On the contribution of remote sensing-based calibration to model multiple hydrological variables in tropical regions

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)

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. While calibration is usually performed with discharge estimates, the internal model processes might be misrepresented, and the model might be getting the "right results for the wrong reasons", thus compromising model reliability. An alternative is to calibrate model parameters with remote sensing (RS) observations of the water cycle. Previous studies highlighted the potential of RS-based calibration to improve discharge estimates, focusing less on other variables of the water cycle. In this study, we analyzed in detail the contribution of five RS-based variables (water level (h), flood extent (A), terrestrial water storage (TWS), evapotranspiration (ET) and soil moisture (W)) to calibrate a coupled hydrologic-hydrodynamic model for a large Amazon sub-basin with extensive floodplains. Single-variable calibration experiments with all variables were able to improve discharge KGE from around 6.1% to 52.9% when compared to a priori parameter sets. Water cycle representation was improved with multi-variable calibration: KGE for all variables were improved in the evaluation period. By analyzing different calibration setups, a consistent selection of

*Present address: Department of Geography, University of Zurich, Zurich, Switzerland.

complementary variables for model calibration resulted in a better performance than incorporating all RS variables into the calibration. By looking at multiple RS observations of the water cycle, inconsistencies in model structure and parameterization were found, which would remain unknown if only discharge observations were considered.

Keywords: hydrological modeling, multi-variable calibration, Amazon, hydrodynamic modeling, large basins.

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, since it does not guarantee an accurate representation of other water cycle components (e.g., soil moisture and evapotranspiration), thus leading to uncertainty in hydrologic predictions (Hrachowitz et al., 2013). Moreover, many parameter sets can provide equally acceptable performances for streamflow evaluation (i.e., the equifinality thesis), 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). However, these are ongoing developments, and stand out as one of the twenty-three unsolved problems in hydrology (Blöschl et al., 2019): "how can we disentangle and reduce model structural/parameter/input uncertainty in hydrological prediction?".

In addition to the presented solutions, an alternative is the use of complementary datasets besides streamflow observations for model calibration (e.g., Crow et al., 2003; Franks et al., 1998; Lo et al., 2010; López et al., 2017; Rajib et al., 2016), 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), or validation (e.g., Alkama et al., 2010; Motovilov et al., 1999; Neal et al., 2012; Siqueira et al., 2018). Such approaches are promising to improve representation of processes in hydrological models (Clark et al., 2015), reduce uncertainty in hydrological predictions (Gharari et al., 2014), understand equifinality (Beven, 2006), and perform predictions in ungauged or poorly-gauged sites (Sivapalan et al., 2003). However, distributed data of complementary hydrological variables (e.g., evapotranspiration, soil moisture) are scarce, and in-situ measurements present poor spatial and temporal representativeness.

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 uniformly estimated over land and oceans (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), while the future SWOT mission will provide at least one water level measurement every 21 days for global rivers wider than 100 m (Biancamaria et al., 2016).

Although previous studies have analyzed the value of integrating RS data into hydrological modeling through calibration or data assimilation (see review by Xu et al., 2014 and Jiang & Wang, 2019), this topic has not been fully explored to its potential yet. Therefore, in section 1.1, we present major knowledge gaps in the context of RS-based calibration of hydrological models through an extensive literature review. In section 1.2, we describe the aims and contributions of this study.

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 (Supplementary Material Table S1). Most publications involved large study areas (> 1000 km²), which is expected because of the usual coarse resolution of RS products. Most studies used RS-derived evapotranspiration for model calibration, followed by soil moisture (Figure 1b), but there were also 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 investigated the added value of RS-derived information to calibrate hydrological models (Figure 1d; Table S1).

Most of the studies (69.35%) used only one RS product for model calibration (Figure 1e, in black), while twelve studies (19.35%) used two products, and five (8.06%) used three products. Only few studies used more than three RS products for model calibration (Demirel et al., 2019; Nijzink et al., 2018). Some studies addressed the use of RS data to estimate discharge in ungauged basins (Kittel et al., 2018; Sun et al., 2010), while others focused on narrowing the parameter search space, and thus equifinality reduction, by combining multiple variables 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 (Figure 1e, in red), providing a more comprehensive discussion on inconsistencies of hydrological models (e.g., Koch et al., 2018; Li et al., 2018).

Regarding how RS is incorporated into the model calibration procedure (Figure 1h), 65.6% of the articles used RS-based spatially distributed information, thus calibrating the model with distributed objective functions (e.g., pixel-by-pixel or by sub-basin). Within these studies, bias-insensitive functions have been recently introduced (e.g., Koch et al., 2018; Demirel et al., 2018; Zink et al., 2018; Dembele et al., 2020), being important for reducing the impact of RS data uncertainty on the parameter estimation procedure. The remaining publications (34.4%) incorporated RS data as an average for the whole basin.

Finally, there is still a need for more studies in tropical regions (especially South America) (Figure 1c), which have particular hydro-climatic characteristics, and so have different requirements than temperate regions 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 the case of basin with complex river-floodplain interactions as in the Amazon, an accurate flood wave routing method is required to correctly depict the water transport along the drainage network. Our analysis shows that most studies used simple flood wave routing schemes such as kinematic wave or Muskingum (Figure 1g). Only 10.4% attempted to couple hydrologic and river hydrodynamic models, highlighting the necessity of better understanding the applicability of RS-based calibration in basins with major flat regions with wetlands (Hodges, 2013; Neal et al., 2012; Pontes et al., 2017).

