Title

A Framework for Assessing Error Heteroscedasticity of Satellite Estimates and Extracting Spatiotemporal Variability from Precipitation Data

Author

Liu, Hao

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

A Framework for Assessing Error Heteroscedasticity of Satellite Estimates and Extracting Spatiotemporal Variability from Precipitation Data

DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the degree of

DOCTOR OF PHILOSOPHY
in Civil Engineering

by

Hao Liu

Dissertation Committee:
Professor Soroosh Sorooshian, Chair
Professor Xiaogang Gao, co-Chair
Professor Kuo-lin Hsu
Professor Amir AghaKouchak

2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AM	Amplitude Modulated
ANN	Artificial Neural Network
ANN-GA	Atificial Neural Network-Genetic Algorithm
cdf	Cumulative Density Function
CI	Confident Interval
CL	Confidence Limit
CMORPH	CPC MORPHing Technique for rainfall estimates
CORR	Correlation Coefficient
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
ENSO	El Niño Southern Oscillation
EOF	Empirical Orthogonal Analysis
EV	Extreme Value
FM	Frequency Modulated
GOES	Geostationary Operational Environmental Satellite
GoF	Goodness of Fit
IR	Infra-Red

KS-test	Kolmogorov-Smirnov test
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MRA	Multi-Resolution Analysis
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NCDC	National Climate Data Center
NCEP	National Centers for Environmental Prediction
NEXRAD	NEXt generation RADar
NM	Nonlinear Mode
NMD	Nonlinear Mode Decomposition
NOAA	National Oceanic and Atmosphere Administration
OF	Objective Function
PC	Principal Component Analysis
PCA	Principal Component
pdf	Probability Density Function
PDO	Pacific Decadal Oscillation
PERSIANN	Precipitation Estimates from Remote Sensinig using Arificial Neural Network

PMW	Passive MicroWave
RMSE	Root Mean Square Error
RS	Remote-Sensing
RT	Real Time
SAWP	Scale-Averages Wavelet Power
SSTemp	Sea Surface Temperature
TFR	Time-Frequency Representation
TMI	Tropical rainfall measurement mission's Microwave Imager
TRMM	Tropical Rainfall Measurement Mission
UCI	University of California, Irvine
WPCA	Wavelet Principal Component Analysis
WSR-88D	Weather Survalliance Radar -1988 Dopler

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisors Prof. Soroosh Sorooshian for the continuous support of my PhD study. He inspired me with his courage, patience and enthusiasm in doing research. I cannot imagine a better place to finish my study without him.

I would never forget each of the advices from my co-advisor, "Master" Prof. Xiaogang Gao. His care, guidance and generosity filled my journey with joy, hope and warmth.

I would like to thank the rest of my committee: Prof. Amir AghaKouchak and Prof. Kuolin Hsu. Their insightful comments greatly improve the quality of this thesis.

My sincere thanks go to Dan Braithwite and Diane Hohnbaum for their administrations in my study and research. I also appreciate my parents: Ruilin Liu and Dingfen Qu, for their support in my life.

I thank Chi-Fan Shih from University Corporation for Atmospheric Research for providing NEXRAD Stage-IV datasets. I also thank those anonymous reviewers for their detailed comments on my manuscript.

I thank my fellow colleagues for making the Center of Hydrometeorology and Remote-Sensing a wonderful place: Nasrin Nasrollahi, Jingjing Li, Qing (Alex) Xia, Tsou-Chun (Joey) Jaw, Jialun Li, Yumeng Tao, Tiantian Yang, Scott Sellars, Phu Nguyen, Hamed

Ashouri, Ali Z, Wei Chu, Sepideh Sarachi, Negar Karbalaee, Andrea Thortensen. I also wish to acknowledge Linyin Cheng, Chiyuan Miao and Zengchao Hao. They continuously remind me to believe in myself and to challenge myself.

Financial support of this thesis was provided by NASA Decision Support project (Grant NA10DAR4310122), the US Army Research Office project (Grant W911NF-11-1-0422), NOAA CICS and NCDC (Prime award number NA09NES440006 and NCSU CICS sub-Award # 2009-1380-01), and NOAA CCDD (NA10OAR4310122).

CURRICULUM VITAE

Hao Liu

EDUCATION

PhD University of California, Irvine, Civil Engineering March 2015

Dissertation: "Frameworks for Assessing Error Heteroscedasticity of Satellite Precipitation and Extracting Spatiotemporal Variability fromPrecipitation Data"

Committee: Soroosh Sorooshian (chair), Xiaogang Gao (co-chair), Kuolin Hsu, Amir AghaKouchak

MS University of California, Irvine, Environmental Engineering June 2010

BS Tongji University, Water and Wastewater Enginnering June 2009

HONORS AND AWARDS

University Scholarship for Excellent Student 2006

Associated Graduate Students Travel Grant 2014

RESEARCH EXPERIENCE

University of California Irvine, Irvine, California 2010 to 2015
Research Assistant, CENTER OF HYDROMETEOROLOGY AND REMOTE SENSING

- Study the Spatiotemporal Variability of Western United States' Seasonal Precipitation
- Assessment of the Error-intensity Relationship of Remote-Sensed Satellite Precipitation Estimates

Tongji University, Shanghai, China, 2007 to 2009
Student Researcher, WATER AND WASTEWATER NETWORK MODELLING LAB

- Removing Atypical Endocrine Disrupters in Tap Water through Advanced Oxidation Processes
- Studying the Hydrodynamic Response of Rainwater Pipeline to Intense Rainfall

TEACHING EXPERIENCE

University of California Irvine, Irvine, California September 2013 to March 2014
Teaching Assitant, Civil Engineering Department
CEE 299 Matlab Programming in Engineering Applications

PUBLICATIONS

Liu, H., Gao, X. , Sorooshian, S. (2015), Assessment of the Spatial and Seasonal Variation of Error-Intensity Relationship in Satellite Precipitation using an Adaptive Parametric Model, Journal of Hydrometeorology (Under review, Major Revision)

Kong, D., C. Miao, A. G. L. Borthwick, Q. Duan, **H. Liu**, Q. Sun, A. Ye, Z. Di & W. Gong (2015) Evolution of the Yellow River Delta and its relationship with runoff and sediment load from 1983 to 2011. *Journal of Hydrology*, 520**,** 157-167.

PRESENTATIONS AND INVITED LECTURES

Poster Presentation, “Analyzing the Variability and Teleconnection of Seasonal Precipitation over Western U.S. Using Empirical Orthogonal Function and Nonlinear ModeDecomposition,” American Geophysical Union Fall Meeting, December, 2014.

Invited Talk, “A New Data Mining Technique and its Application on Water Resource Problems in California,” Institution of Mathematical Behavior Science Seminar Series (UC Irvine), January 2015.

COMPUTER SKILLS

Programming: Python, MatLab, R, C++

Applications: HEC-GeoHMS, HEC-RAS, ArcGIS, ENVI

Platforms: Linux, Windows, Mac OS

ABSTRACT OF THE DISSERTATION

A Framework for Assessing Error Heteroscedasticity of Satellite Estimates and Extracting Spatiotemporal Variability from Precipitation Data

By

Hao Liu

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2015

Professor Soroosh Sorooshian, Chair

Precipitation is an important climate variable influencing human society. This dissertation focuses on developing a framework to assess the Remotely-sensed precipitation error's heteroscedasticity and precipitation's distribution patterns over space and time. The first part of the framework is to establish the joint-probability distribution functions (pdf) between satellite precipitation and ground observations with identical analytic format but adaptive parameters. The adaptability of the proposed model is verified by applying it to three locations (Oklahoma, Montana, and Florida), and by applying it to cold season, warm seasons and the entire year. Then the heteroscedasticities in the errors of satellite precipitations are investigated using my proposed model under those scenarios. The results show the joint-pdfs have the same formulation under these scenarios, whereas their parameter-sets were adaptively adjusted. This parametric model reveals detailed information about the spatial and seasonal

variations of the satellite precipitation. I found that the shape of the conditional pdf shifts across the intensity ranges. At the 10~20 mm d^{-1} range, the conditional pdf is L-shaped, while in the 40~60 mm d^{-1} range, it becomes more bell-shaped. I also conclude that no single satellite precipitation product outperforms others with respect to the different scenarios (i.e., seasons, regions, climates).

The second part of the framework applies Empirical Orthogonal Function (EOF) and Nonlinear Mode Decomposition (NMD) to extract multi-scale spatial patterns and physically-meaningful periodic signals from a space-time array of precipitation measurements. A case study over the southwestern US shows this combined EOF-NMD technique is capable of identifying the teleconnections that the regional precipitation's seasonal cycles are amplitude-modulated (AM) by large-scale climate oscillations: the AM signal from the first PC is correlated (R=0.41) with Pacific Decadal Oscillation index during 1976-2001, the AM signal from the fourth PC is correlated (R=0.56) with Nino 3.4 index during 1985-2000.

1. Introduction and Background

Precipitation, commonly occurred in the form of rain, snow, sleet, or hail, is the product of complicate physical processes. In micro-scale, precipitation forms from the thermodynamic process of atmospheric water vapor that condenses, aggregates, and falls on the ground. In macro-scale, it results from massive interactions between land, atmosphere, and ocean. As the major source of fresh water in the world, precipitation is of vital importance to human society but too much precipitation in a short period can cause flood or flush flood and too little precipitation over a long period can cause drought. People have kept precipitation measurements for a long time using various methods from original gauges on the ground to advanced satellite sensors from the space to obtain precipitation information as complete and accurate as possible.

In this thesis, I make comprehensive evaluations on the uncertainty (reliability) of satellite precipitation products (in reference to ground measurements) using a 2-dimensional joint-pdf model with adaptable parameters. Afterwards, I develop a new data-mining tool to effectively extract the spatiotemporal patterns from long-term precipitation data. Such patterns are the underlying variability that controls the behavior of precipitation events. Therefore, they are crucial for understanding and predicting precipitation at regional scales. This dissertation study is devoted to establishing analysis frameworks on these two important issues related to precipitation datasets.

In this chapter, I first review the current status of studies on remotely-sensed precipitation. Learning from previous works in literatures, I raise two fundamental problems and propose two frameworks for further study: 1) Constructing an improved stochastic error model for evaluating RS precipitation with model parameters adaptable to applied regions and time scales, and 2) Developing a combined Empirical Orthogonal Function and Nonlinear Mode Decomposition (EOF-NMD) method to effectively extract spatiotemporal patterns from long-term regional precipitation data. I proceed to address the hypothesis and objectives of this study and layout the organization of the dissertation.

1.1 Background on Remote Sensing Precipitation

1.1.1 Remote-sensing precipitation

Three types of instruments are employed to measure or estimate precipitation intensity fields on the ground: gauge network, weather radar, and satellite-based remote-sensors. Among them, satellite remotely-sensed precipitation estimation (hereafter RS precipitation) is becoming popular, because its near global scope coverage, high spatial resolution and fine sampling time interval (such as 0.25^{o} grid-box resolution and 3-hourly) establish the unique capability of monitoring large-scale precipitation systems such as hurricanes, typhoons, frontal rain/snow belts as well as fast evolving and intense convective storms on the global scale. Also, satellite precipitation covers mountainous and oceanic regions where radar and gauge network coverage is poor or nonexistent. Such information can help weather forecast, water resources management, and global flooding and drought prediction.

Currently, there are several algorithms from various institutions producing near-real-time precipitation estimations based on satellite information. Full lists of currently available RS precipitation products can be found in many review papers (Stephens and Kummerow 2007; Sorooshian et al. 2011; Tapiador et al. 2012).

1.1.2 Recent Studies on RS Precipitation

Recent studies on RS precipitation can be categorized, based on their focuses, into five general categories: (1) Regional performance; (2) New evaluation approaches including error models; (3) Inter-comparison of RS precipitation products; (4) Improving RS precipitation quality; (5) Exposing precipitation features and applications of RS precipitation data. The studies in each category can be further divide into sub-categories and their references are listed as follows:

1.1.2.1 Studies focus on Regional Performances

North America: USA (AghaKouchak, Bárdossy and Habib 2010b, AghaKouchak et al. 2011, AghaKouchak et al. 2012b, AghaKouchak, Nasrollahi and Habib 2009, Amitai et al. 2012, Behrangi et al. 2011, Chen et al. 2013b, Habib, Henschke and Adler 2009, McCollum et al. 2002, Mehran and AghaKouchak 2014, Zeweldi and Gebremichael 2009)

South America: Brasil (Buarque et al. 2011)

Africa: Ethiopia (Hirpa, Gebremichael and Hopson 2009)

Asia: Taiwan (Chen et al. 2013a), China (Gao and Liu 2013, Li, Zhang and Xu 2014), Pakistan (Khan et al. 2014)

Europe: (Stampoulis and Anagnostou 2012)

Globe or oceans: (Arkin and Ardanuy 1989, Sapiano and Arkin 2009, Sorooshian et al. 2000a)

The abovementioned work did a thorough study focus on regional performances of RS precipitation, but limit their evaluation tools with statistical indices (i.e. correlation coefficient, Mean Square Error, Mean Absolute Error). These statistical indices only reflect single aspect of the general picture about the error. In this thesis, I will provide more detail information about the error (i.e., the heteroscedasticity) with a comprehensive error model.

1.1.2.2 Studies focus on New Evaluation Approach or New Error Model

Object-based approach: (Li et al. 2015, AghaKouchak et al. 2010c)

Joint-cdf: (Gebremichael and Krajewski 2007, Moazami et al. 2014, Yan and Gebremichael 2009, Zhang et al. 2013)

Copula-based approach: (AghaKouchak, Bardossy and Habib 2010a, Gebremichael and Krajewski 2007)

Additive error model: (AghaKouchak et al. 2012b, Gebregiorgis and Hossain 2014, Hong et al. 2006, Kirstetter et al. 2012)

Multiplicative error model: (Müller and Thompson 2013, Tian et al. 2013)

Kernel Density Function Error model: (Gebremichael, Liao and Yan 2011)

Multi-objective approach: (AghaKouchak et al. 2011)

These evaluation techniques and error model provide very important information about how the error changes with intensity of precipitation. However, whether its tested for different scenarios is unknown. This thesis not only develop an error model, but also extensively test it in various scenarios (i.e., different seasons, different regions).

1.1.2.3 Studies focus on Inter-comparisons

Inter-comparison among different measuring instruments (gauge, radar and satellite): (Gourley et al. 2009, Scofield and Kuligowski 2003, Tapiador et al. 2012, Yilmaz et al. 2005)

Inter-comparison among different RS precipitation algorithms (including at least three RS products): (AghaKouchak et al. 2011, Gao and Liu 2013, Gebregiorgis and Hossain 2014, Hirpa et al. 2009, Li, Yang and Hong 2013, McCollum et al. 2002, Sapiano and Arkin 2009, Stisen and Sandholt 2010, Vernimmen et al. 2012, Yilmaz et al. 2005)

These works present an overview about accuracies of all the sensors or of all the RS algorithms. However, RS precipitation products apply the same algorithms for different types of precipitation. Their performances are highly dependent on environmental factors of precipitation events. This thesis aims to reveal the complex error property that is caused by these factors.

1.1.2.4 Studies focus on Improving RS Products

Bias adjustment/bias correction: (Boushaki et al. 2009, Müller and Thompson 2013, Tian, Peters-Lidard and Eylander 2010, Vernimmen et al. 2012)

Improving RS products using new approach: Artificial Neural Network (Turlapaty et al. 2010)

Improving RS products using numeric model forecast: (Kwon et al. 2008)

Improving RS products using gauge: (Li and Shao 2010)

Their works focus on innovative approaches to improve biases of RS products, and they work reasonably well in their study region. But the real problem with high-resolution RS precipitation is the dispersion problem. By studying the heteroscedasticity, I conclude future work will need to fix dispersion problem but not the biases in the RS estimates.

1.1.2.5 Studies focus on Special features or Applications

Hydrological Model response: (Gosset et al. 2013, Hong et al. 2007a, Hong et al. 2006, Stisen and Sandholt 2010)

Extremes: (AghaKouchak et al. 2011, Hsu et al. 2010, Mehran and AghaKouchak 2014, Sun and Barros 2010, Zhang et al. 2013)

Diurnal Pattern: (Hong et al. 2007b, Sorooshian et al. 2002)

Elevation Pattern: (Hong et al. 2007b, Li et al. 2014)

Spatial-temporal sampling issue: temporal sampling (Gebremichael and Krajewski 2005), spatial sampling (North et al. 1994), spatiotemporal resolutions (Steiner 1996)

Downscaling: (Duan and Bastiaanssen 2013, Foufoula-Georgiou et al. 2014)

Extending length of current records: (Ashouri et al. 2014)

Water Budget: (Sheffield et al. 2009)

Their works show the potentials of the RS products and study how these products can help understand precipitation itself. In this thesis, I also develop a pattern extraction tool for understanding the precipitation process.

1.1.3 Limitation in existing evaluations of RS precipitation

The quality of RS precipitation is mostly evaluated by comparing the concurrent measurements at RS pixels and at the co-located ground boxes, in other words, utilizing pixel-based one-to-one comparisons. The ground precipitation measurements are provided by gauge, radar or merged gauge-radar data. Enormous evaluations in literatures basically utilize two types of verification techniques, 1) using statistical indices and 2) using statistical error models. The latter approaches can provide more detailed and complete error information than the former ones. Here, before proceeding to a new stochastic error model of RS precipitation I first summarize several existing error models and their limitations.

In the following description, I use (X, Y) denoting a pair of non-negative random variables in which X stands for the RS precipitation estimate and Y for the corresponding ground measurement at the same time and same location. Samples of (X, Y) are represented in low cases as (x_i, y_i).

1.1.3.1 RS precipitation error models in literatures

a. Additive Error model

It is natural to categorize the error in RS precipitation into deterministic systematic error and random error. According to the theory of error analysis, random errors in experimental measurements are caused by unknown and unpredictable factors, on the other hand, systematic errors usually come from the imperfectness of measuring instruments/procedures (Taylor 1997). Therefore, decomposing the RS precipitation errors into these two categories of error can help correcting the bias in the RS products (AghaKouchak et al. 2012b). Additive Error model is the frequently-used model for the RS precipitation error (Habib et al. 2009, Ebert, Janowiak and Kidd 2007). It states,

$$X_i = a + bY_i + \varepsilon \tag{1.1}$$

Where, a is the offset, b is a scale parameter, and ε is the white noise. Given samples (x_i, y_i), the parameters, a, b and the variance of ε can be determined.

The error analysis also assumes the random error to be completely "random", which means independent and identical (i.i.d.). This assumption makes many statistical indices interpretable to describe the linearity and correspondence including Correlation Coefficient, Mean Square Error (or Root Mean Square Error), Mean Absolute Error, and Bias Ratio.

The first drawback of the error analysis is that it relies on the "system" that is known before the analysis. If the structure of the system is unknown, I merely get an approximated guess about the systematic errors by a presumptuous structure. Secondly, the RS precipitation errors are not i.i.d (AghaKouchak et al. 2010b). they are changing with precipitation intensity, locations and elevations, possessing diurnal cycles, seasonal cycles also sensitive to the spatiotemporal resolution. If these factors are not considered into the error decomposition process, the random error may not be fully random. Even when these factors are added into the systematic structure, the problem of how these factors interact with each other remains unsolved.

b. Multiplicative error model.

In representing daily RS precipitation errors, (Tian et al. 2013) found that the multiplicative error model is a much better choice under three criteria: (1) separation of the systematic and random errors; (2) applicability to the large range of variability in daily precipitation; and (3) better predictive skill. The model is represented as

$$X_i = a \cdot Y_i^b \cdot e^{\varepsilon_i} \tag{1.2}$$

This model can be seen as a logarithm-version of the additive error model therefore, can only make limited improvements on the additive error model.

c. Kernel smooth function error model

A nonparametric method for error estimation of RS precipitation is proposed in attempt to capture the features of conditional density $f(X = x \mid Y = y)$. According to ((Yan and Gebremichael (2009)) , the conditional density sharing the Gamma distribution is applied for all precipitation intensity range,

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} e^{-x/\beta} \tag{1.3}$$

In this study, the framework of evaluating RS precipitation error is developed following this initial assumption and another one to make the error model more comprehensive and more adaptive to applications.

1.1.3.2 Summary on limitations in existing error models

1) Applying the linearity hypothesis

In some of the above-mentioned references, statistical indices (i.e., bias ratio, Correlation Coefficient, Root Mean Square Error, etc.) are used as performance measures which indicate a linear relationship between random variables. Willmott (1982) argued with Fox (1981) that the indices, such as correlations, are "insufficient and often misleading".

2) Neglecting spatiotemporal variability underlying precipitation data

Most of the studies restrain their conclusions within the specific study locations and time periods and neglect the overall performances of RS precipitation products in representing the spatiotemporal variations of precipitation, while studies have shown that regional precipitation exhibits spatial patterns (van Ogtrop, Ahmad and Moeller 2014, LaMarche and Fritts 1971, Ning and Bradley 2014), seasonal variations (Barlow, Nigam and Berbery 2001, Elsanabary and Gan 2013, Jiang et al. 2014) and decadal variations (Cayan et al. 1998, Curtis 2008, Dettinger et al. 1998, Moron, Vautard and Ghil 1998). It is ignorant to proceed to evaluate RS precipitation without account for the spatial and temporal variations. Such attempt only returns a small subset of the complex characters of RS precipitation.

3) Focusing on bias-adjustment

Bias adjustment works effectively when the systematic error is the dominant part in the total error, however, as I will show later, random error is dominant in RS precipitation products.

1.1.4 Source of Errors in RS precipitation

RS precipitation products suffer from serious uncertainty problems which restrict practical applications of RS precipitation products. As in the above review, numerous studies have devoted to evaluations of RS precipitation. The error of RS precipitation is defined as the difference between the RS estimates and the collocated ground measurements of precipitation amounts. According to the previous studies, the major error sources of RS precipitation are:

1.1.4.1 Cause 1: Spatiotemporal sampling differences

In terms of precipitation estimates, the satellite platform has different spatial system and temporal sampling interval from those of ground observations. And these sampling differences inevitably cause errors when the two sources of precipitations are compared under the same spatiotemporal resolution. However, this kind of error can be reduced by decreasing the spatial or temporal resolution. On the other hand, the sampling-based error becomes dominant in high spatiotemporal resolutions.

1.1.4.2 Cause 2: Differences in measuring instruments

Another kind of errors comes from different precipitation measuring instruments onboard satellites and at the ground. RS precipitation basically relies on infrared brightness temperature and passive-microwave emission of cloud. On the ground, gauges measure rainfall amounts and radars estimate rainfall based on the reflectivity of raindrops. The effect of different sensors is twofold. The first effect is difference in capturing the timings of precipitation systems; this effect is more obvious in cases of convective storm. The second effect is difference in capturing the shapes of precipitation systems; this effect can be found in numerous cases including hurricane events and frontal systems. This sensor-based error is the major source of false-alarm and miss-detection error.

1.1.4.3 Cause 3: Use of a unified retrieval algorithm regardless of precipitation seasonality and region

Currently, each of the global RS precipitation products employs the same algorithm to quantify precipitation disregarding the possible influences of time and location on the precipitation amounts. For example, one of the RS algorithms estimates snowfall over Serra-Nevada Mountains, California, and also estimates the hurricane rainfall over

Florida in the same way without any adjustments. Such algorithm-based errors are massive, exhibiting strong spatial and temporal variations in errors.

1.1.4.4 Cause 4: Inability of capturing extreme precipitation events

Precipitation events are discrete in time domain and their amounts follow some non-stationary extreme distributions (AghaKouchak et al. 2012a, Li, Singh and Mishra 2012). Current RS algorithms do not treat extreme events specially therefore usually underestimate or overestimate the precipitation maximums resulting in low capability of flood warning. Under the condition of global climate change and global warming, many studies have called for intensive observations and studies on the extreme precipitation (Cayan et al. 1998, Curtis 2008, Dettinger et al. 1998, Moron et al. 1998).

