple RS variables was able to highlight deficiencies in model structure and parameterization, and observations. In the context of model structure, for instance, calibration with ET highlighted the model inability to represent the root water intake in dry season in this region, thus compensating it by misrepresenting other variables. In the context of model parameterization, for instance, calibration with soil moisture or vegetation ET should not change hydraulic parameters, but it does, which highlights equifinality issues in model parameterization. This outcome was only visible because we used a tightly coupled hydrological-hydrodynamic model setup, which allows hydrological and hydraulic variables and processes to interact during the calibration process.

Besides individual calibration with each RS variable, we conducted two multi-variable calibration experiments: calibration with all variables except discharge, and calibration with two selected variables. Calibration with all variables was useful to some extent, but appropriately selecting complementary variables for model calibration may result in a better overall performance (in this case, soil moisture and water level).

The main conclusions presented here are of great interest for the hydrological community, and agree with previous works in that RS–based calibration is useful to improve water cycle representation in hydrological models. To further investigate the potentiality of RS data, future developments should test the methodology presented here for multiple basins at contrasting hydro-climatic regions.

## Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Grant Number 41161/2017-5. It was conducted in the context of the SWOT-MOD science team project from SWOT satellite mission. We would also like to thank colleagues from the Large Scale Hydrology Group (HGE/IPH) for general discussions about this study. Data presented in this study are available at <https://drive.google.com/open?id=1OEsqOjIGKM2vIwey6CaMOZg2O7TIkK5K> (MGB code in FORTRAN, MGB Input folder, post-processing code in MATLAB).

## References

Aires, F. (2014). Combining Datasets of Satellite-Retrieved Products. Part I: Methodology and Water Budget Closure. *Journal of Hydrometeorology*. https://doi.org/10.1175/jhm-d-13-0148.1

Al Bitar A.; Kerr, Y. H.; Merlin, O.; Cabot, F.; Wigneron, J. (2013). Global drought index from SMOS soil moisture. Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International.

Alkama, R., Decharme, B., Douville, H., Becker, M., Cazenave, A., Sheffield, J., et al. (2010). Global evaluation of the ISBA-TRIP continental hydrological system. Part I: Comparison to GRACE terrestrial water storage estimates and in situ river discharges. *Journal of Hydrometeorology*. https://doi.org/10.1175/2010JHM1211.1

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., et al. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*. https://doi.org/10.13031/2013.42259

Asadzadeh Jarihani, A., Callow, J. N., Johansen, K., & Gouweleeuw, B. (2013). Evaluation of multiple satellite altimetry data for studying inland water bodies and river floods. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2013.09.010

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

Baroni, G., Schalge, B., Rakovec, O., Kumar, R., Schüler, L., Samaniego, L., et al. (2019). A Comprehensive Distributed Hydrological Modeling Intercomparison to Support Process Representation and Data Collection Strategies. *Water Resources Research*. https://doi.org/10.1029/2018WR023941

Bates, P. D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2010.03.027

Biancamaria, S., Lettenmaier, D.P. & Pavelsky, T.M.  (2016). The SWOT mission and its capabilities for land hydrology. In *Remote Sensing and Water Resources* (pp. 117-147). Springer, Cham.

Beven, K. (2006). A manifesto for the equifinality thesis. In *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*. https://doi.org/10.1002/hyp.3360060305

Brêda, J. P. L. F., Paiva, R. C. D., Bravo, J. M., Passaia, O. A., & Moreira, D. M. (2019). Assimilation of Satellite Altimetry Data for Effective River Bathymetry. *Water Resources Research*. https://doi.org/10.1029/2018wr024010

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*. https://doi.org/10.1002/2015WR017096

Collischonn, B., Collischonn, W., & Tucci, C. E. M. (2008). Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2008.07.032

Collischonn, W., Allasia, D., da Silva, B. C., & Tucci, C. E. M. (2007). The MGB-IPH model for large-scale rainfall-runoff modelling. *Hydrological Sciences Journal*. https://doi.org/10.1623/hysj.52.5.878

Croke, B. F. W. (2009). Representing uncertainty in objective functions: Extension to include the influence of serial correlation. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*.

