Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile

Mauricio Zambrano-Bigiarini[1,2], Alexandra Nauditt[3], Christian Birkel[4,5], Koen Verbist[6,7], and Lars Ribbe[3]

[1]Department of Civil Engineering, Universidad de La Frontera, Temuco, Chile
[2]Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile
[3]Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Cologne Technical University, 50679 Köln, Germany
[4]Department of Geography, University of Costa Rica, 2060 San José, Costa Rica
[5]Northern Rivers Institute, University of Aberdeen, AB24 3UF Aberdeen, UK
[6]Hydrological Systems & Global Change Section, UNESCO Chile, Santiago, Chile
[7]International Centre for Eremology, Department of Soil Management, Ghent University, Ghent 9000, Belgium

Correspondence to: Mauricio Zambrano-Bigiarini (mauricio.zambrano@ufrontera.cl)

Abstract. Accurate representation of the real spatio-temporal variability of catchment rainfall inputs is currently severely limited. Moreover, spatially interpolated catchment precipitation is subject to large uncertainties, particularly in developing countries and regions which are difficult to access. Recently, satellite-based rainfall estimates (SRE) provide an unprecedented opportunity for a wide range of hydrological applications, from water resources modelling to monitoring of extreme events such as droughts and floods.

This study attempts to exhaustively evaluate -for the first time- the suitability of seven state-of-the-art SRE products (TMPA 3B42v7, CHIRPSv2, CMORPH, PERSIANN-CDR, PERSIAN-CCS-adj, MSWEPv1.1 and PGFv3) over the complex topography and diverse climatic gradients of Chile. Different temporal scales (daily, monthly, seasonal, annual) are used in a point-to-pixel comparison between precipitation time series measured at 366 stations (from sea level to 4600 m a.s.l. in the Andean Plateau) and the corresponding grid cell of each SRE (rescaled to a 0.25° grid if necessary). The modified Kling-Gupta efficiency was used to identify possible sources of systematic errors in each SRE. In addition, five categorical indices (PC, POD, FAR, ETS, fBIAS) were used to assess the ability of each SRE to correctly identify different precipitation intensities.

Results revealed that most SRE products performed better for the humid South (36.4-43.7°S) and Central Chile (32.18-36.4°S), in particular at low- and mid-elevation zones (0-1000 m a.s.l.) compared to the arid northern regions and the Far South. Seasonally, all products performed best during the wet seasons autumn and winter (MAM-JJA) compared to summer (DJF) and spring (SON). In addition, all SREs were able to correctly identify the occurrence of no rain events, but they presented a low skill in classifying precipitation intensities during rainy days. Overall, PGFv3 exhibited the best performance everywhere and for all time scales, which can be clearly attributed to its bias-correction procedure using 217 stations from Chile. Good results were also obtained by the research products CHIRPSv2, TMPA 3B42v7 and MSWEPv1.1, while CMORPH, PERSIANN-CDR and the real-time PERSIANN-CCS-adj were less skillful in representing observed rainfall. While PGFv3 (currently available up to 2010) might be used in Chile for historical analyses and calibration of hydrological models, the high spatial resolution,

low latency and long data records of CHIRPS and TMPA 3B42v7 (in transition to IMERG) show promising potential to be used in meteorological studies and water resources assessments. We finally conclude that despite improvements of most SRE products, a site-specific assessment is still needed before any use in catchment-scale hydrological studies.

1   Introduction

5   Accurate representation of the real spatio-temporal variability of catchment rainfall inputs is currently severely limited. Traditionally, precipitation (P) data are collected through ground-based observations using raingauges and/or weather radars. Catchment-representative rainfall is usually obtained by interpolation of point rainfall measured at raingauges (e.g. Rogelis and Werner, 2013; Verworn and Haberlandt, 2011; Zhang and Srinivasan, 2009; Kurtzman et al., 2009). However, even in densely monitored regions, precipitation data are highly uncertain (Tian and Peters-Lidard, 2010; Woldemeskel et al., 2013).
10   In developing countries and regions which are difficult to access -such as high elevation areas- there is usually only a sparse network of meteorological stations available, and therefore the obtained spatial rainfall fields are subject to even larger uncertainties (Woldemeskel et al., 2013).

To overcome some of the aforementioned limitations of ground-based rainfall measurements, space-based estimates of precipitation provide a promising alternative source. Several near-global high resolution satellite-based rainfall estimates (*SRE*) have recently become operational, including the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2000; Hsu et al. 1997), the PERSIANN-Cloud Classification System estimation (PERSIANN-CCS; Hong et al. 2004), the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center morphing technique product (CMORPH; Joyce et al. 2004; Janowiak et al. 2005), and the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis products (TMPA, Huffman et al. 2007), among others. The instruments on-board the satellites include passive microwave (PMW), visible (VIS) and infra-red (IR) sensors, and meteorological radar. Recent advances in sensor technology and methods for merging various data sources (e.g., geostationary thermal infrared, passive microwave, radar, and information from the Global Telecommunication System (GTS)) have led to a continuous improvement of temporal and spatial resolution of these precipitation products (e.g. Kidd et al., 2009).

Satellite estimates of precipitation have received different names and acronyms in the literature: satellite precipitation estimates (*SPE*, Scofield and Kuligowski 2007), satellite-based rainfall estimates (*SRFE*, Thiemig et al. 2012), satellite quantitative precipitation estimates (*SQPE*, Lee et al. 2015), satellite rainfall estimates (*SRE*, Abera et al. 2016) and satellite precipitation products (*SPP*, Maggioni et al. 2016; Serrat-Capdevila et al. 2013). Here, we will use *SRE* to name satellite-based rainfall estimates throughout the text.

The emergence of the aforementioned near-global and high-resolution SRE opens up new possibilities for applications in
30   data-scarce or ungauged regions. However, SRE products need evaluation and often calibration before any use in hydrological applications. Recently, Maggioni et al. (2016) published a review on SRE accuracy during the TRMM-Era, evaluating TMPA 3B42 (research and real-time products), CMORPH, GSMAP, PERSIANN and PERSIANN-CCS. They found that topography, seasonality, and climate impacted on the SREs performance, especially in probability of detection and bias. Tian and Peters-

Lidard (2010) studied uncertainties in SRE, by computing the variance from an ensemble of six different TRMM datasets. They found that SRE are more reliable over areas with strong convective precipitation and flat surfaces, such as the tropical oceans and South America. Dinku et al. (2010) evaluated CMORPH and two TMPA products (3B42 and 3B42RT) for mountainous regions of Africa and South America. Both products underestimated the occurrence and amount of rainfall which they attributed to the complex terrain and orographic rain process. Scheel et al. (2011) compared TMPA 3B42v6 estimates with raingauges in the regions of Cuzco (Peru) and La Paz (Bolivia). They detected large biases in the estimation of daily precipitation amounts. The occurrence of strong precipitation events was well represented but their intensities were underestimated. In addition, TMPA estimates for La Paz showed high false alarm ratios. Mantas et al. (2015) validated the research 3B42v7 and the near real-time 3B42RT for the Peruvian Andes of similar complex topography against in situ data. Results also showed a strong regional variability due to different climatic and topographic features. Thiemig et al. (2012) compared six SRE against raingauge data over four African river basins. They found that SRE showed higher performance over the tropical wet and dry zone compared to semiarid mountainous regions, low accuracy in detecting heavy rainfall events over semiarid areas, general underestimation of heavy rainfall events, and overestimation of the number of rainy days in the tropics. Demaria et al. (2011) used an object-based verification method to explore the existence of systematic errors for three SRE in South America (La Plata River Basin): TRMM, CMORPH and PERSIANN. They found that PERSIANN underestimated the observed average rainfall rate and maximum rainfall, CMORPH overestimated the average rainfall rate while the maximum rainfall was slightly underestimated, and the average rainfall rate and volume provided by TRMM correlated well with ground observations, whereas the maximum rainfall was systematically overestimated. In general, there does not seem to exist a single SRE product that performs best always and everywhere. Therefore, the performance of each SRE needs to be assessed for each individual case study. Furthermore, there is little guidance on which performance criteria to use in evaluating SRE products (e.g. Hossain and Huffman, 2008).

Nevertheless, climate and hydrological studies in data-sparse regions can benefit from the spatial coverage and grid-structure of SRE to drive hydrological simulations for water resources management (e.g. Tobin and Bennett, 2014; Meng et al., 2014; Xue et al., 2013; Li et al., 2012; Khan et al., 2011; Hong et al., 2009; Su et al., 2008; Thiemig et al., 2013; Xue et al., 2013) and monitoring of floods (e.g. Werren et al., 2016; Hong et al., 2007) and droughts (e.g. Tao et al., 2016; AghaKouchak et al., 2015; Zhang and Jia, 2013; Naumann et al., 2012). Reliable information on the spatio-temporal variability of rainfall is also one of the main factors to achieve food security, in particular in data scarce regions (Kang et al., 2009; Verdin et al., 2005).

Notwithstanding all the recent improvements of SRE, there are still some issues that need to be addressed before reaching its full potential (Tobin and Bennett, 2014). Several researchers have found that satellite products fail in capturing certain types of precipitation events, have important biases and present false detection of precipitation (e.g. Gebregiorgis and Hossain, 2013; Thiemig et al., 2012; Ebert et al., 2007). When SRE products are used to drive hydrological models, the aforementioned errors are propagated in a non-linear way to simulated streamflows (see e.g. Bisselink et al., 2016; Nikolopoulos et al., 2010; Fekete et al., 2004). In the worst case, such simulations may lead to wrong conclusions and poor management decisions with potentially devastating societal consequences. In addition, different hydrological applications, such as drought monitoring, flood forecasting, water resources management or allocation of long-term water rights, require precipitation at very different time

scales, ranging from hourly to seasonal (Tobin and Bennett, 2014). Also, there is a clear need to better understand how hydrological simulations forced by satellite-derived data depends on the climatological regime and time scale used for simulations (see e.g. Gebregiorgis and Hossain, 2013; Nikolopoulos et al., 2010; Tobin and Bennett, 2014).

In the last decade, droughts of unusual severity have affected the Chilean territory, both because of their intensity and multi-annual duration (Boisier et al., 2016). This is consistent with climate change projections for that region, which indicate central-southern Chile as a global hotspot for increased drought frequency, with likely water security issues in this region (Prudhomme et al., 2014). On the other hand, flood events have a relatively normal occurrence in Chile (Müller et al., 2011) and some particularly intense events have affected the country in recent years. Considering the significant socio-economic costs of extreme water-related hazards (Kundzewicz et al., 2008; Mirza, 2003) we chose the mountainous Chilean territory as case study, due to the presence of extreme elevations (0-6893 m a.s.l.) and its heterogeneous hydroclimatic conditions (hyperarid in the north and extremely wet in the south). Further, there are few climate monitoring stations above 1500 m a.s.l. The present study attempts to exhaustively compare -for the fist time- seven state-of-the-art SRE products against ground-based measurements in this data-scarce and complex mountainous region, in order to provide guidelines about suitable SRE products for future hydrological applications. Spatio-temporal characteristics of different SRE products are compared against that of existing raingauge data, by using different continuous and categorical performance measures. In particular, this study will address the following research questions:

1. what is the overall performance of each SRE?

2. which SRE performs best across the topographic and climate gradient in Chile?

3. which SRE performs best for different time scales (daily, monthly, seasonal, annual)?

4. how does the accuracy of a given SRE change for different precipitation intensities?,

5. when a SRE does not capture the observed precipitation, is it due to a misrepresentation of the shape, magnitude, variability or all the previous properties of the precipitation time series?, and

6. is there any SRE that performs best compared to all the others, everywhere and for all time scales?

Results of this study aim at increasing our knowledge about the suitability of different SRE estimates to characterise the spatio-temporal distribution of precipitation across the climatically and topographically diverse Chilean territory, as a cost-effective complement to ground-based measurement networks in this data-scarce region. In addition, findings of this study aim at providing feedback to the developers of different SRE products for potential use in future releases.

The article is organized as follows: Section 2 presents the study area and datasets, with Section 3 describing the methodology used to compare satellite products against ground observations. Numerical and graphical results are shown in Section 4, whereas Section 5 provides an in-depth discussion in the light of the wider literature. We present concluding remarks in Section 6.

2 Study area and datasets

2.1 Study area

The continental area of Chile has more than 4000 km of latitudinal extension, from 17.50°S to 66.42°S, bounded by the Pacific Ocean to the west (∼76° W) and by the Andes mountain range in the east (∼66° W). Four main morphological units condition the existence of eleven different types of climate (from hot dessert to polar/tundra) and associated vegetation: Coastal Plains, Coastal Mountains, Intermediate Depression and the Andes, with elevations ranging from sea level to 6893 m a.s.l. at the Ojos del Salado volcano. Figure 1 shows the location of the study area, including a digital elevation model and main river basins (panel a), mean annual precipitation (panel b) and temperature (panel c), and the eleven climate types identified based on Köppen classification. The main factors affecting the climate of the Chilean territory are the latitude, topography and the oceanic influence coming from the long Pacific Ocean (INE, 2015). The northern part of Chile is characterized as hyperarid/arid/semiarid, with extremely low precipitation and high temperatures; while abundant precipitation is observed in the south, reaching amounts of up to 5000 mm/year with lower temperatures (Valdés-Pineda et al., 2014). The four traditional seasons of the southern hemisphere are present in Chile: autumn (MAM), winter (JJA), spring (SON), and summer (DJF), with the wet season occurring predominantly during winter in most of the central-southern territory, during summer in the northern regions of Tarapacá and Antofagasta (*Bolivian winter*), and no clearly defined dry season in the southernmost part of Chile (INE, 2015). An exhaustive description and references about the type of precipitations present in Chile and the global patterns that influence its interannual variability can be found on Valdés-Pineda et al. (2014).

To evaluate the performance of SRE, five major macroclimatic zones in Chile were selected, as shown in Figure 1: (i) Far North (17.50-26.00°S), (ii) Near North (26.00-32.18°S), (iii) Central Chile (32.18-36.40°S), (iv) South (36.40-43.70°S) and (v) Austral/Far South (43.70-56.00°S), slightly adapted from DGA (2016).

2.2 Datasets

2.2.1 Raingauges

Time series of observed precipitation were downloaded from an updated dataset of 781 raingauges provided by the Center of Climate and Resilience Research (CR2), with daily data from 01-Jan-1940 to 31-Dec-2015 (available on http://www.cr2.cl/recursos-y-publicaciones/bases-de-datos/). The original raw data were provided by the Chilean Water and Meteorological agencies (DGA and DMC, respectively). However, in order to use only stations with a long record of observations but also to ensure a minium number of stations representative of the Chilean topography and climatic zones, we selected only raingauges with less than 2% of missing data during the period Jan/2003-Dec/2010, which resulted in 366 stations within our study area (see Section 3.1). Figure 1a shows the spatial distribution of the selected raingauges. Daily time series in all the 781 stations analysed in this work can be found in the supplementary material (Zambrano-Bigiarini et al., 2016).