1.2 Research Motivations

Through above review, I found at least one of the following fundamental questions about RS precipitations remain unanswered based on above review of available error models for studying the RS precipitation.

1. How is the error model adaptive to study different location and different seasons?
2. How does the model capture the distinct features of the datasets?
3. What are the patterns in the spatiotemporal variability of RS precipitations?

Inspired by previous work, this dissertation aims to address the fundamental problems on diagnosing the errors in RS precipitation and understanding the precipitation process itself. These problems include:

1. How to account for the heteroscedastic (variance changing) error in the RS precipitation?

2. How to account for the spatiotemporal variability while describing the heteroscedasticity of RS precipitation.
3. What are the distinct characters of RS precipitation? What makes one RS product different from another RS product? How to qualitatively or quantitatively describe these differences.
4. Is precipitation process completely random? Are there any distinct characters in each of the precipitation datasets?
5. Which climate factors are impacting the precipitation and how much they contribute?

1.3 Research Hypothesis and Objectives

The objective of this dissertation study is to exploit two frameworks both related to better understanding precipitation data and the underlying characteristics. The first framework aims to construct a copula-based stochastic error model with adaptive parameters to make a comprehensive evaluation on three influential public RS precipitation products; the second framework is to develop a statistical decomposing technique to extract the spatial variability and temporal periodicity of precipitation from the long-term dataset.

Here, the research hypotheses and objectives are listed as follows:

Hypothesis 1:

Precipitation either from RS estimation or from ground observation is treated as a random variable. When both of them are acting on the same precipitation phenomina, these two random variables are dependent to each other.

Objective 1:

Based on the Hypothesis 1, I develop a framework which leads to an RS precipitation error model in the form of joint cumulative distribution function (cdf) or joint probability density function (pdf). The joint cdf is derived using marginal and copula functions with adjustable parameters adaptive to the applied RS products, time periods, and regions.

Hypothesis 2:

The precipitation data which I am working on contain sufficient information in quality and quantity for us to extract the spatiotemporal patterns of regional precipitation.

Objective 2:

Based on Hypothesis 2, I develop the combined Empirical Orthogonal Function and Nonlinear Mode Decomposition (Combined EOF-NMD) scheme to effective extract spatiotemporal patterns from long-term precipitation data. The extracted patterns consisting of EOF and NMD modes have reduced dimensions and noise, representing the characteristics of spatiotemporal variability of the regional precipitation

1.4 Organization of the dissertation

Chapter 1 is for Introduction

Chapter 2 describes the development and validation of the bi-variant mixture error model for RS precipitation with the emphasis on address model parameters adaptable to applied RS algorithms, locations and seasons. The model is applied to 9-year datasets of three RS precipitations at three locations and two seasons. Error-intensity relationship of these RS precipitation is interpreted from the error-model.

Chapter 3 is devoted to the development and validation of the combined EOF-NMD technique. This approach is applied to the southwestern United States to extract the spatial correlation patterns (EOF modes) and temporal periodic patterns (NMD modes) from the regional precipitation data. The physical meaning of the decomposition results are explained with regard the teleconnection of southwestern precipitation to certain Pacific climate oscillations.

Chapter 4 summarizes the conclusions and discusses about future directions.

2. Framework for Assessing the Error in Remotely-sensed Precipitation

2.1 Introduction

Satellite precipitation estimates are becoming popular for use in monitoring large-scale tropical systems such as hurricanes and typhoons as well as fast evolving and intense convective storms in a global scale. Such information can help the prediction of flood events, assessment of the extent of inundation of the impact regions, and estimation of financial losses associated with extreme weather conditions (Smith and Katz 2013, Ross and Lott 2003). The Satellite precipitation provides imagery over mountainous and oceanic regions, where radar and gauge network coverage is poor or nonexistence. In global scale, Remote-Sensing (RS) precipitation products are also preferred for their higher spatial resolution and finer sampling intervals (0.25^{o} resolution and their 3-hour) than the ground-based precipitation datasets (Huffman et al. 2007, Joyce et al. 2004, Hsu et al. 1997).

The satellite precipitation error are the difference between the RS precipitation estimate and the collocated ground measurement (hereafter, the error), and it limits the potential application of RS precipitation products (Sorooshian et al. 2011, Turk et al. 2008). The error can be quantified through a number of indices such as bias ratio, variance, correlation coefficient and prediction skills such as probability of detection (Buarque et al. 2011, Gebremichael et al. 2005). The error is sensitive to rainfall intensity, spatiotemporal-resolution, location, climate, elevation, season, and other factors (Adler et al. 2003, Buarque et al. 2011, Huffman 1997, Sorooshian et al. 2000b, Steiner et al. 2003, Yuan et al. 2005). The relationships between the error and these factors have been studied by several error models (Gebremichael and Krajewski 2005, Gebremichael et al. 2011, Gebremichael, Over and Krajewski 2006, Hossain and Anagnostou 2006, Yan and Gebremichael 2009).

The error is more related to precipitation intensity than to the other factors, and the relationship is determined using parametric and non-parametric models (Gebremichael et al. 2011, AghaKouchak et al. 2010a, Ciach, Krajewski and Villarini 2007, Gebremichael and Krajewski 2007, Gebremichael and Krajewski 2005, Gebremichael et al. 2006, Hossain and Anagnostou 2006, Moazami et al. 2014, Yan and Gebremichael 2009). In this paper heteroscedasticity refers to the changing variance of the errors (i.e., between satellite estimates and ground-based precipitation observations) associated with the intensity of precipitation measured by ground radar. In some cases modelers have assumed a linear relationship (AghaKouchak et al. 2010a) between the errors as a function of intensity and others have preferred the assumption of nonlinearity (Gebremichael et al. 2011, Hossain and Anagnostou 2006) to represent the heteroscedasticity. To further investigate the nature of heteroscedasticity, the error's probability density function (pdf) and confidence interval (CI) are determined by a non-parametric joint-distribution (Gebremichael et al. 2011). Such information provides a probabilistic form of satellite precipitation. However, it is unclear whether such error-models are suitable for different regions or different seasons. The type and amount of precipitation can shift dramatically in different regions and in different seasons, and that shift should be reflected in the error model. Therefore, a model that can be adaptively adjusted to different scenarios is needed.

To investigate the spatial and seasonal variation of heteroscedasticity, I firstly setup an adaptive parametric error model. I use a mixed joint-pdf model to determine the shape of the radar's pdf and the CI which is conditional at any precipitation intensity. Thus, I obtain a parametric joint-pdf for the random pairs (X,Y) , where X represents the precipitation obtained from a satellite estimation and Y is the ground measurement (i.e., radar). The pairs (X,Y) are discrete because of zero-values, therefore, a discrete joint-cdf is needed. Previous work suggested the discrete joint-cdf can be a combination of four continuous joint-cdf (Serinaldi 2009a, Serinaldi 2009b, Villarini et al. 2014, Villarini, Serinaldi and Krajewski 2008, Herr and Krzysztofowicz 2005). To focus on the continuous part of joint-cdf and to make the model adaptable, I applied a threshold to the variable pairs. After the variable pairs are modeled by a continuous joint-cdf, I derive the conditional-pdf for a range of rainfall intensities from satellite precipitation. Then I analyze three different RS products during three seasons over three locations in the Contiguous United States using my proposed parametric distribution.

2.2 Method and Material

2.2.1 Study locations and datasets

Three satellite precipitation products are included in this study: the PERSIANN from UC Irvine, the TRMM-3B42RT from NASA and the CMORPH from NOAA-Climate Prediction Center. The NEXRAD stage-IV product is used as the ground reference, and a 9-year (2003~2011) overlapping period of Satellite/Radar rainfall record is obtained for evaluation. Because Stage-IV datasets have finer spatiotemporal resolution than satellite, they were transformed to satellite's resolution using arithmetic mean method in the following study. A brief summary and the corresponding references are given in Table 2.1.

Table 2.1 Satellite precipitation algorithms, ground radar products used in this study, along with their resolutions, coverage and primary investigators

Name	Producer	Coverage	Resolutions	Sources	Reference
TRMM-3B42RT	NASA	50^{o}N-50^{o}N 180^{o}E-180^{o}W	3-hourly $0.25°\times0.25°$	GOES-IR+TMI-PMV	(Huffman et al. 2007)
CMORPH	NOAA-CPC	60^{o}N-60^{o}S 180^{o}E-180^{o}W	3-Hourly $0.25°\times0.25°$	GOES-IR+TMI-PMV	(Joyce et al. 2004)
PERSIANN	UCI	60^{o}N-60^{o}S 180^{o}E-180^{o}W	3-Hourly $0.25°\times0.25°$	GOES-IR+TMI-PMV	(Hsu et al. 1997)
NEXRAD Stage-IV	NOAA-NCEP	CONUS	Hourly $4km \times 4km$	NEXRAD	(Fulton et al. 1998)

For the study area, I selected three climatically diverse locations based on Köppen Climate classification as reference. For example, Köppen's classification considers most of Oklahoma and Florida as "Humid subtropical climate", and Montana as "Semi-arid climate" (Peel, Finlayson and McMahon 2007). The three selected locations are tested in warm season (Apr. to Sep.), cold season (Oct. to Mar.) and also over the entire year (both warm and cold together). The uncertainties in the radar measurement increase with distance away from the radar base because of range degradation (Villarini and Krajewski 2010). Therefore, I chose study sites (Figure 2-1) that are either close to radar stations (in Montana and Florida) or have dense gauge network available (in Oklahoma). For each site and season, individual uncertainty models are obtained for each of the RS products.

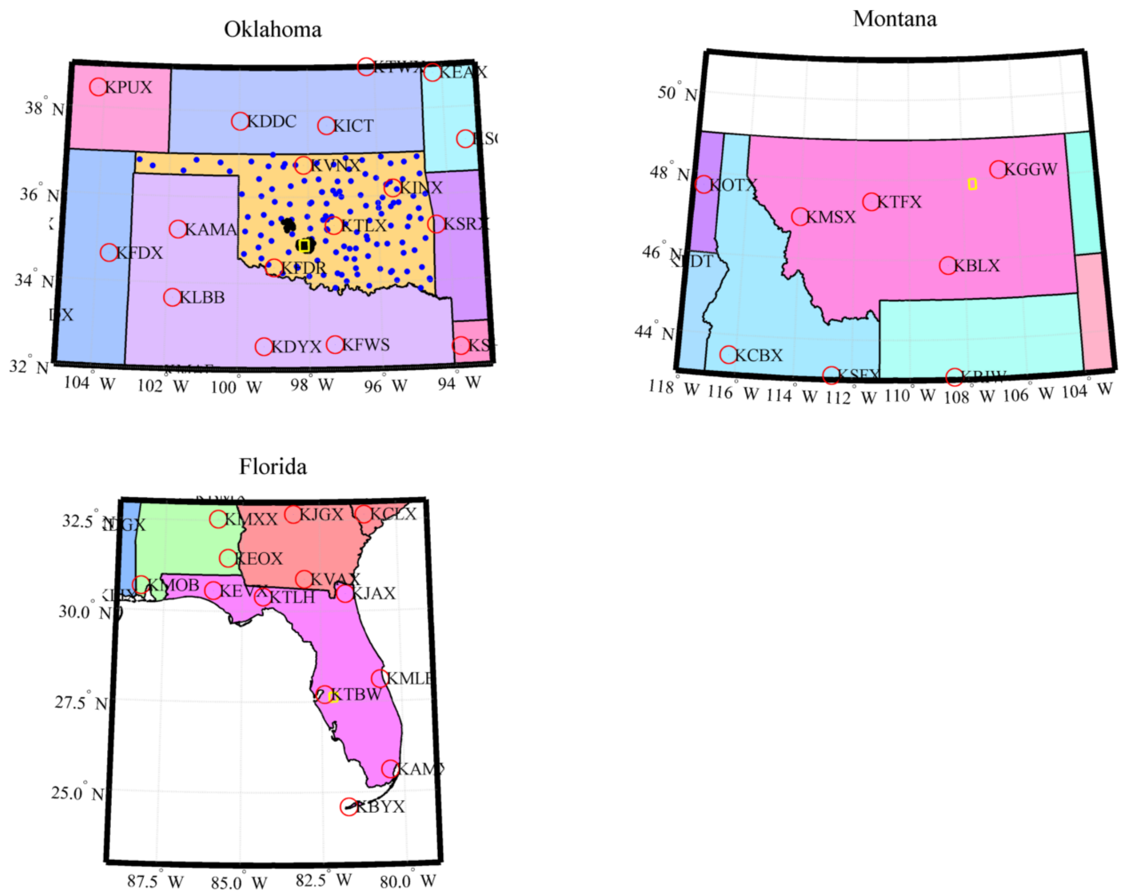

Figure 2-1 The 0.25°×0.25° study area (yellow rectangular box) in Oklahoma (near KFDR), Montana (near KGGW) and Florida (near KTBW). The blue dots are Mesonet and the black dots are Micronet gauge network in Oklahoma. The circles are for WSR-88D radar stations (http://www.nws.noaa.gov/code88d/index.html)

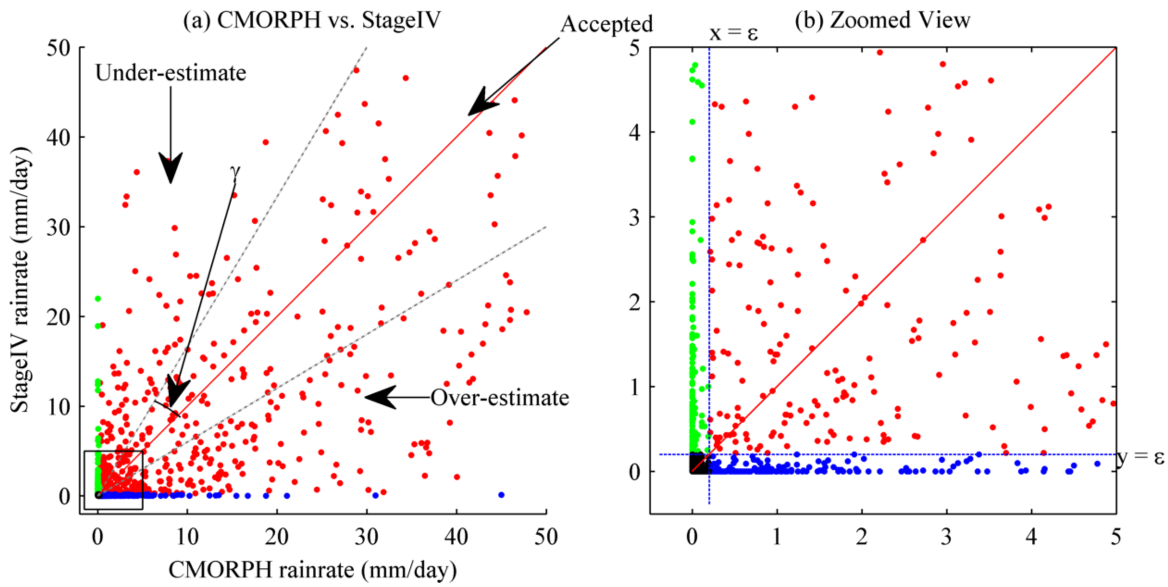

Figure 2-2 Scatter plot of CMORPH daily rainfall estimates vs. Stage IV observations during year 2003~ 2011 over an area of $0.25° \times 0.25°$ at Little-Washita River Experimental Watershed, Oklahoma with dashed lines at $\xi = 0.2$ mm d^{-1} dividing the first quadrant into four domains.

2.2.2 The bivariate-mixed model

Random pairs (X, Y) denote non-negative random variables, of which $\mathbb{P}(X = 0) > 0$ and $\mathbb{P}(Y = 0) > 0$. Here X represents the precipitation estimate belonging to one of the Remote-sensing satellite products and Y is the ground measurement belonging to the same grid. The joint-pdf for (X, Y) are formulated to obtain the pdf of a conditional distribution for Y given $X = x$.

There are actually two concepts under the same name. First concept is the framework that leads to the joint-pdf that describe the continuous part of the datasets, the framework include: (1) threshold selection, (2) copula and marginal distribution selection and test, and (3) parameter calibration. For future clarification I refer it as “the framework”. The framework itself can be rather adaptive enough for various studied that involve two dependent variables, which include other climate variable or the precipitation in other spatiotemporal resolutions.

The second concept is the joint-pdf in the theis which is based on the available RS and ground precipitation datasets, and it is also a result of the first two steps of the framework. In the revised manuscript I call it “the model”. The model can be used for RS precipitation in current resolution and can be applied to various seasons, regions or climates. The benefit of the model is to provide much more details about the error-intensity relationship, and study how this relationship varies with regions and seasons.

In the use of the term “adaptive”, I refer to the definition, “able to adjust or modify to fit into requirements or environmental conditions”. I wish to clarify that the entire frameworks adaptive, and the joint-pdf in Eqn. (2.7) is also adaptive enough for all the scenarios within my study. Furthermore, the framework of constructing the joint-pdf is indeed self-adaptive to any data sets. This framework is also adaptive to other locations or products by adjusting the choices of threshold, copula and marginal distributions and by calibrating its parameter combinations. Previous work also used copula-based joint-pdf for rainfall datasets that has higher or lower spatial and/or temporal resolution (AghaKouchak et al. 2010b, Ghosh 2010, Serinaldi 2009a, Villarini et al. 2008).

The framework that formulating the joint-pdf consists of three steps. First, a threshold is chosen to account for minimum errors in the datasets in preparation for making the dataset ready for obtaining the joint-pdf. Second, a copula-based method is applied to the truncated positive pairs in order to determine the suitable formulation of the joint-cdf. Third, the parameters of the joint-pdf are calibrated by minimizing a multi-objective criterion which measures the differences between the model and the data. The criterion include the four indices (CORR, BIAS, MAE and MSE).

The same threshold ξ is applied to both X and Y , and p_{00}, p_{01} , p_{10} , p_{11} represent the probabilities of that data pairs fall into four categories:

$$\begin{aligned} p_{00} &= \mathbb{P}(X \le \xi, Y \le \xi) \\ p_{10} &= \mathbb{P}(X > \xi, Y \le \xi) \\ p_{01} &= \mathbb{P}(X \le \xi, Y > \xi) \\ p_{11} &= \mathbb{P}(X > \xi, Y > \xi) \end{aligned} \tag{2.1}$$

If the Cumulative Distribution Function (cdf) of bivariate variables (X,Y) is discrete and is defined as $\Phi(x,y) = \mathbb{P}(X \le x, Y \le y \mid x > 0, y > 0)$, I have:

$$\Phi(x,y) = p_{00} + p_{10}H_{X,\xi}(x) + p_{01}H_{Y,\xi}(y) + p_{11}H_{\xi}(x,y) \quad \forall (x,y) \in [\xi, +\infty)^2 \tag{2.2}$$

Where

$$\begin{aligned} H_{X,\xi}(x) &= \mathbb{P}(X \le x \mid x > \xi, 0 < y \le \xi) \\ H_{Y,\xi}(y) &= \mathbb{P}(Y \le y \mid y > \xi, 0 < x \le \xi) \\ H_{\xi}(x,y) &= \mathbb{P}(X \le x, Y \le y \mid X > \xi, Y > \xi) \end{aligned} \tag{2.3}$$

Because:

$$
\begin{aligned}
&p_{00} + p_{10} + p_{01} + p_{11} = 1 \\
&H_{X,\xi}(x) \in [0,1] \\
&H_{Y,\xi}(y) \in [0,1] \\
&H_{\xi}(x,y) \in [0,1]
\end{aligned} \tag{2.4}
$$

That gives:

$$
\begin{aligned}
\Phi(\xi,\xi) &= p_{00} \\
\Phi(+\infty,\xi) &= p_{00} + p_{10} \\
\Phi(\xi,+\infty) &= p_{00} + p_{01} \\
\Phi(+\infty,+\infty) &= 1
\end{aligned} \tag{2.5}
$$

I determine the threshold ξ based on the following procedure. First, I study the metadata of the precipitation datasets to determine the minimum errors of all the products to be studied. Second, I determine the minimum value by using trial-and-error approach when the Goodness-of-fit (GoF) tests are passed for all the scenarios (different RS products, seasons or locations). Third, I choose the bigger value from the above two steps as the threshold ξ. In this way, the threshold ξ not only accounts for observational error, but also ensures the adaptability of joint-cdf $H_{\xi}(x,y)$ under various scenarios.

In the study, The NEXRAD Stage IV rainfall observation system has a minimum error of 0.2 $\mathrm{mm\,d^{-1}}$ (Seo and Krajewski 2010, Xie et al. 2006). 3%~5% of the data pairs fall into the 0~0.2 $\mathrm{mm\,d^{-1}}$ range (Figure 2-2), while the total positive pairs account for 10% ~ 20% of all pairs (Table 2.5). These variable pairs, which only account for less than 5-percent of the total rainfall amount, will hamper the adaptability of a joint-cdf. After a threshold $\xi =$ 0.2 $\mathrm{mm\,d^{-1}}$ is applied to truncate the data pairs, less than 5% of total rainfall amount is discarded from the positive pairs (Table 2.5, Table 2.6, and Table 2.7), and the GoF tests are passed for all the products.

If the joint-cdf $H_\xi(x, y)$ has continuous marginal distribution functions: $F_\xi(x) = \mathbb{P}(X \le x \mid X > \xi)$ and $G_\xi(y) = \mathbb{P}(Y \le y \mid Y > \xi)$, based on Sklar's theorem, there is a unique copula $C_\xi : [0,1]^2 \to [0,1]$ that:

$$H_\xi(x, y) = C_\xi\{F_\xi(x), G_\xi(y)\}, \quad \forall (x, y) \in [\xi, +\infty)^2 \tag{2.6}$$

To determine the type of the copula C_ξ , a GoF test recommended by Genest, Rémillard and Beaudoin (2009) is applied to a group of candidate copulas. To determine the type for marginal distributions $F_\xi(x)$ and $G_\xi(y)$, a 2-sample Komogorov-Smirnorv GoF test is applied on a group of candidate distributions. Details of the GoF test are provided in page 29.

The joint-cdf $H_\xi(x, y)$ is formulated using Gumbel-Hogarrd copula, C_ξ , Weibull distribution $F_\xi(x)$, and Weibull distribution $G_\xi(y)$ based on the result (in the following section). From Eqn. (2.6), a joint-cdf $H_\xi(x, y)$ and joint-pdf $h_\xi(x, y)$, has the copula parameter θ, the shape parameters and scale parameters $k_1, k_2, \lambda_1,\ \lambda_2$, as shown below:

$$H_{\xi}(x,y) = \exp\left\{ -\left[\left(\ln\left\{ 1-\exp\left[-\left(\frac{x-\xi}{\lambda_1} \right)^{k_1} \right] \right\} \right)^{\theta} + \left(\ln\left\{ 1-\exp\left[-\left(\frac{x-\xi}{\lambda_2} \right)^{k_2} \right] \right\} \right)^{\theta} \right]^{1/\theta} \right\}$$

$$h_{\xi}(x,y) = \frac{k_1}{\lambda_1}\left(\frac{x-\xi}{\lambda_1} \right)^{k_1\theta-1} \frac{k_2}{\lambda_2}\left(\frac{y-\xi}{\lambda_2} \right)^{k_2\theta-1} \left\{ \left(\frac{x-\xi}{\lambda_1} \right)^{k_1\theta} + \left(\frac{y-\xi}{\lambda_2} \right)^{k_2\theta} \right\}^{1/\theta-2}$$
$$\times \left\{ \left[\left(\frac{x-\xi}{\lambda_1} \right)^{k_1\theta} + \left(\frac{y-\xi}{\lambda_2} \right)^{k_2\theta} \right]^{1/\theta} + \theta - 1 \right\} \exp\left\{ -\left[\left(\frac{x-\xi}{\lambda_1} \right)^{k_1\theta} + \left(\frac{y-\xi}{\lambda_2} \right)^{k_2\theta} \right]^{1/\theta} \right\}$$

(2.7)

In the last step, the parameter set $\psi = \{\theta, k_1, k_2, \lambda_1, \lambda_2\}$ is further calibrated and fine-tuned. ψ is a parameter set for the joint-cdf $H_{\xi}(x,y)$ which include parameters for copula function C_{ξ} , marginal function $F_{\xi}(x)$, and $G_{\xi}(y)$. The bounds for parameter calibration are $\pm 20\%$ of initially estimated parameter values from the copula's and marginal distributions' selection process.