Crow, W. T., Wood, E. F., & Pan, M. (2003). Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. *Journal of Geophysical Research D: Atmospheres*.

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

Fan, F. M., Buarque, D. C., Pontes, P. R. M., & Collischonn, W. (2015). Um mapa de Unidades de Resposta Hidrológica para a América do Sul. *XXI Simpósio Brasileiro de Recursos Hídricos*.

Foglia, L., Hill, M. C., Mehl, S. W., & Burlando, P. (2009). Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resources Research*. https://doi.org/10.1029/2008WR007255

Franks, S. W., Gineste, P., Beven, K. J., & Merot, P. (1998). On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research*. https://doi.org/10.1029/97WR03041

Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. V., & Savenije, H. H. G. (2014). A constraint-based search algorithm for parameter identification of environmental models. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-18-4861-2014

Gomis-Cebolla, J., Jimenez, J. C., Sobrino, J. A., Corbari, C., & Mancini, M. (2019). Intercomparison of remote-sensing based evapotranspiration algorithms over amazonian forests. *International Journal of Applied Earth Observation and Geoinformation*. https://doi.org/10.1016/j.jag.2019.04.009

Grimaldi, S., Schumann, G. J. P., Shokri, A., Walker, J. P., & Pauwels, V. R. N. (2019). Challenges, Opportunities, and Pitfalls for Global Coupled Hydrologic-Hydraulic Modeling of Floods. *Water Resources Research*. https://doi.org/10.1029/2018WR024289

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. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*. https://doi.org/10.1029/2006GL026047

Hasler, N., & Avissar, R. (2007). What controls evapotranspiration in the Amazon basin? *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM587.1

Herman, M. R., Nejadhashemi, A. P., Abouali, M., Hernandez-suarez, S., Daneshvar, F., Zhang, Z., et al. (2017). Evaluating the Role of Evapotranspiration Remote Sensing Data in Improving Hydrological Modeling Predictability. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2017.11.009

Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2003.04.001

Hodges, B. R. (2013). Challenges in continental river dynamics. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2013.08.010

Holeman, J. N. (1968). The Sediment Yield of Major Rivers of the World. *Water Resources Research*. https://doi.org/10.1029/WR004i004p00737

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*. https://doi.org/10.1029/1998WR900001

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., et al. (2013). A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2013.803183

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM560.1

Jeffrey C. Neal a, ⇑, Nicholas A. Odoni a, b, Mark A. Trigg a, Jim E. Freer a, Javier Garcia-Pintado c, d, David C. Mason e, et al. (2015). Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *Journal of Hydrology*.

Jiang, D., & Wang, K. (2019). The Role of Satellite-Based Remote Sensing in Improving Simulated Streamflow: A Review. *Water*. https://doi.org/10.3390/w11081615

Junk, W. J. (1997). General Aspects of Floodplain Ecology with Special Reference to Amazonian Floodplains. https://doi.org/10.1007/978-3-662-03416-3_1

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., & Wood, E. F. (2017). Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons. *Advances in Water Resources*. https://doi.org/10.1016/j.advwatres.2017.09.010

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. M., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/36.942551

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*. https://doi.org/10.1029/2005WR004362

Kittel, C., Nielsen, K., Tøttrup, C., & Bauer-Gottwein, P. (2018). Informing a hydrological model of the Ogooué with multi-mission remote sensing data. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-22-1453-2018

Koch, J., Demirel, M. C., & Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-11-1873-2018

Koppa, A., Gebremichael, M., & Yeh, W. W. G. (2019). Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. *Advances in Water Resources*. https://doi.org/10.1016/j.advwatres.2019.06.005

Lambin, J., Morrow, R., Fu, L. L., Willis, J. K., Bonekamp, H., Lillibridge, J., et al. (2010). The OSTM/Jason-2 Mission. *Marine Geodesy*. https://doi.org/10.1080/01490419.2010.491030

Lee, H., Jung, H. C., Yuan, T., Beighley, R. E., & Duan, J. (2014). Controls of Terrestrial Water Storage Changes Over the Central Congo Basin Determined by Integrating PALSAR ScanSAR, Envisat Altimetry, and GRACE Data. In *Remote Sensing of the Terrestrial Water Cycle*. https://doi.org/10.1002/9781118872086.ch7

