Atmospheric Research 213 (2018) 34–50

Contents lists available at ScienceDirect

Atmospheric Research

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

Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America

Oscar Manuel Baez-Villanueva[a,e,*], Mauricio Zambrano-Bigiarini[b,c], Lars Ribbe[a], Alexandra Nauditt[a], Juan Diego Giraldo-Osorio[d], Nguyen Xuan Thinh[e]

[a] *Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Technology Arts Sciences TH Köln, Cologne, Germany*
[b] *Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile*
[c] *Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile*
[d] *Pontificia Universidad Javeriana, Bogotá, Colombia*
[e] *Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany*

A R T I C L E I N F O

Keywords:
CHIRPSv2
MSWEPv2
Precipitation
Satellite
Upscaling influence
Validation of SREs

A B S T R A C T

In developing countries, an accurate representation of the spatio-temporal variability of rainfall is currently severely limited, therefore, satellite-based rainfall estimates (SREs) are promising alternatives. In this work, six state-of-the-art SREs (TRMM 3B42v7, TRMM 3B42RT, CHIRPSv2, CMORPHv1, PERSIANN-CDR, and MSWEPv2) are evaluated over three different basins in Latin-America, using a point-to-pixel comparison at daily, monthly, and seasonal timescales. Three continuous (root mean squared error, modified Kling-Gupta efficiency, and percent bias) and three categorical (probability of detection, false alarm ratio, and frequency bias) indices are used to evaluate the performance of the different SREs, and to assess if the upscaling procedure used, in CHIRPSv2 and MSWEPv2, to enable a consistent point-to-pixel comparison affects the evaluation of the SREs performance at different time scales.

Our results show that for Paraiba do Sul in Brazil, MSWEPv2 presented the best performance at daily and monthly time scales, while CHIRPSv2 performed the best at these timescales over the Magdalena River Basin in Colombia. In the Imperial River Basin in Chile, MSWEPv2 and CHIRPSv2 performed the best at daily and monthly time scales, respectively. When the basins were evaluated at seasonal scale, CMORPHv1 performed the best for DJF and SON, TRMM 3B42v7 for MAM, and PERSIANN-CDR for JJA over Imperial Basin. MSWEPv2 performed the best over Paraiba do Sul Basin for all seasons and CHIRPSv2 showed the best performance over Magdalena Basin. The Modified Kling-Gupta efficiency (KGE′) proved to be a useful evaluation index because it decomposes the performance of the SREs into linear correlation, bias, and variability parameters, while the Root Mean Squared Error (RMSE) is not recommended for evaluating SREs performance because it gives more weight to high rainfall events and its results are not comparable between areas with different precipitation regimes.

On the other hand, CHIRPSv2 and MSWEPv2 presented different performance, for some study areas and time scales, when evaluated with their original spatial resolution (0.05° and 0.1, respectively) with respect to the evaluation resulting after applying the spatial upscaling (to a unified 0.25), showing that the upscaling procedure might impact the SRE performance. We finally conclude that a site-specific validation is needed before using any SRE, and we recommend to evaluate the SRE performance before and after applying any upscaling procedure in order to select the SRE that best represents the spatio-temporal precipitation patterns of a site.

1. Introduction

Precipitation is an important variable of the water cycle, it must be assessed carefully due to its changing spatial and temporal distribution. However, an accurate representation of the spatio-temporal variability of catchment rainfall inputs is currently severely limited (Zambrano-Bigiarini et al., 2017). Traditionally, the spatial distribution of precipitation is measured through ground observations forming a rain gauge network (e.g. Berndt et al., 2013; Haberlandt, 2006; Rogelis and Werner, 2013). In developing countries the network of rain gauge stations is sparse, and therefore, the interpolation using point-based rainfall information is subject to a large uncertainty (Woldemeskel

* Corresponding author at: Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Technology Arts Sciences TH Köln, Cologne, Germany.
E-mail addresses: obaezvil@th-koeln.de (O.M. Baez-Villanueva), mauricio.zambrano@ufrontera.cl (M. Zambrano-Bigiarini), lars.ribbe@th-koeln.de (L. Ribbe), alexandra.nauditt@th-koeln.de (A. Nauditt), j.giraldoo@javeriana.edu.co (J.D. Giraldo-Osorio), nguyen.thinh@tu-dortmund.de (N.X. Thinh).

https://doi.org/10.1016/j.atmosres.2018.05.011
Received 16 December 2017; Received in revised form 15 April 2018; Accepted 21 May 2018
Available online 24 May 2018
0169-8095/

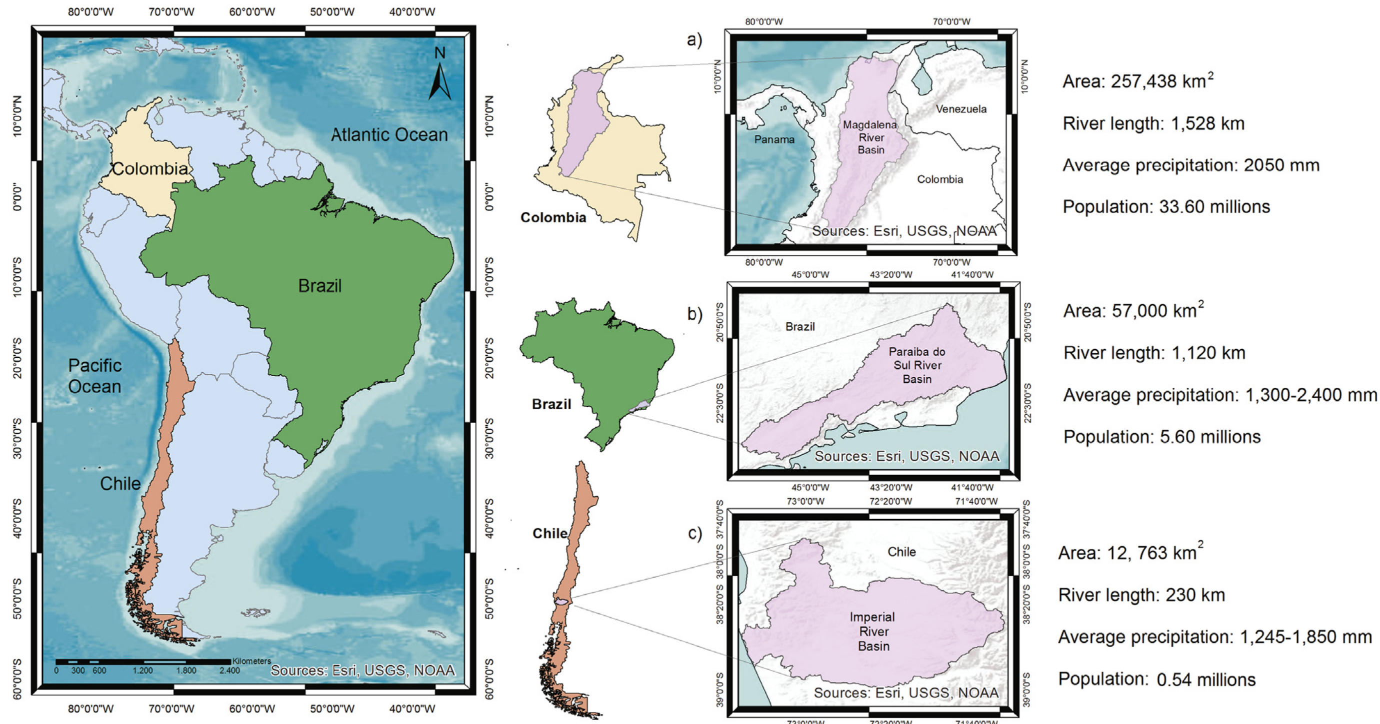

Fig. 1. Localization of the study areas; a) Magdalena River Basin, located in Colombia, b) Paraiba do Sul Basin, in Brazil and c) Imperial River Basin, in Chile.

et al., 2013).

Thanks to the technologic progress, new methodologies are being developed to estimate the amount and spatial distribution of rainfall. In particular, remote sensing techniques based on infrared (IR) and passive microwave (PMW) sensors aboard geostationary (GEO) and Low Earth Orbit (LEO) satellites, respectively.

Several satellite-based precipitation products have been developed and have recently become operational, including the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Hsu et al., 1997; Sorooshian et al., 2000), the PERSIANN Climate Data Record(PERSIANN-CDR; Ashouri et al., 2015), the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007), the Climate Prediction Center Morphing technique product (CMORPHv1; Joyce et al., 2004; Xie et al., 2017), the Climate Hazards group Infrared Precipitation with Stations dataset (CHIRPSv2; Funk et al., 2015), the Multi-Source Weighted-Ensemble Precipitation (MSWEPv2; Beck et al., 2017a), among others.

The evaluation of satellite products over Africa and South America has been very limited (Espinoza Villar Jhan Carlo et al., 2009; Dinku et al., 2010). Satellite rainfall estimates are needed the most in environments with sparse or non-existent rain gauge networks. However, the lack of ground observations and the lack of access to the available information are strong limitations to validate the SREs over those regions. On the other hand, SRE datasets may be used erroneously if a validation process is not taken into account to detect systematic and/or random errors (Gebremichael et al., 2010). Validation will help to quantify the errors in the different SREs in order to choose the one that represents better the precipitation over an area of interest.

Recently, Salio et al. (2015) performed the evaluation of six different SREs over South America (TRMM 3B42v6, TRMM 3B42v7, TRMM3B42RT, CMORPHv0, HYDRO, and the CoSch), concluding that the performance of Satellite Rainfall Estimates that include microwave measurements is higher than the ones with only infrared (IR) information. They also concluded that further research has to be done regarding the effects of topography.

Zambrano-Bigiarini et al. (2017) carried out a temporal and spatial evaluation of different SREs across Chile, evaluating seven different SREs. They found that the calibration of SREs with observed data improves the performance of the datasets but a site-specific validation is still needed. CHIRPSv2, TRMM 3B42v7 and MSWEPv1.2 performed relatively well for the Chilean territory. Also, the satellite products performed the best for humid and low and mid-elevation zones (0–1000 m.a.s.l.).

Melo Davi de et al. (2015) evaluated TRMM 3B42v6 and 3B42v7 over Brazil on a daily, monthly and seasonal basis. The findings were related to the low performance of the products in the daily and seasonal analysis, mostly in the northwest region, which is related to higher rainfall depths. However, both products performed better during the dry seasons and the most recent version improved the performance over the northeast, southeast and southern regions.

Finally, Dinku et al. (2010) evaluated seven different SREs over Colombia, finding a low performance in PERSIANN and TRMM 3B42RT and relatively good performance for CMORPHv0. The study showed that the products were good in detecting the occurrence of rainfall but poor in estimating the amount of daily precipitation.

As observed, there is not a best performing SRE on a general basis. Therefore, a process for validating the satellite-derived information is required for each case study. Even though there are a plethora of studies characterizing and quantifying errors in spatial fields of rainfall measurement, there is no study that comprehensively evaluates the 6 products aforementioned over the studied areas.

For this reason, six state-of-the-art satellite rainfall estimates (SREs) are analyzed (TRMM 3B42v7, TRMM 3B42RT, CHIRPSv2, CMORPHv1, PERSIANN-CDR, and MSWEPv2) for the first time over three different basins (regarding climate, area, and topography) in Latin-America (Imperial River Basin in Chile, Paraiba do Sul Basin in Brazil, and Magdalena River Basin in Colombia) in a point-to-pixel comparison at daily, monthly, and seasonal scales.

The aims of this work are i) to assess the spatio-temporal performance of different SREs over the three different study areas in Latin-America at different time scales, and ii) to evaluate if the upscaling procedure used to ensure a consistent point-to-pixel comparison (in the case of CHIRPSv2 and MSWEPv2) to a spatial resolution of 0.25 affects

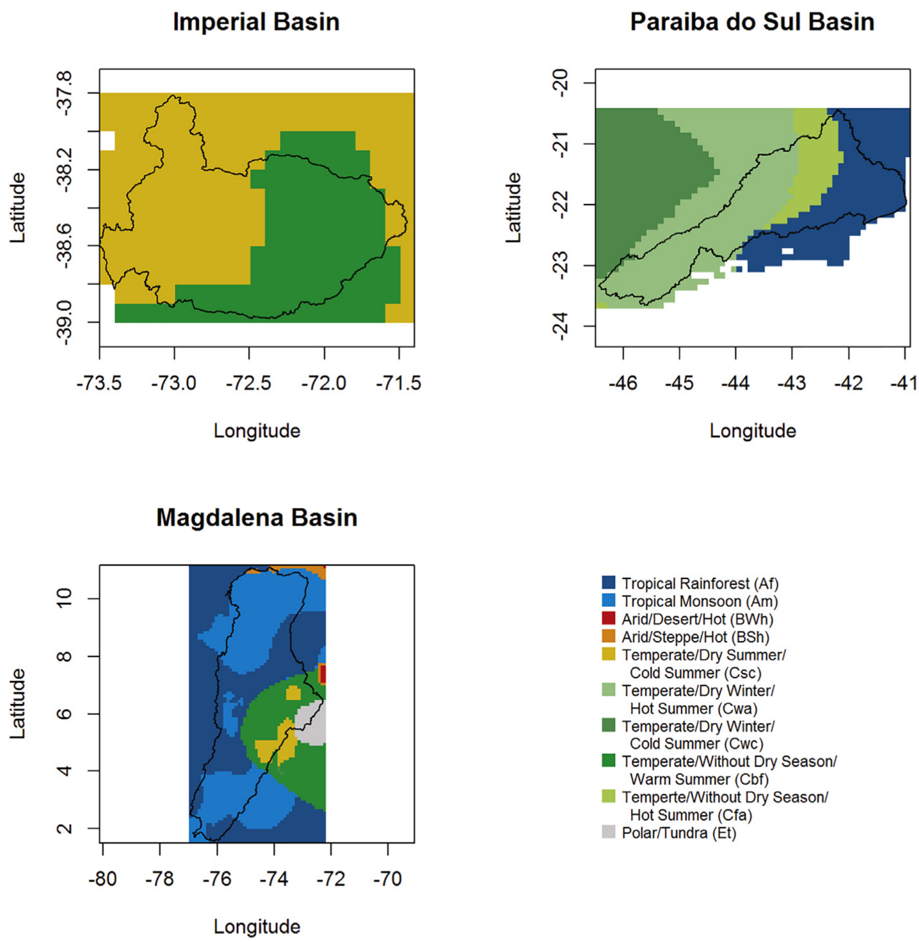

Fig. 2. Köppen-Geiger climate regions over the different study areas; Imperial River Basin (top-left), Paraiba do Sul Basin (top-right) and Magdalena River Basin (bottom-left).

the evaluation of the performance. Once an SRE is validated over an area, it can be used for different purposes such as rainfall-runoff modeling, water allocation or climatic trends evaluation, among others.

