Title

Causal Inference, Nonlinear Dynamics, and Information Theory Applications in Hydrometeorological Systems

Author

Ombadi, Mohammed Omer Salah

Publication Date

2021

UNIVERSITY OF CALIFORNIA,
IRVINE

Causal Inference, Nonlinear Dynamics, and Information Theory Applications in Hydrometeorological Systems

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in Civil and Environmental Engineering

by
Mohammed Ombadi

Dissertation Committee:
Distinguished Professor Soroosh Sorooshian, Co-Chair
Assistant Adjunct Professor Phu Nguyen, Co-Chair
Professor Kuo-lin Hsu
Distinguished Professor Efi Foufoula-Georgiou

2021

DEDICATION

To my parents who have been my first teachers in life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Soroosh Sorooshian for what has been an incredible four years and a half journey. Soroosh has been supportive in every way possible from the first day I met him and even before I've arrived at the United States. This dissertation would have not been possible without the encouragement and freedom he graciously provided me to explore exciting research topics. I would also like to thank my co-advisor and mentor Phu Nguyen for his friendship, support and endless advice on issues related to both research and career goals. I enjoyed the daily, long, several-hours chats that we used to have during my first year of grad school. I am also very grateful to Prof Abdin Salih who guided me through my first steps in research and academia after I graduated from college, and from whom I learned dedication and commitment toward research. Thank you!

I would also like to thank my PhD defense committee members Kuo-lin Hsu and Efi Foufoula-Georgiou for their feedback and comments on my research work. Thanks goes also to Prof. Rina Dechter who has been a committee member of my Advancement to Candidacy exam. It would be remiss of me if I don't thank the amazing professors and teaching staff at UC Irvine from whom I learned a lot about hydrology and water sciences. Thanks goes also to Prof Gamal Abdo at the Water Research Center, University of Khartoum for his continuous support.

Of course, life would have not been so wonderful at UC Irvine without the company of my lab mates at CHRS, classmates, housemates, and many friends. Thanks goes also to Diane Hohnbaum for her incredible support, her patience in submitting countless reference letters for my applications and processing endless travel reimbursements. I would also like to thank April Heath and the rest of the staff at the Department of Civil & Environmental Engineering.

Finally, I thank my parents and my sister Miral for their love, encouragement, advice, and support.

VITA

Education

University of California Irvine, PhD in Civil & Environmental Engineering, Apr 2021
University of California Irvine, MSc in Civil Engineering, Sep 2016 — Dec 2017.
University of Khartoum, BSc in Civil Engineering, Nov 2009 — Nov 2014.

Publications

1. **Ombadi, M.**, Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021), Complexity of hydrologic basins: A chaotic dynamics perspective. ***Journal of Hydrology***, https://doi.org/10.1016/j.jhydrol.2021.126222.

2. **Ombadi, M.**, Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021), How much Information on Precipitation is Contained in Satellite Infrared Imagery?. ***Atmospheric Research***, https://doi.org/10.1016/j.atmosres.2021.105578.

3. **Ombadi, M.**, Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021), Retrospective Analysis and Bayesian Model Averaging of CMIP6 Precipitation in the Nile River Basin**.** ***Journal of Hydrometeorology***, 22(1), 217-229, https://doi.org/10.1175/JHM-D-20-0157.1.

4. **Ombadi, M.**, Nguyen, P., Sorooshian, S., & Hsu, K. L. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. ***Water Resources Research***, 56(7), e2020WR027251, https://doi.org/10.1029/2020WR027251.

5. **Ombadi, M.**, Nguyen, P., Sorooshian, S., & Hsu, K. L. (2018). Developing Intensity-Duration-Frequency (IDF) Curves From Satellite-Based Precipitation: Methodology and Evaluation. ***Water Resources Research***, 54(10), 7752-7766, https://doi.org/10.1029/2018WR022929.

6. Nguyen, P., **Ombadi, M.**, Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., ... & Thorstensen, A. R. (2018). The PERSIANN family of global satellite precipitation data: A review and evaluation of products. ***Hydrology and Earth System Sciences***, 22(11), 5801-5816, https://doi.org/10.5194/hess-22-5801-2018.

