Assessing two precipitation data sources at basins of special interest to hydropower production in Brazil

Avaliação de duas fontes de dados de precipitação com especial interesse para produção hidrelétrica no Brasil

Alberto Assis dos Reis[1,2] , **Wilson dos Santos Fernandes[1]** & **Maria-Helena Ramos[3]**

[1]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]Companhia Energética de Minas Gerais S.A., Belo Horizonte, MG, Brasil
[3]Anthropised Continental Hydrosystems – Resources, Risks, Restoration - Research Unit Equipe, National Research Institute of Science and Technology for Environment and Agriculture, Antony, France

E-mails: aarhidroreis@gmail.com (AAR), wilson@ehr.ufmg.br (WSF), maria-helena.ramos@irstea.fr (MHR)

ABSTRACT

Accurate estimates of precipitation amounts are necessary to evaluate river flows, assess water-related risks (floods and drought) and quantify water availability for a broad range of water uses, such as water supply, agriculture, navigation and energy production. Especially in the context of operations in the Brazilian electricity sector, where the electrical system is essentially hydrothermal and more than 65% of its production comes from hydroelectric generation, real-time observed precipitation plays a key role as a primary input for hydrological models and river flow forecasting. It is thus crucial to build knowledge on and quantify river basin precipitation and its uncertainties. In this paper, we evaluate two sources of real-time (or near real-time) precipitation data, the TRMM-MERGE dataset from the CPETC and the CPC dataset, distributed by NOAA. Our assessment is based on 41 river basins in South America and covers the period 1997-2017. We investigated differences for different time resolutions (daily, monthly and annual precipitation) and their impact on the simulation of streamflows. Substantial differences were found between the two data sources, which seem to be amplified in the second decade. A spatial trend was found towards higher TRMM-MERGE precipitation values than CPC values when moving from north and west in the study area. We also found evidence that differences in precipitation propagate to simulated flows, with large percent differences in precipitation resulting in even larger percent differences in streamflow.

Keywords: Precipitation; Hydrology; Data uncertainty; TRMM-MERGE; CPC; Hydropower.

RESUMO

Estimativas acuradas de precipitação são necessárias para avaliar as vazões dos rios, analisar riscos ligados à água (enchentes e secas) e quantificar a disponibilidade hídrica para uma ampla gama de usos da água, como abastecimento, agricultura, navegação e produção de energia. Especialmente no contexto das operações no setor elétrico brasileiro, onde o sistema elétrico é essencialmente hidrotérmico e mais de 65% de sua produção vem da geração hidrelétrica, a precipitação observada em tempo real desempenha um papel fundamental como principal variável de entrada dos modelos hidrológicos e de previsão de vazões. Assim, é crucial construir conhecimento sobre a precipitação das bacias hidrográficas e quantificar suas incertezas. Neste artigo, avaliamos duas fontes de dados de precipitação em tempo real (ou quase tempo real), o conjunto de dados TRMM-MERGE do CPETC e o conjunto de dados CPC, distribuído pela NOAA. Nossa avaliação é baseada em 41 bacias hidrográficas na América do Sul e abrange o período 1997-2017. Investigamos diferenças para diferentes resoluções de tempo (precipitação diária, mensal e anual) e seu impacto na simulação de vazões. Diferenças substanciais foram encontradas entre as duas fontes de dados, que parecem ser amplificadas na segunda década. Uma tendência espacial foi encontrada em direção a valores mais altos de precipitação TRMM-MERGE, em relação aos valores do CPC, à medida que se desloca para o norte e oeste na área de estudo. Também encontramos evidências de que as diferenças na precipitação se propagam nas vazões simuladas, com resultados que indicam que grandes diferenças percentuais na precipitação resultam em diferenças percentuais ainda maiores nas vazões.

Palavras-chave: Precipitação; Hidrologia; Incerteza dos dados; TRMM-MERGE; CPC; Hidrelétricas.

INTRODUCTION

Knowledge of surface precipitation is particularly important in hydrological research and operations. Accurate estimates of precipitation amounts are necessary to estimate river basin runoff, assess water-related risks (floods and droughts) and evaluate water availability for a broad range of water uses (e.g., water supply, agriculture, hydropower, and environmental protection).

At the global scale, gridded precipitation products have emerged since the late 1990s (Huffman et al., 1997; Adler et al., 2003). These products usually provide monthly estimates of surface precipitation from merged analyses that blend precipitation estimates from satellite data and in-situ rain gauge observations. While they can be useful for global climate change impact studies, finer space and temporal resolutions are often needed for hydrological applications that involve daily decision-making at continental or national scales, such as flood forecasting or hydropower operations (Alfieri et al., 2013; Fan et al., 2016; Emerton et al., 2016; Siqueira et al., 2018).

In Brazil, hydropower generation is responsible for 65.2% of the electric production (Empresa de Pesquisa Energética, 2015). The inflows to hydroelectric plants have a considerable influence on planning the operation of the electrical system, as well as on setting energy prices in the short-term market. Computer models that optimize the system's operation, solving the hydrothermal dispatch problem, run once a week, every Thursday, providing forecasts of inflows on a daily and weekly basis for the first five weeks and on a monthly basis for the next months (Operador Nacional do Sistema Elétrico, 2016).

Currently, the National Operator of the Electric System (ONS) uses historical natural flow records in statistical models for monthly forecasts and daily precipitation data from rain gages to run hydrological models for daily flow forecasts at dozens of hydropower plants distributed around the whole Brazilian territory. Daily precipitation data is provided by the national electricity generators. When real-time precipitation data show gaps or inconsistencies, corrections have to be applied to have a better fit between simulated and observed flows during the warm-up phase of the hydrological models, i.e., before a forecast is issued (Operador Nacional do Sistema Elétrico, 2016). This is an important step to achieve accurate inflow forecasts to the hydropower plants.

Uncertainty in daily precipitation real-time data may come from various sources, such as the low density of the gauging network at some regions, human errors when reading the data, measurement problems at the gauging stations, data communication failures, among others. Observed real-time precipitation will therefore always be an approximation of the actual precipitation falling inside the river basins.

However, accurate real-time precipitation estimates are not the only challenge in hydropower operation. Long-term records of unbiased daily gridded-based observed precipitation data are also crucial when running sub-seasonal to seasonal forecasting systems. This is the case when using the Ensemble Streamflow Prediction (ESP) method, a widely applied technique to generate ensembles of possible future scenarios of streamflow over several weeks and months ahead. The method is based on using a continuous hydrological model to estimate initial hydrological conditions (using real-time meteorological data as input) and future meteorological forecasts (based on historical sequences of meteorological data) to obtain streamflow predictions several months ahead (see recent applications in, for instance, Crochemore et al., 2016; Bennett et al., 2017; Arnal et al., 2018; Harrigan et al., 2018). Reliable and consistent long-term historic meteorological data are therefore also crucial when running seasonal forecasting systems. For the operation of the Brazilian hydropower system at seasonal lead times, a necessary preliminary step to setting up an ESP system is to ensure that a homogeneous long-term precipitation time series over the whole country is available.

Beck et al. (2017) listed a group of 22 gridded rain datasets, but only a few of them have, simultaneously, a daily temporal resolution and a spatial resolution smaller than or equal to 0.5° covering the South America area. Among them, only two datasets were available in real time (or near-real-time), which is a necessary characteristic to use the data in forecasting systems. These datasets are the CPC Unified Gauge-Based Analysis of Global Daily Precipitation of the US National Center for Atmospheric Research (NCAR) and the NASA's IMERG, Integrated Multi-satellitE Retrievals for GPM (Global Precipitation Measurements) data, the successor of the Tropical Rainfall Measuring Mission (TRMM) data products.

Sun et al. (2017) also made a wide review of 30 currently available global datasets, based on gauging stations, satellite estimates and reanalysis. They compared 22 datasets from daily to annual time scales and found hight discrepancies between them. The magnitude of the differences in annual precipitation estimates between two different data sources was found to be as high as 300 mm/yr. The discrepancy in precipitation amounts varied however from region to region and according to the time scale. According to the authors, important differences can limit the capacity of the products to be used for climate monitoring, attribution and model validation.

Another example is provided by Negrón Juárez et al. (2009), who analyzed six databases, some based on radar and others based only on gauging stations. They observed that the largest difference among the datasets over the Brazilian Amazon region was of 8% during the apex of the rainy season (December t0 March). Over the Brazilian northeast region, the maximum difference in the wet season rainfall total (February to April) was 30 mm, or 18%.

Many other authors have studied the uncertainties and the differences among different sources of observed precipitation data in different areas of the world, with some studies focusing on South America (e.g., Demaria et al., 2011; Scheel et al., 2011; Falck et al., 2015; Mantas et al., 2015). Overall, their main conclusion is that differences between different sources of observed precipitation data are common, and often vary in space and time.

The objective of this study is to evaluate two real-time (or near real-time) sources of gridded daily observed precipitation data available over the Brazilian and adjacent territory, namely the TRMM-MERGE (Rozante et al., 2010) and the CPC (Chen et al., 2008) datasets. The differences between these two datasets are investigated in space and time. We evaluated the datasets over 41 river basins of special interest to hydropower production. We also considered the evolution in time of the deviations and the deviations at different temporal resolutions (annual, monthly and daily time steps). The impact of using these different datasets in hydrological modelling is also presented for two case studies.