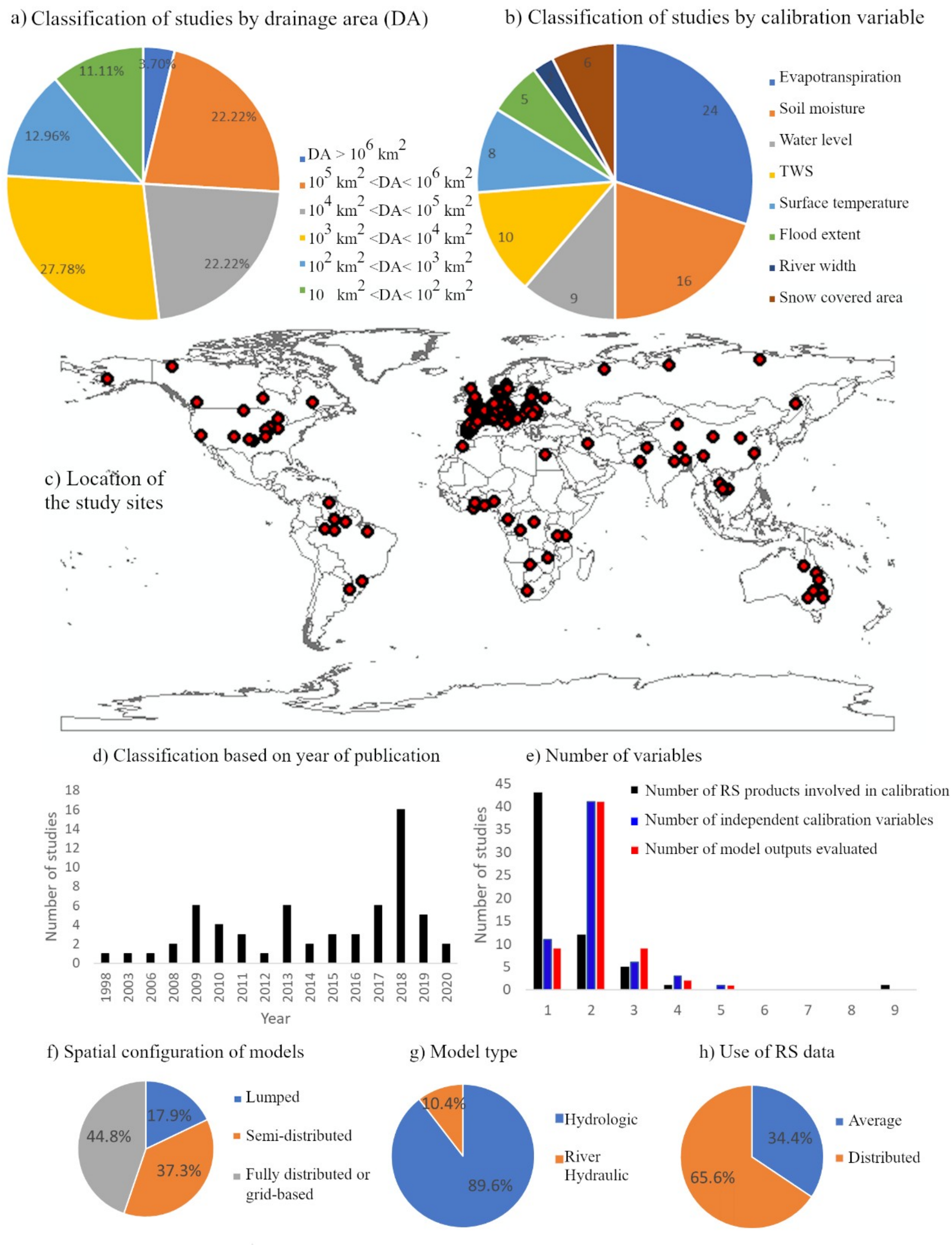

Figure 1. Summary of the literature review on 62 studies that incorporated RS datasets for parameter estimation in hydrological models (see Table S1 in Supplementary Material). (a) Classification of publications based on the drainage area of study sites (an average value was considered for publications that used multiple study sites); (b) distribution of studies based on the calibration variable; (c) geographical distribution of study sites; (d) number of publications

per year; (e) number of RS products involved in calibration (in black), number of independent calibration variables (in blue), and number of model outputs evaluated (in red); (f) classification of models based on their spatial configuration; (g) model type; and (h) use of RS data

1.2 Aims and Contributions of this paper

Our study addresses major knowledge gaps identified in the previous literature review in the context of RS-based calibration of hydrological models. Firstly, most of the studies analyzed two or less variables (Figure 1e). Here, we used RS observations of a large number of variables for model calibration, namely soil moisture, evapotranspiration, terrestrial water storage, flood extent and river water levels, and thus move beyond the contributions of RS for improving only discharge estimates. By simultaneously looking at different variables, we also move towards an improved representation of the water cycle as a whole, enhancing our ability to identify model limitations and inconsistencies. Furthermore, most studies to date focused on European, temperate watersheds (Figure 1c), which largely differ from tropical basins in terms of hydroclimatic characteristics and river-wetland interactions. In this context, large-scale, coupled hydrologic-hydrodynamic models have faced major developments in recent years (Yamazaki et al 2011, Paiva et al 2013, Fleischmann et al 2020), but to our knowledge the complementarity of hydrologic (soil moisture, evapotranspiration, terrestrial water storage) and hydrodynamic (flood extent and river water level) RS observations for model calibration has not yet been addressed in the literature. Here we present a study case in a tropical basin with extensive floodplains in the Amazon with a state-of-the-art coupled hydrologic-hydrodynamic model, which together with the previously mentioned advances provide important contributions to the growing literature of RS-based calibration of hydrological models. This study aims to investigate the applicability of multiple RS observations in an accessible approach to model and represent the water cycle accurately.

2 Methods

2.1 Experimental design

A hydrologic-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 (references in Table S2). The study is then divided into two steps.

Firstly, a sensitivity analysis is performed to understand how different parameter sets (river hydraulic, soil, vegetation) affect model output variables (river discharge, flood extent, river water level, soil moisture, evapotranspiration and terrestrial water storage).

Then, a calibration step is performed in which the model is calibrated with the well-known MOCOM-UA optimization algorithm (Yapo et al., (1998)) considering six variables: (1) in-situ streamflow (one gauge at the basin outlet), and RS freely available, state-of-the-art 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 averaged over the whole basin. The calibration of each variable is performed individually (single-variable), and evaluated for all variables. All calibration experiments are repeated three times with differing initial parameter sets to ensure that convergence is not dependent on the initial parameter sets. Given limitations on the availability of simultaneous RS time coverage, the model is calibrated for one time period (2009-2011), and evaluated for: (i) the same time period of calibration; and (ii) for a different period (2006–2008 for discharge, flood extent, TWS, ET and 2013–2014 for water level and soil moisture). To understand how lumped calibration can retrieve the remotely sensed spatial patterns, a qualitative evaluation is provided additionally. 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 (water level and soil moisture), which are selected for simultaneous calibration for being complementary and having led to satisfactory calibration performance.

2.2 Study area: Purus River Basin

The Purus River Basin (Figure 2) in Amazon has a drainage area of approximately 236,000 km^2, and river discharge ranges from around 1,000 (June-December) to 12,000 m^3/s (January-July) at Canutama gauge. Because of its large area, 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 has minor anthropogenic influences, which simplifies the modeling process. The climate is equatorial (Figure 2d), and mean annual rainfall is 2147 mm/year (according to in-situ gauges). Purus was 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.