7. Nguyen, P., **Ombadi, M.**, Gorooh, V. A., Shearer, E. J., Sorooshian, S., Hsu, K., Bolvin, D., & Ralph, M. F (2020). PERSIANN Dynamic Infrared-Rain Rate (PDIR-Now): A Near-real time, Quasi-Global Satellite Precipitation Dataset. ***Journal of Hydrometeorology***, 21(12), 2893-2906, https://doi.org/10.1175/JHM-D-20-0177.1.

8. Salih, A. and **Ombadi, M.** (2017): Sudan and the Water Sciences Programmes of UNESCO, ***University of Khartoum Engineering Journal***, 6, 20-24.

9. Nguyen, P., Shearer, E. J., **Ombadi, M.**, Gorooh, V. A., Hsu, K., Sorooshian, S., ... & Ralph, M. (2020). PERSIANN Dynamic Infrared–Rain Rate Model (PDIR) for High-Resolution, Real-Time Satellite Precipitation Estimation. ***Bulletin of the American Meteorological Society***, 101(3), E286-E302, https://doi.org/10.1175/BAMS-D-19-0118.1.

10. Tran, H., Nguyen, P., **Ombadi, M.**, Hsu, K. L., Sorooshian, S., & Qing, X. (2019). A cloud-free MODIS snow cover dataset for the contiguous United States from 2000 to 2017. ***Nature Scientific data***, 6, 180300, https://doi.org/10.1038/sdata.2018.300.

11. Tran, H., Nguyen, P., **Ombadi, M.**, Hsu, K., Sorooshian, S., & Andreadis, K. (2019). Improving hydrologic modeling using cloud-free MODIS flood maps. ***Journal of Hydrometeorology***, 20(11), 2203-2214, https://doi.org/10.1175/JHM-D-19-0021.1.

12. Nguyen, P., Ashouri, H., **Ombadi, M.**, Hayatbini, N., Hsu, K. L., & Sorooshian, S. (2020). PERSIANN-CDR for Hydrology and Hydro-climatic Applications. In Satellite Precipitation Measurement (pp. 993-1012). **Springer**, Cham.

13. Nguyen, P., Shearer, E. J., Tran, H., **Ombadi, M.**, Hayatbini, N., Palacios, T., ... & Kuligowski, B. (2019). The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. ***Nature Scientific data***, 6(1), 1-10, https://doi.org/10.1038/sdata.2018.296.

14. Sorooshian, S., Gorooh, V. A., Hayatbini, N., **Ombadi, M.**, Sadeghi, M., Nguyen, P., & Hsu, K. Predictability of Hydrometeorological Extremes and Climate Impacts on Water Resources in Semiarid Zones: Expectations and Reality. ***The Bridge National Academy of Engineering***, 50(1), 33-42.

Under-review

15. **Ombadi** et al., Multivariate Empirical Dynamical Modelling for Streamflow Forecasting and Its Comparison to Long short-term memory (LSTM) Networks.

Career

University of California Irvine, Teaching Assistant (Hydrology CEE 176/276), Fall 2020
University of California Irvine, Teaching Assistant (Hydrology CEE 176/276), Fall 2019
University of California Irvine, Graduate Student Researcher, Fall 2016 - Present
University of Khartoum, Research Assistant, 2015 – 16
University of Khartoum, Teaching Assistant, 2014 – 15

Fellowships & Awards

2020-21 Henry Samueli Endowed Fellowship.
2020 AGU Outstanding Student Presentation Award (OSPA).
2020 Outstanding Young Engineer Award, Orange County Engineering Council (OCEC).
2016 Swedish Institute Study Scholarship (selected).
2018 National Science Foundation (NSF) Award to attend AMS Policy Colloquium.
2019 United Nations Food and Agriculture Organization (FAO) Award to attend Conference.
of Ministry of Water Resources, Sudan.
2020 Associated Graduate Students (AGS) at UCI Travel Grant.