As already mentioned above, the four indices which are used to develop the multi-objective criterion are: Mean Square Error (MSE), Mean Absolute Error (MAE), bias-ratio (BIAS), and correlation coefficient (CORR). These indices are calculated from recorded radar and RS precipitation pairs $(X,Y)_{obs}$, which is denoted as MSE_{obs}, MAE_{obs}, $CORR_{obs}$, and $BIAS_{obs}$. Data pairs $(X,Y)_{sim}$ are also simulated from joint-cdf $H_{\xi}(x,y)$ with a combination of parameter set $\psi = \{\theta, k_1, k_2, \lambda_1, \lambda_2\}$, and the same four indices are calculated from $(X,Y)_{sim}$ and are denoted as $MSE_{sim,\psi}$, $MAE_{sim,\psi}$, $CORR_{sim,\psi}$, and $BIAS_{sim,\psi}$.

$$
\begin{aligned}
MSE_{sim,\psi} &= \frac{1}{n}\sum_{i=1}^{n}(X_{i,sim}-Y_{i,sim})^2 \\
MAE_{sim,\psi} &= \frac{1}{n}\sum_{i=1}^{n}|X_{i,sim}-Y_{i,sim}| \\
BIAS_{sim,\psi} &= \frac{\sum_{i=1}^{n}X_{i,sim}}{\sum_{i=1}^{n}Y_{i,sim}} \\
CORR_{sim,\psi} &= \frac{\sum_{i=1}^{n}(X_{i,sim}-\overline{X}_{sim})(Y_{i,sim}-\overline{Y}_{sim})}{\sqrt{\sum_{i=1}^{n}(X_{i,sim}-\overline{X}_{sim})^2}\sqrt{\sum_{i=1}^{n}(Y_{i,sim}-\overline{Y}_{sim})^2}} \\
MSE_{obs} &= \frac{1}{n}\sum_{i=1}^{n}(X_{i,obs}-Y_{i,obs})^2 \\
MAE_{obs} &= \frac{1}{n}\sum_{i=1}^{n}|X_{i,obs}-Y_{i,obs}| \\
BIAS_{obs} &= \frac{\sum_{i=1}^{n}X_{i,obs}}{\sum_{i=1}^{n}Y_{i,obs}} \\
CORR_{obs} &= \frac{\sum_{i=1}^{n}(X_{i,obs}-\overline{X}_{obs})(Y_{i,obs}-\overline{Y}_{obs})}{\sqrt{\sum_{i=1}^{n}(X_{i,obs}-\overline{X}_{obs})^2}\sqrt{\sum_{i=1}^{n}(Y_{i,obs}-\overline{Y}_{obs})^2}}
\end{aligned}
\tag{2.8}
$$

In the objective function (O.F.), the values of the four indices are normalized and have equal weights. Parameter set of ψ is then optimized by minimizing the O.F.:

$$
\begin{aligned}
\psi &= \text{Arg min}\left(O.F.\right) \\
O.F. &= OF_{MSE}+OF_{MAE}+OF_{BIAS}+OF_{CORR} \\
&= \left(\frac{MSE_{sim,\psi}-MSE_{obs}}{MSE_{obs}}\right)^2+\left(\frac{MAE_{sim,\psi}-MAE_{obs}}{MAE_{obs}}\right)^2+\left(\frac{BIAS_{sim,\psi}-BIAS_{obs}}{BIAS_{obs}}\right)^2 \\
&\quad +\left(\frac{CORR_{sim,\psi}-CORR_{obs}}{CORR_{obs}}\right)^2
\end{aligned}
\tag{2.9}
$$

The Shuffle Complex Evolution Algorithm (Duan, Gupta and Sorooshian 1993, Duan, Sorooshian and Gupta 1992) is adopted to obtain the optimal values of the parameters for $(X, Y)_{obs}$. Sensitivity analysis was also performed for each of the parameters with respect to each of the four components of the multi-objective criterion Eqn. (2.9). The results clearly support that the proposed calibration approach can find the optimal parameter set (see page 32).

The optimal parameter sets are then used in $H_{\xi}(x, y)$. The cdf of the conditional distribution, given $X = x$, is: $H_{\xi}(y \mid x) = \mathbb{P}(Y \le y \mid X = x, Y > \xi, X > \xi)$, and it can be obtained as follows:

$$\begin{aligned} F_{\xi}(x) &= 1 - \exp\left[-\left(\frac{x-\xi}{\lambda}\right)^{k}\right] \\ f_{\xi}(x) &= \mathrm{d}F_{\xi}(x)/\mathrm{d}x \end{aligned} \tag{2.10}$$

Because:

$$\mathbb{P}(Y < y \mid X = x) = \frac{P(Y < y, X = x)}{P(X = x)} = \frac{\partial H_{\xi}(x, y)/\partial x}{\mathrm{d}F_{\xi}(x)/\mathrm{d}x} = \frac{1}{f_{\xi}(x)} \cdot \frac{\partial H_{\xi}(x, y)}{\partial x} \tag{2.11}$$

I have:

$$H_{\xi}(y \mid x) = \frac{1}{f_{\xi}(x)} \cdot \frac{\partial H_{\xi}(x, y)}{\partial x} \tag{2.12}$$

By using Eqn. (2.12), I can derive the pdf and CI of the conditional distribution based on joint-pdf. The pdf $h_{\xi}(y \mid x) = \mathbb{P}(Y = y \mid X = x, Y > \xi)$ is:

$$h_{\xi}(y \mid x) = \frac{h_{\xi}(x, y)}{f_{\xi}(x)} \tag{2.13}$$

The mean value of the conditional pdf $h_{\xi}(y \mid x)$ can be estimated using:

$$E(Y \mid X = x) = \int_{\xi}^{+\infty} y \cdot h_{\xi}(y \mid x)\, \mathrm{d}y \tag{2.14}$$

I define the inverse function of $H_{\xi}(y \mid x) = p$ (p is the conditional cumulative probability) in Eqn. (2.12) as $H_{\xi}^{-1}(p \mid x) = y$. Then the Confidence Limits (CL), which is the endpoints for confidence interval at confidence level of $1-\alpha$, is:

$$\begin{aligned} CL_{upper,1-\alpha} &= H_{\xi}^{-1}\left(1-\frac{\alpha}{2} \mid x\right) \\ CL_{lower,1-\alpha} &= H_{\xi}^{-1}\left(\frac{\alpha}{2} \mid x\right) \end{aligned} \tag{2.15}$$

2.2.3 Determine the type of marginal distributions and copula

The Sklar's theorem suggests that a unique 2-dimensional copula exists when the there is a 2-d joint-cdf and its two marginal cdfs. Thus, to ensure there is a continuous joint-cdf for the positive pairs (X, Y), the continuous marginal distributions that fit the marginal dataset need to be identified. There are a number of distributions reported to be fitted into the daily precipitation (Li, Singh and Mishra 2013). Five candidate distributions (Weibull distribution, Exponential distribution, Log-normal distribution, General-Extreme-Value distribution, Gamma distribution) were tested using Kolmogorov–Smirnov two-sample test (KS-test) for three RS datasets over a region in Oklahoma (Figure 2-1). Then the negative log-likelihoods for these distributions were estimated. Among these distributions, the Weibull-distribution passed the GoF tests for all three products and ground observations (Table 2.2), it also shows the smallest negative-log-likelihood (Table

2.3). Therefore, Weibull distribution is the most suitable distribution among the candidates.

The copula function's type is closely related with the two variable's dependence. I selected eight copula functions of which five functions are Archimedean Copulas and four functions are Extreme-value Copulas. Gumbel-Hougarrd Copula is an Archimedean Copula and an Extreme-value Copula. All of these eight copulas only have one parameter (Nazemi and Elshorbagy 2012). I estimated these copulas' parameter based on Kendall's tau for three RS datasets over a region in Oklahoma and then called GoF tests. I followed a KS-test-based GoF test procedure recommended by Genest et al. (2009). The most suitable copula function is chosen based on the $\mathbb{P}$-value from the hypothesis test. $\mathbb{P}$-value larger than 0.05 means passing the GoF test (Genest et al. 2009). The GoF test results show Gumbel-Hougarrd copula is suitable for all three RS products over the Oklahoma region (Table 2.4).

Table 2.2 Results of KS-test after fitting marginal data into candidate distributions, larger p-values indicate better fitting.

	KS-test p-value					
	PER vs. Stage-IV		CMO vs. Stage-IV		TRMM vs. Stage-IV	
	PER	ST4	CMO	ST4	TRM	ST4
Exponential	0.00	0.00	0.00	0.00	0.00	0.00
Weibull	**0.75**	**0.54**	**0.78**	**0.95**	**0.59**	**0.89**
Gamma	0.28	0.40	0.41	0.41	0.20	0.74
Log-normal	0.06	0.05	0.08	0.21	0.02	0.21
General-EV	0.01	0.02	0.05	0.08	0.32	0.09

Table 2.3 Results of negative log-likelihood after fitting marginal data into different distributions, smaller values indicate better fitting

	Negative loglikelihood					
	PER vs. ST4		CMO vs. ST4		TRM vs. ST4	
	PER	ST4	CMO	ST4	TRM	ST4
Exponential	914.88	842.46	1000.32	908.12	818.58	742.65
Weibull	**873.41**	**812.43**	**970.27**	**872.02**	**802.09**	**725.93**
Gamma	875.61	812.86	972.99	874.96	803.40	726.94
Log-normal	883.66	825.69	985.53	881.76	819.37	737.07
General-EV	903.22	845.09	997.19	900.76	823.27	749.00

Table 2.4 Goodness-of-fit test results for candidate copulas, large p-values indicate better fitting

	Goodness-of-fit test p-value					
	PER vs. radar	PER vs. gauge	CMO vs. radar	CMO vs. gauge	TRM vs. radar	TRM vs. gauge
Normal	0.002	0.032	0.007	0.022	0.087	0.306
Clayton	0.002	0.002	0.002	0.002	0.002	0.002
Frank	0.007	0.022	0.002	0.017	0.107	0.336
Plackett	0.007	0.022	0.002	0.007	0.092	0.381
Gumbel-Hougarrd	**0.396**	**0.590**	**0.142**	**0.425**	**0.545**	**0.878**
Husler-Reiss	0.167	0.460	0.012	0.107	0.197	0.455
Galambos	0.167	0.460	0.012	0.087	0.216	0.550
t-EV	0.162	0.465	0.007	0.097	0.281	0.709

2.2.4 Sensitivity analysis of the parameters in the joint-cdf

To determine the efficiency of the proposed calibration approach, a sensitivity analysis is conducted for each of the five parameters of the joint-cdf. First, I change the value of each parameter while fixing the other parameters to see how those difference-based indices respond to the parameter. I calculate the indices when each of the parameters varies within a small range (±20% of their values which are shown in first column of Table 2.5, for the case of CMORPH in Oklahoma). In this way, I can see the relationship between every difference-based index and the parameters.

The results are summarized in Figure 2-3. The increase in copula parameter θ increase the CORR, reduce the MAE and MSE, but the BIAS is insensitive to θ. Changes in the parameters of marginal distribution (i.e., $k_1, k_2, \lambda_1, \lambda_2$) will change the total amount and dispersion of their rainfall distribution, and therefore cause change in the BIAS, MAE and MSE. However, the CORR is insensitive in response to the parameters of marginal distribution.

Second, I change the value of each parameter while fix the other parameter to see how each of the multi-objective components in the O.F. function respond to the parameter. To do so, I calculate the four components in the Eqn (2.9) (i.e., OF_{CORR}, OF_{BIAS}, OF_{MSE}, OF_{MAE}), when each of the parameters varies within a small range (±20% of their values which are shown in first column of Table 2.5, for the case of CMORPH in Oklahoma).I also analyze the sensitivity of the combined O.F. for each parameter. The results are summarized in Figure 2-4. I can see, by normalize each component of the

O.F., the scaling issue is resolved and the parameters converge to their optimal value through the optimization process.

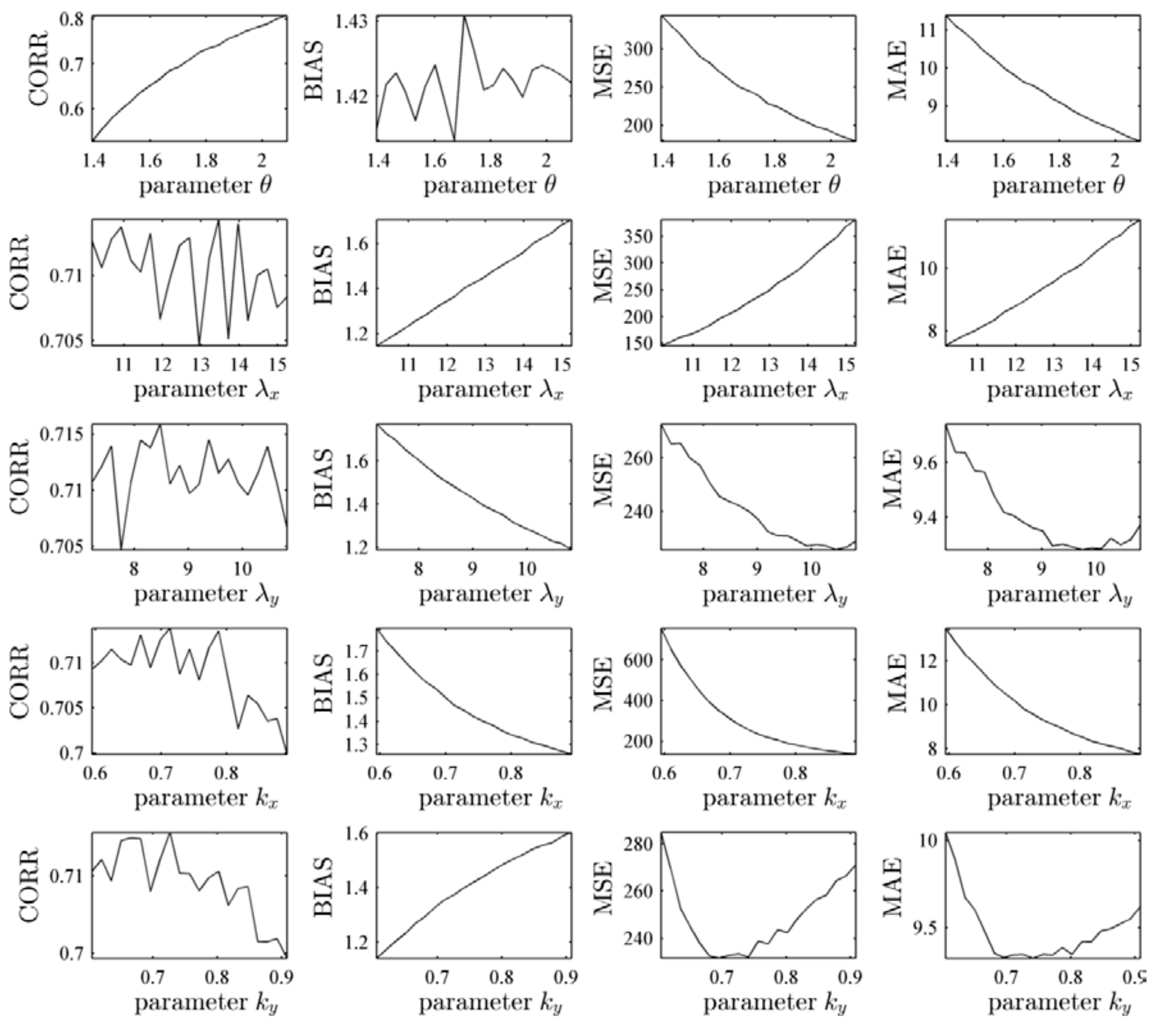

Figure 2-3 Sensitivity test on the five parameters of the joint-pdf (Eqn. (2.7)) using the four difference-based indices: CORR, BIAS, MSE, MAE

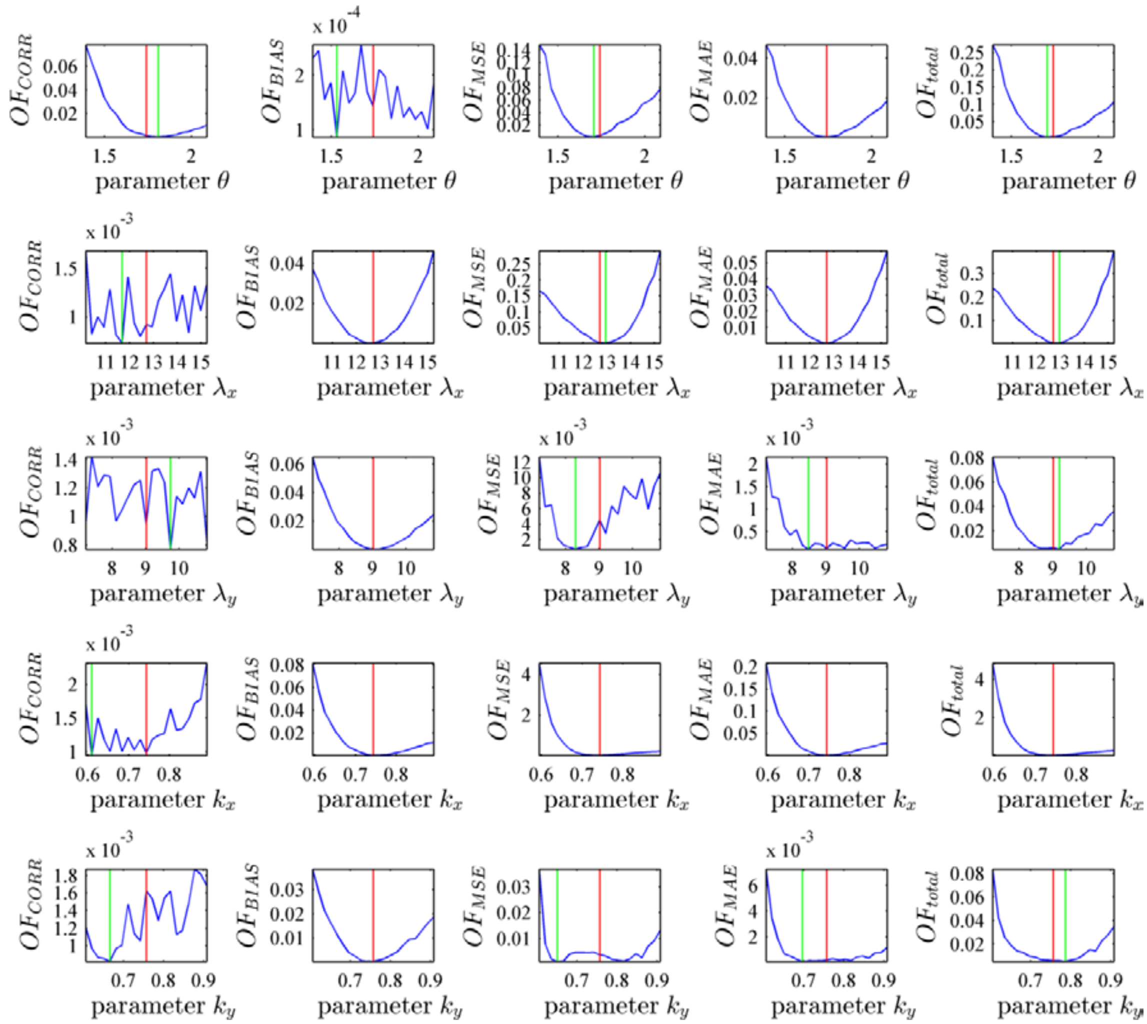

Figure 2-4 O.F. response to each of the five parameter of joint-pdf. Each row represent O.F. responses of a parameter. The first four columns are the four O.F. components in the Eqn (2.9). The last column is for the combined O.F.

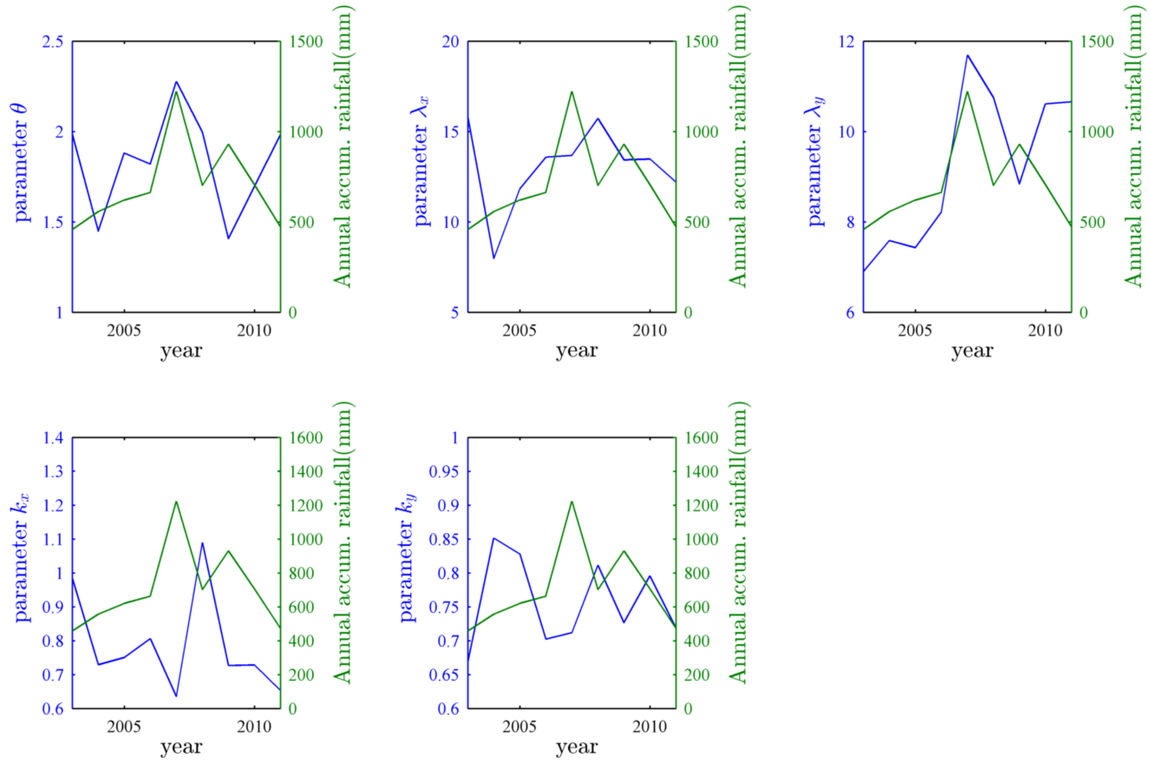

Figure 2-5 The parameters of joint-pdf for each year from 2003 to 2011, (blue lines: parameter values; green lines: annual accumulation of rainfall (mm), study case: CMORPH in Oklahoma

2.3 Results

I evaluated three sets of RS precipitation products over three study locations and during three different periods. These datasets were then used to test the validity of the joint-pdf models for each of the scenarios. I evaluated the probabilities for each of the four categories (p_{10} , p_{01} , p_{11} , p_{00}), and the total precipitation amount and total sample size over the 9-year span (Table 2.5, Table 2.6, and Table 2.7). The results show the dramatic differences in precipitation amount and the number of days with precipitation over the three locations for all three RS datasets. The Florida site received the most amount of rain over the 9 years (9800 mm ~ 12000 mm) which was detected simultaneously by both satellite and radar, with p_{11} of around 25% ~ 30% of 3100 total samples). The Montana site showed the least amount of rainfall over the 9-year period (4600 mm ~ 5300 mm), while the Oklahoma's location exhibited least number of rainy days (p_{11}=15% ~ 18% , from ~2900 samples). I should point out that the sample sizes for different satellites vary depending on the number of the non-applicable records which are discarded. I also noted that the all the RS products have problems over Montana region during cold season. In summary, these three study locations provide a reasonable range of climate variability to be modeled with the joint-pdf.

