

# Validation report of the global, daily precipitation data set DAPAGLOCO

Marloes Gutenstein, Felix Dietzsch, Karsten Fennig, Marc Schröder

April 5, 2019

## 1. Introduction

The global, gridded, daily precipitation data set DAPAGLOCO<sup>1</sup> was developed within the MiKlip project framework for the evaluation of decadal prediction models. The data set essentially consists of two separate data sets – GPCC over land and HOAPS over ocean – combined to achieve global coverage. DAPAGLOCO exists at three spatial resolutions: 2.5°x2.5° and 1°x1° globally, and 0.5°x0.5° over Europe. In this document, a validation of the 1° product with various global and regional data sets is presented. After the description of each data set and the method (Sects. 2 and 3), the results of the validation study based on monthly, yearly, and multi-year averages are presented in Sect. 4. The following section 5 contains the results of the validation of daily precipitation data using ETCCDI indices and Sect. 6 concludes with a discussion of the findings. Referenced literature can be found in Sect. 7, and the acronyms used throughout this document are listed in Sect. 8.

## 2. Data sets

Nine precipitation data sets are compared with DAPAGLOCO and amongst each other to investigate the consistency of the different data sets with regard to spatial and temporal patterns and variability. These data sets are, however, not all completely independent (as follows from the algorithm descriptions in this section and the summary Table 1), which makes a true validation difficult to achieve. In addition, the sparseness of *in situ* data over ocean prohibits a validation with independent data. This means that it cannot be said with certainty which of the satellite-based data sets performs best in terms of precision or accuracy. We here focus on consistency between the precipitation algorithms, without deciding on a certain champion.

Table 1. Characteristics of the observational precipitation data sets used within the presented validation study. Abbreviations: PMW – passive microwave, IR – infrared, PR – precipitation radar.

| Data Record | Ocean input        | Land input             | Used version                |
|-------------|--------------------|------------------------|-----------------------------|
| APHRODITE   | -                  | Gauges                 | V1801R1                     |
| CHIRPS      | -                  | TRMM, IR, gauges       | 2.0                         |
| CMORPH      | PMW, IR            | PMW, IR                | V1.0 CRT Daily              |
| DAPAGLOCO   | HOAPS*             | GPCC                   | 3.1                         |
| GPCC        | -                  | Gauges                 | Full Data Daily V.2018 (V2) |
| GPCP        | PMW, IR            | PMW, IR, GPCC          | 1DD Version 1.3             |
| HOAPS       | PMW                | -                      | Version 4.0                 |
| PACRAIN     | Gauges (on atolls) | Gauges                 |                             |
| PERSIANN    | PMW, IR            | PMW, IR, gauges, radar | CDRv1r1                     |
| TRMM        | PMW, IR, PR        | PMW, IR, PR, GPCC      | TRMM3B42, Version 7         |

\*Due to differences in the input data and the calculation of daily total pr, the HOAPS data in DAPAGLOCO is not identical to HOAPS-4 data (see Sect. 2.4).

<sup>1</sup> For the meaning of acronyms and abbreviations, see Section 8

As mentioned above, the validation was performed for the 1°-product only; the data sets provided on other resolutions (TRMM, CMORPH, APHRODITE, ERA-interim) were re-gridded to a 1° resolution using conservative remapping. Monthly means were calculated from daily data for the first part of the validation, disregarding grid cells with missing values. This approach is only valid if the fraction of missing data is small, hence, (for example) DAPAGLOCO data over ocean observed before 1995 should be treated with some caution as daily global coverage was only achieved with the advent of DMSP's F13 platform in March 1995.

## 2.1. *In situ* data

### 2.1.1. GPCC

The Global Precipitation Climatology Centre's Full Data Daily Product V.2018 (V2) contains daily, gridded precipitation rate (*pr*) derived from relative precipitation anomalies obtained from about 35,000 stations worldwide per month. After rigorous quality control, the gauge data are interpolated to a 1°x1° grid using a modified SPHEREMAP scheme [Schneider et al., 2018]. A comparison of GPCC's monthly data product with other data sets can be found in Schneider et al. (2014).

### 2.1.2. APHRODITE

Asian Precipitation – Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) is a daily gridded precipitation gauge data set for Asia [Yatagai et al., 2012]. After rigorous quality control, the ratio of the daily station data to a climatology is interpolated to a 0.05°x0.05° grid following a sophisticated scheme [Yatagai et al., 2012]. There is quite some overlap in the stations used for the generation of the GPCC and APHRODITE data sets, but the latter additionally obtains data from a large number of stations in Monsoon Asia, particularly in Japan, India, Thailand, and from the Arabian Peninsula. APHRODITE version 1101 spans the time range 1980-2007 and covers most land areas between 20°E - 180°E and 15°S - 80°N. We use the updated version 1801R1 available from 1998-2017, which has a better temporal overlap with the other data sets evaluated here. Currently, version 1801R1 is available over monsoon Asia (60°E - 155°E and 20°S - 55°N) at daily, 0.25°x0.25° and 0.5°x0.5° resolution.

### 2.1.3. PACRAIN

The Pacific Rainfall Database, PACRAIN, consists of daily and monthly rainfall records from many sites located on atolls and islands (Greene, 2007). The database is a collection of observations from a variety of sources, including one, the Schools of the Pacific Rainfall Climate Experiment (SPaRCE), that is unique to PACRAIN, and is updated monthly. The database features extensive quality control, observation and data entry standardization, and enhanced high-resolution data (e.g., on hourly or minute time scales). For our study, the daily PACRAIN data were averaged to monthly means and compared directly with the satellite (or reanalysis) monthly mean in the 1°x1° grid box in which the site was located; and to the mean of the four or nine grid boxes closest to the site (see Sect. 4.5). Atoll sites are chosen to represent over-ocean precipitation, as the influence of atoll-sized islands on precipitation is presumed to be negligible [Sobel et al., 2011].

## 2.2. Satellite data sets and merged products

The data from which the global precipitation algorithms derive the *pr* come from two different sets of sensors. Passive microwave (PMW) data are collected by a suite of polar-orbiting satellite instruments, such as SSM/I and SSMIS on the DMSP series, TMI on TRMM, AMSR-E on Aqua, and

AMSU-B on the NOAA satellites. PMW data is directly related to scattering on and emission by hydrometeors. In contrast to the rather sparse temporal sampling of PMW observations, geostationary infrared (IR) imagers provide excellent coverage in space and time. IR-based *pr* estimates, however, are indirectly derived from brightness temperatures (TB) that depend on cloud top temperature (i.e., cloud height), rather than on precipitating particles. To exploit the advantages of both instrument types, many of the algorithms presented below – GPCP, CMORPH, PERSIANN, TRMM – combine IR *TB* with PMW *pr* to obtain daily *pr* data sets with global or quasi-global coverage.

#### 2.2.1. HOAPS

The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data is a data record purely based on PMW satellite input. Apart from precipitation, HOAPS retrieves evaporation, latent heat flux, and a variety of related parameters representing the global water and energy cycle [Andersson et al., 2010; Graw et al., 2017]. HOAPS variables are derived from PMW radiometers onboard the polar-orbiting satellites on the DMSP platforms. Homogeneous time series are achieved by using CM SAF Fundamental Climate Data Record of inter-calibrated brightness temperatures as input. The algorithm makes use of a neural network retrieval trained with *pr* retrieved from assimilated *TB* in a 1D-Var scheme from the ECMWF. The precipitation retrieval depends only on *TB* as input and does not need first guess or other ancillary data [Anderson et al., 2010, 2011]. For the latest version 4.0, in addition to SSM/I, data from the SSMIS instruments were added, allowing to lengthen the time range from 1987 to 2014. The validation of HOAPS Version 4 (doi: 10.5676/EUM\_SAF\_CM/HOAPS/V002) is presented in [Graw et al., 2017].

#### 2.2.2. TRMM

The TRMM Multisatellite Precipitation Analysis (TMPA) has three components: the merged MW product, the MW-calibrated IR product, and the combined MW-IR product (3B42 [Huffman et al., 2007; Kidd and Huffman, 2011]). Where PMW *pr* estimates are not available, geostationary IR data are used, having been converted to *pr* estimates using a local probability matching between IR *TB* and PMW estimates over a month. The 3B42 Version 7 product used here is bias-corrected against monthly GPCC gauge estimates where available.

#### 2.2.3. GPCP

The Global Precipitation Climatology Project's 1° global data set, GPCP-1DD, derives *pr* from visible and infrared (Vis-IR) observations by instruments on geostationary satellites (and the polar-orbiting AVHRR instrument) [Huffman et al., 1997]. Precipitation rates can, however, not be determined directly in the Vis-IR range, but only via the relationship between cloud top temperature and *pr*. The parameters used to describe this relationship are determined with a monthly frequency and on a 1°x1° grid by combining *pr* determined from collocated PMW observations with *TB* determined in the IR. The monthly parameters are then applied to the 3-hourly IR *TB* to yield a 3-hourly *pr* data set for all latitudes within 40° of the equator. TOVS, AIRS, and AMSU data are used to produce the precipitation estimates at higher latitudes. In a final step, the monthly sum of the daily data set is adjusted to the GPCP-SG monthly total. The latter product is obtained from satellite data and calibrated using monthly GPCC gauge data [Huffman et al., 1997].

#### 2.2.4. CMORPH

The CPC morphing algorithm, CMORPH, was developed at NOAA/CPC to provide precipitation information at high spatial and temporal resolution [Joyce et al., 2004; Joyce and Xie, 2011]. The algorithm first generates advection vectors from 30-minute-interval geostationary IR imagery, then “morphs” precipitation shape and intensity obtained from PMW, meaning that a time-weighted linear interpolation is performed at the times between consecutive PMW sensor overpasses.

#### 2.2.5. PERSIANN

The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN [Hsu et al., 1997, 1999]) algorithm combines information from geostationary and polar-orbiting IR imagers, and PMW instruments. The algorithm uses an artificial neural network model to extract cold-cloud pixels and neighboring features from IR images and associates variations in each pixel’s brightness temperature to estimate the pixel’s surface *pr*.

#### 2.2.6. DAPAGLOCO

The DAPAGLOCO data set was obtained by merging gridded satellite-based observations over ocean with gridded station-based rain gauge data over land. The satellite data consists of *pr* retrieved from a series of passive microwave imagers using the HOAPS algorithm [Anderson et al., 2010]. For DAPAGLOCO, the data from the SSM/I and SSMIS instruments are complemented by TMI and AMSR-E data (on NASA’s TRMM and Aqua platforms, respectively) and constitute a continuous, daily record of precipitation over the ice-free ocean from 1988 to 2015 (with daily global coverage starting from 1995). Instantaneous retrievals are properly integrated to daily accumulated precipitation. Over land, gridded precipitation rates from GPCP Full Data Daily V.2018 product are used. Coastal areas have been interpolated. The combined DAPAGLOCO data record has been doi-referenced (10.5676/DWD\_CDC/HOGP\_XXX/V002, where XXX stands for 050, 100, and 250 for the data sets at 0.5°, 1.0°, and 2.5° resolution, respectively).