Colloquiums

American Meteorological Society (AMS) Summer Policy Colloquium, 2018
NASA JPL Center for Climate Sciences Summer School, 2017
UNESCO IHP 12th Kovacs Colloquium, 2016

Community

Student Member of American Meteorological Society Commission on the Weather, Water and Climate Enterprise (CWWCE) Steering Committee.
Peer review Geophysical Research Letters, Remote Sensing, Water Resources Research
Graduate Peer Mentor. Mentor for incoming international graduate students at University of California Irvine 2020-21.
Young Scientist Member. Sudanese National Academy of Sciences Standing Committee on Water.

Talks & Poster Presentations

2021 American Meteorological Society Annual Meeting: “Using Causal Inference for Skillful Nonlinear State-space Forecasting of Streamflow**”**, Poster Presentation.
2020 American Geophysical Union Fall Meeting: “Toward an Improved Understanding of Hydrologic Complexity”, Oral Presentation.
2020 American Meteorological Society Annual Meeting: “Causal Inference: A pathway for System Identification using Observational Datasets”, Oral Presentation.
2019 American Geophysical Union Fall Meeting: “A Nonlinear Dynamics Approach to the Analysis of Soil Moisture-Rainfall Feedback”, Poster Presentation.
2018 American Geophysical Union Fall Meeting: “Application of Graphical Models for Causal Detection in Hydrometeorological Systems”, Poster Presentation.
2017 American Geophysical Union Fall Meeting: “Assessment of Developing Intensity Duration Frequency Curves using Satellite Observations”, Poster Presentation.

ABSTRACT

Causal Inference, Nonlinear Dynamics, and Information Theory Applications in Hydrometeorological Systems

by

Mohammed Ombadi

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2021

Distinguished Professor Soroosh Sorooshian, Co-Chair

Assistant Adjunct Professor Phu Nguyen, Co-Chair

This dissertation consists of six distinct research studies that are broadly classified into two parts. The first part is concerned with the application of emerging data analysis tools rooted in causal inference, nonlinear chaotic dynamical systems, and information theory to detect associations and characterize patterns of interaction in complex hydrometeorological systems. This part is motivated by the rapid accumulation of hydrometeorological data records in the form of in-situ, remotely sensed observations and climatological reconstructions in addition to the significant advancements in data mining tools that facilitate discovery of interaction patterns solely from observational datasets. More specifically, I present four studies that utilize observational datasets to elucidate patterns of interaction and subsequently improve predictive understanding of the underlying processes. First, I evaluate the performance of four causal inference methods in recovering the causal structure underlying a hydrologic conceptual model and utilize causal

analysis to formulate hypothesis on the differential impact of environmental variables in regulating evapotranspiration. Second, methods rooted in the theory of chaotic dynamical systems are used to examine the dynamical properties of 400 hydrologic basins across the contiguous United States (CONUS) with the aim of developing a catchment classification framework. Third, I propose an algorithm that utilizes causal inference to extend methods of univariate state space forecasting to account for multivariate predictors. The algorithm is applied for daily streamflow forecasting in nine hydrologic basins across CONUS, and the results are compared to that of deep learning models. Finally, concepts of information theory are used to diagnose the complex, nonlinear, space-time varying relationship between infrared brightness temperature and precipitation across different seasons and spatiotemporal scales.

The second part of the dissertation focuses on the use of long historical records of satellite-based precipitation datasets in hydroclimatic research, and it consists of two studies that utilize Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR) dataset. The first study proposes a framework for developing Intensity Duration Frequency (IDF) curves from satellite-based precipitation datasets. The framework accounts for the inherent biases in the estimates of PERSIANN-CDR, and it is used to develop IDF curves over CONUS with evaluation based on in-situ estimates of NOAA Atlas 14. The second study utilizes PERSIANN-CDR dataset over the Nile river basin to constrain future projections of precipitation obtained from climate models. More specifically, a Bayesian Model Averaging (BMA) approach is adopted to constrain future projections of 20 Global Climate Models (GCMs) from phase six of the Coupled Model Intercomparison Project (CMIP6). The results show that annual precipitation is projected to decrease in the upper White Nile basin,

whereas projected change in the Blue Nile basin is highly uncertain both in magnitude and sign of change.

