Title
Toward Improved Understanding of Global Precipitation Variations Using Satellite-based Observations

Permalink
https://escholarship.org/uc/item/9fs8h7sm

Author
Ashouri Talouki, Hamed

Publication Date
2014

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
IRVINE

Toward Improved Understanding of Global Precipitation Variations Using Satellite-based Observations

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Hamed Ashouri

Dissertation Committee:
Professor Soroosh Sorooshian, Chair
Professor Kuo-Lin Hsu
Professor Brett Sanders
Professor Jin-Yi Yu

2014
DEDICATION

To

my dear parents, Fatemeh and Nadali

my wonderful wife, Mina

and my lovely siblings Maede, Zahra, and Amirhossein

in recognition of their worth.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>xii</td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td>xiv</td>
</tr>
<tr>
<td>ABSTRACT OF THE DISSERTATION</td>
<td>xx</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Climate Extremes and Data Issue</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Research Motivations</td>
<td>4</td>
</tr>
<tr>
<td>1.3. Scientific Objectives</td>
<td>6</td>
</tr>
<tr>
<td>1.4. Dissertation Outline</td>
<td>7</td>
</tr>
<tr>
<td>Chapter 2: Reconstruction of Global Daily Precipitation from Multi-Satellite Observations (PERSIANN-CDR)</td>
<td>9</td>
</tr>
<tr>
<td>2.1. Review of Current Satellite-based Precipitation Products</td>
<td>9</td>
</tr>
<tr>
<td>2.2. Problem Statement</td>
<td>11</td>
</tr>
<tr>
<td>2.3. Data</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1. Stage IV Precipitation Data</td>
<td>13</td>
</tr>
<tr>
<td>2.3.2. Gridded Satellite Infrared Data (GridSat-B1)</td>
<td>14</td>
</tr>
<tr>
<td>2.3.3. Global Precipitation Climatology Project (GPCP)</td>
<td>16</td>
</tr>
<tr>
<td>2.4. Methodology</td>
<td>17</td>
</tr>
<tr>
<td>2.4.1. PERSIANN-CDR Precipitation Estimation Algorithm</td>
<td>18</td>
</tr>
<tr>
<td>2.4.2. Adjusting Daily PERSIANN Data Using Monthly GPCP Data</td>
<td>19</td>
</tr>
<tr>
<td>2.5. Demonstration of PERSIANN-CDR Product</td>
<td>22</td>
</tr>
<tr>
<td>2.6. Operational PERSIANN-CDR at NOAA NCDC</td>
<td>29</td>
</tr>
</tbody>
</table>
2.7. Chapter Summary and Conclusions ................................................................. 32

Chapter 3: Verification of PERSIANN-CDR .............................................................. 33

3.1. Introduction ........................................................................................................ 33

3.2. Case Studies ......................................................................................................... 34

3.2.1. Hurricane Katrina ........................................................................................ 34

3.2.2. The 1986 Sydney Flood .............................................................................. 36

3.2.3. Probability Density Function ...................................................................... 38

3.2.4. U.S. Number of Rainy Days ....................................................................... 40

3.2.5. Extreme Precipitation over China ............................................................... 42

3.2.6. Rainfall-Runoff Modeling ........................................................................... 46

3.3. Chapter Summary and Conclusions ................................................................... 57

Chapter 4: Statistical Modeling of Extreme Precipitation Trends .............................. 61

4.1. Introduction ........................................................................................................ 61

4.2. Problem Statement ............................................................................................. 63

4.3. Data ..................................................................................................................... 64

4.3.1. The Climate Prediction Center (CPC) US Unified Precipitation Product ....... 64

4.3.2. The Modern-Era Retrospective-analysis for Research and Applications (MERRA) Precipitation Product ................................................................. 65

4.3.3. The North American Regional Reanalysis (NARR) Precipitation Product ... 66

4.4. Methodology ....................................................................................................... 66

4.4.1. Extreme Value Theory (EVT) ................................................................. 66

4.4.2. Generalized Extreme Value (GEV) Distribution ........................................ 68

4.4.3. Generalized Pareto (GP) Distribution ....................................................... 70

4.5. Results ............................................................................................................... 71

4.5.1. Annual Maximum Daily Precipitation (AMDP) ........................................... 71

4.5.2. Annual Extreme Daily Precipitation (AEDP) ............................................. 79

4.6. Discussion .......................................................................................................... 81

4.6.1. Empirical Probability Density ................................................................. 81

4.6.2. Negative Trend in MERRA over NE-KS .................................................... 84

4.6.3. Seasonality in Extremes ............................................................................. 87
4.7. Chapter Summary and Conclusions ................................................................. 96

Chapter 5: Conclusions and Future Directions ....................................................... 100
  5.1. Summary of Findings .................................................................................. 100
  5.2. Future Extensions ..................................................................................... 108

REFERENCES ........................................................................................................ 111
LIST OF FIGURES

Figure 2.1. Sample image of GridSat-B1 data, IR window (top), water vapor (middle), and visible (bottom) (NOAA/NCDC) .......................................................... 15

Figure 2.2. A schematic of the PERSIANN-CDR algorithm for reconstruction of historical precipitation. .......................................................... 19

Figure 2.3. Sample images (year 1980) of GridSat IR data (top brightness temperature of the cloud in degree Kelvin) which were taken out from the PERSIANN-CDR precipitation estimation process due to quality issues. .................................. 23

Figure 2.4. Global rainfall maps (mm/day) for August 2005 from GPCP 2.5° (top, Adler et al. 2003), PERSIANN-B1 0.25° (middle), and PERSIANN-CDR 0.25° (bottom) monthly datasets.......................................................... 25

Figure 2.5. Mean areal precipitation (mm/day) for Northern (0°-30°N, top) and Southern (0°-30°S, bottom) Tropics from monthly GPCP (red, Adler et al. 2003), PERSIANN-B1 (green), and PERSIANN-CDR (dashed blue) datasets.......................................................... 26

Figure 2.6. Daily global (60°S-60°N) mean areal precipitation (mm/day) for the period of 2007-2009 for GPCP-1DD (red, Huffman et al. 2001), PERSIANN-B1 (green), and PERSIANN-CDR (blue). .......................................................... 27

Figure 2.7. 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, Huffman et al. 2001), PERSIANN-B1 (green), and PERSIANN-CDR (blue). .......................................................... 28

Figure 2.8. Top 5% heavy rainfall (mm/day) maps from GPCP-1DD (top, Huffman et al. 2001), PERSIANN-CDR 1° (middle), and PERSIANN-CDR 0.25° (bottom) for the period of 1997-2012.......................................................... 30
Figure 2.9. Overview of PERSIANN-CDR specifications, input data, and potential applications (NOAA/NCDC/CDRP)........................................................................................................... 31

Figure 3.1. Rainfall (mm/day) over land during Hurricane Katrina on 29 August 2005 from: (a) PERSIANN-B1, (b) PERSIANN-CDR, (c) Stage IV Radar (Lin and Mitchell 2005), and (d) TMPA v7 (Huffman et al. 2007). Black and gray pixels show radar blockages and zero precipitation, respectively. ................................................................. 35

Figure 3.2. Scatterplots of PERSIANN-B1 (top), PERSIANN-CDR (middle), and TMPA v7 (bottom, Huffman et al. 2007) against Stage IV radar data (Lin and Mitchell 2005) for 29 August 2005 during Hurricane Katrina................................................................. 36

Figure 3.3. Rainfall (mm/day) over land during Sydney Australia flood on 5 August 1986 from gauge observation (top left, Jones et al. 2009) and PERSIANN-CDR rain rate estimates (bottom left). The scatter plot and respective statistics are shown in the right................................................................. 37

Figure 3.4. Comparing the Empirical Probability Density Function (PDF) of PERSIANN-B1, PERSIANN-CDR, TMPA v7 (Huffman et al. 2007), and CPC Unified Gauge-Based Analysis of Daily Precipitation (Higgins et al. 2000a) over the CONUS during the period of 1998-2008 .................................................................................. 39

Figure 3.5. Annual average count of days where rainfall ≥ 10 mm (left column) and rainfall ≥ 20 mm (right column) for CPC (top), and PERSIANN-CDR (bottom) for 1983-2011 .................................................. 41

Figure 3.6. Scatterplots of the annual average count of days where rainfall ≥ 10 mm (left column) and rainfall ≥ 20 mm (right) for PERSIANN-CDR against CPC. Correlation coefficient, RMSE, and Bias are shown on the plots.......................................................... 42

Figure 3.7. The distribution of rain-gauge stations in the EA data set ................................................ 43

Figure 3.8. The 99th and 95th percentile indices of extreme daily precipitation from the EA data set (1st column), and PERSIANN-CDR (2nd column). The spatial correlation distribution and the scatterplots of the indices from the EA and PERSIANN-CDR data sets are shown in the 3rd and 4th columns, respectively. The stippled areas in the 3rd column show the significant correlation coefficient at the 95% level.............................. 45

Figure 3.9. The three study basins (SAVOY, ELMSP, and SLOA4) (NOAA/NWS) ................. 46
Figure 3.10. Precipitation comparison plots for SAVOY (top), ELMSP (middle), and SLOA4 (bottom) basins for 2003-2010.

Figure 3.11. Simulated and observed streamflow hydrographs at the outlet of SAVOY basins.

Figure 3.12. Simulated and observed streamflow hydrographs at the outlet of ELMSP basins.

Figure 3.13. Simulated and observed streamflow hydrographs at the outlet of SLOA4 basins.

Figure 3.14. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for SAVOY basin.

Figure 3.15. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for ELMSP basin.

Figure 3.16. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for SLOA4 basin.

Figure 3.17. Scatterplots and Correlation Coefficient, RMSE, and BIAS statistics of PERSIANN-CDR simulated streamflow against USGS observations.

Figure 4.1. A schematic showing how a small shift in the mean of the distribution may lead to large changes in the tail of the distribution (Solomon et al. 2007).

Figure 4.2. Trend (mm/day/year) in the Annual Maximum Daily Precipitation (AMDP) in CPC (top), NARR (middle), and MERRA (bottom) for 1979-2010 period. Right column shows the regions where the trend in AMDP is statistically significant at a 5% level.

Figure 4.3. Location (left column), scale (middle column), and shape (left column) parameters of the time-invariant GEV distribution of AMDP events from CPC (top), NARR (middle), and MERRA (bottom) for 1979-2010.
Figure 4.4. Trend (mm/day/year) in the location parameter of the time-variant GEV distribution of AMDP events in CPC (top), NARR (middle) and MERRA (bottom). Right column shows the regions where the trend in location parameter is statistically significant at a 5% level. ............................................................................................... 76

Figure 4.5. Estimated scale parameter (left Column) of the time-variant GEV distribution for CPC (top), NARR (middle) and MERRA (bottom). Right column shows the maps of the relative difference between the scale parameters in time-variant and time-invariant conditions. ......................................................................................................... 78

Figure 4.6. Trend (number per year) in the number of annual extreme daily precipitation (AEDP) events in CPC (top), NARR (middle), and MERRA (bottom). The right column shows the regions where trend is statistically significant at 5% significance level.... 80

Figure 4.7. Estimated scale (left column) and shape (right column) parameters of GP distribution for AEDP for CPC (top), NARR (middle), and MERRA (bottom) for the period of 1979-2010. ........................................................................................................ 82

Figure 4.8. Empirical probability density over CONUS for AMDP (red) and de-clustered AEDP (black) for CPC (solid), NARR (dashed), and MERRA (dotted) for 1979-2010...... 83

Figure 4.9. Time series, trend lines and respective statistics (top) and anomalies (bottom) of the annual maximum daily precipitation for CPC (solid) and MERRA (dashed) over Kansas and Nebraska. ........................................................................................................ 85

Figure 4.10. Trend (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and HUR (bottom row) for CPC (left column), NARR (middle column) and MERRA (right column) during 1979-2010. ................................................. 89

Figure 4.11. Trend (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for CPC (left column), NARR (middle column) and MERRA (right column) during 1979-2010......................................................................................... 90

Figure 4.12. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and JJA (bottom row) for NARR (left column) and MERRA (right Column) against CPC gridded observation during 1979-2010. All the pixels at which trends are significant in either reanalyses or CPC are considered................................................................. 92
Figure 4.13. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for NARR (left column) and MERRA (right column) against CPC gridded observation during 1979-2010. All the pixels at which trends are significant in either reanalyses or CPC are considered.............................................................................................................. 93

Figure 4.14. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and JJA (bottom row) for NARR (left column) and MERRA (right column) against CPC gridded observation during 1979-2010. Only pixels at which trends are significant in both reanalyses and CPC are considered. .............................................................................................................................................. 94

Figure 4.15. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for NARR (left column) and MERRA (right column) against CPC gridded observation during 1979-2010. Only pixels at which trends are significant in both reanalyses and CPC are considered.............................................................................................................................................. 95
LIST OF TABLES

Table 2.1. Coverage and spatiotemporal resolutions of major satellite precipitation products, including PERSIANN-CDR. ......................................................... 10

Table 3.1. Bias, Correlation Coefficient, and RMSE statistics for simulated streamflow (2003-2010) from stage IV radar data, PERSIANN-CDR, and TMPA against USGS streamflow gauge observations................................................................. 52

Table 3.2. Bias, Correlation Coefficient, and RMSE statistics for simulated streamflow from PERSIANN-CDR against USGS observed streamflow for 1983-2012. ......................... 57
ACKNOWLEDGMENTS

Trying to write the “acknowledgement” part of my dissertation made me travel through time and recalling a journey full of wonderful memories with helpful loving people around me. I would like to take this opportunity and express my sincere appreciations to all these people and individuals whose help and encouragement made this research possible.

First and foremost, I would like to express my deepest appreciations to my academic adviser, Professor Soroosh Sorooshian, for his invaluable guidance, immense knowledge, and endless encouragement throughout my Ph.D. program. Working under the supervision of Professor Sorooshian at the Center for Hydrometeorology and Remote Sensing (CHRS) exposed me to the wonders and state-of-the-art challenges of hydrology, climatology, and remote sensing. Soroosh is a great teacher and a true friend. I am sincerely thankful for all his unconditional help and advices, accompanied by his humble and nice personality. I would also like to extend my sincere gratitude to my co-adviser, Professor Kuo-Lin Hsu, for his immense support during my Ph.D. research. Kuolin’s knowledge and insight in different scientific areas, specifically remote sensing, is exemplary. This dissertation would have not been possible without their exceptional expertise and endless support.

I also wish to express my appreciation to Dr. Michael Bosilovich and Dr. George Huffman at NASA Goddard Space Flight Center, Dr. Robert Adler at the University of Maryland, Dr. Michael Wehner at Lawrence Berkeley National Laboratory and the University of California Berkeley, Dr. Jaechoul Lee at Boise State University, and Dr. DeWayne Cecil at Global Science and Technology Inc. for their invaluable comments throughout my Ph.D. work.

I would also like to thank my Ph.D. committee members, Professors Brett Sanders and Professor Jin-Yi Yu, for their constructive comments and invaluable suggestions in my qualifying exam, as well as their insightful advice and unconditional help during my PhD work, all of which significantly enhanced the quality of this work.

Special thanks to my colleagues and friends at CHRS with whom I share wonderful memories. I appreciate Mr. Dan Braithwaite, our center’s computer specialist and programmer analyst for his data and IT related expertise and assistance, as well as his insightful comments on scientific questions. I would also like to thank Ms. Diane Hohnbaum for her amazing work and attention as our center’s administrative assistant. In addition, past and present CHRS students and affiliates are my dear friends and I would like to thank them all.

I could never find proper words to show how thankful, grateful, and indebted I am to my dear parents, Fatemeh and Nadali. I love you with all my heart and soul, and wish this dissertation can serve as a very tiny gift to you in recognition of your worth in my life. I would also like to express my love and appreciation to my lovely sisters, Maedeh and Zahra, and my dear brother, Amirhossein, as well as my dear extended family and in-laws, for their endless love, support and encouragement throughout my entire life.
Finally, I would like to express my ultimate love and thanks to my dear wife and best friend, Mina, for her unconditional love and support. Mina, I wish to thank you for bringing eternal love, joy and peace into my life. Without your support this dissertation would have not been completed.