It is worth mentioning that several SREs use observed precipitation data from the Global Precipitation Climatology Centre (GPCC, Schneider et al. 2008) to adjust their satellite-only estimates. Here we analysed the number of precipitation gauges

from Chile used in the GPCC dataset. Figure 2 shows that this number has fluctuated over time, reaching a maximum of 146 in the period 1969-1995, dropping to a value of around 30 stations in the last two decades, which is clearly not enough to capture the spatial variability of rainfall in the mountainous Chilean territory. Moreover, the spatial distribution of the gauges used to create the GPCC for Chile has also observed a strong contraction (Figure 2), leaving a large area of the territory without any observation, which subsequently increases the estimation errors of the final product.

2.2.2 Satellite-based data

As the most suitable SRE product found in this work will be used in the future for hydrological modelling in selected basins, we focused only on SREs with long data records and good quality data over the Chilean territory, without giving any priority to low-latency (near)real-time products. In addition, comparability of (near)real-time products and station-based calibrated research products is low (Habib et al., 2009; Huffman et al., 2007). Therefore, seven of the most state-of-the-art SREs with at least 10 years of daily estimates and relatively high spatial resolution are compared against observed precipitation (i.e., raingauges). A brief description of each SRE with some previous applications is given in the next paragraphs, but the interested reader can find more information in the references provided for each product.

CMORPH

The NOAA Climate Prediction Center (CPC) MORPHing technique (Joyce et al., 2004) provides quasi-global estimates of precipitation at relatively high spatial resolution ($0.07° x 0.07°$ and $0.25° x 0.25°$) and frequent temporal resolution (half-hourly and 3-hourly), from $60°N$ to $60°S$. CMORPH estimates are based solely on PMW data (Joyce et al., 2004; Janowiak et al., 2005) with IR imagery not used to estimate precipitation but only to interpolate between two PMW-derived rainfall intensity fields. CMORPH has been reported to outperform other SRE products over the Australian tropics (Ebert et al., 2007; Joyce et al., 2004), central United States (Behrangi et al., 2011) and Europe (Kidd et al., 2012). As many SRE products that do not rely on rainfall gauge data, CMORPH tends to overestimate the amount of precipitation during wet periods (Behrangi et al., 2011). Pereira Filho et al. (2010) compared CMORPH rainfall estimates over South America (Amazon Basin), at 8-km spatial resolution, with available rainfall observations at daily, monthly, and yearly accumulation time scales. Their results show that the correlation between satellite-derived and gauge-measured precipitation increases with the accumulation period, from daily to monthly, especially during the rainy season. In this work we used the 3-hr observations at a spatial resolution of $0.25°$ in NetCDF format, directly downloaded for the Chilean study area from the http://rda.ucar.edu/ website.

PERSIANN-CDR

The PERSIANN algorithm uses an Artificial Neural Network (ANN) model to estimate precipitation using IR. Its accuracy is improved by adaptive adjustment of the network parameters using rainfall estimates from a passive microwave sensor. At the pixel level, the algorithm fits the mean and standard deviation of the brightness temperature of a pixel and the adjacent pixel's temperature texture to the calculated precipitation rate (Hsu et al., 1997). A new product named The PERSIANN-CDR (for Climate Data Record) was developed by applying the PERSIANN algorithm to Gridded Satellite Infrared Data (GridSat-B1) and then bias-correcting estimations using $2.5°$ monthly GPCP precipitation data (Ashouri et al.,

2015). This product provides a 30-year data of near-global (60°S-60°N) daily precipitation data at 0.25° spatial resolution (which is about 25 km at the equator). Mei et al. (2014) compared measured precipitation data in mountainous regions of the Italian Alps with PERSIANN-CDR and found that the product slightly overestimated P in regions with low rainfall and underestimated P in regions with high rainfall. All the daily data for the study area were downloaded from
5   http://www.climatedatalibrary.cl/SOURCES/.UCIrvine/.CHRS/.PERSIANN-CDR.

PERSIANN-CCS-adj

PERSIANN-CCS (Hong et al., 2004) is an infrared-based satellite estimation process, employing image processing and pattern recognition techniques to develop a patch-based cloud classification system (CCS). Combined with an artificial neural network model, hourly pixel precipitation intensity is estimated globally at 0.04°x0.04° and accumulated daily. The PERSIANN-CCS-
10 adj was developed especially for Chile applying a non-parametric Quantile Mapping and Gaussian Weighting (QM-GW) interpolation process to reduce the systematic biases existing in the daily PERSIANN-CCS dataset over the country. In summary, rainfall amounts obtained from 456 raingauges in the period 2009-2014 were used to construct non-parametric cumulative density functions (CDFs) for precipitation amounts during the four different seasons. These observed CDFs were compared with the CDFs of the PERSIANN-CCS over the same time period and then used to correct the biases in the PERSIANN-CCS
15 estimates based on the nonparametric QM-GW approach. Using this climatological bias correction, the PERSIANN-CCS-Adj provides historical and real-time precipitation estimates in closer agreement with the seasonal variation of observed precipitation. A full description of the methodology and validation results can be found in Yang et al. (2016). It should be noted that PERSIANN-CCS-adj is the only (near)real-time product used in this work, and it was considered just because it was especially developed for Chile.

20 **3B42v7**

The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) is intended to provide a "best" estimate of quasi-global precipitation from a wide variety of modern satellite-borne precipitation-related sensors. Rainfall estimates are provided at relatively high spatial resolution (0.25°x0.25°) at 3-hourly time steps, in both real and post-real time, to meet a wide range of research needs (Huffman et al., 2007, 2010). This SRE combines IR data from geosynchronous
25 earth orbit (GEO) satellites with four passive microwave (PMW) sensors, namely TRMM Microwave Imager (TMI), Special Sensor Microwave/Imager (SSM/I), Advanced Microwave Sounding Unit-B (AMSU-B), and the Advanced Microwave Scanning Radiometer-EOS (AMSR-E). The TMPA products include the version 6 (v6) and version 7 (v7) real-time products 3B42RT (3B42RTv6 and 3B42RTv7), the 3-hourly research products 3B42 (3B42v6 and 3B42v7), the daily accumulated rainfall product 3B42_daily (v6 and v7), and the monthly products 3B43 (3B43v6 and 3B42v7). The 3B42 algorithm is exe-
30 cuted in four steps: 1) PMW precipitation estimates are calibrated and combined; 2) IR precipitation estimates are generated using the calibrated PMW data; 3) both IR and PMW data are then combined; and 4) rescaled on a monthly basis using two sources of raingauge data: i) the Global Precipitation Climatological Centre (GPCC) monthly raingauge analysis and, ii) the Climate Assessment and Monitoring System (CAMS) monthly raingauge analysis, developed by CPC (Huffman et al., 2007, 2010). The newest version of TMPA 3B42v7 was released on June 2012, and recent studies show that 3B42v7 esti-

mates improve upon 3B42v6 (e.g. Chen et al., 2013). Today, after more than 17 years of data collection, the instruments on TRMM were turned off on April 8th 2015, but the TMPA 3B42 product will continue to be produced through early 2018 (https://pmm.nasa.gov/data-access/downloads/trmm). In this work we only used the 3B42_daily product, which was accumulated to monthly, seasonal or annual values depending on the analysis.

5   CHIRPSv2

The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS version 2) is a global daily, pentadal and monthly precipitation product explicitly designed for monitoring agricultural drought and global environmental change over land (Funk et al., 2015). CHIRPS combines remotely sensed precipitation of geo-synchronous and polar-orbiting satellites, from five different satellite products, with more than 2000 station records to calibrate global Cold Cloud Duration (CCD) rain-
10 fall estimates (Funk et al., 2015). The product features a spatial resolution of 0.05° from 50°S to 50°N (across all longitudes) with a > 30 years final monthly precipitation record (1981-present). The station data from the near real-time World Meteorological Organization's Global Telecommunication System (GTS) are continuously used to update roughly every two days (a stable product is released every three weeks) and validate the remote sensors using additional information (physiographic and remotely sensed earth energy emissions that correspond to the location and intensity of precipitation) via moving-window
15 geostatistical regression (Funk et al., 2015). In summary, the CHIRPS process involves three main components: i) the 0.05° Climate Hazards group Precipitation climatology (CHPclim Funk et al. 2015), ii) the satellite-only Climate Hazards group Infrared Precipitation (CHIRP), and iii) the station merging procedure. Even though, the continuous development of CHIRPS is mainly in support of drought-related issues in Africa (Climate Hazards Group), there are now other global applications available (e.g., http://ewx.chg.ucsb.edu:8080/EWX/index.html, http://chg.geog.ucsb.edu/tools/geowrsi/index.html) and also papers
20 that have looked at climate dynamics in South America (Ceccherini et al., 2015; Deblauwe et al., 2016).

MSWEPv1.1

In June 2016 a new global precipitation dataset was released, the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset providing data for the period 1979-2015 with a 3-hourly temporal and 0.25° spatial resolution (Beck et al., 2016). It was specifically developed for hydrological modelling with the aim to overcome shortages related to the performance of satellite
25 products in representing precipitation in mountainous, tropical and snowmelt driven regions. It is based on different types of data sources as rain-gauge measurements, satellite observations as well as estimates from atmospheric models. The long-term mean of MSWEP is based on the CHPclim dataset. Where available, data are replaced by more accurate regional datasets. A correction for gauge under-catch and orographic effects is introduced by inferring catchment-average P from streamflow (Q) observations at 13 762 stations across the globe. The temporal variability of MSWEP is determined by weighted averaging of
30 P anomalies from seven datasets; two based solely on interpolation of gauge observations (CPC Unified and GPCC), three on satellite remote sensing (CMORPH, GSMaP-MVK, and TMPA 3B42RT), and two on atmospheric model reanalysis (ERA-Interim and JRA-55). For each grid cell, the weight assigned to the gauge-based estimates is calculated from the gauge network density, while the weights assigned to the satellite- and reanalysis-based estimates are computed from their comparative performance at the surrounding gauges. The MSWEP dataset was validated at the global scale using the conceptual HBV light

rainfall-runoff model with five different precipitation datasets in 9011 catchments (<50000 km^2) across the globe. MSWEP obtained the highest daily correlation coefficient (R) among the five P datasets for 60.0% of the stations and a median R of 0.67 versus 0.44-0.59 for the other datasets and a median calibrated Nash-Sutcliffe efficiency (NSE) of 0.52 versus 0.29-0.39 for other P datasets. In this study we used the version 1.1 of MSWEP, released on August 2nd 2016.

5   **PGFv3**

The PGFv3 product is an improvement over the Princeton University global meteorological forcing (PGF) dataset (Sheffield et al., 2006). In brief, the PGF dataset merges data from the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis (Kalnay et al., 1996) with the Global Precipitation Climatology Project (GPCP; Adler et al. 2003), Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huff-

10   man et al. 2007), observation-based datasets of precipitation, and the Climatic Research Unit (CRU) precipitation dataset (Harris et al., 2013). To ensure adequate representation of rain day anomalies, a correction is applied by resampling daily precipitation data to match the statistics of observation-based precipitation datasets (CRU, GPCP). Downscaling from 2.0° to 1.0° uses a Bayesian probability function considering the fraction of wetted area using a higher resolution dataset (GPCP). The daily precipitation values are also bias-corrected by matching the NCEP-NCAR monthly totals with the monthly values of the

15   CRU dataset. Additional improvements are described in Chaney et al. (2014), performing a spatial downscaling from 1.0° to 0.25° resolution using bilinear interpolation involving 128 stations over Chile from the Global Surface Summary of the Day Version 7 dataset archived at the National Climate Data Center (GSOD, ftp://ftp.ncdc.noaa.gov/pub/data/gsod). The version 3 of PGF used here is finally obtained by merging data on 217 local raingauges (obtained from the Chilean Water Agency - DGA) into the PGF using a Kalman filter approach (Chaney et al., 2014; Peng et al., 2016), for the period 1979-2010. As such,

20   the spatial coverage is greatly improved, as well as the statistical precipitation characteristics (frequency, amount and extreme values) at different temporal scales. The PGF is integrated into the Princeton Latin American Flood and Drought Monitor - LAFDM (http://stream.princeton.edu/) and can be directly downloaded.

Comments on selected SREs

All the aforementioned SREs ingest some type of observed precipitation data in the algorithm used to compute the rainfall

25   estimates, except CMORPH. Among the SREs that use observed precipitation data, we identified two groups: those products that use mostly GPCC observed data to produce the satellite estimates (PERSIANN-CDR, 3B42v7, CHIRPSv2, MSWEPv1.1) and those SREs that use a specific set of Chilean raingauges to bias-correct their satellite estimates (PGFv3 and PERSIANN-CCS-Adj).

Despite that some of the selected SREs products are available at sub-daily resolution (e.g., CMORPH and TMPA 3B42v7),

30   in this study we follow Abera et al. (2016) and only used the daily products for two main reasons: (i) there are no time series of observed precipitation available at sub-daily time scales with enough data length, (ii) pointing at future hydrological simulations for allocation of water resources and drought monitoring tasks, the daily time scale is suitable for capturing the

temporal variation of streamflows at basin scales. Table 1 provides a summary of the satellite-based datasets, including their full names, and spatial and temporal resolutions for the versions used in this study.

3 Assessment of precipitation products

3.1 Methodology for data comparison

5 Following Thiemig et al. (2012), a point-to-pixel analysis was applied to compare time series of data observed at selected raingauges (red circles in Figure 1a) to the corresponding SRE pixel. A comparison at sub-basin spatial scales was not carried out due to the lack of raingauges in the upper Andes, which would involve large uncertainties in any interpolation based on existing point measurements. All SRE products with a spatial resolution higher than 0.25° (i.e., PERSIANN-CCS-adj and CHIRPSv2) were upscaled to a unified grid of 0.25°, in order to enable consistent point-to-pixel comparisons. The upscaling 10 procedure applied in this work consisted in transferring values from the high resolution raster cells to each one of the 0.25° grid cells, by using bi-linear interpolation as implemented in the `resample` function of the `raster` R package (Hijmans, 2016). Considering that the regridding was applied at a daily time scale, where precipitation is assumed to be a smoothly varying variable within each 0.25° grid cell, we consider the bi-linear interpolation an adequate technique with no or low impact on our results.

15 Daily precipitation events were classified and analysed adapting the criteria defined by the World Meteorological Organization (2008), as shown in Table 2, because the same amount of rainfall falling with two different durations will lead to different hydrological processes in a given catchment. Daily observations at the 366 raingauges (see Section 2.2.1) and corresponding satellite estimates were accumulated into monthly, seasonal (DJF, MAM, JJA, SON) and annual values, to assess the accuracy of each precipitation product at different time scales. Based on the availability of satellite data (see Table 1), the evaluation 20 period for this study extends from Jan 2003 to Dec 2010.