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*. https://doi.org/10.1002/2015WR017616

Li, Y., Grimaldi, S., Pauwels, V. R. N., & Walker, J. P. (2018). Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2018.01.013

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*. https://doi.org/10.1029/94jd00483

Lo, M. H., Famiglietti, J. S., Yeh, P. J. F., & Syed, T. H. (2010). Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resources Research*. https://doi.org/10.1029/2009WR007855

López, P. L., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-21-3125-2017

Maeda, E. E., Ma, X., Wagner, F. H., Kim, H., Oki, T., Eamus, D., & Huete, A. (2017). Evapotranspiration seasonality across the Amazon Basin. *Earth System Dynamics*. https://doi.org/10.5194/esd-8-439-2017

Manfreda, S., Mita, L., Dal Sasso, S. F., Samela, C., & Mancusi, L. (2018). Exploiting the use of physical information for the calibration of a lumped hydrological model. *Hydrological Processes*. https://doi.org/10.1002/hyp.11501

Maurer, E. P., Adam, J. C., & Wood, A. W. (2009). Climate model based consensus on the hydrologic impacts of climate change to the Rio Lempa basin of Central America. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-13-183-2009

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

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research D: Atmospheres*.

Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards: Stationarity is immortal! *Water Resources Research*. https://doi.org/10.1002/2014WR016092

Motovilov, Y. G., Gottschalk, L., Engeland, K., & Rodhe, A. (1999). Validation of a distributed hydrological model against spatial observations. *Agricultural and Forest Meteorology*. https://doi.org/10.1016/S0168-1923(99)00102-1

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2011.02.019

Naz, B. S., Frans, C. D., Clarke, G. K. C., Burns, P., & Lettenmaier, D. P. (2014). Modeling the effect of glacier recession on streamflow response using a coupled glacio-hydrological model. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-18-787-2014

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*. https://doi.org/10.1029/2012WR012514

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2016.1183009

Nepstad, D. C., De Carvalho, C. R., Davidson, E. A., Jipp, P. H., Lefebvre, P. A., Negreiros, G. H., et al. (1994). The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. *Nature*. https://doi.org/10.1038/372666a0

New, M., Hulme, M., & Jones, P. (2000). Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate*. https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining Conceptual Hydrological Models With Multiple Information Sources. *Water Resources Research*. https://doi.org/10.1029/2017WR021895

O'Loughlin, F. E., Paiva, R. C. D., Durand, M., Alsdorf, D. E., & Bates, P. D. (2016). A multi-sensor approach towards a global vegetation corrected SRTM DEM product. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2016.04.018

Paiva, R. C.D., Collischonn, W., Bonnet, M. P., De Gonçalves, L. G. G., Calmant, S., Getirana, A., & Santos Da Silva, J. (2013). Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-17-2929-2013

Paiva, Rodrigo C.D., Collischonn, W., & Tucci, C. E. M. (2011). Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2011.06.007

De Paiva, Rodrigo Cauduro Dias, Buarque, D. C., Collischonn, W., Bonnet, M. P., Frappart, F., Calmant, S., & Bulhões Mendes, C. A. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research*. https://doi.org/10.1002/wrcr.20067

Pan, M., & Wood, E. F. (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM495.1

Pan, S., Liu, L., Bai, Z., & Xu, Y. P. (2018). Integration of remote sensing evapotranspiration into multi-objective calibration of distributed hydrology-soil-vegetation model (DHSVM) in a humid region of China. *Water (Switzerland)*. https://doi.org/10.3390/w10121841

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Hydrologic modeling in dynamic catchments: A data assimilation approach. *Water Resources Research*. https://doi.org/10.1002/2015WR017192

Pellet, V., Aires, F., Munier, S., Fernández Prieto, D., Jordá, G., Arnoud Dorigo, W., et al. (2019). Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle - Application to the Mediterranean region. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-23-465-2019

Peña-Arancibia, J. L., Zhang, Y., Pagendam, D. E., Viney, N. R., Lerat, J., van Dijk, A. I. J. M., et al. (2015). Streamflow rating uncertainty: Characterisation and impacts on model calibration and performance. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2014.09.011