2. Study areas

Three basins were selected for this analysis (Fig. 1); Magdalena River Basin in Colombia, Paraiba do Sul Basin in Brazil and Imperial River Basin in Chile. The different study areas were selected in order to compare the results of the best performing SRE in different environments in Latin America. The selected basins present differences mainly in area, climate, and topography. According to the world map of the Köppen-Geiger climate classification (Peel et al., 2007) and as observed in Fig. 2, Imperial River Basin presents a temperate climate with dry and warm summer (Csb) in the North-eastern part and a temperate climate without dry season and warm summer (Cbf) in the South-east. Paraiba do Sul presents a temperate climate with dry winter and hot summer (Cfa) in the upper part of the basin while a tropical rainforest (Af) and tropical monsoon (Am) climates are present in the lower part. Finally, Magdalena Basin shows tropical rainforest (Af), tropical monsoon (Am) and tropical savannah (Aw) climates in the upper and lower parts of the basin and temperate climate without dry season and warm summer (Cbf), temperate climate with dry and warm summer (Csb) and polar tundra (ET) in the elevated zones of the center of the basin.

Magdalena River Basin is the greatest catchment in Colombia and is highly important in the economic development of the nation. Paraiba do Sul, involves Rio de Janeiro, Minas Gerais, and Sao Paulo states being the most important contributor basin to the Brazilian GDP. Finally, Imperial River Basin is essential because the agricultural activity of Chile is mainly developed in the southern region of the country.

The average monthly precipitation cycles for Imperial and Paraiba do Sul basins (period 2001–2015), and for Magdalena Basin (period 2001–2014) are presented in Fig. 3. For Imperial Basin, the rainy season starts in April and ends in September. The opposite is observed in Paraiba do Sul, where the rainy season starts in October and ends in March. For Imperial and Paraiba do Sul basins, the rainy season presents a greater dispersion from the mean values. On the other hand, Magdalena Basin presents a bimodal precipitation pattern caused by the double pass of the Intertropical Convergence Zone (ITCZ) with the highest precipitation values between April–May and October–November.

2.1. Imperial Basin

The Imperial Basin (37° 40′ to 38° 50′ S in latitude and 73° 30′ to 71° 27′ W in longitude) is located in the Araucana region. The area of the basin is 12763 km^2 with 540599 inhabitants. Its river length is about 230 km. The hydrological regime of this basin is pluvial. This characteristic is associated with the low snow accumulation associated with the relatively small altitude in the Andean Mountain range on its latitude (Rivera et al., 2004). The elevations for this basin ranges from 0 to 3095 m.a.s.l. (meters above sea level). The average annual precipitation for this basin ranges from 1500 to 2500 mm.

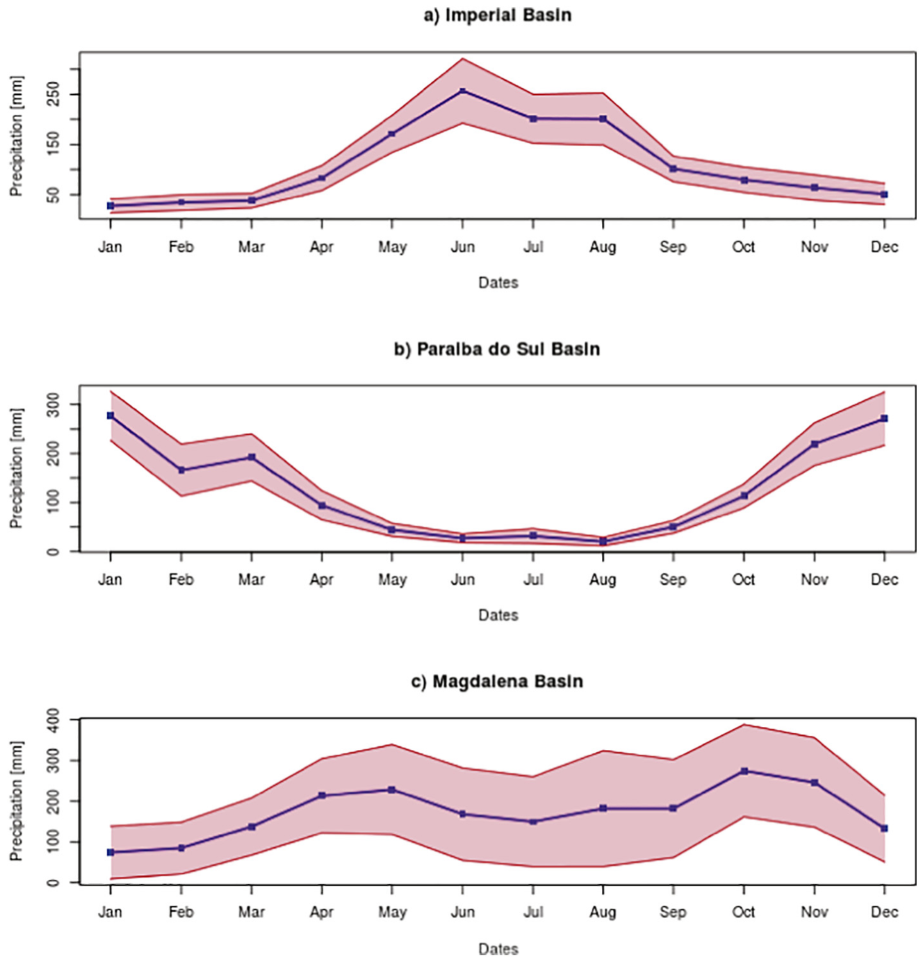

Fig. 3. Annual precipitation cycle for the studied areas. a) Imperial Basin, b) Paraiba do Sul, and c) Magdalena Basin for the period 2001–2015. The blue line represents the mean values for each catchment and the red contour represents one standard deviation above and below the mean value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Paraiba do Sul

Paraiba do Sul is located in Brazil, between the latitudes 20° 26′ and 23° 39′ S and between the longitudes 41° and 46° 30′ W covering an area of nearly 56500 km^2. It has a heterogeneous topography, geomorphology, hydrology, and soil composition (Simões and Barros, 2007). The basin is located among the states of Sao Paulo, Minas Gerais, and Rio de Janeiro and the river length is about 1120 km. The average annual precipitation is 1400 mm, but it exhibits a large inter-annual variability ranging values between 1300 and 2400 mm (Simões and Barros, 2007) and large spatial variability (Soares et al., 2012). It is bordered by two mountain chains; Serra da Mantiqueira and Serra do Mar, presenting variations in elevation between 450 and 2000 m.a.s.l. (Soares et al., 2012). This basin currently accounts for approximately 11% of national gross domestic product (GDP). Therefore, water management is very important for regional development and urban growth. About 8.70 million inhabitants live in Rio de Janeiro Metropolitan area and they depend on Paraiba do Sul Basin for water supply (Silva and Simões, 2014).

2.3. Magdalena Basin

Magdalena Basin is located in Colombia, between the latitudes 11° 06′ and 1° 33′ S and the longitudes 76° 58′ and 72° 22′ W with a range of altitude going from 0 to 5000 m.a.s.l. In this basin, the 85% of the national GDP is generated. Magdalena Basin counts with a population of around 33.60 million of inhabitants. The principal urban centers are located inside of this basin. The Magdalena River is the biggest river system in Colombia, extending for 1612 km. It drains the Andes, which are formed in Colombia by the Western, Central, and Eastern Cordilleras covering an area of 257438 km^2, the 24% of the total area of the country (Restrepo et al., 2005). The annual average rainfall for the basin is 2050 mm. The basin presents two wet (March–May and October–November) and two dry (December–February and June–-September) seasons (IDEAM, 2001).

3. Datasets

3.1. Observed data

Time series of observed precipitation were obtained for each study area. For Imperial and Paraiba do Sul basins the analyzed period starts in January 2001 and ends in December 2015 while for Magdalena River Basin starts in January 2001 and ends in December 2014. The starting date of the analysis is related to TRMM 3B42RT temporal coverage and the ending date is related to the availability of ground-based rainfall.

For Imperial Basin in Chile, 29 daily time series from rain gauge stations were downloaded from the Center of Climate and Resilience Research (CR2) (http://www.cr2.cl/recursos-y-publicaciones/bases-de-datos/). These data were provided by the Chilean Water and Meteorological agencies; Dirección General de Aguas and Dirección Meteorolgica de Chile (DGA and DMC, respectively). One hundred seventy-four stations provided by the National Water Agency of Brazil (ANA) were obtained for Paraiba do Sul Basin (http://hidroweb.ana.

Table 1
Summary of satellite-based products used in this work.

Name	Spatial res.	Temporal res.	Period	Spatial coverage	References
TRMM 3B42RT	0.25°	3 hourly, daily	1998–2015	Global (50°S–50°N)	(Huffman et al., 2007)
TRMM 3B42v7	0.25°	3 hourly, daily	1998–2015	Global (50°S–50°N)	(Huffman et al., 2007)
CHIRPSv2	0.05°	Daily, pentadal, monthly	1981-present	Global (50°S–50°N)	(Funk et al., 2015)
CMORPHv1	0.25°	30 min, 3 hourly, daily	1998-present	Global (60°S–60°N)	(Xie et al., 2017)
PERSIANN-CDR	0.25°	Daily, monthly, yearly	1983–2015	Global (60°S–60°N)	(Ashouri et al., 2015)
MSWEPv2	0.10°	3 hourly, daily	1979–2016	Global	(Beck et al., 2017a, 2017b)

gov.br/default.asp). Finally, for Magdalena River Basin in Colombia, a set of 5864 rain gauge stations were obtained. The time series were provided by the Hydrology, Meteorology and Environmental Studies Institute (IDEAM), which is a public institution responsible for giving technical and scientific support to the Environmental National System.

Rain gauges provide generally reliable point measurements of precipitation and are usually subject to a low measurement error. However, Rain gauge estimates may be affected by instrumental error (Delahaye et al., 2015). Therefore, from the raw datasets, 13, 64, and 124 rain gauge stations were selected for Imperial, Paraiba do Sul, and Magdalena basins, respectively. The selected rain gauge stations have < 2% of missing values during the studied period and showed consistency when analyzed with the double-mass method (Weiss and Wilson, 1953). The double-mass method looks for abnormal differences comparing each station with its neighbors assuming homogeneity between stations.

3.2. Satellite rainfall estimates

In this section, we describe the state-of-the-art SREs used in this work (Table 1; TRMM 3B42v7, TRMM 3B42RT, CHIRPSv2, CMORPHv1, PERSIANN-CDR, and MSWEPv2). The spatial resolution of the products is 0.25 except for CHIRPSv2 (0.05) and MSWEPv2 (0.1), and all SREs have a daily temporal resolution.

The Tropical Rainfall Measuring Mission (TRMM; Huffman et al., 2007) is a joint mission of NASA and JAXA developed to provide a detailed dataset of rainfall distribution over tropical and subtropical regions (Schuster et al., 2011). This mission is intended to provide estimates of quasi-global rainfall at relatively high spatial resolution (0.25°) in both, real-time (TRMM 3B42RT) and post-real time (TRMM 3B42 version 7, hereafter defined as TRMM 3B42v7) which is calibrated with monthly rain gauge data from the Global Precipitation Climatology Center (GPCC) (Huffman et al., 2007). This dataset combines infrared data (IR) from geosynchronous earth orbit (GEO) and passive microwave data (PMW) from LEO satellites. After > 17 years of data collection, the instruments on TRMM were turned off the 8 of April of 2015 but the dataset will be produced until 2018.

The Climate Hazards Group InfraRed Precipitation with Station data version 2 (CHIRPSv2; Funk et al., 2015) is a quasi-global (50 north and south) daily, pentadal, and monthly SRE with a spatial resolution of 0.05°. CHIRPSv2 is designed for monitoring agricultural drought and global environmental changes over land. This dataset uses the Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis version 7 (TRMM 3B42v7) in order to calibrate global Cold Cloud Duration (CCD) rainfall estimates (Funk et al., 2015). CHIRPSv2 uses also a station data merging approach in order to reduce the bias using public and private data archives at monthly scale. According to Funk et al. (2015), CHIRPS uses CHPclim dataset, which is a global 0.05° monthly precipitation climatology, the satellite-only Climate Hazards group Infrared Precipitation (CHIRP), and ground observations from different public and private archives.

The NOAA Climate Prediction Center (CPC) MORPHing technique (CMORPHv1; Joyce et al., 2004; Xie et al., 2017) is a quasi-global (60° north and south) SRE with 0.25° spatial and 3-hourly temporal resolution that relies on infrared (IR) and passive microwave (PMW) information. CMORPHv1 uses motion vectors derived from 3 hourly IR satellite imagery in order to propagate the relatively high-quality precipitation estimates derived from PMW data (Joyce et al., 2004). This dataset is bias corrected through a probability density function (PDF) matching against the CPC daily gauge analysis and the Global Precipitation Climatology Project (GPCP) over land and ocean, respectively.

