Climate Data Record (CDR) Program

Climate Algorithm Theoretical Basis Document (C-ATBD)

Precipitation – PERSIANN-CDR

CDR Program Document Number: CDRP-ATBD-0286
Configuration Item Number: 01B-16
Revision 2 / March 21, 2014

REVISION HISTORY

Rev.	Author	DSR No.	Description	Date
1	Kuolin Hsu, Hamed Ashouri, Dan Braithwaite, and Soroosh Sorooshian from CHRS at UC Irvine	DSR-319	Initial Submission to CDR Program	8/29/2013
2	Kuolin Hsu, Hamed Ashouri, Dan Braithwaite, and Soroosh Sorooshian from CHRS at UC Irvine		Revision for PERSIANN-CDR v01r01	3/21/2014

TABLE of CONTENTS

LIST of FIGURES

LIST of TABLES

1. Introduction

1.1 Purpose

The purpose of this document is to describe the algorithm submitted to the National Climatic Data Center (NCDC) by Soroosh Sorooshian at the Center for Hydrometeorology and Remote Sensing, at University of California, Irvine, that will be used to create the PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information using an Artificial Neural Network-Climate Data Record), using the long wave infrared images from geosynchronous satellites. The actual algorithm is defined by the computer program (code) that accompanies this document, and thus the intent here is to provide a guide to understanding that algorithm, from both a scientific perspective and in order to assist a software engineer or end-user performing an evaluation of the code.

1.2 Definitions

PERSIANN rain rate estimates are generated as a 0.25 degree resolution product that is then calibrated to the monthly 2.5 degree merged in-situ and satellite product produced by the Global Precipitation Climatology Project. GridSat-B1 IRWIN data are used as input to the PERSIANN model at 0.25 degree resolution and a 3-hourly time step. The output from the PERSIANN model (before bias correction) is called PERSIANN-B1. A threshold (*thd*) value needs to be applied to the 3-hourly PERSIANN-B1 rain rate estimates to filter out noisy pixels. These noisy pixels are generally associated with pixels where the rain rate is “zero” but the Neural Network model estimates a very small nonzero value. While the resulting noisy pixels may not affect the adjustment process considerably, they can lead to a very large number of "rainy" days (rain rate > 0 mm/day). The PERSIANN-B1 data will be accumulated to monthly and 2.5 degrees for GPCP bias weight calculation (w). We note that in some locations, such as high latitudes and in dry regions with very low rainfall values, w can become large. This can lead to unreasonably large daily rainfalls in finer resolution. In order to prevent such cases, we applied a cap for the maximum weight. In order to find the best combination of *thd* and maximum w, an optimization model was developed with the objective of finding the combination which gives the minimum Mean Absolute Error (MAE) between GPCP-1DD and PERSIANN-CDR (up-scaled to 1^o). The results show that $thd = 0.1$ and maximum $w = 20$ is perhaps the best combination.

Adjusted 3-hourly 0.25 degree PERSIANN-B1 estimates are obtained using the weight calculation. And, finally, the daily 0.25 degree PERSIANN-CDR is accumulated from the 3-hourly bias-adjusted product. The adjusted daily PERSIANN-CDR is consistent to the GPCP rainfall at the monthly scale. Following symbols are defined to better address the processes.

> $r_{PERSIANN\text{-}B1}(i,j)$ -- *is the 0.25° 3-hourly PERSIANN rain rate estimates from GridSat-B1 IRWIN data at row i and column j.*
>
> $R_{PERSIANN\text{-}B1}$ -- *is the 2.5° monthly accumulation of 0.25° 3-hourly* $r_{PERSIANN\text{-}B1}(i,j) > 0.1$ *mm/hr*

R_{GPCP} -- is the monthly 2.5° Global Precipitation Climatology Project (GPCP)

$w = R_{GPCP} / R_{PERSIANN\text{-}B1}$ -- for each corresponding pixel a weight (w) is calculated. Maximum weight is set to 20.

$Adj\_r_{PERSIANN\text{-}B1}(i,j) = w * r_{PERSIANN\text{-}B1}(i,j)$ -- for each original 0.25° 3-hourly pixel the bias corrected value is calculated using the weight (w) from the corresponding monthly relationship. $Adj\_r_{PERSIANN\text{-}B1}(i,j)$ is the GPCP-based calibrated 0.25° 3-hourly $r_{PERSIANN\text{-}B1}$ rain rate estimate at row i and column j.

PERSIANN-CDR (i,j) – daily accumulation of 0.25° 3-hourly $Adj\_r_{PERSIANN\text{-}B1}(i,j)$ rain rate estimates

1.3 Document Maintenance

As changes are requested by the NCDC CDR program or if our algorithm, programming or data flow require changes we will check this document carefully and make appropriate changes.

2. Observing Systems Overview

2.1 Products Generated

In brief, PERSIANN Precipitation Climate Data Record, here after called PERSIANN-CDR, is a daily near global precipitation product for the period of 1983 to 2012. The data covers from 60°S to 60°N and 0° to 360° longitude at 0.25 degree spatial resolution. This relatively long record of high resolution near global precipitation estimates is particularly useful for climate studies.

2.2 Instrument Characteristics

The primary input data for the PERSIANN-CDR algorithm comes from another CDR: Gridded Satellite Data from ISCCP B1 (GridSat-B1) IR Window Channel. The GridSat-B1 data are combined from the various geostationary satellites available over the years from 1980-present. The infrared (IR) data from the imager instruments on board these satellites varies in wavelength but should be approximately 10.0 - 12.0μm. The GridSat-B1 data set merges and calibrates these inputs to provide the best available near global coverage (70°N to 70°S) for every 3 hour time step.

The other input data is the Global Precipitation Climatology Project (GPCP) v2.2 2.5 degree monthly precipitation analyses derived from both satellite and ground data sources. See documentation at: http://precip.gsfc.nasa.gov/gpcp_v2.2_comb_new.html