STUDY AREA AND DATA

Study area

The study area covers 41 river basins that represent 31 main hydroelectric power plants in Brazil. These river basins vary in size, with drainage areas ranging from 9 300 km² to 38 2000 km². The study area extends from the north (Madeira River, Xingu River, Tapajos River, Tocantins River, and others) to the south of Brazil (Iguaçu River basin), and includes also river basins located at the central part of the country (Paraná River, Grande River, São Francisco River). Figure 1 shows the study area, with the main hydropower plants indicated in the map and the basin areas associated with the hydropower plants delimited in red.

Observed precipitation datasets

Two datasets are evaluated in this study: the TRMM-MERGE and the CPC datasets.

The TRMM partnership project between the US National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) started in November 1997, with the main goal of studying and monitoring precipitation in tropical regions (Kummerow et al., 2000). The TRMM satellite uses several instruments to detect rainfall, including radar, microwave imaging, and lightning sensors (Maggioni et al., 2016; Huffman et al., 2017). Although the TRMM products are considered valuable for numerical validation and simulations (Rozante et al., 2010), systematic errors have been detected, especially on the coast of the northeast of Brazil and in the south region of Brazil, close to the triple frontier between Brazil, Argentina, and Paraguay.

The TRMM-MERGE data was developed by CPTEC/INPE to reduce the interpolation problems observed in regions of low-density of rain gages, causing under- and over-estimations in the TRMM products. It combines gauging station datasets from the Global Telecommunications System (GTS), telemetric stations from various agencies and companies in South America and the real-time TRMM rainfall product (3B42RT), providing an improved quality gridded dataset with a spatial resolution of 0.25° for evaluation of models and operational uses (Rozante et al., 2010). Basically, the merging technique consists in identifying the TRMM grid boxes where the observations are present, discarding the two adjacent grid boxes to the observation point, and, finally, interpolating the TRMM precipitation and the ground observations using the Barnes objective method (Barnes, 1973, apud Rozante et al., 2010).

The TRMM-MERGE daily precipitation data used in this study was obtained from the CPTEC FTP site (Centro de Previsão do Tempo e Estudos Climáticos & Instituto Nacional de Pesquisas
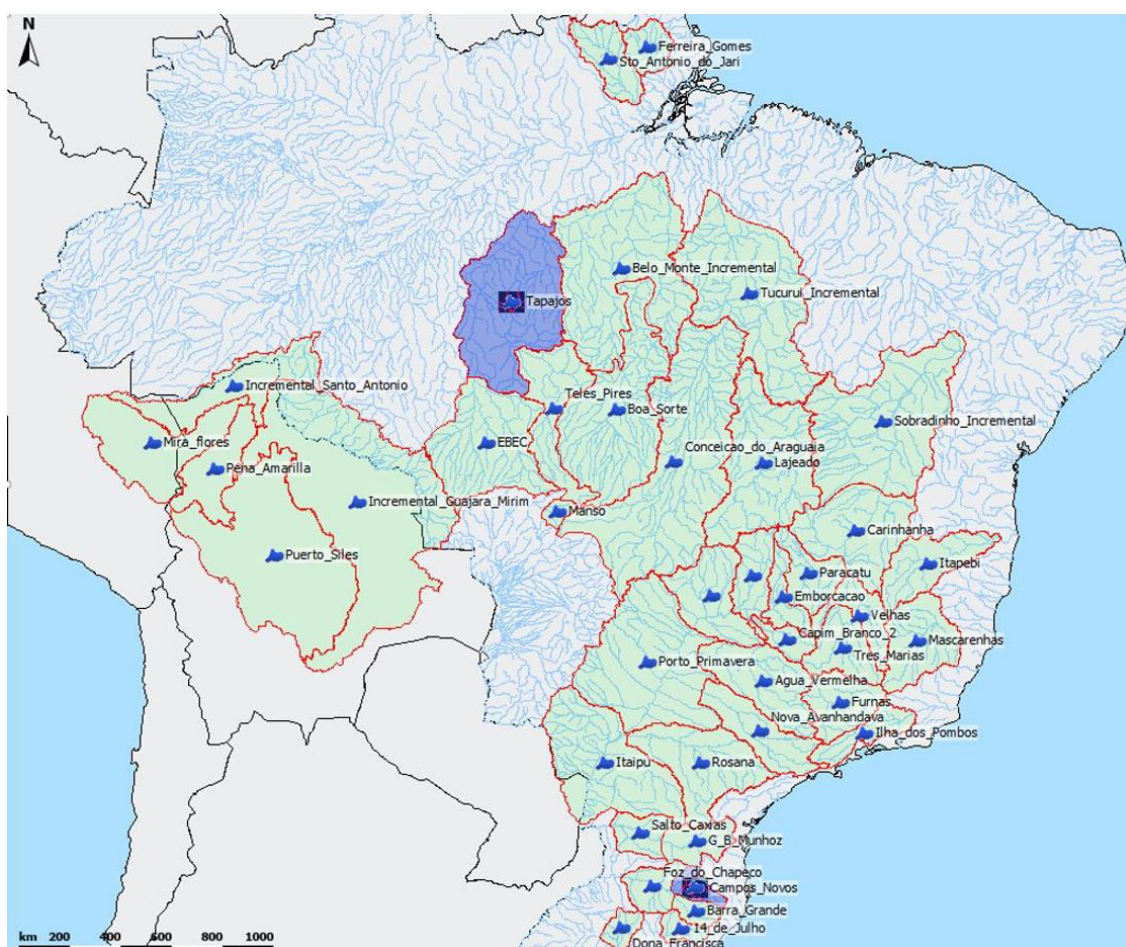

Figure 1. Geographic distribution of the 41 basins of this study, with hydroelectric power plants in Brazil.

Espaciais, 2018). The information is available in the grib2 format, 0.25° resolution, and the historic period covers 1997 to 2017.

The CPC data (Chen et al., 2008) is a product developed by the US NOAA's Climate Prediction Center. It comes from a project created to develop a group of automatic procedures to do quality control for the GTS daily precipitation products, comparing historical gauge records, concomitant observations at nearby stations, satellite estimates and numerical model forecasts. As a product of this project, NOAA/NCEP provides a daily observed precipitation gridded dataset with 0.5° spatial resolution since 1979.

The CPC data used in this study was obtained from the NCEP/NOAA FTP site (National Oceanic and Atmospheric Administration, 2018). The information is available in a grib2 format, and was retrieved for the historic period that covers 1979 to 2017.

METHODOLOGY

Our methodology consists of two main analyses. First, we evaluate basic statistics from the two different precipitation data sources for the average precipitation over each river basin. Secondly, we investigate the impact of the differences in precipitation on streamflow simulations on two selected river basins, one with a low and another with a high difference between the two precipitation data sources. For the flow analysis, we set up and calibrated the HEC-HMS model, as presented below.

Average precipitation for each River Basin

The first step to obtain the average precipitation is to delimit the contour of the basins and generate the shapefile with the basins that will be superposed on the gridded data files. In this study, we used the DELFT FEWS-CEMIG System (Pinto et al., 2013; Werner et al., 2013; Schwanenberg et al., 2015; Gibertoni et al., 2017) to obtain the average precipitation at each basin and time step. We configured the FEWS system to read the shapefiles and to use the geographic information provided to extract the average precipitation. The system uses a workflow of routines to perform the calculations. The methodology used to obtain the average precipitation is the "Average Area", which takes the mean of the data points inside the shape of each basin.

Basic statistics of precipitation differences

We analyzed the daily, monthly and annual precipitation totals of the two data sources, and their basic statistics. We also evaluated the differences in percentage between TRMM-MERGE and CPC precipitation data:

$$\Delta P(t) = \frac{P1(t) - P2(t)}{P1(t)} * 100 \qquad (1)$$

where $P1$ is the precipitation from TRMM-MERGE and $P2$ is the precipitation from CPC at time t. A positive (negative)

difference indicates higher (lower) value of precipitation for the TRMM-MERGE dataset.

In the annual analysis, we considered the hydrological year from October 1st to September 30th to calculate the annual totals. From the time series of the annual percent differences, we estimated the following basic statistics: the minimum, the first quartile, the mean, the third quartile, and the maximum (Naghettin & Pinto, 2007). For the monthly analysis, we used box-plot representations of the monthly percent differences to display the distribution of differences in monthly precipitation.

For the daily analysis, we estimated the Empirical Cumulative Distribution Function ECDF ($F_n(y)$) (Naghettin & Pinto, 2007) of the daily precipitation values for each dataset. We considered only the amounts higher than 1 mm/day over the basin areas. For observations x = (x₁, x₂, ... xₙ), F_n is the fraction of observations less or equal to y.

$$f_n(y) = \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq y} \qquad (2)$$

$$F_n(y) = \int_{-\infty}^{y} f_n(y) \, dy \qquad (3)$$

where the indicator is ($x_i \leq y$) and n = number of data points.

Finally, to visualize the geographic impact of the differences between the datasets and their evolution in time, we used maps to represent the percent differences of the annual average precipitation for two decades: 1998-2007 and 2008-2017.

Flow analysis with the HEC-HMS Model

Streamflow modeling for flow forecasting has to be performed using a continuous model for the simulations in order to evaluate the initial conditions at the onset of the forecasts. For this, it is necessary to choose a more sophisticated configuration of the HEC-HMS model (Feldman, 2000). Below, we describe the modules used and the method chosen for each module in order to build the hydrological modeling approach used in this study (Scharffenberg, 2016).