3.2 Performance indices

An exhaustive evaluation of the seven SREs described in Section 2.2.2 was carried out, using both continuous and categorical indices of model performance at different temporal scales. Continuous indices are described in Appendix A, and they include the modified Kling-Gupta efficiency (Kling et al., 2012; Gupta et al., 2009) along with its three individual components. The 25 modified Kling-Gupta efficiency (*KGE'*, dimensionless) is a relatively new index to compare observations with estimations, which is used here to decompose the total performance of the SREs into a linear correlation (r, Equation A2), bias (β, Equation A3) and a variability (γ, Equation A4) term. We selected the KGE' because forecasting and hydrological applications generally require that rainfall estimates are able to reproduce the temporal dynamics (measured by r) as well as preserving the volume and distribution of precipitation (measured by β and γ, respectively). The optimum value of *KGE'*, r, β and γ is 30 1.0. The Pearson product-moment correlation coefficient r is a measure of the linear correlation between the observations and satellite values, ranging from +1 (perfect positive correlation) to -1 (perfect negative correlation), with zero values indicating

absence of linear correlation. The main drawback of r is its inability to detect changes in location and scale between the two variables. β measures the average tendency of the satellite values to be larger ($\beta > 1$, overestimation) or smaller ($\beta < 1$, underestimation) than their observed counterparts. γ indicates if the dispersion of the satellite estimates is higher or lower compared to the observations. Using CV_s/CV_o for the computation of the variability ratio, instead of using only the σ_s/σ_o, we ensure that the bias and variability ratios are not cross-correlated (Kling et al., 2012).

Most of the studies assessing SREs against ground observations separate rainfall events into *no rain* and *rain* to evaluate the skill of each SRE in reproducing such events (e.g. Guo et al., 2015; Blacutt et al., 2015; Ward et al., 2011; Scheel et al., 2011). Here, we further subdivide daily rainfall in five types of events, which are used to classify precipitation based on its daily intensity, ranging from *no rain* (*dry day*) (< 1 mm d^{-1}) to *violent rain* (> 40 mm d^{-1}), as shown in Table 2. Five categorical indices, described in Appendix B, were used to assess the ability of each satellite product to correctly identify the five aforementioned categories/classes of daily rainfall events (see Table 2). The *percent correct* (*PC*, Equation B1) is a simple measure that indicates the percent of events and no-events that are correctly identified, ranging from zero (absence of (no-)events correctly identified) to one (all (no-)events correctly identified). The *PC* is not useful for low frequency (extreme) events, because misleading high values of the score are usually obtained due to the high frequency of correct negative (CN) events. This shortcoming is compensated for by the next three scores. The *probability of detection* (*POD*, Equation B2, also known as *hit rate*) and the *false alarm ratio* (*FAR*, Equation B3) measure the fraction of events that are correctly and incorrectly identified by the satellite product, respectively. Both indices, *POD* and *FAR*, range from 0 to 1, but 1 is the optimum value for *POD* while a *FAR=0* indicates that no events are incorrectly identified by the SRE. The *equitable threat score* (*ETS*, Equation B4), also known as *Gilbert's Skill Score* (*GSS*), measures the fraction of observed and/or estimated events that are correctly predicted, adjusted by the frequency of hits that would be expected to occur simply by random chance (for example, it is easier to correctly match rain occurrence in a wet climate than in a dry climate). ETS ranges from -1/3 to 1, being 1 its optimal value and scores below 0 indicate that the chance forecast of the event should be preferred over the actual unskilled SRE value. The *frequency bias* (*fBias*, Equation B5) compares the number of events identified by the satellite product to the number of events that actually occurred at the corresponding raingauge. It is commonly referred to as *bias* when there is no possible confusion with other meanings of the term (not in this article). The optimal value of *fBias* is 1.0 (unbiased), i.e., the event was registered by the SRE the same number of times than at the raingauge, with $fBias > 1$ indicating an overestimation of the occurrences by the SRE, whereas $fBias < 1$ reveals that the event was identified by the SRE less times than it was actually observed at the raingauge.

A well-performing satellite product should have a value of *FAR* close to zero, and values of *KGE'*, *r*, β, γ, *PC*, *POD*, *ETS*, and *fBias* close to 1. All the aforementioned indices of SRE performance were computed based on Jolliffe and Stephenson (2003), using the R environment 3.3.1 (R Core Team, 2015) and the `raster` (Hijmans, 2016), `hydroGOF` (Zambrano-Bigiarini, 2016a) and `hydroTSM` (Zambrano-Bigiarini, 2016b) R packages.

3.3 Uncertainties in the verification results

Rainfall measurements using standard gauges and different types of tipping-bucket and weighing gauges have a limited spatial support and often can not accurately determine precipitation in a certain location, because of the impact of wind, flawed installation, wetting losses, evaporation and other random and systematic errors (e.g. Sevruk et al., 2009; Ren and Li, 2007; Sevruk and Chvíla, 2005; Legates and DeLiberty, 1993). Other uncertainties related to precipitation in high-elevation areas might be attributed to missing snow monitoring equipment (i.e., snow scale or pillow to determine the snow-water equivalent), reducing accuracy of precipitation measurements during the winter season. Finally, station density might not be high enough to capture the spatial variability of areal rainfall (e.g. Emmanuel et al., 2015), and additional uncertainties arise from the comparison of the satellite *snapshot* data to the total accumulated over a certain period in the raingauge (Ebert, 2007). Therefore, there are large uncertainties inherent in the ground data used for verification (Scheel et al., 2011), and a method to appropriately take into account errors in the ground observations to then quantify uncertainties in the verification results is still a challenge for the scientific community (Ebert, 2007). We agree with Scheel et al. (2011) in that validated SRE products should not replace gauge-based observations but provide complementary information.

4 Results

4.1 Spatial variability of SRE performance

We computed spatial maps showing the values of continuous indices of SRE performance for each of the 366 selected stations and for different time scales (daily, monthly, annual, four seasons), resulting in 196 different figures of the Chilean territory (7 SREs x 7 time scales x 4 indices of model performance). We put special emphasis on showing SRE performance at different elevation zones and latitudes. Figure 4 is one example of the aforementioned maps, showing the modified Kling-Gupta efficiency (KGE') with CHIRPSv2 estimates compared against observed precipitation at a monthly time scale. It showed a good performance ($KGE' > 0.75$) in low and mid-elevation areas (0-1000 m a.s.l.) of Central and Southern climate zones (32.18-43.70°S) and acceptable performance in the Near North area (26.0-32.18°S). However, it showed a poor performance ($KGE' < 0$) for the high lands of the Andean Plateau (*Altiplano*), specifically between 2000-3500 m a.s.l. To disentangle the origin of differences between observed precipitation and CHIRPSv2 estimates, Figures S1 to S3 (in the supplementary material) show the spatial variation of the individual components of monthly KGE'. First, Figure S1 showed a good linear correlation ($r \geq 0.75$) in most of the Chilean territory, with a few exceptions in the Far North area (17.5-26.0°S). Secondly, Figure S2 revealed that in most of the territory the satellite estimates present a low bias ($0.75 \leq \beta \leq 1.25$), with exception of stations located in the Far North, presenting a large overestimation. In particular, over the Andean Plateau satellite estimates were as high as 94 times the observed precipitation (note that average annual values for those stations is, in general, lower than 1 mm). Finally, Figure S3 showed that CHRIPS slightly underestimated the variability of observed monthly precipitation ($0.75 \leq \gamma \leq 1$) of almost all the stations, with larger values only at the high-elevation zones (2000-4600 m a.s.l.) of the Far

North. Figures similar to 4 and S1 to S3 can be found on the supplementary material (Zambrano-Bigiarini et al., 2016) for the other six SREs and time scales.

As a way of summarizing the findings for different macro-climatic areas, Figure 5 shows boxplots with the modified Kling-Gupta efficiency between different monthly SREs and their corresponding observations. It clearly illustrates that the best performance for all SREs but CMORPH was obtained in the South (36.4-43.7°S), followed by the Central (32.18-36.4°S) and Austral areas (43.7-56°S), with PGFv3 as the best product followed closely by CHIRPSv2, MSWEPv1.1 and 3B42v7, while CMORPH presented the lowest median performance in those climatic areas. On the other hand, all SREs except PERSIANN-CDR presented an acceptable performance in the Near North (26.0-32.18°S) and a poor performance in the Far North (17.5-26.0°S). Similarly to the previous figure, but focusing on elevation zones instead of macro-climatic areas, Figure 6 shows that the best performance for all SREs was obtained in low- and mid-elevation zones (0-1000 m a.s.l.). On the other hand, all SREs except PGFv3 performed poorly for the higher elevations, in particular between 2000 and 3500 m a.s.l. The best SRE product was PGFv3, followed closely by CHIRPSv2, 3B42v7 and MSWEPv1.1, while CMORPH presented the lowest median performance in those elevation areas (Figure 6). Figure 11 compares monthly and annual time series of precipitation, as estimated by the seven SRE products used in this work against the observed values register at the corresponding raingauge at selected sites for the Far North, Near North, Centre, South and Far South climate macrozones.

4.2 Temporal variability of SRE performance

Figures 7 to 10 show boxplots summarizing the modified Kling-Gupta efficiency (KGE') and its individual components of performance (r, β, γ) for each satellite product and for different time scales.

Figure 7 shows the Pearson product-moment correlation coefficient (r) between different satellite estimations and the observations at the corresponding grid cell, for seven different temporal scales: daily, monthly, annual, summer (DJF), autumn (MAM), winter (JJA), and spring (SON). It depicts that correlations are positive, with most daily values ranging from 0.2 to 0.6, and achieving higher median values at monthly (0.60-0.95) than at annual (0.3-0.9) time scales. In general, all the products presented the highest correlation during autumn (MAM) and the lowest values during summer (DJF) and spring (SON). The PGFv3 product was an outlier, which exhibited median correlation values larger than 0.9 for all time scales, followed by MSWEPv1.1, CHIRPSv2, 3B42v7 and PERSIANN-CDR.

Figure 8 depicts the bias ratio of the modified Kling-Gupta efficiency (β) for different satellite estimations against observations at the corresponding grid cell, and for the same temporal scales used before. PGFv3, CHIRPSv2 and 3B42v7 were almost unbiased for all time scales and seasons except for the summer (DJF), for which almost all the products overestimated P compared to the observed precipitation. MSWEPv1.1 and PERSIANN-CDR tended to overestimate P compared to the observed precipitation for all time scales, while CMORPH and PERSIANN-CCS-Adj showed a general underestimation thereof.

Figure 9 shows the variability ratio of the modified Kling-Gupta efficiency (β) between SREs and the observations. Most of the products resulted in underestimating the variability of P for all time scales, with exception of CMORPH and PERSIANN-CCS-Adj which overestimated the observed variability at annual and winter (JJA) time scales. 3B42v7 and PGFv3 are the

products that best resembled the variability of observed precipitation, while PERSIANN-CDR is the one with the largest median underestimation at all time scales.

Figure 10 summarizes the three previous figures into one, showing boxplots of KGE' for all the satellite products and time scales. Most of the SREs, except PGFv3, presented a limited overall performance at daily time scales (median values of $KGE' \leq 0.3$), which improved when aggregated into monthly values (median values of $KGE' \geq 0.5$). PGFv3 and CMORPH were the products with the best and worst performance for almost all scales, while 3B42v7, CHIRPSv2 and MSWEPv1.1 showed the second best performance. At the seasonal scale all the products performed best during autumn (MAM) and winter (JJA), and showed the worst performance during spring (SON) and summer (DJF), with PERSIANN-CDR and MSWEPv1.1 being the worst.

4.3 SRE performance for different precipitation intensities

Figure 12 shows the median values of the five categorical indices of performance computed between different SREs and the observations at the corresponding grid cell, for the five classes of daily rainfall intensity defined in Table 2. First, all SREs presented high values of percent of correct ($PC > 0.78$) in all the precipitation intensities, which is particularly clear for intensities larger than 20 mm d^{-1}, where PC was larger than 0.97 for all SREs. As mentioned in Section 3.2, high values of PC might be misleading due to the high frequency of correct negative (CN) events. Therefore, looking at the other indices of model performance, we observed that most of the products showed a low probability of detection (POD), in general with values lower than 0.4 for all the rainfall events except by *no rain* ([0, 1) mm d^{-1}), which was well captured by all SREs ($POD >= 0.79$). PGFv3 was the product that best captured all classes of rainfall events, while CMORPH resulted in the worst performance for events larger than 5 mm d^{-1}, with almost no skill in detecting rainfall events larger than 20 mm d^{-1}. The false alarm rate (FAR) showed almost the opposite view of the POD results, with most of the SREs with good values of FAR for no rain events and poor performance for intensities larger than 1 mm d^{-1}. However, FAR was less skillful than POD in differentiating the performance of SREs for intensities lower than 20 mm d^{-1}, while it provided a better separation of performance for high daily intensities (> 40 mm d^{-1}). The equitable threat score (ETS) showed that most of the SRE were not able to capture the fraction of observed and/or estimated rainfall events (in general, $ETS < 0.3$), with PGFv3 showing the highest performance for all rainfall intensities. Finally, the frequency bias ($fBIAS$) panel showed that the amount of *no rain* events identified by most of SREs was in excellent agreement with the corresponding observed amount (median $fBIAS \sim 1$). All SREs overestimated the amount of light rain events ([1, 5) mm d^{-1}) and underestimated the number of violent events (> 40 mm d^{-1}). In between, PERSIANN-CDR, MSWEPv1.1, CHIRPSv2, 3B42v7 and PERSIANN-CCS-Adj overestimated moderate rainfall events ([5, 20) mm d^{-1}), while CMORPH underestimated its amount and PGFv3 provided an unbiased estimation of the number of this type of events. The amount of heavy rainfall events ([20, 40) mm d^{-1}) was in general well captured by all the SREs but CMORPH and PERSIANN-CDR, which underestimated it. The supplementary material (Zambrano-Bigiarini et al., 2016) presents detailed boxplots for each categorical performance index used to identify the capabilities of SREs to capture the five types of precipitation events described in Table 2.

Based on the previous analysis, we agree with Hossain and Huffman (2008) that the probability of rain detection (POD_{rain}) and no-rain detection (POD_{norain}) are important measures to quantify the ability of a SRE to detect the magnitude of the rainfall rate. However, we propose to use POD_{rain} for assessing different rainfall intensities and not only to know that *it was raining* when the reference observation is larger than a user-defined threshold t_0.

5 Discussion

This work aimed at providing, for the first time, insights about the performance of seven state-of-the-art SREs (3B42v7, CHIRPSv2, CMORPH, PERSIANN-CDR, PERSIAN-CCS-adj, MSWEPv1.1 and PGFv3) at different temporal scales (daily, monthly, seasonal, annual) over the Chilean territory, using 366 stations located from sea level to 4600 m a.s.l. in the Andean Plateau.

5.1 Overall performance of the evaluated SREs

As expected, the two products that used a Chilean dataset for calibration showed a good performance at all time scales and nation-wide, with PGFv3 showing an outstanding performance in comparison to all SREs, while PERSIANN-CCS-adj had a performance more similar to the other SREs (Figure 10).