3. Algorithm Description

3.1 Algorithm Overview

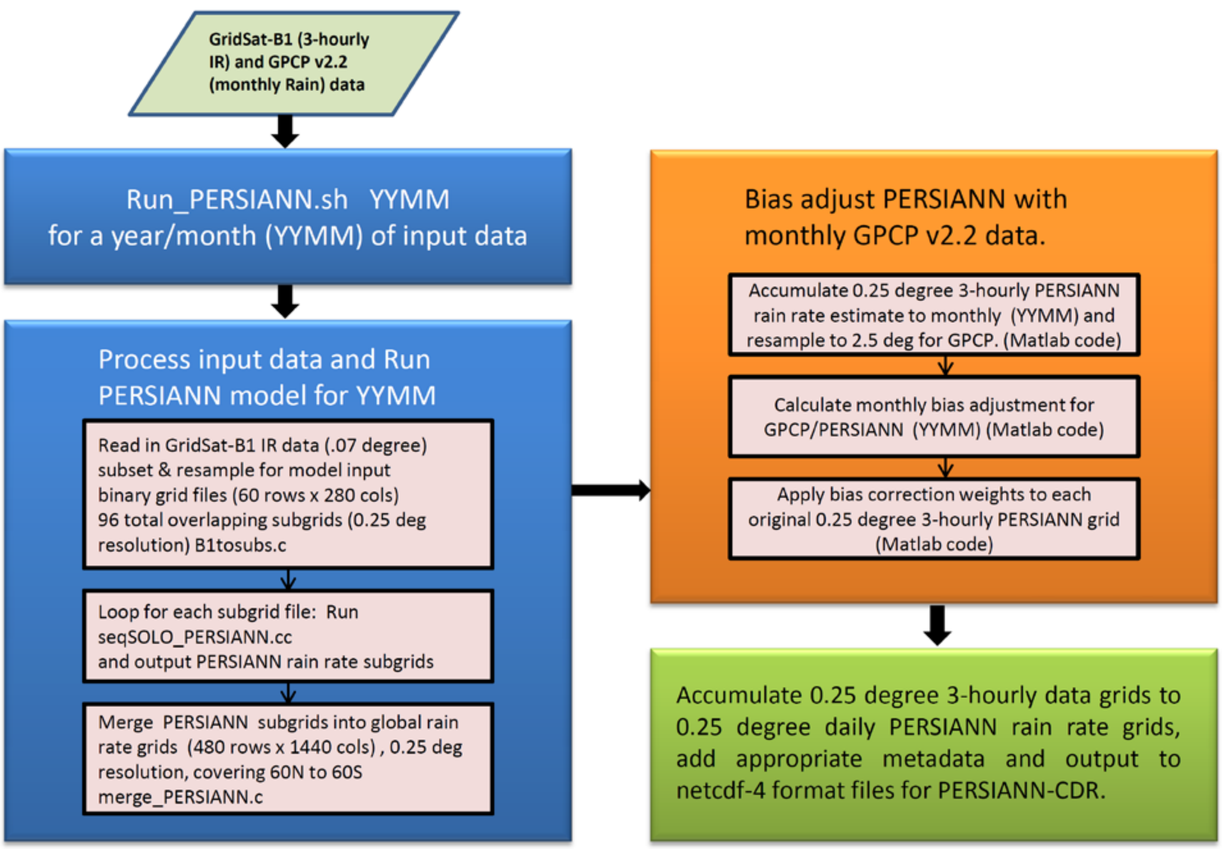
The PERSIANN-CDR product is to be generated for each time step by first estimating precipitation for each GridSat-B1 Infrared Window (IRWIN) file using the basic PERSIANN algorithm which uses an artificial neural network (ANN) to convert the input IR data in degrees Kelvin into rain rate (RR) data in mm/hr. Each month of PERSIANN estimates is then bias corrected with monthly GPCP precipitation data and the final PERSIANN-CDR product results when those bias correct precipitation estimates are accumulated to daily.

3.2 Processing Outline

Current processing consists of shell scripts, C-code, C++ code, and MATLAB code. The final product is produced with bias corrections from the monthly GPCP v2.2 data set, so a month of data is processed at the same time.

Run_persiann.sh YYMM will run the main script for a given year (YY) and month (MM).

The PERSIANN-CDR data flow chart is included here and will be updated as necessary to reflect any future changes.



3.3 Algorithm Input

3.3.1 Primary Sensor Data

The raw sensor data is not used directly. The infrared data gridded from GridSat-B1 is used and converted to a derived format before input into the model.

As the custodian of major climate data sets, the NOAA National Climatic Data Center (NCDC) maintains an historical archive of data from geostationary (GEO) satellites, compiled by the International Satellite Cloud Climatology Project (ISCCP). ISCCP B1, which also is available online (http://www.ncdc.noaa.gov/oa/rsad/isccpb1), is comprised of all-channel observations from a number of international GEO satellites including the Geostationary Operational Environmental Satellite (GOES) series, the European Meteorological satellite (Meteosat) series, the Japanese Geostationary Meteorological Satellite (GMS) series, and the Chinese Fen-yung 2C (FY2) series. The global ISCCP B1 IR brightness temperature from these GEO sources covers the time period from 1979 to present, at space and time resolutions of 10-km and 3-hour intervals. Better coverage began in 1983, albeit with a gap over the Indian Ocean due to the lack of GEO measurements. Other sources, such as AVHRR (Advanced Very High Resolution Radiometer) from the NOAA series of satellites (NOAA-7~NOAA-14), can be used to fill part of the gap before 1997.

Gridded Satellite (GridSat-B1) is the gridded derived product from ISCCP B1 data and is accessible online (http://www.ncdc.noaa.gov/oa/gridsat/index.php). GridSat-B1 data provide data for three channels: visible data, infrared window (IRWIN) data, and infrared water vapor (IRWVP) data. The Infrared Window (IRWIN) data is the main input data for the PERSIANN model. The IRWIN data are gridded on a 0.07 degree latitude equal-angle grid. By selecting the nadir-most observations at each grid point, satellite data are merged. The infrared window brightness temperature – identified as GridSat-B1 CDR – is online. (ftp://eclipse.ncdc.noaa.gov/pub/gridsat/b1-climate-data-record/). GridSat-B1 CDR starts from 1 January 1980 and continues to the current time. It covers the globe from 70S to 70N and 180W to 180E. The data are stored in NetCDF version 4 format. Each time step is provided in a separate NetCDF-4 file (with compression built in) of about 8.2 MB in size.

3.3.2 Ancillary Data

The Global Precipitation Climatology Project (GPCP) was established in 1986 by the World Climate Research Programme (WCRP) to document the spatial and temporal distribution of precipitation at climate scale (Adler et al. 2003). Currently three GPCP (version 2.2) global precipitation products are available: (1) Monthly, 2.5^{o} merged analysis (1979-present), (2) Pentad, 2.5^{o} merged analysis (1979-present), and (3) Daily, 1^{o} merged analysis (Oct. 1996-present). Briefly, the GPCP-v2.1 Monthly 2.5° merged analysis was constructed using multi-satellite (SSMI and IR) precipitation estimates, adjusting the latter using gauge analysis to remove large-scale bias, and then merging satellite and gauge analysis into a final product (Huffman et. al, 1997). The GPCP-v2.2 monthly data are now available at http://precip.gsfc.nasa.gov and currently span January 1979 - June 2011.