Financial support for this research effort is made available through the NASA Earth and Space Science Fellowship (NESSF) award (#NNX12AO11H), the NOAA Climate Change Data and Detection (CCDD) (#NA10DAR4310122), the NOAA Cooperative Institute for Climate and Satellites (CICS) and the NOAA NCDC Climate Data Record program (Prime award number NA09NES440006 and NCSU CICS sub-Award #2009-1380-01), the NASA Energy and Water cycle Study (NEWS) program (NNX06AF93G), and the NASA Decision Support System (NNX09A067G). Without this support this work would not have been possible.
CURRICULUM VITAE

HAMED ASHOURI
Department of Civil and Environmental Engineering
University of California, Irvine
Irvine, CA 92697

RESEARCH INTERESTS

- Hydrology and Climatology
- Remote Sensing and Applications
- Hydrological Rainfall-Runoff Modeling
- Climate Extremes
- Climate Change and Variability
- Climate Data Record (CDR)
- Extreme Value Theory (EVT)
- Statistics and Statistical Modeling
- Data Assimilation
- Streamflow Prediction and Weather Forecasting
- Water Resources Planning and Management
- Machine Learning Techniques

EDUCATION

Ph.D., Civil and Environmental Engineering (Hydrology and Water Resources) University of California, Irvine, USA, 2014

M.Sc., Civil and Environmental Engineering (Water Resources Engineering) Sharif University of Technology, Iran, 2009

B.S., Civil Engineering Isfahan University of Technology, Iran, 2006

PEER-REVIEWED JOURNAL PAPERS


Nguyen P., S. Sellars, A. Thorstensen, Y. Tao, H. Ashouri, D. Braithwaite, K. Hsu, and S.


**CONFERENCE ORAL PRESENTATIONS**


Sorooshian S., K. Hsu, **H. Ashouri**, and D. K. Braithwaite, 2013: Operational Daily, 25km Precipitation from UCI: PERSIANN, Climate Data Record (CDR) Program Annual Meeting, National Oceanic and Atmospheric Administration (NOAA), National Climatic Data Center (NCDC), July 30 - August 1, 2013, Asheville, NC, USA.


University, Tabriz, Iran.


**CONFERENCE POSTER PRESENTATIONS**


Ashouri H., K. Hsu, S. Sorooshian, and D. Braithwaite, 2013: PERSIANN Precipitation Climate Data Record (PERSIANN-CDR), Chapman University Symposium on Big Data and Analytics, 44th Interface Symposium on Computing Science and Statistics, 4-6 April, 2013. Chapman University, Orange, CA, USA.


Forecasting Using Ensemble Streamflow Prediction (ESP) Technique and Large-Scale Climate Signals, International Conference on Water Scarcity, Global Changes and Groundwater Management Responses, December 2008, University of California, Irvine, CA, USA.

**GRANTS/PROPOSALS/TECHNICAL REPORTS**

**Hamed Ashouri**, Renewal proposal for the NASA Earth and Space Science Fellowship (NESSF) Program, Approved on June 2014.


**Hamed Ashouri**, Renewal proposal for the NASA Earth and Space Science Fellowship (NESSF) Program, Approved on March 2013.

Data Submission Agreement between the University of California at Irvine and the National Climatic Data Center for the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Network (PERSIANN) Climate Data Record (CDR), 2013, Submitted to the Climate Data Record (CDR) Program, National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC), PI: Soroosh Sorooshian, Collaborators: Kuolin Hsu, **Hamed Ashouri** (Data Quality Scientist and Expert), Technical Expert: Dan Braithwaite.

PERSIANN Precipitation Climate Data Record, Climate Algorithm Theoretical Basis Document (C-ATBD)”, Kuolin Hsu, **Hamed Ashouri**, Dan Braithwaite, and Soroosh Sorooshian from CHRS at UC Irvine, November 2012, Submitted to Climate Data Record (CDR) Program, National Oceanic and Atmospheric Administration (NOAA), National Climatic Data Center (NCDC), Asheville, NC, USA.


**HONORS & AWARDS**

◊ Recipient of the NASA Earth and Space Science Fellowship (NESSF) Award 2012 - 2014

*National Aeronautics and Space Administration (NASA), Silver Spring, MD.*

◊ Travel award for Community Earth System Model (CESM) tutorial 2014

*National Center for Atmospheric Research (NCAR), Boulder, CO.*
◊ National Fellowship Incentive Program Award
  The Henry Samueli School of Engineering, University of California, Irvine, CA.
  2014

◊ Outstanding achievement in doctoral education
  Association of Professors and Scholars of Iranian Heritage, Los Angeles, CA.
  2014

◊ Travel award for the 2013 AGU Fall Meeting, San Francisco, CA.
  Department of Civil and Environmental Eng., University of California, Irvine, CA.
  2014

◊ Travel grant for the 2013 AGU Fall Meeting, San Francisco, CA.
  Associated Graduate Student, University of California, Irvine, CA.
  2013

◊ Travel award for the 2014 AMS Meeting, Atlanta, GA.
  Department of Civil and Environmental Eng., University of California, Irvine, CA.
  2013

◊ Travel award for the WCRP 4th International Conference on Reanalyses, Silver Spring, MD.
  World Climate Research Program (WCRP)
  2012

◊ Graduate student research and travel grant award for AMS 2012 meeting
  The Henry Samueli School of Engineering, University of California, Irvine, CA.
  2012

◊ Young scientist travel grant for the Pan-GEWEX meeting, Seattle, WA.
  Global Energy and Water Cycle Experiment (GEWEX)
  2010

◊ Ranked 2nd in class of 2008
  Amongst 21 graduate students of Water Resources Engineering, Department of Civil and Environmental Engineering, Sharif University of Technology, Tehran, Iran.
  2009

◊ Exceptional Talented student
  Exceptional Talents Center, Sharif University of Technology, Tehran, Iran.
  2007 - 2008

**Teaching Experience**

◊ Co-lecturer, Surface Water Hydrology (ENGRCEE 176/276)
  Instructor: Prof. Sorooshian
  Dept. Civil and Environmental Engineering, University of California, Irvine.
  Fall 2011

◊ Teaching Assistant, Surface Water Hydrology (ENGRCEE 176/276)
  Instructor: Prof. Sorooshian
  Dept. Civil and Environmental Engineering, University of California, Irvine.
  Fall 2010

◊ Teaching Assistant, Engineering Hydrology
  Instructor: Prof. Abrishamchi
  Dept. Civil and Environmental Engineering, Sharif University of Technology.
  Fall 2007

◊ Instructor, English Language
  Union Department, Sharif University of Technology.
  2006-2007
Instructor, English Language
*Sokhansara Language Institute, Isfahan, Iran*

2005-2006

**PROFESSIONAL EXPERIENCE**

- **Reviewer, Peer-Reviewed Journals**
  - *Bulletin of American Meteorological Society (BAMS)*
  - *Journal of Hydrometeorology (JHM)*
  - *Earth Interactions*
  - 2013 - Present

- **Co-Founder, LearnDataAnalysis (LDA)**
  - *LDA (learndataanalysis.com) is a place for people interested in statistics, data analysis, and machine learning.*
  - Sept. 2011 - Present

- **Intern, Risk Management Solutions (RMS), Inc.**
  - *Modeling flood events in California basin*
  - *Analyzing, processing, and visualizing remote sensing data*
  - *Mathematical modeling of ground motion and earthquake data*
  - June 2014 – Sept. 2014

- **Intern, Pars Jooyab Consulting Engineering Co.**
  - *Designing the Urban Water Distribution System by WaterCAD and EPAnet*
  - June 2006 – Sept. 2006

**COMPUTER SKILLS**

- **Programming**
  - MATLAB, C, Fortran, R, Python, Shell scripting, Java (beginner)

- **Operating Systems**
  - LINUX, UNIX, Windows, Mac OS.

- **Software**
  - Arc GIS, ARC Info
  - HEC-RAS, HEC-HMS,
  - AutoCAD, WaterCAD, EPAnet
  - Loop, Swere

**PROFESSIONAL AFFILIATIONS**

- **American Geophysical Union (AGU)**
  - 2007 - Present

- **American Meteorological Society (AMS)**
  - 2007 - Present

- **American Society of Civil Engineers (ASCE)**
  - 2009 - Present

- **Iranian Committee On Large Dams (IRCOLD)**
  - 2008 - Present

**HOBBIES**

- Semi-professional Volleyball player and swimmer
- Photography, Reading
ABSTRACT OF THE DISSERTATION

Toward Improved Understanding of Global Precipitation Variations Using Satellite-based Observations

by

Hamed Ashouri
Doctor of Philosophy in Civil and Environmental Engineering
University of California, Irvine, 2014
Distinguished Professor Soroosh Sorooshian, Chair

Precipitation is one of the key elements of the Earth’s water cycle. Long-term precipitation observations in global scale and at high spatial and temporal resolutions are imperative for understanding how global warming has affected our climate system, particularly in terms of changes in the characteristics of precipitation. This dissertation contributes to the advancement of our understanding of the water cycle by 1) development of a new long-term high-resolution satellite-based precipitation product to be used for different hydrological and climate studies at a higher spatial and temporal resolution than previously possible, and 2) presenting a probabilistic framework to study the observed variations, trends and changes in global precipitation, particularly extreme precipitation events, in the course of time.

In the first part of the dissertation, using the Geostationary Earth Orbit (GEO) satellites Infrared (IR) channel observations of the brightness temperature of the cloud, and the Global Precipitation Climatology Project (GPCP) monthly product, a retrospective high-resolution (daily, 25km) satellite-based precipitation climate data record is developed and introduced. The
product, namely Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR), provides more than 30 years of global rainfall estimates from 01/01/1983 to delayed present covering the latitude band 60°S-60°N and longitude band 0°-360°. The results from a number of different verification studies (Hurricane Katrina, 1986 Sydney flood, precipitation probability density function, extreme precipitation indices, rainfall-runoff modeling) over different regions (U.S., Australia, China) are presented and are encouraging.

In the second part, a non-stationary probabilistic framework based on Extreme Value Theory (EVT) is developed to assess potential footprints of climate change on characteristics (intensity and frequency) of extreme precipitation events. The main goal was to investigate if there have been statistically significant changes and trends in the Probability Distribution Function (PDF) of precipitation extremes over time. The proposed probabilistic scheme which is based on Generalized Extreme Value (GEV) and Generalized Pareto (GP) distributions proved to be effective in stochastically modeling the behavior of precipitation extremes over the past three decades. The statistically significant trends in the time-variant GEV distribution are tested over the U.S. The results show that Eastern and particularly the Northeastern parts of the U.S. are experiencing positive trends in the intensity and frequency of extreme precipitation events.
Chapter 1: Introduction

1.1. Climate Extremes and Data Issue

Every year, climate extremes (e.g., heat waves, cold snaps, floods, droughts, hurricanes etc.) claim lives and cause billions of dollars in economic damages and human hardship. There are numerous examples of these weather and climate extremes all around the world (e.g., the 2013 Colorado State flood, the 2012 heat waves and hottest year on record in the United States, the 2012 Hurricane Sandy (also known as Superstorm or Hybrid Storm Sandy), and the 2010 Pakistan flood and Russian heat wave). In November 2013, the super typhoon Haiyan struck Southeast Asia and killed more than 6,000 in the Philippines. Hurricane Katrina (August 2005), one of the five most deadliest and the costliest hurricanes ever to strike the U.S., caused inflicted loss of lives and economic damages in the Southeast of the U.S. (Graumann 2006). As reported by the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC), in 2011 and 2012 across the United States alone, there were, respectively, a total of fourteen and eleven “billion-dollar weather and climate disasters” that each of which resulted in losses of human lives and economic impacts of $1 billion or more (NCDC 2013, Smith and Katz 2013).

An important question here is whether or not there have been changes in such events over time due to climate change. As stated by the Intergovernmental Panel on Climate Change (IPCC)
5th Assessment Report (AR5), “Changes in many extreme weather and climate events have been observed since about 1950. It is very likely that the number of cold days and nights has decreased and the number of warm days and nights has increased on the global scale. It is likely that the frequency of heat waves has increased in large parts of Europe, Asia and Australia. There are likely more land regions where the number of heavy precipitation events has increased than where it has decreased. The frequency or intensity of heavy precipitation events has likely increased in North America and Europe.” (Stocker et al. 2013). Brown et al. (2008) detect changes in observed extreme daily temperature, globally, since 1950. The results show that for most regions of the world, the daily maximum and minimum temperatures have increased. Also, as reported in the U.S. 3rd National Climate Assessment (NCA) report, the frequency of extreme precipitation events across the Southeast U.S. has increased (Kunkel et al. 2013b), while a statistically significant trend in the occurrence of extreme precipitation events in the Southwest U.S. is not identified (Kunkel et al. 2013c). In addition, a study by Peterson et al. (2013a) shows that over the last several decades, the heat and the cold waves over the U.S. are increasing and decreasing, respectively.

The key factor in conducting long-term hydrological and climate studies, particularly regarding climate extreme, is availability of reliable datasets. Long-term global observation-based datasets are of primary importance in such studies. As defined by the World Meteorological Organization (WMO), at least 30 years of historical weather data are generally required for long-term climatological studies (Burroughs 2003). Since the focus of this dissertation is on precipitation, it is important to discuss different types of available data sources for this variable. As is the case for almost any weather related variable, ground-based rain gauge
networks are one of the most widely used sensors to measure precipitation. The longest historical precipitation observations are available through rain gauge records (Xie et al. 2003). A significant number of climate studies have focused on precipitation analysis using historical observational data (e.g., Karl et al. 1995; Higgins et al. 2007; among many). Although gauges can directly measure the rain that reaches the ground surface, they are land-based, sparse and point measurements.

Radar is another data source which can provide a viable alternative to gauge measurement. Radar data is a great source of providing high resolution precipitation estimates (Lin and Mitchell 2005); however the data is not available wherever the radar coverage is poor and the radar beam is blocked (Westrick et al. 1999; Maddox et al. 2002). Moreover, the radar data record is not large enough for long-term historical studies.

Satellite observation is another important data source for precipitation. Satellite-based observations have more complete coverage, particularly over oceans, high altitudes, and remote regions where gauge measurements are very limited or unavailable. Geostationary Earth Orbit (GEO) satellites are capable of providing images every 15-30 minutes in multiple spectral bands of the cloud patterns and evolution over time. GEO infrared-based (IR) algorithms are considered to be effective at identifying tropical convective systems in both day and night time, but the performance of an IR-based algorithm is less accurate for warm rain clouds, and cold high non-raining cirrus clouds. Passive/active Microwave (PMW) sensors aboard Low Earth Orbit (LEO) satellites can measure hydrometeor distribution in rain clouds more directly than GEO-based sensors. The low sampling frequency of LEO satellites, however, limits the
effectiveness of PMW-based rainfall data retrieval at short time scales. Integration of multiple LEO satellites can improve this sampling limitation and as a result improve precipitation estimation in short-time scale considerably.

Reanalysis products are also viewed as other sources of “data” for hydrometeorological variables which are developed to study how weather and climate are changing over time. However they are not considered as observation. Reanalysis (or retrospective analysis) is a scientific method which combines historical observations with a model forecast to reconstruct atmospheric variables (Smith et al. 2014). Usually, the temporal coverage of reanalyses products is for the past few decades. Precipitation is one of the most difficult parameters for a reanalysis to produce, and extreme events are likewise challenging for the reanalyses background forecast model.

1.2. Research Motivations

Detecting footprints of climate change on our climate system is a critical and challenging research area. Particularly, due to the extent of the devastating impacts that weather and climate extremes can cause, it is extremely vital to carefully study, investigate, and model the behavior of such catastrophic weather events in a changing climate. One of the main often-asked questions about weather and climate extremes is whether or not the occurrences of these extremes are results of global warming and climate change. In my view, this is not a well-posed question, at least at the present moment. We cannot attribute occurrence of a single extreme event solely to
global warming and long-term climate trends. Climate change can play the role of a performance enhancer for a climate extreme event, injecting more energy and power to the event, but it cannot be the only cause of the occurrence of such event. Natural variability of the climate system plays an important role as well. What we know as of now is the unequivocal fact that our climate system is warming (as been mentioned in the 2013 Intergovernmental Panel on Climate Change (IPCC) report). The warmer atmosphere can and does hold more water vapor than it used to. This increased amount of moisture changes the characteristics (e.g., frequency, duration, return period) of weather events. These changes may be magnified for climate extremes, as a small shift in the mean of the frequency distribution may lead to very large changes in the tail of the distribution where climate extremes are located. This dissertation is initially motivated by the fact that climate change can shift the odds for extreme weather events and therefore, tries to investigate how climate change has affected the tail of precipitation distribution over time. In other words, I seek to understand how climate change has altered the characteristics of precipitation extremes over time.