For the positive pairs, I also tested the GoFs of the copula and the marginal distributions (Table 2.2, Table 2.3, and Table 2.4), and fine-tuned the parameters of the joint-pdfs. To validate the accuracy of the joint-pdfs, 10^6 samples are of positive pairs generated and the generation is repeated for 100 times for each parameter set for estimating the correlation index and difference-based indices of the models (CORR, BIAS, MSE, MAE) (Table 2.5, Table 2.6, and Table 2.7). I compared the correlation index and difference-based indices between the observed and the model-generated data. The indices from model-generated samples fall within the $\pm 5\%$ range of those from the observed data. The results suggest that: 1) the joint-pdf models sufficiently capture the error characteristics of the satellite precipitation, as measured by difference-based indices; and 2) the differences in uncertainty associated with the location, climate, and seasons are also captured by the different parameter sets.

Table 2.5 Climate and location information about the study site in Oklahoma, total rainfall amount, total rainfall events, and its error models' parameters for the entire year, warm season(April to September) and cold season (October to March). Calibration and verification result for these models are also provided.

Koppen Climate	Humid subtropical climate								
State	Oklahoma								
Lat,Lon	34.875,-98.125								
Season	Entire year			Warm season			Cold season		
RS products	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT
p00 (%)	69.23	61.29	71.55	62.53	58.79	67.04	76.36	63.93	76.36
p01(%)	4.74	4.56	7.53	3.94	5.58	7.60	5.59	3.48	7.46
p10(%)	7.82	15.69	5.47	9.84	13.37	5.31	5.66	18.14	5.65
p11(%)	18.21	18.45	15.44	23.69	22.26	20.05	12.38	14.44	10.53
Total samples	2954	2829	2960	1524	1451	1526	1430	1378	1434
Total rainfall amount(mm)	9163.05	9425.16	6781.02	7410.89	7123.45	4662.45	1752.16	2301.71	2118.57
θ	1.74	1.50	1.66	1.92	1.45	1.66	1.76	1.71	1.53
λ_x	12.71	11.38	11.67	15.67	14.79	11.87	7.50	7.09	11.23
λ_y	9.02	8.72	10.78	8.96	9.53	11.02	8.92	7.53	10.55
k_x	0.74	0.69	0.80	0.75	0.72	0.78	0.78	0.70	0.82
k_y	0.76	0.73	0.77	0.74	0.75	0.80	0.78	0.74	0.81
CORR(obs)	0.73	0.59	0.68	0.76	0.57	0.71	0.72	0.69	0.61
CORR(sim)	0.71	0.60	0.68	0.76	0.57	0.67	0.72	0.70	0.61
BIAS(obs)	1.41	1.39	1.08	1.65	1.60	1.09	0.87	0.98	1.05
BIAS(sim)	1.42	1.36	1.06	1.70	1.61	1.09	0.85	1.00	1.05
MSE(obs)	247.29	343.11	177.89	325.10	494.32	182.33	88.60	97.68	168.89
MSE(sim.)	234.23	325.04	178.68	330.93	512.79	188.24	92.63	99.82	175.10
MAE(obs)	9.33	10.69	8.66	10.81	13.14	8.84	6.33	6.70	8.29
MAE(sim.)	9.32	10.58	8.56	10.96	13.33	8.80	6.10	6.12	8.66

Table 2.6 Same as Table 2.5, except for the study area in Montana.

Koppen Climate	Semi-arid climate								
State	Montana								
Lat,Lon	47.875,-107.625								
Season	Entire year			Warm season			Cold season		
RS products	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT
p00 (%)	57.03	44.01	55.70	47.88	47.85	52.71	70.30	38.46	57.88
p01(%)	13.89	9.92	12.65	4.89	10.68	8.32	22.85	6.67	13.55
p10(%)	9.35	22.27	10.81	13.49	13.31	8.72	4.03	35.77	16.38
p11(%)	19.74	23.80	20.84	33.73	28.16	30.25	2.82	19.10	12.19
Total samples	2974	2954	3090	756	721	757	744	780	812
Total rainfall amount(mm)	4671.72	5194.18	5310.36	2532.55	1828.75	1827.66	91.71	1183.52	1409.61
θ	1.37	1.27	1.20	1.61	1.42	1.29	1.64	1.05	1.00
λ_x	5.59	3.80	5.31	7.83	6.56	6.55	2.54	1.97	5.37
λ_y	4.71	3.77	4.85	4.45	4.29	4.62	3.39	2.18	2.97
k_x	0.72	0.67	0.77	0.77	0.75	0.79	0.81	0.65	0.94
k_y	0.73	0.68	0.79	0.82	0.76	0.79	0.76	0.85	1.07
CORR(obs)	0.50	0.44	0.34	0.65	0.54	0.44	0.68	0.09	-0.12
CORR(sim)	0.52	0.43	0.34	0.65	0.54	0.44	0.67	0.09	0.01
BIAS(obs)	1.21	1.04	1.09	1.82	1.54	1.37	0.72	1.11	1.74
BIAS(sim)	1.20	1.01	1.11	1.81	1.53	1.41	0.72	1.14	1.85
MSE(obs)	85.79	65.28	77.00	105.37	92.40	83.93	17.47	23.75	60.12
MSE(sim.)	79.15	65.31	78.95	104.68	87.50	85.42	17.47	24.27	48.38
MAE(obs)	5.47	4.72	5.42	6.22	5.88	5.59	2.90	2.85	4.84
MAE(sim.)	5.46	4.81	5.66	6.12	5.71	5.86	2.56	2.97	4.63

Table 2.7 Same as Table 2.5, except for the study area in Florida.

Koppen Climate	Humid subtropical climate
State	Florida

Lat,Lon	27.625,-82.125								
Season	Entire year			Warm season			Cold season		
RS products	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT	CMORPH	PERSIANN	3B42RT
p00 (%)	47.98	48.34	50.25	30.47	30.41	32.34	64.92	65.46	67.58
p01(%)	16.50	18.40	21.02	18.92	20.16	25.03	14.17	16.72	17.15
p10(%)	5.71	5.34	3.34	4.84	5.02	2.85	6.55	5.64	3.82
p11(%)	29.80	27.92	25.38	45.77	44.40	39.78	14.36	12.18	11.45
Total samples	3151	3016	3144	1549	1473	1546	1602	1543	1598
Total rainfall amount(mm)	10602.06	9838.86	11898.96	8474.86	7662.07	9225.69	2127.20	2176.78	2673.27
θ	1.59	1.44	1.48	1.57	1.39	1.43	2.02	1.56	1.60
λ_x	8.34	9.96	12.82	9.22	10.54	13.70	5.71	8.16	10.18
λ_y	9.07	9.56	10.63	8.97	9.65	10.40	9.03	10.50	10.76
k_x	0.68	0.81	0.81	0.72	0.82	0.87	0.58	0.69	0.68
k_y	0.86	0.76	0.91	0.92	0.85	0.92	0.72	0.73	0.78
CORR(obs)	0.65	0.53	0.57	0.63	0.50	0.53	0.76	0.63	0.66
CORR(sim)	0.64	0.56	0.58	0.62	0.52	0.54	0.79	0.64	0.65
BIAS(obs)	1.07	1.02	1.27	1.17	1.09	1.34	0.79	0.82	1.06
BIAS(sim)	1.11	0.99	1.29	1.21	1.12	1.35	0.81	0.83	1.07
MSE(obs)	151.66	184.55	227.71	162.29	177.10	222.87	118.88	210.46	243.98
MSE(sim.)	161.13	188.26	231.30	162.99	175.33	222.68	113.92	212.83	237.48
MAE(obs)	7.32	8.44	9.32	7.66	8.26	9.41	6.28	9.07	9.00
MAE(sim.)	7.66	8.80	9.75	7.79	8.73	9.88	6.30	8.93	9.39

2.3.1 Results about the Error-Intensity Relationship

According to previous studies, the shapes of the conditional pdfs of radar precipitation change across the intensity range (Gebremichael et al. 2011, Hossain and Anagnostou 2006). To see how the shape varies with increasing precipitation intensity, I plot the radar's conditional pdfs for RS estimates of 10.0 mm d^{-1} , 20.0 mm d^{-1} , 40.0 mm d^{-1}, and 60.0 mm d^{-1} (Figure 2-6) for three locations and for datasets during the whole-year. To evaluate the sensitivity of the difference-based indices as a function of rain intensity, I also calculated the mean-values and variances for these pdfs (Figure 2-6). Doing so allowed me to quantify the accuracy of the RS products as a function of rain intensity. Less difference between the mean value of the pdf and RS estimates implies less over and under estimations, and smaller variance indicates better confidence for the RS estimates.

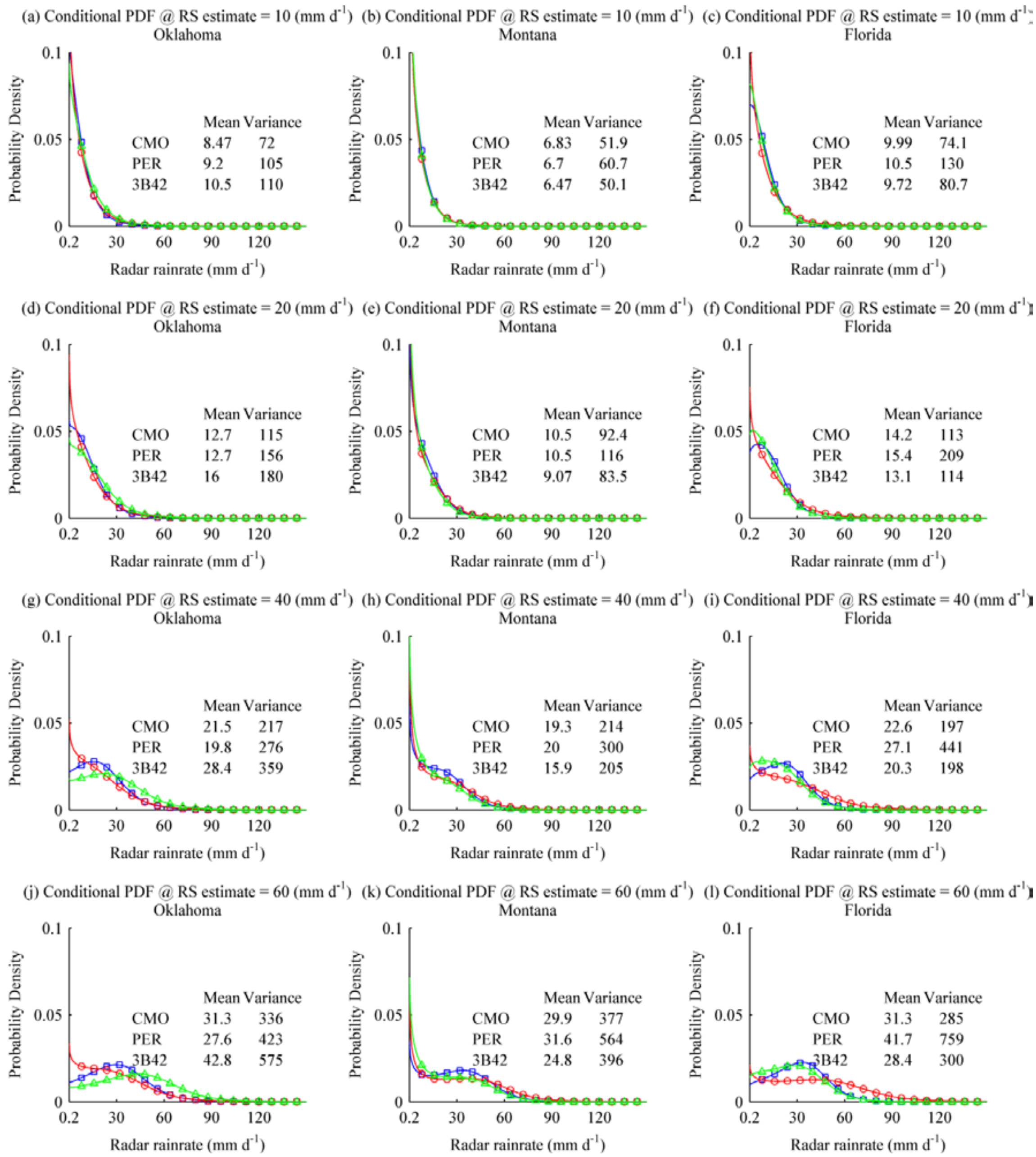

Figure 2-6 Radar's conditional distributions derived from the joint-pdfs for CMORPH (blue-square), PERSIANN (red-circle) and 3B42-RT (green-triangle) with RS estimates at rainfall intensity of 10, 20, 40 ,and 60 mm d−1 over the three study area. First, second, and third column show results for Oklahoma, Montana, and Florida region, respectively.

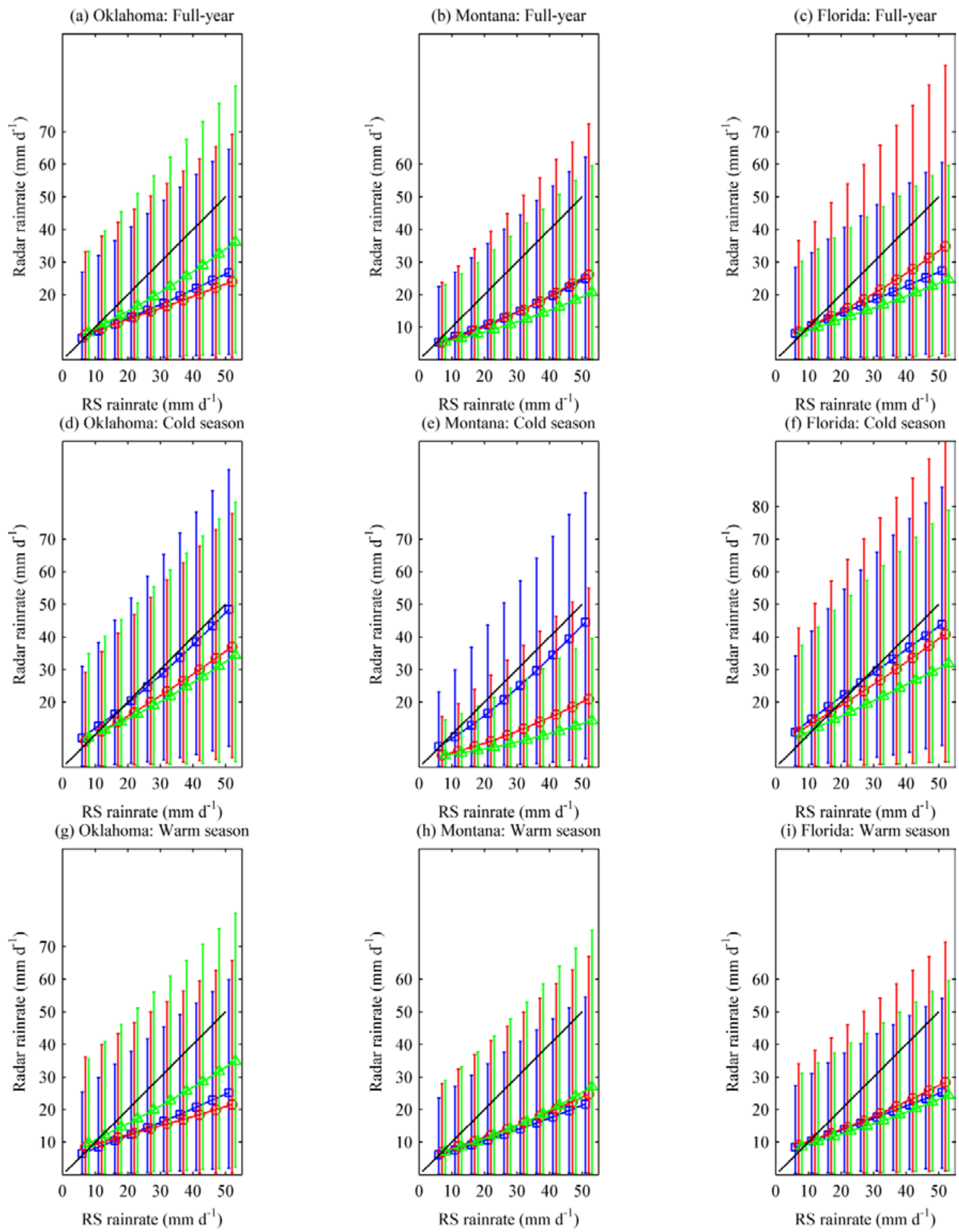

Figure 2-7 The mean value (colored curve) and the 90% CIs (colored vertical lines) derived from conditional pdf from joint-pdf for CMORPH (blue-square), PERSIANN (red-circle) and 3B42-RT (green-triangle). Each column represents different study area (Oklahoma,

Montana and Florida), while each row shows result for different season of the year (Entire-year, cold season and warm season)

The results show when the three satellite algorithms estimate rainfall at 10.0~20.0 $\mathrm{mm\ d^{-1}}$ intensity (Figures 2-6a through 2-6f), the conditional pdfs of radar measurements are similar and concentrated in the 0~30 $\mathrm{mm\ d^{-1}}$ range and skewed towards zero. At the intensity range from 40.0 $\mathrm{mm\ d^{-1}}$ to 60.0 $\mathrm{mm\ d^{-1}}$ (Figures 2-6g through 2-6l), the shapes of the conditional pdfs become bell-like but they start deviating from each other.

2.3.2 Results about Spatiotemporal Variations and Intercomparisons

To further investigate the heteroscedasticity of satellite precipitation errors as a function of intensity, CIs are estimated at various rain rates for the three products. This is achieved by obtaining the confidence range from the conditional pdfs driven from the joint-pdf of simultaneous radar and satellite precipitation values. For each individual rain estimate of a satellite product, the CI represents the range in which 90% of the corresponding radar observed intensities fall. I calculate the upper- and lower-CLs for the three RS products over three locations and three seasons listed above, at the confidence level of 90% ($\alpha = 0.10$) using Eqn. (2.15).

The results (Figure 2-7) are described separately for the three aspects considered (i.e, seasons, regions and different RS algorithms). To facilitate clarity for the interpretation of results, it is important to describe the information depicted in Figure 2-7. The bold black lines give the unbiased situation across the intensity range. The colored curves represent the expected values of the conditional pdfs across the intensity range from 5 mm d^{-1} to 55 mm d^{-1}, and they are assciated with symbols. If the colored curves are below the black line, which means their corresponding satellite precipitations generally estimate rainfall intensity higher than what ground radar does. In this case, flatter (i.e., less inclined) curve means more positively biased. The colored vertical lines represent the CIs which are associated with discretely sampled intensities (every 5 mm d^{-1}). The shorter the vertical lines are, the more confident their associated satellite precipitation is. In summary, the colored curve lines of mean values are used to evaluate how accurately these satellite precipitations *tend* to estimate, and the colored vertical lines of confidence intervals (CIs) are used to evaluate how much these estimates *disperse.*

a) Discussion of seasonal and spatial results:

In regard with variations from warm seasons (April to September) to cold seasons (October to March), each satellite product has distinct features within a season. These features are also sensitive to the regional variations (Figure 2.7d-i, Table 2.5, 2.6 and 2.7). Specifically:

1. CMORPH over-estimated somewhere between 17% ~ 65% of total precipitation in warm season and under-estimated 13% ~21% of total precipitation in cold season (Except for Montana). However, its CIs were 31% ~ 33% smaller in warm season than in cold season.
2. PERSIANN's seasonal variation depends on the region under study. It over-estimated (9%) in Florida's warm season and under-estimated (18%) in the cold season over there. In Oklahoma region, it over-estimated (60%) in warm season

and under-estimated (2%) in cold season. In both Florida and Oklahoma, it also showed 13~30% smaller CI in warm season than in cold season.

3. TRMM shows consistent performance from warm season to cold season in Oklahoma region (-4% change in bias ratio and +2% change in CI). However, over Florida it exhibited similar behavior as other RS products. It showed more bias in the warm season (34%) than in the cold season (6%).

With respect to Montana, the three RS products showed very poor correspondence with ground-based observations in the cold season. In the analytical result of observed data sets (Table 2.6), PERSIANN and TRMM had CORR of 0.09 and -0.12, respectively. In addition, CMORPH showed only 2.82-percent out of 744 days (~21 days in 9-year's winter season) when both CMORPH and radar estimate positive precipitation intensities. The number of rainy days captured by CMORPH was 86-percent less than PERSIANN and 78-percent less than TRMM during the same period. These problems are caused by poor performance of PERSIANN and TRMM during winter season over the Montana region, and CMORPH has serious detection problems as well.

b) Discussion of intercomparison among RS Products:

RS products also show distinct features in terms of their biases and CIs. Here I discuss the overall relative performance of the three RS products under various scenarios (Figure 2-7, Table 2.5, 2.6 and 2.7).

1. Over Oklahoma, CMORPH had 20-percent less CIs than TRMM and 9-percent less CIs than PERSIANN. But TRMM showed 80-percent less biased than

CMORPH and 78-percent less biased than PERSIANN over the same region. The PERSIANN outperformed the TRMM in the winter time over Oklahoma in both bias (PERSIANN's -2% vs TRMM's -13%) and CIs (PERSIANN is 3% less than TRMM).

2. Over Florida, PERSIANN performed the best in term of bias (+2%) but had the largest CI among the three products. On the other hand, TRMM had the least CI but showed the most biased (+27%) estimates among the three.

The take-home message is:

- At least during the warm season over Florida neither product showed clear superiority over the others.
- The Passive-Microwave-based CMORPH and TRMM products had (2% ~25%) less CIs in hurricane-dominated Florida region as opposed to Oklahoma (Figure 2-7d, f, g and i), while the Infrared-based PERSIANN showed (5% ~ 21%) more CIs over Florida than over the convective-storm-dominated Oklahoma (Figure 2-7, f, g and i).

2.4 Summary

The following conclusions can be drawn from my investigations:

1. The parametric joint-pdf proposed for modeling the errors from three RS products over three regions and during three seasons has proven to be an effective framework for studying the heteroscedasticity of the errors as a function of precipitation intensity. The forms of the pdfs' are similar and thus can easily be

parameterized and adapted to various situations such as different satellite algorithms, climates and seasons.

2. The joint-pdf form can be calibrated and validated using the indices such as CORR, BIAS, MSE, and MAE. After calibration, there are less than 5% differences between these indices from the simulated samples and those from the observed data sets.
3. Using the joint-pdf, one can observe that the shape of the conditional pdf shifts across the intensity ranges. At the 10~20 mm d^{-1} range, the conditional pdf is L-shaped, while in the 40 ~60 mm d^{-1} range, the conditional pdf becomes more bell-shaped. This implies the characteristics of errors in the satellite precipitation vary over different intensity ranges.
4. This study results on seasonal variations are consistent with previously reported studies such as Ebert et al. (2007), and provide additional information. In Ebert et al. (2007), the seasonal fluctuation in the accuracy of the satellite precipitation is usually discussed using CORR and RMSE. These two indices are quite sensitive to the number of rainfall events. Therefore the seasonal variation of accuracy could not be justified without compensating for the number of events. On the other hand, my parametric error model is insensitive to the number of samples, thus it compensates for the seasonal fluctuations in rainfall events. Additionally, my model provides details about the degree of heteroscedasticity of the conditional pdfs with their associated mean values and CIs. This information provides the foundation for comparative and comprehensive studies related to RS precipitations' accuracy.