Keywords: Causal Inference, Chaos, Dynamical Systems, Information Theory, Entropy, Graph Theory, Hydrologic Complexity, Evapotranspiration, Streamflow Forecasting, Satellite-based Precipitation, Intensity-Duration-Frequency (IDF) curves, Infrared Brightness Temperature, Future Projections of Precipitation, Climate Change.

Chapter 1

1. Introduction

1.1 Thesis Overview

This dissertation is comprised of six distinct studies that broadly fall within two general research themes. The first and primary theme explores the application of emerging methods of causal inference, nonlinear dynamics and information theory in addressing a variety of research questions in hydrometeorology, whereas the second theme investigates the utility of integrating satellite-based precipitation datasets in decision-making, planning and design of infrastructure in a changing climate. The dissertation is, therefore, structured in alignment with this classification, and it consists of two main parts. Part I includes four chapters (chapters 2 to 5) each of which presents an application of a data-driven method that is either rooted in the theory of causal inference, chaotic dynamics, or information theory. On the other hand, Part II is comprised of two studies (chapters 6 and 7) that illustrate the utility of satellite-based precipitation records in deriving design rainfalls and constraining future projections of precipitation.

Chapter 2 of this dissertation presents a review of four general methods of causal inference, evaluates their performance in recovering the causal structure of hydrometeorological systems, and investigates the impact of sample length, process noise, and observational noise in the performance of these methods. Subsequently, a causal inference method is then used to develop a hypothesis on the differential impact of environmental variables (e.g. net radiation, vapor pressure deficit, soil moisture …etc.) in regulating evapotranspiration using observational timeseries from a FLUXNET tower site in Arizona. Finally, the results are compared to what would have been obtained using classical analysis of correlation, and the hypothesis is interpreted based on our understanding of canopy seasonal dynamics and evapotranspiration processes.

The next two chapters of the dissertation (chapters 3 and 4) are closely related as they introduce the application of methods rooted in the theory of chaotic dynamical systems in analysis of hydrologic timeseries. More specifically, chapter 3 utilizes the theory of phase space reconstruction based on time-delayed univariate time series along with daily streamflow observations to examine the properties of dynamical behavior in hydrologic basins. The analysis is carried out over approximately 400 hydrologic basins across the contiguous United States (CONUS). Next, an attempt is made to relate the properties of dynamical behavior with the physical properties of basins (e.g. Area, Slope, Elevation, Vegetation Cover …etc.) where we found coherent relationships that link basin dynamical behavior to its physical properties. Overall, the results form the basis of a catchment classification framework, and they bear significance for improved streamflow forecasting and rainfall-runoff modelling in ungauged hydrologic basins. Chapter 4 follows along the same lines presented in chapter 3 and utilizes the theory of phase space reconstruction for streamflow forecasting. More specifically, we present an extension of state-space forecasting based on univariate timeseries to the case of multivariate predictors. We propose a search algorithm that uses causal inference to navigate the search space of possible embedding coordinates and converge to the optimum set of embedding coordinates which is then used for forecasting. The algorithm is applied in a case study that consists of 9 hydrologic basins across CONUS, and the results are compared with deep learning approaches of Long Short-Term Memory (LSTM) networks.

Chapter 5 brings the first part of the dissertation to a conclusion. In particular, the chapter presents the results of an analysis that utilizes information-theoretic methods in diagnosing the relationship between Infrared (IR) brightness temperature and surface precipitation rate. This study is motivated by the need to understand the spatial, temporal, and seasonal variability in the complex, indirect relationship between IR and precipitation. IR is generally a good proxy of the properties of cloud tops, and it is often used in operational algorithms of satellite-based precipitation estimation. However, the relationship is complex, nonlinear and exhibits significant variability across time and space scales as well as over different seasons. Therefore, in this study, we used an information-theoretic measure named Maximal Information Coefficient (MIC) to diagnose this complex relationship. We argue that this analysis provides insights that cannot be obtained by evaluation of satellite-based precipitation estimates against in-situ observations because in such a case the errors resulting from the inherent dependence are inextricable from those induced by the assumptions encoded in the estimation algorithms.