Overall, PGFv3 was the best performing product in terms of KGE' and its individual components (see Figures 7 to 10), which was expected due to the use of 217 local raingagues in the bias correction procedure used to adjust its estimates. PGFv3 was followed closely by CHIRPSv2, 3B42v7 and MSWEPv1.1, all of them using the GPCC dataset to calibrate their precipitation estimates, not any local dataset. Generally, SREs performed better for wetter periods (i.e., MAM-JJA, Figure 10) and southern and central regions (Figure 5). All SREs except PGFv3 performed poorly for the higher elevations, in particular between 2000 and 3500 m a.s.l. (Figure 6) and also for the Far North desert region (Figure 5). These findings are in line with the review on globally varying performance of SRE products by Maggioni et al. (2016).

5.2 Which SRE performs best across the topographic and climate gradient in Chile?

Mountainous regions pose important challenges across all seasons to satellite estimates derived from both TIR and PM observations (Tian and Peters-Lidard, 2010; Dinku et al., 2010; Scheel et al., 2011; Mantas et al., 2015). Figure 6 illustrated that all SREs except PGFv3 performed poorly for the higher elevations, in particular between 2000 and 3500 m a.s.l. However, for the 23 stations located in the highest elevations (3500-4600 m a s.l.) results are better, especially for PERSIANN-CCS-Adj and PGFv3 which can be explained by the use of Chilean datasets in the bias-correction procedure of these products. On the other hand, an extreme variation was observed for the results along the latitudinal gradient, covering eleven types of climate, from hot dessert to polar/tundra (Figure 1). Figure 5 showed that all SRE perform best in the humid South (36.4-43.7°S), followed by Central Chile (32.18-36.4°S) and Far South (43.7-56.0°S), compared to the arid northern regions, in particular in the desert Far North where satellite estimates were as high as 94 times the observed precipitation (e.g., Figure S2 in the supplementary material for CHIRPSv2 and Figure 11a). This can be attributed to the very low precipitation values in that area

(some raingauges have annual values lower than 1 mm/year), where one single event not correctly identified can be responsible of 100% of over or underestimation. This overestimation has been observed as well in the scope of other studies (Dinku et al., 2010; Maggioni et al., 2016). Time series comparison between raingauges and SRE data can be found in the supplementary material (Zambrano-Bigiarini et al., 2016) for each of the 366 selected stations described in Section 2.2.1 (see also comparison of selected stations in Figure 11).

It is worth mentioning that the number of stations in each macroclimate zone (Figure 5) and elevation range (Figure 6) was not the same, which hampered an unbiased comparison among different climate and elevation zones. A higher number of stations used in our analysis was located below 1000 m a.s.l (276 out of 366) and in Central and Near South of Chile (223 out of 366). In particular, Figure 4 showed that stations with high number of days with information and located above 2000 m a.s.l. were concentrated in the Far North, making clear the need of additional monitoring stations in elevations above 2000 m a.s.l from the Near North to the southern extreme of Chile.

5.3 Which SRE performs best for different time scales?

The PGFv3 exhibited the best KGE' performance for all time scales, followed by CHIRPSv2, 3B42v7 and MSWEPv1.1, while the worst performing were CMORPH, PERSIANN-CCS-Adj and PERSIANN-CDR (Figure 10). In line with previous work by Scheel et al. (2011) and Pereira Filho et al. (2010), better performance was obtained for monthly time scales compared to the daily scale. Unexpected, however, was the fact that the annual scale resulted worse compared to the daily scale, in particular for CMORPH and PERSIANN-CCS-Adj, which was due to the amplification of small systematic biases at the daily time scale when aggregated into the annual one, passing from 2922 days to only 8 annual values. Figure 10 illustrated that most SREs showed a distinct behaviour during different seasons, performing better during the more humid seasons (MAM-JJA) than during the dry ones (SON-DJF). However, while PGFv3, CHIRPSv2, 3B42v7 and CMORPH presented minor differences in performance along the seasons, MSWEPv1.1, PERSIANN-CDR and PERSIANN-CCS-Adj showed large seasonal differences and large spread of performance during summer (DJF) and spring (SON).

5.4 How does the accuracy of a given SRE change for different precipitation intensities?

Evaluating the SRE performance for different precipitation intensities resulted in low values of most categorical indices. This was partly due to the demanding classification criteria: we used 5 types of precipitation intensities instead of the typical two classes (rain/no-rain) (e.g. Dinku et al., 2010; Scheel et al., 2011). All SREs were able to correctly identify the occurrence of no rain events, but during rainy days they presented a low skill in providing an accurate classification of precipitation intensities (Figure 12). On the other hand, all SREs underestimated the number of violent rain events (> 40 mm d^{-1}) and overestimated the amount of light rain events ($[1, 5)$ mm d^{-1}) (Figure 12). In between, PGFv3 provided an unbiased estimation of the number of moderate rainfall events,$[5, 20)$ mm d^{-1}, while CMORPH underestimated that amount and PERSIANN-CDR, MSWEPv1.1, CHIRPSv2, 3B42v7 and PERSIANN-CCS-Adj overestimated it. The number of heavy rainfall events ($[20, 40)$ mm d^{-1}) was well captured by all the SREs but CMORPH and PERSIANN-CDR, which largely underestimated its amount.

5.5 How well does the aggregated KGE and its components evaluate SRE performance?

Modified Kling-Gupta efficiency (KGE', Kling et al. 2012; Gupta et al. 2009) proved to be a useful index to assess the performance of SRE products, because it captures in a single number a pseudo multi-objective evaluation of the linear correlation, bias and variability of the satellite estimations compared to its corresponding observations at raingauges.

Figure 7 showed that linear correlation (r) between SRE values and their corresponding observations were in general acceptable, with higher values at monthly times scales (0.60-0.95) than at annual (0.3-0.9) and daily (0.35-0.9) ones. In general, all the SREs presented the highest correlation during autumn (MAM) and the lowest values during summer (DJF) and spring (SON). Figure 8 showed that PGFv3, CHIRPSv2 and 3B42v7 were almost unbiased (β) for all time scales and seasons except for the summer (DJF), where almost all the products overestimated P. CMORPH and PERSIANN-CCS-Adj showed a general underestimation of P, while MSWEPv1.1 and PERSIANN-CDR tended to overestimate it. The aforementioned underestimation of CMORPH was in agreement with previous studies (e.g. Abera et al., 2016; Ringard et al., 2015; Lo Conti et al., 2014; Dinku et al., 2010) as well as the acceptable linear correlation (Abera et al., 2016), in particular at monthly time scales. The fact that CMORPH presented the highest underestimation of rainfall for almost all time scales very likely is because CMORPH does not use any observed precipitation data to compute its estimates, in contrast to all the other analysed SREs. Figure 9 illustrated that most of the products underestimated the variability of P for all time scales, with exception of CMORPH and PERSIANN-CCS-Adj which overestimated the observed variability at annual and winter (JJA) time scales. PGFv3 and 3B42v7 were the products that best captured the variability of observed precipitation, while PERSIANN-CDR presented the largest median underestimation at all time scales. From the analysis described above, it is clear that providing a better representation of the variability of observed precipitation should be a major concern in future releases of the analysed SREs. In particular, PGFv3, CHIRPSv2 and 3B42v7 which presented high linear correlation and were almost unbiased for monthly and annual scales.

Based on the results presented in Section 4.2, we consider that the combined use of the modified Kling-Gupta efficiency (KGE') and its three components (r, β, γ) is suitable for identifying if a SRE is able to capture the temporal dynamics (r) as well as preserving the total amount and distribution of precipitation (β and γ, respectively). For example, two SREs with the same KGE' value might have different values of r and β, however reducing a systematic bias in the SRE is easier to correct (e.g., using a bias correction method) than improving the linear correlation between the observed and SRE value of precipitation.

5.6 Is there any SRE that performs best compared to all the others, everywhere and for all time scales?

PGv3F vas the best product for all time scales, macro-climatic areas and elevation zones, which was expected because this product used 217 raingauges to bias-correct the original PGFv3 estimates using a Kalman filter approach. Therefore, it is not entirely fair to compare PGFv3 to other products that did not use local data for calibration. However, PGFv3 was included as potentially useful for some hydrological applications in the Chilean territory. Interesting is the case of PERSIANN-CCS-Adj, which also used raingauge data, but which was unable to obtain results similar to those of PGFv3. The latter might be due to the bias-correction of the original PERSIANN-CCS estimates using only climatology (Yang et al., 2016)). However,

PERSIANN-CCS-adj is still under development through improved QM-GW calibration procedures using monthly cumulative distribution functions, as well as a real-time assimilation of precipitation measurements to improve performance at daily and monthly timescales in the near future.

6   Conclusions

Satellite-based rainfall estimates (SRE) provide an unprecedented opportunity to be applied in a wide range of meteorological and hydrological applications. Despite that most of the existing products are continuously improving their algorithms and data sources to adapt to particular environments, catchment-specific studies should still be carried out before any hydrological application thereof.

In this article, the performance of seven state-of-the-art SRE datasets (3B42v7, CHIRPSv2, CMORPH, PERSIANN-CDR, PERSIAN-CCS-adj, MSWEPv1.1, PGFv3) was compared against observations from 366 raingauges over the Chilean territory, which is extremely challenging due to its complex topography (elevations range from sea level to 6893 m a.s.l.) and the existence of 11 different types of climates (from hot dessert to polar/tundra) along the large latitudinal extent of 4300 km. Seven different temporal scales relevant for hydrological applications were addressed in the evaluation: daily, monthly, annual, DJF (summer), MAM (autumn), JJA (winter), SON (spring). The modified Kling-Gupta efficiency (KGE') was used as continuous index of performance, along with its three individual components (linear correlation, bias and variability), to identify possible sources of systematic errors in each SRE. In addition, five categorical indices (PC, POD, FAR, $fBias$, ETS) were used to assess the ability of each SRE to correctly identify different intensities of daily precipitation.

Results revealed that the best performing SRE product was PGFv3, calibrated with a Chilean dataset, followed closely by CHIRPSv2, 3B42v7 and MSWEPv1.1, while CMORPH presented the lowest median performance everywhere, which was expected because it is the only product that does not use any type of measured rainfall for calibration. All SREs except PGFv3 performed poorly for the higher elevations, in particular between 2000 and 3500 m a.s.l. and also for the extreme Northern desert region. Surprisingly, all products performed worse for the annual scale compared to the monthly scale and even daily scale.

The following paragraphs summarise the key findings:

– Most SRE products performed the best for the humid South (36.4-43.7°S) and Central Chile (32.18-36.4°S), in particular at low- and mid-elevation zones (0-1000 m a.s.l.), compared to the arid northern regions and the Far South.

– Most SREs performed worst in the high elevation areas ($\geq$ 2000 m a.s.l.) of the hyper-arid Far North (17.5-26.0°S).

– All SRE present positive values of linear correlation with ground observations, with best median values at monthly time scales ($0.4 \leq r \leq 0.95$) than at daily and annual scales. Correlations are also higher during autumn (MAM) and lower during summer (DJF) and spring (SON).

- PGFv3, CHIRPSv2 and 3B42v7 are almost unbiased at all time scales (except summer). MSWEPv1.1 and PERSIANN-CDR tend to overestimate P compared to observations, while CMORPH and PERSIANN-CCS-Adj tend to underestimation at all time scales.

- All SREs underestimated the observed variability of P at all time scales, except for CMORPH and PERSIANN-CCS-Adj which overestimated P at the annual scale and winter (JJA). 3B42v7 and PGFv3 are the products that best captured the observed variability of P, while PERSIANN-CDR presented the poorest performance.

- KGE' and its three individual components are recommended for further comparisons, because they do not only provide an overall assessment of the SRE performance, but also allowed us to understand if mismatches are due to errors in the representation of shape, magnitude or variability of the observed precipitation. If detected errors are mostly due to under or over-estimation of observed precipitation, it is very likely that applying a bias-correction procedure will significantly improve the performance of the product (e.g., PGFv3). Providing a better representation of the variability of observed precipitation should be a major concern in future releases of the analysed SREs, in particular for PGFv3, CHIRPSv2 and 3B42v7 which presented high linear correlations with observed P and were almost unbiased at monthly and annual time scales.

- All SREs performed best in terms of KGE' during the wet seasons autumn and winter (MAM-JJA) compared to summer (DJF) and autumn (SON).

- Overall, the best SRE product was PGFv3 followed by CHIRPSv2, 3B42v7 and MSWEPv1.1, which is expected because PGFv3 used 217 Chilean raingauges in its bias-correction procedure used to obtain the final estimations.

- After evaluating five different categorical indices of performance (PC, POD, $fBias$, FAR and ETS), we recommend the use of POD and $fBias$ to assess the ability of a SRE product to capture different rainfall intensities.

- All SREs were able to correctly identify the occurrence of no rain events (i.e., dry days). However, during rainy days they presented a low skill in providing an accurate classification of different precipitation intensities.

- All SREs overestimated the occurrence of light rain events ([1, 5) mm d^{-1}), while underestimated the amount of violent rain events (> 40 mm d^{-1}). High precipitation events were well captured by all products.

- In general, CMORPH, PERSIANN-CDR and PERSIANN-CCS-Adj were considered less suitable to be used in hydrological applications in Chile at this point in time, due to their large biases when compared to observed precipitation records. Although the PERSIANN-CCS-Adj has the benefit of a real-time product, it has only been bias-corrected using climatology, which explains its lower performance compared to products incorporating gauge observations in its calibration (3B42v7, CHIRPSv2, MSWEPv1.1, PGFv3).

- Lack of raingauges at higher elevation zones (over 2000 m a.s.l.) over most of the Chilean territory (south of 26.0°S), prevented from an exhaustive assessment of SRE products in such areas. The same applied for mid-elevation zones (200-1000 m a.s.l.) in the Far North (17.0-26.0°S) and Far South (southern to 43.7°S).

Outlook

Despite continuous improvements of most SRE products, a site-specific assessment is still recommended before any use in hydrological studies. Different types of mismatches between SREs and ground observations might be reduced by using local observations to calibrate the satellite estimates. This was evidenced by PGFv3, which showed the best performance compared
5 to the other products due to the statistical merging of 217 raingauges, highlighting the need for adequate observation networks to complement the SRE products. This is especially valid for high elevation areas, where only limited precipitation records are available, and where well-calibrated SRE products could provide the highest added-value.