5. I also draw the following broader conclusion that in general no single satellite precipitation product outperforms others with respect to the different test scenarios (i.e., seasons, regions, climates) used in the study. Similarly no individual product showed superior bias and CIs across the intensity ranges, even under the same test scenario. This conclusion captures the essence of other reported studies which show that the accuracy rankings of the various satellite precipitation products are highly dependent on the seasonal extreme precipitations, regions, elevations, and the evaluation metrics used (AghaKouchak et al. 2011, Gao and Liu 2013, Hirpa et al. 2009). Future work should focus on identifying and combining the strength of satellite precipitations for various scenarios, rather than merely providing rankings for specific scenarios or with specific evaluation metrics.

This study also has its own limitations. For one, perhaps a large number of observations which fell below 0.2mm/day were not included in the process of modeling $H_{\xi}(x, y)$. Potentially some useful information in the false-alarms $H_{X,\xi}(x)$ and missing rainfalls $H_{Y,\xi}(y)$ values were not considered. Extension of this model to incorporate small values into the joint-pdf to form a complete picture of uncertainty is the next step currently underway. Furthermore, one additional assumption built into the model is that each sample is treated independent of the other samples even though there is relatively strong dependence between satellite and radar precipitation estimates. The validity of this assumption can be questioned because of the time-dependence (i.e. seasonal variation) in the daily precipitation datasets. A possible justification in defense of the sampleindependence assumption made in my model is that I consider the seasonal variations by splitting the year into warm and cold seasons as well as considering the two seasons together as the entire year. The results are also limited to the pixel-level and spatial-results will be presented in the next-step.

In summary, my findings suggest that the proposed 2-dimensional joint-pdf can effectively quantify the conditional pdf and CI of the satellite precipitation's error under different scenarios. The proposed parametric error model can also be applied to other climate and hydrometeorological variables.

3. Framework for Assessing the Patterns in the Precipitation Dataset

3.1 Introduction

Empirical Orthogonal Function (EOF) is a powerful tool for analysis, reconstruction, and prediction of climate variables (Chen et al. 2002, Reason and Rouault 2002, Weare, Navato and Newell 1976, Lorenz 1956, Elsanabary and Gan 2013). Numerous studies have applied EOF and EOF-based methods to analyze precipitation fields and their teleconnection patterns associated to climate oscillations (Elsanabary and Gan 2013, Giannini, Saravanan and Chang 2003, Jiang et al. 2014, van Ogtrop et al. 2014, Watterson 2009). The precipitation patterns over the western US were firstly determined using EOF method (Sellers 1968). Sheaffer and Reiter (1985) displayed the dominant modes of Pacific Sea Surface Temperature (SSTemp) influencing the seasonal precipitation over the same region. Based on EOF analysis, McCabe, et al. (2004) determined a close-relationship between US precipitation and climate oceanic indices through rotated EOF. Chu, et al. (2011) demonstrates western US precipitation possesses strong spatial correlation patterns which are consistent across 50-year period and independent with spatial resolutions. The previous studies mostly focused on extracting the spatial correlation patterns (EOFs) and explaining the long-term climatic features from the smoothed (time averaged) PCs, of which many periodic patterns in the PC time series were not analyzed systematically as I will do in this study using the Nonlinear Mode Decomposition method (NMD). In the NMD analysis, the periodic patterns in the PC time series showing highly inconsistent from year to year indicate interactive impacts from climate oscillations at different scales. Because the periodic patterns can be influenced by various factors, they can be extracted from different groups of leading EOF modes. The goal is to study the periodic patterns in the leading PCs in four aspects: waveform, time period, amplitude-modulation, and corresponding leading EOFs.

3.2 Review of literatures

To improve the usability of EOF, a great deal of work applied combined wavelet transformation and EOF (Wavelet EOF, WEOF or WPCA) for the analysis and prediction of the precipitation (Mwale, Yew Gan and Shen 2004, Jiang et al. 2014, Elsanabary and Gan 2013, Mwale and Gan 2010, Kuo, Gan and Yu 2010). Wavelet Transform and wavelet analysis has been out since 1980s, and gains its popularity in geophysics in the 1990s and 2000s (Kumar and Foufoula-Georgiou 1997). Notably, Elsanabary and Gan (2013) developed a predicative tool that incorporate WPCA, Artificial Neural Networks-Genetic Algorithm (ANN-GA) and statistical disaggregation. Mwale and Gan (2010) applied WEOF and statistical disaggregation to predict river runoff in Africa. Some studies attempted to statistically predict rainfall and streamflow by using wavelet based multiple-linear regression framework (He et al. 2014, Kisi 2010).

3.2.1 WPCA-ANN-GA method

The WPCA-ANN-GA is developed by several previous work (Elsanabary and Gan 2013, Kuo et al. 2010, Mwale and Gan 2010, Mwale et al. 2004, Mwale and Gan 2005) . It is a combination of Continuous Wavelet Transform (CWT), PCA and generic algorithm calibrated- ANN (ANN-GA). It firstly applies CWT to the SSTemp and rainfall time-series to calculate the Scale-Averages Wavelet Power (SAWP), then treat SAWPs as input for EOF. The output from EOF is referred as Wavelet PCs. Wavelet PCs from SSTemp and rainfall is used to train the ANN-GA. The ANN-GA model is then used to predict rainfalls base on the input Wavelet PCs which is calculated from newly acquired SSTemps and rainfalls.

3.2.2 Wavelet-based Multiple Linear Regression model

WMLR model is a combination of wavelet-based multi-resolution analysis (MRA) and Multi-Linear Regression (MLR) developed by (He et al. 2014). It uses DWT-based MRA to decompose the rainfall and climate-index signals into a number of component time-series at different time-scales. Because rainfall is assumed to respond to climate-signals, a proper time-lag between the rainfall and climate-indices is determined. Then these components are auto-regressed with Multiple linear regression model for future rainfall.

3.2.3 Summary

WEOF method and wavelet-regression method applies firstly wavelet transform procedure and passes the output into EOF procedure or Multiple-Linear Regression (MLR) procedure, respectively. This works very well when the numbers of parameters and study grids/sites are relatively small. But they suffer from two potential problems. Firstly, the input for EOF or MLR will increase multiple times after wavelet transform. That means, when study region get larger or spatial resolution get finer, the calculation demand increase exponentially. Secondly, wavelet and wavelet-assisted procedure output remain unapprehended and are blindly forced into the regression or EOF procedure. Usually the outputs from wavelet transform remain noisy because of the nature of precipitation processes, and most part of the output is still not informative.

3.3 Methodology

After closely examine the WPCA-ANN-GA model and WMLR model, I try to propose a new model from the following aspects:

- The model should be able to handle high-dimensional dataset which contains details in spatial distribution.
- The model should be able to provide prediction with sufficient spatial details
- The model should be able to give analytical error term along with the prediction
- The model should be easily interpreted and lead to better understanding of precipitation field.

To achieve this goal, I combined the EOF method with NMD method. EOF method has been widely applied in the atmospheric science and shows its effectiveness in statistical weather prediction (Lorenz 1956, Lorenz 1965, Wilks 2011). On the other hand, recent development in signal processing realm enhances the possibility for better understanding and better modeling over the precipitation field. As mentioned before, the NMD method is chosen for its following advantages (Iatsenko, McClintock and Stefanovska 2012):

1. The method is mathematically grounded. The NMD utilizes wavelet-transform and synchrosqueezing transform (SSTrans) to extract localized frequency signals. It adaptively select the harmonics from the TFR. Thanks to the nature of wavelet-based TFR, theoretically the summation of all the harmonic components from the whole TFR domain will lead to a full reconstruction of the original signal. I only extract small number of harmonic to construct the NMs as they already represent a majority variation in the original signal. Comparatively, the Hilbert-Huang

Transform (Huang and Wu 2008) is empirical-based and cannot provide full reconstruction of the original signal. The method is adaptive and is insensitive to user-defined parameters. The algorithm of NMD adaptively extracts and validates the candidate harmonics.

2. The method is noise-robust, because the localization property of the wavelet-based TFR, the method is robust to slowly-varying amplitude and slowly-varying frequency signals. Therefore, relatively small noises are considered as locally change amplitudes or frequencies. This feature leaves the user worry-free from insignificant noises.

3. The method returns physically meaningful modes, every NM has its fundamental frequency and can be a combination of several component of related frequencies. Therefore, the NMs can exhibit complex but invariant waveforms. Comparatively, the recently developed Synchrosqueezing Transform can only extract cosine waveform (Chen, Cheng and Wu 2014, Daubechies, Lu and Wu 2011, Thakur et al. 2013), and the HHT is only able to return irregular and time-variant waveform (Huang and Wu 2008).

The proposed framework used in the study to analyze precipitation field was combining the strength of Empirical Orthogonal Function (EOF) with the Nonlinear Mode Decomposition (NMD) technique. This combined approach provides several advantages and first and foremost among them is providing the mechanism to investigate the physical meanings behind the EOF results and provide answer to questions about which phenomenon dominate each of the EOF's spatial pattern under study. The NMD method is an advanced signal-processing technique used to study Amplitude and Frequency Modulation (AM/FM) signals of complex waveform. It has been applied to separate the echo of the complex ultrasonic waves to estimate arrival time of the signal (Avanesians and Momayez 2015). Second advantage of EOF is in its effectiveness to maintain the spatial distribution details while reducing the dimensionality. The third advantage of the approach is in its ability to filter out noise and amplify signals at each step. Through the EOF, the temporal signals in the variables are usually amplified (North et al. 1982). This is especially favorable to precipitation as the observation data is usually stochastic and the signal to noise ratio is low. Amplified signals are further extracted through NMD method. In summary, the combined EOF-NMD technique effectively extracts physically-meaningful patterns and shows how these patterns contribute to the original observations.

3.3.1 Empirical Orthogonal Function method

EOF (or PCA) analysis explores a linear approximation that explains the dominant variance (both spatial patterns and temporal variations) in a space-time array of data, $\mathbf{X}_{M\times N}$ here, M stands for the number of spatial variables (i.e., the number of gauges in this case) and N for the number of samples in each variable (i.e., the number of rainfall measurements at each gauge in this case). The goal of EOF analysis is to transform $\mathbf{X}_{M\times N}$ into the spatial correlation mode matrix $\mathbf{E}_{M\times M}$ with the i^{th} column vector as ranked EOF modes $\mathbf{e}^i_{M\times 1}$ and the PC matrix $\mathbf{U}_{M\times N}$ with the i^{th} row vector $\mathbf{u}^i_{1\times N}$ as the corresponding time series,

$$\mathbf{X}_{M\times N} = \mathbf{E}_{M\times M}\mathbf{U}_{M\times N} = \sum_{i=1}^{M}\mathbf{e}^i_{M\times 1}\mathbf{u}^i_{1\times N} \quad (3.1)$$

Usually, only the first m rows in $\mathbf{E}_{M\times M}$ (the leading EOFs) and the first m columns in $\mathbf{U}_{M\times N}$ (the leading PCs) are sufficient to represent the dominant systematic spatiotemporal variance in $\mathbf{X}$,

$$\mathbf{X}_{M\times N} = \sum_{i=1}^{m} \mathbf{e}^{i}_{M\times 1}\mathbf{u}^{i}_{1\times N} + \boldsymbol{\varepsilon}_{M\times N} \tag{3.2}$$

Where, the residue $\boldsymbol{\varepsilon}_{M\times N}$ contains the truncated (insignificant) variance and random noise. Given the data matrix X, the EOF matrix $\mathbf{E}$ and PC matrix $\mathbf{U}$ can be solved.

3.3.2 Nonlinear Mode Decomposition method

The physically meaningful periodic patterns in an oscillatory time series are usually not simple sine or cosine waves but have more complicated waveforms. For example, a cubed cosine-wave may be an interesting waveform, which can be decomposed in the frequency domain with two components,

$$[A(t)\cos(\varphi(t))]^3 = \frac{3}{4}A^3(t)[\cos(\varphi(t)) + \frac{1}{3}\cos(3\varphi(t))] \tag{3.3}$$

where waveform which is nonlinearly transformed from sine/cosine function is referred as a *nonlinear mode*. NMD method defines the full *Nonlinear Mode* $v(\varphi(t))$ (NM) as a sum of all corresponding component to the same *base frequency* ($h \in \mathbb{N}^+$):

$$c(t) = A(t)v(\varphi(t)) = A(t)\sum_{h} a_h \cos(h\varphi(t) + \varphi_h) \tag{3.4}$$

The NMD method can handle a signal that is a combination of a number of NMs $c_j(t)$ plus noise $\eta(t)$, ($j \in \mathbb{N}^+$):

$$s(t)=\sum_j c_j(t)+\eta(t) \tag{3.5}$$

The main goal is to extract a given signal into a set of NMs. The procedures of NMD method is summarized as below:

(a) Manually or adaptively decide which Time-Frequency Representation method (Wavelet transform or Windowed-Fourier Transform) is best suited for the input signals

(b) Extract a fundamental harmonic of an NM from the signals TFR ;

(c) Find all possible candidate for this NM;

(d) Indentify the true harmonics among the candidates;

(e) Construct the full NM by summing together all the true harmonics;

Subtract the NM from the signal, and iteratively return the residual as input for step (b), until no more fundamental harmonics are extra.

3.3.3 Combined EOF-NMD method

By incorporating the Equations together, I have,

$$\begin{aligned}\mathbf{X}_{M\times N} &= \sum_{i=1}^{m}\mathbf{e}^i_{M\times 1}\mathbf{u}^i_{1\times N}+\boldsymbol{\varepsilon}_{M\times N}\\ &= \sum_{i=1}^{m}\mathbf{e}^i_{M\times 1}\left(\sum_j^n c_j(t)+\eta(t)\right)^i_{1\times N}+\boldsymbol{\varepsilon}_{M\times N}\\ &= \sum_{i=1}^{m}\mathbf{e}_i\left(\sum_j^n A_{ij}(\mathbf{t})\left(\sum_h a_{ijh}\cos\left(h\varphi_{ijh}(\mathbf{t})\right)\right)+\eta_i(\mathbf{t})\right)+\boldsymbol{\varepsilon}_{M\times N}\end{aligned} \tag{3.6}$$

Here, $A_{ij}(\mathbf{t})$ denotes the time-varying amplitude for each NM, and $\varphi_{ij}(\mathbf{t})$ denotes the time-varying phase. In the following section, I will show the EOF result of orthogonal functions as $\mathbf{e}_i \in \mathbb{R}^{M\times 1}$ and show how to find $A_{ij}(\mathbf{t})$ and $\varphi_{ij}(\mathbf{t})$ to represent the PCs. I also calculated instantaneous frequency of NM:

$$f_{ij}(\mathbf{t}) = \frac{\varphi_{ij1}(\mathbf{t}+\Delta t) - \varphi_{ij1}(\mathbf{t})}{2\pi \cdot \Delta t} \tag{3.7}$$

3.4 Case study over southwestern US

3.4.1 Data

3.4.1.1 Precipitation data

The precipitation dataset used for this study is the Climate Prediction Center's Unified Gauge-Based Analysis of Daily Precipitation over the Contiguous United States (CPC gauge data). The CPC gauge data has a continuous spatial grid of $0.25° \times 0.25°$ degrees at daily temporal resolution from 1948 to the present. The data is derived from over 13,000 daily rain gauges reported across the contiguous United States. Quality control procedures including ground-based radar precipitation correction from the NEXRAD array as well as satellite corrections from the GOES 8 and GOES 10 IR data.

I choose the Western US (mainly California and Nevada) as the study area. The previous study which applied EOF over California has shown its precipitation is dominant by elevation and pacific decadal pattern (Cayan 1996, Knowles, Dettinger and Cayan 2006, Sheaffer and Reiter 1985).

The CPC gauge dataset is preprocessed for the purposes of EOF and following NMD approach. The first step is to remove the grids over the oceans. Then I rearrange the datasets into a two-dimensional matrix, the matrix has row vectors as different locations (variable), and has column vectors as measurements of daily rainfall accumulation in each day of the study period.

3.4.1.2 Climate indices

To explore the interconnection between the El-Nino and Southern Oscillation (ENSO), Oceanic Nino Index, Southern Oscillation Index, Nino 4 and Nino3.4, Pacific Decadal Oscillation (PDO) are selected in this study. Descriptions of these climate indices are documented in NOAA Climate Prediction Center website (http://www.esrl.noaa.gov/psd/data/climateindices/).

3.4.2 Result

3.4.2.1 El-Nino and PDO signals in the precipitation data.

Precipitation process is considered as an output from atmosphere-ocean-land interaction. These interactions follow several stationary or non-stationary oscillations. Therefore, spatial and seasonal patterns should be reflected along with their own frequencies in the precipitation filed. Those patterns are extracted by applying the EOF-NMD method to the large-scale and long-term precipitation measurements.

Therefore, I analyzed the leading PC time-series with NMD method and extracted amplitude-modulated (AM) $A_{ij}(\mathbf{t})$ and frequency modulated (FM) $f_{ij}(\mathbf{t})$ signals. The variance explained by first 40^{th} EOFs along with the first four EOFs' spatial patterns, and the first four EOFs already explained 90 percent of the monthly data's total variance (Figure 3-1). Then the spatial patterns $\mathbf{e}_i$ (Figure 3-1) and their corresponding PCs (Figure 3-2) are obtained. Based on Eqn (3.6), the first three periodic curve out from each leading PC (Figure 3-3, Figure 3-4, and Figure 3-5), along with their time-varying amplitudes $A_{ij}(\mathbf{t})$, monotonic-increasing phases $\varphi_{ijh}(\mathbf{t})$ of the oscillation, and time-varying frequencies. (Figure 3-6, Figure 3-7, Figure 3-8).

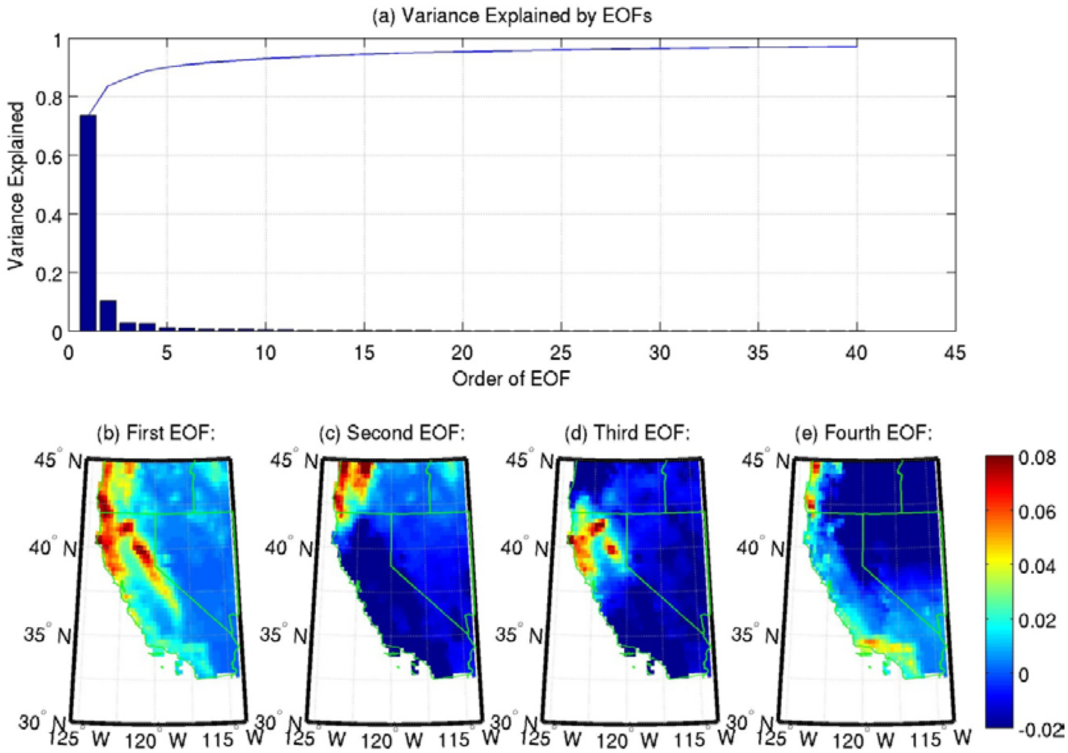

Figure 3-1 (a) the bar graph shows variance explained by each of the EOF in descending order, the line graph shows the accumulative variance explained by these EOFs. (b)~(e) gives the spatial patterns of first four EOFs.

I am going to interpret the result from the following aspects:

(a) The dominant frequency of the curves in each PC.

(b) The AM signals in each PC.

(c) The dominant waveform in each PC.

(d) The EOF spatial patterns and the variance explained.

3.4.2.2 The frequency in each PC

The frequency information about the PC is revealed after the NMD analysis. The frequency of first extracted curve from the first PC ($f_{11}(\mathbf{t})$) is fluctuating within a range of [0.95,1.05] (Figure 3-6). Therefore, the first PC is exhibiting an annual cycle. The frequency of first extracted curve from the second PC ($f_{21}(\mathbf{t})$) is within the range of $[1.5, 2.5]$, and 70-percent of the time it remains in the range of $[1.8, 2.2]$ (Figure 3-7). Then the frequency is half-annual cycle, or twice every year. The frequency of first extracted curve from the fourth PC ($f_{41}(\mathbf{t})$) is within the range of [0.8,1.1] (Figure 3-8), therefore the fouth PC has an annual cycle. The third PC, however, which does not have a stable frequency, will be discussed in the following section.

3.4.2.3 The AM signals in the dominant periodic curves from each PC

The first curves from the first PC, second PC and fourth PC show significantly stable periodicity, therefore these PCs are dominant by periodic oscillation (annually or bi-annually), and their amplitudes and waveforms could retain important information about dominant climate signals. The AM signal in PC-1's first curve ($A_{11}(\mathbf{t})$) has shown high correlation (R=0.41) with the PDO index (Figure 3-9). The AM signal in PC-4's first curve ($A_{41}(\mathbf{t})$) has shown significant correlation (R=0.56) with the Nino 3.4 index (Figure 3-10).

3.4.2.4 The waveforms in the dominant periodic curves from each PC

The dominant waveforms from the periodic PCs reflect the oscillation patterns in which the western US rainfall follows. Valley-shaped waveform from the first PC's $c_1(\mathbf{t})$ (Figure 3-11) has the period of 1-year. The valley reaches bottom at July-August, and rise to peak at December-January. W-shaped waveform from the second PC's $c_2(\mathbf{t})$ (Figure 3-12) has the period of 1-year. The $c_2(\mathbf{t})$ waveform has two lowest point located in the March and September, and has two highest point in the June and December. A complex waveform is extracted from the fourth PC (Figure 3-13). The complex waveform has two peak in the March and August and reached to the bottom at November.

3.4.2.5 The spatial patterns in each EOF

Besides the periodic curves' frequencies, and waveforms, the EOF spatial distribution is also static to time-varying PC's coefficient. I focus on where the highest coefficients occur in the EOF and look for gradient of the coefficient. In the Figure 3-1, the coefficient value of EOF is represented in gradient color. High-values of coefficients are represented in red or yellow, whereas low-values of coefficients are presented in blue or dark. First EOF (explained 74% variance) has a high-coefficient region over the mountainous area, and has a northwest to southeast gradient (Figure 3-1b). Second EOF (explained 9% variance) has a high-coefficient region in the Oregon state (Figure 3-1c), the impact gradually decrease from north to south. Third EOF (explained 4% variance) has an high-coefficient region in the Northern California (Figure 3-1d), the spatial coefficient is quite intense in San Fransisco Bay area and radial gradient from center to nearby region. Fourth EOF (explained 3% variance) has high-coefficient region in the Coastal line (Figure 3-1e), the spatial coefficient is quite intense in coastal region and decrease from ocean to inland.