- Canopy method: Simple Canopy - the precipitation is intercepted until the canopy storage capacity of the surface is filled. All excess of precipitation falls to the surface. The potential evapotranspiration is used to empty the canopy storage;

- Surface method: Simple Surface - the precipitation that arrives on the soil is captured until the storage capacity of the surface is filled, then the runoff starts with the excess of precipitation. The water in the surface infiltrates into the soil, according to the soil's infiltration capacity;

- Loss method: Soil Moisture Accounting - SMA - this loss method uses three layers (soil storage, upper groundwater, and lower groundwater) to represent the dynamics of the water movement in the soil. For the given precipitation and evapotranspiration the model calculates surface runoff of the basin, groundwater flux, losses and the deep percolation over the whole basin. The method is capable to simulate wet and dry cycles and can be used for long periods of continuous simulation;

- Transformation method: Clark Unit Hydrograph - this method is a synthetic unit hydrograph and the principal components are the time of concentration defining the travel time in the sub-basin and the storage coefficient used to account for storage effects on the linear reservoir;

- Base-flow method: Linear Reservoir - uses a linear reservoir to model the recession of the base-flow after a precipitation event, conserving the mass. The lateral outflow of the groundwater is connected with the infiltration from the soil moisture accounting loss method.

The HEC-HMS configuration selected has 26 free parameters to be calibrated against observed flow data. In our study, observed streamflow data comes from ONS (National Operator of Electric System). The calibration was performed manually, by comparing simulations with observations and minimizing volume errors.

For the analysis of the impact of the different precipitation data sources on the simulations of streamflow, we calibrated two selected basins. They represent the extremes of precipitation differences observed during the annual precipitation analysis. This basins are: i) the UHE Campos Novos, located at the south region, at Canoas River (Uruguay river basin), where the differences between the precipitation data sources are smaller than 5% and there is not a strong seasonality in precipitation, and ii) the UHE Tapajos, located at the north region, at Tapajos river (Amazon basin), with annual differences between precipitation data sources higher than 40%, and with a strong precipitation seasonality. These basins are full blue colored in Figure 1.

In the flow analysis, we want to investigate what happens if we calibrate the hydrological model with one dataset and simulate it with another dataset. The hydrological model is first calibrated for the complete data period, from October 1997 to September 2017, with one climatic forcing and then run to simulate streamflow using the other climatic forcing, over the same period. Since we have two climatic forcing datasets, the calibration and simulation procedure is done twice.

In order to analyze the effects of the different precipitation datasets used as input to the hydrological models, we represented the ECDF of both TRMM-MERGE and CPC simulated time series. We also evaluated the simulated flows against observed flows using four numerical criteria as performance indicators: NSE, RMSE, KGE and R^2.

NSE – The Nash- Sutcliffe efficiency (Nash & Sutcliffe, 1970) measures how good the results of the model are when compared with a simulation represented by the mean observed flow. Values equal to 1 indicate a perfect fit and values smaller than zero indicate that the mean is a better predictor than the model.

RMSE – The Root Mean Square Error is a common measure of the accuracy of a model. It is calculated by taking the square root of the average of the sum of squared differences between observed and simulated values. It can be interpreted as the standard deviation of the model prediction error. A smaller value of RMSE indicates better model performance.

KGE – The Kling-Gupta efficiency (Gupta et al., 2009) is an alternative criterion to the NSE, and was proposed to assess the qualities of a model in terms of its ability to represent the water balance,

flow variability and correlation. Values range between $-\infty$ and 1, and as for the NSE, values close to 1 indicate a more accurate model.

R^2 – The coefficient of determination is an indication of how well one variable correlates with the other. It is represented as a value between 0 and 1. The closer the value is to 1, the better the fit, or relationship, between the two variables.

RESULTS

Evaluation of annual precipitation totals

Table 1 shows the basic statistics of the percent differences for annual precipitation totals for each basin. From this table, we can see that differences in quantiles range between –31% and +42% in mm of total annual rainfall. High differences are more frequently observed towards positive differences. This means that when differences are high, it is more frequently due to higher values of rainfall given by the TRMM-MERGE dataset.

From Table 1, we also note that the magnitude of the differences in annual precipitation varies according to the basin. Some basins tend to exhibit a similar behavior in terms of basic statistics of the percent differences between annual precipitation totals. We detected eight groups of similar behavior (Basin Groups I to VIII, hereafter, BGI to BGVIII). These groups are indicated in Table 1. They are formed by the following basins: for north basins - Group I: Inc. Santo Antonio, Mira Flores, Pena Amarilla, Puerto Siles, and Guajara Mirim and Group II: Ferreira Gomes, Santo Antônio do Jari, Belo Monte Inc., Boa Sorte, Tapajos, Ebec, Teles Pires, Manso, Tucurui Inc., and Conceição do Araguaia. For northeast and east Basins - Group III: Sobradinho Inc., Carinhanha, Paracatu, Velhas, Três Marias, Itapebi, and Mascarenhas. For the basins in the center-west and southeast regions - Group IV: Lajeado, Emborcação, Itumbiara, São Simão, Capim Branco 2, Ilha dos Pombos, and Furnas; Group V: Agua Vermelha, Nova Avanhandava and Rosana; and Group VI: Itaipu and Porto Primavera. For the basins in the south region - Group VII: Salto Caxias, G. B. Munhoz, Barra Grande, and Campos Novos; Group VIII: Foz do Chapecó, Quatorze de Julho, and Dona Francisca.

Figure 2 shows how the differences in annual precipitation totals between data sources evolve along the years. We selected one basin representative of each BG. We can see that the basins in the north region (BGI and BGII) display more often high positive differences, with values tending to increase with time, mainly after 2010. These BGs are generally affected by strong variations of one of the sources: either TRMM-MERGE presents very high annual precipitation totals all over the period (BGI) or CPC presents very low annual totals for a more recent period (BGII). These results illustrate how big the uncertainties in precipitation can be in this region, where the density of gauges is low. For the other regions, where the gauge density is higher, the variation of the differences is smaller and tends to be more linked to specific time periods. BGIII, BGIV and BGVI (northeast, central-west and southeast regions), for instance, display high values more often for the negative differences, with these occurring either at the initial years or at the final years of the study period. In the south region, BGVII and BGVIII present alternated years of positive and negative differences along the study period.

Table 1. Basic statistics from the time series of percent differences between annual precipitation totals from TRMM-MERGE and CPC datasets.

Basin	Min.	1st Qu.	Mean	3rd Qu.	Max.
GROUP I	27%	25%	34%	39%	38%
Mira_flores					
Puerto_Siles	22%	17%	27%	35%	36%
Pena_Amarilla	8%	2%	8%	11%	0%
Inc._Guajara_Mirim	17%	11%	12%	10%	14%
Inc._Santo_Antonio	19%	10%	10%	13%	9%
GROUP II	28%	21%	10%	5%	2%
Ferreira_Gomes					
Sto_Antonio_do_Jari	16%	4%	6%	8%	11%
Belo_Monte_Inc.	28%	15%	11%	7%	-1%
Boa_Sorte	30%	14%	13%	9%	0%
Tapajos	42%	16%	14%	10%	4%
EBEC	22%	13%	5%	-2%	2%
Teles_Pires	18%	4%	3%	-1%	-4%
Manso	26%	4%	5%	4%	-5%
Tucurui_Inc.	8%	9%	6%	6%	1%
Conceicao_do_Araguaia	16%	8%	6%	4%	1%
GROUP III	6%	2%	0%	1%	-1%
Sobradinho_Inc.					
Carinhanha	6%	9%	1%	-3%	0%
Paracatu	-4%	4%	0%	-2%	-1%
Velhas	3%	2%	-1%	-1%	-7%
Tres_Marias	1%	-6%	-4%	-4%	-9%
Itapebi	3%	2%	1%	-1%	4%
Mascarenhas	-1%	-1%	-3%	-5%	-4%
GROUP IV	2%	1%	-1%	-4%	-3%
Lajeado					
Emborcacao	-2%	-4%	-6%	-4%	-11%
Inc._Itumbiara	-3%	-1%	-5%	-2%	-9%
Inc._Sao_Simao	2%	-3%	-3%	-3%	2%
Capim_Branco_2	-7%	-2%	-4%	-9%	-1%
Ilha_dos_Pombos	5%	-9%	-7%	-8%	-6%
Furnas	2%	2%	-1%	-4%	-2%
GROUP V	-5%	-10%	-5%	-6%	5%
Agua_Vermelha					
Nova_Avanhandava	3%	-14%	-5%	1%	1%
Rosana	-8%	-4%	-3%	-2%	-3%
GROUP VI	4%	-1%	2%	4%	6%
Itaipu					
Porto_Primavera	11%	7%	4%	3%	5%
GROUP VII	28%	4%	5%	3%	11%
Salto_Caxias					
G_B_Munhoz	4%	0%	1%	-2%	5%
Barra_Grande	-31%	-6%	-2%	-1%	-2%
Campos_Novos	5%	1%	1%	0%	5%
GROUP VIII	-3%	-4%	0%	0%	2%
Foz_do_Chapeco					
Quatorze_de_Julho	1%	5%	3%	0%	6%
Dona_Francisca	9%	7%	4%	2%	2%