To conclude, the PGFv3 is the best performing product for Chile at all time scales and locations tested, but it is currently only available up to 2010, making it a relevant dataset for water balance estimations but not for real-time applications. Nevertheless,
10 it is expected that PGFv3 will be used to bias-correct the 3B42v7 as part of the Latin American Flood and Drought Monitor. Since the PGFv3 dataset is performing well for all the indices tested, it is expected that the calibrated 3B42v7 will also improve its performance. As this process is still under development, it could not be tested in this study and will become the subject of future research. For (quasi)real-time and hydrological applications, we suggest to consider the use of CHIRPSv2 , because it has a long data record (1981-present), a low latency (1 day-3 weeks, depending on the product), and high spatial resolution
15 (0.05° instead of the 0.25° of most SREs). MSWEPv1.1 is also a promising product, and new versions are expected to be released at least once a year. The next version of MSWEPv1.1 is anticipated to incorporate corrections for drizzle and provide estimates for the entire globe, including ocean areas (Hylke Beck, personal communication). The 3B42v7 has a lower spatial resolution and covers a shorter period (1998-present), but showed a very similar performance to CHIRPSv2 across all indices. The 3B42 dataset will be superseded by the Global Precipitation Measurement (GPM) mission product IMERG, launched by
20 NASA and the Japanese Aerospace Exploration Agency (JAXA) on February 27th 2014. To overcome the short data length of the new IMERG dataset, it is planned to provide a retrospectively processed IMERG datasets from 1998 onward, which are expected to be released in early 2017 (Huffman, 2015). The real-time precipitation product PERSIANN-CCS-Adj is currently being enhanced through real time merging of station data. The first results indicate that this procedure will significantly improve the performance of the product over Chile (Yang et al., 2016), which will be evaluated in future research.
25 Given the high dependency of most SRE products on the GPCC dataset for calibration, the authors would like to recommend to the Chilean authorities in charge of the collection and analysis of meteorological data (e.g., DGA and DMC) to make additional efforts to share their data with GPCC, in order to improve the performance of SREs that ingest observed precipitation data.

Appendix A: Continuous indices of model performance

30 $$KGE = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{A1}$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}} \tag{A2}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{A3}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{A4}$$

where N is the number of observations; O_i and S_i are the observed and the corresponding satellite precipitation values at day i, respectively; $\overline{O}$ and $\overline{S}$ are the arithmetic mean of the observations ans satellite estimates, respectively; O_{max} and O_{min} are the maximum and minimum observed value, respectively.

Appendix B: Categorical indices of model performance

$$PC = \frac{H + CN}{Ne} \tag{B1}$$

$$POD = \frac{H}{H + M} \tag{B2}$$

$$FAR = \frac{FA}{H + FA} \tag{B3}$$

$$ETS = \frac{H - H_e}{(H + F + M) - H_e} \tag{B4}$$

$$fBias = \frac{H + F}{H + M} \tag{B5}$$

where Ne is the total number of verification points (number of events), H indicates a *hit*, i.e., a satellite estimate that correctly identifies the type of rainfall event measured at the raingauge; M is a *miss*, i.e., an event recorded at the raingauge but not correctly identified by the satellite product; F is a *false alarm*, i.e., a rainfall event detected by the satellite product but not recorded at the corresponding raingauge; CN represents the *correct negatives* (or *correct rejections*),i.e., an event not registered either by the raingauge or the satellite; and H_e indicate a *hit that could occur by chance*, computed as $H_e = (H + M)(H + F)/Ne$. Table 3 summarises the aforementioned concepts used to compute the different categorical indices of performance.

Acknowledgements. The authors thank FONDECYT 11150861 "Understanding the relationship between the spatio-temporal characteristics of meteorological drought and the availability of water resources, by using satellite-based rainfall and snow-cover data. A case study in a data-scarce Andean Chilean catchment" for financial support to the main author; the FONDAP 15110009 for providing the precipitation dataset

used in this study; the Center for Hydrometeorology and Remote Sensing at the University of Irvine for providing the PERSIANN-CCS-adj and PERSIAN-CDR datasets, the Terrestrial Hydrology Research Group at Princeton University for providing the PGFv3 datasets; Thomas Cram (NCAR/CISL/DSS) for technical support about CMORPH data, Daniel Wunder (NOAA) for technical support related to PERSIANN-CDR, and Hylke Beck for providing an updated version 1.1 of MSWEP data. The main author is also grateful to the active R community for unselfish and prompt support, in particular to Robert J. Hijmans for developing and maintaining the `raster` R package. In the framework of this special issue, the authors also wish to thank Eric Wood, Justin Sheffield and Liqing Peng from the Terrestrial Hydrology Research Group at Princeton University for the local calibration of the PGFv3 dataset for Chile and its integration into the Latin American Flood and Drought Monitor. Finally, the authors thank V. Maggioni, C. Massari and another anonymous referee which constructive comments helped to improve the quality of the final manuscript.

References

Abera, W., Brocca, L., and Rigon, R.: Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin, Atmospheric Research, doi:10.1016/j.atmosres.2016.04.017, 2016.

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

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities, Reviews of Geophysics, 53, 452–480, doi:10.1002/2014RG000456, 2015.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P.: PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies, Bulletin of the American Meteorological Society, 96, 69–83, doi:10.1175/BAMS-D-13-00068.1, 2015.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences Discussions, p. 1, doi:10.5194/hess-2016-236, 2016.

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K., and Sorooshian, S.: Hydrologic evaluation of satellite precipitation products over a mid-size basin, Journal of Hydrology, 397, 225–237, doi:10.1016/j.jhydrol.2010.11.043, 2011.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., and de Roo, A.: Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions, Journal of Hydrology: Regional Studies, 8, 112–129, doi:10.1016/j.ejrh.2016.09.003, 2016.

Blacutt, L. A., Herdies, D. L., de Gonçalves, L. G. G., Vila, D. A., and Andrade, M.: Precipitation comparison for the CFSR, MERRA, TRMM3B42 and Combined Scheme datasets in Bolivia, Atmospheric Research, 163, 117–131, doi:10.1016/j.atmosres.2015.02.002, 2015.

Boisier, J. P., Rondanelli, R., Garreaud, R. D., and Muñoz, F.: Anthropogenic and natural contributions to the Southeast Pacific precipitation decline and recent megadrought in central Chile, Geophysical Research Letters, 43, 413–421, doi:10.1002/2015GL067265, 2016.

Ceccherini, G., Ameztoy, I., Hernández, C., and Moreno, C.: High-Resolution Precipitation Datasets in South America and West Africa based on Satellite-Derived Rainfall, Enhanced Vegetation Index and Digital Elevation Model, Remote Sensing, 7, 6454–6488, doi:10.3390/rs70506454, 2015.

Chaney, N. W., Sheffield, J., Villarini, G., and Wood, E. F.: Development of a High-Resolution Gridded Daily Meteorological Dataset over Sub-Saharan Africa: Spatial Analysis of Trends in Climate Extremes, Journal of Climate, 27, 5815–5835, doi:10.1175/JCLI-D-13-00423.1, 2014.

Chen, S., Hong, Y., Gourley, J. J., Huffman, G. J., Tian, Y., Cao, Q., Yong, B., Kirstetter, P.-E., Hu, J., Hardy, J., Li, Z., Khan, S. I., and Xue, X.: Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States, Water Resources Research, 49, 8174, doi:10.1002/2012WR012795, 2013.

Climate Hazards Group: CHIRPSv2.0 webpage, http://chg.ucsb.edu/data/chirps/, doi:10.15780/G2RP4Q.

CPC-NCEP-NWS-NOAA-USDC: NOAA CPC Morphing Technique (CMORPH) Global Precipitation Analyses, Tech. rep., Boulder CO, doi:10.5065/D6CZ356W, [Last Accessed: 25.Jan.2016], 2011.

Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J. C., Wieringa, J. J., Ramesh, B. R., Stévart, T., and Couvreur, T. L. P.: Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics, Global Ecology and Biogeography, 2016.

Demaria, E. M. C., Rodriguez, D. A., Ebert, E. E., Salio, P., Su, F., and Valdes, J. B.: Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach, Journal of Geophysical Research: Atmospheres, 116, 1–13, doi:10.1029/2010JD015157, 2011.

DGA: Atlas del Agua 2016, Santiago, Chile, available on line at http://www.dga.cl/atlasdelagua/. [Accessed on 29-Aug-2016], 2016.

Dinku, T., Connor, S. J., and Ceccato, P.: Comparison of CMORPH and TRMM-3B42 over Mountainous Regions of Africa and South America, in: Satellite Rainfall Applications for Surface Hydrology, edited by Gebremichael, M. and Hossain, F., pp. 193–204, Springer Netherlands, doi:10.1007/978-90-481-2915-7_11, 2010.

Ebert, E. E.: Methods for verifying satellite precipitation estimates, in: Measuring precipitation from space, pp. 345–356, Springer, 2007.

Ebert, E. E., Janowiak, J. E., and Kidd, C.: Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models, Bulletin of the American Meteorological Society, 88, 47–64, doi:10.1175/BAMS-88-1-47, 2007.

Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., and Payrastre, O.: Influence of rainfall spatial variability on rainfall-runoff modelling: Benefit of a simulation approach?, Journal of Hydrology, 531, 337–348, doi:10.1016/j.jhydrol.2015.04.058, 2015.

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., and Willmott, C. J.: Uncertainties in Precipitation and Their Impacts on Runoff Estimates, Journal of Climate, 17, 294–304, doi:10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2, 2004.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes, Sci Data, 2, 150 066, doi:10.1038/sdata.2015.66, 2015.

Gebregiorgis, A. S. and Hossain, F.: Understanding the Dependence of Satellite Rainfall Uncertainty on Topography and Climate for Hydrologic Model Simulation, IEEE Transactions on Geoscience and Remote Sensing, vol. 51, issue 1, pp. 704-718, 51, 704–718, doi:10.1109/TGRS.2012.2196282, 2013.

Guo, H., Chen, S., Bao, A., Hu, J., Gebregiorgis, A., Xue, X., and Zhang, X.: Inter-Comparison of High-Resolution Satellite Precipitation Products over Central Asia, Remote Sensing, 7, 7181–7212, doi:10.3390/rs70607181, 2015.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Habib, E., Henschke, A., and Adler, R. F.: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA, Atmospheric Research, 94, 373–388, doi:10.1016/j.atmosres.2009.06.015, 2009.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, doi:10.1002/joc.3711, 2013.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling, https://CRAN.R-project.org/package=raster, r package version 2.5-8, 2016.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, International Journal of Climatology, 25, 1965–1978, doi:10.1002/joc.1276, 2005.

Hong, Y., Hsu, K.-L., Sorooshian, S., and Gao, X.: Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system., Journal of Applied Meteorology, 43, 1834–1853, 2004.

Hong, Y., Adler, R. F., Negri, A., and Huffman, G. J.: Flood and landslide applications of near real-time satellite rainfall products, Natural Hazards, 43, 285, doi:10.1007/s11069-006-9106-x, 2007.

Hong, Y., Adler, R. F., Huffman, G. J., and Pierce, H.: Applications of TRMM-Based Multi-Satellite Precipitation Estimation for Global Runoff Prediction: Prototyping a Global Flood Modeling System, in: Satellite Rainfall Applications for Surface Hydrology, pp. 245–265, Springer Netherlands, doi:10.1007/978-90-481-2915-7_15, 2009.

Hossain, F. and Huffman, G. J.: Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales, Journal of Hydrometeorology, 9, 563–575, doi:10.1175/2007JHM925.1, 2008.

Hsu, K.-L., Gao, X., Sorooshian, S., and Gupta, H. V.: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, Journal of Applied Meteorology, 36, 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.

Huffman, G. J.: The Transition in Multi-Satellite Products from TRMM to GPM (TMPA to IMERG), Tech. rep., NASA, available online at https://pmm.nasa.gov/sites/default/files/document_files/TMPA-to-IMERG_transition.pdf. [Last accessed 20-Aug-2016], 2015.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, Journal of Hydrometeorology, 8, 38, doi:10.1175/JHM560.1, 2007.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J.: The TRMM Multi-Satellite Precipitation Analysis (TMPA), in: Satellite Rainfall Applications for Surface Hydrology, edited by Gebremichael, M. and Hossain, F., pp. 3–22, Springer Dordrecht Heidelberg, London New York, doi:10.1007/978-90-481-2915-7_1, 2010.

INE: Medio Ambiente - Informe Anual 2015, Chile, iSBN: 978-956-323-173-1. Available on line at http://www.ine.cl/canales/chile_estadistico/estadisticas_medio_ambiente/2015/informe-medio-ambiente2015.pdf. [Accessed on 29-Aug-2016], 2015.

Janowiak, J. E., Kousky, V. E., and Joyce, R. J.: Diurnal cycle of precipitation determined from the CMORPH high spatial and temporal resolution global precipitation analyses, Journal of Geophysical Research: Atmospheres, 110, 23 105, doi:10.1029/2005JD006156, 2005.

Jarvis, A., Reuter, H., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org, [Online; last accessed May-2012], 2008.

Jolliffe, I. T. and Stephenson, D. B., eds.: Forecast verification: A practitioner's guide in atmospheric science, John Wiley & Sons Ltd, England, 2003.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution., Journal of Hydrometeorology, 5, 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–472, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kang, Y., Khan, S., and Ma, X.: Climate change impacts on crop yield, crop water productivity and food security – A review, Progress in Natural Science, 19, 1665–1674, doi:10.1016/j.pnsc.2009.08.001, 2009.

Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policelli, F., Habib, S., and Irwin, D.: Satellite Remote Sensing and Hydrologic Modeling for Flood Inundation Mapping in Lake Victoria Basin: Implications for Hydrologic Prediction in Ungauged Basins, IEEE Transactions on Geoscience and Remote Sensing, vol. 49, issue 1, pp. 85-95, 49, 85–95, doi:10.1109/TGRS.2010.2057513, 2011.

Kidd, C., Levizzani, V., Turk, J., and Ferraro, R.: Satellite Precipitation Measurements for Water Resource Monitoring, JAWRA Journal of the American Water Resources Association, 45, 567–579, doi:10.1111/j.1752-1688.2009.00326.x, 2009.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K.-L., and Braithwaite, D.: Intercomparison of High-Resolution Precipitation Products over Northwest Europe, Journal of Hydrometeorology, 13, 67–83, doi:10.1175/JHM-D-11-042.1, 2012.

5   Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424-425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Jimenez, B., Miller, K., Oki, T., Şen, Z., and Shiklomanov, I.: The implications of projected climate change for freshwater resources and their management, Hydrological Sciences Journal, 53, 3–10, doi:10.1623/hysj.53.1.3, 2008.

10   Kurtzman, D., Navon, S., and Morin, E.: Improving interpolation of daily precipitation for hydrologic modelling: spatial patterns of preferred interpolators, Hydrological Processes, 23, 3281–3291, doi:10.1002/hyp.7442, 2009.

Lee, H., Zhang, Y., Seo, D.-J., and Xie, P.: Utilizing satellite precipitation estimates for streamflow forecasting via adjustment of mean field bias in precipitation data and assimilation of streamflow observations, Journal of Hydrology, 529, 779–794, doi:10.1016/j.jhydrol.2015.08.057, 2015.

15   Legates, D. R. and DeLiberty, T. L.: Precipitation Measurement Biases in the United States, Journal of the American Water Resources Association, 29, 855–861, doi:10.1111/j.1752-1688.1993.tb03245.x, 1993.

Li, X.-H., Zhang, Q., and Xu, C.-Y.: Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin, Journal of Hydrology, 426-427, 28–38, doi:10.1016/j.jhydrol.2012.01.013, 2012.

Lo Conti, F., Hsu, K.-L., Noto, L. V., and Sorooshian, S.: Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea, Atmospheric Research, 138, 189–204, doi:10.1016/j.atmosres.2013.11.011, 2014.