#### 2.2.7. CHIRPS

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a quasi-global (land-only, 50°S-50°N) precipitation data set. CHIRPS combines geostationary satellite IR imagery with *in situ* station data to create gridded precipitation time series for trend analysis and seasonal drought monitoring [Funk et al., 2015]. The coefficients needed to transform IR brightness temperatures into *pr* are determined on a monthly, 0.05°x0.05° grid using the TRMM3B42 product.

#### 2.3. ERA-interim reanalysis

ERA-interim is the version of ECMWF Re-Analysis that succeeded ERA-40 after several shortcomings were identified in the latter. Based on Cycle 31r2 of the IFS model, ERA-interim applies four-dimensional variational data assimilation, uses a completely automated scheme to adjust for biases in satellite radiance observations, and executes modified convective and boundary layer cloud schemes, increasing the atmospheric stability and producing less precipitation compared to ERA-40 (Dee et al., 2011). A wide variety of observations are assimilated to obtain a global state of the atmosphere. Satellite data, including those from SSM/I, are an important input variable for water vapour profiles, and, thus, implicitly influence precipitation forecasts. The native grid resolution of ERA-interim is T255, which corresponds to about 0.75°x0.75°; the temporal resolution is 6 hours.

### 3. Method

The validation of DAPAGLOCO was performed using monthly mean precipitation rates on a  $1^\circ \times 1^\circ$  grid. The Tool for Intercomparison of Gridded Data Sets (TIGriDS), which was developed for this purpose, was used to produce maps and time series of means, differences, and biases between the various data sets. Prior to the calculation of statistics, TIGriDS synchronizes all data sets, i.e., grid points are only considered for analysis if valid values are available for each data set. If TRMM, for example, is one of the data sets used for analysis, values at latitudes greater than  $50^\circ\text{N}$  and smaller than  $50^\circ\text{S}$  are eliminated from all data sets, as TRMM does not contain data there.

For the validation, monthly mean  $pr$  were determined from daily means, disregarding grid boxes containing invalid (missing) data. This approach is only valid if the amount of missing data is small, hence, DAPAGLOCO data over ocean observed before 1995 should be treated with some care, as daily global coverage was only achieved with the advent of DMSP's F13 platform in March 1995.

An important characteristic of precipitation is its variability, both in time and space. In many situations (changes in) precipitation extremes, such as droughts or floods, are of high importance - but monthly means provide only limited information of such events. To assess the consistency of precipitation extremes among all data sets ETCCDI (Expert Team on Climate Change Detection Indices, [Peterson and Manton, 2008; Zhang et al., 2011]) were determined from daily  $pr$  data and compared (Sect. 5). The subset of ETCCDI presented in this report are: the simple daily intensity index (SDII), the maximum cumulative 5-day amount of rain (Rx5), and the highest number of successive excessively wet days (days with  $pr$  exceeding 20 mm, R20mm), determined within a calendar year.

### 4. Results

#### 4.1. DAPAGLOCO means and stability

The global, multi-year mean  $pr$  field from DAPAGLOCO, averaged over the years 1995-2015, is shown in Fig. 1A. The highest mean  $pr$  are found at the Inter-Tropical Convergence Zone (ITCZ) near the equator, with mean rain rates exceeding 10 mm/day. Regional maxima are found over the tropical Indian Ocean and the South Pacific Convergence Zone (SPCZ). High  $pr$  within the North Atlantic and Pacific storm tracks are clearly visible, whereas minima can be found in the world's land deserts and the subtropical oceanic deserts in the eastern Atlantic and Pacific Oceans.

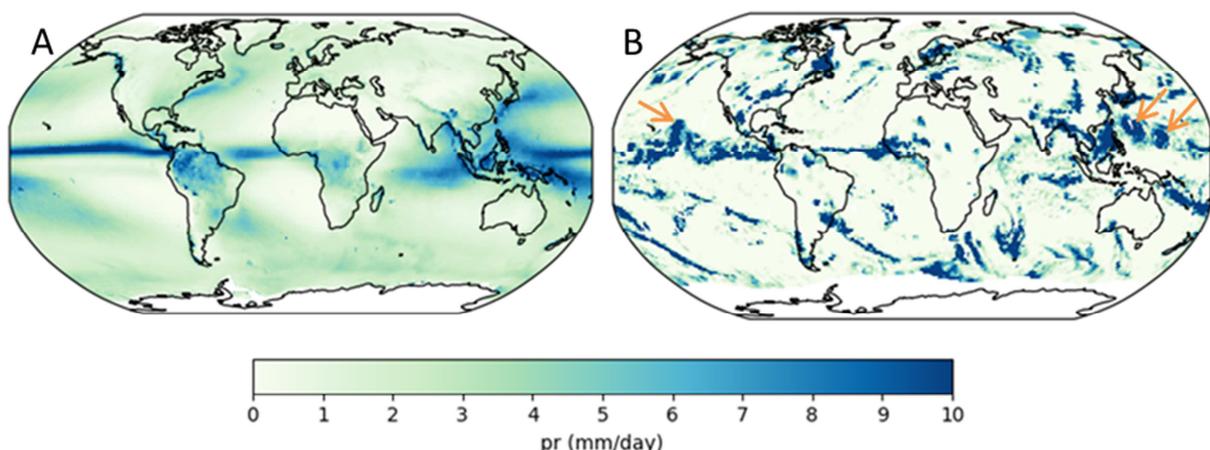


Figure 1. DAPAGLOCO mean precipitation rates. A: Multi-year mean of 1995-2015, B: Precipitation rate on July 8, 2015. Orange arrows in panel B indicate the position of three tropical storms.

An important feature of Fig.1A is the absence of seemingly unphysical land/sea gradients, indicating that the merging of the two data sets – GPCC over land, HOAPS over ocean – was successful. This is even more apparent in daily data, for example on 8/7/2015 (Fig. 1B), where several precipitation systems are seen to cross from land to ocean (or *vice versa*) without apparent inconsistencies. Another obvious feature seen in Fig.1 is the lack of data at and around the North and South poles. This is due to the fact that the HOAPS algorithm retrieves *pr* only over the ice-free ocean where GPCC does not include data.

Figure 2 shows global 12-month running mean precipitation anomalies with respect to the mean of the entire time series. The mean was determined for the whole global data set (black line), for ocean data only (blue), and for all land data (green). A slight positive trend of 0.0012 mm/day per decade was found for the whole global data set, but is not significant in view of the variability of the data and the limited length of the data set.

Most of the year-to-year variability in global mean precipitation can be explained by the El Niño Southern Oscillation (ENSO) phase, as seen by comparison with the monthly Multivariate ENSO Index (MEI [Wolter and Timlin, 1999]) indicated by grey shaded areas in Fig. 2. The good correlation between global *pr* and MEI is caused by the strong connection between ENSO and *pr* near the ITCZ and the fact that the region around the ITCZ contributes most to the global mean (as seen in Fig. 1).

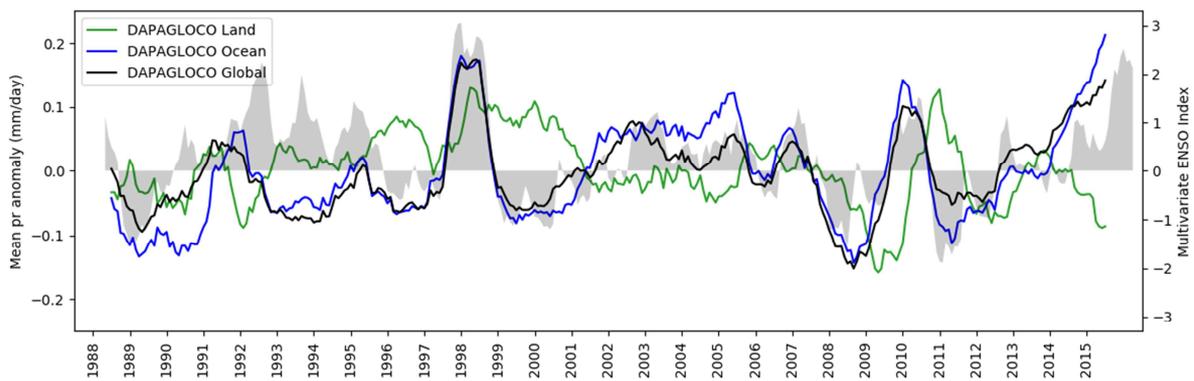


Figure 2. DAPAGLOCO mean yearly anomalies. The anomalies were obtained by subtracting the multi-year mean (1988-2015) from the yearly means: 2.52 mm/day for global (black), 2.26 mm/day for land only (green), 2.82 mm/day for ocean only (blue). Greyshade and right y-axis: monthly Multivariate ENSO Index (MEI). Numbers on the x-axis denote January of a certain year. The MEI time scale is shifted by +4 months to account for the lag in precipitation reaction to changes in sea surface temperature [Jost, 2000].

#### 4.2. Comparison with global data sets

DAPAGLOCO global means were compared with global means of four other global *pr* observational data sets: TRMM, GPCP, PERSIANN, and CMORPH; and with reanalysis data from ERA-interim. A validation time period of 1998-2015 was chosen, as the TRMM and CMORPH data sets commence in 1998. The left panel of Fig. 3 shows the yearly quasi-global (50°S - 50°N) mean *pr* of each of the six data sets and the mean of all observational data sets (DS\_mean); the right panel presents the corresponding yearly *pr* anomalies with respect to each data set's mean over the whole time period (1998-2015). All observational data sets are in general agreement, although CMORPH (in purple) is biased low by about 0.1 mm/day. DAPAGLOCO, indicated by a blue line, displays the largest year-to-year variation, with anomalies as large as  $\pm 0.15$  mm/day, whereas the other data sets show maximum anomalies of about  $\pm 0.05$  mm/day. The yearly variability of TRMM data is markedly

different from that shown by the other data sets (best seen in Fig.3B). The reanalysis data are rather similar to the observations, but the global mean shows a constant bias of +0.3 mm/day from 1998-2005, after which it steadily increases to a constant value of almost +0.5 mm/day from 2010 on.