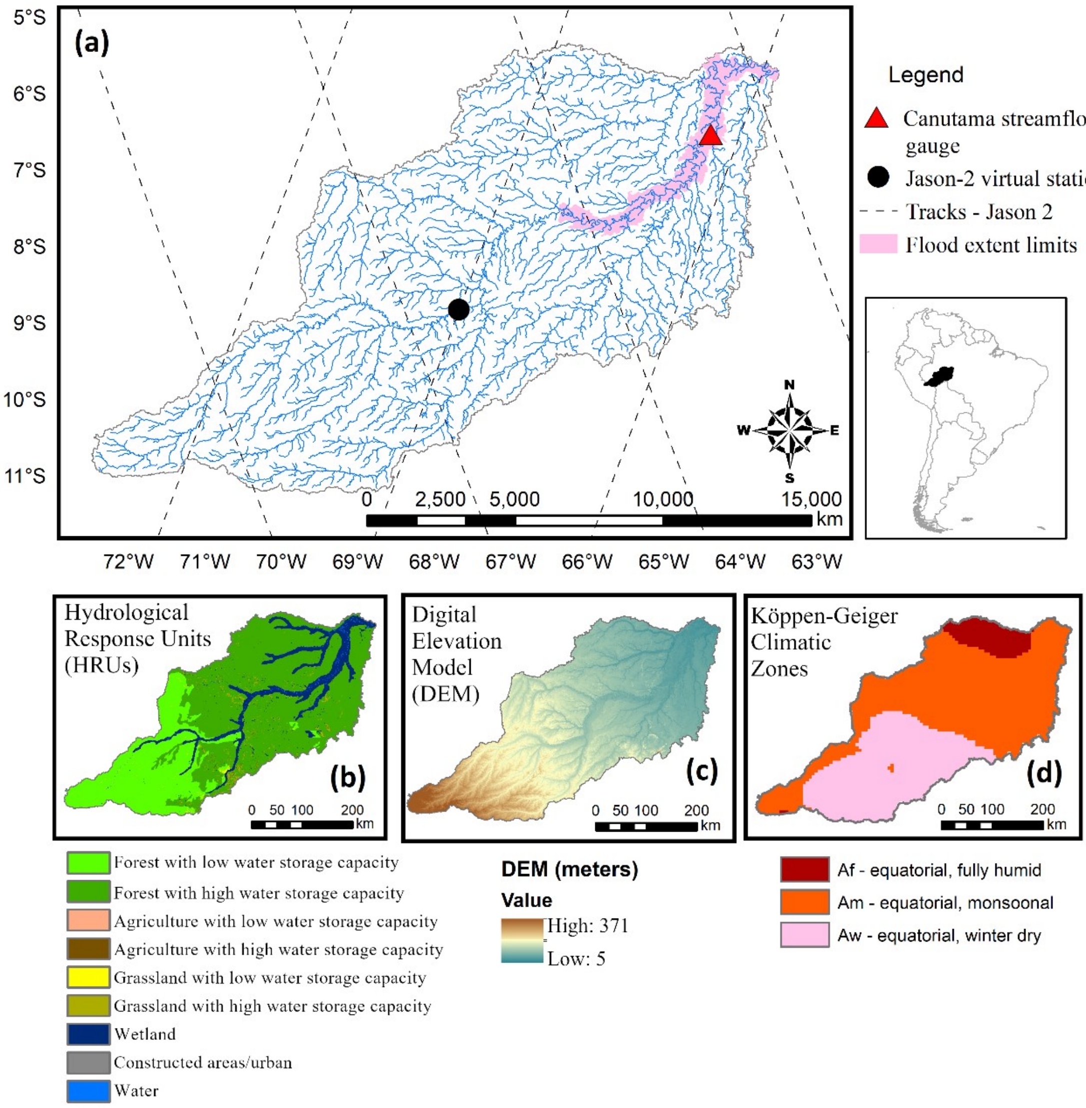

Figure 2. Study area: Purus River Basin. (a) drainage network (in blue), location of the discharge gauge (Canutama, triangle in red), tracks of the spatial altimetry mission Jason 2 (dashed black lines), location of the altimetry virtual station (circle, in black), and the area used for extraction of flood extent (Lower Purus, pink polygons); (b) Hydrological Response Units (Fan et al., 2015); (c) Bare Earth Digital Elevation Model (O'Loughlin et al., 2016); (d) Köppen-Geiger Climatic Zones (Kottek et al., 2006).

2.3 Hydrologic-hydrodynamic model: MGB

The MGB ("Modelo de Grandes Bacias", a Portuguese acronym for "Large Basin Model") is a semi-distributed, hydrologic-hydrodynamic model (Collischonn et al., 2007; Pontes et al., 2017). It was chosen for this study because (1) it has been widely 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 like VIC (Liang et al., 1994) and SWAT; 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. Model parameters are specific for each of the HRUs. A vertical water balance is performed for each HRU, considering canopy interception, soil infiltration, evapotranspiration, and generation of surface, subsurface and groundwater flows. Soil is represented as a bucket model with a single layer. 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. The model has 19 parameters, which are further detailed in the next section. Other model inputs are precipitation, climate data, soil type and land use maps, which are further described in section *2.6 Model Setup.*

2.4 A priori uncertainty of model parameters

Within MGB model, there are parameters related to vegetation cover (albedo, 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, XL, CAP, Wc, CI, CS, CB), which are further detailed in Supplementary Material (Table S2). Out of the 19 model parameters, six are fixed and 13 are calibrated.

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

2.5 Sensitivity analysis

In order to understand how different parameter sets (river hydraulic, soil, vegetation) affect model output variables (river discharge, flood extent, river water level, soil moisture, evapotranspiration and terrestrial water storage), 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 to ensure that convergence is not dependent on the initial parameter sets, thus resulting in 300 runs for each setup. In this step, no RS-based calibration is performed yet.

Parameters were varied considering a uniform distribution, and results were analyzed in terms of mean 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

Supplementary Material Table S2). 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 dispersion of model outputs was also compared to uncertainty in the observations, as derived from literature.

To understand which variables are inter-related in the model, we further analyzed the results of setup "(4) varying all parameters together". This was done by firstly computing the Kling-Gupta Efficiency metric (KGE; Gupta et al., (2009)) between the perturbed runs and a reference one (i.e., run with the initial parameter set) for each variable, and then calculating the Pearson correlation (r) between KGE values for each pair of variables (e.g., discharge and water level, discharge and flood extent, and so forth). This experiment is relevant to evaluate whether two variables get improved or get worsened together, or whether a variable improvement impacts on the deterioration of another. In other words, this approach allows to evaluate the correlation between the variables.

2.6 Model setup

The Bare Earth Digital Elevation Model (DEM; O'Loughlin et al., 2016) (Figure 2c) was used for stream network computation and basin discretization with the IPH-HydroTools GIS package (Siqueira et al., 2016). The original DEM resolution is 90 m, and it was resampled to 500 m to facilitate GIS processing. An upstream area threshold of 100 km^2 was adopted to delineate the drainage network, and unit-catchments were discretized by dividing the stream network into fixed reach length 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) (Figure 2b): (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, where “deep soils” refer to soils with high water storage capacity, and “shallow soils” are those with low water storage capacity. 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; available at: <https://gpm.nasa.gov/data-access/downloads/trmm>), and were interpolated with the nearest neighbor 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 were obtained from the Climatic Research Unit database (New et al., 2000; available at: <http://www.cru.uea.ac.uk/data>), at a spatial resolution of 10’, and also interpolated with the nearest neighbor method.

2.7 Model calibration

The MOCOM-UA calibration algorithm (Yapo et al., 1998; Multi-objective global optimization for hydrologic models) was adopted 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. The algorithm converges towards the Pareto optimum, when all points in the population become non-dominated. The model population consists of randomly distributed points within the parameter search space, and it reflects the a priori uncertainty of model parameters. Here, the population size was set to 100 individuals. The altered model parameters and their respective ranges are described in Supplementary Material Table S2. All calibration experiments are repeated three times (totaling 300 initial runs) with differing initial parameter sets to ensure that convergence is not dependent on the initial parameter sets. Initial parameters are set as the mean of their literature-based range (Table S2).