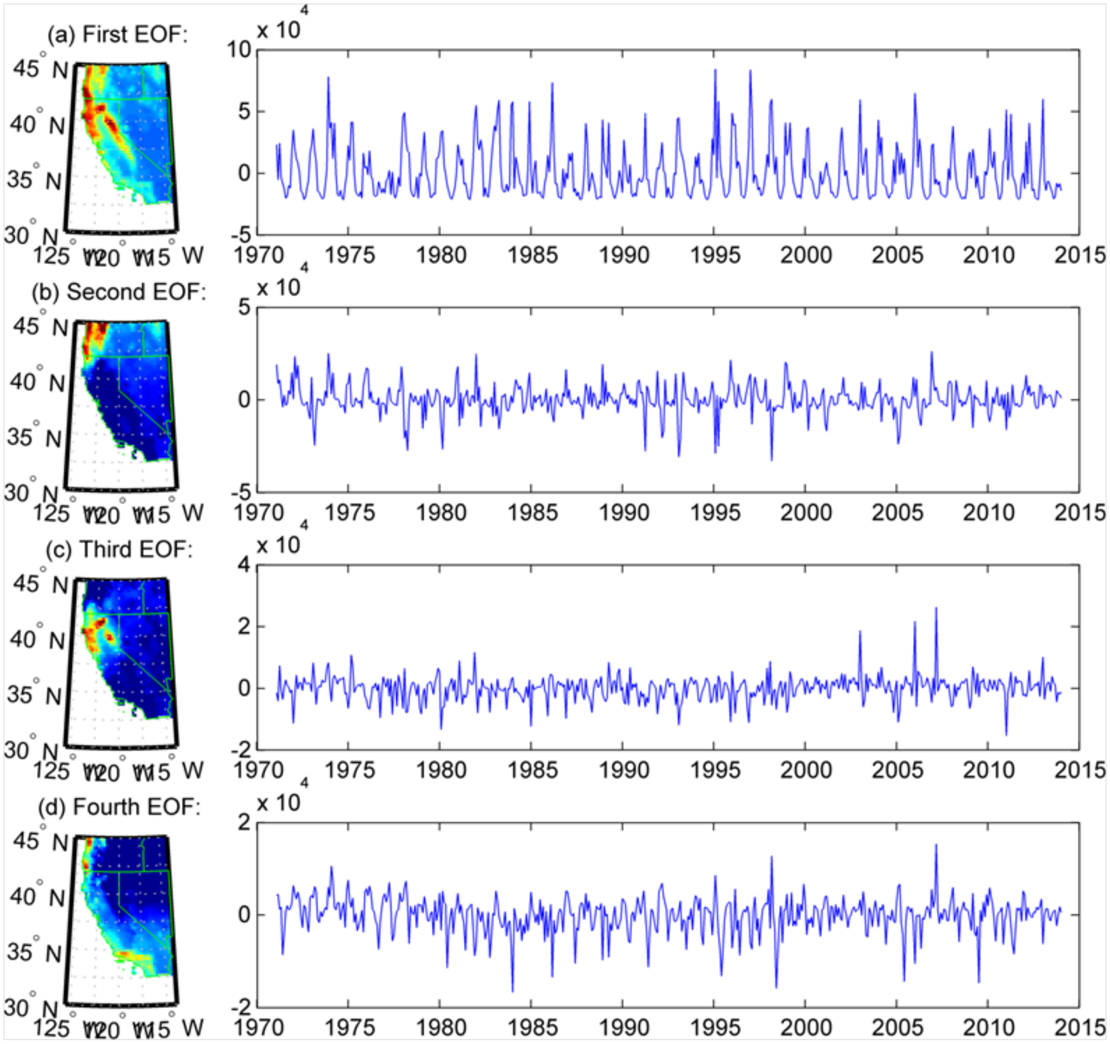

Figure 3-2 The first four EOFs with their corresponding PCs' time-series

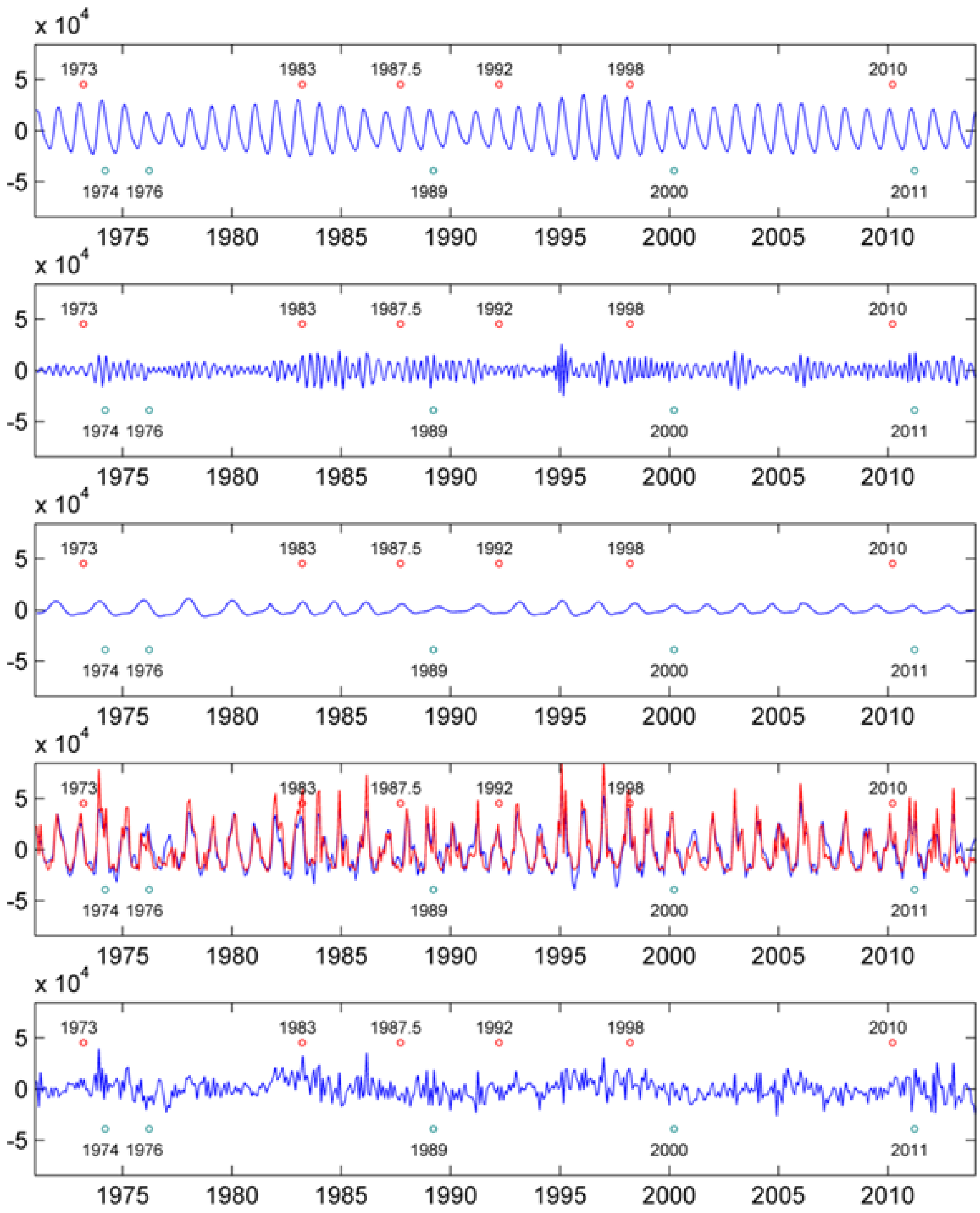

Figure 3-3 The first PC is analyzed by NMD method, row 1-3 : extracted three oscillatory curves from the first PC, row 4: the summation of all the three curves (blue) compared with original PC (in red), row 5: the difference between the blue line and red line in row 4. Historical El-Niño years (red dots) and La-Niña year (green dots) are annotated above or below the curves.

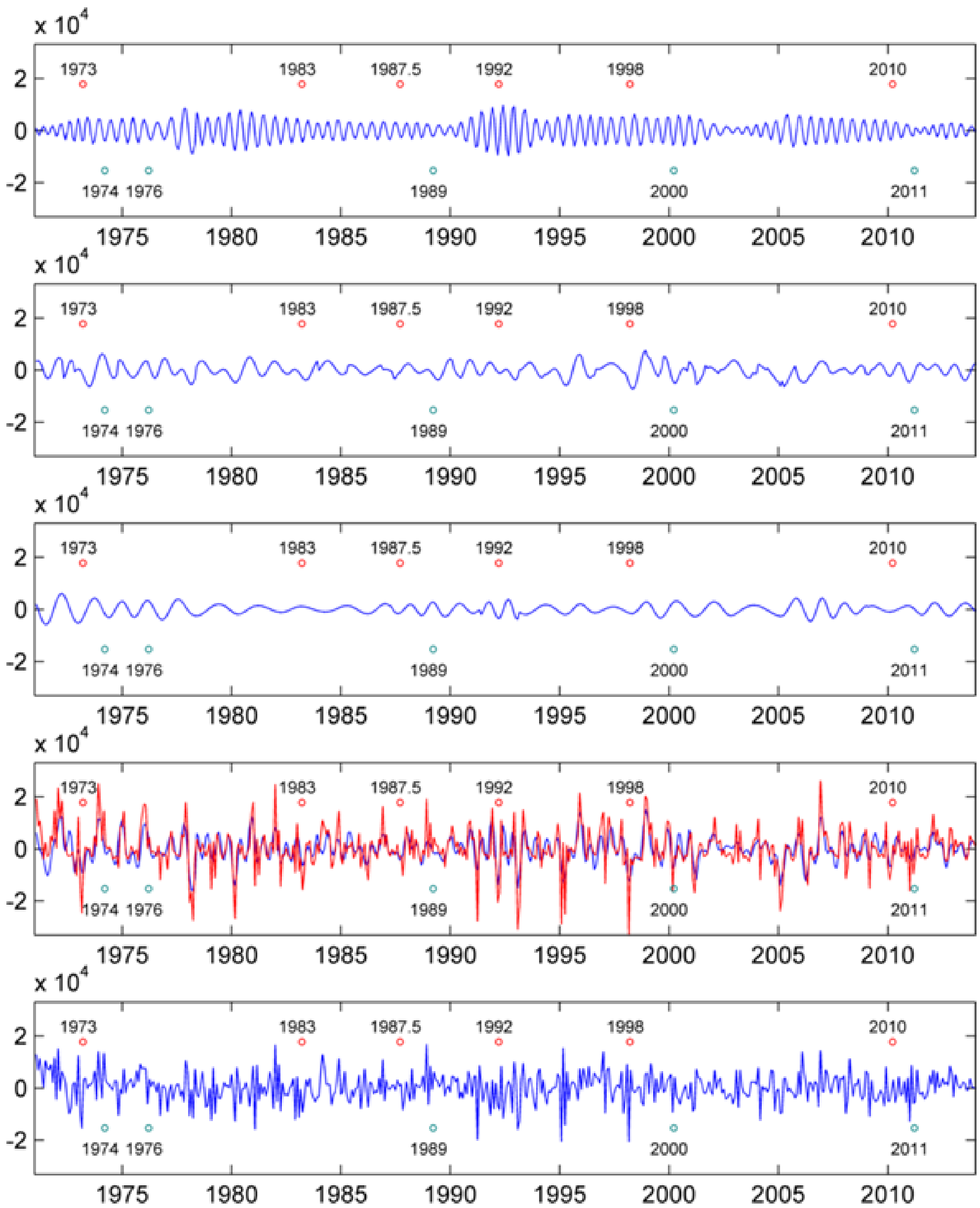

Figure 3-4 Same as Figure 3-3, except for second PC

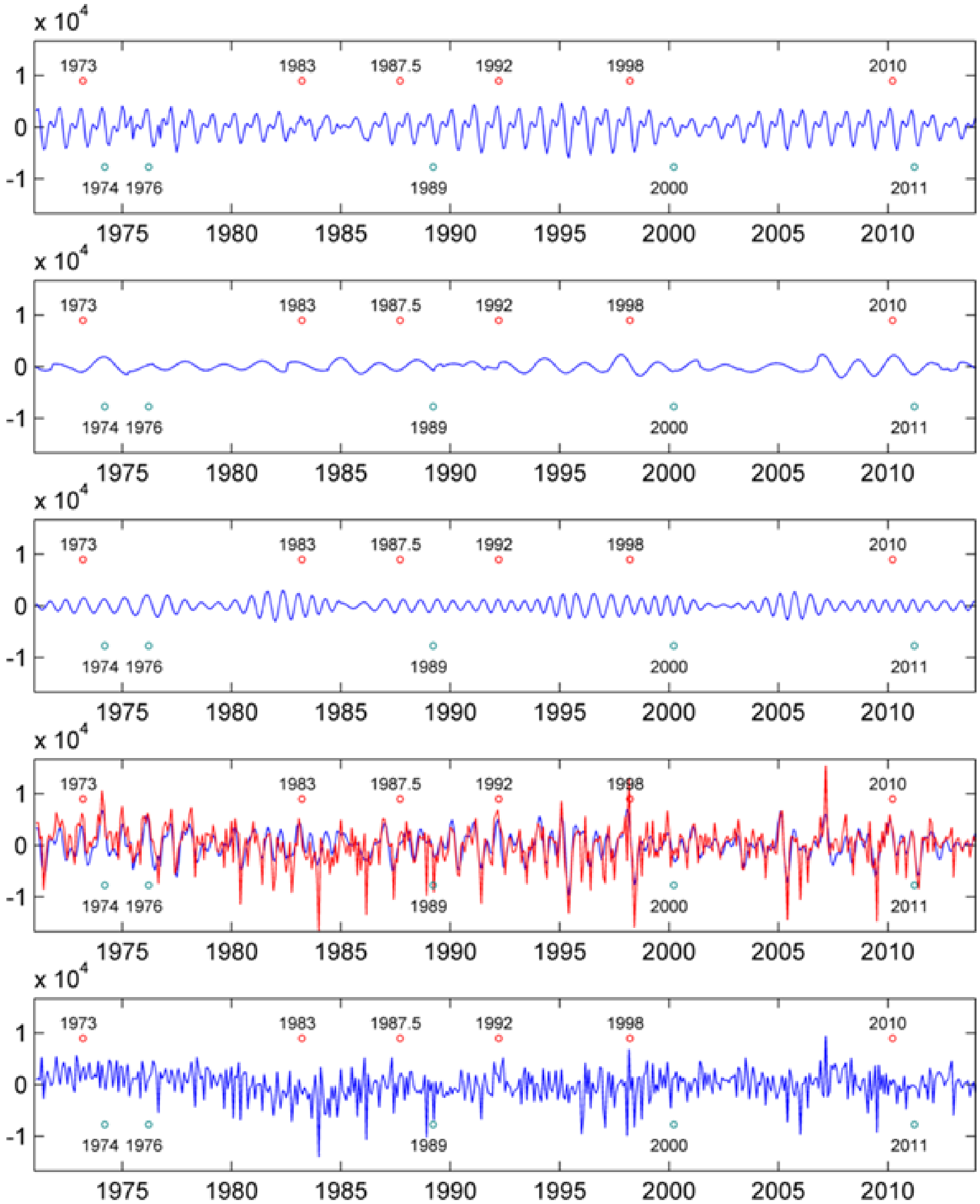

Figure 3-5 Same as Figure 3-3, except for fourth PC

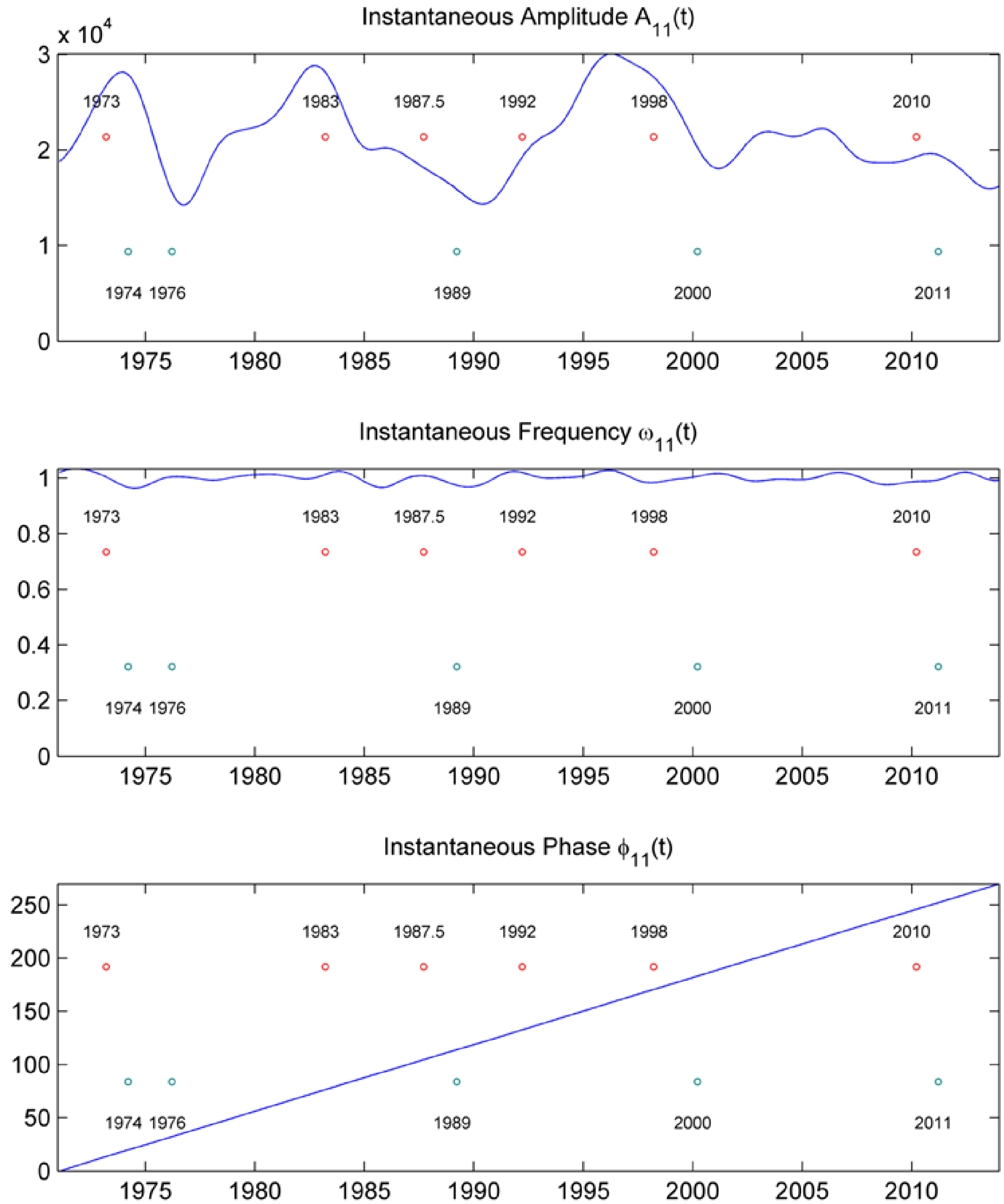

Figure 3-6 (a) Instantaneous Amplitude of first oscillatory curve which is extracted from first PC, (b) instantaneous frequency of the same curve. Historical El-Niño years (red dots) and La-Niña year (green dots) are annotated above or below the curves.

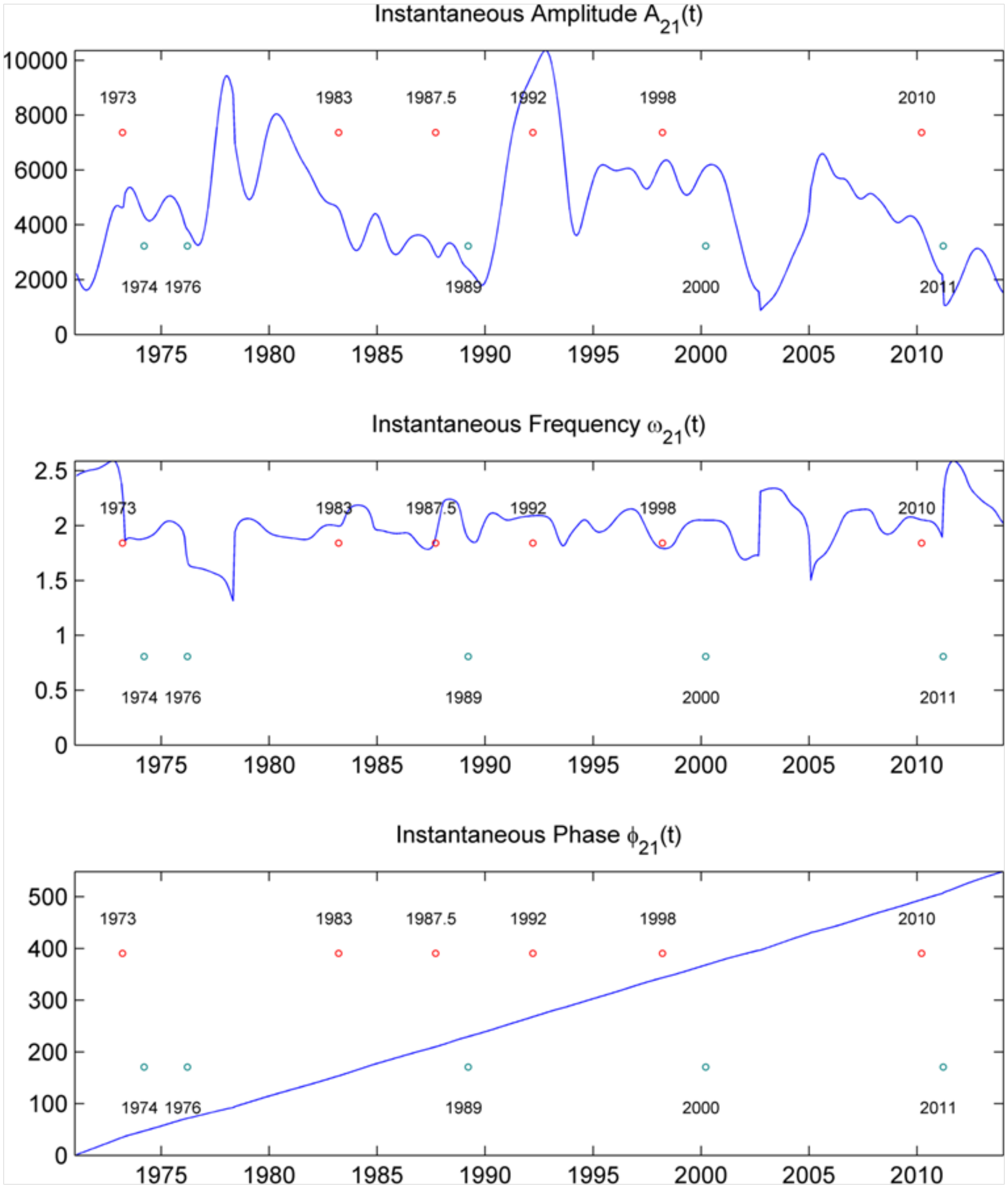

Figure 3-7 Same as Figure 3-6, except for the second PC.