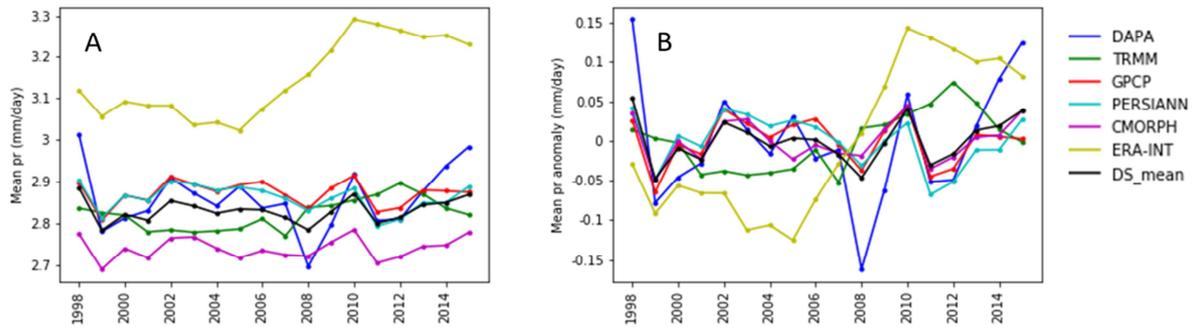


Figure 3. Yearly quasi-global (50°S - 50°N) means (A) and anomalies with respect to the mean over the whole time period (B) for the five observational data sets, their mean (DS\_mean), and ERA-interim reanalysis data.

The mean seasonal cycle and meridional profile of all six data sets and DS\_mean, determined for the time period 1998-2015, are shown in Fig.4. CMORPH data are in good agreement with all other data sets (particularly GPCP and PERSIANN), except for the constant offset of 0.1 mm/day. As already seen in Fig.3, DAPAGLOCO's variability is larger than that of the other data sets, particularly in the winter months, December – February. The second half of the seasonal cycle shown by TRMM is in good agreement with DAPAGLOCO, with the exception of December. The first half (January – May), however, shows pr nearly as small as those seen in the CMORPH data set. The seasonal cycle of ERA-interim data is in good agreement with the observations, if the bias of 0.3 mm/day is disregarded.

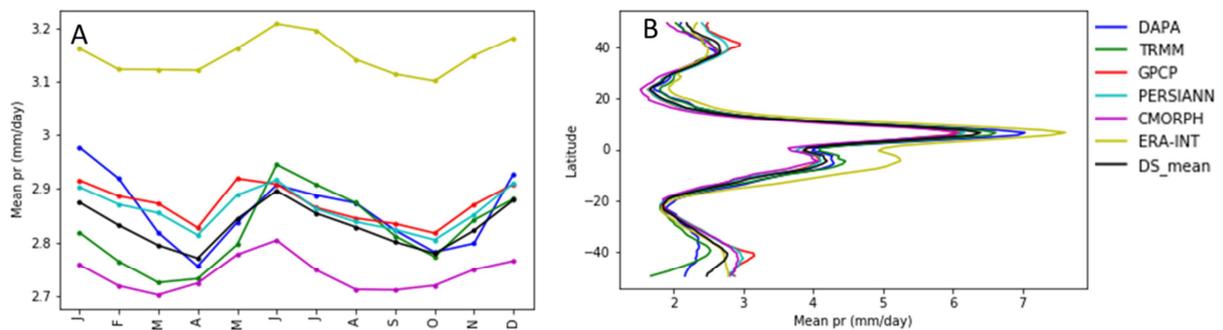


Figure 4. Quasi-global (50°S - 50°N), climatological (1998-2015) mean seasonal cycle (A) and meridional profile (B) of five pr observation data sets, the mean of all observation data sets (DS\_mean), and ERA-interim reanalysis data.

The meridional profiles of all data sets are similarly shaped, as can be seen in Fig. 4B. ERA-interim displays clearly larger pr in the tropics, but is in agreement with the observations at higher latitudes. DAPAGLOCO yields the largest mean pr at the northern hemisphere summer ITCZ, whereas TRMM shows higher mean pr at the winter ITCZ (around 5° S). There is a large variation between the data sets at higher latitudes – particularly south of 30° S. This is mainly due to the insensitivity of satellite retrievals to light rain and the complexity of treating falling snow. The scarcity of station data at these latitudes additionally makes validation or calibration of satellite data difficult.

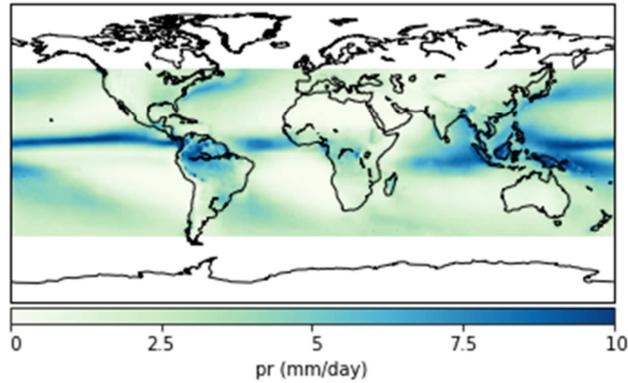


Figure 5. Climatological mean (1998-2015) of mean of five observational data sets, DAPAGLOCO, TRMM, GPCP, PERSIANN, and CMORPH.

To investigate the origin of the differences among the global pr data sets, climatological mean maps were determined from each data set separately and from DS\_mean (the mean of the observational data sets). The latter map is shown in Fig. 5. At a first glance, the differences between the climatological means of DS\_mean and DAPAGLOCO (depicted in Fig. 1A, albeit with a different time range) are small. Figure 6A, however, shows that there are systematic differences on the order of 1-2 mm/day between DAPAGLOCO and DS\_mean, particularly in regions with high mean pr. In comparison to the other data sets, DAPAGLOCO consistently shows higher pr at the ITCZ, the southern Indian Ocean, and at the northeastern Pacific. Over land, higher pr (with respect to DS\_mean) are observed over mountain ranges, such as the Himalayas and Andes, and over Central Africa. The TRMM data set also shows deviations from DS\_mean on the order of 1-2 mm/day (Fig. 6C), but the global pattern is distinctly different from that found for DAPAGLOCO. TRMM finds higher pr mostly near the equator, over the Atlantic, Indian, and western Pacific Ocean. At the ITCZ in the eastern Pacific, Fig. 6C shows a clear underestimation of pr by TRMM (in comparison to DS\_mean), where DAPAGLOCO strongly overestimates mean pr.

The three remaining observational data sets, GPCP, CMORPH, and PERSIANN, each show much smaller differences to DS\_mean, which is to a large extent due to the fact that their algorithms and input data are very similar (see Sect. 2). At this point, it is important to note that the better agreement of a data set to DS\_mean does not necessarily indicate a better quality of that particular data set, but, instead, indicates a higher degree of similarity to the data sets used to generate DS\_mean.

The differences between DS\_mean and ERA-interim reanalysis data sets are clearly much larger than those amongst the observational data sets themselves. The large bias of ERA-interim data with respect to the other data sets, found in the yearly means and seasonal cycle (Figs. 3A and 4A), is clearly due to the overestimation of pr in the tropics (Figs. 4B and 6F).

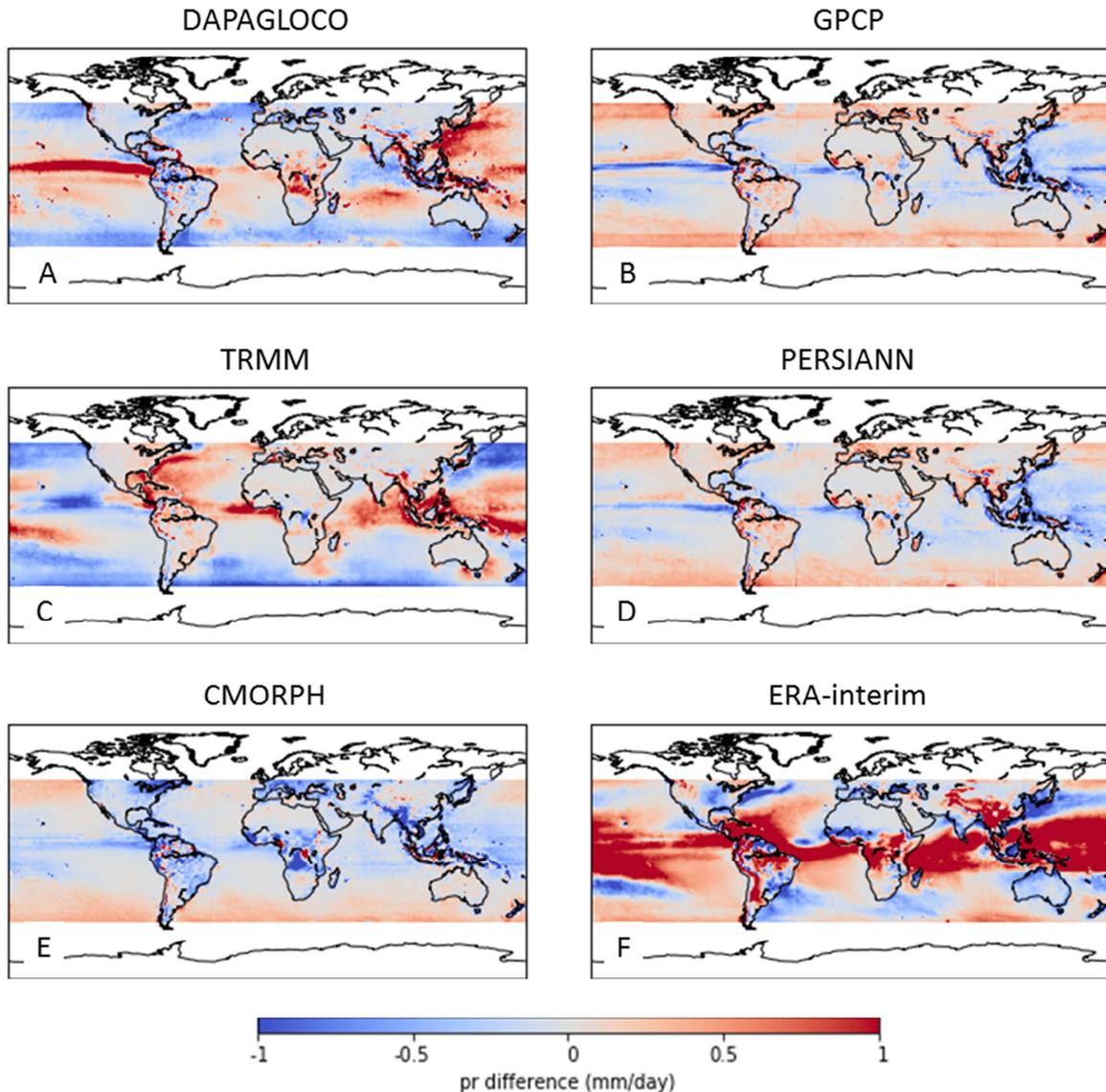


Figure 6. Differences between multi-year means (1998-2015) of five observational data sets and reanalysis data with DS\_mean, respectively. A: DAPAGLOCO, B: GPCP, C: TRMM, D: PERSIANN, E: CMORPH, F: ERA-interim.