GPCP v2.2 monthly precipitation files are used to bias correct the PERSIANN model output data. These GPCP files are described in the GPCP documentation:
http://precip.gsfc.nasa.gov/gpcp_v2.2_comb_new.html

"The data set archive consists of yearly unformatted REAL*4 binary files with ASCII headers, each of which holds 12 monthly fields. Each file occupies almost 0.5 MB. The grid on which each field of values is presented is a 2.5°x2.5° latitude--longitude (Cylindrical Equal Distance) global array of points. It is size 144x72"

3.3.3 Derived Data

GridSat-B1 data NetCDF-4 files are converted from 0.07 degrees lat/lon to a 0.25 degree resolution in a flat binary format for input to the PERSIANN algorithm system. This global coverage array is a 2-byte short integer binary format grid 480 rows x 1440 columns covering 60N to 60S and 0 to 360 long.

The GridSat-B1 IRWIN brightness temperature data are gridded as 0.07 degree resolution lat/lon grids. However, the input files to the PERSIANN model should have a 0.25 degree spatial resolution. A preprocessing code is used to extract the GridSat-B1 IRWIN data, to alter the resolution of the IR data and to subset the global coverage into four major sub-regions. These sub-regions overlap about 10^{o} degrees in longitude. Further subsetting of these 4 major subs can be done for further parallelization into 24 overlapping sub-regions each.

3.3.4 Forward Models

Not applicable

3.4 Theoretical Description

The data record for the PERSIANN Precipitation CDR spans from 1/1/1982 to 12/31/2011. This precipitation CDR addresses the spatial and temporal precipitation variability of precipitation over the past several decades. Our approach builds upon existing monthly product of the Global Precipitation Climatology Program (GPCP) product at coarse resolution (2.5^{o}x2.5^{o}), and applies the PERSIANN algorithm to the historical archive of GridSat-B1 infrared window (IRWIN) observations from GEO satellites to generate daily precipitation (1982-2011) at 0.25° daily scales for the area between (60^{o}S to 60^{o}N).

3.4.1 Physical and Mathematical Description

PERSIANN Precipitation

The PERSIANN system provides global precipitation estimates using combined IR and passive microwave (PMW) information from multiple GEO and LEO (Low Earth Orbiting) satellites (Hsu et al., 1997; 1999; Sorooshian et al., 2000). The algorithm uses an Artificial Neural Network (ANN) to extract cold cloud pixels and neighboring features from GEO IR images, and associates variations in each pixel's brightness temperature to estimate the pixel's surface rain rate (Hsu et al., 1997; 1999; Sorooshian et al., 2000). IR-based precipitation estimates from the neural network are further adjusted by the PMW precipitation estimates produced using the data from LEO satellites.

Figure 1 shows a simplified flowchart of the current operational PERSIANN system. Because very limited PMW sampling data are available prior to 2000, the effectiveness of error correction will be reduced in the reconstructed data. The data generation framework incorporates with GPCP monthly rainfall data at 2.5°x2.5° lat-lon scale to adjust 3-hour PERSIANN rainfall to ensure data consistency and quality.

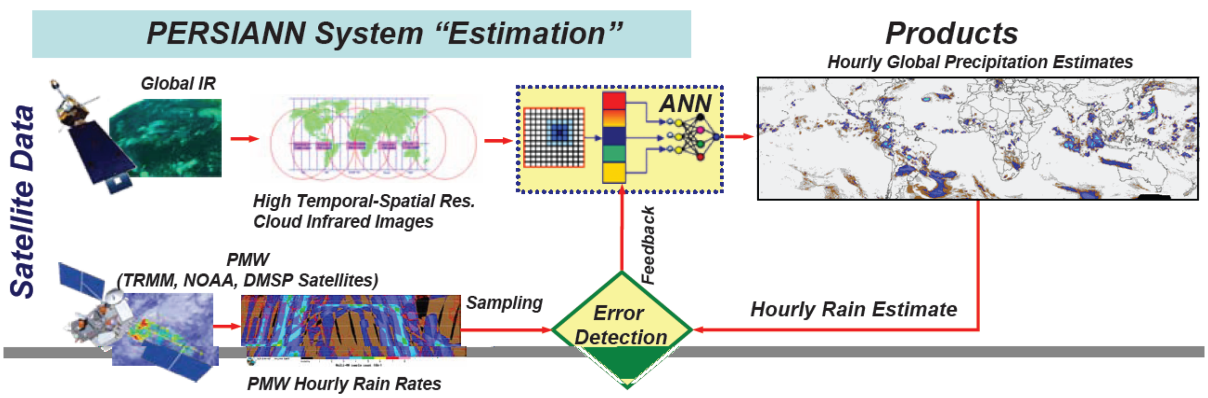


Figure 1: A flowchart of the current operational PERSIANN system

Adjusting Daily PERSIANN Data Using Monthly GPCP Data

The GPCP was established by the World Climate Research Programme (WCRP) to document the spatial and temporal distribution of precipitation at climate scale (WCRP, 1986, and Adler et al. 2003). The GPCP-v2.2 Monthly 2.5° merged analyses were constructed using multi-satellite (SSMI [Special Sensor Microwave Imager] and IR) precipitation estimates and then merging satellite and GPCC gauge analyses into a final product (Huffman et. al, 1997).

To reduce bias of PERSIANN rainfall at 3-hour scale, while at the same time preserving spatial and temporal patterns in the high resolution estimates, GPCP monthly rainfall at 2.5-degree resolution are used to adjust the high resolution PERSIANN estimates using a correction factor below:

$$w = R_{GPCP} / R_{PERSIANN} \qquad (1)$$

where R_{GPCP} is the 2.5° monthly GPCP precipitation and $R_{PERSIANN}$ is the accumulated monthly PERSIANN precipitation aggregated to 2.5°. As explained in section "1.2. Definitions", a *thd* = 0.1 and maximum $w = 20$ is used for PERSIANN accumulation and weight calculation, respectively. Subsequently, the monthly bias is spatially downscaled and removed from PERSIANN estimate at 0.25° resolution estimates using the above correction factor. The GPCP adjusted monthly 0.25° PERSIANN precipitation $aR_{PERSIANN}$ for each (i,j) pixel inside the 2.5° grid cell is calculated as follows:

$$aR_{PERSIANN}(i, j) = w \times R_{PERSIANN}(i, j) \qquad (2)$$

where $R_{PERSIANN}(i,j)$ is the PERSIANN monthly precipitation estimate at pixel (i,j). Equation 2 can be applied to higher temporal resolutions (e.g. daily), which will result in distributing the monthly bias removal for each day proportional to the day's contribution to the total monthly precipitation.