The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Climate Data Record (PERSIANN-CDR; Sorooshian et al., 2000; Ashouri et al., 2015) is a quasi-global (60° north and south) daily product with a spatial resolution of 0.25° providing 30 years of daily precipitation estimates. PERSIANN-CDR estimates precipitation using IR and its accuracy is improved using rainfall estimates from a PMW sensor. It is adjusted using the monthly GPCC product to maintain consistency and to reduce the bias while preserving the high spatial resolution (Ceccherini et al., 2015; Guo et al., 2015).

The Multi-Source Weighted-Ensemble Precipitation version 2 (MSWEPv2; Beck et al., 2017a, 2017b) is a global precipitation dataset with 3-hourly temporal and 0.1° spatial resolution. MSWEPv2 uses a wide range of data sources, such as rain gauge stations, satellite information, and atmospheric reanalysis models to obtain precipitation estimates and it was developed for hydrological modeling (Beck et al., 2017a). MSWEPv2 uses the Global Historical Climatology Network-Daily (GHCN-D), the Global Summary of the Day (GSOD) database, the Latin American Climate Assesment & Dataset (LACA&D) database, the Chilean Climate Data Library, and national databases for Mexico, Brazil, Peru, and Iran to perform a daily bias correction.

The detailed description of the algorithms used for each SRE to derive rainfall estimates can be found in the specific literature of each product.

4. Methodology

A point-to-pixel analysis was applied to the studied areas to compare the time series of the selected rain gauge stations against the corresponding pixel value of the selected SREs at daily, monthly, and seasonal timescales. This methodology has been widely used for SREs validation (Reis, Renn, and Lopes, 2017; Thiemig et al., 2012; Zambrano-Bigiarini et al., 2017). The implicit assumption of this methodology is that the rain gauge stations are representative observations of the respective pixels of the poducts.

This comparison is not completely fair due that for some SREs, ground data are used to correct the bias of the precipitation estimations. This is the case of TRMM 3Bv7, CMORPH, and PERSIANN-CDR which use the GPCC dataset; CHIRPSv2 and MSWEPv2 which use the Global Historical Climate Network (CHCN) and the Global Surface Summary of the Day (GSOD). MSWEP also uses the Latin American Climate Assesment & Dataset (LACA&D) and national databases from Mexico, Brazil, Peru, and Iran. This bias corrected datasets are expected to present a better performance than the SREs that do not use rain gauge stations (i.e. TRMM 3B42RT).

In Fig. 4, the number of rain gauges in Imperial, Paraiba do Sul, and Magdalena basins used in the GPCC dataset (Peterson and Vose, 1997) can be observed. For the studied period on average, the GPCC uses

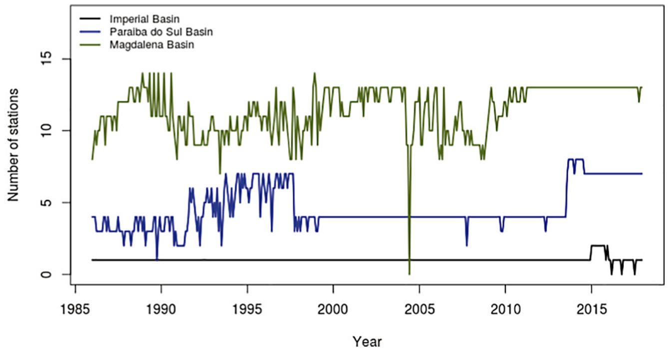

Fig. 4. Number of rain gauges used in GPCC dataset for the period 1985–2017 over the different studied areas.

around 1, 6, and 12 ground stations for Imperial, Paraiba do Sul and Magdalena basins, respectively. Also, the CHCN uses a similar amount of rain gauge stations as the GPCC dataset over the 3 different study areas (Menne et al., 2012). As the GPCC does not provide the information of the specific used stations they could not be removed. However, this number of stations are clearly not enough to capture the spatial variability of rainfall over these basins. This consequently, increases the estimation errors of the final products. On the other hand, the SREs that use rain gauge stations from the studied areas are prone to present a better performance compared to the datasets that do not.

The spatial resolution of all SREs except CHIRPSv2 and MSWEPv2 is 0.25°. Therefore, CHIRPSv2 and MSWEPv2 were upscaled from 0.05° and 0.1°, respectively to a new spatial resolution of 0.25° by using bilinear interpolation to enable consistent point-to-pixel comparisons among SRE products. However, in order to determine the effect of the upscaling process in the SREs performance evaluation, we used two CHIRPSv2 and MSWEPv2 datasets i) One with its original spatial resolution (CHIRPSv2, MSWEPv2) and the other ii) with the datasets upscaled to 0.25° (hereafter defined as CHIRPSv2 upscaled and MSWEPv2 upscaled).

Six different indices of performance (three continuous and three categorical) were applied to the data over the different regions. Continuous indices are described in Appendix A, these indices are the modified Kling-Gupta Efficiency (KGE′), the Root Mean Square Error (RMSE), and the percent bias (PBIAS). The KGE′ (Eq. (A1)) is an index used to compare observed data with estimations. It decomposes the total performance or SREs into three different terms with the same weight; the linear correlation (r) measuring, in this case, the temporal rainfall dynamics, the bias ratio (β) used to measure the overestimation or underestimation compared to the ground observations, and the variability ratio (γ) which is a relative measure of the dispersion (Gupta et al., 2009; Kling et al., 2012). The optimum value for KGE′ is the unity. For KGE′, the linear correlation (r, Eq. (A2)) presents its optimum value at the unity (perfect correlation) being the minimum value -1.0 (perfect negative correlation), the 0 indicates absence of correlation. The bias term (β, Eq. (A3)) measures the average tendency of the SRE values to overestimate ($\beta > 1$) or underestimate ($\beta < 1$) the observed values. The variability ratio (γ, Eq. (A4)) evaluates the dispersion of the satellite estimates compared to the observed data and also presents its optimum value at the unity.

Besides KGE′ evaluation, the RMSE (Eq. (A5)) and PBIAS (Eq. (A6)) were computed and compared to the results obtained with KGE′. The RMSE is widely used in the performance evaluation of SREs and it was included in this study to assess whether is a useful measure of performance or not. The PBIAS measures the average tendency of the simulated values to be larger or smaller than their observed ones. The optimal value of PBIAS index is 0, with low values indicating an accurate simulation, while positive values indicating overestimation and negative values indicating underestimation of the SREs.

Table 2
Classification of rainfall events based on daily intensity (i) according to Zambrano-Bigiarini et al. (2017).

Rainfall event	Intensity (i), [mm d^{-1}]
No rain	[0, 1)
Light rain	[1, 5)
Moderate rain	[5, 20)
Heavy rain	[20, 40)
Violent rain	≥ 40

The SREs performance evaluation was implemented on a daily, monthly, and seasonal timescales. In an inter-comparison between observed and satellite rainfall data, the usage of the KGE′ makes sense because it offers a way of evaluating the different components of performance separately. However, the KGE′ does not allow to identify the performance of a given SRE for different rainfall intensities. For this reason, diverse categorical indices were also used to complement the evaluation of the satellite products.

Moreover, three categorical indices (Appendix B) were applied for different rainfall intensities (Table 2) as recommended by Zambrano-Bigiarini et al. (2017); the probability of detection (POD, Eq. (B1)), the frequency bias (fBias, Eq. (B2)), and the false alarm ratio (FAR, Eq. (B3)). The POD calculates the fraction of the observed rainfall events that are correctly detected by the satellite products ranging from 0 (no detection) to 1 (detection of all events). The classification of rainfall events is based on its intensity. Table 2 shows the classification of the events based on daily intensity obtained from Zambrano-Bigiarini et al. (2017). The false alarm ratio (FAR) measures the fraction of the events that are incorrectly identified by the SREs. This index ranges from 0 (no events are incorrectly identified), to 1 (all events are incorrectly identified). Also, The frequency bias (fBias), uses the same classification of rainfall intensities observed in Table 2. This index is used to compare the number of events identified by the satellite product to the number of events that actually occurred at the corresponding rain gauge. The optimal value of the fBias is 1.0, indicating no bias, fBias > 1 indicates an overestimation of the SRE occurrences, and fBias < 1 points to an underestimation.

The aforementioned methodology was applied for the studied areas using the R environment 3.3.1 (Team, 2011) and the raster (Hijmans et al., 2017), hydroGOF (Zambrano-Bigiarini, 2017a), and hydroTSM (Zambrano-Bigiarini, 2017b) R packages.

5. Results

5.1. Spatio-temporal performance of SREs

We computed spatial maps to represent the spatial variation of SRE performance for each study area at a daily, monthly and seasonal temporal scales. The seasonal temporal scale is divided as follows: December–January–February (DJF), March–April–May (MAM), June–July–August (JJA), and September–October–November (SON). Six different indices were applied to identify the best performing SRE for each region (KGE′, RMSE, PBIAS, POD, FAR, and fBias).

5.1.1. Performance of SREs at daily scale

The evaluation with KGE′ at daily scale showed that for Imperial Basin, MSWEPv2 performed the best at daily scale followed by PERSIANN-CDR, MSWEPv2 upscaled, and CHIRPSv2 upscaled [0.4 < KGE′ < 0.5]. TRMM 3B47RT had the lowest performance followed by CMORPHv1 [0.1 < KGE′ < 0.2]. All SREs performed the similar at the different elevations. As observed in Fig. 5, when the performance of the selected SREs was evaluated with RMSE, PERSIANN-CDR presented the best performance followed closely by TRMM 3B42RT, with values of 7.66 mm and 7.98 mm, respectively. MSWEPv2, which was the best performing SRE for KGE′, was ranked third when

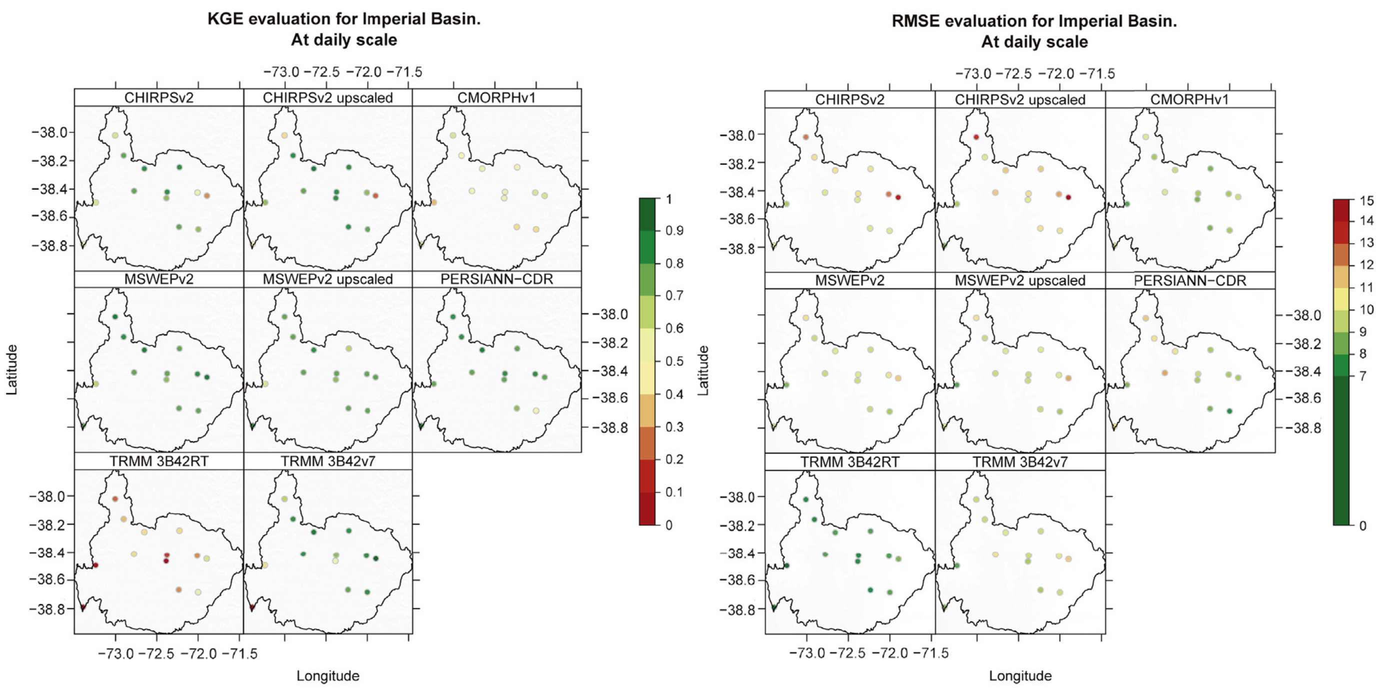

Fig. 5. Daily comparison between KGE′ (left) and RMSE (right) indices for analyzed SREs over Imperial River Basin at daily scale.