#### 4.3. Comparison over ocean

Over ocean, DAPAGLOCO data deviate only little from the operational HOAPS-4 pr product, which was validated by Graw et al. (2017). Here, we discuss global, ocean-only yearly means and anomalies of three observational data sets and ERA-interim reanalysis data, shown in Fig. 7. For this comparison we include only data sets covering 80°S - 80°N, and consider a larger time range from 1996-2014. The bias between DAPAGLOCO and HOAPS-4 of about 0.1 mm/day is mainly due to the difference in the way instantaneous satellite measurements are integrated to daily precipitation totals. Apart from this bias, the data are very similar, except for a stronger ENSO response, most clearly seen in the yearly anomalies in 1999 and 2000. The comparison between DAPAGLOCO and GPCP data over ocean is very similar to that found for the global comparison (Fig. 3): GPCP means are slightly higher and the year-to-year variability is smaller. Also the comparison of ERA-interim with the observations is similar, with a bias of about 0.3 mm/day which increases to 0.4 mm/day over time. In fact, the yearly anomalies determined from reanalysis data are in agreement with observations (particularly GPCP) up to 2001, after which the reanalysis shows a behavior very different from the observations.

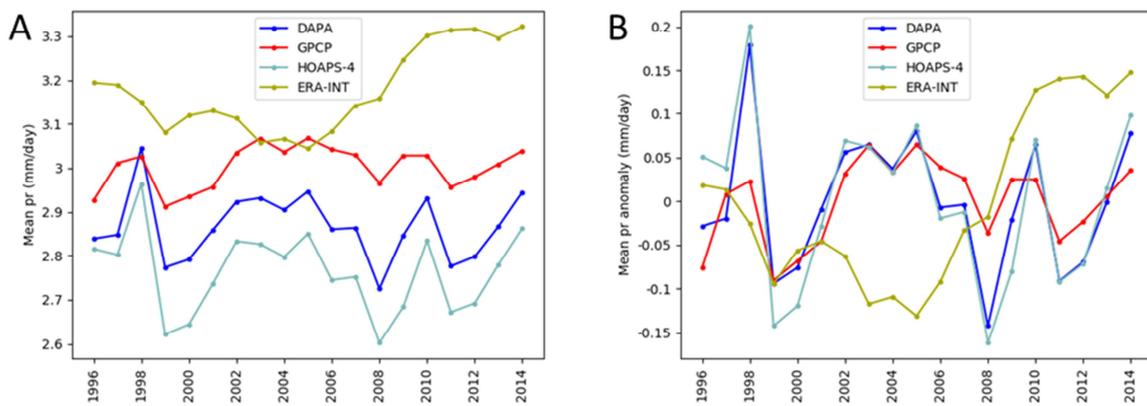


Figure 7. Yearly global (80°S - 80°N) means (A) and anomalies w.r.t. the mean over the whole time period (B) for four data sets evaluated over ocean during 1996-2014.

The mean meridional profiles, shown in Fig. 8, are very similar in structure to those found for the global data sets (in Fig. 4B), with the reanalysis data showing the highest maxima at the summer ITCZ (at 5°N), closely followed by DAPAGLOCO and HOAPS-4. At 5°S, ERA-interim yields mean pr that are nearly 1 mm/day (25%) larger than DAPAGLOCO, which shows the second-highest mean pr there. The differences at latitudes >40° between DAPAGLOCO/HOAPS-4 and GPCP are more pronounced in Fig. 8. This is probably related to the algorithm change that takes place within GPCP at those latitudes: as geostationary data are not available at high latitudes, data from polar-orbiting instruments are used for the retrieval instead (see Sect. 2). Whereas reanalysis data are in good agreement with DAPAGLOCO and HOAPS-4 on the Northern Hemisphere, but are significantly higher at the ITCZ and at latitudes between 30° and 60° S.

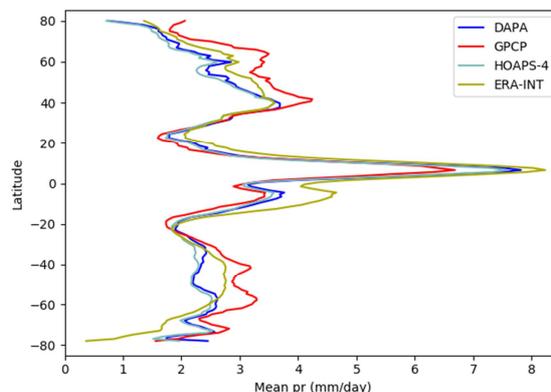


Figure 8. Multi-year (1996-2014) meridional profile of four global (80°S - 80°N) data sets.

#### 4.4. Comparison over land

Over land DAPAGLOCO data consist of gridded, station-based GPCC data (see Sect. 2). As GPCC incorporates data from many stations, it is difficult to find independent data with which to compare. The Japanese project APHRODITE, whose gridded product covers most of Southeast Asia, from 60°E - 155°E and 20°S - 55°N, incorporates many gauges not used within GPCC. The multi-year pr climatology, created from APHRODITE data, is shown in Fig. 9.

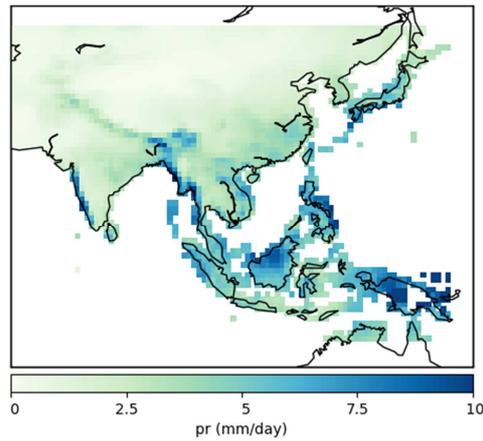


Figure 9. Multi-year mean (1998-2015) of APHRODITE gridded pr.

A comparison of selected global data sets with APHRODITE is shown in Fig. 10. Compared to DAPAGLOCO, GPCP, TRMM, and ERA-interim, APHRODITE finds significantly lower pr over Indonesia. As the number of stations used by GPCP and APHRODITE is similar in Indonesia, the deviations must be caused by differences in quality control and interpolation methods [Yatagai et al., 2012]. The Himalayan mountain range is visible in all difference maps – as a positive bias for DAPAGLOCO and ERA-interim, a negative bias in case of GPCP – except for TRMM. In addition, GPCP finds a positive bias over Thailand and ERA-interim a large positive bias over most of the southeastern tip of the Asian mainland.

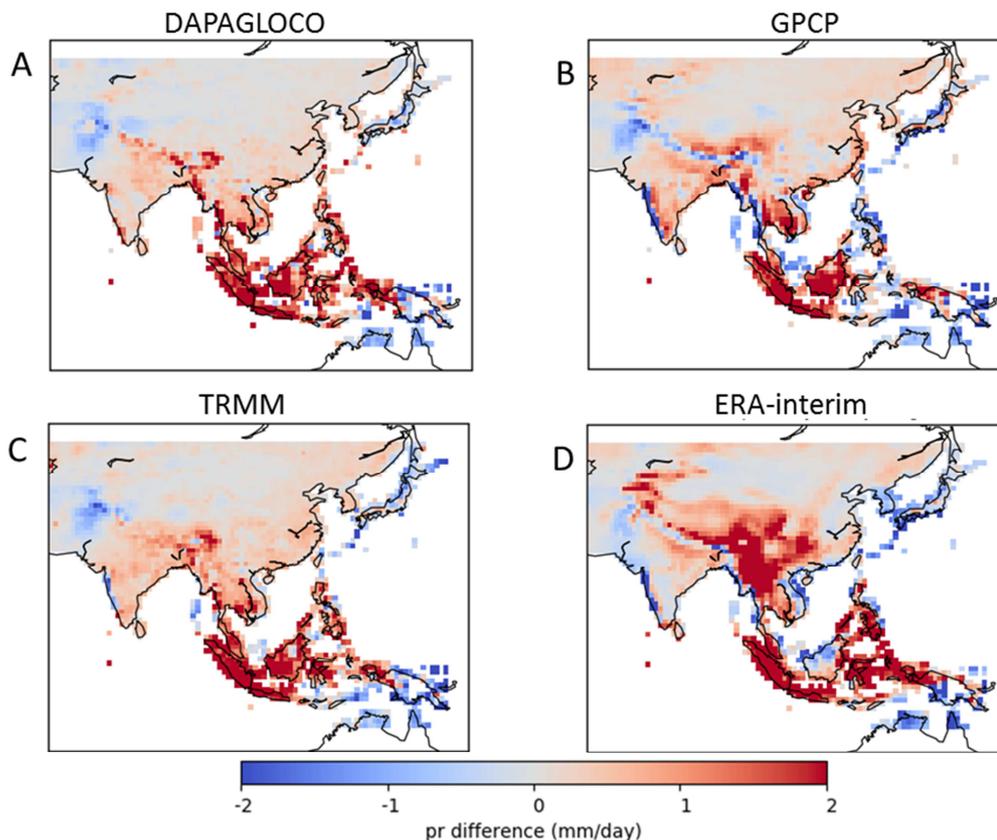


Figure 10. Differences between multi-year means (1998-2015) of four data sets with APHRODITE. A: DAPAGLOCO, B: GPCP, C: TRMM3B42, D: ERA-interim.

The results of a more comprehensive comparison involving eight data sets (DAPAGLOCO, TRMM, GPCP, PERSIANN, CMORPH, CHIRPS, APHRODITE, and ERA-interim) are shown in Fig. 11. DS\_mean was calculated from the satellite data sets (i.e., all except CHIRPS, APHRODITE, and ERA-interim). To avoid cluttering the figures, only DS\_mean, DAPAGLOCO, and the data sets not included in DS\_mean are shown in solid lines; the grey shading around DS\_mean indicates the mean absolute bias of the DS\_mean data sets with respect to DS\_mean, and the error bars show the minimum and maximum values of the data sets contained in DS\_mean.

The inter-annual variation is very similar in all data sets, although ERA-interim somewhat overestimates the variability. The bias between reanalysis data and the other data sets is the order of 0.3 mm/day, but over the selected region this bias stays rather constant over time. The observations are very similar, which is mostly due to the fact that they all use GPCP data to some extent. APHRODITE lies within the range of the global observational products, although it is biased low by 0.2 mm/day compared to DS\_mean. The anomalies agree very well (Fig. 11B), except for the higher year-to-year variability found by ERA-interim and the first three and last three years of APHRODITE data.