20   Maggioni, V., Meyers, P. C., and Robinson, M. D.: A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era, Journal of Hydrometeorology, 17, 1101–1117, doi:10.1175/JHM-D-15-0190.1, 2016.

Mantas, V., Liu, Z., Caro, C., and Pereira, A.: Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes, Atmospheric Research, 163, 132–145, doi:10.1016/j.atmosres.2014.11.012, 2015.

25   Mei, Y., Anagnostou, E. N., Nikolopoulos, E. I., and Borga, M.: Error Analysis of Satellite Precipitation Products in Mountainous Basins, Journal of Hydrometeorology, 15, 1778, doi:10.1175/JHM-D-13-0194.1, 2014.

Meng, J., Li, L., Hao, Z., Wang, J., and Shao, Q.: Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River, Journal of Hydrology, 509, 320, doi:10.1016/j.jhydrol.2013.11.049, 2014.

Mirza, M. M. Q.: Climate change and extreme weather events: can developing countries adapt?, Climate Policy, 3, 233–248, doi:10.3763/cpol.2003.0330, 2003.

30   Müller, A., Reiter, J., and Weiland, U.: Assessment of urban vulnerability towards floods using an indicator-based approach - a case study for Santiago de Chile, Natural Hazards and Earth System Science, 11, 2107–2123, doi:10.5194/nhess-11-2107-2011, 2011.

Naumann, G., Barbosa, P., Carrao, H., Singleton, A., and Vogt, J.: Monitoring Drought Conditions and Their Uncertainties in Africa Using TRMM Data, Journal of Applied Meteorology and Climatology, 51, 1867–1874, doi:10.1175/JAMC-D-12-0113.1, 2012.

35   Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., and Borga, M.: Understanding the Scale Relationships of Uncertainty Propagation of Satellite Rainfall through a Distributed Hydrologic Model, Journal of Hydrometeorology, 11, 520–532, doi:10.1175/2009JHM1169.1, 2010.

Peng, L., Sheffield, J., and Verbist, K. M. J.: Merging station observations with large-scale gridded data to improve hydrological predictions over Chile, in: 2016 AGU Fall Meeting Abstract, 12-16 December 2016, San Francisco, CA, USA, 2016.

Pereira Filho, A. J., Carbone, R. E., Janowiak, J. E., Arkin, P., Joyce, R., Hallak, R., and Ramos, C. G.: Satellite Rainfall Estimates Over South America - Possible Applicability to the Water Management of Large Watersheds, JAWRA Journal of the American Water Resources Association, 46, 344–360, doi:10.1111/j.1752-1688.2009.00406.x, 2010.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proc. Natl. Acad. Sci. U.S.A., 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014.

R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2015.

Ren, Z. and Li, M.: Errors and correction of precipitation measurements in China, Advances in Atmospheric Sciences, Volume 24, Issue 3, pp.449-458, 24, 449–458, doi:10.1007/s00376-007-0449-3, 2007.

Ringard, J., Becker, M., Seyler, F., and Linguet, L.: Temporal and Spatial Assessment of Four Satellite Rainfall Estimates over French Guiana and North Brazil, Remote Sensing, 7, 16 441–16 459, doi:10.3390/rs71215831, 2015.

Rogelis, M. C. and Werner, M. G. F.: Spatial Interpolation for Real-Time Rainfall Field Estimation in Areas with Complex Topography, Journal of Hydrometeorology, 14, 85–104, doi:10.1175/JHM-D-11-0150.1, 2013.

Schamm, K., Ziese, M., Raykova, K., Becker, A., Finger, P., Meyer-Christoffer, A., and Schneider, U.: GPCC Full Data Daily Version 1.0 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data, doi:10.5676/DWD_GPCC/FD_D_V1_100, 2015.

Scheel, M. L. M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E., and Huffman, G. J.: 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, 2649–2663, doi:10.5194/hess-15-2649-2011, 2011.

Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B.: Global precipitation analysis products of the GPCC, Internet publikation, Global Precipitation Climatology Centre (GPCC), DWD, 2008.

Scofield, R. A. and Kuligowski, R. J.: Satellite Precipitation Algorithms for Extreme Precipitation Events, pp. 485–495, doi:10.1007/978-1-4020-5835-6_38, 2007.

Serrat-Capdevila, A., Valdes, J. B., and Stakhiv, E. Z.: Water Management Applications for Satellite Precipitation Products: Synthesis and Recommendations, JAWRA Journal of the American Water Resources Association, 50, 509–525, doi:10.1111/jawr.12140, 2013.

Sevruk, B. and Chvíla, B.: Error sources of precipitation measurements using electronic weight systems, Atmospheric Research, 77, 39–47, doi:10.1016/j.atmosres.2004.10.026, 2005.

Sevruk, B., Ondrás, M., and Chvíla, B.: The WMO precipitation measurement intercomparisons, Atmospheric Research, 92, 376–380, doi:10.1016/j.atmosres.2009.01.016, 2009.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088, doi:10.1175/JCLI3790.1, 2006.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall, Bulletin of the American Meteorological Society, vol. 81, Issue 9, pp.2035-2046, 81, 2035–2046, doi:10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2, 2000.

Sorooshian, S., Hsu, K., Braithwaite, D., Ashouri, H., and NOAA CDR Program : NOAA Climate Data Record (CDR) of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN-CDR), Version 1 Revision 1. [2003-2014], Tech. rep., NOAA National Centers for Environmental Information., doi:10.7289/V51V5BWQ, [access date: 30-Jan-2016], 2014.

Su, F., Hong, Y., and Lettenmaier, D. P.: Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin, Journal of Hydrometeorology, 9, 622–640, doi:10.1175/2007JHM944.1, 2008.

Tao, H., Fischer, T., Zeng, Y., and Fraedrich, K.: Evaluation of TRMM 3B43 Precipitation Data for Drought Monitoring in Jiangsu Province, China, Water, 8, 221, doi:10.3390/w8060221, 2016.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and De Roo, A.: Validation of Satellite-Based Precipitation Products over Sparsely Gauged African River Basins, Journal of Hydrometeorology, 13, 1760–1783, doi:10.1175/JHM-D-12-032.1, 2012.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin, Journal of Hydrology, 499, 324–338, doi:10.1016/j.jhydrol.2013.07.012, 2013.

Tian, Y. and Peters-Lidard, C. D.: A global map of uncertainties in satellite-based precipitation measurements, Geophysical Research Letters, 37, 24 407, doi:10.1029/2010GL046008, 2010.

Tobin, K. J. and Bennett, M. E.: Satellite precipitation products and hydrologic applications, Water International, 39, 360–380, doi:10.1080/02508060.2013.870423, 2014.

Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J. B., Olivares, C., Vera, M., Balocchi, F., Pérez, F., Vallejos, C., Fuentes, R., Abarza, A., and Helwig, B.: Water governance in Chile: Availability, management and climate change, Journal of Hydrology, doi:10.1016/j.jhydrol.2014.04.016, 2014.

Verdin, J., Funk, C., Senay, G., and Choularton, R.: Climate science and famine early warning, Philos. Trans. R. Soc. Lond., B, Biol. Sci., 360, 2155–68, doi:10.1098/rstb.2005.1754, 2005.

Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall - effect of additional information, variogram inference and storm properties, Hydrology and Earth System Sciences, 15, 569–584, doi:10.5194/hess-15-569-2011, 2011.

Ward, E., Buytaert, W., Peaver, L., and Wheater, H.: Evaluation of precipitation products over complex mountainous terrain: A water resources perspective, Advances in Water Resources, 34, 1222–1231, doi:10.1016/j.advwatres.2011.05.007, 2011.

Werren, G., Reynard, E., Lane, S. N., and Balin, D.: Flood hazard assessment and mapping in semi-arid piedmont areas: a case study in Beni Mellal, Morocco, Natural Hazards, 81, 481–511, doi:10.1007/s11069-015-2092-0, 2016.

Woldemeskel, F. M., Sivakumar, B., and Sharma, A.: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia, Journal of Hydrology, 499, 167–176, doi:10.1016/j.jhydrol.2013.06.039, 2013.

World Meteorological Organization: Guide to meteorological instruments and methods of observation. WMO-No. 8, chap. Present and past weather; state of the ground, pp. I.14–1–I.14–9, World Meteorological Organization, Geneva, Switzerland, Seventh edn., 2008.

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Dorji, C., and Chen, S.: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?, Journal of Hydrology, 499, 91–99, doi:10.1016/j.jhydrol.2013.06.042, 2013.

Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., and Verbist, K. M. J.: Bias Adjustment of Satellite-based Precipitation Estimation using Gauge Observations-A Case Study in Chile, Journal of Geophysical Research: Atmospheres, doi:10.1002/2015JD024540, 2016.

Zambrano-Bigiarini, M.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, https://CRAN.R-project.org/package=hydroGOF, R package version 0.3-8-7, 2016a.

Zambrano-Bigiarini, M.: hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling, http://CRAN. R-project.org/package=hydroTSM, R package version 0.4-8, 2016b.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall esti- mates across the complex topographical and climatic gradients of Chile (suplemmentary material), doi:10.5281/zenodo.251069, 2016.

5 Zhang, A. and Jia, G.: Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data, Remote Sensing of Environment, 134, 12–23, doi:10.1016/j.rse.2013.02.023, 2013.

Zhang, X. and Srinivasan, R.: GIS-Based Spatial Precipitation Estimation: A Comparison of Geostatistical Approaches, Journal of The American Water Resources Association, 45, 894–906, doi:10.1111/j.1752-1688.2009.00335.x, 2009.

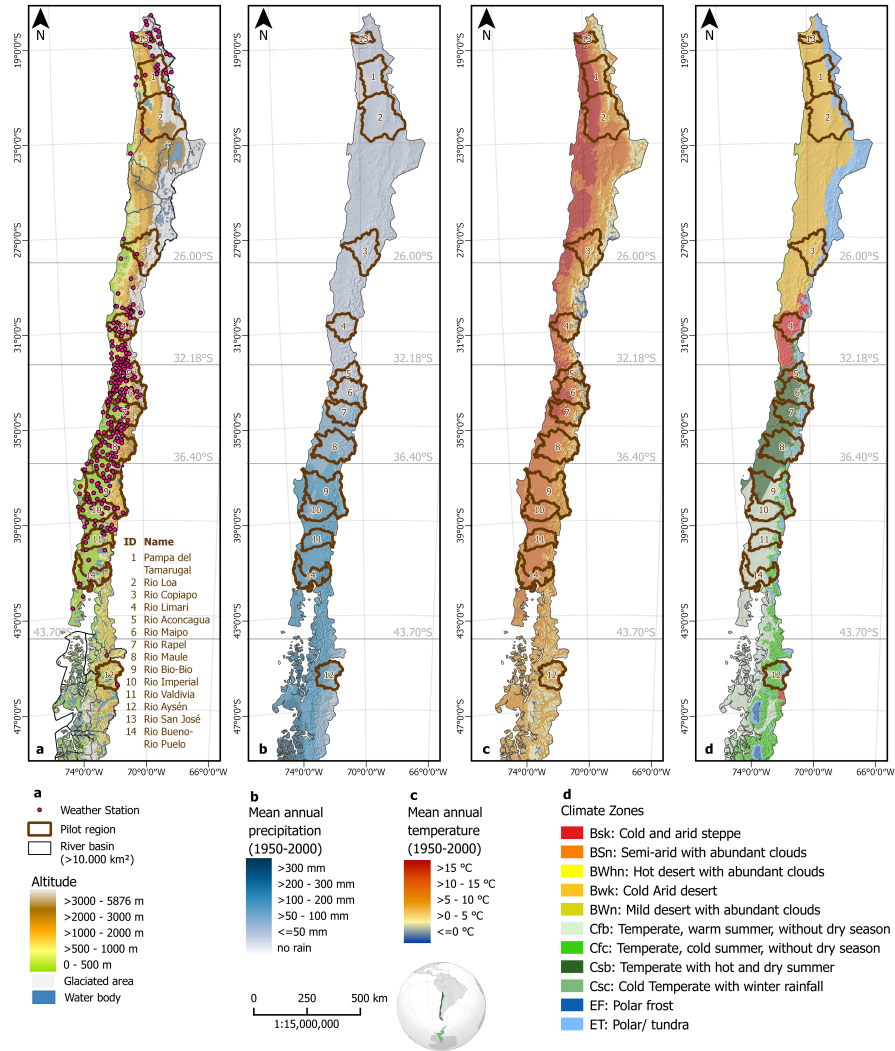



Figure 1. Study Area. From left to right: (a) digital elevation model (Jarvis et al., 2008), main Chilean basins, and location of 289 selected raingauge; (b) mean annual precipitation (Hijmans et al., 2005); (c) mean annual temperature (Hijmans et al., 2005); and (d) climate zones based on Köppen classification.

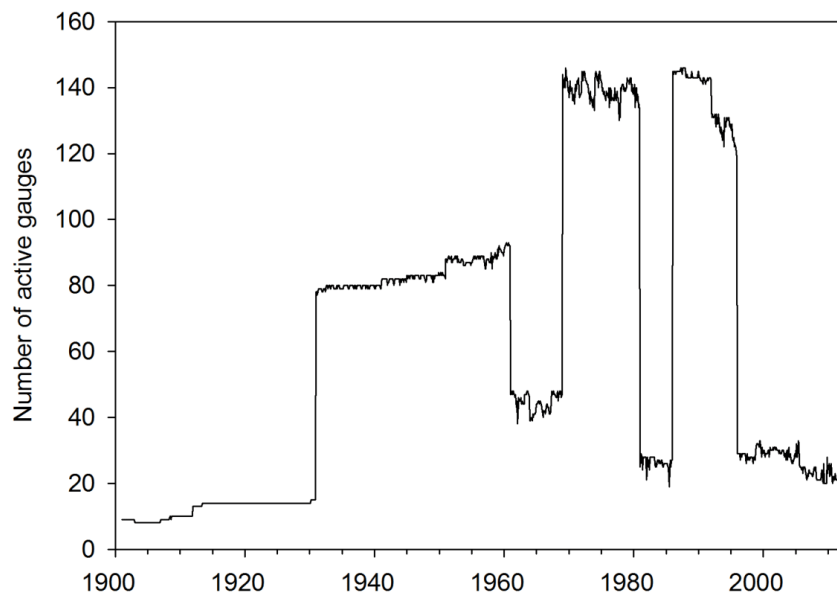



Figure 2. Number of raingauges in Chile used in the GPCC dataset (Schamm et al., 2015).