To find robust answers to the above questions, availability of a long-term, high-resolution, consistent dataset is crucial. As explained in previous section, the current data sources are either sparse land-based measurements (e.g., ground-based measurements), or they have limited spatial coverage (e.g., radar observations in mountainous regions). Even in the case of satellite-based precipitation products, current datasets are either of insufficient duration for climate studies, or their temporal and/or spatial resolution is too coarse for analysis of climate extremes. According to the assessment of current global precipitation products by the World Climate Research Programme (WCRP) Global Energy and Water Exchanges (GEWEX),
improving and extending the current global precipitation products is identified as the biggest challenge facing the scientific community and pursuing efforts for obtaining higher spatial and temporal resolution precipitation data is recognized as being of great importance and need (Gruber and Levizzani 2008). This research is greatly motivated by this need. Using observations from multiple satellites, I will develop a consistent, long-term, high-resolution, and global satellite-based precipitation dataset to be used for long-term hydrological and climate studies.

1.3. Scientific Objectives

The main scientific objectives of this dissertation can be divided into two main parts. As the first objective, this dissertation tries to fill the existing gap in current precipitation data products with respect to long-term hydro-climatological studies. As the second objective, the observed trends and changes in extreme precipitation events over time in a changing climate are modeled using satellite-based observations. These objectives are very well aligned with the NOAA’s long-term goal as addressed in the NOAA’s Next Generation Strategic Plan. To accomplish these goals, this study aims to address the following main objectives and questions:

i. Developing a consistent, long-term (multi-decadal), high-resolution, global precipitation product from multiple satellite observations as a precipitation Climate Data Record (CDR) to study the water cycle at higher resolutions than previously possible.
ii. Conducting different validation studies for testing the efficacy and efficiency of the newly developed precipitation CDR product when compared to ground-truth and other currently available high-resolution satellite-based precipitation products.

iii. Statistical modeling of extreme precipitation events using satellite-based precipitation products. This study seeks to investigate potential changes and trends in the probability distribution of extreme precipitation events over time. The main goal is to model the observed behavior of extreme precipitation events and investigate whether there has been a statistically significant trend in the frequency distribution of heavy precipitation events over time due to climate change. In other words, this study seeks to find out how climate change has altered the characteristics (e.g. frequency, intensity, duration) of climate extremes.

1.4. Dissertation Outline

This dissertation consists of five main chapters. Chapter 2 presents and discusses the methodology and data used for reconstruction of global precipitation from multi-satellite observations. Chapter 3 details the various different verification studies for testing the performance of the newly developed product against ground-truth and other satellite-based precipitation products are provided. Chapter 4 provides detailed explanations about developing
different statistical models to study the behavior of precipitation extremes in a changing climate. Chapter 6 summarized the findings of this study and discuss about future extensions of this work.
Chapter 2: Reconstruction of Global Daily Precipitation from Multi-Satellite Observations (PERSIANN-CDR)

2.1. Review of Current Satellite-based Precipitation Products

Reliable observations are needed for sustainable global-scale monitoring of precipitation variability and trends in space and time at resolutions suitable for climate studies (e.g., Solomon et al. 2007; Wentz et al. 2007). Among various different data sources, satellite observations can provide a complete global spatial coverage. Development of global satellite-based precipitation datasets has been an emerging research area in the past three decades. Since 1997, the long-lived Tropical Rainfall Measurement Mission (TRMM) has significantly improved rainfall retrievals over the tropical regions (Kummerow et al. 1998; Simpson et al. 1988; Kummerow et al. 2000). Though TRMM satellite is running out of fuel and is slowly descending preparing for re-entry to the Earth’s atmosphere, it still continues to collect data. The Global Precipitation Measurement (GPM) mission, which was launched in 27 February 2014, combines observations from multiple microwave (MW) sensors on LEO satellites to provide information on global precipitation distributions in 3-hour periods (Hou et al. 2008; Hou et al. 2013). The Global Precipitation Climatology Project (GPCP, Huffman et al. 1997; Huffman et al. 2001; Adler et al. 2003) and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP—Xie and Arkin 1997; Xie et al. 2003) are two other datasets with long records of data. These two datasets, with
their global coverage (over both ocean and land) and long-term time periods, albeit at 2.5° and monthly spatial and temporal resolutions respectively, have contributed greatly to climate change studies (e.g., Yilmaz and DelSole 2010; Allan et al. 2010; Peterson et al. 2012; Bourassa et al. 2013; Ma et al. 2013; Kucera et al. 2013; Rossow et al. 2013; among many). However, their coarse spatial and temporal resolution limits their ability to capture the spatial details and dynamics of extreme precipitation events, particularly hurricanes and convective storms, whose life cycles range from hours to days. Table 2.1 shows the time coverage and spatiotemporal resolution of current major satellite-based precipitation datasets (the last row shows the specifications of the product that is developed and presented in this study).

Table 2.1. Coverage and spatiotemporal resolutions of major satellite precipitation products, including PERSIANN-CDR.

<table>
<thead>
<tr>
<th>Product</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Period</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPCP</td>
<td>Monthly/Pentad</td>
<td>2.5°</td>
<td>1979 – (delayed) present</td>
<td>90S - 90N</td>
</tr>
<tr>
<td>GPCP-1DD</td>
<td>Daily</td>
<td>1°</td>
<td>1996 – (delayed) present</td>
<td>90S - 90N</td>
</tr>
<tr>
<td>CMAP</td>
<td>Monthly/Pentad</td>
<td>2.5°</td>
<td>1979 – (delayed) present</td>
<td>90S - 90N</td>
</tr>
<tr>
<td>TMPA v7</td>
<td>3 hourly</td>
<td>0.25°</td>
<td>1998 – (delayed) present</td>
<td>50S - 50N</td>
</tr>
<tr>
<td>CMORPH</td>
<td>0.5 hour</td>
<td>~0.07°*</td>
<td>2002 – present</td>
<td>60S - 60N</td>
</tr>
<tr>
<td>PERSIANN</td>
<td>0.5 hour</td>
<td>0.25°</td>
<td>2000 – present</td>
<td>60S - 60N</td>
</tr>
<tr>
<td>PERSIANN-CCS</td>
<td>0.5 hour</td>
<td>0.04°</td>
<td>2003 – present</td>
<td>60S - 60N</td>
</tr>
<tr>
<td>PERSIANN-CDR</td>
<td>Daily</td>
<td>0.25°</td>
<td>1983 – (delayed) present</td>
<td>60S - 60N</td>
</tr>
</tbody>
</table>

*CMORPH resolution is ~8km.
CMORPH produces high-resolution global precipitation analysis based on LEO-based PMW observations from different sources, such as the Defense Meteorological Satellite Program (DMSP) F-13, F-14, F-15 (Special Sensor Microwave Imager; SSM/I), and F-16 (Special Sensor Microwave Imager/Sounder; SSMI/S), the NOAA-15, 16, 17, and 18 (Advanced Microwave Sounding Unit-B; AMSU), Aqua (Advanced Microwave Scanning Radiometer for the Earth Observing System; AMSR-E), and the TRMM Microwave Imager (TMI). Half-hourly CMORPH data at 8-km spatial resolution has been operationally produced since 22 November 2002 (Joyce et al. 2004). PERSIANN primarily uses infrared brightness temperature data from geostationary satellites to estimate rainfall rate, updating its parameters using PMW observations from low-orbital satellites. The PERSIANN half-hourly 0.25° rain rate product is available for March 2000-present (Hsu et al. 1997). The version 7 TMPA data product (Huffman et al. 2007) is in three-hourly and 0.25° temporal and spatial resolutions, respectively, starting from January 1998. The TMPA algorithm combines high-quality PMW observations and IR data from geostationary satellites to derive precipitation. The NRL-Blend satellite precipitation dataset is another precipitation product based on both geostationary visible and infrared data and PMW observations (Turk et al. 2010). The NRL global precipitation accumulation product is available at 0.25° and 3-hourly spatiotemporal resolutions starting in January 2004.

2.2. Problem Statement

As can be seen with the aforementioned products, high-resolution, satellite-based, precipitation estimation algorithms generally need PMW observations as a major source of input.
data. Such algorithms would be unable to provide reliable precipitation estimates when PMW samples are limited or unavailable. This is particularly the case for the pre-1997 period, where only 1 or 2 PMW observations are available daily. This constraint limits the application of these algorithms in the reconstruction of precipitation records during the pre-PMW era necessary for more significant global climate studies.

Among high-resolution, satellite-based, precipitation estimation algorithms, PERSIANN, because of its primary reliance on infrared information which dates back to 1979, is very suitable for estimating historical precipitation over the past three decades. To meet the calibration requirement of PERSIANN, the model is pre-trained using the National Centers for Environmental Prediction (NCEP) Stage IV hourly precipitation data. Then the parameters of the model are kept fixed, and the model is run for the full historical record of IR data. The archive of global IR data is available through the International Satellite Cloud Climatology Project (ISCCP). To reduce the biases in the PERSIANN-estimated precipitation, while preserving the spatial and temporal patterns in high resolution, 2.5° monthly GPCP precipitation data were utilized. The bias-corrected PERSIANN precipitation estimates maintain a monthly total consistent with the monthly GPCP product. The final product, called the PERSIANN-CDR (PERSIANN-Climate Data Record), provides more than 30 years of near-global (60°S-60°N) daily precipitation data at 0.25° spatial resolution. Consistency of PERSIANN-CDR precipitation data over the entire record was maintained throughout the modeling process. All of these characteristics make PERSIANN-CDR a useful product for global climate studies at a scale relevant to extreme weather events.
The scope of this chapter is organized as follows: Section 2.3 presents information regarding the utilized data. Section 2.4 provides a detailed explanation of the methodology. Section 2.5 describes the details of the PERSIANN-CDR product. Section 2.6 provides detailed information about the availability of PERSIANN-CDR as an operation precipitation climate data record at the NOAA NCDC website. Section 2.7 summarizes the findings of this chapter.

2.3. Data

2.3.1. Stage IV Precipitation Data

The Stage IV precipitation product is made available by the NCEP Environmental Modeling Center (EMC) from high-resolution Doppler NEXRADs and hourly rain gauge data over the continental United States. Stage IV data are provided over the 4x4 km² Hydrologic Rainfall Analysis Project (HRAP) national grid system and are made available at hourly, 6-hourly, and 24-hourly scales. The 12 River Forecast Centers (RFCs) of the National Weather Service (NWS) do manual quality control and NCEP further mosaics all the data received from the RFCs. Different studies have investigated the uncertainties associated with Stage IV data (Westrick et al. 1999; Young et al. 2000; Maddox et al. 2002; Young and Brusnell 2008; Habib et al. 2009), and great efforts have been made to improve the quality of the data (Lin and Mitchell 2005). More information about Stage IV data can be obtained from Fulton et al. (1998), and this webpage link: http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/. This dataset has been widely used as a reference data for evaluation of satellite based precipitation estimations.
(Ebert et al. 2007; Zeweldi and Gebremichael 2009; Anagnostou et al. 2010; AghaKouchak et al. 2011). In this study, Stage IV radar data are used for the initial training of the Neural Network (NN) model, as well as evaluating the performance of PERSIANN-CDR.

2.3.2. Gridded Satellite Infrared Data (GridSat-B1)

As the custodian of major climate datasets, the NOAA National Climatic Data Center (NCDC) maintains a historical archive of data from GEO satellites as compiled by the International Satellite Cloud Climatology Project (ISCCP). ISCCP B1 global geostationary observations (Knapp 2008a) are 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 2 (FY2) series. The ISCCP B1 IR brightness temperature data available from these GEO sources cover the time period from 1979-present at space and time resolutions of 10-km and 3-hour intervals. Better global coverage began in 1983, albeit with a gap over the Indian Ocean due to a lack of GEO data (Rossow and Schiffer 1991; Rossow and Garder 1993; Knapp 2008a).

Gridded Satellite (GridSat-B1) data are derived from merging ISCCP B1 IR data (see Knapp et al. 2011 for complete details). GridSat-B1 provides near-global data for three channels: visible, infrared window (IRWIN), and infrared water vapor (IRWVP). A sample image of each of these channels is shown in Figure 2.1.
Figure 2.1. Sample image of GridSat-B1 data, IR window (top), water vapor (middle), and visible (bottom) (NOAA/NCDC)
GridSat-B1 IRWIN data, the main input data to the PERSIANN-B1 model, are merged using the nadir-most observations at each grid point and adjusted for different biases in satellite sensors (see Knapp 2008b for details regarding inter-sensor differences). The infrared window brightness temperature data, GridSat-B1 CDR (Knapp et al. 2011), spans from 1 January 1980 to the current time, covering the globe from 70°S-70°N and 180°W-180°E. The GridSat-B1 IRWIN brightness-temperature data are gridded to a 0.07° resolution latitude-longitude grid and are available at the 3-hourly time scale. To make GridSat data compatible with the input structure of the PERSIANN model, these data were averaged and upscaled to a 0.25° resolution and filtered to remove data values out of the normal range for IR data.

2.3.3. Global Precipitation Climatology Project (GPCP)

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 (WCRP 1986; Adler et al. 2003). Currently, three GPCP global precipitation products are available (see Table 1): (1) monthly, 2.5° merged analysis (1979-present), (2) pentad, 2.5° merged analysis (1979-present), and (3) 1° daily (1DD) merged analysis (October 1996-present).

The GPCP monthly 2.5° merged analysis was constructed using multi-satellite (SSM/I and IR) precipitation estimates, adjusting the latter using gauge analysis to remove large-scale bias, and then merging satellite and gauge analyses into a final product (Huffman et al. 1997;
Huffman et al. 2001; Adler et al. 2003; Huffman et al. 2009). GPCP monthly rainfall includes precipitation gauge analysis provided by the Global Precipitation Climatology Centre (GPCC) (Rudolf et al. 1993; Rudolf et al. 1994; Schneider et al. 2008). The existing long-term monthly GPCP product has been widely used for climatology studies on a global scale. In this study, the latest version of the GPCP monthly 2.5° product (version 2.2) was used for correcting the biases of the PERSIANN rain rate estimates. GPCP version 2.2 is available at http://precip.gsfc.nasa.gov and currently spans January 1979- November 2012. Documentation of the GPCP version 2.2 is accessible at http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/V2.2_doc.pdf. In addition to the monthly product, the GPCP 1° daily precipitation product (Huffman et al. 2001) was used for evaluation purposes. GPCP 1DD documentation is available at http://www1.ncdc.noaa.gov/pub/data/gpcp/1dd-v1.1/1DD_v1.1_doc.pdf.

2.4. Methodology

The development of the PERSIANN-CDR precipitation product is aimed at addressing the need for a consistent, long-term, high-resolution near-global dataset to study the spatial and temporal characteristics of precipitation in a scale relevant to climate studies. In this study, the PERSIANN algorithm is applied to the historical archive of GridSat-B1 infrared window observations from GEO satellites to generate 3-hourly rain rate estimates (1980-2012) at 0.25° for the region between (60°S-60°N). To be consistent throughout this paper, the output from the PERSIANN model using GridSat-B1 data with no PMW training and no bias correction is called PERSIANN-B1. The GPCP monthly product is then used to remove the biases of the
PERSIANN-B1 rain rate estimates, making it consistent with the GPCP monthly product. Adjusted PERSIANN-B1 rain rate estimates resulting from this stage are represented as $\text{Adj}_r\text{PERSIANN-B1}$ in this paper. Lastly, the 3-hourly $\text{Adj}_r\text{PERSIANN-B1}$ precipitation data are accumulated to the daily scale to produce the PERSIANN-CDR product. Detailed information regarding the PERSIANN-CDR algorithm and the bias-adjustment process is presented in the following subsections.

### 2.4.1. PERSIANN-CDR Precipitation Estimation Algorithm

The existing PERSIANN algorithm provides global precipitation estimates using combined IR and PMW information from multiple GEO and LEO satellites. The algorithm uses an Artificial Neural Network (ANN) model to extract cold-cloud pixels and neighboring features from GEO long-wave infrared images (~10.2-11.2 μm) and associates variations in each pixel’s brightness temperature to estimate the pixel’s surface rainfall rate (Hsu et al. 1997; Hsu et al. 1999; Sorooshian et al. 2000). The PMW information from LEO satellites and the CPC globally-merged, full-resolution (4 km, 1/2 hourly) IR data from GEO satellites (Janowiak et al. 2001) are processed to 0.25°x0.25° lat-long spatial resolution for rainfall estimation using the PERSIANN model.

In this CDR product, in order to eliminate the need for PMW observations, the nonlinear regression parameters of the ANN model are trained and remain fixed when PERSIANN is used for retrospective estimation of rainfall rates using the 3-hourly GridSat-B1 IRWIN data.
Furthermore, a bias-adjustment stage based on GPCP 2.5° monthly precipitation data is incorporated into the reconstruction process. The data-generation framework incorporates GPCP monthly rainfall data to adjust 3-hourly PERSIANN-B1 rainfall estimates and therefore ensures data consistency and quality. Figure 2.2 shows a simplified schematic of the current operational PERSIANN-CDR system.