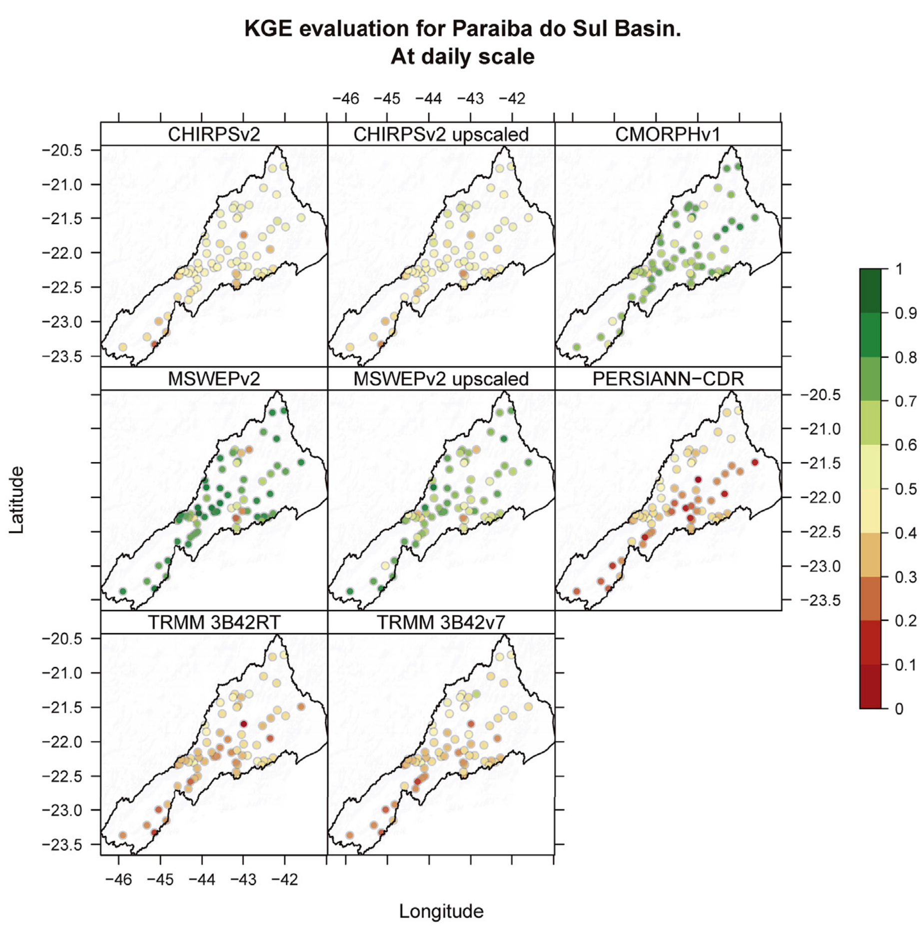

Fig. 6. KGE′ evaluation of the SREs used in this study over Paraiba do Sul at daily scale.

RMSE was used.

For Paraiba do Sul Basin (Fig. 6), MSWEPv2 was selected as the best performing dataset when KGE′ was used, followed by CMORPHv1 and MSWEPv2 upscaled [0.55 < KGE′ < 0.65]. CHIRPSv2 and CHIRPSv2 upscaled obtained a KGE′ value of 0.37. The lowest evaluations were obtained by PERSIANN-CDR, TRMM 3B42v7, and TRMM 3B42RT with KGE′ values lower than 0.3. CMORPHv1 obtained the best evaluation with the RMSE followed closely by MSWEPv2, which in agreement with the KGE′ result. However, for this index CHIRPSv2 performed worse than PERSIANN-CDR which was the worst evaluated SRE when the KGE′ was used. The performance of the SREs was slightly lower in elevated areas.

Finally, for Magdalena Basin CHIRPSv2 presented the best performance when evaluated with KGE′, followed by CMORPHv1 and CHIRPSv2 upscaled [0.25 < KGE′ < 0.30]. The daily performance of all SREs was lower in Magdalena Basin than over the other regions. PERSIANN-CDR and MSWEPv2 presented the lowest performance [0.0 < KGE′ < 0.1] All SREs presented low KGE′ values in the lower area which corresponds to the northern part of the basin.

5.1.2. Performance of SREs at the monthly scale

The KGE′ evaluation for Imperial Basin at monthly scale showed that CHIRPSv2 is the best performing SRE, followed by TRMM 3B42v7 [0.75 < KGE′ < 0.80]. Almost all the SREs scored higher than 0.7 with the exception of TRMM 3B42RT with a KGE′ value of 0.3. For RMSE, MSWEPv2 and MSWEPv2 upscaled were the best-evaluated SREs, presenting the lowest values (38.90 mm and 42.17 mm, respectively). MSWEPv2 and MSWEPv2 upscaled occupied the fifth and sixth position of performance when evaluated with the KGE′.

Paraiba do Sul Basin presented the highest KGE′ values compared to Imperial and Paraiba do Sul basins. The best performing SRE in this area was MSWEPv2 upscaled, followed by CHIRPSv2 with KGE′ values of 0.83 and 0.82 respectively. MSWEPv2, CHIRPSv2 upscaled, and CMORPHv1 presented higher KGE′ values than 0.80. For both indices (KGE′ and RMSE), TRMM 3B42RT, TRMM 3B42v7, and PERSIANN-CDR presented the lowest performance.

The SREs presented KGE′ values ranging from 0.3 to 0.7 when evaluated over Magdalena Basin (Fig. 7). CHIRPSv2 and CHIRPSv2 upscaled presented the highest KGE′ values (0.70 and 0.66, respectively). PERSIANN-CDR showed the lowest performance for both indices. However, MSWEPv2, MSWEPv2 upscaled and TRMM 3B42RT presented a higher performance when evaluated with the KGE′ than with the RMSE.

5.1.3. Performance of SREs at seasonal scale

The KGE′ evaluation for Imperial Basin at seasonal scale CMORPHv1 and TRMM 3B42v7 presented a good performance for DJF and SON, while TRMM 3B42v7 and MSWEPv2 performed better for MAM and PERSIANN-CDR for JJA. TRMM 3B42RT presented the lowest performance over Imperial (for KGE′ and RMSE) in all seasons. All SREs had a lower performance in the rainy season JJA (winter) over Imperial basin.

Paraiba do Sul presented the highest values of KGE′ for the selected SREs. MSWEPv2 and MSWEPv2 upscaled performed the best for all seasons and for the KGE′ and RMSE indices, followed by CHIRPSv2 and CMORPHv1. The SREs with the lowest performance for all seasons were PERSIANN-CDR and TRMM 3B42RT. Also, all the satellite datasets performed similarly over the four different seasons.

For Magdalena Basin CHIRPSv2, CHIRPSv2 upscaled, and TRMM 3B42RT performed the best for DJF and SON for both indices (KGE′ and RMSE). These products also performed the best for MAM when the KGE′ was used, while CHIRPSv2, CHIRPSv2 upscaled, and TRMM 3B42v7 performed the best when the RMSE was used. For SON, CHIRPSv2 and CHIRPSv2 upscaled showed a good performance for both indices. The SREs presented a lower performance for MAM and JJA (spring and summer, respectively) over this basin.

5.2. Components of the KGE′

The total performance of the KGE′ (as seen in Appendix A) can be decomposed into linear correlation (r), bias (β), and a variability (γ) term. It is a useful index because the correlation component (r) is able

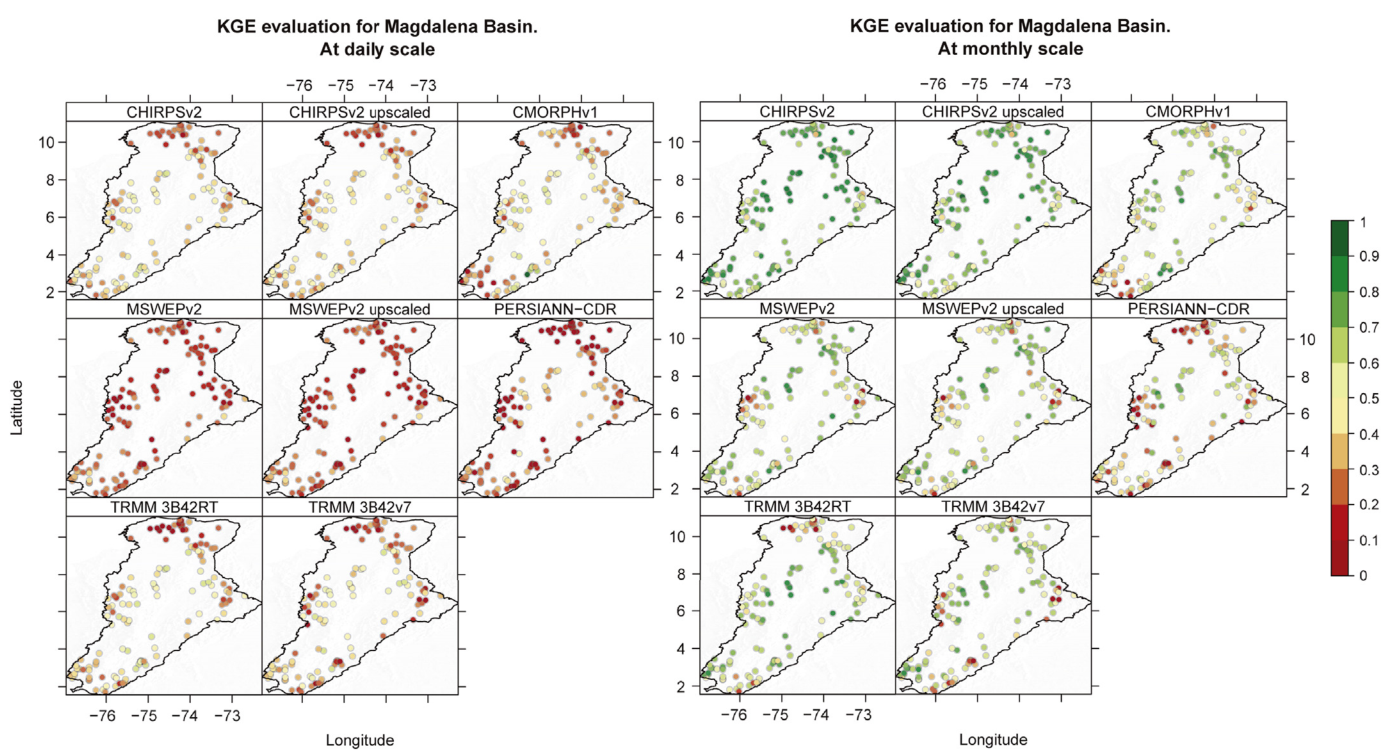

Fig. 7. KGE′ comparison between the used SREs over Magdalena River Basin at daily (left) and monthly (right) temporal scales.

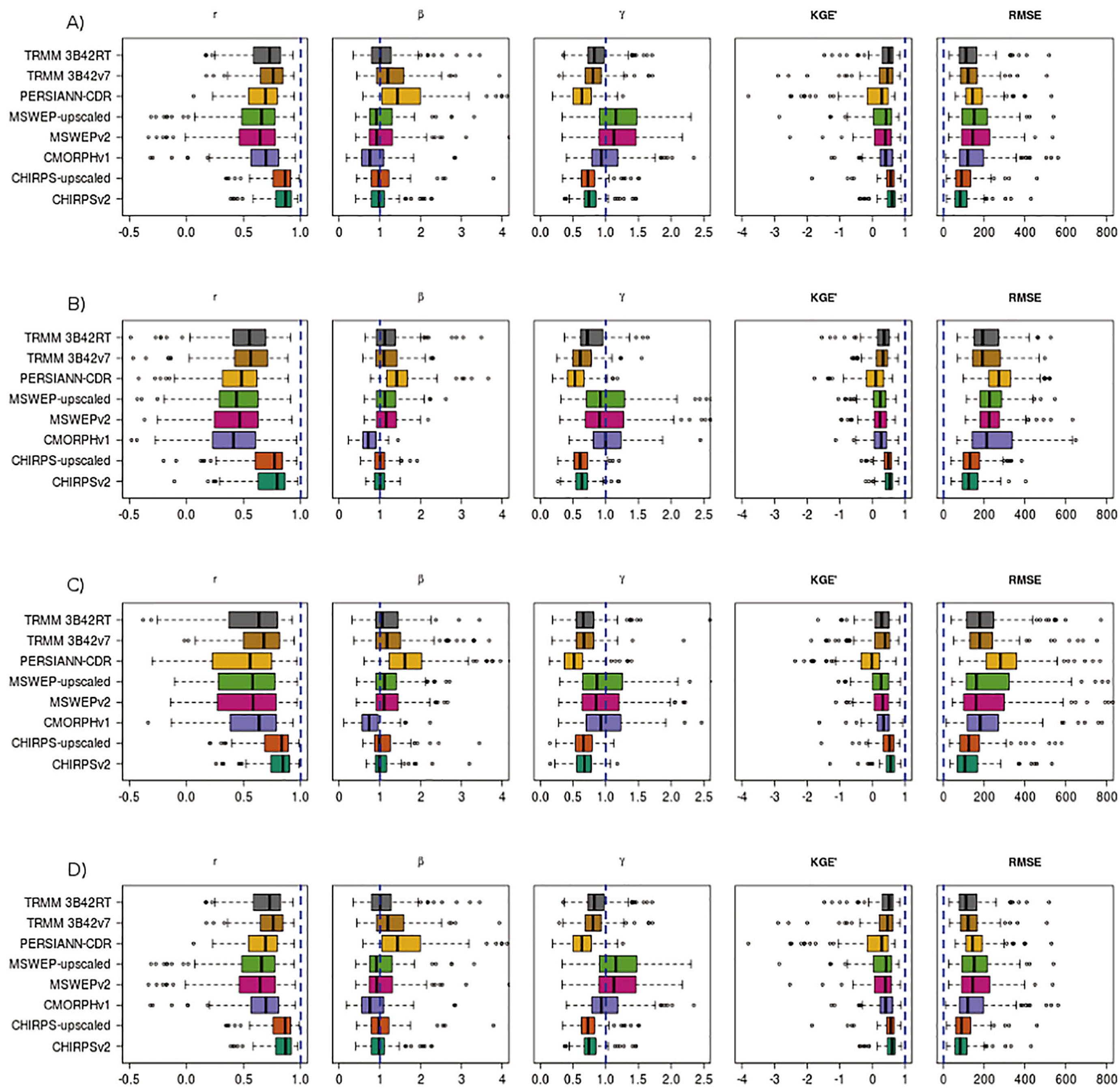

Fig. 8. Boxplots of the different components of KGE′ with the correspondent KGE′ and RMSE values for the Magdalena Basin at seasonal scale, where the seasons are divided as follows: a) December–January–February (DJF), b) March–April–May (MAM), c) June–July–August (JJA), and d) September–October–November (SON). The blue lines represent the optimal value for each parameter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to evaluate the temporal dynamics of precipitation, while β and γ are able to evaluate the volume and variability of rainfall, respectively.