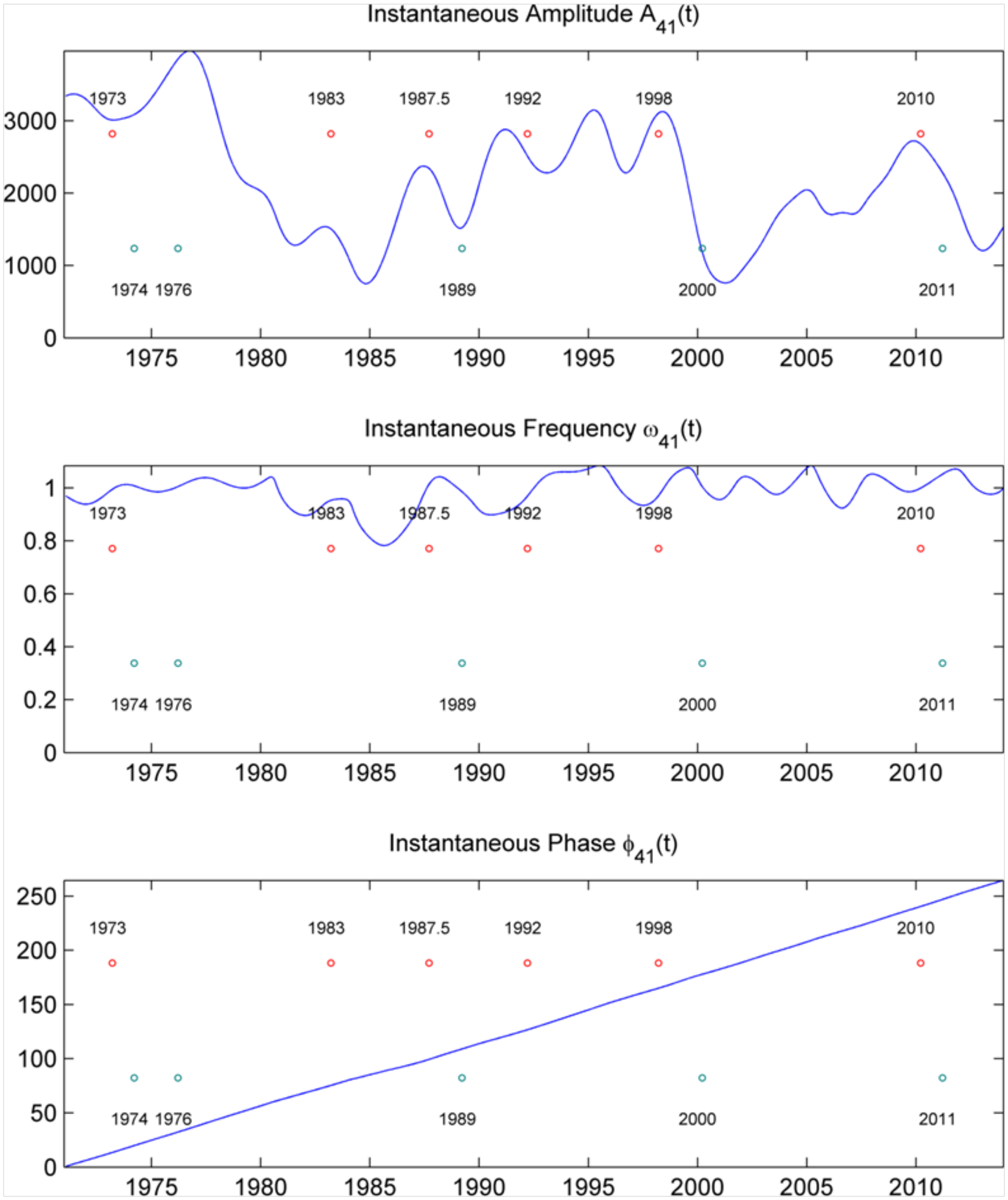

Figure 3-8 Same as Figure 3-6, except for the fourth PC.

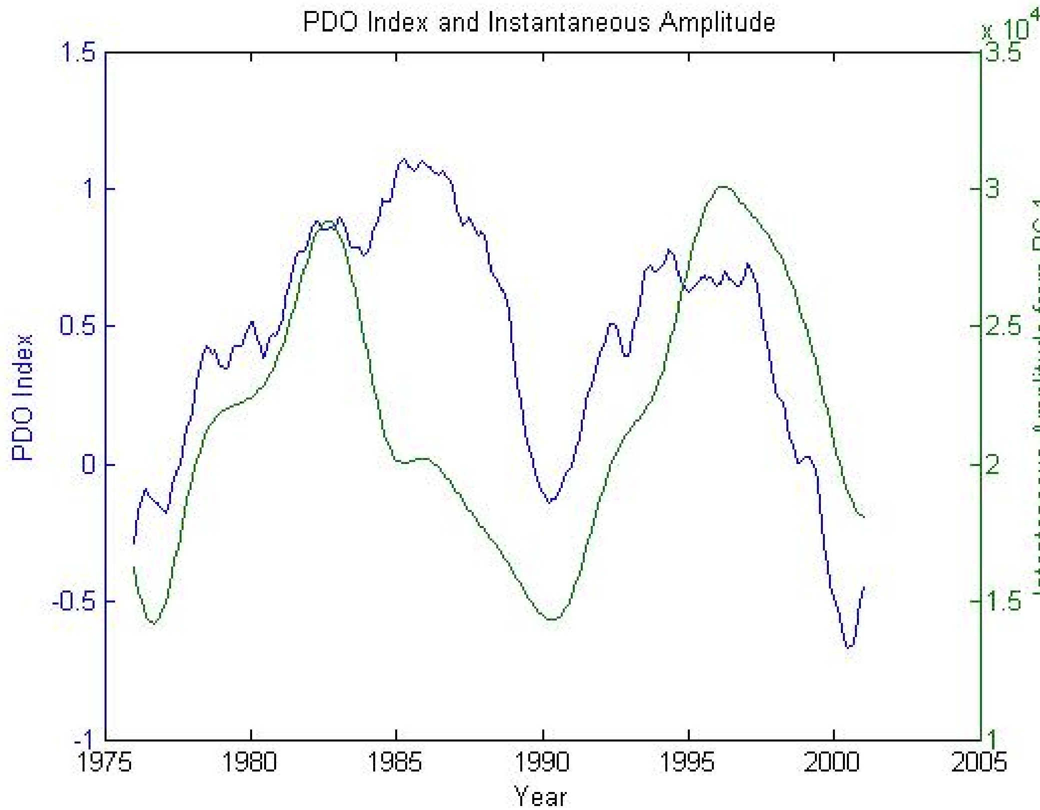

Figure 3-9 The PDO index 13-month running average and Instantaneous Amplitude of first curve from PC-1 during the 1976~2001 year period. (R=0.41)

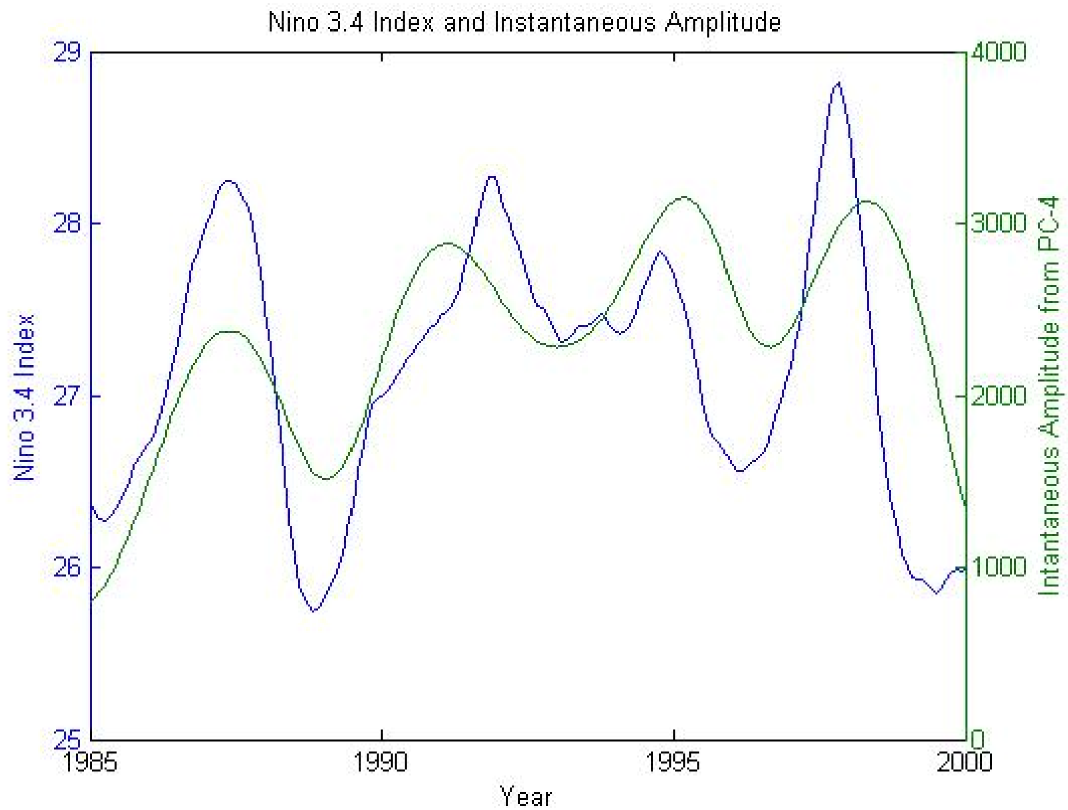

Figure 3-10 The Nino3.4 index 13-month running average and Instantaneous Amplitude of first curve from PC-4 during the 1985~2000 year period. (R=0.56)

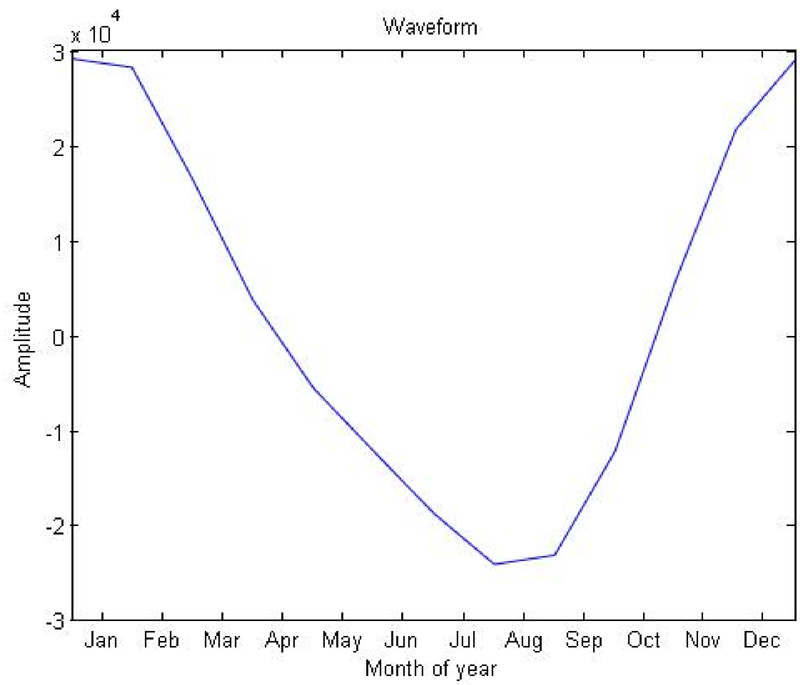

Figure 3-11 Waveform of the first curve extracted from PC-1

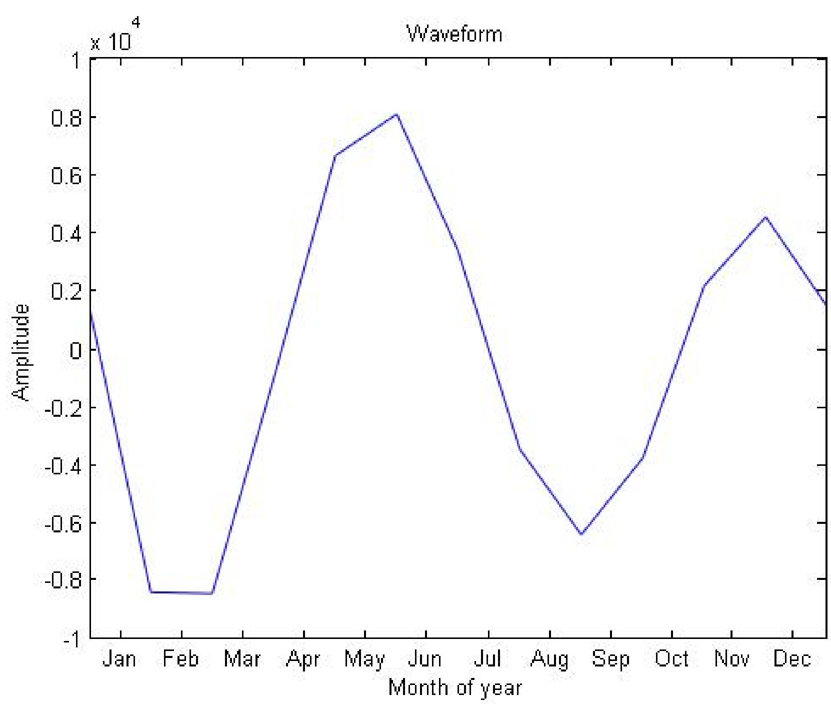

Figure 3-12 Waveform of the first curve extracted from PC-2

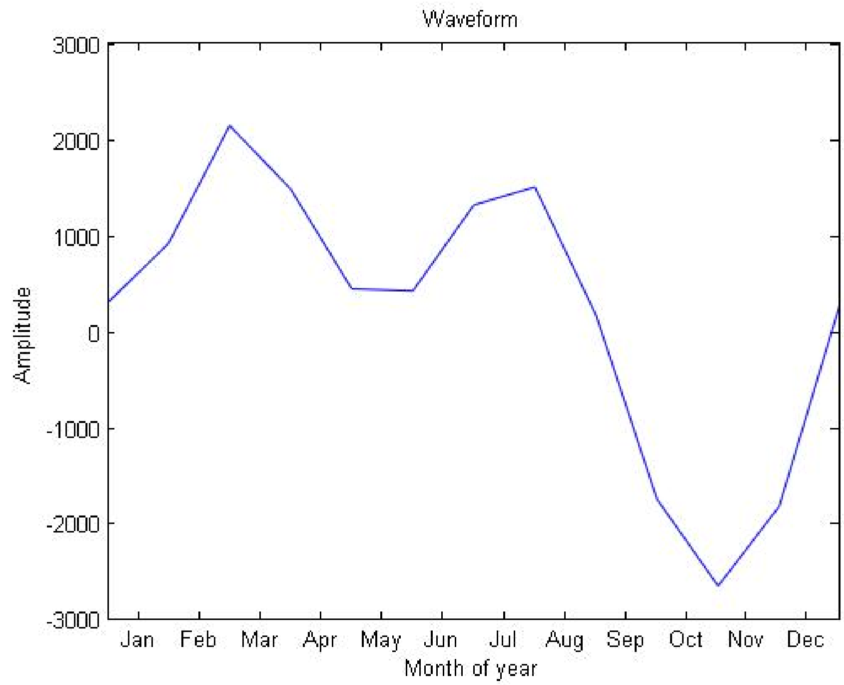

Figure 3-13 Waveform of the first curve extracted from PC-4

3.5 Discussions

I introduce EOF-NMD method for studying the dominant spatial patterns and seasonal patterns in space-time array of precipitation measurements. As a case study, I applied EOF analysis to long-term, ground-observed precipitation dataset over western US, and then applied NMD method to analyze the leading PCs from the output of EOF and extracted essential seasonal patterns which are have significant correlation with several climatic indices. I summarize my finding as follows:

1. The leading four EOFs and PCs explain more than 90-percent of the total variance in the space-time array of precipitation. And the NMD method can retrieve three significant oscillating signals in the leading PCs.

2. First EOF and PC represent the impact from Pacific Decadal Oscillation and annual cycle. Because its PC has 1-year period and its AM signals is significantly correlated with PDO index (R= 0.41), while its EOF spatial patterns reflect a mountainous texture and northwest-to-southeast gradient.
3. Second EOF and PC represent the impact of Southwest United States monsoon, because their waveform has a drastic decrease from June to August, which is corresponding to the dry June (first peak of the waveform) and rainy July (second valley of the waveform), followed by a dry season after mid-September (second peak of the waveform).
4. The third PC does not exhibit stable frequency, but the PC-3 account for 3-percent of the total variance. The third PC is dominated by the atmospheric river events because of the third EOF's pattern show high impact region overlap with frequent atmospheric river's route. Its pattern has very concentrated high-coefficient region near the San Francisco Bay area. The region also shows a clear radial gradient. The atmospheric river event is not stably periodic, but often create large amount of rainfall over its impact region.
5. The fourth EOF and PC exhibit a clear impact from ocean to land. The dominant periodic pattern in fourth PC is significantly correlated with El-Nino index (R=0.56). The fourth leading spatial pattern implies the impacts are from the ocean to land. The effect of El-Nino event over the western United State is very negligible because this PC account for about 3% of the total variance. Kunkel and Angel (1999) reported that during La-Nina year the northwest US has 10% more snowfall in the winter season, but stated the anomaly is not significant.

Our study provides mathematically-grounded evidences to several works that attempt to relate Western US's precipitation to ENSO and PDO. Redmond and Koch (1991) investigated the precipitation patterns over the western US with split sample analysis, and pointed out there is a strong correlation between season-averaged SOI from July to November and precipitations in October-March. My findings provide more details and supporting evidence to their conclusion.

The EOF-NMD, when considered as a whole, is a transformation for a space-time array which emphasizes the original spatiotemporal variability. In the EOF step and the NMD step, the random parts of the variation are filtered out and the dominant spatial/temporal variation patterns are preserved. Because EOF and NMD's mathematical nature, the EOF-NMD can be applied reversely and those dominant spatial and seasonal patterns can reconstruct majority of the spatiotemporal variability. Also, the amount of unexplained random variation can be quantified. Future work will combine EOF-NMD and auto-regression techniques for short-term climate forecasting.

4. Conclusions and Future Directions

4.1 Conclusions

Precipitation process, in its micro-scale, exhibit randomness and need a statistical model to capture its heteroscedasticity. In its macro-scale, it has patterns in its spatiotemporal variability and need factorization and signal-processing approach to extract those spatial and temporal patterns. In the thesis, I used a joint-pdf to study the error's heteroscedasticity of the RS precipitation. I also used EOF and NMD approach to study the multi-scale spatial patterns and temporal patterns. With these two approaches, the motivating questions of this thesis are fully or partially addressed:

1. ***How to account for the heteroscedastic (variance changing) error in the RS precipitation?***

For the study on high-resolution RS precipitation, I find the error models in the previous studies heavily rely on restraining assumptions. These assumptions, from symmetric error distribution to constant error distribution, are not verified. Then I proposed a joint-pdf model that relies on a much loosened assumptions: "*The precipitation products are assumed to follow some stationary random processes and to have stationary dependence with each other*". Driven by this assumption, a copula-based joint-pdf is proposed as the error model. In the joint-pdf model, the error and its heteroscedasticity are implicitly expressed. This is different from previous studies that tend to explicitly express the error and its conditional distribution. In this way, I revealed the changing error distribution across the intensity. *I found that the shape of the conditional pdf shifts across the* *intensity ranges. At the 10~20 mm d^{-1} range, the conditional pdf is L-shaped, while in the 40~60 mm d^{-1} range, it becomes more bell-shaped.*

2. ***How to account for the spatiotemporal variability while describing the heteroscedasticity of RS precipitation.***

After I verified the adaptability of the joint-pdf model, I evaluated three sets of RS precipitation products over three study locations and during three different periods. I found under different scenarios, these RS precipitation behave differently. Their performance rankings in bias and dispersion are inconstant under different seasons or different regions. Although these results only limit to pixel level and only three regions are studied, it is enough to show the complex variability in the error of RS precipitation error. *It becomes evident that no single satellite precipitation product outperforms others with respect to the different scenarios (i.e., seasons, regions, climates).*

3. ***What are the distinct characters of RS precipitation? What makes one RS product different from another RS product? How to qualitatively or quantitatively describe these differences.***

The characters of RS precipitations are the characters of their errors. Given the fine resolution of RS precipitation, their errors are heteroscedastistic. That means the error's conditional variances are changing with the intensity. I used errors' conditional mean values and conditional confidence intervals to quantatively measure the errors' characters. *The results suggest the dominant problems in high-resolution RS precipitations are their large dispersions, not their biases. Therefore, bias-correction is not recommended as the first priority in the current state of the RS precipitation study.*

4. ***Is precipitation process completely random? Are there any distinct characters in each of the precipitation datasets?***

For the study on spatiotemporal variability in the precipitation data, I found many of the previous studies focus their study on certain season or months. These segmentations on original data assume only part of the data is teleconnected with oceanic or atmospheric change. However, it is not well-expained why certain seasons or months are chosen for study. Therefore, very limited evidences are provided for the hypothesis of "*There are spatial and temporal variations in the precipitation datasets. These variations are result from climatic processes including interactions among land, ocean and atmosphere. Because precipitation is physical result of the climatic processes, the signal of these processes can be separated from the precipitation datasets*".

To overcome the limitation of previous study, I used EOF-NMD approach to process the whole observed long-term data without segmentation. Amplitude-modulated seasonal variations are revealed by this approach. I found precipitations exhibit seasonal patterns and spatial patterns by studying the long-term and monthly gauge data over western United States. That means the precipitation process, in its macro-scale, is auto-correlated in space and time. By proving this hypothesis using mathematically stable approaches (i.e., EOF and NMD), people can investigate the teleconnections between precipitation and oceanic/atmospheric oscillations in a much finer resolution than previous studies.

These patterns and teleconnections, hidden in the precipitation datasets, are the defining characters of these datasets.

5. ***Which climate factors are impacting the precipitation and how much they contribute?***

By applying the EOF-NMD method to Southwestern US, several patterns can be extracted and analyzed. The results show that part of the dominant climate factors are PDO, Monsoon and El-Nino, whereas the other factors are still under investigation and are yet to be confirmed.

4.2 Future directions

I believe the two frameworks will provide promising improvement for the RS precipitation, as well as enhanced understanding of the precipitation process itself. In the shadow of these two goals, my study has several limitations. I suggest the following possible directions for future research:

Incorporating the false-alarm and missing part into the joint-pdf model

Perhaps a large number of observations which fell below 0.2mm/day were not included in the process of modeling $H_{\xi}(x, y)$. Potentially some useful information in the false-alarms $H_{X,\xi}(x)$ and missing rainfalls $H_{Y,\xi}(y)$ values were not considered. Extension of my model to incorporate small values into the joint-pdf to form a complete picture of uncertainty is the next step.

Improving the pattern extraction technique

For the EOF-NMD approach, there are two directions to continue the study. First, the current framework is numerically stable but is not optimized for extracting the patterns. Therefore, I will continue to seek how to optimally delineate the patterns from the original data. Second, the Amplitude-Modulated signals are yet to be fully explained. While preliminary results show some level of correlation with existing ocean state indices. Further insight is needed into this matter and it could possibly reveal the full picture of how precipitation is affected by the large scale interactions among land, ocean and atmosphere.

REFERENCES

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin & E. Nelkin (2003) The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, 4**,** 1147-1167.

AghaKouchak, A., A. Bardossy & E. Habib (2010a) Copula-based uncertainty modelling: application to multisensor precipitation estimates. *Hydrological Processes*, 24**,** 2111-2124.

AghaKouchak, A., A. Bárdossy & E. Habib (2010b) Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. *Advances in Water Resources*, 33**,** 624-634.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu & E. Amitai (2011) Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, 116**,** D02115.

AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert & S. Sorooshian. 2012a. *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*. Springer Netherlands.

AghaKouchak, A., A. Mehran, H. Norouzi & A. Behrangi (2012b) Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, 39**,** L09406.

AghaKouchak, A., N. Nasrollahi & E. Habib (2009) Accounting for Uncertainties of the TRMM Satellite Estimates. *Remote Sensing,* 1**,** 606-619.

AghaKouchak, A., N. Nasrollahi, J. Li, B. Imam & S. Sorooshian (2010c) Geometrical Characterization of Precipitation Patterns. *Journal of Hydrometeorology,* 12**,** 274-285.