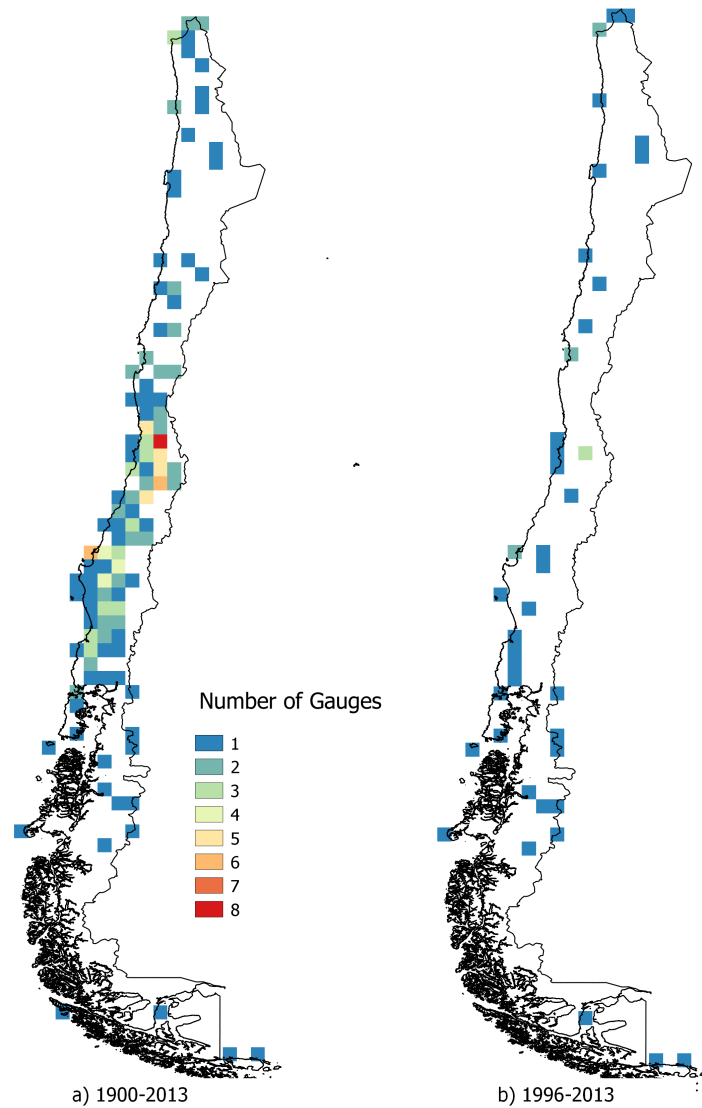



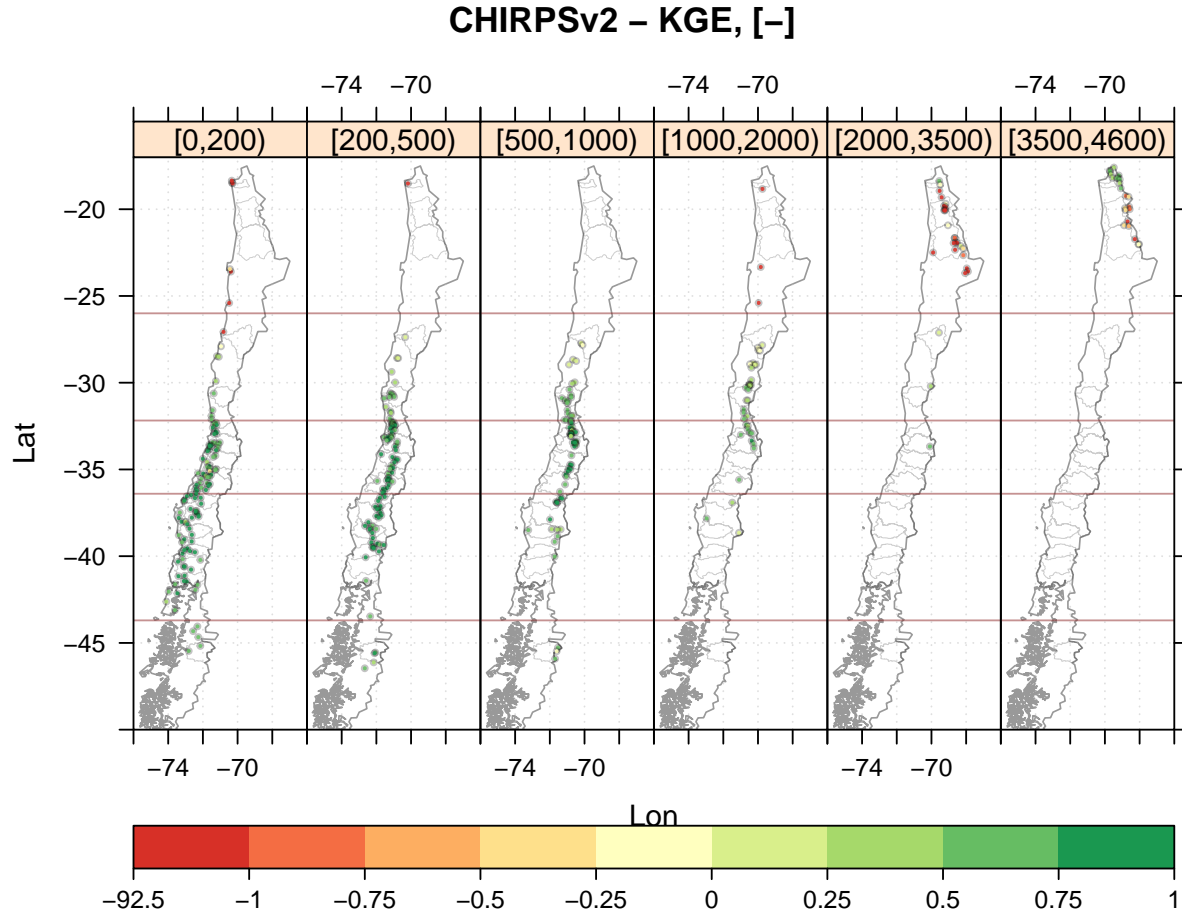
Figure 3. Spatial distribution of the maximum number of raingauges used in Chile in GPCC over the full period and the maximum number of raingauges used in the last two decades (Schamm et al., 2015).

CHIRPSv2 – KGE, [–]

Figure 4. Modified Kling-Gupta efficiency (*KGE'*) between monthly satellite estimations (CHIRPSv2) and observations at the corresponding grid cell, for six different elevation zones: [0,200), [200,500), [500,1000), [1000,2000), [2000,3500), [3500,4560) m a.s.l. Colours for KGE' range from intense red to dark green, representing very poor and optimum performance, respectively.

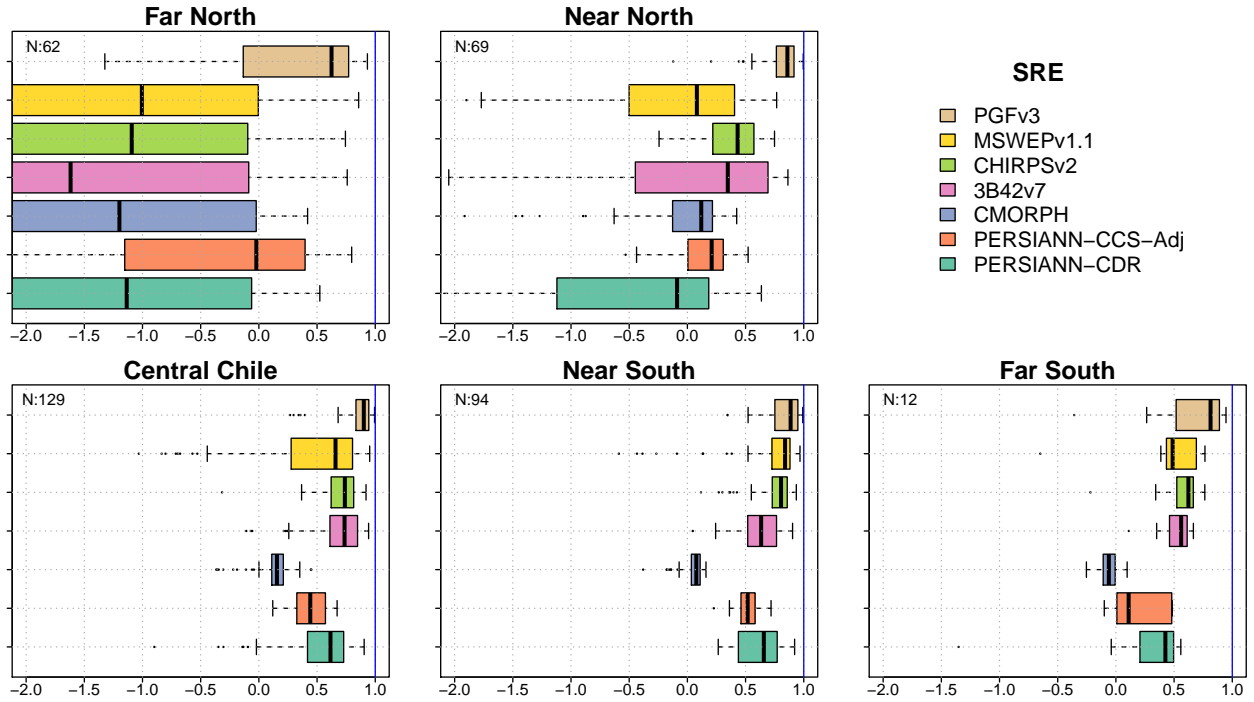



Figure 5. Modified Kling-Gupta efficiency (*KGE'*) between different monthly satellite estimations and observations at the corresponding grid cell, for five different macroclimate zones: Far North, Near North, Centre, South and Far South. The vertical blue line indicates the optimum value for KGE', while N indicates the number of stations in each macroclimate zone.

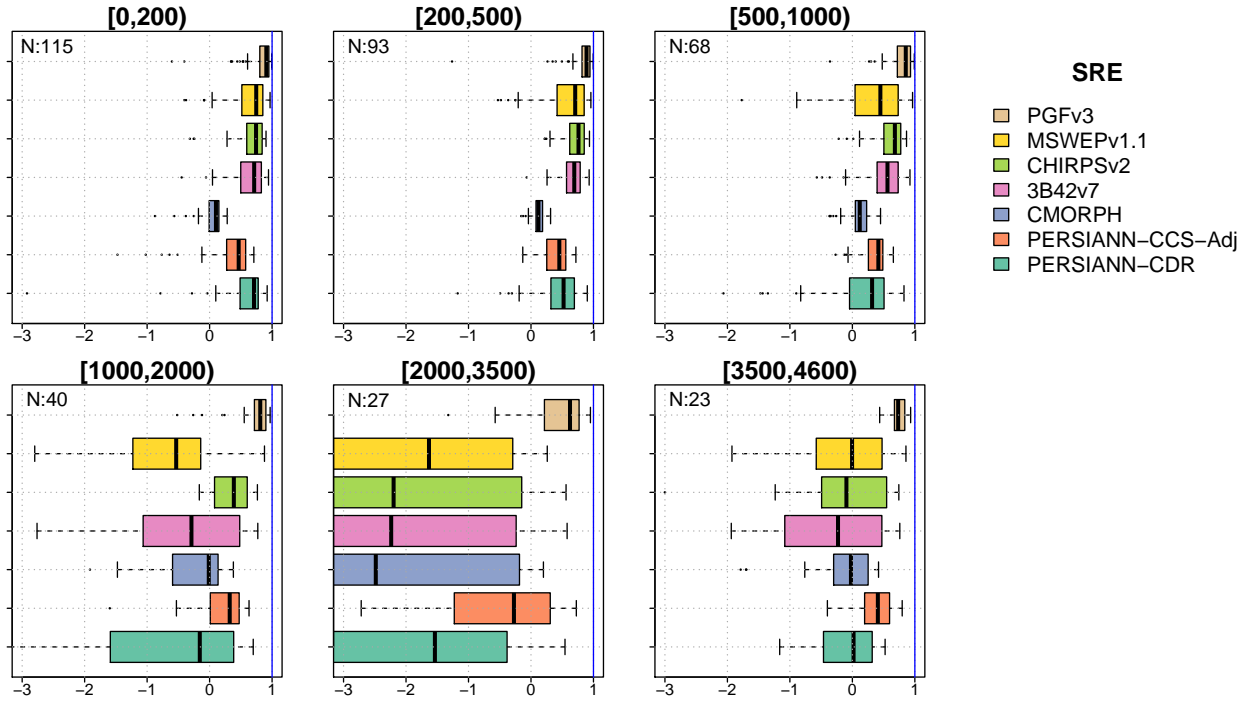



Figure 6. Modified Kling-Gupta efficiency (KGE') between different monthly satellite estimations and observations at the corresponding grid cell, for six different elevation zones: [0,200), [200,500), [500,1000), [1000,2000), [2000,3500), [3500,4560) m a.s.l. The vertical blue line indicates the optimum value for KGE', while N indicates the number of stations in each elevation zone.

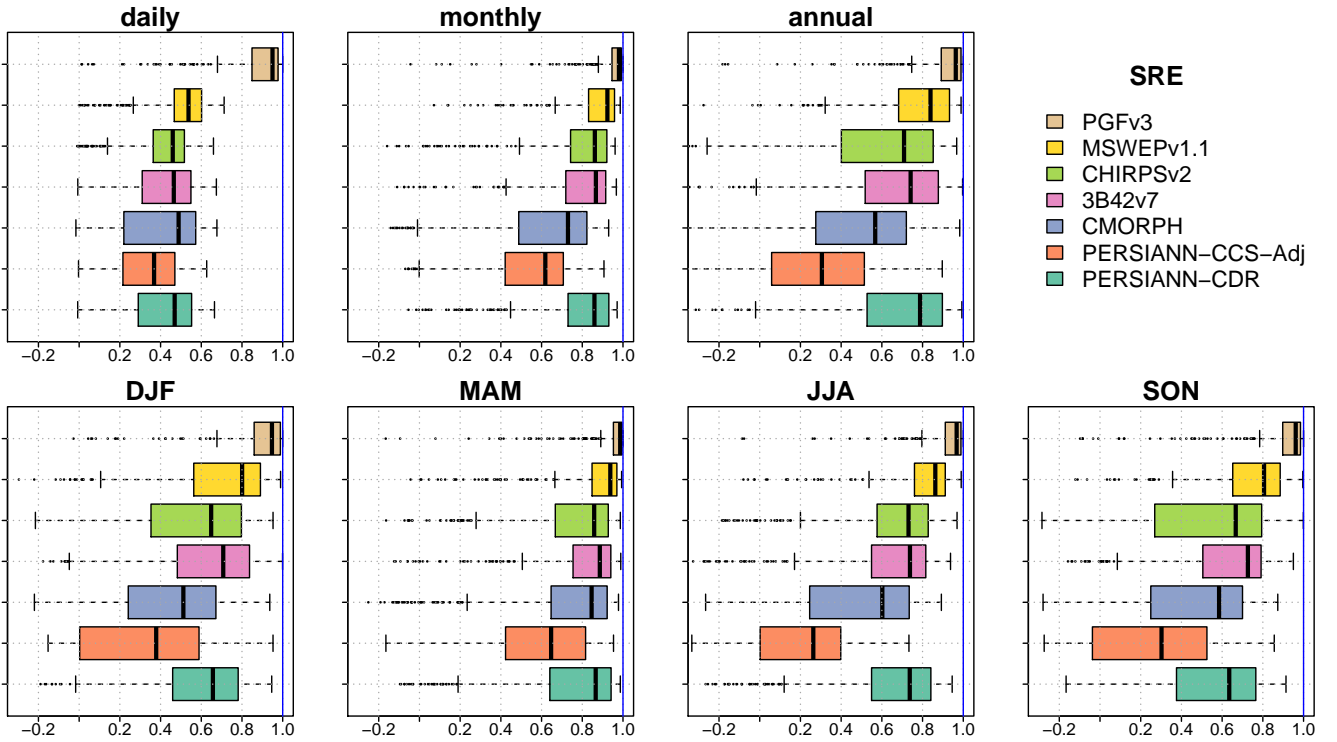



Figure 7. Pearson product-moment correlation coefficient (r) between different satellite estimations and the observations at the corresponding grid cell, for six different temporal scales. From left to right and up to bottom: daily, monthly, annual, summer (DJF), autumn (MAM), winter (JJA), and spring (SON). The vertical blue line indicates the optimum value for r.