In the seasonal analysis for Imperial Basin, MSWEPv2 presents the highest values of r for all seasons. CHIRPSv2 and CHIRPSv2 upscaled performed good for β for all the seasons and also CMORPHv1 for DJF and SON. CMORPHv1 presented the best performance for γ in DJF and MAM while CHIRPSv2 and CHIRPSv2 upscaled for JJA and SON. Also, the correlation component showed the lowest values for JJA which corresponds to the rainy season.

For Paraiba do Sul MSWEPv2 performed the best for r and γ for all the seasons and CHIRPSv2, CHIRPSv2 upscaled, CMORPHv1 and also, MSWEPv2 and MSWEPv2 upscaled presented a good performance for the β component. The components performed similarly for the different seasons over this basin.

The RMSE, KGE′, and its components are shown in Fig. 8 for Magdalena Basin at seasonal time scale. CHIRPSv2 and CHIRPSv2 upscaled perform the best for all seasons, followed by TRMM 3B42RT for DJF, MAM, and SON, and TRMM 3B42v7 for JJA. CHIRPSv2 and CHIRPSv2 upscaled present higher values of r and β for all seasons and MSWEPv2, MSWEPv2 upscaled, and CMORPHv1 present values near 1.0 fo the γ component for all seasons meaning that these SREs are able to capture the distribution of precipitation over the Magdalena Basin. However, MSWEPv2 present a low performance over the Magdalena River Basin. The correlation component shows the lowest values in JJA, while in

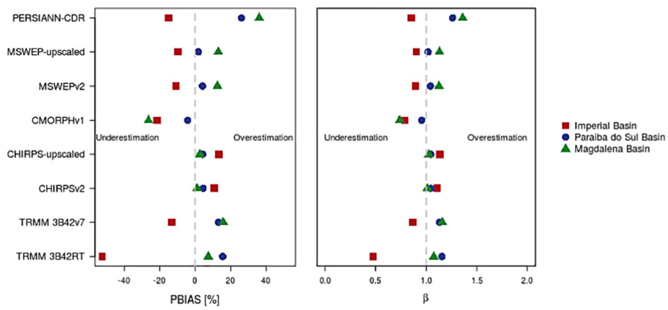

Fig. 9. Comparison of the percent bias (PBIAS; left) and the bias component of KGE′ (β; right) for the studied areas.

DJF and SON, β performed worse. The γ component show the lowest values in MAM and JJA.

In Fig. 9 the PBIAS is compared to the bias (β) component of KGE′ for the three study areas. Although the PBIAS is represented in percentage and β is represented as the ratio between the correspondent SRE mean and the observed mean (Eq. A3), there is a total agreement between PBIAS and β. For Paraiba do Sul and Magdalena basins, almost all SREs overestimate the rainfall compared to observations with the exception of CMORPHv1 which underestimates. On the other hand, for Imperial Basin, the SREs tend to underestimate rainfall except for CHIRPSv2 and CHIRPSv2 upscaled.

5.3. SREs performance for categorical indices

Fig. 10 shows the probability of detection (POD) for five different precipitation intensities (described in Table 2). All SREs show a relatively high POD in no rain events ([0, 1) mm d^{-1}) with values higher than 0.6, except in the case of Magdalena Basin for PERSIANN-CDR. All SREs performed better at capturing moderate rainfall intensities (between 5 and 20 mm d^{-1}) in all cases. For Magdalena Basin, all SREs present almost no skill capturing high intensities. For Paraiba do Sul Basin, MSWEPv2 performs the best for all rainfall intensities. Also, For Imperial River Basin is observed that CHIRPSv2, CHIRPSv2 upscaled, MSWEPv2, and MSWEPv2 upscaled perform relatively better at capturing high rainfall intensities.

For light rainfall intensities ([1, 5) mm d^{-1}) almost all the SREs presented a decrease in the POD, with values lower than 0.4 in all study areas (except for MSWEPv2 and CMORPHv1 in Paraiba do Sul) showing that light rain events are difficult to capture for the current satellite products. For Imperial and Magdalena basins there was not a single SRE which performed the best for all intensities. In the case of Paraiba do Sul, MSWEPv2 performed considerably better. For no rain events, all products except PERSIANN-CDR (in Magdalena Basin) presented a high POD. For rain intensities higher or equal to 1, PERSIANN-CDR, CHIRPSv2, and MSWEPv2 products presented a better performance.

Fig. 11 shows the false alarm ratio (FAR) for the precipitation intensities mentioned in Table 2. These results show consistency with the POD. The no-rain events show low values over the different basins,

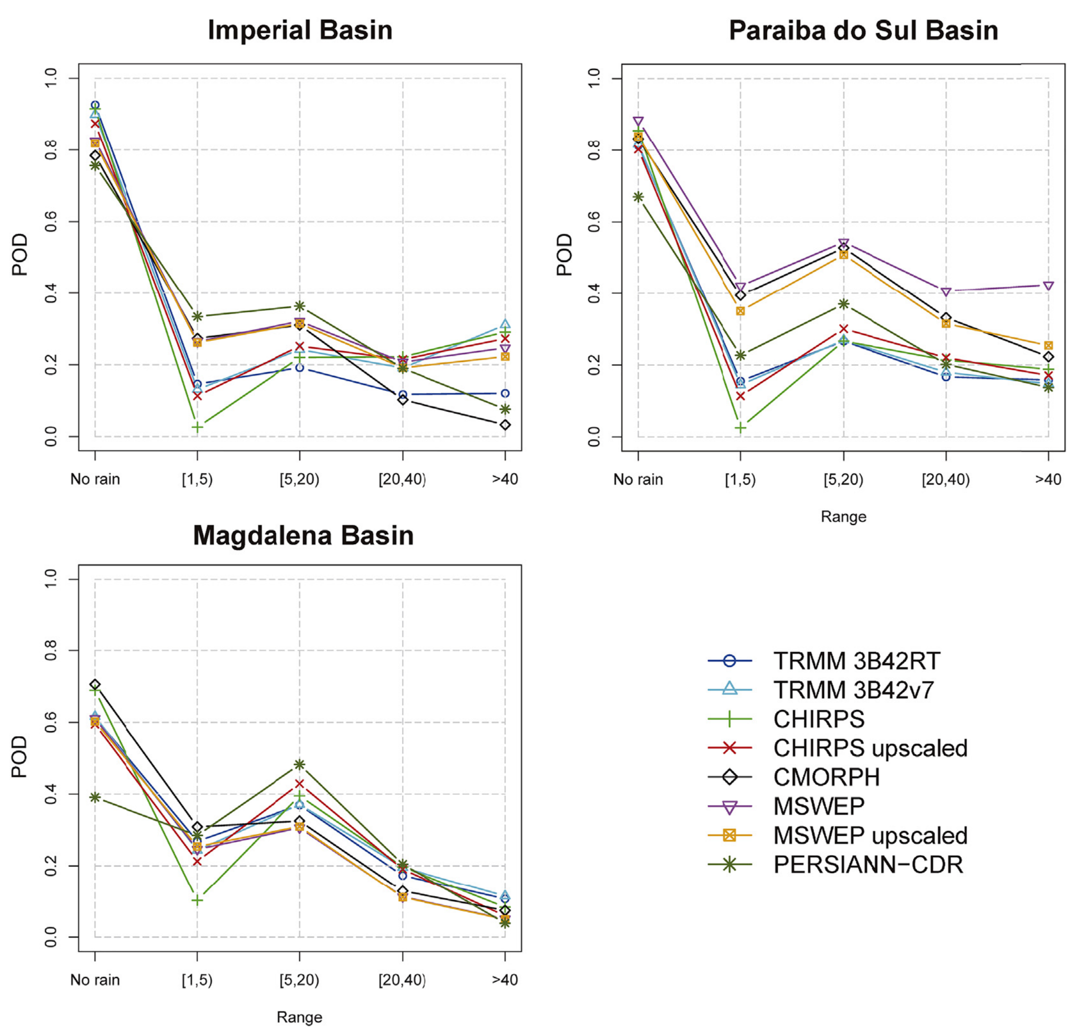

Fig. 10. Mean values of the categorical index of performance probability of detection (POD) and the five classes of rainfall intensity defined in Table 2 for the different studied areas.

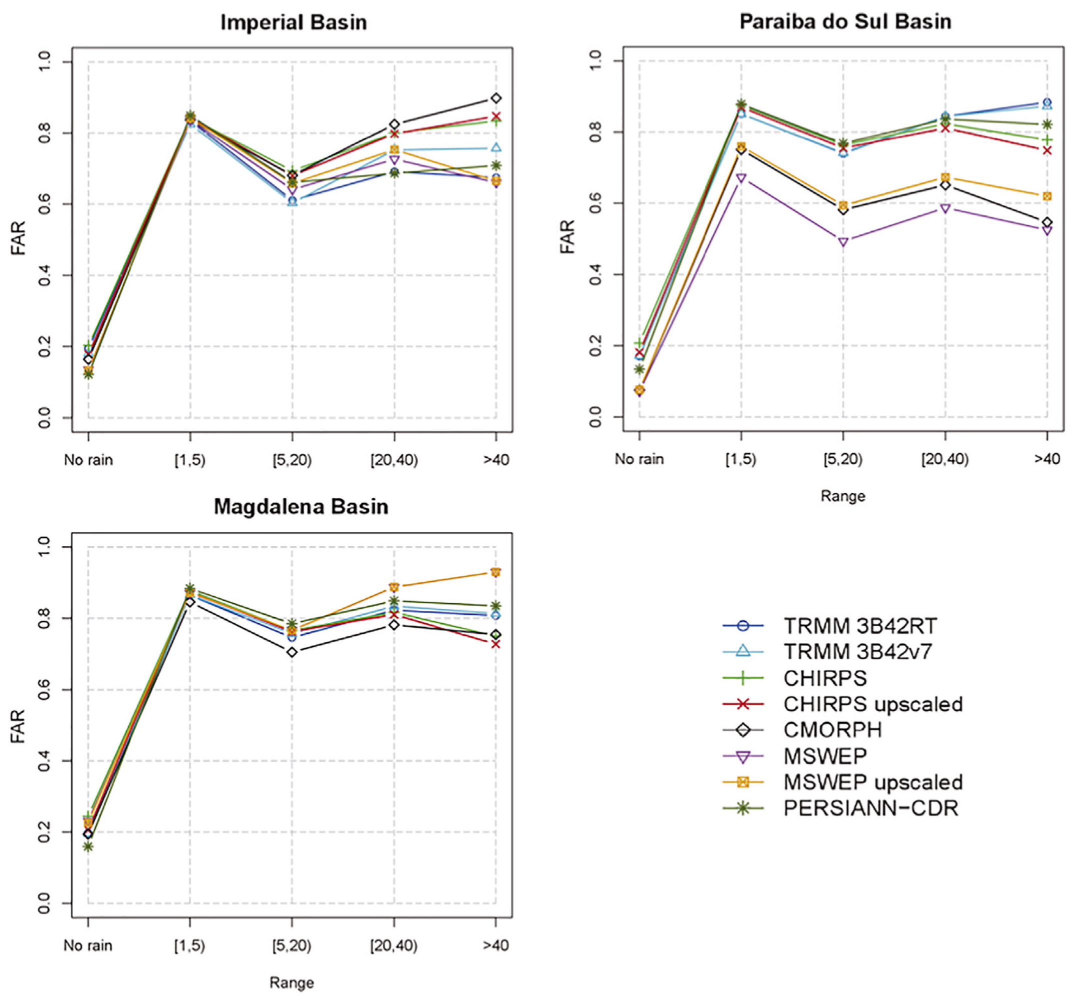

Fig. 11. Mean values of the categorical index of performance false alarm ratio (FAR) and the five classes of rainfall intensity defined in Table 2 for the different studied areas.

while the events with precipitation, present values between 0.5 and 0.95 in all cases. This shows that for the analyzed SREs, the days with precipitation are more difficult to identify. In particular, for the Imperial Basin MSWEPv2, MSWEPv2 upscaled, and TRMM 3B42RT showed low FAR values, while in Magdalena Basin CMORPHv1, CHIRPSv2, and CHIRPSv2 upscaled performed better in high rain intensities. For Paraiba do Sul Basin, MSWEPv2 (as observed in POD) was the SRE which presented the best performance in all the rain intensities.

The frequency Bias (fBias; Fig. 12) shows an excellent agreement to POD index in no-rain events with values of fBias ˜ 1. For Imperial and Paraiba do Sul basins, all products except CHIRPSv2 and CHIRPSv2 upscaled overestimated the light rain events, while for Magdalena Basin, all products showed a high overestimation. In Magdalena, almost all products presented an underestimation for violent rain events (equal or higher than 40 mm d^{-1}), with MSWEPv2 and MSWEPv2 upscaled as an exception. In the case of Paraiba do Sul, TRMM 3B42RT, and TRMM 3B42v7 overestimated and for Imperial, CMORPHv1, PERSIANN-CDR, TRMM 3B42RT, and MSWEPv2 underestimated them. The violent rainfall events presented relatively low fBias in all regions, and moderate events ([5, 20) mm d^{-1}) were overestimated for Paraiba do Sul and Magdalena basins, while for Imperial Basin only PERSIANN-CDR showed a slight overestimation. It is also observed that Paraiba do Sul present the lowest fBias values followed closely by the Chilean basin. Magdalena presented the highest fBias values for light rainfall intensities.