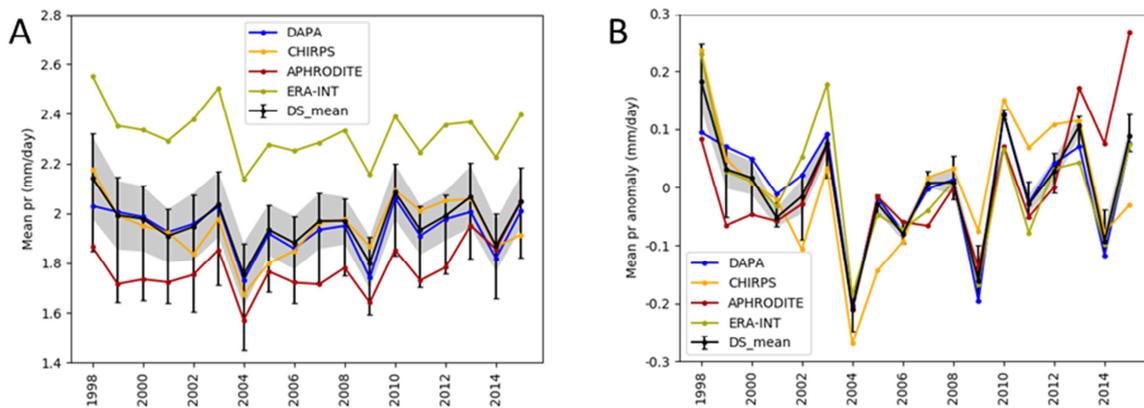


Figure 11. Yearly regional (60°E – 155°E and 20°S – 55°N) means (A) and anomalies w.r.t. the mean over the whole time period (B) for four data sets evaluated over land during 1998-2015. Grey shading around DS\_mean indicates the mean absolute bias of the DS\_mean data sets (DAPAGLOCO, TRMM, GPCP, PERSIANN, and CMORPH) w.r.t. DS\_mean; the error bars show the minimum and maximum values of the DS\_mean data sets.

#### 4.5. Comparison with station data: PACRAIN

The western Pacific is a region that exhibits large differences among different satellite products (see, e.g., Fig. 6 and [Pfeifroth et al., 2013]). The region is characterized by heavy, convection-induced precipitation, contributing disproportionately to the total global precipitation amount.

PACRAIN is currently not used for calibration of merged satellite products. Data from PACRAIN stations are included in GPCP Full Data Daily V.2018, but DAPAGLOCO prefers HOAPS data over ocean (even if GPCP data are available), so that for the purpose of our comparison PACRAIN can be considered as an independent data set. To ensure that no GPCP data were considered in the comparison, grid boxes containing either GPCP or interpolated GPCP/HOAPS data (found near coast lines) were removed from the data set prior to comparison, as shown in Fig. 13.

The length of PACRAIN data records varies widely from station to station, from a few years of measurements to several decades. For this investigation, we directly compare monthly averaged atoll station data with gridded satellite and reanalysis data. Two of such comparisons are shown in Fig.12, for DAPAGLOCO with the PACRAIN stations with IDs US14892 and NZ53300 (marked on the

map in Fig.13). As precipitation is highly variable in space and time, several collocation windows were applied to the satellite data: in the first case, only the grid box containing the station was regarded; in the other cases, data from the four and nine grid boxes surrounding the station were averaged. Panel A of Fig. 12 shows that there is a high degree of agreement between DAPAGLOCO and PACRAIN: the correlation coefficient is 0.69-0.82, depending on which collocation criterion is used; in this case, averaging the four grid boxes surrounding the PACRAIN station yields the highest correlation. Station US14892 has a time series that almost completely coincides with the DAPAGLOCO time range, and is nearly complete (apart from a larger gap in 2012/2013). Station NZ53300, depicted in panel B, on the other hand, has a much shorter data record, which ends in 1999. Measurements may have continued at this station after 1999 under a different station name, but they could not be found in the downloaded data set. The correlation between DAPAGLOCO and PACRAIN is very high at this station – up to 0.86 for the 9-grid box average. The inter-annual variability found in this case is picked up very well by the satellite retrieval, and particularly the low pr are in good agreement.

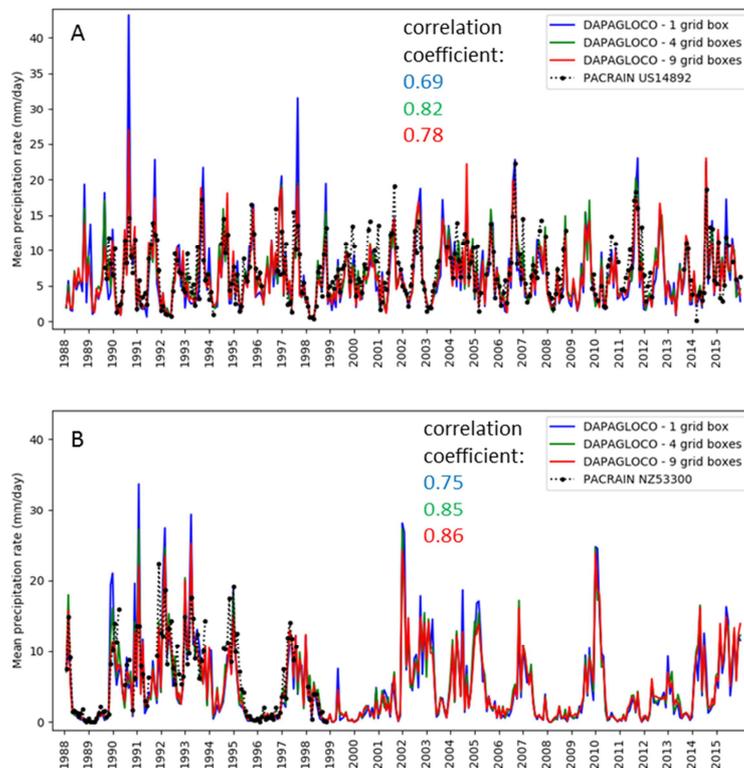


Figure 12. Time series of monthly mean pr. DAPAGLOCO data were averaged over 1, 4, and 9 grid boxes directly surrounding the PACRAIN stations US14892 (at 10°N / 140°E; panel A) and NZ53300 (at 1°S / 170°E; panel B). The corresponding correlation coefficients are given in the figure.

Correlation coefficients with DAPAGLOCO data were computed for monthly mean pr from all PACRAIN atoll stations with a relative temporal coverage of at least 33% (i.e., data for at least 1/3 of the months within the selected time period are available), and the results are shown in Fig. 13. Out of 30 stations, 28 yield correlation coefficients above 0.6; the highest correlation was found for NZ48700 (0.897, red dot in the northeastern section of Fig. 13).

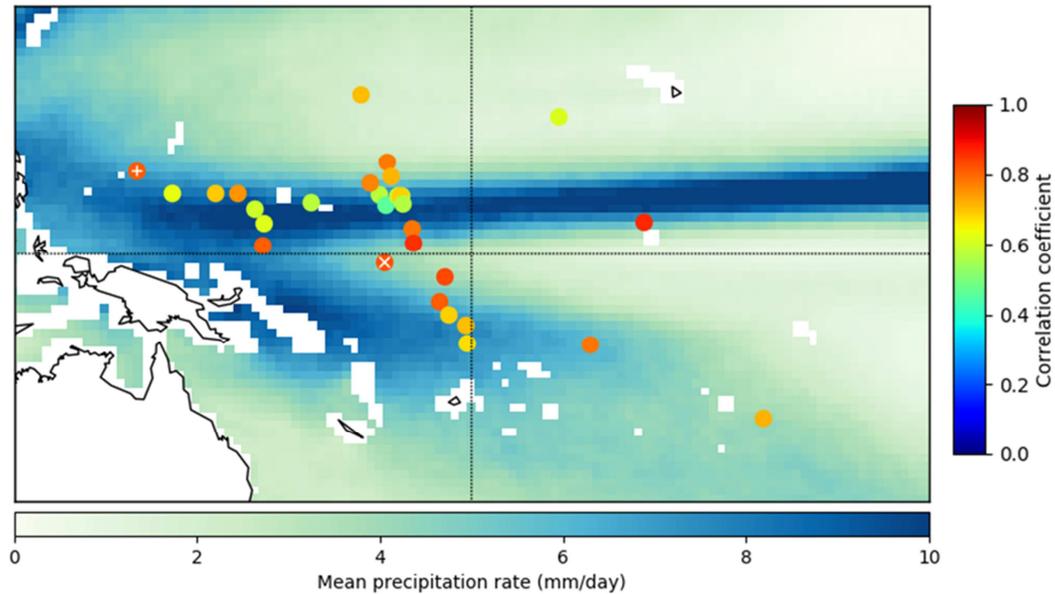


Figure 13. Climatological mean (1988-2015) precipitation rate of DAPAGLOCO (data from HOAPS retrieval only) over the western Pacific Ocean (125°E-235°E / 30°S - 30°N) overlaid with correlation coefficients from selected PACRAIN atoll stations superimposed as filled circles. Only stations where data were available for at least 1/3 of the months within the selected time period are shown. PACRAIN stations US14892 and NZ53300 are marked by a white plus and a white cross, respectively.

Comparisons of PACRAIN monthly mean pr from 30 atoll stations between 1998-2015 with four gridded global data sets are shown in Fig. 14, together with the line of 1:1 correspondence (black), a linear fit to the data (light blue), and a piecewise linear fit with one allowed breakpoint (purple). The total number of collocated data points amounts to 3240. Statistics are summarised in Table 2. From panel A it can be seen that DAPAGLOCO data are highly correlated with PACRAIN, although the retrieval slightly underestimates the higher precipitation rates. The piecewise linear fit (shown in purple) does not deviate strongly from the linear fit, indicating that the correspondence between DAPAGLOCO and PACRAIN does not depend on precipitation rate. This is also the case for TRMM, which yields a smaller slope than the DAPAGLOCO comparison (0.677), but a slightly higher correlation and smaller standard error (0.78 and 0.010, respectively). In contrast, GPCP data (panel C) show a clear break with  $pr < 5$  mm/day in good agreement with PACRAIN (slope<sub>1</sub> = 1.01), and a clear underestimation of higher pr (slope<sub>2</sub> = 0.44). Similar behavior is found for ERA-interim, although the reanalysis overestimates light rain by 40% and underestimates heavy rain even more than GPCP (slope<sub>2</sub> = 0.34). The fact that the gridded data sets underestimate higher pr is not so much an error of the retrieval, but is at least partly caused by the larger footprint of satellite observations, which acts to smooth extreme signals [Pfeifroth et al., 2013].