Objective functions to be optimized depend on the calibration setup. In the single-variable calibration, for each variable, three objective functions (*OF*) that summarize the agreement between simulated and observed (RS) time-series are simultaneously optimized: Pearson correlation (r), ratio of averages ($\mu_{¿}/\mu_{obs}$), and ratio of standard deviations ($\sigma_{¿}/\sigma_{obs}$), which are associated to the individual terms of KGE metric. These 3 objective functions are depicted in Equations 1 to 3, where X denotes the assessed variables (Q, h, A, TWS, ET or W).

$$OF_1=\left(\frac{\mu_{¿}}{\mu_{obs}}\right)_X (1); OF_2=\left(\frac{\sigma_{¿}}{\sigma_{obs}}\right)_X (2); OF_3=r_X (3)$$

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), as depicted in Equations 4 to 8.

$$OF_1=KGE_h (4); OF_2=KGE_A (5); OF_3=KGE_{TWS} (6); OF_4=KGE_{ET} (7); OF_5=KGE_W (8)$$

Secondly, two objective functions were adopted and simultaneously calibrated (KGE of selected variables 1 (x) and 2 (y)), as depicted in Equations 9 and 10.

$$OF_1=KGE_x (9); OF_2=KGE_y (10)$$

Results are expressed in terms of a Skill Score (S) (Equation 11; Zajac et al., 2017), in order to evaluate the improvement (or deterioration) in the representation of a variable when the model is calibrated with a given variable, compared to the uncalibrated setup.

$$S = \frac{KGE_{calibrated} - KGE_{initial}}{1 - KGE_{initial}} \quad (11)$$

$KGE_{calibrated}$ is the mean KGE resulting from running the model with the calibrated parameters. $KGE_{initial}$ is the mean KGE resulting from running the model with the a priori parameter sets (i.e., randomly selected parameters within an a priori range of parameter values).

2.8 Calibration/Evaluation Data

In the next paragraphs we introduce the data used for model calibration and evaluation, as well as how MGB outputs were evaluated in comparison to them.

-*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 km², period of data availability: 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). Discharge was evaluated on a daily basis.

- *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 number 165. Processed data for this study were downloaded from the Hydroweb/Theia database (available at: <http://hydroweb.theia-land.fr>). Water level was computed in MGB at the unit-catchment associated to the altimetry virtual station, being an advantage of using the hydrodynamic scheme for flood routing instead of the Muskingum simplification. Simulated and RS water level data were compared every 10 days in terms of anomaly (values subtracted from long term average).

- *Satellite flood extent data* were derived from ALOS-PALSAR imagery, which presents 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 uncertainty 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. ALOS-PALSAR presents a recurrence cycle of 46 days (from 2006 to 2011), and flood extent data were available and compared to MGB for 21 dates.

- *Satellite-based terrestrial water storage (TWS) anomalies* were extracted from GRACE mission, launched in March 2002. GRACE provides monthly TWS estimates based on anomalies in gravitational potential, at a resolution of 300-400 km, with a 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 the mean value based on the three products was averaged for the whole basin. In MGB, TWS values were computed as the sum of water storage of all hydrological compartments: river, floodplains, soil, groundwater and vegetation canopy. Simulated and RS-based TWS were compared in terms of anomaly (values subtracted from long term average) at a monthly time-scale.

- *Satellite-based evapotranspiration* estimates were retrieved from the MOD16 product, derived by an algorithm presented by Mu et al. (2011) based on the Penman-Monteith equation. The dataset covers the period 2000-2010 with a spatial resolution of 1 km for global vegetated land areas. Because of that, even though MGB evapotranspiration is calculated for flooded areas (open water evaporation in 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. Accuracy of MOD16 along the Amazon basin is estimated as 0.76 mm/day (Gomis-Cebolla et al., 2019). MOD16 data is available at: < https://www.ntsg.umt.edu/project/modis/mod16.php>. In MGB, evapotranspiration is computed via Penman-Monteith equation, based on the climate input variables.

- *Satellite-based soil moisture* is derived from the SMOS mission (Kerr et al., 2001), processed by the Centre Aval de Traitement des Données SMOS (CATDS), and downloaded in processing level 4, which combines lower level products with data from other sensors and modeling/data assimilation techniques. The daily L4 root zone soil moisture product at 0-1 m soil depth (Al Bitar et al., 2013) were used (available at: <https://www.catds.fr/Products/Available-products-from-CEC-SM/L4-Land-research-products>), and data from ascending and descending orbits were averaged for the whole basin. In MGB, soil moisture as a saturation degree was computed as the water in the soil compartment divided by the maximum water capacity of the soil (*Wm* parameter). Since MGB estimates saturation degree values for a soil bucket reservoir, SMOS values were rescaled for the range 0 - 100% 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), and them compared to MGB on a daily time-scale as an average for the whole basin.

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 the water cycle representation as a whole (Section 3.2).

3.1 Sensitivity analysis

A sensitivity analysis was carried out to understand how different parameter types (river hydraulic, soil, vegetation, and all together) affect the variation of different hydrological processes in MGB (Figure 3a). This was performed by analyzing the dispersion of six output variables (discharge, water level, flood extent, TWS anomalies, vegetation ET, and soil moisture). These results are also compared with an estimate of the uncertainties of observations (values provided in section *2.8 Calibration/Evaluation Data*), and are discussed in the subsequent sections.

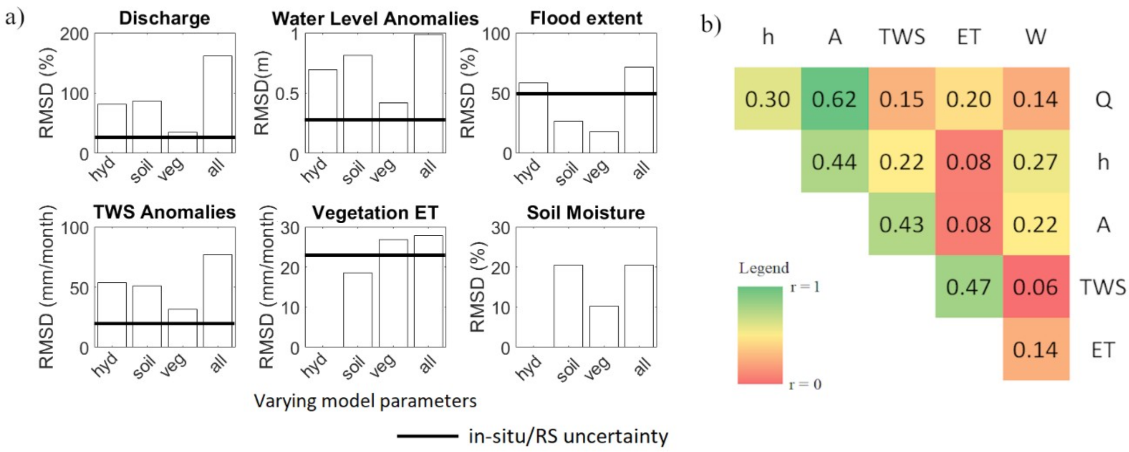

Figure 3. **a)** Sensitivity analysis of different model output variables to varying sets of parameters (hyd=hydraulics, soil, veg=vegetation, and all together). The a priori dispersion of the model parameters, for each output variable, is compared to the reported uncertainty for the in-situ / RS product estimates, 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). **b)** Correlation matrix (Pearson coefficient) between performance metrics (KGE) for the six analyzed variables, by varying all parameters together. KGE values are computed by comparing multiple runs with the reference simulation (i.e., the initial run with the initial parameter set as defined in Supplementary Material Table S2). Q = discharge, h = water level, A = flood extent, TWS = total water storage anomalies, ET = vegetation evapotranspiration, W = soil moisture.