5.4. Upscaling comparison for CHIRPSv2 and MSWEPv2

Upscaled versions of CHIRPSv2 and MSWEPv2 were computed from 0.5° and 0.1°, respectively. This, in order to ensure a consistent point-to-pixel comparison with the other SREs at 0.25° spatial resolution. However, the datasets were also analyzed using their original high spatial resolution to evaluate if the upscaling process can interfere with the evaluation of performance.

As observed in Fig. 13, the upscaling procedure increased the KGE′ values of CHIRPSv2 for Imperial Basin in Chile and slightly decreased the values of MSWEPv2 at daily scale. For this basin at the monthly scale, CHIRPSv2 upscaled performed worse than the regular CHIRPS2 and MSWEPv2 upscaled improved when compared to MSWEPv2. Also, CHIRPSv2 upscaled presented a decrease when compared to CHIRPSv2 at monthly scale over Magdalena Basin. However, there were not important changes at daily scale over Magdalena Basin for both products and for MSWEPv2 at monthly scale. For Paraiba do Sul, CHIRPSv2 products presented a similar behavior at both time scales while MSWEPv2 upscaled had a lower performance compared to the original MSWEPv2.

Fig. 7 shows that CHIRPSv2 upscaled presents a higher performance in the probability of detection (POD) over all the study areas and almost for all rainfall intensities, except for violent rain events (> 40 mm). MSWEPv2 performed better than MSWEPv2 upscaled over Paraiba do Sul and Imperial basins, while no differences were observed in Magdalena Basin.

Fig. 12 shows the frequency bias (fBias) for the three study areas. In

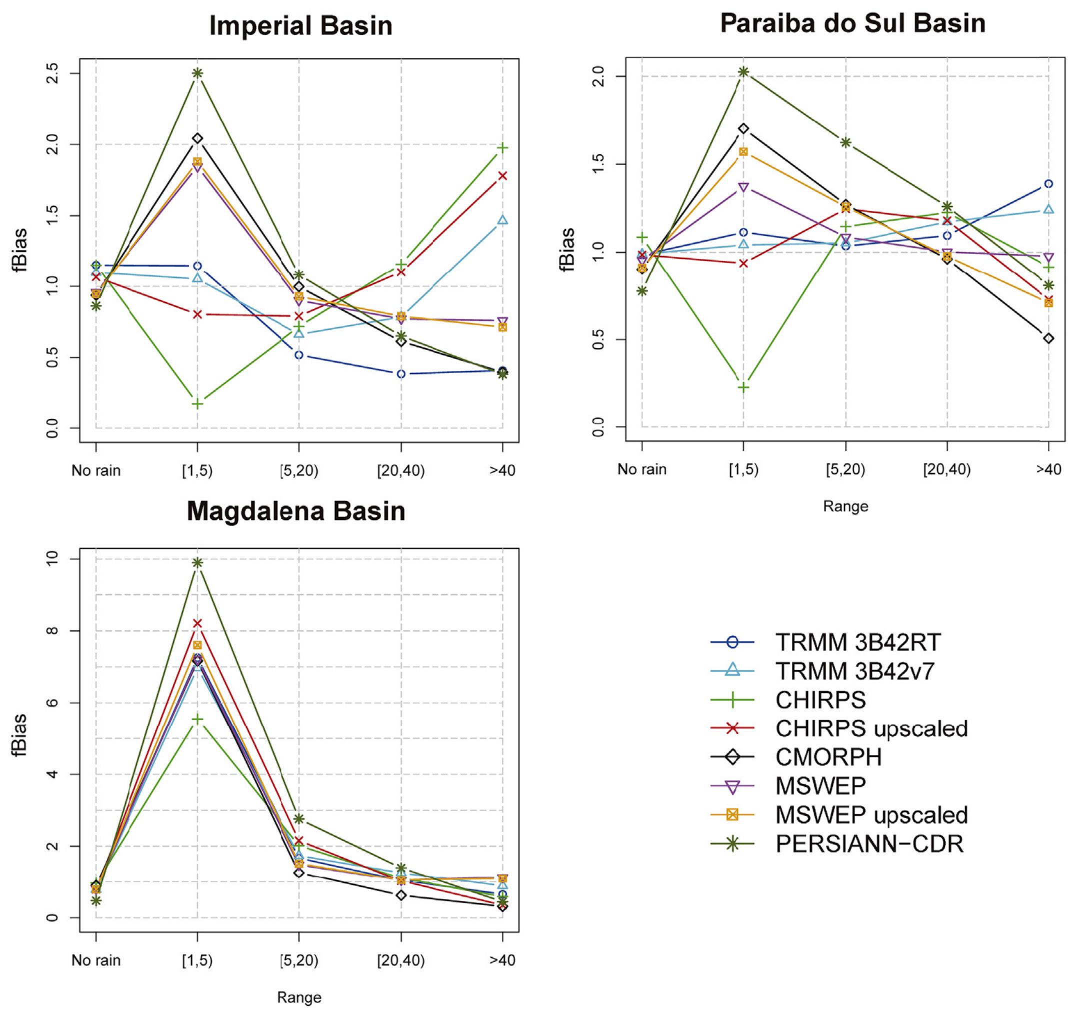

Fig. 12. Mean values of the categorical index of performance frequency bias (fBias) and the five classes of rainfall intensity defined in Table 2 for the different studied areas.

Imperial and Paraiba do Sul basins, CHIRPSv2 upscaled performed better than CHIRPSv2, while the opposite was observed over Magdalena Basin. MSWEPv2 and MSWEPv2 upscaled showed similar behavior over Imperial and Magdalena basins while in Paraiba do Sul MSWEPv2 upscaled obtained higher fBias values.

6. Discussion

The aim of this work is to evaluate for the very first time, the spatio-temporal performance of six state-of-the-art SREs (TRMM 3B42RT, TRMM 3B42v7, CMORPHv1, CHIRPSv2, PERSIANN-CDR, and MSWEPv2) over different areas in Latin-America using 201 rain gauge stations in total, and to assess if the upscaling procedure used to enable a consistent point-to-pixel comparison affects the evaluation of the upscaled SRE datasets performance (only for CHIRPSv2 and for MSWEPv2).

6.1. Performance of the evaluated SREs

As expected, the SREs performed differently over each study area. For Imperial Basin, MSWEPv2 presented the highest performance at daily time scale, followed by PERSIANN-CDR, MSWEPv2 upscaled and CHIRPSv2 (see Fig. 5). When the SREs were evaluated at monthly scale, CHIRPSv2 showed the best performance, followed by TRMM 3B42v7. The seasonal evaluation showed that CMORPHv1 and TRMM 3B42v7 performed the best for DJF and SON, TRMM 3B42v7 and MSWEPv2 performed better for MAM, and PERSIAN-CDR and MSWEPv2 for JJA. On the other hand, the lowest performing SRE for all time scales was TRMM 3B42RT. These results are in agreement with the work carried out by Zambrano-Bigiarini et al. (2017) where CHIRPSv2, MSWEPv1.2, and TRMM 3B42v7 showed a good performance over Chile. However, PERSIANN-CDR over Imperial Basin presented a better performance than Central-Southern Chile evaluated in Zambrano-Bigiarini (2017a, 2017b).

For Paraiba do Sul, MSWEPv2 performed the best at daily time scale, with CMORPHv1 and MSWEPv2 upscaled showing good performance as well (see Fig. 6). MSWEPv2, CHIRPSv2, and CMORPHv1 presented the highest performance when evaluated at monthly and seasonal scales. The lowest SREs over Paraiba do Sul at daily, monthly, and seasonal timescales were PERSIANN-CDR, TRMM 3B42v7 and TRMM 3B42RT. The results are in total agreement with Melo Davi de et al. (2015) where TRMM 3B42RT and TRMM 3B32v7 showed a low performance over Brazil.

Finally, CHIRPSv2, CMORPHv1, and CHIRPSv2 upscaled presented the best performance at daily scale over Magdalena River Basin (see Fig. 7), while CHIRPSv2 and CHIRPSv2 upscaled performed the best at monthly and seasonal timescales. Dinku et al. (2010) concluded that PERSIANN and TRMM 3B42RT had a low performance over Colombia, while in this work we used PERSIANN-CDR which also presented a low performance over this area.

Our results confirm that there is no a single best performing product for all regions, and therefore, a site-specific evaluation is still

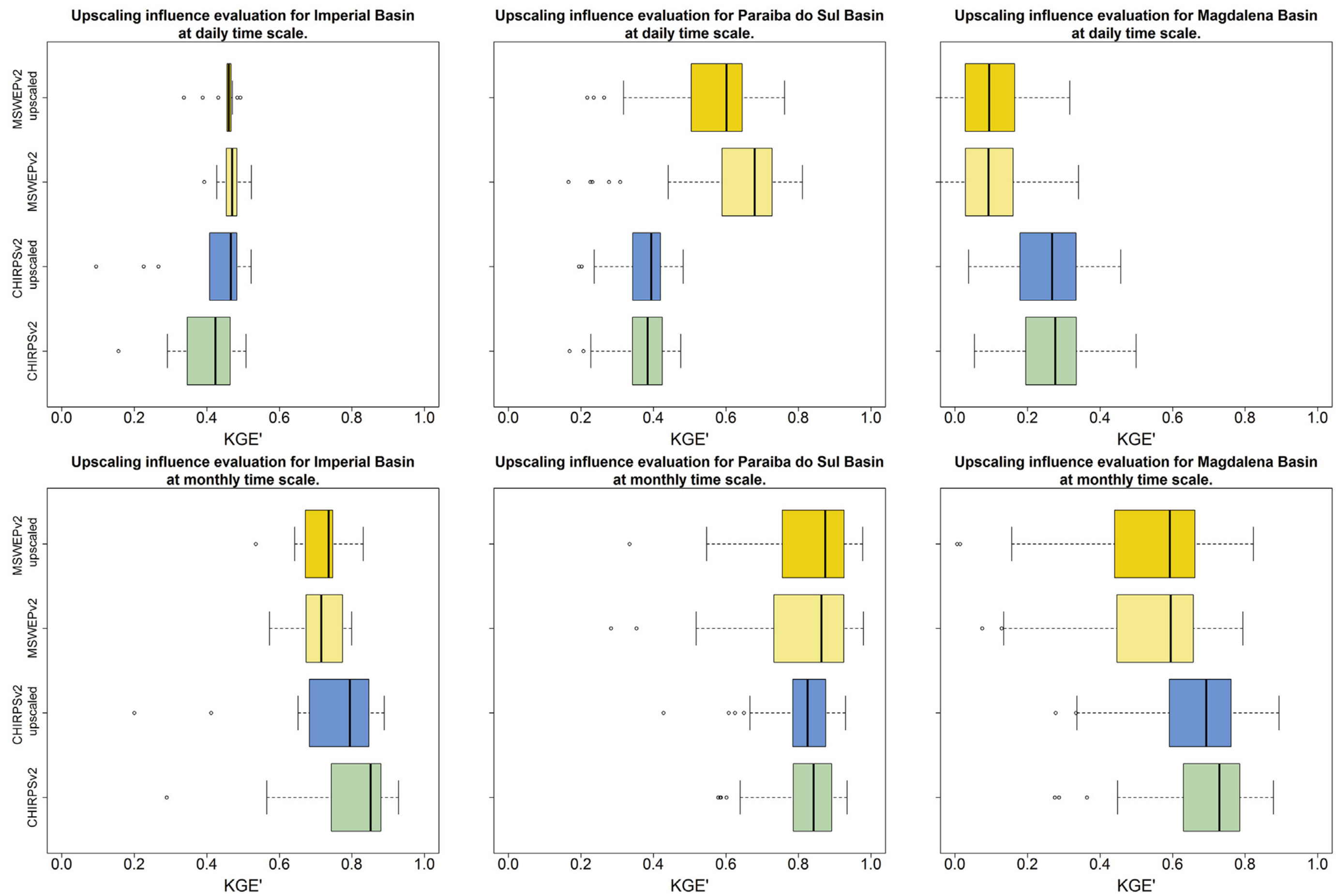

Fig. 13. Evaluation of the performance of MSWEPv2 and CHIRPSv2 before and after the upscaling procedure at daily and monthly scales.

recommended to identify the SRE that best represent the spatio-temporal characteristics of the precipitation field over a specific study area. The results show that gauge-adjusted algorithms tend to perform better in general than those with no ground-based adjustment. In some cases, an independent validation dataset is difficult to obtain. This is the case of MSWEPv2 in Paraiba do Sul Basin. MSWEPv2 performed the best over this basin at daily, monthly and seasonal scales. However, this dataset also uses the National Water Agency (ANA) dataset to calibrate the daily rainfall estimates and therefore, we expected a high performance of this dataset over this basin. The independence of the validation dataset should be analyzed before any comparison among SREs due that this may affect the results of the validation of the satellite estimates. Moreover, the performance of the SREs increase when the daily values are aggregated into monthly or seasonal values.

6.2. Does the upscaling process affects the evaluation of the performance?

To the best of our knowledge, this is the very first time that CHIRPSv2 and MSWEPv2 are evaluated at a different spatial resolution over the same study areas. For Imperial Basin, the upscaling procedure improved the performance of CHIRPSv2 and decreased the performance of MSWEPv2 at daily scale. The opposite was observed at monthly time scale where CHIRPSv2 upscaled presented a lower performance than CHIRPSv2 and MSWEPv2 upscaled performed higher than the original MSWEPv2. In the case of Paraiba do Sul, CHIRPSv2 performed better after the upscaling procedure while MSWEPv2 was not affected when evaluated at daily scale. At the monthly scale, CHIRPSv2 upscaled performed slightly better than CHIRPSv2 while MSWEPv2 upscaled had a lower performance. Finally, in the Magdalena Basin, the upscaling procedure did not have any impact on the performance at daily scale for the two SREs; however, at monthly time scale CHIRPSv2 upscaled performed worse than the regular product and no changes were observed for MSWEPv2.