Evaluation of monthly and daily precipitation

Figure 3 shows the statistical distribution (boxplot) of the monthly differences. The line in red represents the monthly average precipitation of the TRMM-MERGE data. It provides a reference to compare the totals to the magnitude of the deviations. With this information, it is possible to visualize the differences along the months and the seasons in each of the eight groups of basins with similar behavior. We can see that monthly precipitation differences are higher for the basins of the north
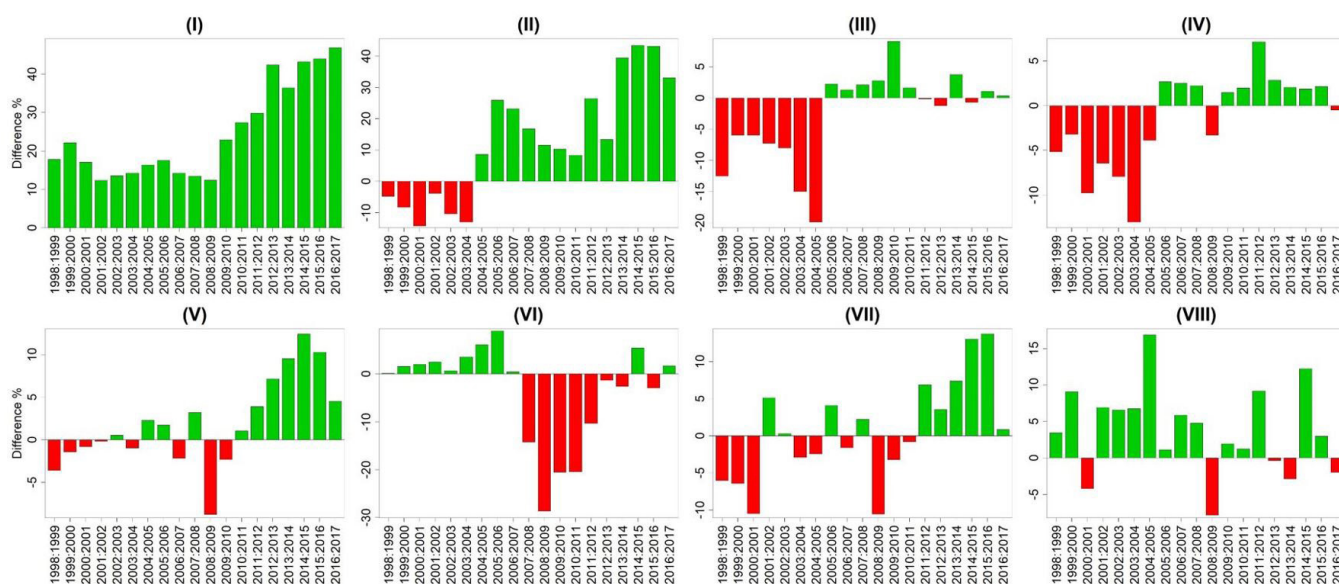
Figure 2. Percent differences in annual precipitation totals between TRMM-MERGE and CPC datasets for a representative basin in each Basin Group (BGI to BGVIII) (period 1998-2017).
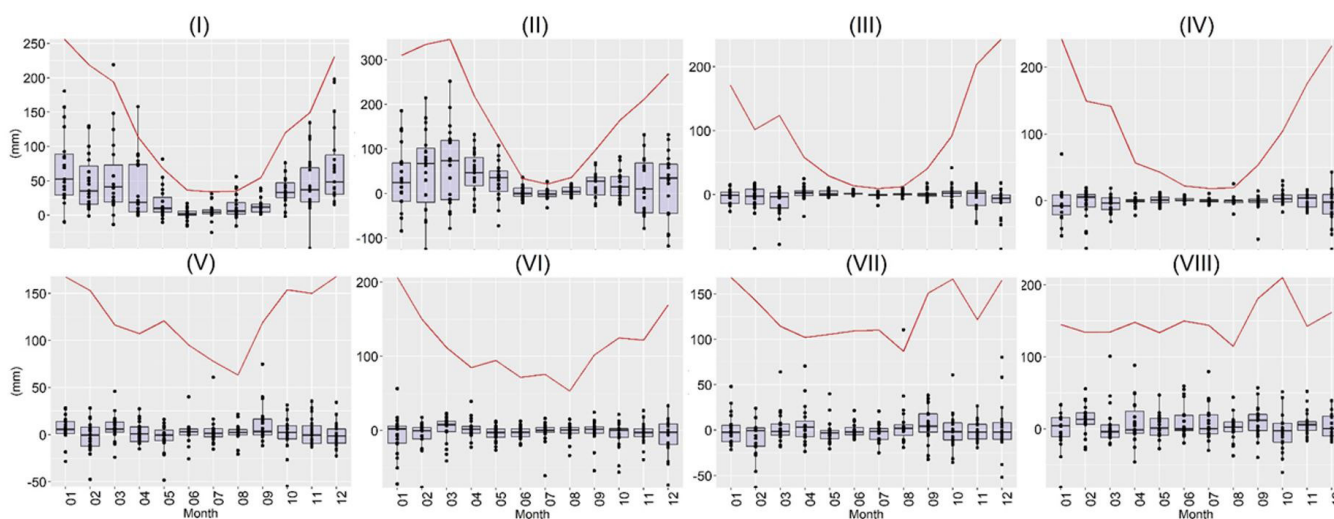

Figure 3. Boxplots of percent differences (%) in monthly precipitation between TRMM-MERGE and CPC datasets for a representative basin in each Basin Group (BGI to BGVIII). The red line represents the monthly average precipitation (mm/month) of the TRMM-MERGE data (period 1998-2017).

region (BGI and BGII). For these groups, the TRMM-MERGE data source presents more precipitation than the CPC data, in practically all months and especially during the rainy months (November to March). For the other regions, the differences vary around zero, with the higher variations occurring during the wet months. TRMM-MERGE dataset can display either wetter or dryer months than CPC, depending on the region.

Figure 4 shows the ECDF curves of the daily precipitation values greater than 1mm/day from the two data sources. The line in red represents the TRMM-MERGE data and the blue line, the CPC data. Each graph shows one basin representative of the

basin groups defined in Table 1. We can see that the cumulative distribution functions of daily precipitation are very similar. Differences can only be seen in the basins in the north region (BGI and BG II) and in the extreme south region (BG VIII), where the TRMM-MERGE dataset presents higher values of precipitation for almost all probabilities. This analysis illustrates the tendency of the basins in the more central regions to present more similarity between the data sources than the basins located in the extreme north and south regions.

Table 2 shows an example of the differences we can expect in precipitation quantiles (mm/day) for the probability
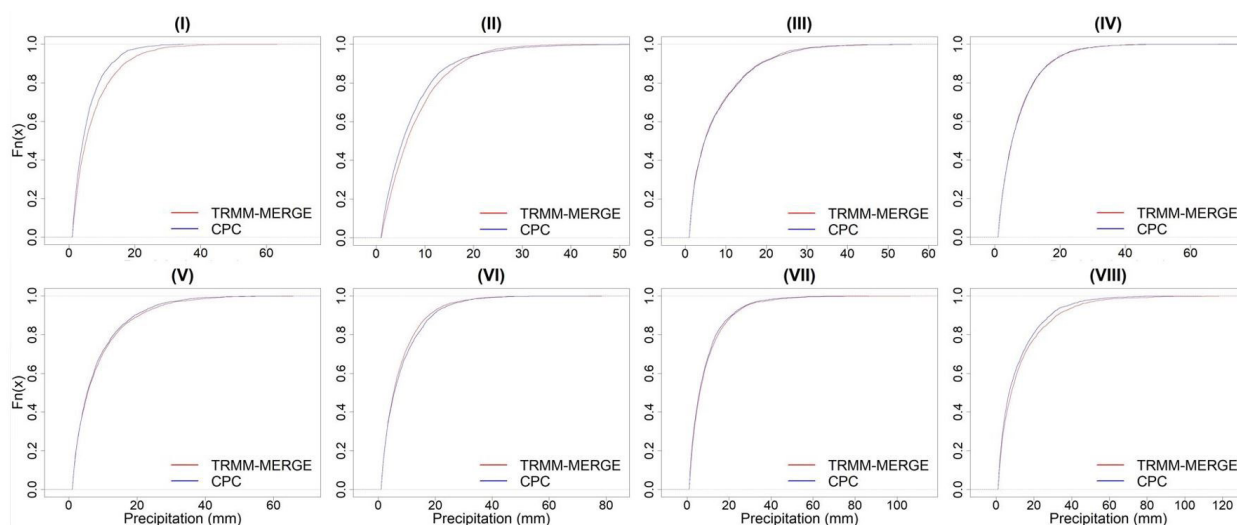
Figure 4. ECDF of daily precipitation greater than 1mm/day for the TRMM-MERGE (red) and the CPC (blue) datasets for a representative basin in each Basin Group (BGI to BGVIII) (period 1998-2017).

Table 2. Daily precipitation quantile (mm/day) for the probability of non-exceedance P=0.9 extracted from the ECDF curves of each basin and percent difference between TRMM-MERGE and CPC data sources.