3.1.1 How do varying model outputs relate to uncertainty in the observations?

Some variables present in-situ/RS observations that have uncertainties significantly lower than the overall dispersion of the model, e.g., 25 % for discharge observations, while model overall parameter dispersion 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. Nonetheless, uncertainties in RS products of flood extent (~50%) and vegetation ET (~23%) are in the same order of magnitude of model overall parameter dispersion, 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 model dispersions are related to different 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, being insensitive to hydraulic parameters. Soil moisture (W) is more sensitive to soil parameters, while vegetation ET is more sensitive to vegetation parameters. These results are very useful to understand the RS-based calibration experiments addressed in section 3.2. For instance, if model calibration with ET or W is achieved through optimization of hydraulic parameters, it would highlight that the model would have "gotten the right results for the wrong reasons". The same would occur if flood extent calibration targeted soil or vegetation parameters.

3.1.3 Which variables are inter-related?

By varying all parameters together, there is a high correlation (greater or equal to 0.4) between the performance of discharge and flood extent, water level and flood extent, flood extent and TWS, and ET and TWS (Figure 3b). High correlations between discharge, water level and flood extent are expected because of their strong association through river transport processes. However, correlation between discharge and water level is not too high (0.30), and this is probably due to high uncertainties in hydraulic parameters, and to the large distance separating the water level virtual station and the streamflow gauge. Furthermore, high correlations between TWS and flood extent might be related to surface water storage dynamics which are especially relevant in regions with floodplains.

In general, a high correlation between variables in Figure 3b should be reflected in positive results when calibrating with a given variable and evaluating with the other highly correlated variable (single-variable calibration). This may also indicate that observations of these variables are redundant if used simultaneously in a multi-calibration framework. However, high correlations in Figure 3b followed by deterioration after the single-variable calibration process might indicate structural errors in the model, or in the observations. We stress however that this study did not attempt to quantify structural errors. Conversely, low correlations in Figure 3b, followed by improvement in performances with the calibration with multiple variables, might indicate complementarity between variables.

3.2 Model calibration

3.2.1 How RS-based model calibration improves discharge estimates?

For the evaluation time period (2006–2008 for discharge, flood extent, TWS, ET and 2013–2014 for water level and soil moisture), calibration with all RS products led to improvements in discharge estimates (Figure 4a). For the calibration time period (2009-2012), TWS, ET and soil moisture RS products also led to improvements in discharge estimates, while water level and flood extent led to discharge overestimation in wet periods (Figure 4a). This could be due to high uncertainties in the observations (Figure 3a), 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 4b), 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.

Nonetheless, 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 four virtual stations, and flood extent for stream segments at different locations in the basin.

RS variables as TWS, ET, and soil moisture were able to improve discharge estimates by S =13.7% (KGE_{cal}=0.36), S= 52.9% (KGE_{cal}=0.64), and S = 27.0% (KGE_{cal}=0.45) (Figure 5-I, calibration period) or S = 27.4% (KGE_{eval}=0.52), S= 6.1% (KGE_{eval}=0.37), S= 12.3% (KGE_{eval}=0.43) (Figure 5-II, evaluation period), 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 5b) resulted in a Skill Score of S = 17.4% (KGE_{eval}=0.45) for discharge in the evaluation period. This is relevant for estimating discharge in poorly gauged basins. Nonetheless, for the calibration period, Skill Score had a low value (S = 1.7%, KGE_{cal}=0.25), reflecting some limitations when retrieving discharges, probably because of potential trade-offs between variables (Koppa et al., 2019). RS uncertainties can be reduced in model calibration, for instance by using bias-insensitive metrics (e.g., Demirel et al., 2018; Zink et al., 2018; Dembele et al., 2020), or explicitly including them into the objective functions (Aires, 2014; Croke, 2009; Foglia et al., 2009; Peña-Arancibia et al., 2015).

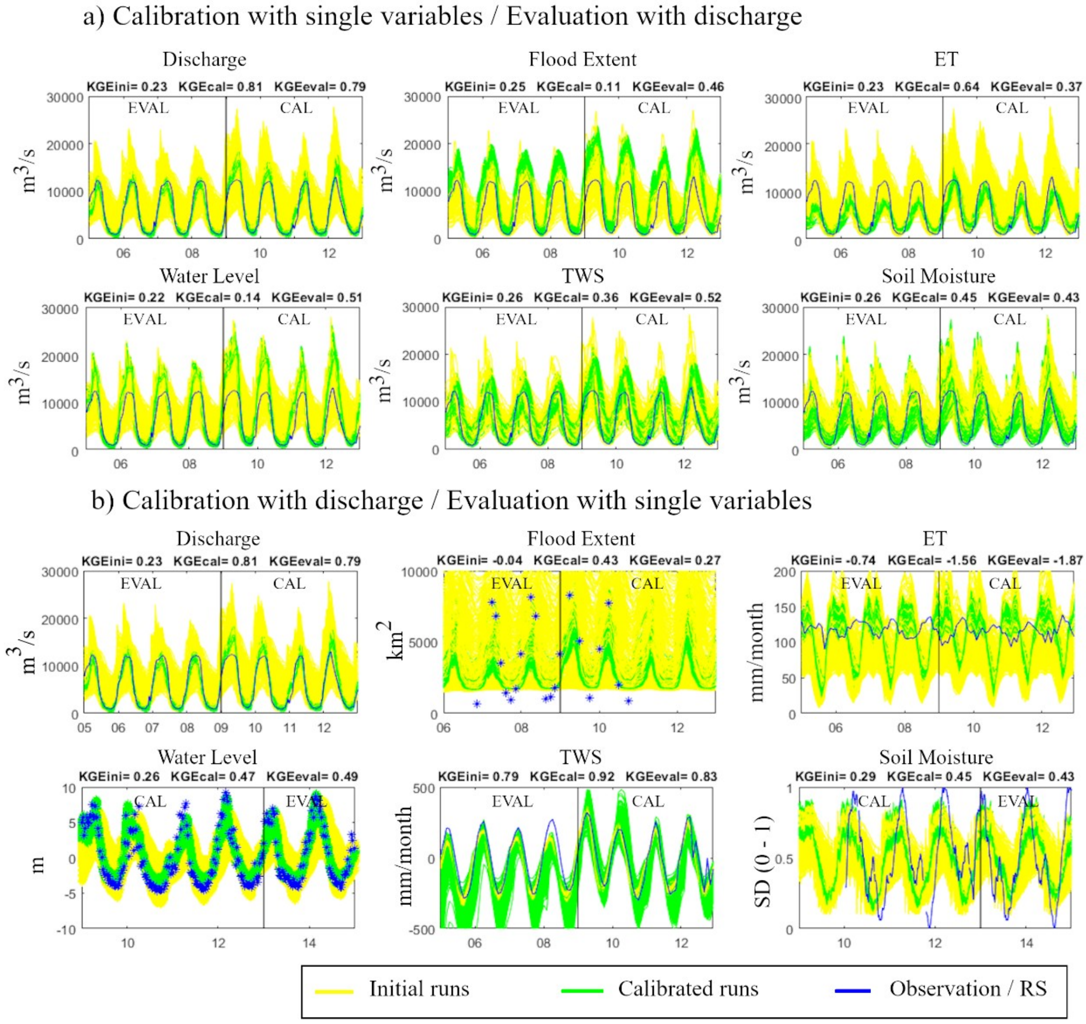

Figure 4. (a) Daily time series of discharge, when calibrating the model with six different variables. (b) Time series of the six variables when calibrating the model with discharge observations only (discharge, water level, flood extent and soil moisture are at a daily time step, while TWS and ET are at a monthly time step). *KGEini* is the mean KGE of initial runs, *KGEcal* is the mean KGE of calibrated runs, evaluated for the same period of calibration, *KGEeval* is the mean KGE of calibrated runs, evaluated for a different period than calibration. Time series for all variables by calibrating the model with all setups is presented in Supplementary Material (Figure S1).