Part II of the dissertation consists of two studies that demonstrate the utility of using long, historical records of satellite-based precipitation in derivation of infrastructure design rainfalls and constraining future projections of precipitation. In both studies, the dataset used is Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR). Specifically, chapter 6 presents a methodological framework to derive Intensity-Duration-Frequency (IDF) curves from satellite-based precipitation observations. The framework accounts for the inherent biases in the estimates of extreme precipitation obtained from PERISANN-CDR, namely elevation-dependent underestimation bias and areal to point estimation biases. We then use the framework to develop daily IDF curves over CONUS and evaluate the results using NOAA Atlas 14 of IDF curves estimated from in-situ observations. We also discuss issues related to uncertainty and the potential of extending the application of the framework to other regions of the world—especially developing countries.

In chapter 7, we present the application of Bayesian Model Averaging (BMA) to constrain future projections of precipitation in the Nile River basin. The study is motivated by the current intense debate on the future of water availability in the Nile river amongst the geopolitical tensions resulting from transboundary water management and the mega dams being built in the basin. Specifically, we use the outputs of 20 Global Circulation Models (GCMs) from phase six of the Coupled Model Intercomparison Project (CMIP6) together with a long historical record of precipitation from PERSIANN-CDR (1983 – 2014). The obtained results provide a more informative estimate of future projections than the ensemble mean as well as providing probability distributions of future projections instead of a single value. Our results show that annual precipitation is more likely to decrease in the White Nile river basin, whereas projected change in the Blue Nile is highly uncertain both in magnitude and sign of change.

Finally, Chapter 8 of the dissertation provides concluding remarks and future directions regarding the application of data-driven exploration methods in hydrometeorological systems. More specifically, the chapter summarizes the main findings of the six studies and provides a brief discussion on the possibility of extending the application of the presented methods to address some of the important research questions in hydrologic research.

1.2 Contributions

Five of the six studies that constitute this dissertation have been published in peer-reviewed articles as follows:

- Chapter 2: Ombadi et al., 2020, Evaluation of methods for causal discovery in hydrometeorological systems, ***Water Resources Research***.
- Chapter 3: Ombadi et al., 2021a, Complexity of hydrologic basins: A chaotic dynamics perspective. ***Journal of Hydrology***.
- Chapter 5: Ombadi et al., 2021b, How much Information on Precipitation is Contained in Satellite Infrared Imagery? ***Atmospheric Research***.
- Chapter 6: Ombadi et al., 2018, Developing Intensity-Duration-Frequency (IDF) Curves From Satellite-Based Precipitation: Methodology and Evaluation, ***Water Resources Research***.
- Chapter 7: Ombadi et al., 2021c, Retrospective Analysis and Bayesian Model Averaging of CMIP6 Precipitation in the Nile River Basin, ***Journal of Hydrometeorology***.

The sixth study which represents the fourth chapter of the dissertation is currently under-review. In addition to these peer-reviewed articles, the studies reported in the dissertation have been presented in several conferences including the American Geophysical Union (AGU) Fall

Meeting and the American Meteorological Society (AMS) Annual Meeting. Furthermore, the results reported in chapter 3 of the dissertation have been awarded an AGU Outstanding Student Presentation Award (OSPA) for an oral talk at the 2020 AGU Fall Meeting.

1.3 Motivation

The primary research theme in this dissertation, namely using data-driven methods to understand scale-dependent interactions in hydrometeorological systems is, on one hand, motivated by the unprecedented availability of observational data records in the form of in-situ, remotely sensed observations and climatological reconstructions. On the other hand, it is aimed to address the gap of knowledge in hydrological systems regarding the change in dynamical behavior of hydrological systems across the hierarchy of space and time scales, and to satisfy the need for establishing scale-dependent relationships that take into account the emergent behavior of such systems at the macroscale level. In particular, it has long been argued that significant advances in hydrology are hindered due to the current status of hydrological research which focuses on incorporating heterogeneity in highly calibrated, sophisticated, distributed hydrological models rather than seeking organizing principles and laws that govern the variability in heterogeneity and control the macroscopic behavior of the watersheds (Klemeš, 1983; Dooge, 1986; McDonnell et al., 2007). A recent hydrologic community initiative that involved a large number of researchers in the field of hydrology has identified twenty-three unsolved problems in hydrology, and explicitly acknowledged the importance of scale (time and space) in formulating hydrologic laws and relationships (Blöschl et al., 2019).