Figure 8. Bias ratio of the modified Kling-Gupta efficiency (β) between different satellite estimations and the observations at the corresponding grid cell, for six different temporal scales. From left to right and up to bottom: daily, monthly, annual, summer (DJF), autumn (MAM), winter (JJA), and spring (SON). The vertical blue line indicates the optimum value for β.

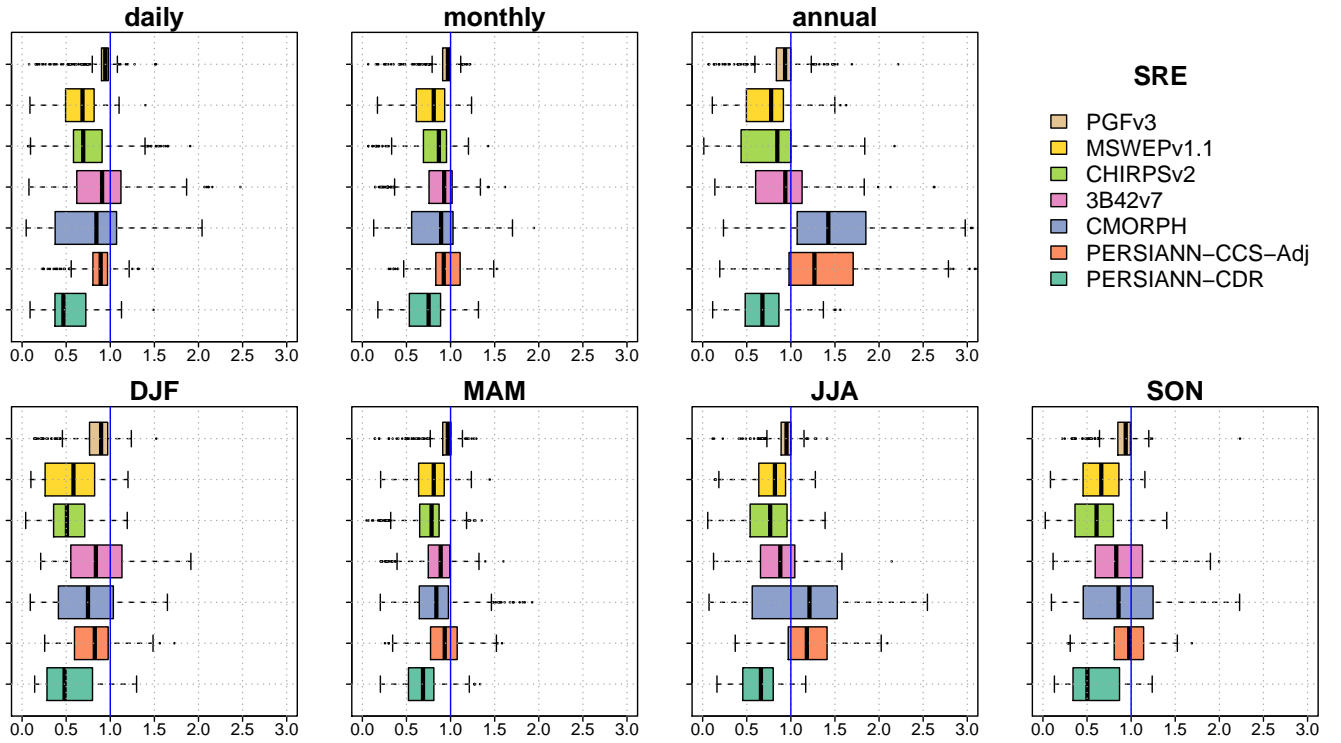



Figure 9. Variability ratio of the modified Kling-Gupta efficiency (γ) between different satellite estimations and the observations at the corresponding grid cell, for six different temporal scales. From left to right and up to bottom: daily, monthly, annual, summer (DJF), autumn (MAM), winter (JJA), and spring (SON). The vertical blue line indicates the optimum value for γ.

Figure 10. Modified Kling-Gupta efficiency (*KGE'*) between different satellite estimations and the observations at the corresponding grid cell, for six different temporal scales. From left to right and up to bottom: daily, monthly, annual, summer (DJF), autumn (MAM), winter (JJA), and spring (SON). The vertical blue line indicates the optimum value for KGE'.

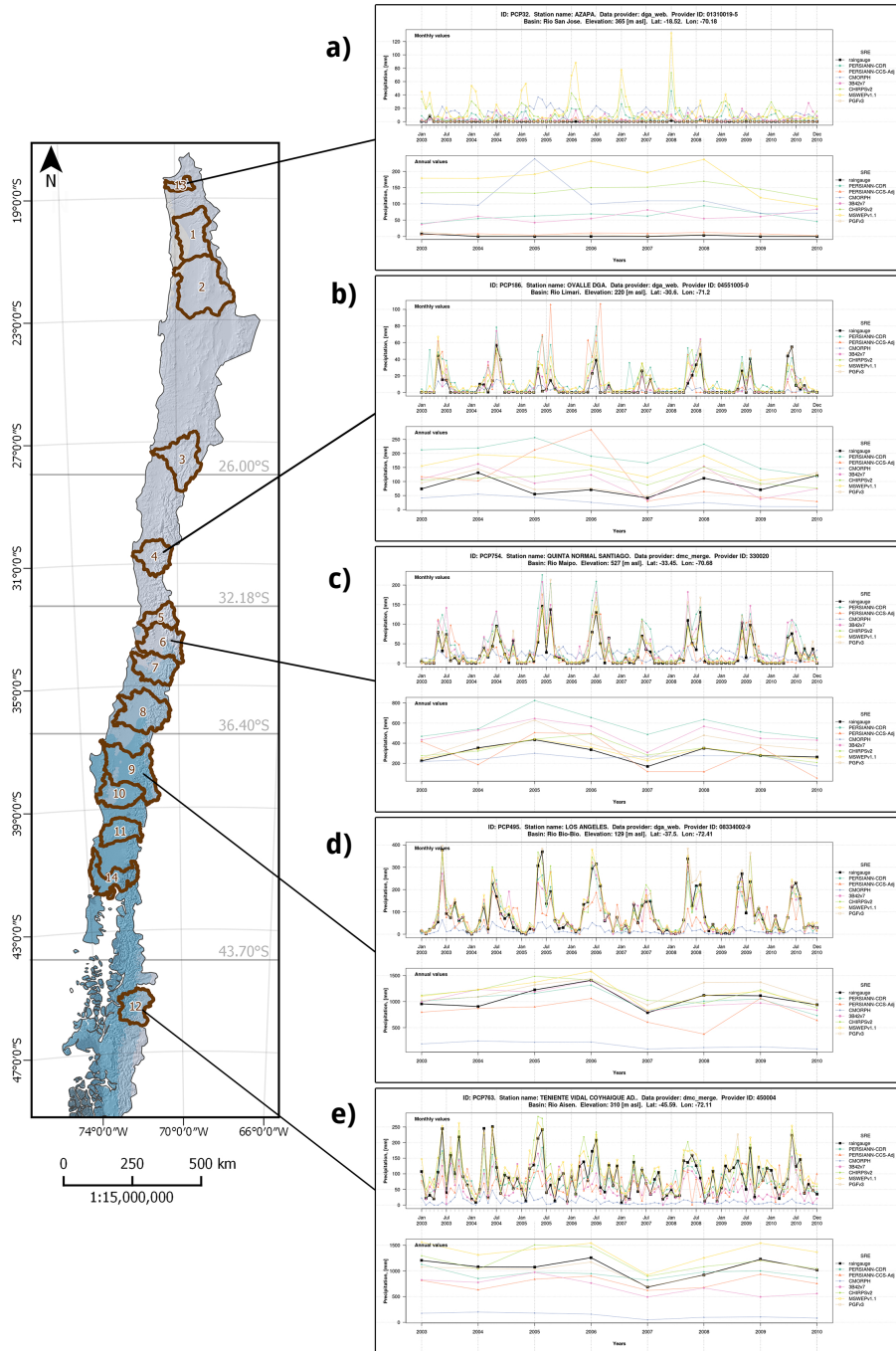



Figure 11. Monthly and annual precipitation time series as estimated by the seven SRE products compared to the observed values registered at the corresponding raingauge. Panels a), b), c), d) and e) show station examples for the Far North, Near North, Centre, South and Far South climate macrozones, respectively.

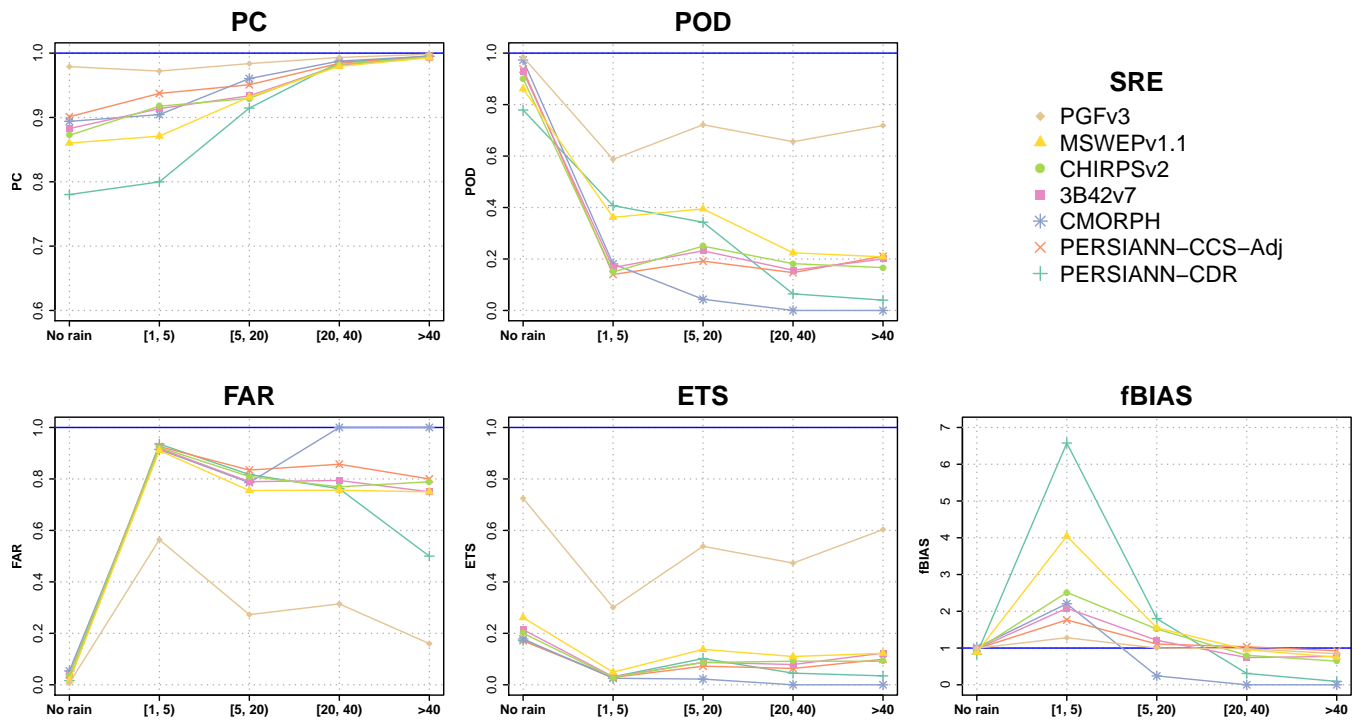



Figure 12. Median values of categorical indices of performance for different SREs and the five classes of rainfall intensity defined in Table 2. From left to right and up to bottom: percent of correct *PC*), probability of detection (*POD*), false alarm ratio (*FAR*), equitable threat score (*ETS*) and frequency bias (*fBias*)..

Table 1. Summary of satellite-based products providing (sub)daily data at a quasi-global scale

Dataset	Full Name	Latitudinal Coverage	Spatial Resolution	Temporal Coverage	Temporal Resolutions	References
CMORPH	NOAA **C**limate Prediction Center (CPC) **MORPH**ing technique	60°N-60°S	0.07°, 0.25°	Dec-2002 - present	3-hourly, daily	Joyce et al. 2004; CPC-NCEP-NWS-NOAA-USDC 2011
PERSIANN-CDR	PERSIANN **C**limate **D**ata **R**ecord, Version 1 Revision 1	60°N-60°S	0.25°	Jan-1983 - present	daily	Sorooshian et al. 2014; Ashouri et al. 2015
PERSIANN-CCS-adj	**P**recipitation **E**stimation from **R**emotely **S**ensed **I**nformation using **A**rtificial **N**eural **N**etworks	50°N-50°S	0.04°	Jan-2003 - present	daily	Yang et al. 2016; Hong et al. 2004
3B42v7	**TRMM M**ulti-satellite **P**recipitation **A**nalysis research product 3B42 Version 7	50°N-50°S	0.25°	Jan-1998 - present	3-hourly, daily	Huffman et al. 2007, 2010
CHIRPSv2	**C**limate **H**azards group **I**nf**r**ared **P**recipitation with **S**tations Version 2.0	50°N-50°S	0.05°	Jan-1981 - present	daily, pentadal, monthly	Funk et al. 2015
MSWEPv1.1	**M**ulti-**S**ource **W**eighted-**E**nsemble **P**recipitation Version 1.1	90°N-90°S	0.25°	Jan-1979 Dec-2014	daily	Beck et al. 2016
PGFv3	**P**rinceton University **G**lobal Meteorological **F**orcing Version 3	17°S-57°S	0.25°	Jan-1979 Dec-2010	daily	Peng et al. 2016; Sheffield et al. 2006

Table 2. Classification of rainfall events based on its daily intensity i. Modified for daily values from World Meteorological Organization (2008).

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

Table 3. Contingency table used to compute the categorical performance indices for each rainfall event shown in Table 2.

Satellite-product	Observed rainfall		
	Yes	No	**Total**
Yes	Hit (H)	False Alarm (*FA*)	$H + FA$
No	Miss(*M*)	Correct Negative (*CN*)	$M + CN$
Total	$H + M$	$FA + CN$	*Ne*