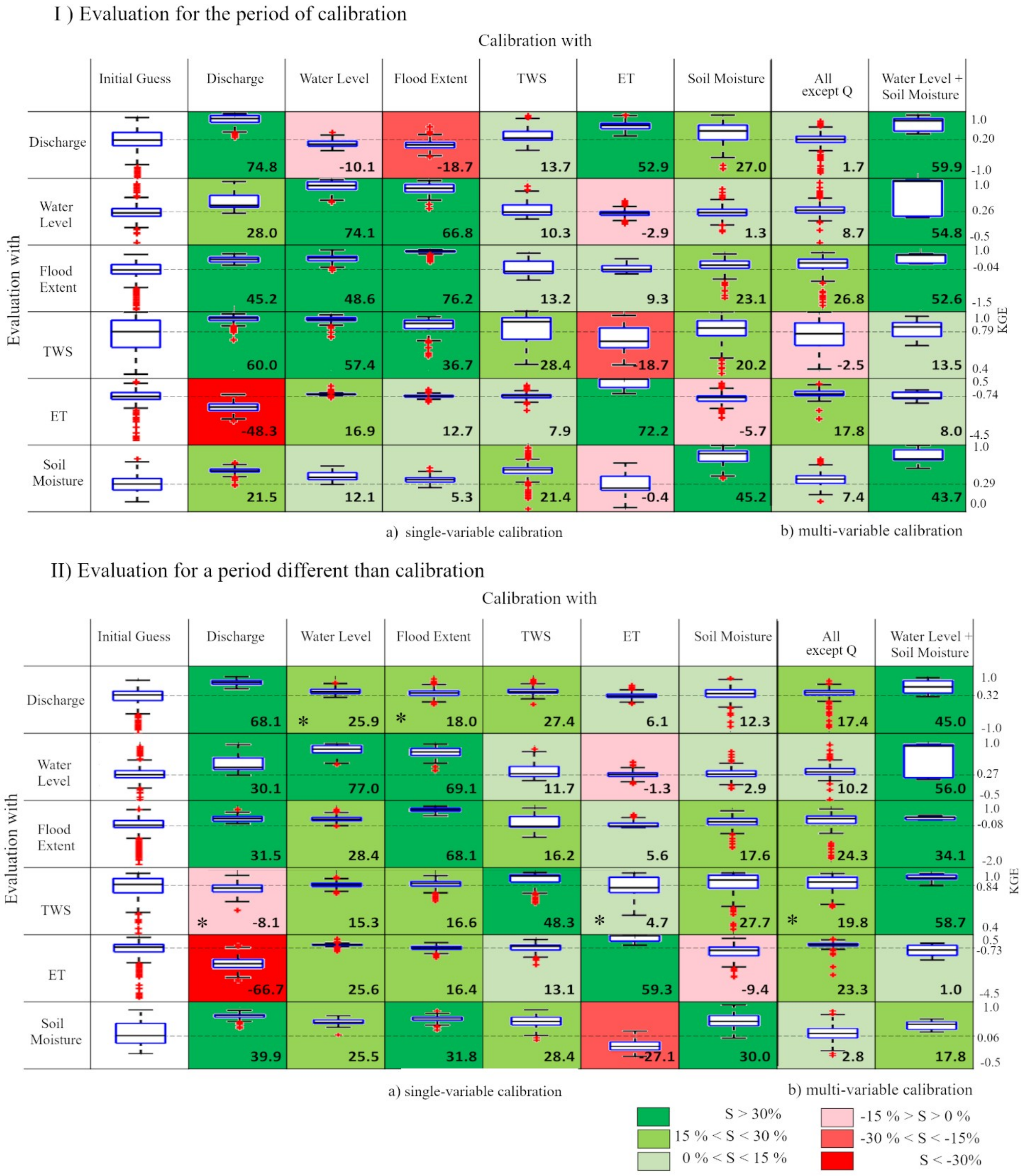

Figure 5. Boxplots of mean KGE for the evaluation of multiple variables with different calibration strategies. (I) Evaluation for the period of calibration (2009 – 2012); (II) Evaluation for a different period than calibration (2006 – 2008 for Q, A, TWS, ET; 2013 – 2014 for h and W). "Initial guess" refers to model runs with the a priori parameter sets. (a) Single-variable (discharge, water level, flood extent, TWS, vegetation ET, soil moisture) and (b) multi-variable

calibration (all except discharge, water level + soil moisture). The spread of the values in the boxplots stems from 300 model runs (100 for each of three calibration experiments). Numbers next to the boxplots represent Skill Score (%). Colors refer to classes of skill score. Please note that the KGE scales are different for each variable. Asterisks refer to cases when the evaluation period resulted in a different performance than the calibration period (i.e., positive Skill Score in calibration followed by negative Skill Score in evaluation, or vice-versa). Please note that Skill Score values are computed based on mean values, while the boxplots depict median values.

3.2.2 How does RS-based model calibration improve the water cycle representation?

When performing a single-variable calibration, the performance of the variable itself always improves, which is evidenced by the positive values in the main diagonal (Figure 5-I-a, for calibration period, and Figure 5-II-a, for evaluation period). Calibration with water level was also able to improve estimates of flood extent, TWS, ET and soil moisture (cal period), and all variables (eval period). Calibration with flood extent improved water level, TWS, ET and soil moisture. Calibration with TWS improved all variables. Calibration with ET was able to improve discharge and flood extent. Calibration with soil moisture improved all variables but ET. Results for calibration and evaluation periods agree (i.e., improvement (positive Skill Score) or deterioration (negative Skill Score) for both cal and eval) in 43 out of the 48 cases (89.6%). In the five remaining cases (10.4%), results between calibration and evaluation periods differ: three of them are in the evaluation with TWS, and two of them are in the discharge evaluation (calibration with water level and flood extent).

In the best modeling scenario, 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, in the observations, or the integration techniques in model calibration (Dembele et al., 2020). Previous studies have also found significant advantages in using RS-based model calibration 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 Figure 4b and Supplementary Material (Figure S1), calibration with discharge improved estimates of almost all variables. However, calibration with discharge deteriorated the performance for vegetation ET time series. Vegetation ET estimated by MOD16 varies at maximum 30mm/month. MGB calibration with discharge led to ET variations of 100 mm/month, 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 4b). 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) and Pan et al (2020) 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 was not water-limited 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 it were only compared to discharge time series.

Even though the calibration with discharge observations was not able to accurately estimate ET, calibration with the remaining variables (except for soil moisture) was able to improve ET estimates. For instance, in Figure 3b, ET and water level presented low correlation (r = 0.08), but calibration with water level improved ET estimates by S = 16.9% (cal period) and S = 25.6% (eval period). However, in Figure 3b, ET and TWS presented high correlation (r=0.47), but calibration with TWS improved ET estimates by only S = 7.9% (cal period) and S = 13.1% (eval period).

In general, calibration with TWS did not present much influence on any of the variables. In spite of some improvements, skill scores were usually low. Consistently, TWS estimates got relatively easily improved by calibration with any variable (except for ET, for cal period; or discharge, for eval period). 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 value 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 the initial guess (uncalibrated) setup TWS performances were already 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.