BG	Basin	Quantile (mm/day) P=0.9		
		TRMM	CPC	Diff
I	Inc._Guajara_Mirim	14.5	12.6	13%
	Inc._Santo_Antonio	18.0	16.2	10%
	Mira_flores	20.8	17.7	15%
	Pena_Amarilla	20.8	21.5	-4%
	Puerto_Siles	16.9	12.9	24%
II	Belo_Monte_Inc.	16.2	15.2	6%
	Boa_Sorte	17.6	16.2	8%
	Conceicao_do_Araguaia	15.7	15.0	4%
	EBEC	17.5	16.6	5%
	Ferreira_Gomes	20.9	19.5	6%
	Manso	20.6	19.4	6%
	Sto_Antonio_do_Jari	19.5	17.3	11%
	Tapajos	17.1	15.8	8%
	Teles_Pires	17.8	18.3	-3%
	Tucurui_Inc.	15.8	14.8	6%
III	Carinhanha	17.5	17.4	0%
	Itapebi	17.2	16.2	5%
	Mascarenhas	18.3	18.6	-1%
	Paracatu	19.2	19.0	1%
	Sobradinho_Inc.	13.9	14.5	-4%
	Tres_Marias	17.9	17.9	0%
	Velhas	19.0	17.9	6%
IV	Capim_Branco_2	21.0	19.4	8%
	Emborcacao	16.8	17.1	-1%
	Furnas	16.7	17.0	-2%
	Ilha_dos_Pombos	15.9	16.3	-3%
	Inc._Itumbiara	15.8	15.8	0%
	Inc._Sao_Simao	15.2	15.0	2%
	Lajeado	14.6	15.0	-2%
V	Agua_Vermelha	15.5	17.2	-11%
	Nova_Avanhandava	19.2	19.0	1%
	Rosana	17.8	19.1	-7%
VI	Itaipu	20.7	19.2	7%
	Porto_Primavera	15.3	14.5	5%
VII	Barra_Grande	26.4	25.3	4%
	Campos_Novos	25.9	25.2	3%
	G_B_Munhoz	21.5	21.2	2%
	Salto_Caxias	28.8	24.5	15%
VIII	Quatorze_de_Julho	29.0	25.5	12%
	Dona_Francisca	32.2	28.5	11%
	Foz_do_Chapeco	28.9	27.6	5%

of non-exceedance of P=0.9. The quantiles were extracted from the ECDF curves for all basins. The percent differences confirm, in numbers, the behavior showed in the ECDF plots (Figure 4), i.e., a tendency of the extreme regions (north and south) to have higher differences (TRMM-MERGE daily precipitation greater than CPC precipitation) and the center regions to have similar values for the same probability.

Variation of annual precipitation differences in space and time

We investigated if the differences between precipitation data from TRMM-MERGE and CPC vary when considering two time periods: Figure 5 shows a map with the average percent differences of annual precipitation for the period 1998-2007 and Figure 6 shows the same but for the period 2008-2017. The shadows of green represent positive differences and the shadows of red represent negative differences. Table 3 shows the specific values for the average difference for each basin for each decade.

The differences between the TRMM-MERGE and the CPC data sources present a clear spatial and temporal behavior. For the first period (1998-2007), some basins in the north region (BGI and BGII) and in the extreme south region (BGVIII) display the most important positive average values of percent differences, while basins in the south-east region (BGIII and BGIV) display the highest negative differences. Positive average percent differences become higher and spread over the north and central regions in the last period (2008-2017). Negative average percent differences do not spread over the area in the second decade. They are however higher at the basins of the BGVI.

Impact of different precipitation data on flow simulations

The results of the experiment with the HEC-HMS hydrological model calibrated for the UHE Campos Novos and the Tapajos basins are shown in Table 4. It shows model performance when the model is calibrated with the TRMM-MERGE precipitation
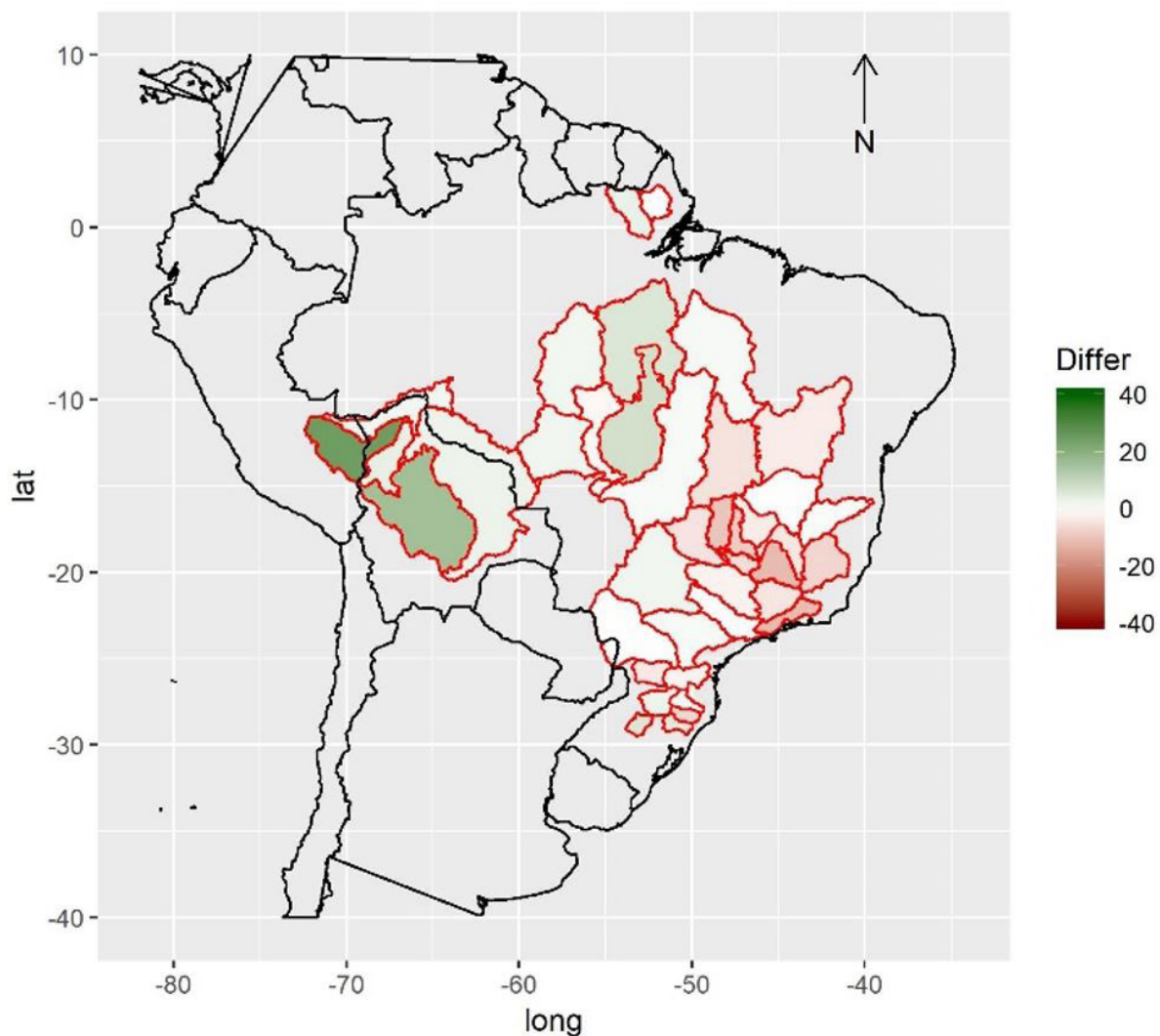

Figure 5. Map of the average percent difference (%) of annual precipitation between TRMM-MERGE and CPC for the period 1998-2007.
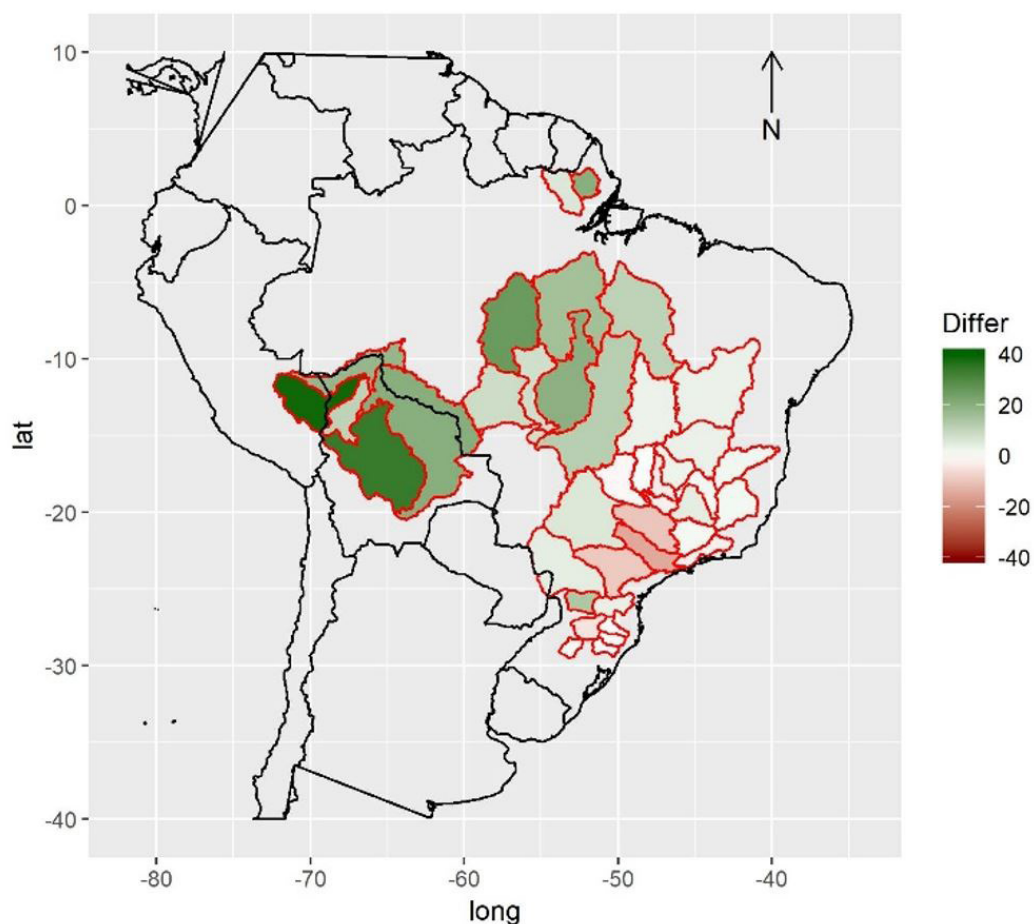
Figure 6. Map of average percent difference (%) of annual precipitation between TRMM-MERGE and CPC for the period 2008-2017.