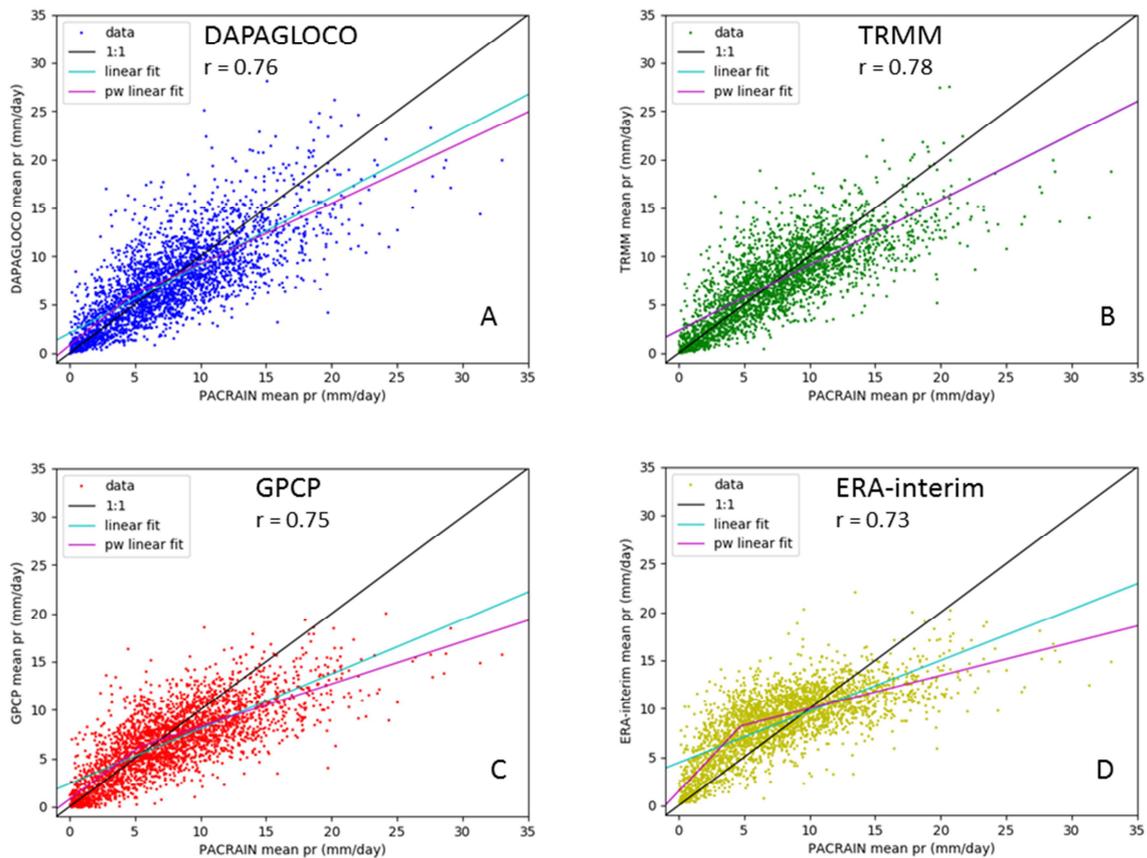


Figure 14. Scatter plots of monthly mean *pr* (1998-2015) from 30 PACRAIN atoll stations with (A) DAPAGLOCO (ocean only), (B) TRMM, (C) GPCP, and (D) ERA-interim. The line of 1:1 correspondence is shown in black, light blue and purple lines show the linear and the piecewise linear fit, respectively (see text for details); the corresponding correlation coefficient of linear regression is given at the top of each panel.

The results shown in Fig. 14 and Table 2 are very similar to those found by [Pfeifroth et al., 2013], who performed a comprehensive comparison of PACRAIN with HOAPS (Version 3), GPCP, ERA-interim, and MERRA reanalyses for 1989-2005. There, correlations of 0.70, 0.77, and 0.75 were found for HOAPS 3, GPCP, and ERA-interim, respectively.

Table 2. Evaluation of monthly mean *pr* from four data sets with PACRAIN stations during 1998-2015. Measures are: slope and intercept of the linear fits: slope\_n and interc\_n (n = 0 for the linear fit, n = 1,2 for the first and second segment of the piecewise linear fit, respectively), linear correlation coefficient r, and the standard error, stderr.

| Data set    | slope_0 | interc_0 | r    | stderr | slope_1 | interc_1 | slope_2 | interc_2 |
|-------------|---------|----------|------|--------|---------|----------|---------|----------|
| DAPAGLOCO   | 0.705   | 2.04     | 0.76 | 0.011  | 1.09    | 4.70     | 0.63    | 5.89     |
| TRMM        | 0.677   | 2.31     | 0.78 | 0.010  | 4.34    | -12.55   | 0.68    | -6.18    |
| GPCP        | 0.566   | 2.40     | 0.75 | 0.0092 | 1.01    | 5.41     | 0.44    | 6.25     |
| ERA-interim | 0.529   | 4.43     | 0.73 | 0.0093 | 1.42    | 4.75     | 0.34    | 8.21     |

## 5. ETCCDI

Up to this point, monthly, yearly, and multi-year mean fields were compared to assess the degree of correspondence among various precipitation data sets. The calculation of average values, however, inevitably leads to the smoothing of extremes, such as intense precipitation events or periods of drought. In addition, since DAPAGLOCO is available at daily resolution, it should be validated at that resolution (the 'scope principle' [Tapiador et al., 2017]). Rather than comparing the daily data themselves in this section a comparison is made of a small selection of precipitation indices as defined by the WCRP Expert Team on Climate Change Detection: ETCCDI [Peterson and Manton, 2008; Zhang et al., 2011] (see also: [www.wcrp-climate.org/etccdi](http://www.wcrp-climate.org/etccdi)).

ETCCDI have been used previously for the investigation of extremes; particularly when observations are compared to model results for the purpose of model evaluation [Alexander and Arblaster, 2008; van den Besselaar et al., 2012; Donat et al., 2013, Herold et al., 2016]. From DAPAGLOCO's predecessor, DAPACLIP, an ETCCDI climatology was determined and presented in [Dietzsch et al., 2017].

Of the 17 precipitation-specific indices, the ETCCDI selected for the current analysis are, first: the simple daily intensity index (SDII), which is essentially a measure for how much, on average, it rains when it rains. That is to say: the mean pr of all days (within a year) with  $pr \geq 1$  mm/day. Second: the highest cumulative amount of precipitation on five consecutive days in a given year (Rx5day); and third: the number of days with very heavy precipitation ( $pr \geq 20$  mm/day) in a year, R20mm. Maps of the SDII of 2014 from DAPAGLOCO, TRMM, GPCP, and ERA-interim data are shown in Figure 15. Panel A shows a pattern that – at least over ocean – bears some similarity with the mean precipitation climatology in Fig. 5, with the highest values over the ITCZ and the northwestern Pacific Ocean.

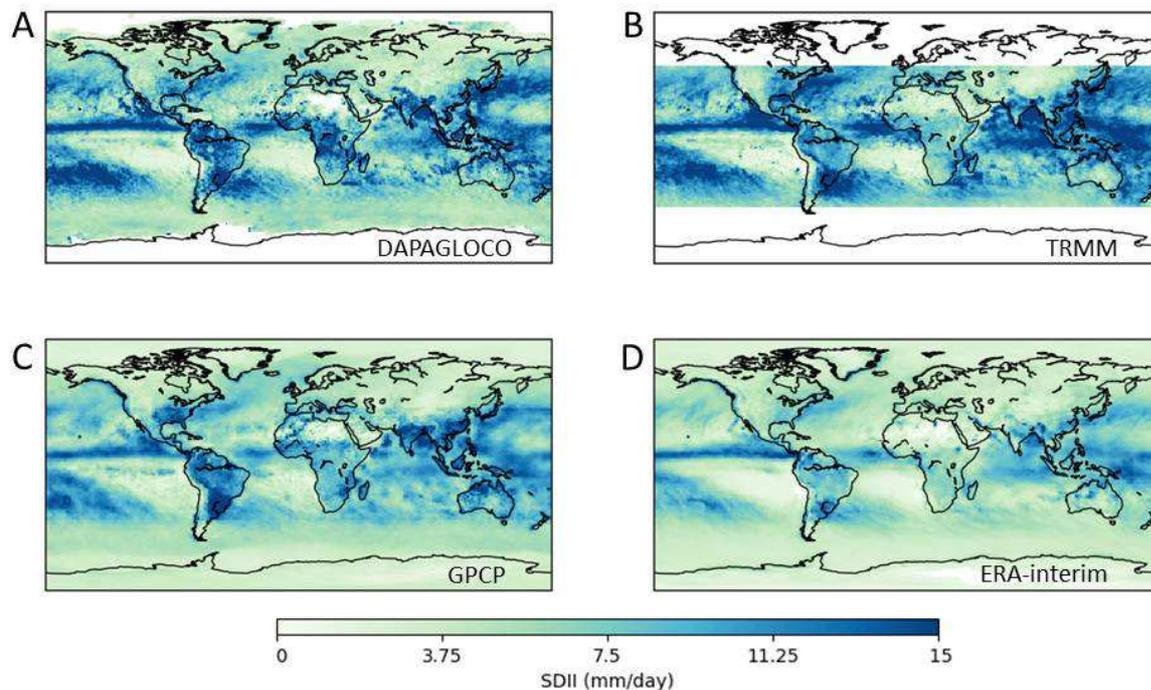


Figure 15. Simple daily intensity index (SDII) of the year 2014 from daily pr from (A) DAPAGLOCO, (B) TRMM, (C) GPCP, and (D) ERA-interim.

Over land, the highest SDII are found over South America, India, and central Africa. The picture provided by TRMM (panel B) is similar, although TRMM systematically finds higher values over ocean

and somewhat smaller values over land (particularly over central Africa). The differences between DAPAGLOCO and GPCP (panel C) are smaller and more random – at least over ocean. In contrast, several land regions show strong, systematic deviations, which is surprising because GPCP is calibrated against GPCC. The largest differences, found over Africa, can be attributed to the scarcity of stations, but this cannot be said for Argentina or the southeastern United States. ERA-interim (panel D) finds similar patterns to DAPAGLOCO, but with a large negative bias of almost 2.5 mm/day. This can also be seen in Fig. 15, where the near-global (50°S - 50°N) yearly mean SDII is shown for the whole overlapping time range, 1998-2015. The bias remains rather constant throughout the time series, owing to the fact that neither ERA-interim, nor DAPAGLOCO display a trend in SDII. In contrast, TRMM data yields a small positive trend and GPCP a negative trend, which leads to a difference between the two data sets of more than 1 mm/day by the end of the time series.

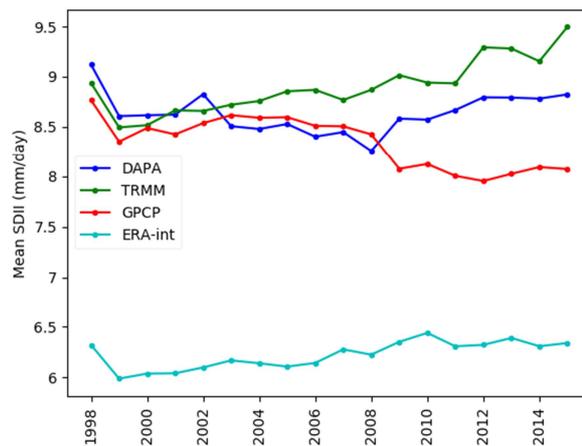


Figure 16. Yearly quasi-global (50°S - 50°N) means of the simple daily intensity index from four data sets.