![Figure 2.2. A schematic of the PERSIANN-CDR algorithm for reconstruction of historical precipitation.](image)

2.4.2. Adjusting Daily PERSIANN Data Using Monthly GPCP Data

To reduce any biases in the 3-hourly PERSIANN-B1 estimates, while at the same time preserving spatial and temporal patterns in the high-resolution precipitation estimates, GPCP monthly rainfall at 2.5° resolution is used to adjust the high-resolution PERSIANN-B1 estimates.
A separate correction is performed for each 2.5° grid box of PERSIANN-B1 data for each month of each year. For each 2.5° \((i', j')\) grid box, the corresponding 0.25° 3-hourly PERSIANN-B1 rain rate estimates \(r_{\text{PERSIANN-B1}}(i, j)\) are spatially and temporally aggregated to 2.5° monthly scale \((R_{\text{PERSIANN-B1}}(i', j'))\). In doing so, a threshold value \((thd\text{ in Eq. (2.1)})\) 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).

\[
R_{\text{PERSIANN-B1}}(i', j') = \sum_{i=1}^{nd} \sum_{j=1}^{nh} \left( \sum_{i=1}^{10} \sum_{j=1}^{10} \left[ r_{\text{PERSIANN-B1}}(i, j) \geq thd \right] \right) 
\]  

(2.1)

In Eq. (2.1), \(i\) and \(j\) are the high resolution (0.25°) latitude-longitude, \(i'\) and \(j'\) are the low resolution (2.5°) latitude-longitude, \(nh\) is the number of 3-hourly PERSIANN-B1 samples in each day, and \(nd\) is the number of days in each month. The correction factor for each monthly 2.5° grid cell is then calculated as follows;

\[
w(i', j') = \frac{R_{\text{GPCP}}(i', j')}{R_{\text{PERSIANN-B1}}(i', j')} 
\]  

(2.2)

where \(R_{\text{GPCP}}\) is the 2.5° monthly GPCP precipitation for a given pixel. We note that in some locations, such as high latitudes and in dry regions with very low rainfall values, the weight \((w)\) 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\(^\circ\)). The results show that \( thd = 0.1 \) and maximum \( w = 20 \) is perhaps the best combination.

The monthly bias is then spatially downscaled and removed from the PERSIANN-B1 estimates at 0.25\(^\circ\) resolution using the correction factor. Each 2.5\(^\circ\) grid cell covers 10x10 pixels of PERSIANN-B1 estimation. To prevent discontinuities at the edges of the 10x10 pixels after adjustment, the correction factor \( (w(i, j)) \) is assigned to the center of each 10x10 pixel block and then a linear interpolation method is applied to find the correction factor at each 0.25\(^\circ\) pixel \( (w(i, j)) \). Thus, each 0.25\(^\circ\) pixel is corrected with a separate adjustment factor which results in a smooth and continuous transition over the edge of the 10x10 pixels. The GPCP-adjusted 0.25\(^\circ\) monthly PERSIANN-B1 precipitation, \( Adj_{R_{\text{PERSIANN-B1}}}(i, j) \), is then calculated as follows:

\[
Adj_{R_{\text{PERSIANN-B1}}}(i, j) = w(i, j) \ast R_{\text{PERSIANN-B1}}(i, j)
\]  

(2.3)

Due to the linearity of the bias adjustment process, the correction factor can be applied to the higher temporal resolution PERSIANN-B1 estimates. Thus, the GPCP-adjusted 0.25\(^\circ\) 3-hourly PERSIANN-B1 precipitation, \( Adj_{r_{\text{PERSIANN-B1}}} \), is calculated as follows:
\[ Adj \_ r_{\text{PERSIANN-}B1}(i, j) = w(i, j) \times r_{\text{PERSIANN-}B1}(i, j) \] (2.4)

Eq. (2.4) is applied to each 3-hourly PERSIANN-B1 estimate, which results in distributing the monthly bias correction to the 3-hourly precipitation estimates. The bias adjusted PERSIANN-B1 precipitation data maintains GPCP monthly total precipitation. To reduce uncertainty, the 3-hourly \( Adj \_ r_{\text{PERSIANN-}B1} \) rain rate data are accumulated to daily scale to produce the PERSIANN-CDR product.

\[ \text{PERSIANN-CDR}(i, j) = \sum_{n=1}^{N} Adj \_ r_{\text{PERSIANN}}(i, j, n) \] (2.5)

where \( N \) is the number of \( Adj \_ r_{\text{PERSIANN-}B1} \) samples per day.

2.5. Demonstration of PERSIANN-CDR Product

PERSIANN-CDR is a daily 0.25° precipitation product that covers the area between 60°S-60°N latitude and 0°-360° longitude. The data product spans the period of 01/01/1983 to (delayed) present time. It is noteworthy that before using GridSat-B1 IR data as the input to the model, an extensive quality control procedure was performed on whole record of 3-hourly GridSat IR images. Figure 2.3 shows few sample images in year 1980 which were taken out from the input record. Work is in progress to complete the years of 1980, 1981 and 1982, where the quality of the GridSat-B1 data for few months of each year needs further improvement.
Figure 2.3. Sample images (year 1980) of GridSat IR data (top brightness temperature of the cloud in degree Kelvin) which were taken out from the PERSIANN-CDR precipitation estimation process due to quality issues.
PERSIANN-CDR is updated to near current time as soon as both GridSat-B1 IR data and GPCP monthly rainfall data updates become available. Figure 2.4 shows the results of the adjustment of PERSIANN-B1 estimates for August 2005 compared to GPCP (top), before adjustment (middle), and after adjustment (bottom). After GPCP adjustment, even by visual comparison, it is clear that there is an improvement of the PERSIANN-CDR estimates toward GPCP estimates.

To investigate whether or not the GPCP precipitation product was properly assimilated into the PERSIANN-CDR product, the Mean Areal Precipitation (MAP) for both the northern and southern tropical regions from monthly PERSIANN-B1, PERSIANN-CDR, and GPCP data were calculated. As shown in, Figure 2.5 the resulting time series of the mean areal precipitation from PERSIANN-CDR matches well with that of GPCP monthly product.

PERSIANN-CDR was also compared with GPCP-1DD in daily scale. Mean areal precipitation estimates 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 2.6. As shown, improvements become evident after applying the bias-adjustment algorithm to PERSIANN-B1 estimates (green line, Figure 2.6). 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 SSMI/S) in its estimate. It is noteworthy that no PMW data is used in PERSIANN-CDR, except indirectly from GPCP monthly data.
Figure 2.4. Global rainfall maps (mm/day) for August 2005 from GPCP 2.5° (top, Adler et al. 2003), PERSIANN-B1 0.25° (middle), and PERSIANN-CDR 0.25° (bottom) monthly datasets.
Figure 2.5. Mean areal precipitation (mm/day) for Northern (0°-30°N, top) and Southern (0°-30°S, bottom) Tropics from monthly GPCP (red, Adler et al. 2003), PERSIANN-B1 (green), and PERSIANN-CDR (dashed blue) datasets.
Figure 2.6. Daily global (60°S-60°N) mean areal precipitation (mm/day) for the period of 2007-2009 for GPCP-1DD (red, Huffman et al. 2001), PERSIANN-B1 (green), and PERSIANN-CDR (blue).

Similar graphs for the Northern Hemisphere (0°-60°N), Tropics (0°-30°N), and Extratropics (30°N-60°N) for the period of 1997-2012 are displayed in Figure 2.7. As shown, even without any GPCP-adjustment, the PERSIANN-B1 rain rate estimates show good agreement with GPCP 1DD in tropical regions. As shown in Figure 2.7, 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-60°S).
Figure 2.7. 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, Huffman et al. 2001), PERSIANN-B1 (green), and PERSIANN-CDR (blue).

To examine the performance of PERSIANN-CDR in the case of extreme precipitation events, the top 5% heavy rainfall (mm/day) patterns from GPCP-1DD (Figure 2.8, top), PERSIANN-CDR up-scaled to 1° (Figure 2.8, middle), and PERSIANN-CDR 0.25° (Figure 2.8, bottom).
bottom) for the period of 1997-2012 are extracted. As shown in Figure 2.8, the global patterns of extreme precipitation events are closely similar. PERSIANN-CDR depicts larger rainfall for some extreme precipitation events particularly over the Intertropical Convergence Zone (ITCZ). Some of the observable differences are due to the spatial resolution differences between the two products, as extremes are smoothed and their intensity dampened in coarser resolution. This results in less intense extremes in the GPCP-1DD when compared to PERSIANN-CDR. Answering which of these two products is closer to the truth cannot be fully tested due to the lack of available measurements over the oceans. With the 0.25° rainfall data from PERSIANN-CDR (Figure 2.8, bottom), the distribution of extremes can be seen at a finer resolution. These results show that the PERSIANN-CDR precipitation dataset have high potential to be a useful product for long-term climate studies at a much finer time scale than previously available.

2.6. Operational PERSIANN-CDR at NOAA NCDC

PERSIANN-CDR provides more than 30 years of daily, 0.25° and near-global satellite-based precipitation data. As a high-resolution satellite-based dataset, PERSIANN-CDR provides time series of precipitation of sufficient length, consistency and continuity to study trends and observed changes in global and regional precipitation patterns due to climate change and natural variability. These characteristics are consistent with the CDR definition as described in the National Research Council report (2004). After close collaborations between the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine and the NOAA NCDC scientists, PERSIANN-CDR is made available to the public as an operational
Figure 2.8. Top 5% heavy rainfall (mm/day) maps from GPCP-1DD (top, Huffman et al. 2001), PERSIANN-CDR 1° (middle), and PERSIANN-CDR 0.25° (bottom) for the period of 1997-2012.
climate data record on the NOAA NCDC CDR Program website under the Atmospheric CDRs category via http://www.ncdc.noaa.gov/cdr/operationalcdrs.html. Figure 2.9 provides a concise overview of PERSIANN-CDR specification and potential applications to the scientific community and the users. In addition, a brief description of PERSIANN-CDR can be found on the CHRS website (http://www.chrs.web.uci.edu/).

Figure 2.9. Overview of PERSIANN-CDR specifications, input data, and potential applications (NOAA/NCDC/CDRP)
2.7. Chapter Summary and Conclusions

The main scientific goal of this part of the dissertation was to fulfill the need and gap among the current precipitation datasets. As a result, a new retrospective satellite-based precipitation dataset is constructed as a climate data record for hydrological and climate studies. Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR) provides daily and 0.25° rainfall estimates for the latitude band 60°S–60°N for the period of 01/01/1983 to delayed present. PERSIANN-CDR is aimed at addressing the need for a consistent, long-term, high-resolution and global precipitation dataset for studying the changes and trends in daily precipitation, especially extreme precipitation events, due to climate change and natural variability. PERSIANN-CDR is generated from the PERSIANN algorithm using GridSat-B1 historical multi-satellite infrared observations. It is adjusted using the Global Precipitation Climatology Project (GPCP) monthly product to maintain consistency of the two datasets at 2.5° monthly scale throughout the entire record. The PERSIANN-CDR 0.25°daily rainfall estimations can be used to study the water cycle at higher resolutions than previously possible.

PERSIANN-CDR should prove to be a useful dataset for addressing various key hydrological and climatological research questions which require longer and finer resolution (daily, 0.25°) data than previously available. The next chapter provides deep insights into the efficacy and accuracy of PERSIANN-CDR rainfall data in different applications. Different case studies over different parts of the world are carried out to test the accuracy and efficacy of the PERSIANN-CDR long-term 0.25° daily rainfall estimations.
Chapter 3: Verification of PERSIANN-CDR

3.1. Introduction

This chapter focuses on different verifications studies to test the validity of the PERSIANN-CDR rainfall estimations. The following verification studies have been carried out. The first case study is related to the Hurricane Katrina that struck the U.S. in August 2005. Gauge-adjusted Stage IV radar data and TMPA v7 satellite-based precipitation data are used for comparison purposes (Section 3.2.1). The second case study targets the pre-1997 period when current high-resolution satellite-based precipitation products are not available. It focuses on a major flood event over Sydney, Australia in 1986, using the gridded daily gauge data provided by the Australia’s Bureau of Meteorology (Section 3.2.2). The third case study shows the comparison of the probability density function of PERSIANN-CDR with those of the gauge observations and the TMPA v7 over the Contiguous United States (CONUS) (Section 3.2.3). The forth study tests the capability of PERSIANN-CDR in reproducing the annual count of rainy days over the CONUS (Section 3.2.4). The fifth verification study focuses on evaluating the performance of PERSIANN-CDR in capturing the behavior of daily extreme precipitation events in China during the period of 1983-2006 (Section 3.2.5). The sixth study, evaluates the performance of daily PERSIANN-CDR rainfall estimations in simulating streamflow in a rainfall-runoff model (Section 3.2.6). Details of each of the above studies are provided below.
3.2. Case Studies

3.2.1. Hurricane Katrina

Katrina occurred in August 2005 and is considered the most costly extreme weather event ever to strike the United States (Graumann 2006). PERSIANN-CDR precipitation data are compared with Stage IV radar data at the 0.25° spatial scale during Hurricane Katrina. In order to examine the performance of PERSIANN-CDR with other satellite-based precipitation products, TMPA v7 was also included in the analysis. As shown in Figure 3.1, PERSIANN-CDR (Figure 3.1b) shows similar precipitation patterns to the radar data (Figure 3.1c). 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.

The scatterplots of PERSIANN-B1, PERSIANN-CDR, and TMPA v7 against Stage IV radar data are plotted and the relevant statistics are calculated. As shown in Figure 3.2, the correlation coefficient between PERSIANN-B1 and radar data is significant (0.84), implying that even before doing any adjustment, PERSIANN-B1 is depicting similar patterns, albeit with lesser intensity (Figure 3.1a) when compared to radar data (Figure 3.1c). However, the RMSE (17.73) and Bias (-0.4624) are rather large. After applying the GPCP-adjustment algorithm, the PERSIANN-CDR Bias (-0.0883) is significantly reduced. PERSIANN-CDR and TMPA were each compared to Stage IV radar data.
Figure 3.1. Rainfall (mm/day) over land during Hurricane Katrina on 29 August 2005 from: (a) PERSIANN-B1, (b) PERSIANN-CDR, (c) Stage IV Radar (Lin and Mitchell 2005), and (d) TMPA v7 (Huffman et al. 2007). Black and gray pixels show radar blockages and zero precipitation, respectively.

As shown in Figure 3.2, PERSIANN-CDR shows a higher correlation coefficient than TMPA however, the bias in TMPA is lower than that in PERSIANN-CDR. These results show that each of the products have their respective strength and shortcomings.
Figure 3.2. Scatterplots of PERSIANN-B1 (top), PERSIANN-CDR (middle), and TMPA v7 (bottom, Huffman et al. 2007) against Stage IV radar data (Lin and Mitchell 2005) for 29 August 2005 during Hurricane Katrina.

3.2.2. The 1986 Sydney Flood

Sydney, Australia experienced a historic flood from precipitation received on 5-6 August 1986 resulting in significant losses and disruptions to transportation systems (Handmer 1988). The performance of PERSIANN-CDR during this flood event was evaluated against interpolated daily rainfall gauge data from the Australian Bureau of Meteorology available at 0.05° spatial resolution (Jones et al. 2009). The 0.05° gridded data was resampled to 0.25° for compatibility with PERSIANN-CDR estimates. The rainfall maps from gauge data and PERSIANN-CDR, along with the respective scatter plot and statistics are presented in Figure 3.3. As shown, during the 1986 Sydney flood, PERSIANN-CDR shows a good correlation coefficient (0.62) with gauge data.
Figure 3.3. Rainfall (mm/day) over land during Sydney Australia flood on 5 August 1986 from gauge observation (top left, Jones et al. 2009) and PERSIANN-CDR rain rate estimates (bottom left). The scatter plot and respective statistics are shown in the right.

Regarding the observed differences in Figure 3.3, besides uncertainties in PERSIANN-CDR and the sparsely distributed gauge network, some of the differences could be due to the temporal differences between the two datasets. PERSIANN-CDR daily grids correspond to a given 00:00-23:59 UTC time period, however the Australian interpolated gauge data represents the 24-hour accumulation of observations taken at 09:00 local time (Jones et al. 2009). The local
09:00-09:00 daily rainfall accumulation in New South Wales and Victoria corresponds with 23:00-23:00 UTC. This is a one-hour difference from the PERSIANN-CDR daily grids. Therefore, for the case of Sydney flood presented in this study, the temporal differences probably do not contribute significantly to the rainfall differences, but could be an issue in cases that cover larger areas and span from the east to the west with different local time. Such temporal inconsistencies in gauge observations may not be critical during dry seasons and drought years, but can introduce large differences in wet years, especially in cases of heavy rainfall events such as the August 1986 flood. In addition to the temporal differences, since gauge data are point measurements, the interpolated product might not be able to represent the spatial coverage properly which could also generate some significant differences between the satellite and gauge observations.