Flood extent and water level performances were improved by calibration with 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.

Calibration with soil moisture improves performances of all variables (water level to a lesser extent), except for ET. Consistently, calibration with all variables (except ET) are able to improve soil moisture to some extent.

3.2.3 What is the added value of complementary RS observations?

By calibrating with all variables together except Q (Figure 5b), we found improvements for almost all variables, with the most significant improvements for flood extent (S = 25% for cal and eval periods) and ET (S = 20% for cal and eval periods). For discharge, performance for the evaluation period was improved (S = 17.4%), which is important for estimating discharge in poorly gauged basins. However, for the calibration period, Skill Score for discharge performance was S = 1.7%, which might reflect some limitations in retrieving discharge 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 5-I, calibration with water level only improves all variables but discharge (and soil moisture to a lesser extent), while calibration with soil moisture only improves all variables, but ET (and water level to a lesser extent).

The calibration arrangement of water level and soil moisture led to improvements not only to soil moisture and water level themselves, but also to all other variables (ET to a lesser extent). For instance, flood extent was improved by S = 52.6% and S = 34.1% (cal and eval period, respectively). Discharge was improved by S = 59.9%, with a resulting mean KGE = 0.70 for the calibration period (S = 45.0% and mean KGE = 0.35 for evaluation period). 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 RS-based calibration method for poorly-gauged regions.

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 does not represent deep root water intake during dry season, as discussed previously.

By comparing the two frameworks for multi-variable calibration (all except Q versus h+W calibration), we found that calibration with all variables except Q 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 6 for a few selected parameters, Supplementary Material Figure S2 for all calibrated parameters), it can be observed that the range of parameters varies largely depending on the calibration variable. For instance, Wm is a soil conceptual parameter related to maximum water storage in the soil. In the calibration based on single variables (except ET) it converged to low values (300 mm), while in the calibration with ET it reached high values (2000 mm). This probably occurred in order to compensate, by overparameterization, 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).

The surface resistance parameter also resulted in a wide range of values depending on the calibration target variable. When calibrated with water level, flood extent, or ‘all except Q’ experiments, it reached median values higher than 150 s/m, but calibration with h+W led to median values lower than 50 s/m. Surface resistance is a vegetation parameter directly related to ET dynamics, so it is important to note that calibration with ET was able to reduce the dispersion of this parameter, reaching a median value of about 80 s/m (similar to calibration with Q and W).

Another interesting result relates to channel Manning's coefficient, which presented different values for each calibration experiment. This agrees with previous findings about Manning parameter being often used as an effective parameter that compensates for neglected hydrodynamic processes as localized channel head losses, poor cross section representation, or non-represented 2D processes (Neal et al 2015).

Many previous studies have highlighted the use of multi-variable calibration to narrow parameters' search space (Nijzink et al., 2018; W. Sun et al., 2018), but this was not observed in our results. Based on the limited multi-variable calibration experiments performed here ('all except Q' and h+W), no narrowing in parameters' search space was found. For most parameters (except for Wm), calibration with 'all except Q' and h+W resulted in a wide range of values. This can be due to differing convergence sets of parameters between each of the triplicate runs. A more robust experiment comparing more multi-variable calibration strategies (e.g., Q + different R-based variables) might provide better understanding on this topic.

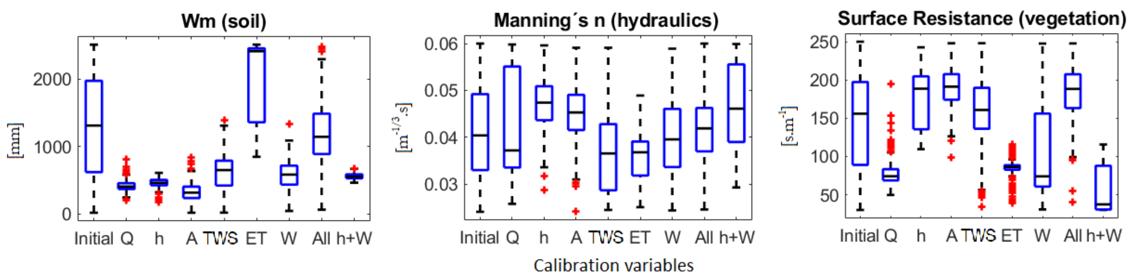

Figure 6. Boxplots of dispersion of three model parameters before (Initial) and after the single-variable calibration (Q – discharge; h – water level; A – flood extent; TWS – total water storage anomalies; ET - vegetation ET; W – soil moisture), and multi-variable calibration (All – variables except discharge; h+W – water level and soil moisture). The spread of the values in the boxplots stems from 300 model runs (100 for each calibration experiment). Description of

parameters is presented in Supplementary Material Table S2. A complete figure with boxplots for all parameters is presented in Supplementary Material Figure S2.

3.4 Spatial Evaluation

For model calibration, we used one streamflow gauge for discharge, one virtual station for water level, and averaged RS data for the whole basin for TWS, ET and soil moisture. However, many recent studies investigated the potential for using RS spatially distributed information in model calibration, for instance with bias-insensitive metrics (Demirel et al., 2018; Zink et al., 2018; Dembele et al., 2020). Here we further analyze how the lumped calibration affected the simulated spatial patterns (Figure 7; Figure S3 in Supplementary Material).

For discharge, water level, flood extent and TWS, spatial patterns are well reproduced even when running the model with the initial parameter set, because the spatial patterns of these variables are determined by intrinsic characteristics of the basin. Nonetheless, for ET, the spatial patterns are completely different between the initial parameter set and the calibrated setup. In this case, the calibration with spatially aggregated ET was able to recover the spatial representation of MOD16. A similar result was found for soil moisture spatial representation by Demirel et al. (2019), that calibrated a model with spatially aggregated soil moisture and TWS data.

In summary, these results highlight the overall model capability to retrieve the ET spatial pattern even by using a lumped calibration approach. However, for other variables, the spatial pattern was not considerably affected by the differing model calibration strategies.