GPCP monthly rainfall includes gauge measurements over land (GPCC, Rudolf et. al. 1993, and Rudolf et. al. 1994); the bias correction of PERSIANN will maintain total monthly precipitation estimates that are consistent with GPCP monthly rainfall, while retaining the spatial and temporal details made available through PERSIANN estimations (0.25-degree and daily).

Merging IR data from multiple satellites is important for accurate and constant precipitation estimation. Figure 2, illustrates the results of the experiment, which consisted of applying PERSIANN estimates, obtained from GridSat-B1 IRWIN (without correction) and those obtained using NOAA Climate Prediction Center (CPC) global IR data (with correction for zenith angle viewing effects and inter-satellite calibration differences), to estimate precipitation. The two different PERSIANN estimates were then each adjusted using the bias-removal algorithm and downscaled to the corresponding GPCP monthly data. The adjusted PERSIANN daily precipitations calculated from the GridSat-B1 IRWIN and NOAA CPC gridded IR datasets are similar, and the scatter-plot shows that the adjusted daily precipitation estimations are well distributed along the 1:1 line with reasonable spread.

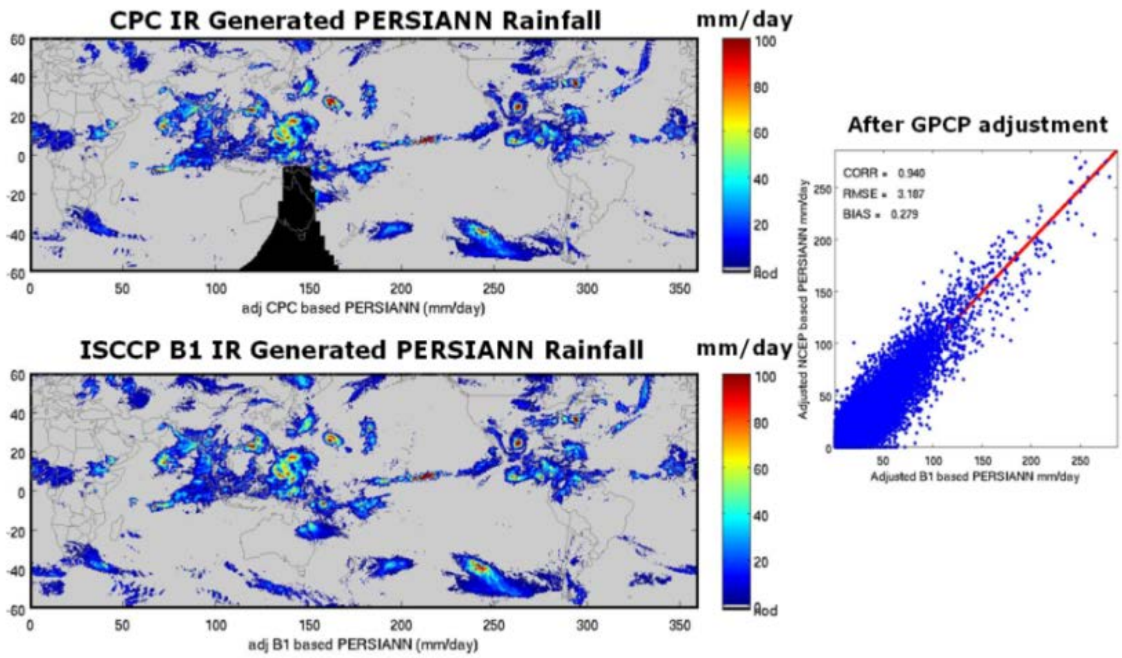


Figure 2: Comparison between adjusted PERSIANN data generated using zenith corrected CPC-IR (top left panel) data set and using non-corrected ISCCP B1 IR (lower left panel) data set for July 20, 2005. Comparison statistics are shown in the right panel scatter-plot.

3.4.2 Data Merging Strategy

The data record for the GridSat-B1 IRWIN spans from 1980 to 2012. The other data from GPCP is available from 1979 to delayed present time and more recent data will be provided for climate studies.

3.4.3 Numerical Strategy

3.4.4 Calculations

Most of the algorithm calculations are discussed in other sections. The basic steps are outlined in 3.2 and theory in 3.4.1. For each subarea IR data are read in and for each pixel a sliding window of surrounding pixels is used to generate 4 other values for each pixel, including the mean and standard deviation for both a 3x3 pixel neighborhood and a 5x5 neighborhood. These values are used to determine what IR to rain rate (RR) conversion will be used to estimate the precipitation for each pixel. The output subareas are merged together if necessary. The output 0.25 degree resolution grids need to be resampled to 2.5

degrees and accumulated to monthly for adjustment with GPCP data. The adjustment outputs are then disaggregated to back to 0.25 degrees and 3-hourly interval temporal resolution data grids as described in 3.4.1. Lastly, the 3-hourly files will be accumulated to daily mm/day precipitation estimates.

3.4.5 Look-Up Table Description

Not applicable

3.4.6 Parameterization

Not applicable

3.4.7 Algorithm Output

The data files produced from the complete PERSIANN-CDR algorithm system consists of a grid 480 rows x 1440 columns of precipitation estimates in units of mm/day. The data will be near global coverage from 60N to 60S and 0 to 360 longitude. Each daily precipitation grid is incorporated into NetCDF-4 format files with appropriate metadata, including georeference coordinates, to facilitate the use of sophisticated viewers and database searches. Each file will be labeled with the date that the data represents for example: PERSIANN-CDR_v01r01_2010-07-01.nc represents the PERSIANN-CDR precipitation estimates in mm/day accumulated for July 1, 2010 from 00:00 to 23:59 UTC.

4. Test Datasets and Outputs

4.1 Test Input Datasets

In order to analyze the performance of the algorithm, different methods of validation were applied. At the very beginning, the monthly accumulation precipitation for August 2005 was considered. The comparison was done to investigate whether or not the algorithm was capable of making PERSIANN estimates consistent with GPCP monthly data in total monthly precipitation using the GridSat-B1 data as input. As it is shown in Figure 3, the improvements after applying the algorithm are even visually evident.