3.2.3. Probability Density Function

For validation of PERSIANN-CDR, the probability density functions (PDF) of PERSIANN-B1, PERSIANN-CDR, and TMPA v7 over CONUS for the period of 1998-2008 were extracted and compared. CPC unified gauge-based analysis of daily precipitation over CONUS was included as a reference to compare the performance of other products. The CPC US Unified Precipitation data is provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at http://www.esrl.noaa.gov/psd/data/gridded/data.unified.html (Higgins et al. 2000a). The PDFs of these products are plotted and shown in Figure 3.4. As shown, the PERSIANN-B1 PDF moves toward the CPC Gauge PDF after applying the bias-adjustment
Figure 3.4. Comparing the Empirical Probability Density Function (PDF) of PERSIANN-B1, PERSIANN-CDR, TMPA v7 (Huffman et al. 2007), and CPC Unified Gauge-Based Analysis of Daily Precipitation (Higgins et al. 2000a) over the CONUS during the period of 1998-2008.
algorithm. Comparing the PDFs of TMPA and PERSIANN-CDR with the PDF of CPC Gauge data indicates that both of these satellite-derived precipitation products are doing a reasonable job of matching CPC Gauge PDF, although they both overestimate or underestimate the probability for some precipitation ranges.

3.2.4. U.S. Number of Rainy Days

Expert Team on Climate Change Detection and Indices (ETCCDI) (Klein Tank et al., 2009), sponsored jointly by the World Meteorological Organization (WMO) Commission for Climatology (CCl), the Joint Commission for Oceanography and Marine Meteorology (JCOMM), and the World Climate Research Program (WCRP) on Climate Variability and Predictability (CLIVAR), has defined different precipitation indices for studying extremes. With respect to the number of rainy days, two indices are defined as the annual average count of days when rainfall ≥ 10 mm (R10mm), and the annual average count of days when rainfall ≥ 20 mm (R20mm). Using the CPC U.S. Unified gridded observational data, the performance of PERSIANN-CDR in reproducing the number of rainy days over the U.S. for 1983-2011 is tested. Figure 3.5 shows the results for CPC (top) and PERSIANN-CDR (bottom). In general, PERSIANN-CDR seems to follow the same patterns as depicted in CPC, underestimating on the west coast of the U.S. PERSIANN-CDR underestimation over the Sierra Nevada Mountains could be most likely due to 1) the type of precipitation in this region, being more toward snow rather than rain which PERSIANN-CDR detects, and/or 2) being orographic rain which satellite and radar might miss.
Figure 3.5. Annual average count of days where rainfall $\geq 10$ mm (left column) and rainfall $\geq 20$ mm (right column) for CPC (top), and PERSIANN-CDR (bottom) for 1983-2011.

The scatterplots and the statistics (Correlation Coefficient (Corr. Coef.), Root Mean Square Error (RMSE), and Bias) of PERSIANN-CDR against CPC are shown in Figure 3.6. As shown, correlation coefficient is high in both cases (greater than 0.9). PERSIANN-CDR tends to show a lower bias for R10mm.
3.2.5. Extreme Precipitation over China

To evaluate the performance of PERSIANN-CDR in capturing the behavior of precipitation extremes over China, two percentile-based extreme precipitation indices are calculated and compared respectively with the same indices based on the East Asia (EA, Xie et al., 2007) ground-based gridded daily precipitation data set. EA dataset contains observations from more than 1,300 ground-based stations across China. The distribution of rain-gauge stations in the EA data set in China are shown in Figure 3.7. Rain-gauge distribution in terms of areal coverage is highly uneven. There are approximately 1.8 gauges per 10,000 km$^2$ in the eastern monsoon region of China, but the number reduces to ~0.4 in the western and northwestern regions of China. The procedure used by Xie et al. (2007) to develop the gridded EA data set was through the interpolation of point measurements into 0.5°×0.5° grid boxes using the Optimal

![Gauge station distribution](image)

**Figure 3.7.** The distribution of rain-gauge stations in the EA data set.

The performances of PERSIANN-CDR in capturing the 99\textsuperscript{th} (RR99p) and 95\textsuperscript{th} (RR95p) percentile indices of the daily precipitation during the period of 1983–2006 at 0.5° × 0.5° grid box over China are shown in Figure 3.8. High values of annual RR99p and RR95p appear in southeastern China where generally extreme precipitation falls. In general, PERSIANN-CDR captures the spatial distribution of RR99p and RR95p similar to what the EA data set shows, with the increasing patterns in RR99p and RR95p from North to South and from East to West.
As far as the correlation coefficients between PERSIANN-CDR and the EA data set are concerned, the significance of correlation coefficient values was tested at the 0.05 level using the two-tailed t-test. As shown in Figure 3.8 (Pixel correlation column), high positive and statistically significant correlations are found in most regions for both indices, especially over the southeast regions, where extreme precipitation events occur more frequently. The scatterplot shows that the PERSIANN-CDR has good agreement with gridded-gauge data. PERSIANN-CDR tends to underestimate the low-value percentile indices over the dry and arid regions in the Western (Tibetan Plateau) and Northwestern (Taklamakan Desert) China. An important factor that may have influenced the results in this region is that the ground-based stations from which EA gridded data were produced are very sparse in this region. On the contrary, in the station-rich regions in Eastern China, the performance of PERSIANN-CDR is significant. PERSIANN-CDR slightly underestimates the values of very intense precipitation.
Figure 3.8. The 99th and 95th percentile indices of extreme daily precipitation from the EA data set (1st column), and PERSIANN-CDR (2nd column). The spatial correlation distribution and the scatterplots of the indices from the EA and PERSIANN-CDR data sets are shown in the 3rd and 4th columns, respectively. The stippled areas in the 3rd column show the significant correlation coefficient at the 95% level.
3.2.6. Rainfall-Runoff Modeling

The main goal of this study is to test the performance of PERSIANN-CDR in a hydrological rainfall-runoff modeling, and compare the result with that of another high-resolution satellite-based precipitation product, TMPA v7. Three basins from the Oklahoma Test Basins from the NOAA’s National Weather Service (NWS) Distributed Hydrologic Model Intercomparison Project - Phase 2 (DMIP2) are chosen (Figure 3.9).

Figure 3.9. The three study basins (SAVOY, ELMSP, and SLOA4) (NOAA/NWS)

Stage IV radar data is used as a reference data for evaluating the PERSIANN-CDR and TMPA precipitation data. All products are scaled to 0.25° and daily spatiotemporal resolution. Precipitation evaluation results, presented using Taylor Diagrams (Figure 3.10), show that TMPA and PERSIANN-CDR have close performances. In addition, PERSIANN-CDR shows improvements over PERSIANN product. It is noteworthy that while PERSIANN-CDR
outperforms PERSIANN due to the benefit from gauge adjustment, PERSIANN data is available in real-time while PERSIANN-CDR is available with few months delay.

Figure 3.10. Precipitation comparison plots for SAVOY (top), ELMSP (middle), and SLOA4 (bottom) basins for 2003-2010.
The NWS Office of Hydrologic Development (OHD) Hydrology Laboratory-Research Distributed Hydrologic Model (HL-RDHM) is used to simulate the streamflow in the test basins. HL-RDHM has been widely used for hydrologic studies (e.g., Smith et al. 2004, Moreda et al. 2006, Reed et al. 2007, Tang et al. 2007, Yilmaz et al. 2008, Wagener et al. 2009, Khakbaz et al. 2012, Smith et al. 2012a, and Smith et al. 2012b). Detailed information about this model can be found in its user manual (NWS, 2011, Koren and Barrett 1995). HL-RDHM has been expertly calibrated by NOAA NWS for the DMIP2 basins. The same parameters are used in this study.

The streamflow simulation process is performed for 2003-2010 time period where all the precipitation datasets are available. The HL-RDHM is then forced with the PERSIANN-CDR and the TMPA precipitation products, as well as the stage IV radar data, to simulate streamflow at the outlet of the basins. The United States Geological Survey (USGS) Streamflow observations are used as the reference streamflow data. The resulted streamflow simulations for SAVOY, ELMSP, and SLOA4 basins are shown in Figure 3.11, Figure 3.12, and Figure 3.13, respectively. As shown, in general in all the three DMIP 2 basins the simulated hydrographs when model is forced with PERSIANN-CDR and TMPA show good agreement. In addition, with respect to the streamflow peaks, PERSIANN-CDR shows better performance than Stage IV radar data in capturing the extreme streamflow magnitudes. For quantitative comparisons, different statistics, Bias, Correlation Coefficient, and RMSE are calculated for PERSIANN-CDR and TMPA against USGS streamflow gauge observations and shown in Table 3.1. As shown, stage IV outperforms the other precipitation datasets. The performance of PERSIANN-CDR and TMPA are very close with higher correlation coefficients and better RMSEs for TMPA and lower Biases for PERSIANN-CDR. The lower Bias in PERSIANN-CDR shows the effectiveness
Figure 3.11. Simulated and observed streamflow hydrographs at the outlet of SAVOY basins.
Figure 3.12. Simulated and observed streamflow hydrographs at the outlet of ELMSP basins.
Figure 3.13. Simulated and observed streamflow hydrographs at the outlet of SLOA4 basins.
Table 3.1. Bias, Correlation Coefficient, and RMSE statistics for simulated streamflow (2003-2010) from stage IV radar data, PERSIANN-CDR, and TMPA against USGS streamflow gauge observations.

<table>
<thead>
<tr>
<th></th>
<th>Statistics</th>
<th>Stage IV</th>
<th>PERSIANN-CDR</th>
<th>TMPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAVOY</td>
<td>Bias (%)</td>
<td>-13.03</td>
<td>14.59</td>
<td>21.19</td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td>0.8079</td>
<td>0.6861</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>RMSE (m³/s)</td>
<td>10.3176</td>
<td>11.4826</td>
<td>10.247</td>
</tr>
<tr>
<td>ELMSP</td>
<td>Bias (%)</td>
<td>-41.05</td>
<td>-10.08</td>
<td>-16.84</td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td>0.7206</td>
<td>0.6408</td>
<td>0.6217</td>
</tr>
<tr>
<td></td>
<td>RMSE (m³/s)</td>
<td>5.8556</td>
<td>6.3745</td>
<td>7.1079</td>
</tr>
<tr>
<td>SLOA4</td>
<td>Bias (%)</td>
<td>-26.34</td>
<td>9.91</td>
<td>8.08</td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficient</td>
<td>0.7334</td>
<td>0.6768</td>
<td>0.7239</td>
</tr>
<tr>
<td></td>
<td>RMSE (m³/s)</td>
<td>23.6014</td>
<td>24.4071</td>
<td>23.837</td>
</tr>
</tbody>
</table>

of the bias-adjustment algorithm in this product.

The results of this case study reveal strong potentials for PERSIANN-CDR rainfall estimations to be used for historical rainfall-runoff modeling at different basins around the world, specifically at remote regions and ungauged basins. HL-RDHM is forced with 1983-2012 daily PERSIANN-CDR rainfall estimation to produce historical streamflow at the three study basins. The resulted hydrographs for SAVOY, ELMSP, and SLOA4 are shown in Figure 3.14, Figure 3.15, and Figure 3.16, respectively. As shown, before around year 1996 streamflow observations are not available in the three study basins. PERSIANN-CDR, however, provide a complete historical daily streamflow record from 1983 to 2012.
Figure 3.14. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for SAVOY basin.
Figure 3.15. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for ELMS Basin.
Figure 3.16. Long-term (1983-2012) historical simulated streamflow from PERSIANN-CDR daily precipitation data (blue) versus USGS streamflow observations (black) for SLOA4 basin.
The scatterplots and different quantitative statistics such as Correlation Coefficient, RMSE, and Bias are shown in Figure 3.17.

Figure 3.17. Scatterplots and Correlation Coefficient, RMSE, and BIAS statistics of PERSIANN-CDR simulated streamflow against USGS observations.
The results depict high correlation (~0.7 in the three basins) and relatively low bias (~5-10%) between PERSIANN-CDR simulated daily streamflow and USGS daily observations. The PERSIANN-CDR simulated streamflow seems to overestimate for low flow conditions. Table 3.2 summarizes the resulted statistics.

Table 3.2. Bias, Correlation Coefficient, and RMSE statistics for simulated streamflow from PERSIANN-CDR against USGS observed streamflow for 1983-2012.

<table>
<thead>
<tr>
<th></th>
<th>Bias (%)</th>
<th>Correlation Coefficient</th>
<th>RMSE (m$^3$/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAVOY</td>
<td>12.2</td>
<td>0.6752</td>
<td>13.4</td>
</tr>
<tr>
<td>ELMSP</td>
<td>-10.9</td>
<td>0.734</td>
<td>6.5</td>
</tr>
<tr>
<td>SLOA4</td>
<td>5.26</td>
<td>0.734</td>
<td>29.4</td>
</tr>
</tbody>
</table>

3.3. Chapter Summary and Conclusions

In this chapter, six different verification studies over different parts of the world were carried out for testing the accuracy and efficiency of the PERSIANN-CDR daily rainfall product in different applications. In the first verification study, performance of PERSIANN-CDR was tested during Hurricane Katrina (day 29 August 2005). The results show that PERSIANN-CDR depicts good agreement with the Stage IV radar data, noting that PERSIANN-CDR has more complete spatial coverage than the radar data. In the second study, comparison of PERSIANN-CDR against gridded gauge observations during the 1986 Sydney flood in Australia reaffirms the capability of PERSIANN-CDR to provide reasonably accurate rainfall estimates. In the third
study, the PDF of the PERSIANN-CDR daily precipitation over the CONUS is compared with
the PDFs of the CPC gridded gauge data and the TMPA satellite-based precipitation product.
PERSIANN-CDR exhibits good agreements with other products. In the fourth study, using the
CPC U.S. Unified gridded observational data, the performance of PERSIANN-CDR in
reproducing the annual average count of days when rainfall ≥ 10 mm (R10mm), and rainfall ≥ 20
mm (R20mm) over the U.S. during 1983-2011 is tested. The results show that PERSIANN-CDR
reasonably reproduces the same patterns as the observational data. PERSIANN-CDR tends to
underestimate these patterns over the Sierra Nevada Mountains, most likely due to the nature of
precipitation in this region, being snow or orographic rain which is hard for satellite products to
hit and capture.

In the fifth study, very extreme precipitation events defined as the 99th (RR99p) and 95th
(RR95p) top percentiles of the rainfall distributions over China are studied. The 1983-2006
period is chosen. The results show that in general PERSIANN-CDR captures the spatial
distribution of RR99p and RR95p similar to what the EA observational data shows, particularly
in the Eastern China monsoon region where the intensity and frequency of heavy rainfall events
are very high. PERSIANN-CDR depicts similar results as the ground-based EA data with the
increasing patterns in both indices from North to South and from East to West. The disagreement
between PERSIANN-CDR and the EA data set is relatively obvious in dry regions such as the
Tibetan Plateau in the West and the Taklamakan Desert in the Northwest. An important factor
that may have influenced the results is that the ground-based stations from which EA gridded
data were produced are very sparse. On the contrary, in the station-rich regions in Eastern China,
the performance of PERSIANN-CDR is significant. PERSIANN-CDR slightly underestimates the values of extreme heavy precipitation.

In the sixth study, daily rainfall data from PERSIANN-CDR are used in the HL-RDHM rainfall-runoff model to simulate streamflow in three DMIP2 basins. The study was conducted in two phases. In phase 1 (2003-2010 period), where Stage IV radar data, PERSIANN-CDR, and TMPA data are all available, the simulated hydrographs from HL-RDHM model when forced with PERSIANN-CDR and TMPA show good agreement when compared with USGS streamflow observations. In general, the lower bias for PERSIANN-CDR shows the effectiveness of the bias removal process when producing this product. The results show strong potentials for PERSIANN-CDR rainfall estimations to be used for historical rainfall-runoff modeling. In the second phase of the study, PERSIANN-CDR daily rainfall data for 1983-2012 are used as the input to the HL-RDHM to simulate historical streamflow for this period. The resulted high correlation (~0.7 in the three basins) and relatively low bias (~5-10%) between PERSIANN-CDR simulated daily streamflow and USGS daily observations suggest usefulness of PERSIANN-CDR in reconstruction of historical streamflow at the test basins. This is an important outcome particularly for hydrological modeling in remote regions and ungauged basins.