The maximum cumulative 5-day precipitation index (Rx5day) indicates where high amounts of rain fall over an extended time period, which, over land, is associated with enhanced risk of flooding. Moreover, Rx5day is less sensitive to outliers than R1xday, the maximum daily precipitation index. Figure 17 shows a high degree of agreement regarding the spatial pattern of Rx5day (among all data sets but GPCP), and a large spread in values for the four compared data sets. For this reason, panels A and B have a different color scale (0-200 mm) than panels C and D (0-100 mm). It was noted before, in the comparisons with monthly mean data, that DAPAGLOCO displays the highest variability, and this is clearly seen by the high Rx5day values that this data set attains. In 2014, more than 1750 grid boxes (3.3% of all valid entries) yield Rx5day > 200 mm; for TRMM, this number is 281 (0.8%), for ERA-interim 131 (0.2%); the maximum Rx5day found for GPCP is less than 100 mm.

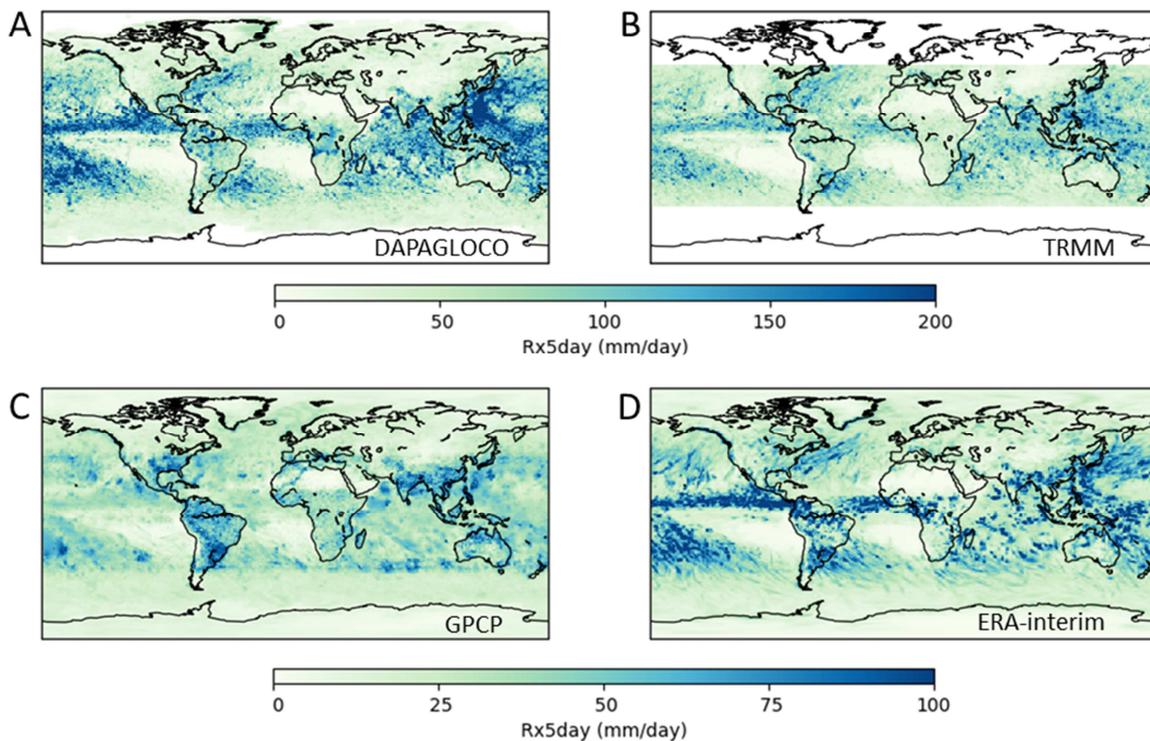


Figure 17. Comparison of Rx5mm (maximum cumulative 5-day precipitation in 2014) between four data sets. (A) DAPAGLOCO, (B) TRMM, (C) GPCP, (D) ERA-interim. Note that the color scale of panels A and B spans 200 mm, whereas that of panels C and D goes up to 100 mm.

The index R20mm represents the number of days in a year when  $pr \geq 20$  mm/day and is shown for DAPAGLOCO in panel A of Figure 18. Again, the spatial pattern is very similar to that found for the multi-year pr climatology or that of the other ETCCDI studied here: the ITCZ, the northwestern Pacific and Atlantic yield the largest number of days with heavy precipitation. Over land, northern South America (Amazon region) experiences by far the most very wet days, followed by Indonesia and Central Africa. The difference plots between R20mm of the three remaining data sets on the one hand, and DAPAGLOCO on the other, are shown in panels B-D of Fig. 18. Somewhat surprisingly – in view of the findings with Rx5day – TRMM and GPCP are in good agreement with DAPAGLOCO, with TRMM even showing systematically higher R20mm than DAPAGLOCO over most of the ITCZ and the western North Atlantic. ERA-interim yields R20mm almost a factor of 2 smaller over South America, Central Africa, and Indonesia, but is overall in good agreement with the observations.

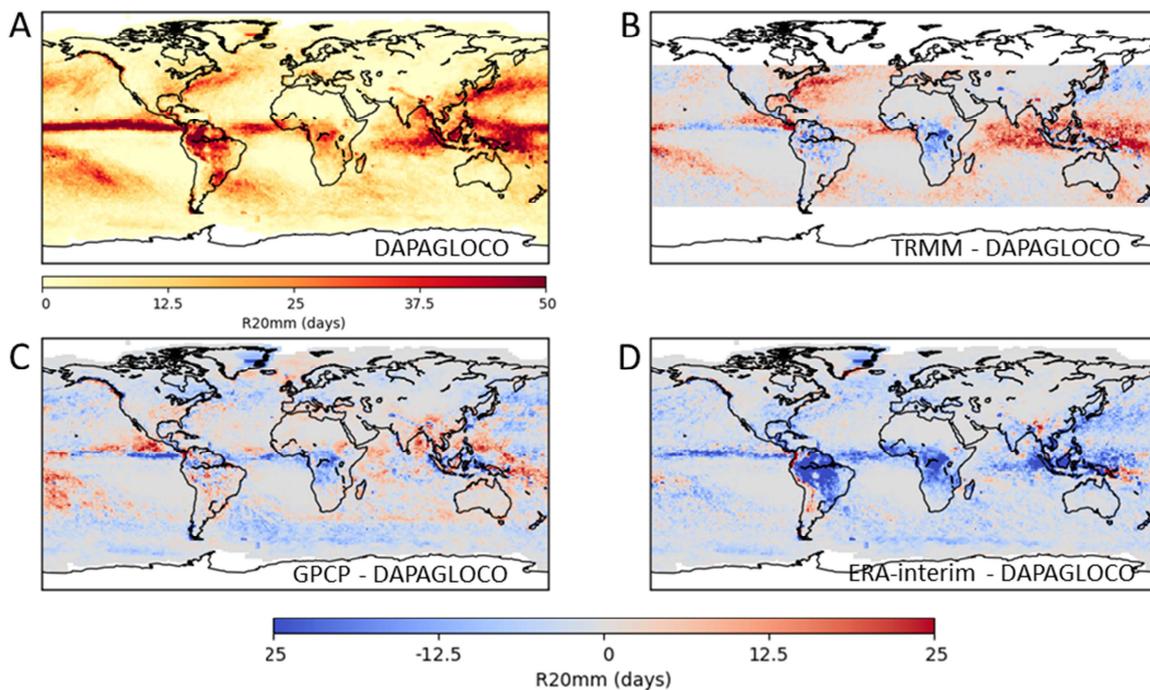


Figure 18. Comparison of R20mm (number of days in 2014 when  $pr \geq 20$  mm/day) between four data sets. (A) R20mm of DAPAGLOCO, (B) R20mm difference of TRMM - DAPAGLOCO, (C) R20mm difference of GPCP - DAPAGLOCO, (D) R20mm difference of ERA-interim - DAPAGLOCO.

## 6. Discussion

A large number of validation and comparison studies of satellite precipitation products (spp) has been performed (see, e.g., the reviews by [Kidd and Huffman, 2011; Maggioni et al., 2016; Tapiador et al., 2017]), and there is a general consensus on several points:

- Largest absolute differences are found where  $pr$  is high (e.g., at the ITCZ), largest relative differences at latitudes  $>40^\circ$ .
- Orographic enhancement of rainfall cannot be well resolved from satellite.
- On monthly and longer time scales, spatial variability is rather consistent among spp, as these means are governed by moderate  $pr$  (1-10 mm/hr), which can be well retrieved by spp; the intensities, however, vary significantly.

With respect to reanalyses:

- In validation studies, reanalysis data are generally outperformed by spp.
- Reanalyses tend to overestimate small to medium  $pr$ , and underestimate higher  $pr$ .

No single technique or methodology is superior to any other in all regions and under all conditions, as each has its own particular strengths and weaknesses. For example, light precipitation and warm rain events are typically underestimated by PMW retrievals, whereas reanalyses have trouble correctly representing convective precipitation. PMW retrievals are good at retrieving instantaneous precipitation rates, whereas IR data better capture spatial and temporal patterns.

Our current findings lie along the same lines. The global and seasonal patterns of the compared satellite-based  $pr$  data sets are very similar. Apart from CMORPH, which is biased low by 0.1 mm/day, the yearly global means (Fig. 3A) of all observational data sets lie within  $\pm 0.05$  mm/day of DS\_mean –

with the exceptions of DAPAGLOCO during the El Niño events in 1997/98 and 2014/15, and the La Niña year 2008. ERA-interim data shows a positive bias of nearly 0.3 mm/day (about 10%), which steadily increases between 2006 and 2010 to almost 0.5 mm/day. The interannual variability of the global mean of all investigated satellite-based data sets is very similar (Fig. 3B), apart from the stronger responses to the ENSO phase by DAPAGLOCO. The interannual variability displayed by TRMM, although being of the same magnitude, appears to be out of phase with the other data sets. The agreement among the determined climatological mean seasonal cycles is high (Fig. 4A). Both the shape and the amplitude (about 1 mm/day between minimum and maximum) are very similar for GPCP, CMORPH, PERSIANN, and ERA-interim. DAPAGLOCO and TRMM both show a more pronounced seasonal cycle (about 1.5 and 2.0 mm/day, respectively), which – at least in the case of DAPAGLOCO – can be attributed to the stronger ENSO response.

The positive bias in the global mean pr from ERA-interim is due to an overestimation of pr in the rain-rich tropics (Fig. 6). In view of the fact that SDII, Rx5day, and R20mm of ERA-interim data are all smaller than those found for DAPAGLOCO (Figs. 15-18), the cause of this overestimation must be a higher frequency of days with moderate rainfall. DAPAGLOCO's high bias of pr over the ITCZ and the Kurushio current, on the other hand, is characterized by elevated SDII, Rx5day, and R20mm (compared to the other observations). The same is valid for TRMM, although the smaller Rx5day (compared to DAPAGLOCO) indicates that the maximum pr retrieved by TRMM are below those retrieved by DAPAGLOCO.