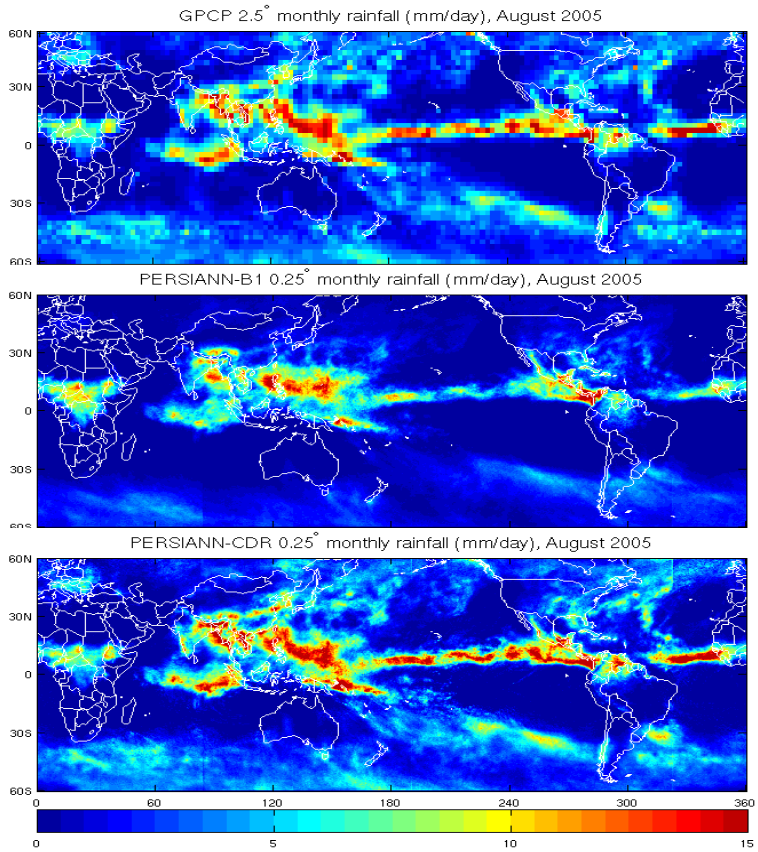


Figure 3: Global rain rate for August 2005 for GPCP Monthly 2.5° (top), GridSat-B1 based PERSIANN (PERSIANN-B1) Monthly 0.25° (middle), and Bias-Adjusted PERSIANN (PERSIANN-CDR) Monthly 0.25° (bottom)

4.2 Test Output Analysis

4.2.1 Reproducibility

As stated in previous section, different characteristics and statistics were tested and checked to evaluate the performance of the produced precipitation dataset. After getting the promising preliminary results for test dates (e.g. August 2005), the model and correction algorithm were forced with almost the whole record of GridSat-B1 IRWIN data for 1983 to 2012, excluding the problematic months. To investigate whether or not GPCP precipitation product is properly assimilated into the PERSIANN-CDR product, the Mean Areal Precipitation (MAP) for Northern and Southern Tropical regions were calculated. Following graphs (Figure 4) show the strength of the algorithm in matching accumulated monthly PERSIANN rain rate estimates to GPCP monthly product and making the PERSIANN-CDR product.

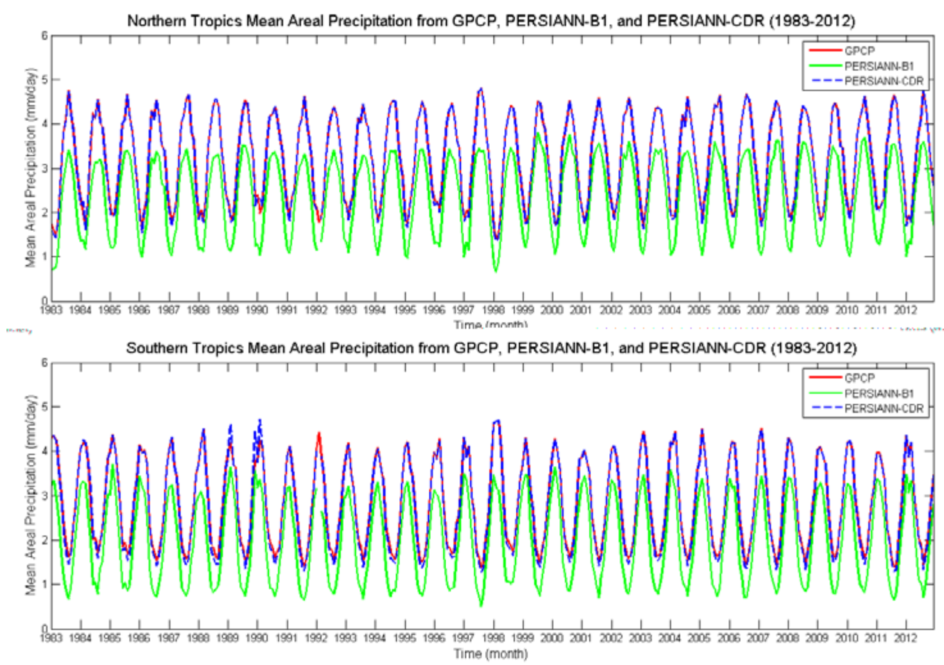


Figure 4: Comparing Mean Areal Precipitation (MAP) for Northern (0-30N, top figure) and Southern (0-30S, bottom figure) Tropical Regions. MAP calculated from monthly GPCP (red), PERSIANN-B1 (green), and PERSIANN-CDR (dashed blue) data sets.

As it's clear in the above Figure 4, the mean areal precipitation from PERSIANN-CDR perfectly matches the same variable extracted from GPCP monthly product.

For more detailed investigations, comparison at the daily scale was performed. For this purpose, GPCP daily product known as GPCP-1DD was included into the analysis. The Mean Areal Precipitation (MAP) for different regions of the globe for daily PERSIANN-B1, PERSIANN-CDR, and GPCP-1DD were calculated. The result over the whole globe for the period of 2007-2009 is shown in Figure 5. Improvements become evident after applying the bias-adjustment algorithm to PERSIANN-B1 estimates (green line, Figure 5). It shows that PERSIANN-CDR performs well in estimating the same global precipitation as GPCP-1DD which benefits from the incorporation of PMW information (such as SSM/I, and Special Sensor Microwave Imager/Sounder (SSMIS)) in its estimate. It is noteworthy that no PMW data is used in PERSIANN-CDR, except indirectly from GPCP monthly data.