The main goal of this chapter was to provide deeper insights into the accuracy and efficacy of the newly developed product, PERSIANN-CDR, by testing its performance in different case studies over different parts of the world. In an ideal situation, one hopes to provide a “perfect” and error free dataset to the scientific and user communities. While this is an
admirable goal to strive towards, achieving it is a difficult one. This is particularly true in the case of satellite-based products. Not only do the algorithms need continual enhancement and re-calibration, but the raw data to be utilized as inputs to the algorithms needs to be improved and expanded (e.g., including more relevant spectral bands from LEO and GEO satellites) and be periodically re-evaluated and quality controlled. Therefore, as is the case for any new environmental dataset, improving the efficacy and accuracy of PERSIANN-CDR will continue to be a work-in-progress. In this regard, while we have presented reasonably good results based on the presented testing, more evaluation and verification over different regions and seasons are desirable. The feedback from the user and scientific communities is of immense value for further improvements to the product.
Chapter 4: Statistical Modeling of Extreme Precipitation Trends

4.1. Introduction

As explained in Chapter 1, climate change and global warming can shift the odds for extreme weather events. As shown schematically in Figure 4.1, a small increase in the mean of the frequency distribution can lead to very large changes and shifts in the tail of the distribution where extreme events reside.

Figure 4.1. A schematic showing how a small shift in the mean of the distribution may lead to large changes in the tail of the distribution (Solomon et al. 2007)

Investigating the observed and projected trends and changes in catastrophic climate and weather events has been a growing research area in the recent past. Different studies have been

Change Assessment Program (NARCCAP) regional climate models. The results show that the mean and extreme precipitation exhibits decreases in the West and the Southwest U.S. Lee et al. (2014) developed statistical models with extreme value and changepoint features to estimate trends in monthly extreme temperature events in the 48 conterminous United States. The results showed that while monthly maximum temperatures are not significantly changing, monthly minimum temperatures depict significant warming.

4.2. Problem Statement

In this study, the goal is to assess potential changes in the Probability Distribution Function (PDF) of the extreme precipitation events in the U.S. The main goal is to model the observed behavior of extreme precipitation events and investigate whether there has been a statistically significant trend in the PDF of heavy precipitation over time. In addition, since precipitation and specifically extreme precipitation is one of the most difficult parameters for a reanalysis to produce, I will also investigate how two widely-used regional and global reanalysis products perform in reproducing such trends. Of specific interest is investigating the performance of NASA’s Modern-Era Retrospective-analysis for Research and Applications (MERRA) in reproducing the changes and trends in extreme precipitation events over time. The temporal changes, if any, in the parameters of a time dependent tail-focused PDF fitted to the observed and the reanalyses extreme precipitation events is investigated. The study considers the most recent three decades where reanalysis products are available. A global and a regional reanalysis product with long-term records (more than 30 years) and high temporal resolution are
used in this study. The Climate Prediction Center (CPC) gridded data is also included to evaluate
the performances of such reanalysis.

This Chapter is organized as follows: Section 4.3 presents a brief description of the data
that is used. Section 4.4 provides detailed explanations about the methodology of investigating
possible changes in the distribution of extreme precipitation events. Section 4.5 represents the
results. Section 4.6 provides detailed discussion on the results. Section 4.7 summarizes the key
findings of this research, along with future research plans.

4.3. Data

4.3.1. The Climate Prediction Center (CPC) US Unified Precipitation Product

The Climate Prediction Center (CPC) US Unified Precipitation data product spans from
1948 to recent years and provides precipitation data at daily scale over the Contiguous United
States (CONUS) 0.25° x 0.25° scale (Higgins et al. 2000a). Three data sources including 1)
NOAA’s National Climatic Data Center (NCDC), 2) River Forecast Centers, and 3) daily
accumulation from hourly precipitation are used in generating CPC precipitation. The data
product covers the 140W – 60W 20N – 60N region. In this study, CPC gridded data is
considered as the reference for comparing the performances of MERRA and NARR in modeling
extreme precipitation events.
4.3.2. The Modern-Era Retrospective-analysis for Research and Applications (MERRA) Precipitation Product

NASA’s Modern-Era Retrospective-analysis for Research and Applications (MERRA) product is designed to support NASA’s earth science research interests by producing a global long-term dataset for the satellite era from 1979 to present (Rienecker et al. 2011). Scientists at the Global Modeling and Assimilation Office (GMAO) at the Goddard Space Flight Center (GSFC) use the Goddard Earth Observing System Model - Version 5 (GEOS-5) and data assimilation techniques to generate the MERRA product at a spatial resolution of 1/2° latitude by 2/3° longitude with 72 model vertical levels (Rienecker et al. 2007, Suarez et al. 2011). By the use of an incremental analysis update which minimizes the spin down effects of the water vapor analysis, and also by providing an extensive number of variables at a relatively high spatial resolution, MERRA has shown improvements in representing large-scale global precipitation, particularly in the tropical regions (Bosilovich et al. 2011, Bloom et al. 1996). However, continental scale precipitation remains a challenge for all global reanalyses, including MERRA, when compared to the Global Precipitation Climatology Project (GPCP). Recently, Bosilovich (2013) analyzed summer seasonal precipitation in recent reanalyses showing that MERRA was not able to produce highs and lows in the summer seasonal time series, especially in the Midwestern US.
4.3.3. **The North American Regional Reanalysis (NARR) Precipitation Product**

The National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) product combines model and assimilation techniques to produce high resolution (~0.3° or 32 km at the lowest altitude) data over the North American regions (Mesinger et al. 2006). NARR benefits from using the very high resolution NCEP model together with the Regional Data Assimilation System (RDAS) for assimilating different measurements. Different sources of observed precipitation data are assimilated, including the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP—Xie and Arkin 1997, Xie et al. 2003, Xie et al. 2007), the CPC CONUS Precipitation (Higgins et al. 2000a), and daily gauge-based measurements in Canada and Mexico are used. NARR precipitation data are available in 3-hourly time steps for the period of 1979 to near present. Accumulated total precipitation in daily scale is used in this study. For consistency, the three daily precipitation data products are re-gridded to 0.5° x 0.5° grid boxes for the period spanning 1979-2010.

4.4. **Methodology**

4.4.1. **Extreme Value Theory (EVT)**

Extreme Value Theory (EVT) is a statistical theory which deals with maximum and minimum (i.e., extreme values) of a sample of independent and identically distributed (iid) random variables, as the sample size increases. EVT is used for modeling the events with very
low probability of occurrence. Practical explanations and documentations about EVT are presented in Coles (2001) and Castillo et al. (2004). In general, three different types of extreme value distributions exist. Types I, II, and III are, respectively, named after Gumbel (1891-1966), Frechet (1878-1973), and Weibull (1887-1979), the three scientists who worked with extreme value distributions for years. The Fisher-Tippett theorem (1928) is the core of the EVT. Based on this theory, the asymptotic distribution of the extreme values belongs to one of the three EVT distributions, regardless of the underlying distribution of the observed random variables. The role of the family of extreme value distributions as the limiting distributions of a sample’s extremes in the Fisher-Tippett theorem is similar to the role of the normal distribution as the limiting distributions for a sample’s sums or averages in the Central Limit theorem.

Our extreme daily precipitation study uses the two most well-adopted extreme value analysis approaches. In the first approach, known as the “Block Maxima” approach, the maximum daily precipitation of each year in 1979-2010 in CPC, NARR and MERRA is considered as the extreme precipitation event of any given year. In the second approach, known as the “Exceedance-Above-Threshold (EAT)” or “Peak-Over-Threshold (POT)” method, a certain threshold is set and all the rainfall values greater than or equal to that threshold are included in the modeling process. Details of each approach are further illustrated in the following subsections. MATLAB (2011) and R (R Core Team 2013) environments are used for computer coding and building the statistical models.
4.4.2. Generalized Extreme Value (GEV) Distribution

In this method, a “block” is defined as one year and the maximum daily precipitation in each year is considered as the block maxima. This variable is hereafter called Annual Maximum Daily Precipitation (AMDP). For each 0.5° pixel of the CONUS, I assume that the sequence of AMDP \( \{X_t\} \), extracted from each of the CPC, NARR and MERRA datasets, is independent in time. As the block size (365, or 366 days in the case of leap years) is large enough, according to the Fisher-Tippett theorem (1928) the marginal cumulative distribution function for AMDP \( \{X\} \) can be well approximated by the Generalized Extreme Value (GEV) distribution:

\[
G(x; \mu, \sigma, \xi) = P(X \leq x) = \exp\left\{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}
\]

For \( z \) with \( 1 + \xi \left(\frac{x - \mu}{\sigma}\right) > 0 \) where \( \mu, \sigma \) and \( \xi \) are unknown location, scale, and shape parameters, respectively. If \( \xi < 1 \), the mean (expected) value of \( X \) is

\[
E(X) = \mu + \frac{\sigma}{\xi} \left[\Gamma(1 - \xi) - 1\right]
\]

To examine possible changes in the distribution of the AMDP events over time, the location parameter in Eq. (4.1) and (4.2) is parameterized as a linearly time-variant parameter inside the GEV distribution function, relaxing the requirement that the PDF be stationary in time:

\[
\mu_t = \beta_0 + \beta_1 t
\]
where $\beta_0$ and $\beta_1$ are unknown location and slope parameters (Coles 2001). $\beta_1$ is interpreted as the linear trend in extreme precipitations as the expected change in $X_t$ from time $t$ to $t+1$ is

$$E(X_{t+1}) - E(X_t) = \mu_{t+1} - \mu_t = \beta_1$$

(4.4)

The Maximum Likelihood (ML) method is used to find the estimates of the unknown parameters in the GEV distribution function. For this, I numerically find the values of the parameters that maximize the GEV likelihood function:

$$L(\beta_0, \beta_1, \sigma, \xi; x_1, ..., x_n) = \prod_t f(x_t; \beta_0, \beta_1, \sigma, \xi)$$

(4.5)

where $f(x; \beta_0, \beta_1, \sigma, \xi) = \frac{d}{dx}G(x; \beta_0, \beta_1, \sigma, \xi)$ is the probability density function of the GEV distribution. The standard errors for the ML estimates are obtained from the information matrix of the likelihood fit and will be used for significance tests for the estimated linear trends in the AMDP events. These methods are applied to each pixel of the CONUS for the CPC, MERRA and NARR datasets. In this study, both the time-invariant and the time-variant GEV models are considered in modeling process.
4.4.3. Generalized Pareto (GP) Distribution

A caveat pertinent to the first approach is the possibility of wasting useful data, as the GEV distribution only considers the maximum precipitation of each year. To overcome this drawback, at each pixel the top 1% heavy rainfall of all daily precipitation (rain ≥ 1 mm/day) data for 1979-2010 is selected as the threshold in our EAT method. The points above that threshold are considered for extreme precipitation modeling. This variable is called Annual Extreme Daily Precipitation (AEDP). I also applied a de-clustered scheme to the AEDP variables by clustering the peaks which belong to the same cluster and only choosing the maximum peak as the representative of the cluster, so that the Poisson assumption is better maintained (Coles 2001). In EVT, the Generalized Pareto (GP) distribution is considered as an appropriate limiting probability distribution to model these threshold de-clustered exceedances. Therefore, the precipitation exceedance \( Y = X - u \) conditioned on that \( X > u \) has the following GP distribution:

\[
H(y; \tilde{\sigma}, \xi) = P(X \leq y + u | X > u) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{4.6}
\]

for \( y > 0 \) and \( 1 + \xi y / \tilde{\sigma} > 0 \). It is noteworthy that \( \tilde{\sigma} = \sigma + \xi (u - \mu) \). Like the GEV method, maximizing the likelihood function

\[
L(\tilde{\sigma}, \xi; y_1, \ldots, y_n) = \prod_t \frac{d}{dy_t} H(y_t; \tilde{\sigma}, \xi) \tag{4.7}
\]
produces the ML estimates for $\hat{\sigma}$ and $\hat{\xi}$. This GP model is fitted to every pixel of the CONUS for the CPC, MERRA and NARR datasets and the parameters of the distribution are estimated. Only time-invariant form of the GP model is considered.

### 4.5. Results

#### 4.5.1. Annual Maximum Daily Precipitation (AMDP)

Before describing the fitted extreme value distributions, it is worth exploring potential linear trends in the AMDP time series by fitting a linear trend regression model to each 0.5° x 0.5° grid box. Figure 4.2 left column shows the slope of the linear model fitted to the AMDP for CPC (top), NARR (middle) and MERRA (bottom). As shown, generally all three datasets depict positive trends in AMDP in the Eastern U.S. MERRA shows statistically significant negative trends in the central U.S. while such trends are not identified in the CPC gridded data. NARR seems to perform better than MERRA in capturing the trend patterns similar to CPC. To identify the regions where the trends in AMDP events are statistically significant, a two-sided significance test was performed at 5% significance level for each and every grid point. The results are shown in Figure 4.2 right column. As shown, AMDP exhibits an overall positive trend across the Eastern U.S. in three products, although some discrepancies exist location-wise. According to the CPC results, in the Southwest U.S. AMDP shows significant increasing trends in a few pixels. NARR and MERRA underestimate the trend in this region. The inconsistent
Figure 4.2. Trend (mm/day/year) in the Annual Maximum Daily Precipitation (AMDP) in CPC (top), NARR (middle), and MERRA (bottom) for 1979-2010 period. Right column shows the regions where the trend in AMDP is statistically significant at a 5% level.
trend in MERRA over Kansas and Nebraska is a place for question and is discussed in the “Discussion” section.

As noted in the previous section, AMDP sequences follow the GEV distribution. The time-invariant location, scale and shape parameters of the GEV distribution at each pixel for the three data products are estimated and shown in Figure 4.3. As shown, for both location and scale parameter estimates, NARR and MERRA exhibit similar patterns to CPC; however they both underestimate these parameters, MERRA more than NARR. Most of the discrepancies between the reanalysis products and the CPC gridded data are in and around the Gulf Coast states. Since the GEV location parameter relates to the mean of the time series via Eq. (4.2), it can be stated that NARR and MERRA generally underestimate the AMDP mean, specifically over the west coast, the east coast, and the Gulf Coast regions. It is difficult to make a robust statement about the accuracy of the shape parameter estimates (Figure 4.3 right column) as the discrepancies between the $\xi$ patterns in the reanalyses and CPC are large.

To test the GEV distribution for goodness of fit, we used the Kolmogorov-Smirnov (K-S) test (Massey 1951). For each rainfall product (i.e., CPC, NARR, and MERRA), the empirical and theoretical GEV Cumulative Distribution Functions (CDF) at each pixel were built and considered for this purpose. The null hypothesis that the AMDP data are from their theoretical GEV distribution is tested. The K-S test results showed that the null hypothesis was not rejected anywhere, meaning that the GEV distribution is indeed the good fit for the AMDP data.
Figure 4.3. Location (left column), scale (middle column), and shape (left column) parameters of the time-invariant GEV distribution of AMDP events from CPC (top), NARR (middle), and MERRA (bottom) for 1979-2010.
So far, parameters of the GEV distribution were considered constant and not changing over time, however due to the impacts of climate change the GEV distribution might have been changing over time. To investigate the possibility of the existence of such changes, we consider a time-variant GEV distribution model. Specifically, I assume the GEV location parameter changes over time as modeled in Eq. (4.3), implying that the climate change induces a change in the precipitation mean. The time-variant GEV model in Eq. (4.1) with $\mu$ parameterized as $\mu_t$ via Eq. (4.3) is fitted to each pixel in CPC, NARR, and MERRA and the four parameters $\beta_0, \beta_1, \sigma, \text{and } \xi$ are estimated. Figure 4.4 (left column) displays the estimated trends in the location parameter of the time-variant GEV distribution of AMDP events in the three products. All three datasets generally depict positive trends in the Eastern United States, with the exception of Southern Louisiana and Alabama and a few local and isolated locations. In contrast, the Western US has negative trends in AMDP for the three datasets, with some regional exceptions such as cities of Los Angeles and Sacramento in California. The central U.S. has substantially inconsistent patterns among the reanalyses and the CPC dataset, more so for MERRA than NARR. The South Coast of Texas shows positive trends for CPC, NARR, and MERRA. Kansas, Missouri, southeast Illinois, Southwest Indiana, and the Northern North Dakota depict the strongest substantial increasing trends in AMDP for the CPC and NARR datasets but not for the MERRA dataset. In contrast, the Central U.S. shows strong negative trend in MERRA AMDP.
Figure 4.4. Trend (mm/day/year) in the location parameter of the time-variant GEV distribution of AMDP events in CPC (top), NARR (middle) and MERRA (bottom). Right column shows the regions where the trend in location parameter is statistically significant at a 5% level.
To evaluate statistical significance for the estimated GEV trends, the z-scores of the trend estimates are calculated and tested. Z-score shows how many standard errors the estimated trend differs from zero. For a 5% significance test (two-sided), a statistically significant trend is the one whose respective z-score is beyond the 2.5th and 97.5th percentile range of the standard normal distribution (±1.96). While patterns of the trend in $\beta_1$ parameter with complete spatial coverage are shown in Figure 4.4 left column, locations where the $\beta_1$ trend is statistically significant at 5% significance level are shown in Figure 4.4 right column. As experienced before, increasing trends in AMDP are more apparent in the Eastern U.S. than in the Western U.S. As shown in the AMDP trend maps, the reanalysis datasets depict different patterns for significant trends, more so for MERRA than NARR. MERRA shows decreasing AMDP trends for the Central U.S. and more increasing trends for the South U.S., following a path from eastern Texas to Maryland.