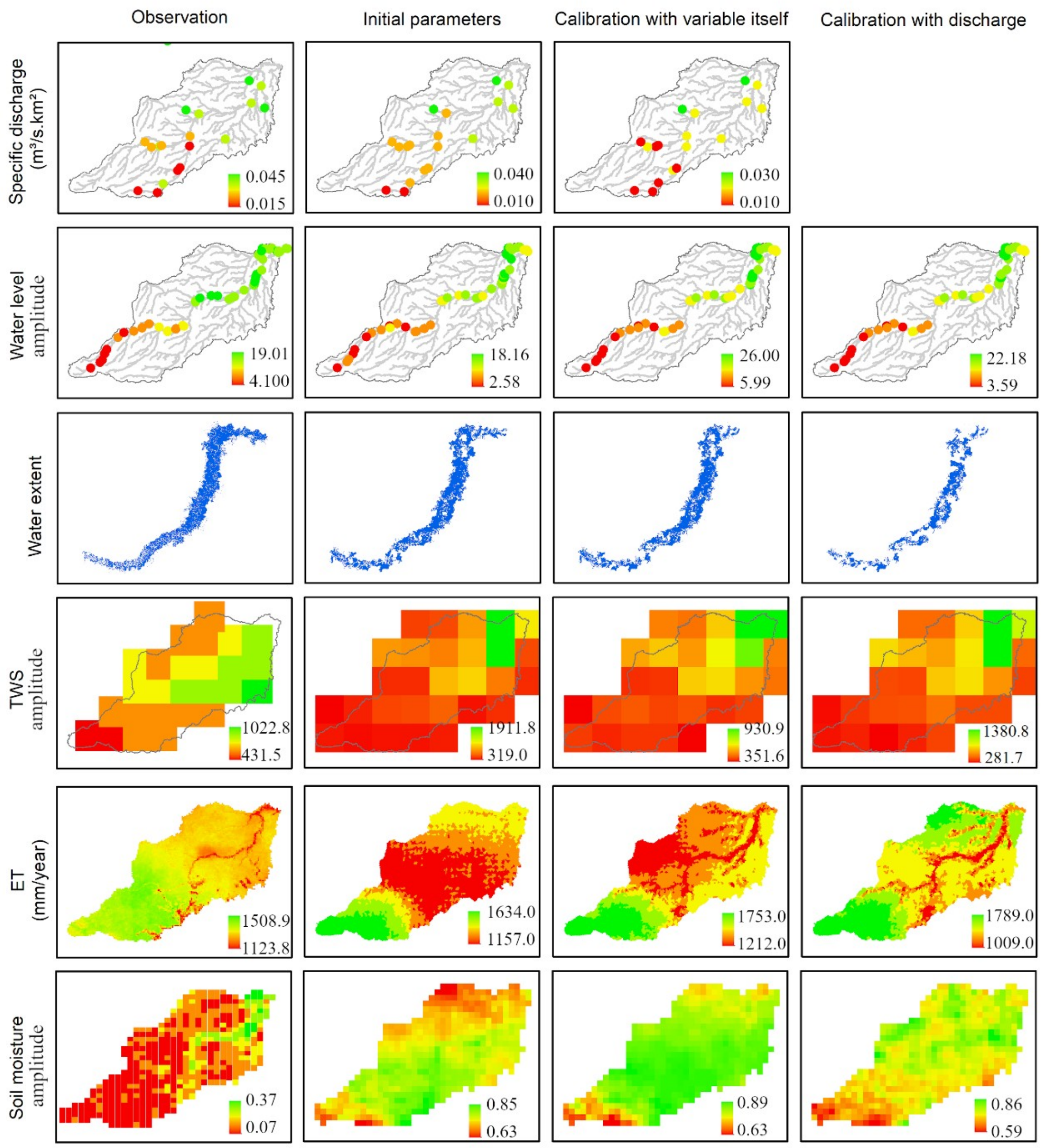

Figure 7. Spatial distribution of variables. Columns: RS observation, model run with the initial parameter set, model run with the best parameter set (calibrated for each variable), model run with the best parameter set (when calibrated with discharge). Complete figure is presented in Figure S3 (Supplementary Material).

4 Conclusion

We calibrated and evaluated a hydrologic-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, in the uncalibrated setup (a priori parameter sets), average performances for discharge were around KGE = 0.30. By calibrating the model with ET from MOD16 (and evaluating for the same time period), discharge average performance was improved to KGE = 0.64, representing a Skill Score of S = 52.9%. Also in the calibration period, a joint scheme of calibration with water level + soil moisture led to discharge improvements of S = 59.9%. When evaluating for a different time period, discharge performance was improved by calibration with water level, TWS and a joint scheme of all RS-variables (S = 25.9%, S = 27.9% and S = 17.4%, respectively). We conclude that RS observations are useful to predict discharge estimates. However, the utility of each RS variable might depend on the study area characteristics and the time period considered.

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.

Moreover, calibration with multiple RS variables was able to highlight deficiencies that might be related to model structure, parameterization, observations, and data integration techniques in model calibration. 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, we found a wide range of different parameters by varying the calibration target variable.

Besides individual calibration with each RS variable, we conducted two multi-variable calibration experiments: calibration with all variables except discharge, and calibration with water level and soil moisture. Calibration with all variables was useful to some extent, but appropriately selecting complementary variables for model calibration may result in a better overall performance. Even though we used a lumped calibration approach, results highlighted the overall model capability to retrieve ET spatial pattern, but not for TWS and soil moisture.

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 the 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. Here, we assessed an Amazonian Equatorial basin, with particular climate and land cover characteristics and an overall spatial homogeneity of rainfall-runoff processes. Other basins with different hydroclimatic regimes could be also assessed, e.g., in arid basins subject to long dry periods, more erratic precipitation patterns, and different runoff generation mechanisms than the Amazon, which require different model structures.

Finally, here we used one state-of-the-art RS product for each variable, but future developments should explore other missions like SWOT for surface water observation (Biancamaria et al., 2016), as well as considering different products for representing each variable (e.g., ET could be estimated by GLEAM, MODIS, SSEBop, SEBS, ALEXI, METRIC, etc., besides MOD16).

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://doi.org/10.5281/zenodo.3956609> (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

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

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

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

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

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three Unsolved Problems in Hydrology (UPH)—A community perspective. *Hydrological Sciences Journal.* https://doi.org/10.1080/02626667.2019.1620507

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*. https://doi.org/10.1029/2002JD003292

Dembele, M., Ceperley, N., Zwart, S.J., Salvadore, E., Mariethoz, G., Schaefli, B. (2020). Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Advances in Water Resources.* https://doi.org/10.1016/j.advwatres.2020.103667

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences.* https://doi.org/10.5194/hess-22-1299-2018

Demirel, M. C., Özen, A., Orta, S., Toker, E., Demir, H. K., Ekmekcioglu, Ö., Taysi, H., Eruçar, S., Sag, A. B., Sari, Ö., Tuncer, E., Hanci, H., Özcan, T. I., Erdem, H., Kosucu, M. M., Basakin, E. E., Ahmed, K., Anwar, A., Avcuoglu, M. B., Vanli, Ö., Stisen, S., & Booij, M. J. (2019). Additional value of using satellite-based soil moisture and two sources of groundwater data for hydrological model calibration. *Water.* https://doi.org/10.3390/w11102083

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

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*.

Fleischmann, A.S., Paiva, R.C.D., Collischonn, W., Siqueira, V.A., Paris, A., Moreira, D.M., Papa, F., Bitar, A.A., Parrens, M., Aires, F. & Garambois, P.A. (2020). Trade−offs between 1D and 2D regional river hydrodynamic models. *Water Resources Research.* https://doi.org/10.1029/2019WR026812

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/JHM560.1Jiang, 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

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*. https://doi.org/10.1127/0941-2948/2006/0130

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*. https://doi.org/10.1029/2003JD003823

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.C., Odoni, N. A., Trigg, M.A., Freer, J. E., Garcia-Pintado, J., & Mason, D. C. (2015). Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2015.07.026

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, R. 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

Paiva, R. C. D., 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

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J.E.M.S., Ottlé, C., Poulter, B., Zaehle, S., Running, S.W. (2020). Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.* https://doi.org/10.5194/hess-24-1485-2020

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

Schattan, P., Schwaizer, G., Schöber, J., & Achleitner, S. (2020). The complementary value of cosmic-ray neutron sensing and snow covered area products for snow hydrological modelling. *Remote Sensing of Environment.* https://doi.org/10.1016/j.rse.2019.111603

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

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*. https://doi.org/10.1029/2010WR009726

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