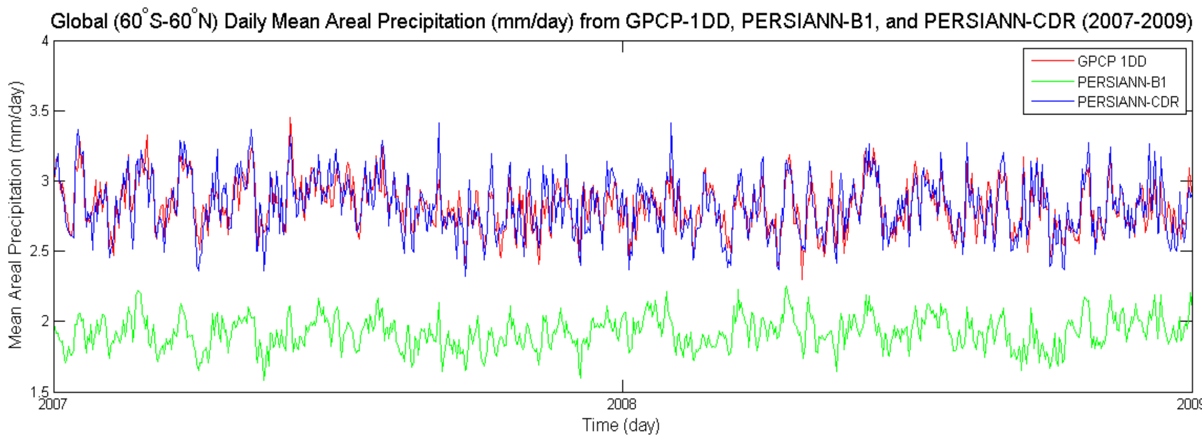


Figure 5: Daily comparison of mean areal precipitation (MAP) for 60S-60N region for the period of 2007-2009. MAP calculated from GPCP-1DD (red), daily PERSIANN-B1 (green), and daily PERSIANN-CDR (blue)

Similar graphs for the Northern Hemisphere (0°-60°N), Tropics (0°-30°N), and Extratropics (30°N-60N°) for the period of 1997-2012 are displayed in Figure 6. As shown, even without any GPCP-adjustment, the PERSIANN-B1 rain rate estimates show good agreement with GPCP 1DD in tropical regions. The performance of PERSIANN-B1 over extratropical regions is poorer. In these regions, PERSIANN-B1 underestimates the precipitation. However, after applying the GPCP bias-adjustment, PERSIANN-CDR captures the precipitation patterns and demonstrates considerable consistency with both GPCP daily and monthly precipitation products. The results are similar for the Southern Hemisphere (0°-60°S), Tropics (0°-30°S), and Extratropics (30°S-60S°).

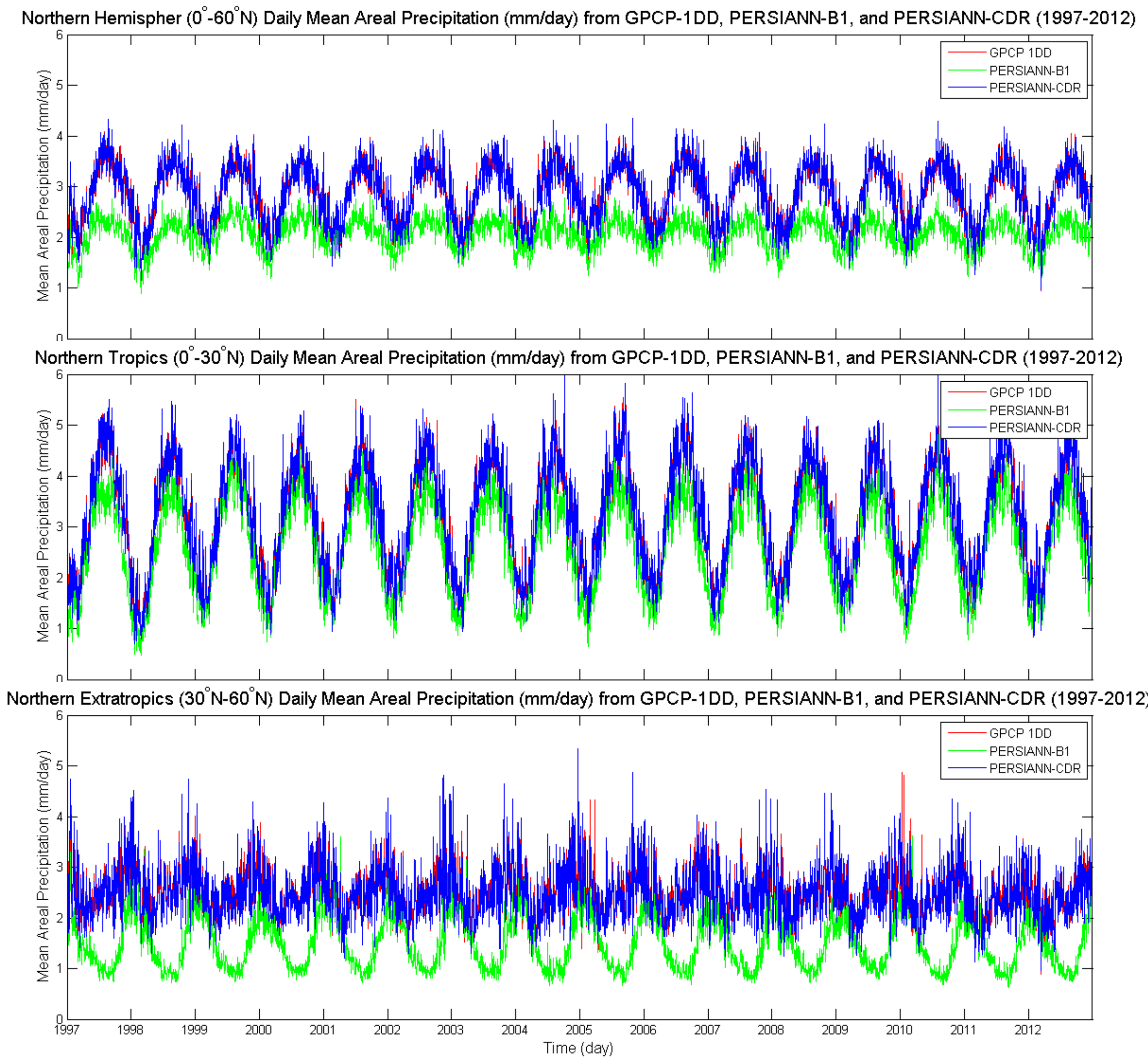


Figure 6: Daily mean areal precipitation (mm/day) for the Northern Hemisphere (0°-60°N, top), Tropics (0°-30°N, middle), and Extratropics (30°N-60°N, bottom) for the period of 1997-2012 from GPCP-1DD (red), PERSIANN-B1 (green), and PERSIANN-CDR (blue).

For further investigation, PERSIANN-CDR estimates were compared to Stage IV radar data at the 0.25° spatial scale during Hurricane Katrina (2005). Katrina happened in August 2005 and is considered as the most costly natural disaster ever to strike United States. Figure 8 below shows the one-day comparison (August 29, 2005) of four different products; 1) PERSIANN-B1, 2) PERSIANN-CDR, 3) Stage IV Radar, and 4) TMPA V7 3B42 daily precipitation in the southeastern part of the U.S.