The estimated scale parameters in this non-stationary condition for the three products are also shown in Figure 4.5 (left column). In aggregate, the West has similar patterns for the scale parameter estimates among the three datasets, with smaller values across the mountainous areas and greater values in the West coast regions, and with somewhat smaller values across the coast in the reanalyses when compared to CPC, more so in MERRA. In contrast, the East represents substantial differences for the three datasets. For CPC, the Gulf Coast States have larger values of the GEV scale parameters, spreading this pattern up to Dakotas in the north and up to New England in the east.
Figure 4.5. Estimated scale parameter (left Column) of the time-variant GEV distribution for CPC (top), NARR (middle) and MERRA (bottom). Right column shows the maps of the relative difference between the scale parameters in time-variant and time-invariant conditions.
In aggregate, this CPC pattern appears similar in NARR and MERRA but reduced in magnitude, especially for MERRA. The relative differences between the estimated scale parameters in the time-invariant and time-variant GEV distributions are also tested. As shown in Figure 4.5 right column, in most locations there is no difference between them. However, in pixels where a statistically significant trend is identified (such as east of Texas, Georgia, North Carolina, Virginia for MERRA), the scale parameter in the time-variant condition is estimated approximately 10 - 20% less than the time-invariant condition.

4.5.2. Annual Extreme Daily Precipitation (AEDP)

Unlike the AMDP data, a linear trend in the de-clustered peaks above the 99th percentile (i.e. AMDP data) is not easily practical since the time difference between these events is not a constant value over time. Instead, I looked at the number of AEDP events at each year of the record and investigated if any statistically significant trend can be detected. The results are shown in Figure 4.6. As shown in Figure 4.6 left column, the number of AEDP event increased for Eastern and Northeastern U.S., specifically for the states of Main, New York, Pennsylvania, West Virginia, Indiana, Georgia, and south of Wisconsin. Although both NARR and MERRA show an increasing trend in Eastern U.S., there exist discrepancies location-wise, more so for MERRA than NARR. MERRA depicts a spurious increasing trend in Texas, Arkansas, North Carolina, Virginia and Maryland, and a spurious negative trend in Nebraska and Kansas.
Figure 4.6. Trend (number per year) in the number of annual extreme daily precipitation (AEDP) events in CPC (top), NARR (middle), and MERRA (bottom). The right column shows the regions where trend is statistically significant at 5% significance level.
As discussed in the methodology section, the GP distribution is fitted to the AEDP sequences. The estimated scale parameters for NARR and MERRA, as shown in Figure 4.7 left column, depict similar patterns to the CPC; however, like AMDP, the two reanalysis products both underestimate the CPC scale parameter. Significant discrepancies and underestimations are identified in the Gulf Coast regions and the Western U.S. As for the shape parameter, large discrepancies in the Eastern U.S. and the Gulf Coast states are identified between the two reanalyses and the observational data. Similar to block maxima approach, a goodness of fit test was performed using the K-S test. The null hypothesis was defined as de-clustered AEDP data in the empirical and the theoretical CDFs belongs to the same distribution. Using the empirical CDF and theoretical GP CDF in the K-S test, I found out the K-S null hypothesis was not rejected anywhere, meaning that the GP distribution is indeed a good fit to the AEDP data.

4.6. Discussion

4.6.1. Empirical Probability Density

In previous sections, potential changes in the probability distribution of extreme precipitation events over time are tested. However, it is still worthwhile to look at the entire empirical PDF of the precipitation extremes over the entire record. Figure 4.8 shows the empirical probability density over CONUS for both AMDP and de-clustered AEDP for CPC, NARR, and MERRA for 1979-2010. As shown, with respect to AMDP sequences both MERRA
Figure 4.7. Estimated scale (left column) and shape (right column) parameters of GP distribution for AEDP for CPC (top), NARR (middle), and MERRA (bottom) for the period of 1979-2010.
Figure 4.8. Empirical probability density over CONUS for AMDP (red) and de-clustered AEDP (black) for CPC (solid), NARR (dashed), and MERRA (dotted) for 1979-2010.
and NARR tends to be consistent with CPC up until rain rate ~ 15 mm/day, then overestimate until rain rate ~50mm/day and after that they both underestimate, more so for MERRA than NARR. For AEDP, similar performance as to AMDP is observed except at larger rain rates.

As shown in Figure 4.8, both MERRA and NARR are biased low in the higher magnitudes of the extremes, more so for MERRA than NARR. MERRA is likely affected by its relatively coarse scale. The convective parameterization is scale dependent and therefore it cannot reproduce extreme rainfall rates to the fine point scale of the gauges used in CPC. Thinking about this in the concept of trend analysis tells us that in an upward trend MEERA will eventually stop increasing because it cannot resolve the higher rain rates and thus, its upward trends would be smaller than observational data when observations are going upward. This effect can be seen in Figure 4.8 where the empirical density functions for MERRA AMDP and MERRA AEDP become almost the same after a certain rain rate for extreme rainfall (in our case over the CONUS this is happening after 60mm/day).

### 4.6.2. Negative Trend in MERRA over NE-KS

A feature that has been repeatedly identified (in Figure 4.2, Figure 4.4, and also in Figure 4.6) in MERRA is a negative trend over Nebraska and Kansas. This is the opposite of what observations show. To further investigate the potential reasons for such a spurious trend, it is worthwhile to look at the time series of annual maximum daily precipitation for MERRA and CPC over this region. As shown in Figure 4.9 (top), while CPC shows an increasing trend
Figure 4.9. Time series, trend lines and respective statistics (top) and anomalies (bottom) of the annual maximum daily precipitation for CPC (solid) and MERRA (dashed) over Kansas and Nebraska.

(+0.125 mm/day/year), MERRA is decreasing (-0.284 mm/day/year). We used a two tailed z-test to investigate if these trends are statistically different from zero. Based on the two-sided z-test at a 5% significance level, if the p-value for the trend estimate is less than 0.05, then the trend is
statistically significant. The associated p-values for the trend for this region are shown in the legend of Figure 4.9. Based on these values, the negative trend for MERRA is statistically significant while the CPC trend is not statistically significant. It is however noteworthy that some of the interannual variability seems well represented in MERRA, as shown in the Figure 4.9 (bottom). It is also noteworthy that obvious structural discontinuities in precipitation mean do not appear in the CPC and MERRA AMDP series.

There are a number of potential reasons for the MERRA precipitation to not replicate the observations, such as boundary layer parameterization, land-atmosphere interactions and/or convective precipitation parameterization. Robertson et al. (2011) has studied the effects of the two observing system epoch changes, the Advanced Microwave Sounding Unit-A (AMSU-A) series in late 1998 and the Spatial Sensor Microwave Imager (SSM/I) in late 1987. The results show that precipitation is very sensitive to the changing observing system. In addition, Bosilovich (2013) found that range of seasonal precipitation over the Central U.S. decreases relative to the observation, considering that to be related to a deficiency in the model land atmosphere interactions. In addition, a negative correlation between the analysis increment and precipitation in the annual mean time series was identified in this region, however it is not yet completely understood if such correlation is causal or not. In another study, Trenberth et al. (2011) showed that the mean atmosphere moisture divergence (expressed as evaporation minus precipitation) extracted from MERRA data in this region (over land) is positive which is unrealistic. This triggered a study by Bosilovich et al (2014) where they evaluated the observation influence on regional water budgets in different reanalyses. The study shows that there has been an effect of the changing observations on the Central US water cycle. A satellite
instrument on the Advanced Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (ATOVS) changed the water vapor analysis at 06Z and 18Z when no radiosondes are available to anchor the analysis. This has apparently affected the water vapor analysis, and consequently the local water cycle in the Central US, more so than in other regions.

4.6.3. Seasonality in Extremes

Another interesting result from MERRA is the increasing trend in the number of AEDP events over the Gulf Coast states (e.g., Texas, Louisiana, and Arkansas) and along the East Coast of the U.S (as is clear in Figure 4.6). Similar trends are also clear in the trend map of the time-variant GEV distribution for MERRA (Figure 4.4). Again, CPC does not show such a pattern. We found that these two tracks seem very similar to the tracks of the East Coast and the Gulf Coast cyclones that hit East Southeast U.S. (Rauber et al. 2002), suggesting a bias in the model land atmosphere interactions toward these cyclones specially for extreme rainfalls.

To further investigate the above issue, we studied the seasonality of the extreme precipitation events. The following different seasons are defined: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON). In addition, since the MERRA trajectories in Figure 4.6 show high correlation with the Gulf Coast and East Coast cyclones, we also looked at the Hurricane Season (hereafter called HUR). Based on the definition of the National Hurricane Center (NHC), HUR starts from the beginning of June and ends at the end of November. The time series of seasonal maximum daily
rainfall for each of the above five seasons are constructed and statistically significant trends in seasonal maxima are investigated. Figure 4.10 and Figure 4.11 shows the significant trends (mm/day/year) at 95% confidence interval in seasonal (DJF, MAM, JJA, SON, and HUR) maximum daily precipitation in CPC, NARR, and MERRA during 1979-2010. As shown, in general all three datasets show an increasing trend in seasonal extreme rainfall in all the seasons in the Eastern part of the country, with the largest increasing trends to happen in SON, as well as the HUR season. In the Southwest U.S., a significant decreasing trend in extreme rainfall was identified in MAM and JJA seasons. These results seem consistent with the trend results in seasonal maximum 5-day precipitation totals presented by Alexander et al. (2006).

With respect to the spurious negative trend in the Central U.S. in MERRA, as is shown in Figure 4.10, MAM season is the main season in which a major portion of this trend occurs. The HUR, JJA and SON seasons also partly show the negative trend in this region. In addition to that, MERRA shows an increasing trend in the maximum DJF precipitation, whereas CPC shows the opposite. The increasing seasonal trend in DJF help reinforce the increasing annual trend in the MERRA AMDP, but in the CPC, the decreasing DJF trend works against the SON, making the CPC AMDP trend weaker. For the years where there are few tropical cyclones (TCs) over the land, then the DJF trends contributes as a much stronger signal to the annual trend, especially in weak spots in MERRA.
Figure 4.10. Trend (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and HUR (bottom row) for CPC (left column), NARR (middle column) and MERRA (right column) during 1979-2010.
Figure 4.11. Trend (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for CPC (left column), NARR (middle column) and MERRA (right column) during 1979-2010.
Scatterplots, correlation coefficients and the Root Mean Square Errors (RMSEs) of the seasonal trend patterns in MERRA and NARR against the CPC seasonal trend patterns are also calculated. Figure 4.12 and Figure 4.13 show the scatterplots and the statistics for different seasons for all the pixels with statistically significant trend in either observation or reanalysis product. As shown, there is not a good agreement between the reanalyses and the observational data. There are many pixels where either reanalysis or observation depicts no trend while the other data product shows a significant trend. It is worth to also consider only the pixels where the trend is significant in both the reanalysis and the observational product. As shown in Figure 4.14 and Figure 4.15, the correlation coefficient has increased considerably. NARR shows better performance than MERRA. It can be concluded that one of the main issues here is not the overestimation or underestimation of the seasonal trend by reanalyses, but more importantly lack of good skills in both NARR and MERRA in identifying the correct trend location-wise.
Figure 4.12. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and JJA (bottom row) for NARR (left column) and MERRA (right Column) against CPC gridded observation during 1979-2010. All the pixels at which trends are significant in either reanalyses or CPC are considered.
Figure 4.13. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for NARR (left column) and MERRA (right column) against CPC gridded observation during 1979-2010. All the pixels at which trends are significant in either reanalyses or CPC are considered.
Figure 4.14. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in DJF (top row), MAM (middle row) and JJA (bottom row) for NARR (left column) and MERRA (right Column) against CPC gridded observation during 1979-2010. Only pixels at which trends are significant in both reanalyses and CPC are considered.
Figure 4.15. Correlation coefficient, RMSE and scatterplots of the trends (mm/day/year) in seasonal maximum daily precipitation in SON (top row) and HUR (bottom row) for NARR (left column) and MERRA (right Column) against CPC gridded observation during 1979-2010. Only pixels at which trends are significant in both reanalyses and CPC are considered.
4.7. Chapter Summary and Conclusions

In this study, we developed and applied a non-stationary statistical model based on Extreme Value Theory (EVT) for investigating the observed trends and changes in the patterns of extreme precipitation events over time. The main goal was to model the observed trends in extreme precipitation events in the past three decades and investigate if there have been statistically significant trends in the PDFs of heavy precipitation events over time due to climate change. For this purpose, the location parameter of the GEV distribution is parameterized as a time-variant model. Therefore, we will have a time-variant probability distribution whose parameter can change over time. This enables us to test the potential impacts of climate change on the probability distribution of extreme events.

Being a data-rich region, the United States is chosen as the study region and the CPC gridded gauge data product is used for trend detection and also validation purposes. Two widely-used reanalysis products, one being a regional reanalysis (NARR) and the other being a global reanalysis product (MERRA) are used. We defined extremes as 1) Annual Maximum Daily Precipitation (AMDP), and 2) Annual Extreme Daily Precipitation (AEDP, representing the top 1% heavy rainfall events during 1979-2010) events.

Based on the CPC gridded data, the Eastern and particularly the Northeastern parts of the U.S. are experiencing positive trends in the AMDP intensity, while in the Western parts of the country a mix of both negative and positive trends are identified. NARR and MERRA tend to depict similar patterns in general but discrepancies are identified location-wise. With respect to a fitted time-invariant GEV distribution, both NARR and MERRA capture the behavior of
extremes across the CONUS. However, they both tend to underestimate the location and scale parameters of the GEV distribution, particularly over the Gulf Coast region with the discrepancy greater in MERRA than in NARR. A fitted time-variant GEV distribution of the CPC observations shows that Eastern and Western part of the U.S. exhibit positive and negative trends, respectively, in the trends of precipitation extremes. Like with the raw AMDP trends, similar patterns are seen in MERRA and NARR but not always at the pixels that CPC shows. Similar to the results from the time-invariant GEV distribution, in the time-variant case both NARR and MERRA underestimate the scale parameters of the time-variant GEV distribution.

With respect to the peaks above 99\textsuperscript{th} percentile (i.e. AEDP) sequences, similar underestimation in the scale parameter of the GP distribution of reanalyses is identified. Looking at the trend in the number of AEDP events over time, we find that the North, the Northeast, and the Central U.S. have experienced increasing trends from 1979 to 2010. This trend is negative for the West, the Southwest, and the Northwest U.S.

Looking at the empirical probability density function (PDF) of three products over CONUS and for the entire period, we find that for dry regions where the values for both AMDP and AEDP are small (in our case in orders of 10-20 mm/day), the PDFs of reanalyses and CPC matches pretty well. However, for wet regions where the AMDP and AEDP rainfall values are high, both NARR and MERRA are biased low when compared to CPC. MERRA is likely affected by its relatively coarse scale, preventing it from reproducing extreme rainfall rates to the fine point scale of the gauges used in CPC.
Our analysis also shows that MERRA depicts a spurious negative trend in the Central U.S., particularly over Nebraska and Kansas. We discuss that this artifact is due to the changes in the observation system in and around the Central U.S which has apparently affected the water cycle in this region. Time series of extreme rainfall over this region implicates that while the negative trend in MERRA is statistically significant; some of the interannual variability is well represented in MERRA.

Unlike CPC, MERRA shows two tracks of increasing trend, one over the Gulf Coast region and the other along the East Coast. These tracks seem to be consistent with the paths that the East Coast and the Gulf Coast tropical cyclones take when striking the U.S. MERRA’s precipitation product seems to be biased toward these cyclones. Looking at the trends in the seasonal maximum precipitation events reveal that the increasing annual trend simulated by MERRA in the Gulf Coast region is due to a trend of an incorrect sign in winter precipitation extremes. In addition, a major portion of the negative trends in the central U.S. occurs in the spring season, although other seasons also contribute to the negative trend. Detailed investigation into the GEOS-5 model and the assimilation techniques used in MERRA product is required to enable the community to make more detailed inference about the reasons for such pattern.