Over land, the compared observational data sets show a very high degree of similarity – which is not surprising, considering the fact that all depend on GPCP data. The APHRODITE data set, which partially relies on the same station data as GPCP, shows a very similar interannual variability, but is biased low. Precipitation over the East Asian mainland is somewhat lower in APHRODITE, but most of the bias comes from the discrepancy over Indonesia. In that region, both networks use a comparable amount of stations, and it remains unclear where the difference originates: possible sources are differences in quality control and in the interpolation of station data to a regular grid [Yatagai, 2012]. ERA-interim performs poorly in Southeast Asia, which is due to the complexities of parameterizing convective precipitation.

Our findings from the comparison of satellite and reanalysis data sets with PACRAIN corroborate the findings of Pfeifroth et al. (2013), who performed a similar exercise with an earlier version of HOAPS (3.2). Similar good correlations of 0.73-0.76 (0.70-0.77 in the previous study) were found for monthly means of DAPAGLOCO/HOAPS, GPCP, and ERA-interim with PACRAIN station data. Interestingly, the correlation of DAPAGLOCO with PACRAIN was significantly better if four or nine 1°x1° grid boxes around the station were averaged, than if the value from the grid box containing the station was taken, a finding also described in Pfeifroth et al., (2013). Perhaps most importantly, both studies found a good correlation and a reasonable slope (0.7) for DAPAGLOCO. There was no noticeable breakpoint in scatter plots in either DAPAGLOCO or TRMM data – in contrast to GPCP and ERA-interim, which both strongly underestimate pr > 5-10 mm/day in this region.

Some of the conclusions from the ETCCDI comparison have been discussed above, but a few interesting aspects remain. First, the observation that GPCP's algorithm switch at 40° from the equator is much more apparent in the SDII and Rx5day than in the multi-year means: both drop abruptly by 50-100% when going towards higher latitudes. Second, that the global mean SDII of DAPAGLOCO, TRMM, and GPCP were very similar until 2009, when TRMM started drifting upwards, and GPCP in the opposite direction. It is interesting to note that the inter-annual variability of SDII is much smaller than that of mean pr (compare Figs. 3 and 16).

Third, the general underestimation of strong precipitation events by GPCP compared to DAPAGLOCO and TRMM, particularly over the oceans, is evident in the small values of Rx5day (Fig. 17).

Nevertheless, the frequency of days with heavy precipitation (R20mm) is very similar among the three satellite products (Fig. 18).

To conclude: DAPAGLOCO is a precipitation data set that compares very well with a suite of existing data sets. Although it is a merged data set of gauge data over land and satellite data over ocean, the two algorithms are not merged or blended. Over land, DAPAGLOCO purely depends on gauges; over ocean, purely on PMW satellite measurements. This leads to overall good agreement with many other observational data sets over land, which rely on GPCC data to some extent. Considering the fact that no single retrieval is superior to all others in all respects, DAPAGLOCO is a useful addition to the suite of existing satellite precipitation data sets.

## 7. References

- Alexander, L.V., and J.M. Arblaster, 2009: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol.*, 29, 3, 417-435, doi: 10.1002/joc.1730
- Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., and Schulz, J., 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data – HOAPS-3. *Earth Syst. Sci. Data*, 2, 215-234, doi: 10.5194/essd-2-215-2010.
- Anderson, A., K. Fennig, M. Schröder, R. Hollmann, and M. Werscheck, 2011: CM SAF Algorithm Theoretical Basis Document Issue 1.1. HOAPS release 3.2, SAF/CM/DWD/ATBD/HOAPS
- Besselaar, E.J.M. van den, A.M.G. Klein Tank, and T.A. Buishand, 2012: Trends in European Precipitation Extremes over 1951-2010. *Int. J. Climatol.*, 33, 12, 2682-2689, doi: 10.1002/joc.3619
- Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011: The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System. *Q. J. R. Meteorol. Sci.*, 137, 553-597, doi: 10.1002/qj.828
- Dietzsch, F., A. Andersson, M. Ziese, M. Schröder, K. Raykova, K. Schamm, and A. Becker, 2017: A Global ETCCDI-Based Precipitation Climatology from Satellite and Rain Gauge Measurements. *Climate*, 5(1), 9, doi: 10.3390/cli5010009
- Donat, M.G., L.V. Alexander, and H. Yang, 2013: Global Land-Based Datasets for Monitoring Climatic Extremes. *Bull. Am. Met. Soc.*, 94, 997-1006, doi: 10.1175/BAMS-D-12-00109.1
- Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen, 2015: The Climate Hazards Infrared Precipitation with Stations – A New Environmental Record for Monitoring Extremes. *Scientific Data* 2, 150066, doi: 10.1038/sdata.2015.66
- Graw, K., Kinzel, J., Fennig, K., Schröder, M., Hollmann, R., and M. Werscheck, 2017: Validation report SSM/I and SSMIS products – HOAPS Version 4.0. EUMETSAT, SAF/CM/DWD/PUM/HOAPS/2
- Greene, J.S., M. Klatt, M. Morrissey, and S. Postawko, 2007: The Comprehensive Pacific Rainfall Database. *J. Atmos. Oceanic Technol.* 25(1), 71-82, doi: 10.1175/2007JTECHA904.1
- Herold, N., A. Behrangi, and L.V. Alexander, 2016: Large Uncertainties in Observed Daily Precipitation Extremes over Land. *J. Geophys. Res.*, 122, 668-681, doi: 10.1002/2016JD025842
- Huffman, G.J., R.F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider, 1997: The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bull. Am. Met. Soc.*, 78, 5-20, doi: 10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2
- Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, and D.B. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, 8, 38-55, doi: 10.1175/JHM560.1
- Hsu, K., X. Gao, S. Sorooshian, and H.V. Gupta, 1997: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteor.*, 36, 1176-1190, doi: 10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2
- Hsu, K., H.V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of Physical Variables from Multi-Channel Remotely Sensed Imagery Using a Neural Network: Application to Rainfall Estimation. *Water Resour. Res.*, 35, 1605-1618, doi: 10.1029/1999WR900032
- Jost, V., 2000: HOAPS: Eine neue Klimatologie des Süßwasserflusses an der Meeresoberfläche abgeleitet aus Satellitendaten, PhD Thesis, University of Hamburg, ISSN 0938-5177
- Maggioni, V., P.C. Meyers, and M.D. Robinson, 2016: A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *J. Hydrometeor.*, 17, 1101-1117, doi: 10.1175/JHM-D-15-0190.1
- Peterson, T.C., and M.J. Manton, 2008: Monitoring Changes in Climate Extremes: A Tale of International Collaboration. *Bull. Am. Met. Soc.*, 89, 9, 1266-1271, doi: 10.1175/2008BAMS2501.1

- Pfeifroth, U., Mueller, R., Ahrens, B., 2013: Evaluation of Satellite-Based and Reanalysis Precipitation Data in the Tropical Pacific. *J. Appl. Meteor. Climatol.*, 52, 634-644, doi: 10.1175/JAMC-D-12-049.1
- Schamm, K., M. Ziese, A. Becker, P. Finger, A. Meyer-Christoffer, U. Schneider, M. Schröder, and P. Stender, 2014: Global Gridded Precipitation over Land: A Description of the New GPCP First Guess Daily Product. *Earth Syst. Sci. Data*, 6, 49-60, doi: 10.5194/essd-6-49-2014
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCP's New Land Surface Precipitation Climatology Based on Quality-Controlled In Situ Data and Its Role in Quantifying the Global Water Cycle. *Theor. Appl. Climatol.*, 115, 15-40, doi: 10.1007/s00704-013-0860-x
- Schneider, U., M. Ziese, and A. Becker, 2018: Global Precipitation Analysis Products of the GPCP, Deutscher Wetterdienst, Weltzentrum für Niederschlagsklimatologie, [https://opendata.dwd.de/climate\\_environment/GPCP/PDF/GPCP\\_intro\\_products\\_v2018.pdf](https://opendata.dwd.de/climate_environment/GPCP/PDF/GPCP_intro_products_v2018.pdf)
- Volter, K. and M.S. Timlin, 1998: Measuring the strength of ENSO events: How does 1997/98 rank? *Weather*, 53: 315-324. doi:10.1002/j.1477-8696.1998.tb06408.x
- Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasuomi, and Akia Kitoh, 2012: APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bull. Am. Met. Soc.*, 93, 1401-1415, doi: 10.1175/BAMS-D-11-00122.1
- Zhang, X., L. Alexander, G.C. Hegerl, P. Jones, A. Klein Tank, T.C. Peterson, B. Trewin, and F.W. Zwiers, 2011: Indices for Monitoring Changes in Extremes Based on Daily Temperature and Precipitation Data. *Wiley Interdiscip. Rev. Clim. Change*, 2, 6, 851-870, doi: 10.1002/wcc.147

## 8. Glossary

|           |  |
|-----------|--|
| AIRS      | Atmospheric Infrared Sounder   |
| AMSR-E    | Advanced Microwave Scanning Radiometer - Earth Observing System  |
| AMSU      | Advanced Microwave Sounding Unit   |
| AVHRR     | Advanced Very High Resolution Radiometer   |
| CMORPH    | CPC morphing technique   |
| CM SAF    | Climate Modelling Satellite Application Facility   |
| CPC       | Climate Prediction Center  |
| DAPAGLOCO | Daily Precipitation Analysis for the Validation of Global Medium-Range Climate Predictions Operationalised |
| DMSF      | Defense Meteorological Satellite Program   |
| ECMWF     | European Centre for Medium-Range Weather Forecasts   |
| ETCCDI    | Expert Team on Climate Change Detection Indices  |
| GPCC      | Global Precipitation Climatology Centre  |
| GPCP      | Global Precipitation Climatology Project   |
| GPCP-SG   | Global Precipitation Climatology Project Satellite-Gauge product   |
| HOAPS     | Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data   |
| IFS       | Integrated Forecasting System  |
| ITCZ      | Inter-tropical convergence zone  |
| MiKlip    | Mittelfristige Klimaprognosen – a project funded by the German Ministry of Education and Research          |
| NOAA      | National Oceanic and Atmospheric Administration  |
| PERSIANN  | Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks                 |
| pr        | precipitation rate   |
| PR        | Precipitation Radar  |
| Rx5       | Maximum cumulative 5-day precipitation   |
| R20mm     | Number of days in a year with $pr \geq 20$ mm/day  |
| SDII      | Simple Daily Intensity Index   |
| SSM/I     | Special Sensor Microwave/ Imager   |
| SSMIS     | Special Sensor Microwave Imager/ Sounder   |
| TB        | Brightness Temperature   |
| TIGriDS   | Tool for the Intercomparison of Gridded Data Sets  |
| TMI       | TRMM Microwave Imager  |
| TMPA      | TRMM Multi-satellite Precipitation Algorithm   |
| TOVS      | Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder                         |
| TRMM      | Tropical Rainfall Measuring Mission  |
| WCRP      | World Climate Research Programme   |