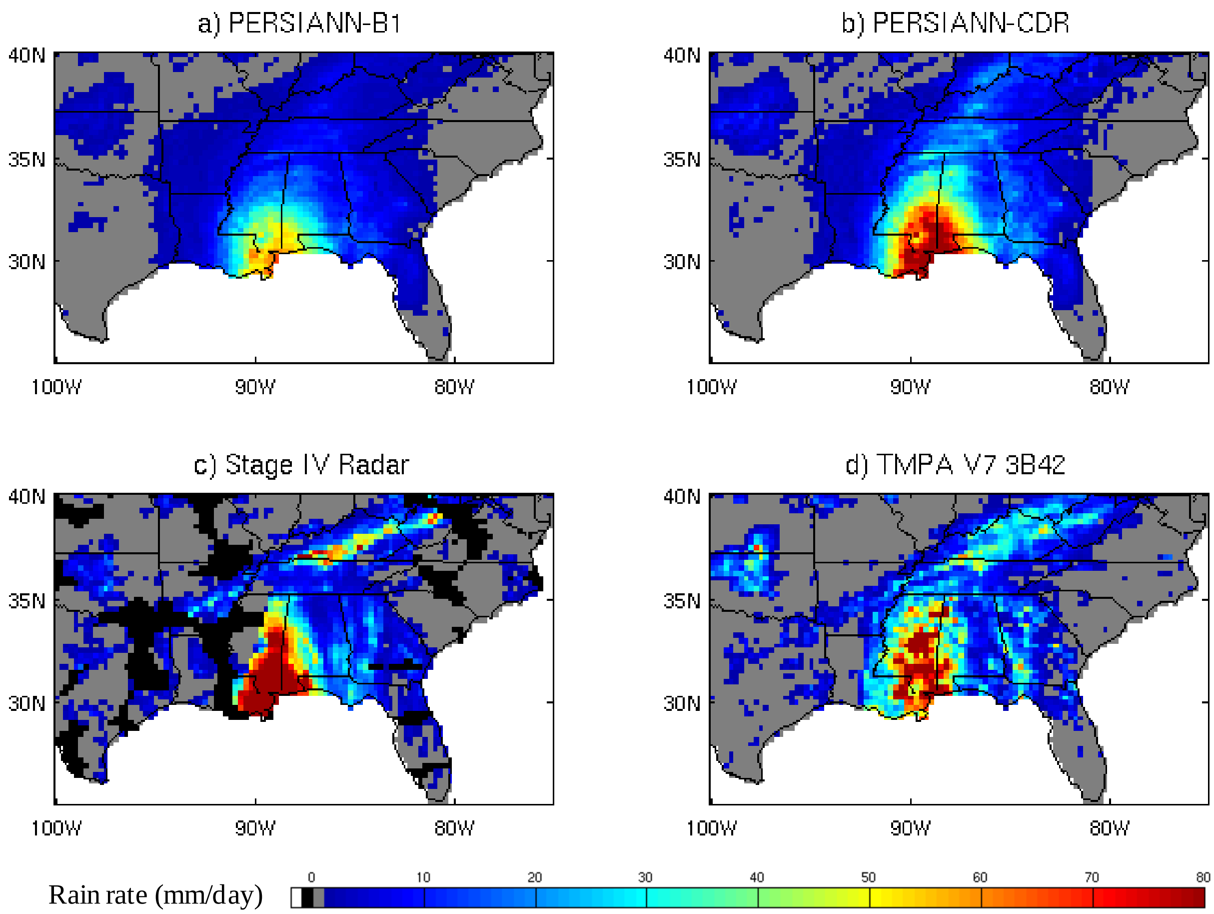


Figure 7: Rainfall (mm/day) over land during Hurricane Katrina on 29 August 2005 from: (a) PERSIANN-B1, (b) PERSIANN-CDR, (c) Stage IV Radar, and (d) TMPA v7. Black and gray pixels show radar blockages and zero precipitation, respectively.

As shown in Figure 7, PERSIANN-CDR (Figure 7b) shows similar precipitation patterns to the radar data (Figure 7c). Moreover, in regions where radars are blocked by mountains or a radar site is down (e.g., the Lake Charles radar site in Southwest Louisiana during Katrina), the spatial coverage provided by PERSIANN-CDR is very valuable and captures a wide view of the precipitation and hurricane landfall.

For further investigation of the performance of PERSIANN-CDR with more recent satellite-based precipitation products, TMPA v7 was included in the analysis (Figure 7d). As can be seen, PERSIANN-CDR and TMPA precipitation data both have some similarities and differences when compared to Stage IV radar data. PERSIANN-CDR seems to better show the distribution of the hurricane landfall, while underestimating the rainfall over KY when compared to radar and TMPA.

4.2.2 Precision and Accuracy

PERSIANN-CDR_v01r01 has been generated and made available to NCDC for archive. Evaluation of PERSIANN-CDR using CPC Unified gauge observations and Stage IV radar data over CONUS is ongoing. The quality of PERSIANN-CDR will be covered in the 2014 project period and will be reported by June 2014.

4.2.3 Error Budget

Not applicable

5. Practical Considerations

5.1 Numerical Computation Considerations

The PERSIANN algorithm is currently run using parallel jobs running on 4 major subareas which then are split up into 24 subareas, but this is not necessary and could be removed to make a cleaner set of codes.

5.2 Programming and Procedural Considerations

Unknown.

5.3 Quality Assessment and Diagnostics

During the GPCP bias adjustment phase, some code could be included to quantify the level of bias correction applied and flag any data well out of the ordinary range for further examination. Also, newer PERSIANN-CDR data files can be compared to data from other sources to check anomalies.

5.4 Exception Handling

Most of the expected exceptions will be due to missing files. At each stage that a file is input, code is included to check if the file is missing and handle it by using a NODATA file or flagging that time step for possible reprocessing. Corrupt data files from the GridSat-B1 CDR data set have been identified and that data will need to be reprocessed. Noise can be found in some satellite data and the code checks abnormally connected (unnatural) IR data pixels outside of expected normal ranges and extreme differences between neighboring pixels. These connected "bad" pixels are usually in a linear row and can be removed and missing data interpolated if necessary to meet the resolution for input into the model.

Due to the length of time required to run the PERSIANN model for the 30+ year record, we will, at this time, remove rather than reprocess any time steps that were generated from problem files prior to bias correction. This version and revision of PERSIANN-CDR (v01r01) will eventually be reprocessed with improved data filtering to remove erroneous parts of each B1 grid prior to running PERSIANN.