This study highlights some areas for which reanalyses products need improvements, specifically with respect to precipitation extremes. As the results showed, NARR generally outperformed MERRA; however a very important point that should be considered here is that NARR benefits from gauge adjustments and data assimilation of precipitation observations, while MERRA does not have that correction component. This is a very important point to
consider while comparing the performance of NARR and MERRA. For precipitation studies over CONUS or North America, NARR proves to be a better product than MERRA. However, for international readership interested in studying other parts of the world other than North America, or even interested in performing global studies, MERRA would be the choice rather than NARR.

Future plans are to expand the study to include other reanalyses such as MERRA-land, the European Center for Medium range Weather Forecasting (ECMWF) Interim reanalyses (ERA-Interim), the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR), and the Japanese 55-year Reanalysis (JRA-55). Moreover, follow-on studies will examine seasonal extreme precipitation of the recent past with time varying GEV and GP parameters for the individual seasons. In addition, with the availability of a newly developed satellite-based precipitation product, PERSIANN-CDR (Ashouri et al. 2014), there will be other opportunities to further test the ability of reanalyses and model simulations in capturing the behavior of extreme precipitation events.
Chapter 5: Conclusions and Future Directions

5.1. Summary of Findings

The main thrust behind this dissertation is to improve our understanding of long-term variations and changes in precipitation patterns and distributions in a changing climate using satellite imagery and remote sensing. In the very preliminary stages of the work, it was understood that “data” and specifically “observational data” is a major challenging issue confronting the goals of this study. It is only after having reliable observational data that improving our modeling approaches would lead in robust results. Therefore, in this dissertation, two major aspects, namely; 1) “data” and 2) “modeling” components are addressed. In the first stage, in order to overcome the data limitations, a retrospective satellite-based precipitation climate data record for long-term hydrological and climate studies is developed. The product, called PERSIANN-CDR, provides the capability of studying water cycle at a higher spatiotemporal resolution than previously possible. Different verification studies over different parts of the world are performed to test the accuracy and usefulness of the daily and 0.25° PERSIANN-CDR rainfall estimations. The results are promising.

In the “modeling” stage, the focus is on developing suitable statistical schemes for modeling the stochastic behavior of extreme precipitation events within the past three decades of changes in the climate system. In specific, the focus is on catastrophic and hazardous extreme
precipitation events which by definition are rare events and for which very limited observational data of sufficient resolution (in time and space) and length (more than 30 years) has been available until now. The main interest here is to understand how climate change has altered the characteristics of extreme precipitation events (such as frequency, intensity, duration etc.) over the course of time. Rather than looking at the full frequency distribution of precipitation, the efforts are focused on studying the tail of the distribution where extreme precipitation events reside and a small shift in the mean of the frequency distribution can lead to very large changes in this part of the distribution. Following is a summary of the key findings addressed in this dissertation.

**PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies**

In this dissertation, by developing the PERSIANN-CDR rainfall estimation algorithm and applying it to the historical archive of the GridSat-B1 infrared data from multi-satellite observations, available through the International Satellite Cloud Climatology Project (ISCCP), a new retrospective satellite-based precipitation dataset is constructed as a climate data record. Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR) provides more than 30 years of daily and 0.25° rainfall estimates. PERSIANN-CDR rainfall estimations are adjusted based the Global Precipitation Climatology Project (GPCP) monthly product to maintain consistency of the two datasets at 2.5° monthly scale throughout the entire record. The spatial and temporal coverages of PERSIANN-
CDR are, respectively, the latitude band 60°S–60°N and the longitude band 0°–360°, and from 01/01/1983 to the early 2014 for which GPCP monthly product and quality-controlled GridSat-B1 data were available.

PERSIANN-CDR is part of the U.S. National Climate Data Record Program and is publicly available as an operational CDR at the website of the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) via http://www.ncdc.noaa.gov/cdr/operationalcdrs.html. PERSIANN-CDR rainfall data is updated on a quarterly basis with about three month lag time.

Testing and Verification Results

In order to test the accuracy and efficacy of PERSIANN-CDR, six different verification studies with different objectives and over different parts of the world are performed. The first two studies were focused on two extreme rainfall events, Hurricane Katrina (2005) in the U.S and Sydney flood (1986) in Australia. The evaluation is based on comparing the PERSIANN-CDR daily rainfall estimations with ground-truth observations.

In the case of Hurricane Katrina, with the availability of Stage IV gauge-adjusted radar data, a quantitative comparison is performed on day 29, August 2005 where Katrina produced significant amounts of rainfall over the Southeast U.S. The results show good agreement between PERSIANN-CDR and Stage IV radar data, noting that PERSIANN-CDR has more
complete spatial coverage than the radar data. In addition, PERSIANN-CDR shows a higher correlation coefficient than TMPA when compared to Stage IV radar data. However; the bias in TMPA is lower than that in PERSIANN-CDR. These results show that each of the products have their respective strengths and shortcomings.

In the second study, the 1986 Sydney flood in Australia was chosen. The goal was to test the performance of PERSIANN-CDR for 1) an extreme event outside the U.S. and 2) for a time period for which other currently available high-resolution microwave satellite-based precipitation products were not available (i.e. prior to ~1997 where PMW observations are very limited and even not available). The interpolated daily rainfall gauge data from the Australian Bureau of Meteorology available at 0.05° spatial resolution was downloaded, processed, and treated as the in-situ observation for comparison purpose. While PERSIANN-CDR shows a relatively good correlation coefficient (0.62) with gridded gauge data, we note that there are spatiotemporal differences between the two products. Needless to say that any satellite-based precipitation observation estimates, including PERSIANN-CDR, have their own inherent inaccuracies at this time in their evolution, but ground-based observations have their own shortcomings too. In this particular example, there are two aspects of the in-situ measurements contributing to the uncertainties. Firstly, there are temporal differences between the two datasets due to the way gauge data is processed. The PERSIANN-CDR daily grids correspond to a given 00:00-23:59 UTC time period while the Australian interpolated gauge data represents the 24-hour accumulation of observations taken at 09:00 local time. Secondly, since gauge data are point measurements, it is unlikely that the interpolated in-situ estimates represent the spatial
coverage properly and hence could result in some significant differences between the satellite and gridded gauge observations.

In the third study, the probability density functions (PDF) of PERSIANN-CDR over CONUS during 1998-2008 was compared to those of TMPA v7 and CPC U.S. Unified gridded observations. PERSIANN-CDR exhibits good agreements with other products.

In the fourth and fifth studies, the performances of PERSIANN-CDR in reproducing different extreme precipitation indices 1) over the U.S. and 2) over China are tested against gauge-based observations. For the U.S. case study, using the CPC U.S. Unified gridded observational data, the performance of PERSIANN-CDR in reproducing the annual average count of days when rainfall $\geq 10$ mm (R10mm), and rainfall $\geq 20$ mm (R20mm) during 1983-2011 is tested. The results show that PERSIANN-CDR reasonably reproduces the same patterns as the observational data with the exception of underestimating over the Sierra Nevada Mountains. This underestimation could be due to 1) the nature of precipitation over mountainous regions, being snow rather than rain which PERSIANN-CDR algorithm estimates, and 2) orographic rainfalls in the mountainous areas which are generally hard to hit for satellites and radars.

For the China case study, extreme precipitation events, defined as the 99th (RR99p) and 95th (RR95p) top percentiles of the rainfall distributions for the period of 1983-2006 are studied. The East Asia (EA) gridded gauge-based dataset is chosen as the reference dataset. The results show that in the station-rich regions, where the number of interpolated gauges into EA dataset is large enough, the performance of PERSIANN-CDR is significant. PERSIANN-CDR depicts
similar results as the ground-based EA data with the increasing patterns in both indices from North to South and from East to West. In regions such as Tibetan Plateau in the West and the Taklamakan Desert in the Northwest, the performance of PERSIANN-CDR degrades. An important factor that may have influenced the results here is perhaps the fact that the ground-based stations from which EA gridded data were produced are very sparse in these regions.

In the sixth and final verification study, the National Weather Service (NWS) Hydrology Laboratory-Research Distributed Hydrologic Model (HL-RDHM) is forced with the PERSIANN-CDR, the TMPA v7, and the Stage IV radar rainfall data to simulate streamflow in three DMIP2 basins. The first phase of the study is conducted for the period of 2003-2010 where all the rainfall products are available. The hydrologic model produces very similar hydrographs for both PERSIANN-CDR and TMPA as compared to USGS gauge data. Relatively speaking, TMPA exhibits a higher correlation coefficient and a lower RMSE, while PERSIANN-CDR exhibits a lower bias suggesting the effectiveness of the bias-adjustment algorithm in this product. These results reveal strong potential for PERSIANN-CDR rainfall estimations to be used for long-term historical rainfall-runoff modeling at the study basins. In this regard, HL-RDHM is forced with 1983-2012 daily PERSIANN-CDR rainfall estimation to produce historical streamflow at the study basins. The results depict high correlation (~0.7 in the three basins) and relatively low bias (~5-10%) between PERSIANN-CDR simulated daily streamflow and USGS daily observations, suggesting the usefulness of PERSIANN-CDR rainfall data in rainfall-runoff hydrological modeling.
Statistical modeling of extreme precipitation trends in a changing climate

Categorized in the “modeling” stage of this dissertation, efforts were focused to develop and apply stationary and non-stationary statistical models based on Extreme Value Theory (EVT) to investigate the observed trends and changes in the patterns of extreme precipitation events over time due to climate change. The main goal is to model the observed behavior of extreme precipitation events and investigate whether there has been a statistically significant trend in the PDF of heavy precipitation over time. Testing the potential changes in the PDF of heavy precipitation is equivalent to testing the potential changes in the parameters of the distribution. In this study the location parameter of the GEV distribution is considered as a function of time. Therefore, we will have a time-variant probability distribution whose parameter can change over time. This enables us to test the potential impacts of climate change on the probability distribution of extreme events. Statistically speaking, any significant trend in the location parameter of the GEV distribution can be interpreted as the trend in the GEV. The CPC U.S. Unified gridded gauge data product was used in this study. In addition, to extend the study, two widely-used reanalysis products, one a regional reanalysis (NARR) and the other a global reanalysis product (MERRA) are chosen to test the effectiveness of the developed methodology. In this study, extremes are defined as 1) Annual Maximum Daily Precipitation (AMDP), and 2) Annual Extreme Daily Precipitation (AEDP, representing the top 1% heavy rainfall events during 1979-2010) events.

The results show that the Eastern and particularly the Northeastern parts of the U.S. are experiencing positive trends in the AMDP intensity, while in the Western parts of the country a
mix of both negative and positive trends are identified. NARR and MERRA tend to depict similar patterns in general but discrepancies are identified location-wise. With respect to a fitted time-invariant GEV distribution, though both NARR and MERRA depict similar patterns, they both tend to underestimate the location and scale parameters of the GEV distribution, particularly over the Gulf Coast region with the discrepancy greater in MERRA than in NARR. With respect to the developed time-variant GEV model, CPC observations shows that Eastern and Western part of the U.S. exhibit positive and negative trends, respectively, in the trends of precipitation extremes. Like the time-invariant GEV results, both NARR and MERRA underestimate the scale parameters of the time-variant GEV distribution.

The analysis shows that MERRA depicts a spurious negative trend in the Central U.S., particularly over Nebraska and Kansas. In this dissertation, it has been discussed that this artifact is due to the changes in the observation system in and around the Central U.S which has apparently affected the water cycle in this region. Time series of extreme rainfall over this region implicates that while the spurious negative trend in MERRA is statistically significant; some of the interannual variability is well represented in MERRA. In addition, it is found that the MERRA’s precipitation is biased toward the East Coast and the Gulf Coast tropical cyclones which strike the U.S. in the Gulf Coast and Eastern U.S. states. The results have been brought to the attention of the MERRA developers and scientists at the NASA Global Modeling and Assimilation Office and works are in progress for further tests and for improving the quality of the product.
It is very important to point out that NARR benefits from gauge adjustments and data assimilation of precipitation observations, while MERRA does not have this correction component. This is a very important point to consider while comparing the performance of NARR and MERRA and choosing one.

5.2. Future Extensions

This dissertation focused on 1) the development of a high-resolution satellite-derived precipitation climate data record, and 2) the development of time-variant statistical schemes for studying the behavior of precipitation extremes over time with consideration of potential impacts of climate change. Needless to say that further extensions and improvements can be made and following are some of the recommended areas for future studies:

Retrieving Rainfall using Multiple Channel Imagery

The current operational version of the PERSIANN-CDR algorithm use single IR channel with a gauge adjustment process to retrieve historical rainfall estimations. Using multiple channels including, but not limited to, visible (VIS) and water vapor (WV) information can result in further improvement in the quality of the data. In daytime hours, VIS image can be included with sun angle corrections (Arkin et al., 1985; King et al., 1995; Hsu et al., 1999; Behrangi et al. 2009), while in the nighttime hours, IR and WV can be used for precipitation retrieval. VIS and
IRWV are available from GridSat archive making it feasible to extend PERSIANN-CDR precipitation retrieval from the single IR channel retrieval to multiple channels (i.e. GridSat VIS and IRWVP).

Adjustment Using GPROF PMW Rainfall Product

Effective integration of LEO PMW rainfall with more frequent samples from GEO images can improve the sampling of precipitation events in space and time. One of the future extensions of this work will be the use of most recent available PMW rainfall generated from the Goddard profiling algorithm (GPROF, Kummerow, 1993; Kummerow et al. 1996, 2001) for the adjustment of the PERSIANN-B1 rainfall estimations, before using GPCP monthly data for adjustment. GPROF includes PMW information from AMSE-E, SSM/I, SSMIS, and TMI from Aqua, DMSP, NOAA, and TRMM satellites. Compilation of all available concurrent and co-located PMW and PERSIANN-B1 precipitation samples after 1997 can serve as a very beneficial dataset. The probability matching method (PMM) can serve as a useful approach to transfer PERSIANN-CDR rainfall toward PMW rainfall distribution based on all the available co-located samples (Hong et al., 2004).

Multivariate Statistical Modeling of Precipitation Extremes
To further improve the time-variant GEV statistical model, climate related covariates from the large-scale teleconnection signals, such as El Niño Southern Oscillation (ENSO), Pacific decadal oscillation (PDO), North Atlantic Oscillation (NAO), Madden–Julian Oscillation (MJO), and/or Indian Ocean Dipole (IOD), can be included into the analysis. Moreover, not only may the location parameter of the GEV distribution be correlated with these covariates, but also the scale parameters of the GEV distribution can be conditioned on both time and covariates. However, it is extremely important to consider 1) the physics behind these phenomena and 2) the computational expenses.
REFERENCES


Anagnostou, E. N., V. Maggioni, E. I. Nikolopoulos, T. Meskele, F. Hossain, and A. Papadopoulos, 2010: Benchmarking high resolution global satellite rainfall products to


Dulière, V., Y. Zhang, E. P. Salathé, 2013: Changes in twentieth-century extreme temperature and precipitation over the western United States Based on observations and regional climate model simulations, *J. Climate*, 26, 8556–8575. doi: http://dx.doi.org/10.1175/JCLI-D-12-00818.1


(TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, 
*J. Hydrometeor.*, 8(1), 38-55.


Klein Tank A. M.G., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, climate data and monitoring, WCDMP-No 72. WMO-TD No 1500, 56pp


National Climatic Data Center (NCDC), 2012; Billion-dollar weather/climate events. [Available online: http://www.ncdc.noaa.gov/billions]


National Weather Service, Distributed Model Intercomparison Project – Phase 2 (DMIP 2) 
http://www.nws.noaa.gov/oh/hrl/dmip/2/images/dmip%20test%20basins%20web%20page_Slide2.JPG


C. Stewart, A. V. Vecchia, G. Villarini, R. S. Vose, J. Walsh, M. Wehner, D. Wolock, K.
Wolter, C. A. Woodhouse, and D. Wuebbles, 2013b: Monitoring and understanding
changes in heat waves, cold waves, floods, and droughts in the United States: State of
doi: http://dx.doi.org/10.1175/BAMS-D-12-00066.1

Rauber, R. M., J. Walsh, and D. Charlevoix, 2002: Severe and Hazardous Weather, Kendall

Rienecker, M., and Coauthors, 2007: The GEOS-5 data assimilation system—Documentation of
versions 5.0.1 and 5.1.0. NASA GSFC Tech. Rep. Series on Global Modeling and Data

Rienecker, M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S.
D. Schubert, L. Takacs, G. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva,
W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion,
Woollen, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and


130