Amitai, E., C. L. Unkrich, D. C. Goodrich, E. Habib & B. Thill (2012) Assessing Satellite-Based Rainfall Estimates in Semiarid Watersheds Using the USDA-ARS Walnut Gulch Gauge Network and TRMM PR. *Journal of Hydrometeorology,* 13**,** 1579-1588.

Arkin, P. A. & P. E. Ardanuy (1989) Estimating Climatic-Scale Precipitation from Space: A Review. *Journal of Climate,* 2**,** 1229-1238.

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

Avanesians, P. & M. Momayez (2015) Wave separation: Application for arrival time detection in ultrasonic signals. *Ultrasonics,* 55**,** 15-25.

Barlow, M., S. Nigam & E. H. Berbery (2001) ENSO, Pacific Decadal Variability, and U.S. Summertime Precipitation, Drought, and Stream Flow. *Journal of Climate,* 14**,** 2105-2128.

Behrangi, A., B. Khakbaz, T. C. Jaw, A. AghaKouchak, K. L. Hsu & S. Sorooshian (2011) Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology,* 397**,** 225-237.

Boushaki, F. I., K.-L. Hsu, S. Sorooshian, G.-H. Park, S. Mahani & W. Shi (2009) Bias Adjustment of Satellite Precipitation Estimation Using Ground-Based Measurement: A Case Study Evaluation over the Southwestern United States. *Journal of Hydrometeorology,* 10**,** 1231-1242.

Buarque, D. C., R. C. D. de Paiva, R. T. Clarke & C. A. B. Mendes (2011) A comparison of Amazon rainfall characteristics derived from TRMM, CMORPH and the Brazilian national rain gauge network. *Journal of Geophysical Research: Atmospheres,* 116**,** D19105.

Cayan, D. R. (1996) Interannual Climate Variability and Snowpack in the Western United States. *Journal of Climate,* 9**,** 928-948.

Cayan, D. R., M. D. Dettinger, H. F. Diaz & N. E. Graham (1998) Decadal Variability of Precipitation over Western North America. *Journal of Climate,* 11**,** 3148-3166.

Chen, M., P. Xie, J. E. Janowiak & P. A. Arkin (2002) Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations. *Journal of Hydrometeorology,* 3**,** 249-266.

Chen, S., Y. Hong, Q. Cao, P.-E. Kirstetter, J. J. Gourley, Y. Qi, J. Zhang, K. Howard, J. Hu & J. Wang (2013a) Performance evaluation of radar and satellite rainfalls for

Typhoon Morakot over Taiwan: Are remote-sensing products ready for gauge denial scenario of extreme events? *Journal of Hydrology,* 506**,** 4-13.

Chen, S., Y. Hong, J. J. Gourley, G. J. Huffman, Y. Tian, Q. Cao, B. Yong, P.-E. Kirstetter, J. Hu, J. Hardy, Z. Li, S. I. Khan & X. Xue (2013b) Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States. *Water Resources Research,* 49**,** 8174-8186.

Chen, Y.-C., M.-Y. Cheng & H.-T. Wu (2014) Non-parametric and adaptive modelling of dynamic periodicity and trend with heteroscedastic and dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 76**,** 651-682.

Ciach, G. J., W. F. Krajewski & G. Villarini (2007) Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *Journal of Hydrometeorology,* 8**,** 1325-1347.

Curtis, S. (2008) The Atlantic multidecadal oscillation and extreme daily precipitation over the US and Mexico during the hurricane season. *Climate Dynamics,* 30**,** 343-351.

Daubechies, I., J. Lu & H.-T. Wu (2011) Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis,* 30**,** 243-261.

Dettinger, M. D., D. R. Cayan, H. F. Diaz & D. M. Meko (1998) North–South Precipitation Patterns in Western North America on Interannual-to-Decadal Timescales. *Journal of Climate,* 11**,** 3095-3111.

Duan, Q., S. Sorooshian & V. Gupta (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research,* 28**,** 1015-1031.

Duan, Q. Y., V. K. Gupta & S. Sorooshian (1993) Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications,* 76**,** 501-521.

Duan, Z. & W. G. M. Bastiaanssen (2013) First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure. *Remote Sensing of Environment,* 131**,** 1-13.

Ebert, E. E., J. E. Janowiak & C. Kidd (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society,* 88**,** 47-+.

Elsanabary, M. H. & T. Y. Gan (2013) Wavelet Analysis of Seasonal Rainfall Variability of the Upper Blue Nile Basin, Its Teleconnection to Global Sea Surface Temperature, and Its Forecasting by an Artificial Neural Network. *Monthly Weather Review,* 142**,** 1771-1791.

Foufoula-Georgiou, E., A. M. Ebtehaj, S. Q. Zhang & A. Y. Hou (2014) Downscaling Satellite Precipitation with Emphasis on Extremes: A Variational ℓ1-Norm Regularization in the Derivative Domain. *Surveys in Geophysics,* 35**,** 765-783.

Fox, D. G. (1981) Judging Air Quality Model Performance. *Bulletin of the American Meteorological Society,* 62**,** 599-609.

Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller & T. O'Bannon (1998) The WSR-88D Rainfall Algorithm. *Weather and Forecasting,* 13**,** 377-395.

Gao, Y. C. & M. F. Liu (2013) Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrol. Earth Syst. Sci.,* 17**,** 837-849.

Gebregiorgis, A. S. & F. Hossain (2014) Estimation of Satellite Rainfall Error Variance Using Readily Available Geophysical Features. *Geoscience and Remote Sensing, IEEE Transactions on,* 52**,** 288-304.

Gebremichael, M. & W. Krajewski (2007) Application of Copulas to Modeling Temporal Sampling Errors in Satellite-Derived Rainfall Estimates. *Journal of Hydrologic Engineering,* 12**,** 404-408.

Gebremichael, M. & W. F. Krajewski (2005) Modeling distribution of temporal sampling errors in area-time-averaged rainfall estimates. *Atmospheric Research,* 73**,** 243-259.

Gebremichael, M., W. F. Krajewski, M. L. Morrissey, G. J. Huffman & R. F. Adler (2005) A detailed evaluation of GPCP 1 degrees daily rainfall estimates over the Mississippi river basin. *Journal of Applied Meteorology,* 44**,** 665-681.

Gebremichael, M., G.-Y. Liao & J. Yan (2011) Nonparametric error model for a high resolution satellite rainfall product. *Water Resources Research,* 47**,** W07504.

Gebremichael, M., T. M. Over & W. F. Krajewski (2006) Comparison of the Scaling Characteristics of Rainfall Derived from Space-Based and Ground-Based Radar Observations. *Journal of Hydrometeorology,* 7**,** 1277-1294.

Genest, C., B. Rémillard & D. Beaudoin (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics,* 44**,** 199-213.

Ghosh, S. (2010) Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula. *Hydrological Processes,* 24**,** 3558-3567.

Giannini, A., R. Saravanan & P. Chang (2003) Oceanic Forcing of Sahel Rainfall on Interannual to Interdecadal Time Scales. *Science,* 302**,** 1027-1030.

Gosset, M., J. Viarre, G. Quantin & M. Alcoba (2013) Evaluation of several rainfall products used for hydrological applications over West Africa using two high-resolution gauge networks. *Quarterly Journal of the Royal Meteorological Society,* 139**,** 923-940.

Gourley, J. J., Y. Hong, Z. L. Flamig, L. Li & J. Wang (2009) Intercomparison of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations for a Season of Record Rainfall. *Journal of Applied Meteorology and Climatology,* 49**,** 437-452.

Habib, E., A. Henschke & R. F. Adler (2009) Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmospheric Research,* 94**,** 373-388.

He, X., H. Guan, X. Zhang & C. T. Simmons (2014) A wavelet-based multiple linear regression model for forecasting monthly rainfall. *International Journal of Climatology,* 34**,** 1898-1912.

Herr, H. D. & R. Krzysztofowicz (2005) Generic probability distribution of rainfall in space: the bivariate model. *Journal of Hydrology,* 306**,** 234-263.

Hirpa, F. A., M. Gebremichael & T. Hopson (2009) Evaluation of High-Resolution Satellite Precipitation Products over Very Complex Terrain in Ethiopia. *Journal of Applied Meteorology and Climatology,* 49**,** 1044-1051.

Hong, Y., R. F. Adler, F. Hossain, S. Curtis & G. J. Huffman (2007a) A first approach to global runoff simulation using satellite rainfall estimation. *Water Resources Research,* 43**,** W08502.

Hong, Y., D. Gochis, J. T. Cheng, K. L. Hsu & S. Sorooshian (2007b) Evaluation of PERSIANN-CCS rainfall measurement using the NAME Event Rain Gauge Network. *Journal of Hydrometeorology,* 8**,** 469-482.

Hong, Y., K.-l. Hsu, H. Moradkhani & S. Sorooshian (2006) Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resources Research,* 42**,** W08421.

Hossain, F. & E. N. Anagnostou (2006) A two-dimensional satellite rainfall error model. *Geoscience and Remote Sensing, IEEE Transactions on,* 44**,** 1511-1522.

Hsu, K.-L., A. Behrangi, B. Imam & S. Sorooshian. 2010. Extreme Precipitation Estimation Using Satellite-Based PERSIANN-CCS Algorithm. In *Satellite Rainfall Applications for Surface Hydrology,* eds. M. Gebremichael & F. Hossain, 49-67. Springer Netherlands.

Hsu, K.-l., X. Gao, S. Sorooshian & H. V. Gupta (1997) Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *Journal of Applied Meteorology,* 36**,** 1176-1190.

Huang, N. E. & Z. Wu (2008) A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics,* 46**,** RG2006.

Huffman, G. J. (1997) Estimates of Root-Mean-Square Random Error for Finite Samples of Estimated Precipitation. *Journal of Applied Meteorology,* 36**,** 1191-1201.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman & E. F. Stocker (2007) The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *Journal of Hydrometeorology,* 8**,** 38-55.

Iatsenko, D., P. V. E. McClintock & A. Stefanovska (2012) Nonlinear Mode Decomposition: a new noise-robust, adaptive decomposition method. *ArXiv e-prints*.

Jiang, R., T. Y. Gan, J. Xie & N. Wang (2014) Spatiotemporal variability of Alberta's seasonal precipitation, their teleconnection with large-scale climate anomalies and sea surface temperature. *International Journal of Climatology,* 34**,** 2899-2917.

Joyce, R. J., J. E. Janowiak, P. A. Arkin & P. Xie (2004) CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology,* 5**,** 487-503.

Khan, S. I., Y. Hong, J. J. Gourley, M. U. K. Khattak, B. Yong & H. J. Vergara (2014) Evaluation of three high-resolution satellite precipitation estimates: Potential for monsoon monitoring over Pakistan. *Advances in Space Research,* 54**,** 670-684.

Kirstetter, P.-E., Y. Hong, J. J. Gourley, S. Chen, Z. Flamig, J. Zhang, M. Schwaller, W. Petersen & E. Amitai (2012) Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–Based National Mosaic QPE. *Journal of Hydrometeorology,* 13**,** 1285-1300.

Kisi, O. (2010) Wavelet regression model for short-term streamflow forecasting. *Journal of Hydrology,* 389**,** 344-353.

Knowles, N., M. D. Dettinger & D. R. Cayan (2006) Trends in Snowfall versus Rainfall in the Western United States. *Journal of Climate,* 19**,** 4545-4559.

Kumar, P. & E. Foufoula-Georgiou (1997) Wavelet analysis for geophysical applications. *Reviews of Geophysics,* 35**,** 385-412.

Kunkel, K. E. & J. R. Angel (1999) Relationship of ENSO to snowfall and related cyclone activity in the contiguous United States. *Journal of Geophysical Research: Atmospheres,* 104**,** 19425-19434.

Kuo, C.-C., T. Y. Gan & P.-S. Yu (2010) Wavelet Analysis on the Variability, Teleconnectivity, and Predictability of the Seasonal Rainfall of Taiwan. *Monthly Weather Review,* 138**,** 162-175.

Kwon, E. H., B. J. Sohn, D. E. Chang, M. H. Ahn & S. Yang (2008) Use of numerical forecasts for improving TMI rain retrievals over the mountainous area in Korea. *Journal of Applied Meteorology and Climatology,* 47**,** 1995-2007.

LaMarche, V. C. & H. C. Fritts (1971) ANOMALY PATTERNS OF CLIMATE OVER THE WESTERN UNITED STATES, 1700–1930, DERIVED FROM PRINCIPAL COMPONENT ANALYSIS OF TREE-RING DATA. *Monthly Weather Review,* 99**,** 138-142.

Li, C., V. P. Singh & A. K. Mishra (2012) Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resources Research,* 48**,** W03521.

--- (2013) A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. *Water Resources Research,* 49**,** 767-789.

Li, J., K. Hsu, A. AghaKouchak & S. Sorooshian (2015) An object-based approach for verification of precipitation estimation. *International Journal of Remote Sensing,* 36**,** 513-529.

Li, M. & Q. X. Shao (2010) An improved statistical approach to merge satellite rainfall estimates and raingauge data. *Journal of Hydrology,* 385**,** 51-64.

Li, X., Q. Zhang & C.-Y. Xu (2014) Assessing the performance of satellite-based precipitation products and its dependence on topography over Poyang Lake basin. *Theoretical and Applied Climatology,* 115**,** 713-729.

Li, Z., D. Yang & Y. Hong (2013) Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the Yangtze River. *Journal of Hydrology,* 500**,** 157-169.

Lorenz, E. N. (1956) Empirical orthogonal functions and statistical weather prediction.

Lorenz, E. N. (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus,* 17**,** 321-333.

McCollum, J. R., W. F. Krajewski, R. R. Ferraro & M. B. Ba (2002) Evaluation of biases of satellite rainfall estimation algorithms over the continental United States. *Journal of Applied Meteorology,* 41**,** 1065-1080.

Mehran, A. & A. AghaKouchak (2014) Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations. *Hydrological Processes,* 28**,** 2262-2270.

Moazami, S., S. Golian, M. R. Kavianpour & Y. Hong (2014) Uncertainty analysis of bias from satellite rainfall estimates using copula method. *Atmospheric Research,* 137**,** 145-166.

Moron, V., R. Vautard & M. Ghil (1998) Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics,* 14**,** 545-569.

Müller, M. F. & S. E. Thompson (2013) Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal. *Advances in Water Resources,* 60**,** 121-134.

Mwale, D. & T. Gan (2010) Integrating Wavelet Empirical Orthogonal Functions and Statistical Disaggregation for Predicting Weekly Runoff for the Upper Kafue Basin in Zambia, Africa. *Journal of Hydrologic Engineering,* 15**,** 822-833.

Mwale, D. & T. Y. Gan (2005) Wavelet Analysis of Variability, Teleconnectivity, and Predictability of the September–November East African Rainfall. *Journal of Applied Meteorology,* 44**,** 256-269.

Mwale, D., T. Yew Gan & S. S. P. Shen (2004) A new analysis of variability and predictability of seasonal rainfall of central southern Africa for 1950–94. *International Journal of Climatology,* 24**,** 1509-1530.

Nazemi, A. & A. Elshorbagy (2012) Application of copula modelling to the performance assessment of reconstructed watersheds. *Stochastic Environmental Research and Risk Assessment,* 26**,** 189-205.

Ning, L. & R. S. Bradley (2014) Winter precipitation variability and corresponding teleconnections over the northeastern United States. *Journal of Geophysical Research: Atmospheres,* 119**,** 2014JD021591.

North, G. R., T. L. Bell, R. F. Cahalan & F. J. Moeng (1982) Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Monthly Weather Review,* 110**,** 699-706.

North, G. R., J. B. Valdés, E. Ha & S. S. P. Shen (1994) The Ground-Truth Problem for Satellite Estimates of Rain Rate. *Journal of Atmospheric and Oceanic Technology,* 11**,** 1035-1041.

Peel, M. C., B. L. Finlayson & T. A. McMahon (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.,* 11**,** 1633-1644.

Reason, C. J. C. & M. Rouault (2002) ENSO-like decadal variability and South African rainfall. *Geophysical Research Letters,* 29**,** 16-1-16-4.

Redmond, K. T. & R. W. Koch (1991) Surface Climate and Streamflow Variability in the Western United States and Their Relationship to Large-Scale Circulation Indices. *Water Resources Research,* 27**,** 2381-2399.

Ross, T. & N. Lott. 2003. *A climatology of 1980-2003 extreme weather and climate events*.

Sapiano, M. R. P. & P. A. Arkin (2009) An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data. *Journal of Hydrometeorology,* 10**,** 149-166.

Scofield, R. A. & R. J. Kuligowski (2003) Status and Outlook of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events. *Weather and Forecasting,* 18**,** 1037-1051.

Sellers, W. D. (1968) CLIMATOLOGY OF MONTHLY PRECIPITATION PATTERNS IN THE WESTERN UNITED STATES, 1931–1966. *Monthly Weather Review,* 96**,** 585-595.

Seo, B. C. & W. F. Krajewski (2010) Scale Dependence of Radar Rainfall Uncertainty: Initial Evaluation of NEXRAD's New Super-Resolution Data for Hydrologic Applications. *Journal of Hydrometeorology,* 11**,** 1191-1198.

Serinaldi, F. (2009a) Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. *Stochastic Environmental Research and Risk Assessment,* 23**,** 677-693.

--- (2009b) A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. *Journal of Geophysical Research-Atmospheres,* 114.

Sheaffer, J. D. & E. R. Reiter (1985) Influence of Pacific Sea surface temperatures on seasonal precipitation over the western plateau of the United States. *Archives for meteorology, geophysics, and bioclimatology, Series A,* 34**,** 111-130.

Sheffield, J., C. R. Ferguson, T. J. Troy, E. F. Wood & M. F. McCabe (2009) Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters,* 36.

Smith, A. B. & R. W. Katz (2013) US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural hazards,* 67**,** 387-410.

Sorooshian, S., A. AghaKouchak, P. Arkin, J. Eylander, E. Foufoula-Georgiou, R. Harmon, J. M. H. Hendrickx, B. Imam, R. Kuligowski, B. Skahill & G. Skofronick-Jackson (2011) Advanced Concepts on Remote Sensing of Precipitation at Multiple Scales. *Bulletin of the American Meteorological Society,* 92**,** 1353-1357.

Sorooshian, S., X. Gao, K. Hsu, R. A. Maddox, Y. Hong, H. V. Gupta & B. Imam (2002) Diurnal variability of tropical rainfall retrieved from combined GOES and TRMM satellite information. *Journal of Climate,* 15**,** 983-1001.

Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam & D. Braithwaite (2000a) Evaluation of PERSIANN System Satellite–Based Estimates of Tropical Rainfall. *Bulletin of the American Meteorological Society,* 81**,** 2035-2046.

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

Stampoulis, D. & E. N. Anagnostou (2012) Evaluation of Global Satellite Rainfall Products over Continental Europe. *Journal of Hydrometeorology,* 13**,** 588-603.

Steiner, M. (1996) Uncertainty of Estimates of Monthly Areal Rainfall for Temporally Sparse Remote Observations. *Water Resources Research,* 32**,** 373-388.

Steiner, M., T. L. Bell, Y. Zhang & E. F. Wood (2003) Comparison of Two Methods for Estimating the Sampling-Related Uncertainty of Satellite Rainfall Averages Based on a Large Radar Dataset. *Journal of Climate,* 16**,** 3759-3778.

Stisen, S. & I. Sandholt (2010) Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrological Processes,* 24**,** 879-891.

Sun, X. M. & A. P. Barros (2010) An Evaluation of the Statistics of Rainfall Extremes in Rain Gauge Observations, and Satellite-Based and Reanalysis Products Using Universal Multifractals. *Journal of Hydrometeorology,* 11**,** 388-404.

Tapiador, F. J., F. J. Turk, W. Petersen, A. Y. Hou, E. García-Ortega, L. A. T. Machado, C. F. Angelis, P. Salio, C. Kidd, G. J. Huffman & M. de Castro (2012) Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research,* 104–105**,** 70-97.

Taylor, J. R. 1997. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books.

Thakur, G., E. Brevdo, N. S. Fučkar & H.-T. Wu (2013) The Synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications. *Signal Processing,* 93**,** 1079-1094.

Tian, Y., G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni & H. Wu (2013) Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophysical Research Letters,* 40**,** 2060-2065.

Tian, Y., C. D. Peters-Lidard & J. B. Eylander (2010) Real-Time Bias Reduction for Satellite-Based Precipitation Estimates. *Journal of Hydrometeorology,* 11**,** 1275-1285.

Turk, F. J., P. Arkin, M. R. P. Sapiano & E. E. Ebert (2008) Evaluating High-Resolution Precipitation Products. *Bulletin of the American Meteorological Society,* 89**,** 1911-1916.

Turlapaty, A. C., V. G. Anantharaj, N. H. Younan & F. Joseph Turk (2010) Precipitation data fusion using vector space transformation and artificial neural networks. *Pattern Recognition Letters,* 31**,** 1184-1200.

van Ogtrop, F., M. Ahmad & C. Moeller (2014) Principal components of sea surface temperatures as predictors of seasonal rainfall in rainfed wheat growing areas of Pakistan. *Meteorological Applications,* 21**,** 431-443.

Vernimmen, R. R. E., A. Hooijer, Mamenun, E. Aldrian & A. I. J. M. van Dijk (2012) Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. *Hydrol. Earth Syst. Sci.,* 16**,** 133-146.

Villarini, G. & W. Krajewski (2010) Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics,* 31**,** 107-129.

Villarini, G., B.-C. Seo, F. Serinaldi & W. F. Krajewski (2014) Spatial and temporal modeling of radar rainfall uncertainties. *Atmospheric Research,* 135–136**,** 91-101.

Villarini, G., F. Serinaldi & W. F. Krajewski (2008) Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Advances in Water Resources,* 31**,** 1674-1686.

Watterson, I. G. (2009) Components of precipitation and temperature anomalies and change associated with modes of the Southern Hemisphere. *International Journal of Climatology,* 29**,** 809-826.

Weare, B. C., A. R. Navato & R. E. Newell (1976) Empirical Orthogonal Analysis of Pacific Sea Surface Temperatures. *Journal of Physical Oceanography,* 6**,** 671-678.

Wilks, D. S. 2011. *Statistical methods in the atmospheric sciences*. Academic press.

Willmott, C. J. (1982) Some Comments on the Evaluation of Model Performance. *Bulletin of the American Meteorological Society,* 63**,** 1309-1313.

Xie, H., X. Zhou, J. M. H. Hendrickx, E. R. Vivoni, H. Guan, Y. Q. Tian & E. E. Small (2006) EVALUATION OF NEXRAD STAGE III PRECIPITATION DATA OVER A SEMIARID REGION1. *JAWRA Journal of the American Water Resources Association,* 42**,** 237-256.

Yan, J. & M. Gebremichael (2009) Estimating actual rainfall from satellite rainfall products. *Atmospheric Research,* 92**,** 481-488.

Yilmaz, K. K., T. S. Hogue, K.-l. Hsu, S. Sorooshian, H. V. Gupta & T. Wagener (2005) Intercomparison of Rain Gauge, Radar, and Satellite-Based Precipitation Estimates with Emphasis on Hydrologic Forecasting. *Journal of Hydrometeorology,* 6**,** 497-517.

Yuan, H., S. L. Mullen, X. Gao, S. Sorooshian, J. Du & H. M. H. Juang (2005) Verification of probabilistic quantitative precipitation forecasts over the southwest United States during winter 2002/03 by the RSM ensemble system. *Monthly Weather Review*, 133**,** 279-294.

Zeweldi, D. A. & M. Gebremichael (2009) Evaluation of CMORPH Precipitation Products at Fine Space–Time Scales. *Journal of Hydrometeorology*, 10**,** 300-307.

Zhang, Y., E. Habib, R. J. Kuligowski & D. Kim (2013) Joint distribution of multiplicative errors in radar and satellite QPEs and its use in estimating the conditional exceedance probability. *Advances in Water Resources*, 59**,** 133-145.