Our results show that there is not a general conclusion about the impact of the upscaling procedure on the performance of a given SRE dataset, with different impacts regarding products, temporal scales, and regions (see Fig. 13). However, the topography plays an important role when upscaling an SRE. The stations with higher differences in the performance of SREs compared to observations before and after the upscaling process are located closer to elevation gradients. For this reason, if a validation in a mountainous area is required, the upscaling procedure may affect the results. On the other hand, if the topography is not rugged, an upscaling procedure can enable a fair point-to-pixel comparison among SREs.

6.3. Methodology used for SRE performance evaluation

Based on the results of Section 5.1 we consider that the Modified Kling-Gupta efficiency (KGE′) has proved to be a useful evaluation index. It decomposes the performance of the SREs into linear correlation (*r*), bias (*β*) and variability (γ) components. For this reason, it is easier to understand the source(s) of the mismatches between SREs and their corresponding ground observations. Also, the Probability of Detection (POD), the false alarm ratio (FAR), and the frequency bias (fBias) are important indices in order to evaluate the accuracy of the SREs at identifying different rainfall intensities. For all study areas, the highest performance of the evaluated SREs was obtained for No-rain events (see Figs. 10, 11, and 12). Also, for the days with rain (from 1 mm d^{-1}), the intensities between 5 and 20 mm d^{-1} performed higher over all the basins than the other rainfall intensities. For Imperial and

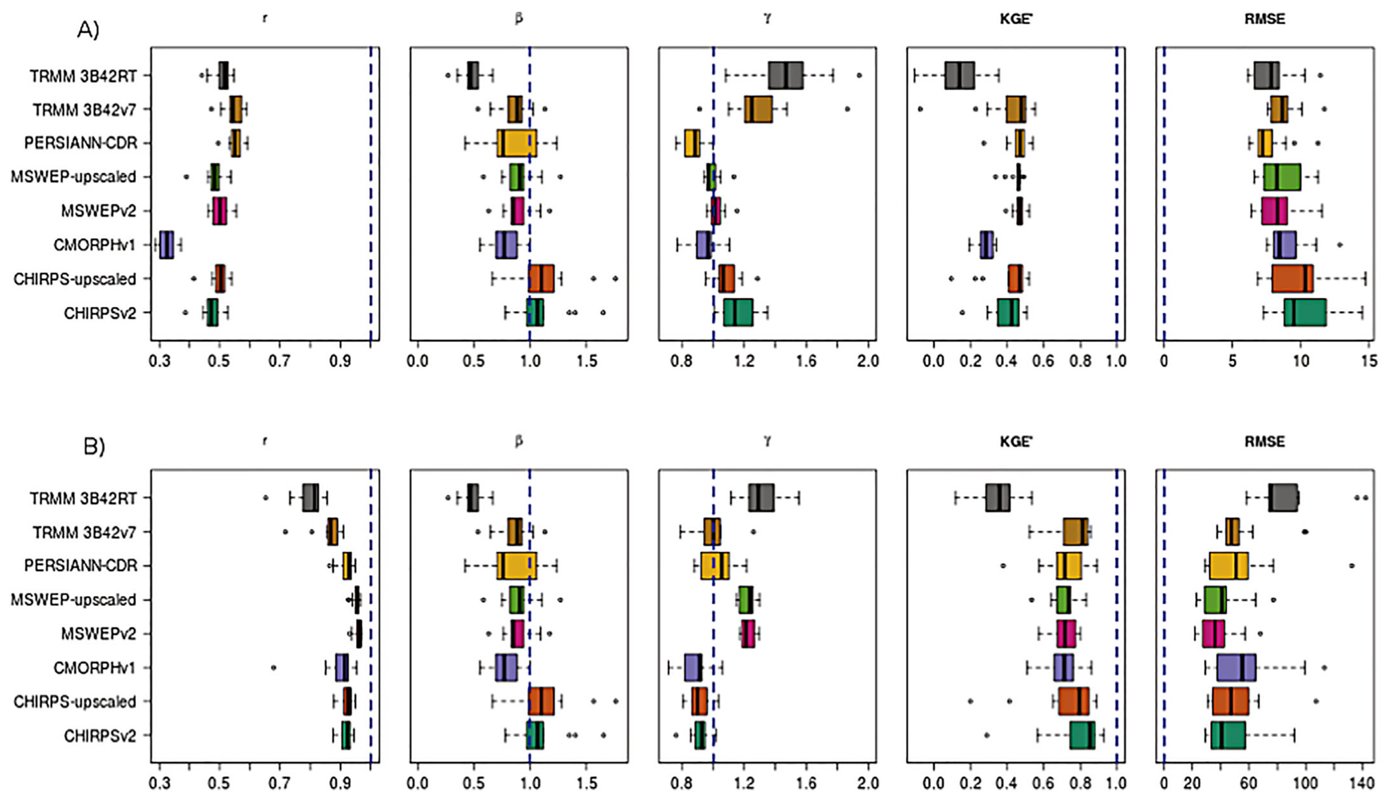

Fig. 14. Boxplots of the different components of KGE′ with the correspondent KGE′ and RMSE values for the Magdalena Basin at daily and monthly timescales, where a) is the daily evaluation and b) is the monthly evaluation.

Paraiba do Sul basins, the light rain intensities (from 1 to 5 mm d^{-1}) obtained the lowest POD, while in Magdalena Basin the lowest POD was obtained by the violent rain events (higher than 40 mm). The fBias results seen in Fig. 12, are in agreement with the POD showing the largest fBias over the three study areas for the light rain events.

RMSE results were different from KGE′ ones. Its formula gives more weight to the mismatches between the observed and satellite estimates of precipitation in two cases: i) when a low rainfall intensity is constantly not detected, and ii) during high rain events because it squares the difference of high values.

For example, a small error in a high rain event can be more important for RMSE than a total miss of several low rainfall events. As an example, the RMSE, KGE′, and its components are shown in Fig. 14 for the daily and monthly evaluation over Imperial Basin. As mentioned in the Section 5.1.1, at daily basis; MSWEPv2 performed the best followed by PERSIANN-CDR, MSWEPv2 upscaled, and CHIRPSv2 upscaled when evaluated with KGE′. For the RMSE, PERSIANN-CDR and TRMM 3B42RT had the best performance. TRMM 3B42RT presented the lowest performance for the KGE′. MSWEPv2, MSWEPv2 upscaled, and CHIRPSv2 upscaled are relatively close to 1.0 in β and γ while PESRIANN-CDR presents a better correlation and γ. In the case of TRMM 3B42RT, the β and γ components are farther but it has a relatively good correlation. For RMSE, TRMM 3B42RT performs better than the gauge corrected TRMM (3B42v7). However, TRMM 3B42v7 shows a better performance in the KGE′ components than TRMM 3B42RT.

This difference at daily scale over Imperial Basin can also be observed in Fig. 15. The boxplot titled Difference is calculated per station and is the subtraction of the daily observed values and the selected SRE values (observation - SRE) while the top-right boxplot titled Squared difference shows this daily differences between observations and the corresponding SRE squared. A positive value of the differences represents underestimation values and a negative value represents overestimation values of the correspondent SRE. In the case of TRMM 3B42RT, is observed that this difference is greater than the obtained

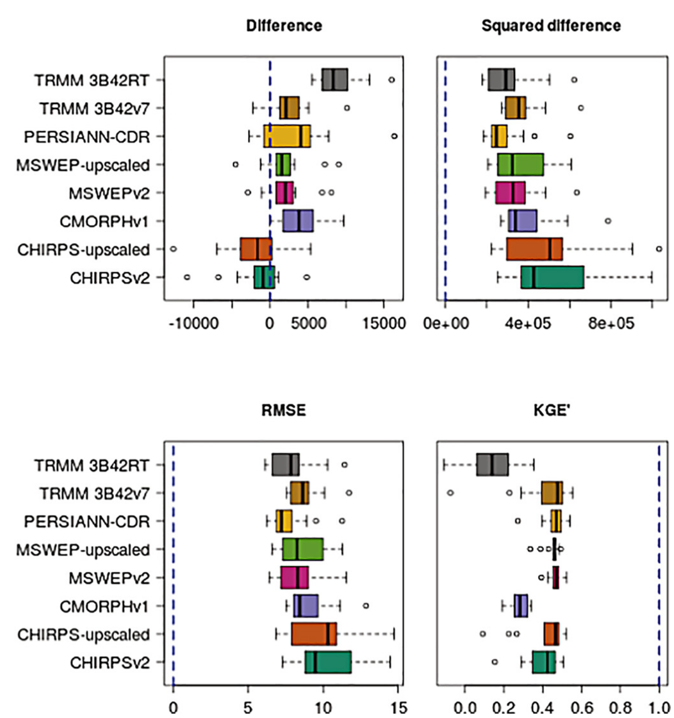

Fig. 15. Boxplots of the difference between observed values and the correspondent SRE (observation - SRE values; top-left), difference between observed values and the correspondent SRE to the squared at daily basis (top-right), RMSE (bottom-left), and KGE′ values (bottom-right) for Imperial River Basin at daily scale. The blue lines represent the optimal value for each parameter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with TRMM 3B42v7. However, due that RMSE penalizes the mismatches of heavy rain rate, the boxplot of the squared differences shows the opposite. It is observed that TRMM 3B42RT performs better than TRMM 3B42v7 when the RMSE and the squared differences are applied. Also, CMORPHv1 shows a greater positive difference compared to CHIRPSv2 and CHIRPSv2 upscaled but performs better using the RMSE. In both cases, the KGE′ is able to capture the behavior of the SREs by combining different measures of performance such as the linear correlation, bias and variability ratio.

In Fig. 14 the evaluation of performance over Imperial Basin at monthly scale can be also observed. For KGE, CHIRPSv2 and TRMM 3B42v7 performed the best, while for RMSE; MSWEPv2 and MSWEPv2 upscaled presented the best performance. MSWEPv2 and MSWEPv2 upscaled have a better correlation to the observations. However, CHIRPSv2 and CHIRPSv2 upscaled present higher value of β and γ.

As observed, RMSE and KGE′ can present different results when evaluating the spatio-temporal performance of different SREs. However, the KGE′ has proven to be a better index of performance due to its ability of not penalize mismatches of heavy rain rate and its ability of decomposing the total performance of the SREs into a linear correlation (r), a bias (β), and a variability (γ) terms (Fig. 15).

7. Conclusions

In this article, the performance of six state-of-the-art SREs (TRMM 3B42RT, TRMM 3B42v7, CMORPHv1, CHIRPSv2, PERSIANN-CDR and MSWEPv2) was compared against 201 rain gauge stations over three different catchments in Latin-America (Chile: Imperial River Basin, Brazil: Paraiba do Sul Basin and Colombia: Magdalena River Basin). The modified Kling-Gupta efficiency (KGE′) was used to evaluate the performance of the different SREs at daily, monthly, and seasonal scales. This index evaluates three different measures of performance between the satellite and the observed data; the linear correlation (r), bias (β), and variability ratio (γ) components which are able to reproduce the temporal dynamics as well as preserving the volume and variability of precipitation, respectively. The RMSE was also included in the evaluation because it is widely used in literature, and it was used to assess whether it is a useful measure of performance or not. Also, the percent bias (PBIAS) was computed to measure the average tendency of the simulated values to be larger or smaller than the observations. In addition, three categorical indices (POD, FAR, and fBias) were used to evaluate the performance of SREs for different precipitation intensities. The main findings were:

Each evaluated satellite product performed differently for each catchment and each time scale. MSWEPv2 performed the best over Imperial and Paraiba do Sul and CHIRPSv2 for Magdalena River Basin, when evaluated at daily scale. CHIRPSv2 presented the highest performance over Imperial and Magdalena basins and MSWEPv2 performed the best over Paraiba do Sul, when evaluated at monthly timescale. When the basins were evaluated at seasonal scale, CMORPHv1 performed the best for DJF and SON, TRMM 3B42v7 for MAM, and PERSIANN-CDR for JJA over Imperial Basin. MSWEPv2 performed the best over Paraiba do Sul Basin for all seasons. CHIRPSv2 showed the best performance over Magdalena Basin. Finally, it is worth to mention that the SREs, in general, the gauge-adjusted algorithms tend to perform better than those with no ground-based adjustment. For this reason, the independence of the validation dataset must be taken into account before any comparison among SREs due that it may affect the results of the evaluation.

The highest probability of detection (POD ˜ 1) was obtained for the no-rain intensities for all SREs over all the study areas. In general, all SREs presented a low POD for high rain events. The moderate-rain events (from 5 to 20 mm d^{-1}) were the best captured in all basins when precipitation was larger than 1 mm d^{-1}. In particular, MSWEPv2 performed considerably better for all rainfall intensities in Paraiba do Sul compared to the rest of the SREs. The fBias presented higher variations for the violent (higher than 40 mm) and the light rain events ([1, 5) mm d^{-1}) which is in agreement with the observed in the POD index. The results obtained with the false alarm ratio (FAR) showed consistency with the results obtained with POD showing low values for the no-rain events.

Despite the evolution of the satellite rainfall estimates, our results confirm that a catchment-specific validation is still needed in order to select a suitable SRE dataset for hydrological purposes. The results of this validation cannot be extrapolated to other basins, this comparison shows that the same satellite products present different behavior over different areas. We invite the readers to validate the SREs before using one over an area.

The Modified Kling-Gupta efficiency (KGE′) has proved to be a useful evaluation index. It decomposes the performance of the SREs into linear correlation (r), bias (β) and variability (γ) parameters. These parameters allowed us to understand the origin of mismatches between SREs and observations.

The Root Mean Squared Error (RMSE) gives more weight to the mismatches in two cases: i) when a low rainfall intensity is constantly not detected, and ii) in high rain events because it gives more weight squaring the difference of high values. Therefore, the RMSE is not recommended for evaluating SREs over an area for the aforementioned reasons and because its results are not comparable between areas with different precipitation regimes.