Table 3. Average percent difference (%) of annual precipitation between TRMM-MERGE and CPC for the study basins and for the periods 1998-2007 and 2008-2017.

BG	Basin	1998-2008	2008-2017
I	Inc._Guajara_Mirim	2.9	20.3
	Inc._Santo_Antonio	1.8	18.0
	Mira_flores	24.9	39.0
	Pena_Amarilla	3.7	10.7
	Puerto_Siles	15.9	33.9
II	Belo_Monte_Inc.	6.0	15.7
	Boa_Sorte	7.7	19.6
	Conceicao_do_Araguaia	1.3	12.0
	EBEC	2.2	8.6
	Ferreira_Gomes	0.0	20.0
	Manso	1.8	7.9
	Sto_Antonio_do_Jari	3.3	5.7
	Tapajos	2.0	25.4
	Teles_Pires	-1.3	7.8
	Tucurui_Inc.	1.6	11.1
III	Carinhanha	0.4	3.2
	Itapebi	0.8	2.1
	Mascarenhas	-6.9	1.8
	Paracatu	-3.1	2.7
	Sobradinho_Inc.	-3.0	3.2
	Tres_Marias	-11.2	4.2
	Velhas	-4.0	1.7

Table 3. Continued...

BG	Basin	1998-2008	2008-2017
IV	Capim_Branco_2	-5.4	-2.1
	Emborcacao	-8.4	-2.8
	Furnas	-4.2	1.7
	Ilha_dos_Pombos	-11.1	-2.1
	Inc._Itumbiara	-9.5	0.2
	Inc._Sao_Simao	-4.4	-1.1
	Lajeado	-4.7	3.1
V	Agua_Vermelha	-1.9	-9.5
	Nova_Avanhandava	0.3	-14.3
	Rosana	1.2	-8.9
VI	Itaipu	-0.1	4.2
	Porto_Primavera	2.3	5.9
VII	Barra_Grande	-6.5	-0.5
	Campos_Novos	1.7	0.2
	G_B_Munhoz	-1.8	3.4
	Salto_Caxias	-3.4	13.7
VIII	Quatorze_de_Julho	5.0	-0.5
	Dona_Francisca	5.7	1.6
	Foz_do_Chapeco	2.7	-4.2

Table 4. Performance indicators (RMSE, NSE, KGE and R^2) and difference in performance (Diff) when calibrating the hydrological model HEC-HMS with TRMM-MERGE (TRMM cal) and CPC (CPC cal) precipitation and simulating river flows with the other dataset (CPC sim and TRMM sim, respectively) for two basins (Campos Novos and Tapajos) and considering the whole period 1997-2017 for calibration and simulation. Diff values indicate the percentage loss (negative values) or gain (positive values) in performance when moving from calibration with one dataset to simulation with the other dataset.

Basin	Campos Novos			Tapajos		
Perf. indicator	TRMM cal	CPC sim	Diff.	TRMM cal	CPC sim	Diff
RMSE (m³/s)	203	214	-6%	1959	3578	-83%
NSE	0.73	0.70	-4%	0.88	0.61	-31%
KGE	0.83	0.79	-4%	0.94	0.70	-26%
R²	0.65	0.60	-8%	0.87	0.50	-42%
	TRMM sim	CPC cal	Diff.	TRMM sim	CPC cal	Diff
RMSE (m³/s)	210	212	1%	3858	3383	-14%
NSE	0.71	0.72	-1%	0.55	0.66	-16%
KGE	0.82	0.79	3%	0.45	0.82	-44%
R²	0.65	0.63	3%	0.67	0.60	11%

data and then used in simulation with the CPC precipitation data as forcing and vice versa. Both calibration and simulation runs are performed over the same period, 1997-2017. We can see that model performance is good in both basins, with NSE and KGE values ranging between 0.71 and 0.94 in calibration and between 0.45 and 0.79 in simulation. As expected, the performance of the model behaves according to the magnitude of the differences between the precipitation data sources: for the basin where the differences are small (Campos Novos), the performance in calibration is similar to the performance in simulation (differences in performance indicators are between -4% and -8% when calibrating with TRMM-MERGE, and simulating with CPC and between -1% and 3% when calibrating with CPC and simulating with TRMM-MERGE). For the basin with a high difference in precipitation datasets (Tapajos), the decrease in performance from calibration to simulation is clear, with stronger differences in performance indicators. The most important losses are in accuracy (RMSE).

Figure 7 shows the ECDF of observed flows and of daily flows when calibrating the model with TRMM-MERGE and simulating with CPC. We can see that for the basin where the precipitation data of both sources are similar, the ECDF curves are also very similar. However, for the basin with a higher difference between the two sources, the ECDF curves show a clear difference. For the same probability level, the flows simulated with the CPC data are lower than the observed flows and the flows simulated with the TRMM-MERGE data. The ECDF curves obtained when using the CPC data for calibration and the TRMM-MERGE for simulation present a similar behavior (not shown in this paper).

Finally, in Figure 8, we present a comparison between the percent differences of annual precipitation (TRMM-MERGE minus CPC) and the percent differences of annual streamflow simulations, when using TRMM-MERGE data in calibration and CPC data in simulation. The points in the first quadrant indicate the situations where the values of precipitation and simulated flow are smaller when using the CPC data. The third quadrant
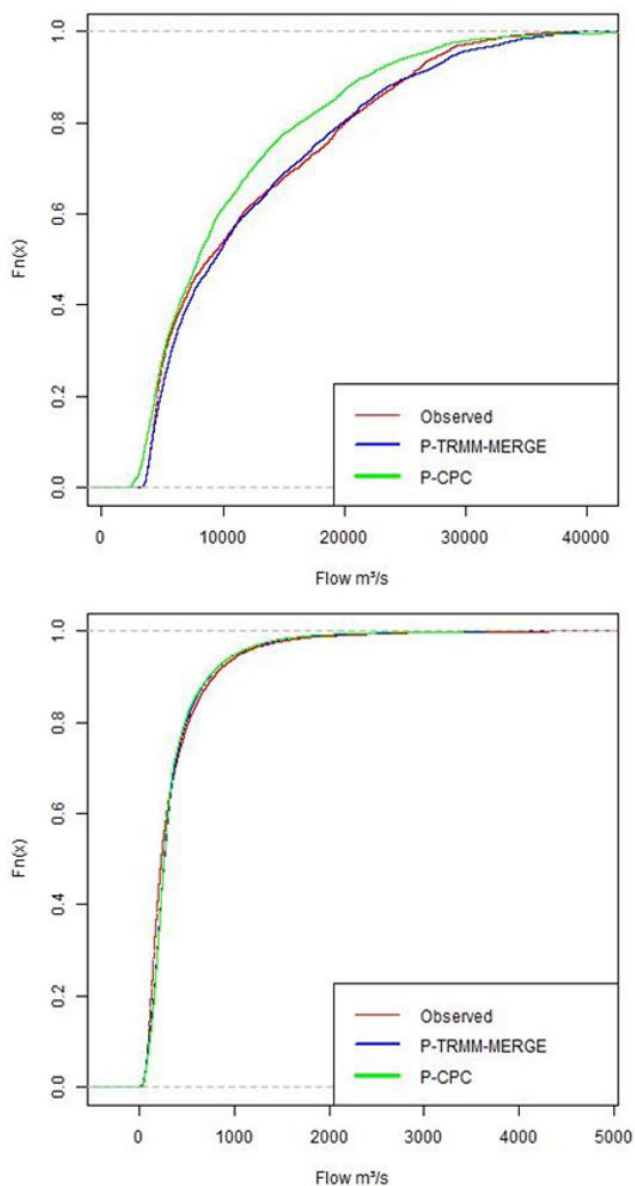
Figure 7. ECDF curves of the observed (red) and simulated daily flows (m³/s) with the TRMM-MERGE (blue) and the CPC (green) precipitation data as forcing in the HEC-HMS model for two basins (Tapajos, above, and Campos Novos, below).

indicates the opposite, the situations where the CPC precipitation is more intense than the TRMM-MERGE and therefore also the streamflow simulations based on CPC data.

We can see that in the basin with small differences between the precipitation datasets (Campos Novos), the relation between the differences of precipitation and the differences of flow (orange lines) tends to be closer to the diagonal line, with a slope that is higher when the CPC precipitation is higher than the TRMM-MERGE precipitation. For the basin with the high difference in precipitation datasets (Tapajos), the slopes of the regression lines (blue lines) are higher than 1 for both quadrants, first and third. The angular coefficient of the line of the third quadrant is also bigger than
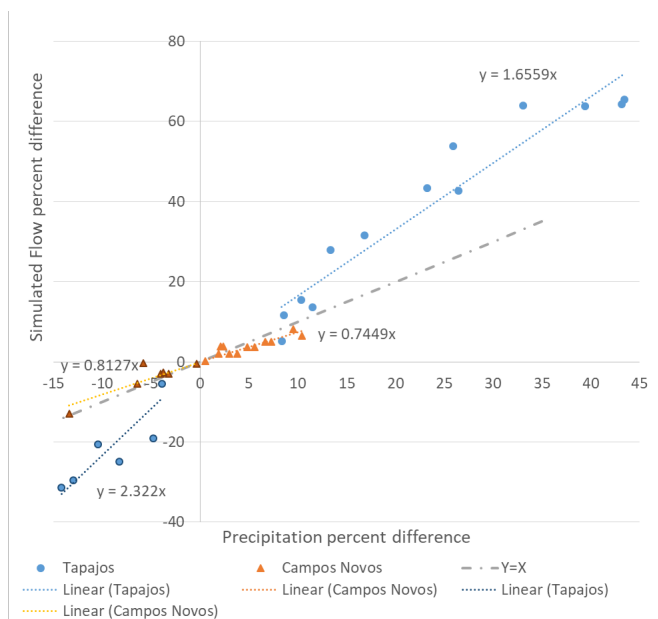

Figure 8. Percent differences (%) in annual simulated flows as a function of percent differences in annual precipitation between TRMM-MERGE and CPC datasets for two basins (Tapajos in blue and Campos Novos in orange).

the one of the line in the first quadrant. The graphs with the CPC data used in calibration and the TRMM-MERGE data used in simulation (not shown in this paper) present a similar behavior for the Campos Novos basin. For the Tapajos basin, where the precipitation of the simulation dataset is more intense then the precipitation in the CPC calibration dataset, the differences in flows are amplified.