Poméon, T., Diekkrüger, B., & Kumar, R. (2018). Computationally efficient multivariate calibration and validation of a grid-based hydrologic model in sparsely gauged West African river basins. *Water (Switzerland)*. https://doi.org/10.3390/w10101418

Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., de Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., et al. (2017). MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2017.03.029

Rajib, M. A., Merwade, V., & Yu, Z. (2016). Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2016.02.037

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*. https://doi.org/10.1002/2016WR019430

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*. https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2

Rosenqvist, A., Shimada, M., Ito, N., & Watanabe, M. (2007). ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment. In *IEEE*

*Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2007.901027

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*. https://doi.org/10.1029/2008WR007327

Schneider, R., Nygaard Godiksen, P., Villadsen, H., Madsen, H., & Bauer-Gottwein, P. (2017). Application of CryoSat-2 altimetry data for river analysis and modelling. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-21-751-2017

Schumacher, M., Forootan, E., van Dijk, A. I. J. M., Müller Schmied, H., Crosbie, R. S., Kusche, J., & Döll, P. (2018). Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.10.029

Semenova, O., & Beven, K. (2015). Barriers to progress in distributed hydrological modelling. *Hydrological Processes*. https://doi.org/10.1002/hyp.10434

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*. https://doi.org/10.1002/2014WR016520

Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., & Boni, G. (2015). Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-19-1727-2015

Siqueira, V., Fleischmann, A., Jardim, P., Fan, F., & Collischonn, W. (2016). IPH-Hydro Tools: a GIS coupled tool for watershed topology acquisition in an open-source environment. *Revista Brasileira de Recursos Hídricos*. https://doi.org/10.21168/rbrh.v21n1.p274-287

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., et al. (2018). Toward continental hydrologic-hydrodynamic modeling in South America. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-22-4815-2018

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*. https://doi.org/10.1623/hysj.48.6.857.51421

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

Sun, W., Fan, J., Wang, G., Ishidaira, H., Bastola, S., Yu, J., et al. (2018). Calibrating a hydrological model in a regional river of the Qinghai–Tibet plateau using river water width determined from high spatial resolution satellite images. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2018.05.020

Sun, W. C., 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*. https://doi.org/10.5194/hess-14-2011-2010

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*. https://doi.org/10.1126/science.1099192

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*. https://doi.org/10.1002/hyp.9550

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2012.05.052

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*. https://doi.org/10.1029/2004WR003059

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. *Hydrological Processes*. https://doi.org/10.1002/hyp.1135

Wambura, F. J., Dietrich, O., & Lischeid, G. (2018). Improving a distributed hydrological model using evapotranspiration-related boundary conditions as additional constraints in a data-scarce river basin. *Hydrological Processes*. https://doi.org/10.1002/hyp.11453

Werth, S., & Güntner, A. (2010). Calibration analysis for water storage variability of the global hydrological model WGHM. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-14-59-2010

Werth, S., Güntner, A., Petrovic, S., & Schmidt, R. (2009). Integration of GRACE mass variations into a global hydrological model. *Earth and Planetary Science Letters*. https://doi.org/10.1016/j.epsl.2008.10.021

Willem Vervoort, R., Miechels, S. F., van Ogtrop, F. F., & Guillaume, J. H. A. (2014). Remotely sensed evapotranspiration to calibrate a lumped conceptual model: Pitfalls and opportunities. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2014.10.034

Winsemius, H. C., G. Savenije, H. H., & M. Bastiaanssen, W. G. (2008). Constraining model parameters on remotely sensed evaporation: Justification for distribution in ungauged basins? *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-12-1403-2008

Xu, C. Y., Widén, E., & Halldin, S. (2005). Modelling hydrological consequences of climate change - Progress and challenges. *Advances in Atmospheric Sciences*. https://doi.org/10.1007/BF02918679

Xu, X., Li, J., & Tolson, B. A. (2014). Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography*. https://doi.org/10.1177/0309133314536583

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*. https://doi.org/10.1016/S0022-1694(97)00107-8

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2017.03.022

Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature. *Water Resources Research*. https://doi.org/10.1002/2017WR021346