The upscaling procedure can affect the SRE performance and it may vary between products, timescales, and regions. If an upscaling procedure is performed while validating SREs over an area, the selection of the best performing product can be affected by it. We recommend to evaluate the performance before and after an upscaling procedure is computed in order to select the most representative SRE of the spatio-temporal precipitation patterns of an area. Also, the topography plays an important role when upscaling an SRE. The stations with higher performance differences between the original SRE and the upscaled one were located close to elevation gradients. For this reason, if a validation of different satellite rainfall estimates is needed in a mountainous area, the upscaling procedure may affect the results, but if the topography is not rugged, an upscaling procedure can enable a fair point-to-pixel comparison among the evaluated SREs.

Declarations of interest

None.

Funding

The authors thank FONDECYT (11150861) for partially funding the collaboration between the Universidad de la Frontera (Temuco, Chile) and the Institute for Technology and Resources Management in the Tropics and Subtropics (Köln, Germany). The authors also thank the Centers for Natural Resources and Development (CNRD) Ph.D. program for the financial support to the main author (Ph.D. scholarship).

Acknowledgements

Mauricio Zambrano-Bigiarini thanks the Fondap-Conicyt (15110009) for the Chilean precipitation dataset. Finally, the authors thank two anonymous referees, whose constructive comments helped to improve the quality of this manuscript.

Appendix A. Continous indices of model performance

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad (A1)$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})}} \quad (A2)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (A3)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (A4)$$

where n is the number of observations; O_i and S_i are the observed and the corresponding SRE values at day i, respectively; $\bar{O}$ and S- are the arithmetic mean of the observations and the corresponding SRE, respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (A5)$$

where n is the number of observations; yi and ŷi are the observed and the corresponding SRE values at day i, respectively.

$$PBIAS = (100)\frac{\sum_{i=1}^{n}(S_i - O_i)}{\sum_{i=1}^{n}(O_i)} \quad (A6)$$

where n is the number of observations; Oi and Si are the observed and the corresponding SRE values at day i, respectively.

Appendix B. Categorical indices of model performance

$$POD = \frac{H}{H+M} \quad (B1)$$

$$fBias = \frac{H+F}{H+M} \quad (B2)$$

$$FAR = \frac{F}{H+F} \quad (B3)$$

where H indicates a hit (a satellite estimate that correctly identifies the type of precipitation event measured at the rain gauge station); M indicates a miss (an event recorded at the rain gauge but not correctly identified by the SRE); F represents a false alarm (a precipitation event detected by the SRE but not recorded at the corresponding rain gauge).

References

Ashouri, Hamed, Hsu, Kuo-Lin, Sorooshian, Soroosh, Braithwaite, Dan K., Knapp, Kenneth R., Dewayne Cecil, L., Nelson, Brian R., Prat, Olivier P., 2015. PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Am. Meteorol. Soc. 96 (1), 69–83. http://dx.doi.org/10.1175/BAMS-D-13-00068.1.

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615. http://dx.doi.org/10.5194/hess-21-589-2017.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21 (12), 6201–6217. http://dx.doi.org/10.5194/hess-21-6201-2017.

Berndt, Christian, Rabiei, Ehsan, Haberlandt, Uwe, 2013. Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. J. Hydrol. 508 (January), 88–101. http://dx.doi.org/10.1016/j.jhydrol.2013.10.028.

Carlo, Espinoza Villar Jhan, Josyane, Ronchail, Loup, Guyot Jean, Gerard, Cochonneau, Filizola, Naziano, Waldo, Lavado, Eurides, De Oliveira, Rodrigo, Pombosa, Philippe, Vauchel, 2009. Spatio-temporal rainfall variability in the Amazon Basin countries Brazil, Peru, Bolivia, Colombia, and Ecuador. Int. J. Climatol. 29 (11), 1574–1594. http://dx.doi.org/10.1002/joc.1791.

Ceccherini, Guido, Ameztoy, Iban, Hernndez, Claudia Patricia Romero, Moreno, Cesar Carmona, 2015. High-resolution precipitation datasets in South America and West Africa based on satellite-derived rainfall, enhanced vegetation index and digital elevation model. Remote Sens. 7 (5), 6454–6488. http://dx.doi.org/10.3390/rs70506454.

Delahaye, Florian, Kirstetter, Pierre-Emmanuel, Dubreuil, Vincent, Machado, Luiz A.T., Vila, Daniel A., Clark, Robert, 2015. A consistent gauge database for daily rainfall analysis over the legal Brazilian Amazon. J. Hydrol. 527 (August), 292–304. http://dx.doi.org/10.1016/j.jhydrol.2015.04.012.

Dinku, Tufa, Ruiz, Franklyn, Connor, Stephen J., Ceccato, Pietro, 2010. Validation and intercomparison of satellite rainfall estimates over Colombia. J. Appl. Meteorol. Climatol. 49 (5), 1004–1014. http://dx.doi.org/10.1175/2009JAMC2260.1.

Reis dos, João Bosco Coura, Renn, Camilo Daleles, Lopes, Eymar Silva Sampaio, 2017. Validation of satellite rainfall products over a mountainous watershed in a humid subtropical climate region of Brazil. Remote Sens. 9 (12), 1240. http://dx.doi.org/10.3390/rs9121240.

Funk, Chris, Peterson, Pete, Landsfeld, Martin, Pedreros, Diego, Verdin, James, Shukla, Shraddhanand, Husak, Gregory, et al., 2015. The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes. Sci. Data 2 (December), 150066. http://dx.doi.org/10.1038/sdata.2015.66.

Gebremichael, M., Anagnostou, E.N., Bitew, M.M., et al., 2010. Critical steps for continuing advancement of satellite rainfall applications for surface hydrology in the nile river basin. J. Am. Water Resour. Assoc. 361–366. http://dx.doi.org/10.1111/j.1752-1688.2010.00428.x.

Guo, Hao, Chen, Sheng, Bao, Anming, Jujun, Hu, Gebregiorgis, Abebe S., Xue, Xianwu, Zhang, Xinhua, 2015. Inter-comparison of high-resolution satellite precipitation products over Central Asia. Remote Sens. 7 (6), 7181–7211. http://dx.doi.org/10.3390/rs70607181.

Gupta, Hoshin V., Kling, Harald, Yilmaz, Koray K., Martinez, Guillermo F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1), 80–91. http://dx.doi.org/10.1016/j.jhydrol.2009.08.003.

Haberlandt, Uwe, 2006. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J. Hydrol. 332 (January), 144–157. http://dx.doi.org/10.1016/j.jhydrol.2006.06.028.

Hijmans, Robert J., van Etten, Jacob, Cheng, Joe, Mattiuzzi, Matteo, Sumner, Michael, Greenberg, Jonathan A., Lamigueiro, Oscar Perpinan, et al., 2017. Raster: Geographic Data Analysis and Modeling. https://CRAN.R-project.org/package=raster.

Hsu, Kou-lin, Gao, Xiaogang, Sorooshian, Soroosh, Gupta, Hoshin V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteorol. 36 (9), 1176–1190. http://dx.doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

Huffman, George J., Bolvin, David T., Nelkin, Eric J., Wolff, David B., Adler, Robert F., Guojun, Gu, Hong, Yang, Bowman, Kenneth P., Stocker, Erich F., 2007. The TRMM multisatellite precipitation analysis TMPA: quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8 (1), 38–55. http://dx.doi.org/10.1175/JHM560.1.

IDEAM, 2001. Estudio Ambiental de La Cuenca Magdalena-Cauca Y Elementos Para Su Ordenamiento Territorial. Technical Report and Arcinfo Database. Instituto de Hidrologa, Metereologa Y Estudios Ambientales (IDEAM), Bogot D.C., Colombia. http://documentacion.ideam.gov.co/cgi-bin/koha/opac-detail.pl?biblionumber=6122&shelfbrowse_itemnumber=6272.

Joyce, Robert J., Janowiak, John E., Arkin, Phillip A., Xie, Pingping, 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5 (3), 487–503. http://dx.doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kling, Harald, Fuchs, Martin, Paulin, Maria, 2012. Runoff conditions in the upper Danube Basin under an ensemble of climate change scenarios. J. Hydrol. 424–425 (March), 264–277. http://dx.doi.org/10.1016/j.jhydrol.2012.01.011.

Melo Davi de, C.D., Xavier Alexandre, C., Thiago, Bianchi, Oliveira Paulo, T.S., Scanlon Bridget, R., Lucas Murilo, C., Edson, Wendland, 2015. Performance evaluation of rainfall estimates by TRMM multisatellite precipitation analysis 3b42v6 and V7 over Brazil. J. Geophys. Res. 120 (18), 9426–9436. http://dx.doi.org/10.1002/2015JD023797.

Menne, Matthew J., Durre, Imke, Vose, Russell S., Gleason, Byron E., Houston, Tamara G., 2012. An overview of the global historical climatology network-daily database. J. Atmos. Ocean. Technol. 29 (7), 897–910. http://dx.doi.org/10.1175/JTECH-D-11-00103.1.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Kppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11 (5), 1633–1644. http://dx.doi.org/10.5194/hess-11-1633-2007.

Peterson, Thomas C., Vose, Russell S., 1997. An overview of the global historical climatology network temperature database. Bull. Am. Meteorol. Soc. 78 (12), 2837–2850 doi:10.1175/1520-0477(1997)078 < 2837:AOOTGH > 2.0.CO;2.

Restrepo, Juan D., Kjerfve, Bjrn, Hermelin, Michel, Restrepo, Juan C., 2005. Factors controlling sediment yield in a major south American Drainage Basin: the Magdalena River, Colombia. J. Hydrol. 316 (1), 213–232. http://dx.doi.org/10.1016/j.jhydrol.2005.05.002.

Rivera, N.R., Encina, F., Muoz-Pedreros, A., Mejias, P., 2004. La Calidad de Las Aguas En Los Ros Cautn E Imperial, IX Regin-Chile. Información Tecnolgica 15 (5), 89–101. http://dx.doi.org/10.4067/S0718-07642004000500013.

Rogelis, M.C., Werner, M.G.F., 2013. Spatial interpolation for real-time rainfall field estimation in areas with complex topography. J. Hydrometeorol. 14 (1), 85–104. http://dx.doi.org/10.1175/JHM-D-11-0150.1.

Salio, Paola, Hobouchian, Mara, Skabar, Yanina Garca, Vila, Daniel, 2015. Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. Atmos. Res. 163 (November). http://dx.doi.org/10.1016/j.atmosres.2014.11.017.

Schuster, Grit, Ebert, Elizabeth E., Stevenson, Mark A., Corner, Robert J., Johansen, Cheryl A., 2011. Application of satellite precipitation data to analyse and model arbovirus activity in the tropics. Int. J. Health Geogr. 10 (January), 8. http://dx.doi.org/10.1186/1476-072X-10-8.

Silva, Walquiria Marins, Simões, Silvio Jorge C., 2014. Spatial intra-annual variability of precipitation based on geostatistics. A case study for the Paraiba do Sul Basin, southeastern Brazil. Int. J. Geosci. 05 (04), 408. http://dx.doi.org/10.4236/ijg.2014.54039.

Simões, Silvio, Barros, Ana, 2007. Regional hydroclimatic variability and Brazil's 2001 energy crisis. Manag. Environ. Qual. 18 (April), 263–273. http://dx.doi.org/10.1108/14777830710731734.

Soares, Paulo, Yoshinaga, Sueli, Simes, Silvio, Bernardes, George, Barbosa, Srgio, Trannin, Isabel, 2012. The definition of potential infiltration areas in Guaratinguet watershed, Paraba Do Sul Basin, southeastern Brazil: an integrated approach using physical and land-use elements. Environ. Earth Sci. 67 (November). http://dx.doi.org/10.1007/s12665-012-1610-6.

Sorooshian, Soroosh, Hsu, Kuo-Lin, Gao, Xiaogang, Gupta, Hoshin V., Imam, Bisher, Braithwaite, Dan, 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. Bull. Am. Meteorol. Soc. 81 (9), 2035–2046. http://dx.doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.

Team, R Development Core, 2011. R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

Thiemig, Vera, Rojas, Rodrigo, Zambrano-Bigiarini, Mauricio, Levizzani, Vincenzo, Ad, De Roo, 2012. Validation of satellite-based precipitation products over sparsely gauged African River basins. J. Hydrometeorol. 13 (6), 1760–1783. http://dx.doi.org/10.1175/JHM-D-12-032.1.

Weiss, L.L., Wilson, W.T., 1953. Evaluation of significance of slope changes in double mass curves. EOS Trans. Am. Geophys. Union 34 (6), 893–896. http://dx.doi.org/10.1029/TR034i006p00893.

Woldemeskel, Fitsum, Sivakumar, B., Sharma, Ashish, 2013. Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. J. Hydrol. 499 (August), 167–176. http://dx.doi.org/10.1016/j.jhydrol.2013.06.039.

Xie, Pingping, Joyce, Robert, Wu, Shaorong, Yoo, Soo-Hyun, Yarosh, Yelena, Sun, Fengying, Lin, Roger, 2017. Reprocessed, Bias-corrected CMORPH global high-resolution precipitation estimates from 1998. J. Hydrometeorol. 18 (6), 1617–1641. http://dx.doi.org/10.1175/JHM-D-16-0168.1.

Zambrano-Bigiarini, Mauricio, 2017a. hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. https://CRAN.R-project.org/package=hydroGOF.

Zambrano-Bigiarini, Mauricio, 2017b. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. https://CRAN.R-project.org/package=hydroTSM.

Zambrano-Bigiarini, Mauricio, Nauditt, Alexandra, Birkel, Christian, Verbist, Koen, Ribbe, Lars, 2017. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. Hydrol. Earth Syst. Sci. 21 (March), 1295–1320. http://dx.doi.org/10.5194/hess-21-1295-2017.