5.5 Algorithm Validation

PERSIANN-CDR is under evaluation using CPC Unified gauge observation and Stage IV radar data over CONUS. Statistics will cover the probability of detecting extremes, the volume of correctly identified precipitation, and false alarm ratio of various extreme thresholds (see AghaKouchak et. al., 2011 for evaluation statistics of extreme precipitation).

5.6 Processing Environment and Resources

The PERSIANN-CDR data processing system is configured to run on LINUX servers on generic (x86 for byte order) hardware. The operating system is currently CENTOS 5.5, but the codes have also been run on much older Fedora Core 3 and other variations of Red Hat LINUX. Programs are written in, C, C++, bash shell, and MATLAB. Compilers are Gnu gcc/g++ with HDF5 libraries.

The data go through many stages of processing: 125 GB/year raw data compressed are processed to provide inputs to the model scripts --> those inputs reduced the volume to 5 GB/year compressed --> PERSIANN output from model generate 5.5 GB/year compressed --> the final daily 0.25° PERSIANN-CDR netcdf-4 output 500MB/year compressed.

Due to the high volume of data to process we ran the PERSIANN model on 3 different systems. A year of data on one system running 4-cores in parallel can take 5 days to finish. Another system with faster Xeon 8-cores and better compiler optimized code can now finish a year of data in 10-12 hrs. Other stages, raw input processing, bias corrections, accumulation to daily and netcdf-4 output, take minimal time in comparison, with bias correction taking at most 30 minutes/year being the longest

NOTE: as configured the PERSIANN codes can be run in 4 major subareas of 24 subs each in parallel, but the code will not handle running 2 of the same "small subs" at the same time. So, each time step is processed one by one with the subs running in parallel or duplicate directory structures can be set up to process in parallel. Disk I/O is also a bottleneck so parallel runs on faster and separate disks improve speed.

6. Assumptions and Limitations

6.1 Algorithm Performance

See section 5.6 above for a discussion on the actual performance of the algorithm on different systems and how improvements were made. In future a rewrite of the general model code to specifically handle this PERSIANN-CDR process would reduce the unnecessary code to handle training and other options. This task is not expected to be essential at this time.

6.2 Sensor Performance

Although PERSIANN-CDR does not use the raw sensor data directly the problems with noise, pixel navigation (geo-location) errors, inter-calibration, and time shifting between the different satellite scans and satellites in the GridSat-B1 data give the PERSIANN algorithm false cloud structure and texture signatures as input. Also, any cold "data" pixels that come from non-data sources will cause an estimate of rainfall regardless. Many of these data problems are apparent in a view of the IR data, but cannot be easily distinguished from normal data programmatically in the model code. Satellite data can be noisy and the older data files in the ISSCP GEO historical archive are probably not as clean or as easily integrated and merged as the newer years. A planned new version of the GridSat-B1 CDR product addressing some of the errors in the current version will help with some of these issues. A new PERSIANN-CDR version will be produced using the improved GridSat-B1 data.

7. Future Enhancements

Future enhancements are not being considered at this time.

8. References

Adler, R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, 2003. The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979 present). Journal of Hydrometeorology 4, 11471167.

AghaKouchak A., Behrangi A., Sorooshian S., Hsu K., Amitai E., 2011, Evaluation of satellite-retrieved extreme precipitation rates across the Central United States, *Journal of Geophysical Research*, 116, D02115, doi:10.1029/2010JD014741.

Ashouri H., K. Hsu, S. Sorooshian, D. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, O. P. Prat, 2014: "PERSIANN-CDR: Daily Precipitation Climate Data Record from Multi-Satellite Observations for Hydrological and Climate Studies", Bulletin of American Meteorological Society (BAMS), Submitted.

Ashouri H., K. Hsu, S. Sorooshian, D. Braithwaite, 2012, "A Satellite-based High Resolution Precipitation Product for Studying Climate Extremes", American Geophysical Union Fall Meeting, San Francisco, California, December 3-7, 2012.

Hsu, K., X. Gao, S. Sorooshian, and H.V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks, *J. of Applied Meteorology*, **36**(9), 1176-1190.

Hsu, K., H.V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of physical variables from multiple channel remotely sensed imagery using a neural network: application to rainfall estimation, *Water Resources Research*, **35**(5), 1605-1618.

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. *Amer. Meteor. Soc.*, **78**, 5–20.

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Stocker, E., Wolff, D., 2007. The TRMM multi-satellite precipitation analysis: quasi-global, multiyear, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol*ogy. **8**, 3855.

Rudolf, B., 1993. Management and analysis of precipitation on a routine basis. Proc. Internat. WMO/IAHS/ETH SYMP. on Precip. and Evap., Slovak Hydromet. Inst., Bratislava, Sept. 1993, **1**, 69-76

Rudolf, B., H. Hauschild, W. Rueth and U. Schneider, 1994. Terrestrial Precipitation Analysis: Operational Method and Required Density of Point Measurements. In: Global Precipitations and Climate Change (Ed. M. Desbois, F. Desalmond), NATO ASI Series I, Vol. 26, Springer-Verlag, p. 173-186.

Sorooshian, S., K. Hsu, X. Gao, H.V. Gupta, B. Imam, and Dan Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bulletin of the American Meteorological Society*, **81**(9), 2035-2046.

Appendix A. Acronyms and Abbreviations

Acronym or Abbreviation	Meaning
ANN	Artificial Neural Network
AVHRR	Advanced Very High Resolution Radiometer
C-ATBD	Climate Algorithm Theoretical Basis Document
CDR	Climate Data Record
CHRS	Center for Hydrometeorology and Remote Sensing
CPC	Climate Prediction Center (National Weather Service)
FY2	Feng Yun 2 (Chinese satellite)
GOES	Geostationary Operational Environmental Satellite
GEO	Geostationary Satellites
GMS	Geostationary Meteorological Satellite
GPCP	Global Precipitation Climatology Project
HDF5	Hierarchical Data Format (version 5)
IR	Infrared data
IRWIN	Infrared Window
IRWVP	Infrared water vapor
ISCCP	International Satellite Cloud Climatology Project
LEO	Low Earth Orbiting Satellites
MAP	Mean Areal Precipitation
Meteosat	Family of satellites from the European Space Agency
MM	2-digit month
MW	Microwave data from LEO satellites (also PMW)
NCDC	National Climatic Data Center
NetCDF-4	Network Common Data Form (version 4)
NOAA	National Oceanic and Atmospheric Administration
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
RR	Rain Rate
SSMI	Special Sensor Microwave Imager
thd	Threshold on 3-hourly 0.25 degree PERSIANN-B1 estimates
UCI	University of California, Irvine
UTC	Coordinated Universal Time

w	Ratio of GPCP and PERSIANN-B1 at monthly 2.5 degree
WCRP	World Climate Research Programme
YY	2-digit year