DISCUSSION

The TRMM-MERGE precipitation dataset uses raingauge data to calibrate the system. They have quality control of the sensors and use the telemetry rain data to improve the interpolation of the grid data (Huffman et al., 2017). The CPC precipitation data uses satellite information to perform the quality control of the rain gauge data and to create a better and trustworthy grid of precipitation data (Chen et al., 2008). Despite the fact that both sources use information from satellite and rain gauges in different degrees and ways, they both try to represent the same variable and one could expect they would provide similar datasets.

In our comparative study, the first signal of the differences between TRMM-MERGE and CPC precipitation datasets appears in the results from the basic statistics of annual precipitation totals. We found high differences, up to 42%, for the minimum annual precipitation, as well as for the first quartile (up to 25%). The other values (mean, third quartile and maximum) also show differences, but they are not as strong. This shows that, in the absence of a more accurate dataset of ground precipitation data, and considering that the observed precipitation in a certain basin is an estimation of the actual precipitation, uncertainties

are present in any dataset and, for our study, both datasets need to be considered.

Our study provides an extensive analysis of differences in precipitation data for a wide range of basins in South America, covering the continent from north (3°N) to south (30°S) and a variety of climatic conditions. Our results show that the 41 basins studied can be grouped into eight groups of similar behavior. The results from the group I, the Madeira river basin, show that the precipitation values from the TRMM-MERGE are higher than the values obtained from the CPC dataset for all the study period (1997-2017), with differences that tend to increase in the last years. The analysis of monthly precipitation totals shows that the highest differences are observed during the wet season, which goes from November to April. During the wettest months, it is common to observe differences higher than 100 mm/month and some maximum values higher than 200 mm/month, which can be higher than the monthly average precipitation. The analysis of the statistical distribution of daily precipitation shows that the TRMM-MERGE daily precipitation quantiles tend to be higher than the CPC quantiles for the same probability of occurrence. This is observed for the majority of the probability quantiles.

The results from the group II, representing other basins in the north region, exhibit smaller differences between the precipitation datasets for the first decade of the data period (1998-2007), with alternation of positive and negative differences. However, in the last decade (2008-2017), the differences become more often positive, showing that CPC precipitation tends to be lower than TRMM-MERGE precipitation. These differences in the last decade translate into high variations during the wet season, from January to March. It is common to observe differences higher than 100 mm/month. The maximum differences occur at the basins in the extreme north, with values near to 200 mm/month, which can be around 70% and 80% of the monthly averages. The ECDF curves of daily precipitation also exhibit the tendency of TRMM-MERGE daily precipitations higher than the CPC values for the same probability of occurrence. The magnitude of the differences depends on the basin.

These results for the north region are coherent with the results obtained by Negrón Juárez et al. (2009). The authors show positive differences when comparing TRMM-MERGE and CPC precipitation data. Their differences were smaller, but the values were only computed considering the first decade of our data period.

The analysis of the northeast and east basins, group III, shows annual precipitation differences smaller than in the north region, with differences that, on the majority of the basins, are lower than 10%. The values tend to be slightly negative on the first decade of the study period and become slightly positive along the second decade. For the major part of the basins, the smaller differences are observed during the last five years, which can be an indication that both data sources may be getting more similar along the years in terms of annual precipitation totals. The differences in monthly precipitation show that the majority of differences in this group of basins are between -25 mm/month and 25 mm/month. The highest differences occur again during the wet season, November-March, but they are often smaller than 20% of the monthly average precipitation. In terms of daily

precipitation, the ECDF curves do not indicate a great difference between the two data sources.

The basins in the center-west and southeast regions, group IV, show a tendency to have negative values of the differences between datasets in the first decade, varying from small values near 5% until higher values near 25% for some basins. In the second decade of the study period, the differences become positive (i.e., TRMM-MERGE precipitation data are higher than CPC data), but the values are smaller than 10%. For the monthly precipitation differences, the highest differences occur in November, December, January and February, with highlights for strong negative values that sometimes are close to -100 mm/month. For the daily precipitation values, the probability distribution curves do not indicate a great difference between the two sources of precipitation data.

The three basins in group V, in the southeast region, display a different behavior from the others. The differences in annual precipitation are slightly positive in the first decade and tend to be negative during the second decade. The highest differences occur between the years 2008-2013, with values near -30%. During the last five years of the study period, the differences become small again, near 5%. The monthly differences are higher from November to February. In terms of daily values, there is a tendency for TRMM-MERGE daily precipitation to be lower than the CPC precipitation for the same probability of occurrence.

In the group VI, basins on the neighborhood of the South region, the differences in annual precipitation are very small during the first-decade and become positive along the second decade, with values higher than 10%. The highest differences occur during the wet season, varying, in the majority of basins, between -50 mm/month and 50 mm/month. The ECDF curves show that the daily precipitation from TRMM-MERGE tends to be higher than CPC precipitation for the same probability of occurrence.

The group VII, basins in the south region, displays small negative differences during the first decade, which then become positive along of the last decade, with some values higher than 10%. These basins do not have a clear seasonality and differences in monthly precipitations spread all over the year, with the majority of values between -50mm/month and 50mm/month. For daily precipitation values, the ECDF curves show that the TRMM-MERGE daily precipitations are higher than the CPC values for the same probability of occurrence.

Finally, at the extreme south region, the group VIII exhibits a change of behavior in terms of differences in annual precipitation, when compared with the other south basins. Differences are more often positive in the first decade and tend to reduce along the second decade. In terms of monthly values, this group has the same behavior as group VII, without a wet season with higher differences. For daily precipitation values, the statistical distributions show that the TRMM-MERGE daily precipitations are often higher than the CPC values for the same probability of occurrence.

Our analyses show clearly a regional pattern on the differences between the two precipitation data sources. As we move to the north and west regions of the study area, the annual differences tend to become more positive. This spatial variability in annual precipitation differences is amplified in the second and

most recent decade of the study period. The basins located in the northeast, east, and southeast regions have smaller differences and these differences tend to become more positive in the second decade, although at a smaller degree. The exceptions are the basins of the group VI, which tend to display more negative differences in annual precipitation during the last decade.

We also evaluated the impact of the observed differences on the simulation of streamflows, using calibrated hydrological models in two basins (Campos Novos and Tapajos), representative of the lowest and highest differneces between precipitation data sources. The analysis showed how the models are sensitive to changes in precipitation, confirming the general findings in Fan (2015). If the two precipitation data sources used in calibration and simulation are similar, model performance is also similar. However, when they are very different, the performance indicators showed that the hydrological model tends to lose performance. The amount of loss in performance may vary according to the quality of the data source used in the calibration. In our study, in the Tapajos basin, where the differences between TRMM-MERGE and CPC data are high, the performance loss was stronger when calibration was performed with TRMM-MERGE and CPC was used in simulation, comparatively to the opposite situation. The empirical relationship between annual precipitation and flow values shows that the dispersion is higher when dealing with the CPC data. This indicates that CPC data has more uncertainty, which impacts the results when the CPC data is used to simulate flows in a model that was calibrated with another data source. More uncertainty in the CPC data in this basin can be explained by the low density of raingauge stations in this area. The use of satellite information may give more accuracy to the TRMM-MERGE dataset in this case.

Another result of our study is that the hydrological model seems to propagate and amplify the differences in precipitation data into differences in streamflow simulations. Small differences in precipitation result in similar small differences in streamflow. However, large differences in precipitation seem to result in even larger differences in streamflow.

CONCLUSIONS

This study aimed to evaluate the differences between precipitation data obtained from two sources over 41 river basins in South America, the TRMM-MERGE and the CPC datasets, for the period 1997-2017. We investigated differences for different time resolutions (daily, monthly and annual precipitation), at different locations and according to their impact on the simulation of streamflow.

The results show that differences vary in space and time, and according to the temporal aggregation of the precipitation values. The second decade tends to amplify the observed differences in the majority of the basins.

Some basins show considerable differences, notably in terms of daily and monthly precipitation values, with an expected impact on the simulation of daily streamflows, which are also affected by the uncertainty of each precipitation data source. In addition, a spatial behavior of the differences between the precipitation sources was detected, with differences becoming more positive

(i.e., TRMM-MERGE values are higher than CPC values) as we move to north and west in the study area.

With the results of this study, we recommend being cautious when working with a unique source of historic precipitation data to calibrate hydrological models, since this source can display uncertainties and errors that vary in space and time. In our study, we showed that it is a complex problem to determine a precipitation data source that is the best for all situations, especially when no observed data set can be used as ground truth or reference, as in the case of large continental areas such as South America.

When it comes to maximize the performance of streamflow simulations, it becomes important to extract information from all data sources available. In this study, we illustrated how hydrological models can be sensitive to changes in the precipitation data, especially when these changes reflect high differences between different forcing data sources. The use of observed streamflow is an alternative to help selecting the best precipitation data source. The comparison between observed and simulated streamflows is an indirect way to carry out the precipitation data analysis, but it can, nevertheless, be useful in hydrological applications at large river basins.

Further research will focus on accessing the uncertainties and investigating how data sources such as TRMM-MERGE and CPC can be combined, with varying weights according to basin location and time of the year, to provide a more robust long time series of precipitation data for hydrological model calibration and simulation. The goal is to have time series of forcing data that minimize the errors between observed and simulated flows in the past, so that these time series can be used for seasonal forecasting in the hydropower sector within the traditional ESP (Ensemble Streamflow Prediction) method, where hydrological models and historical precipitation are used to generate a set, or an ensemble, of possible flow scenarios dependent on the initial states of a given basin in real-time.

ACKNOWLEDGEMENTS

REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-Present). *Journal of Hydrometeorology*, 4(6), 1147-1167. http://dx.doi.org/ 10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS – Global Ensemble Streamflow Forecasting and Flood Early Warning. *Hydrology and*

Earth System Sciences, *17*(3), 1161-1175. http://dx.doi.org/10.5194/hess-17-1161-2013.

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., & Pappenberger, F. (2018). Skillful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, *22*(4), 2057-2072. http://dx.doi.org/10.5194/hess-22-2057-2018.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25_ global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21, 589-615. http://dx.doi.org/10.5194/hess-21-589.

Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., & Michael, K. (2017). Assessment of an ensemble seasonal streamflow forecasting system for Australia. *Hydrology and Earth System Sciences*, *21*(12), 6007-6030. http://dx.doi.org/10.5194/hess-21-6007-2017.

Chen, M., Xie, P., Shi, W., Silva, V., Kousky, V., Higgins, W., & Janowiak, J. E. (2008). Quality control of daily precipitation reports at NOAA/CPC. In *Proceedings of the 12th Conference on IOAS-AOLS Adjourns* (pp. 1-7). Maryland: NOAA.

Centro de Previsão do Tempo e Estudos Climáticos – CPTEC, & Instituto Nacional de Pesquisas Espaciais – INPE. (2018). *TRMM-MERGE Dataset ftp*. Retrieved in 2018, January 22, from ftp:ftp1.cptec.inpe.br/modelos/io/produtos/MERGETRMM-MERGE

Crochemore, L., Ramos, M.-H., & Pappenberger, F. (2016). Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *20*(9), 3601-3618. http://dx.doi.org/10.5194/hess-20-3601-2016.

Demaria, E. M. C., Rodriguez, D. A., Ebert, E. E., Salio, P., Su, F., & Valdes, J. B. (2011). Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach. *Journal of Geophysical Research*, *116*(D8), D08103. http://dx.doi.org/10.1029/2010JD015157.

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., & Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev. Water*, *3*(3), 391-418. http://dx.doi.org/10.1002/wat2.1137.

Empresa de Pesquisa Energética – EPE. (2015). *Balanço Energético Nacional 2015: Ano base 2014* (292 p.). Rio de Janeiro: EPE.

Falck, A., Maggioni, V., Tomasella, J., Vila, D. A., & Diniz, F. L. R. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: a case study in the Tocantins–Araguaia basin in Brazil. *Journal of Hydrology*, 527, 943-957, http://dx.doi.org/10.1016/j.jhydrol.2015.05.042.

Fan, F. M. (2015). *Previsão por conjunto de vazões afluentes a reservatórios em grandes bacias hidrográficas brasileiras* (Tese de doutorado). Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental, Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Fan, F. M., Schwanenberg, D., Alvarado, R., Reis, A. A. D., Collischonn, W., & Naumman, S. (2016). Performance of deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower reservoir. *Water Resour Manage*, 30, 3609-3625. http://dx.doi.org/10.1007/s11269-016-1377-8.

Feldman, A. D. (2000). *Hydrologic Modeling System – HEC-HMS – Technical Reference Manual*. USA: USACE – US Army Corps of Engineers - Hydrologic Engineering Center.

Gibertoni, R. F. C., Saboia, J. P. J., Daru, R. L., Kan, A., Oening, A. P., Detzel, D. H. M., Talsma, J., & Reis, A. A. (2017). Previsão de vazões e operação de reservatórios através do Sistema Fews-CEMIG. In *Anais do Congresso de Inovação Tecnológica em Energia Elétrica – IX CITENEL*. Brasília: CITENEL.

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 modeling. *Journal of Hydrology*, 377(1-2), 80-91. http://dx.doi.org/10.1016/J.JHYDROL.2009.08.003.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., & Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, *22*(3), 2023-2039. http://dx.doi.org/10.5194/hess-22-2023-2018.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J, McNab, A., Rudolf, B., & Schneider, U. (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, 78(1), 5-20. http://dx.doi.org/10.1175/15200477.

Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2017). *Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation*. USA: NASA Goddard Space Flight Center.

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A. T. C., Stocker, E., Adler, R. F., Hou, A., Kakar, R., Wentz, F., Ashcroft, P., Kozu, T., Hong, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z., Huffman, G., Ferrier, B., Olson, W. S., Zipser, E., Smith, E. A., Wilheit, T. T., North, G., Krishnamurti, T., & Nakamura, K. (2000). The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *Journal of Applied Meteorology and Climatology*, *39*(12), 1965-1982. http://dx.doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2.

Maggioni, V., Mayers, P. C., & Robinson, M. D. (2016). A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *Journal of Hydrometeorology*, 17, 1101-1117. https://doi.org/10.1175/JHM-D-15-0190.1.

Mantas, V. M., Liu, Z., Caro, C., & Pereira, A. J. S. C. (2015). Validation of TRMM Multi-satellite Precipitation Analysis (TMPA) products in the Peruvian Andes. *Atmospheric Research*, 163, 132-145. http://dx.doi.org/10.1016/j.atmosres.2014.11.012.

Naghettin, M., & Pinto, E. J. A. (2007). *Hidrologia estatística*. Belo Horizonte: CPRM.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology (Amsterdam)*, *10*(3), 282-290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.

National Oceanic and Atmospheric Administration – NOAA. (2018). *Gauge Global Dataset ftp*. Retrieved in 2018, January 18, from ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB

Negrón Juárez, R. I., Li, W., Fu, R., Fernandes, K., & Cardoso, A. O. (2009). Comparison of precipitation datasets over the Tropical South American and African. *Journal of Hydrometeorology*, *10*(1), 289-299. http://dx.doi.org/10.1175/2008JHM1023.1.

Operador Nacional do Sistema Elétrico – ONS. (2016). *Procedimentos de rede*. Retrieved in 2016, September 20, from http://apps05.ons.org.br/procedimentorede/procedimento_rede/procedimento_rede.aspx

Pinto, R. B., Gibertoni, R. F. C., Rego, J. L., Schwanenberg, D., & Reis, A. A. (2013). Sistema Integrador DELFT_FEWS: usos e aplicações no cenário brasileiro. In *Anais do XX Simpósio Brasileiro de Recursos Hídricos*. São Paulo: ABRH.

Rozante, J. R., Soares, M. D., Gonçalves, L. G. G., & Vila, D. A. (2010). Combining TRMM and surface observations of precipitation: technique and validation over South America. *American Meteorological Society*, 25, 885-894. https://doi.org/10.1175/2010WAF2222325.1.

Scharffenberg, W. (2016). *Hydrologic Modeling System – HEC-HMS – User's Manual, version 4.2*. USA: USACE – US Army Corps of Engineers - Hydrologic Engineering Center.

Scheel, M. L. M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E., & Huffman, G. J. (2011). Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution. *Hydrology and Earth System Sciences*, *15*(8), 2649-2663. http://dx.doi.org/10.5194/hess-15-2649-2011.

Schwanenberg, D., Fan, F. M., Naumann, S., Kuwajima, J. I., Montero, R. A., & Reis, A. A. (2015). Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty - application to the Três Marias Reservoir in Brazil. *Water Resources Management*, 29, 1635-1651. http://dx.doi.org/10.1007/s11269-014-0899-1.

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., Paris, A., Calmant, S., & Collischonn, W. (2018). Toward continental hydrologic-hydrodynamic modeling in South America. *Hydrology and Earth System Sciences*, 1-50. https://doi.org/10.5194/hess-2018-225.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, H.-L. (2017). A review of global precipitation data sets: data sources, estimation and Inter-comparisons. *Reviews of Geophysics*, *56*(1), 79-107.

Werner, M., Schellekens, J., Gijsbers, P., Van Dijk, M., van den Akker, O., & Heynert, K. (2013). The Delft-FEWS Flow Forecasting System. *Environmental Modelling & Software*, 40, 65-77. http://dx.doi.org/10.1016/j.envsoft.2012.07.010.

Authors contributions

Alberto Assis dos Reis: Lead author of the article and PhD research developer, responsible for data acquisition, analysis, interpretation, methodology and drafting of manuscript.

Wilson dos Santos Fernandes: Co-author, Doctoral Research Advisor, general review of the text, the methodology used in the evaluation of precipitation data and presentation of results and discussion.

Maria-Helena Ramos: Co-Author, Co-Advisor of the doctoral research, general review of the text, the methodology used in the evaluation of the precipitation data and the presentation of the results and discussion.