In this dissertation, we argue that methods of causal inference rooted in networks, chaos and information theories (as will be discussed in chapter 2) provide a strong basis for a framework to discover patterns and regularities in hydrological systems at the macroscale level. These patterns can subsequently be utilized to improve our predictive understanding of hydrometeorological systems. More specifically, in chapter 2, we show an example of using causal inference methods to understand the relative contributions of environmental variables in regulating evapotranspiration. In chapter 3, we present another example of using observational datasets of streamflow to detect macroscale patterns of dynamical behavior in hydrologic basins across the contiguous United States. Chapter 4 presents a case study of using observational data solely to improve our capacity of streamflow forecasting using data-driven methods. This latter case illustrates how improved understanding of scale-dependent interactions can translate to enhanced capacity of forecasting and prediction of system behavior.

The second theme of the dissertation focuses on the use of long historical records of satellite-based precipitation for planning and design of infrastructure as well as for constraining the projections of future precipitation obtained from climate models. This chapter is motivated by the increasing availability of satellite-based precipitation datasets and the significant advancement in remote sensing of precipitation. While research in remote sensing of precipitation in the recent two decades have mostly focused on improving algorithms for estimation of precipitation rate from infrared brightness temperature and passive microwave observations, it is important to conduct research on the inherent biases of the datasets and the possibility of integrating the long historical records in engineering applications of immense socioeconomic value. This has been the impetus behind the two research studies in part II of the dissertation.

1.4 Table of Datasets

Several datasets have been used in the six studies that constitute this dissertation. Table A.1 in the appendix lists all the datasets used in this dissertation and provides link to repositories for data download. Moreover, Table A.2 in the appendix provides a list of datasets that were generated in some of the studies with links to repositories for data download.

1.5 Table of Symbols

For each chapter of the dissertation (chapters 2 to 7), a table is provided in the appendix with symbols and notations used in the chapter. These tables are numbered in the appendix (Table A.3 to A.8); please refer to these tables for the definition of any given symbol.

Part I: Causal Inference, Nonlinear Dynamics, and Information Theory

Chapter 2

2. Causal Inference in Hydrometeorological Systems

“This chapter is extracted from Ombadi et al. (2020) with few edits incorporated for brevity and clarity”

2.1 Introduction

In hydrometeorology, as in other branches of natural sciences, causal inference plays a central role in the acquisition of objective scientific knowledge. Arguably, most questions encountered in hydrometeorology can be framed in the context of cause and effect. Causal questions in the form of hypothesis formulation and validation regarding the interaction of variables and processes are ubiquitous in the literature of hydrology and climate. Most notable are studies concerned with understanding the impact of climate change on the hydrologic cycle (e.g., Barnett et al., 2005; Held & Soden, 2006; Milliman et al., 2008; Piao et al., 2010), resolving ambiguities in the interactions of the coupled Land-Ocean-Atmosphere system (e.g., Charney, 1975; Eltahir, 1998; Entekhabi et al., 1996; Koster et al., 2004) and understanding the impact of land cover and anthropogenic land-use on atmospheric circulations (e.g., de Noblet-Ducoudré et al., 2012; Findell et al., 2017). The underlying question among all these studies is a causal one, and the aim is to understand how a specific variable will change as a result of an intervention in the system. It is important to distinguish causal interactions from associations; the latter characterize the dependence between variables as in standard statistical analysis (regression and correlation), while the former extends the analysis by identification of variables in dependence relationships as cause and effect. In order to identify causal interactions, all causal inference methods rest on different sets of assumptions that identify invariant relationships in systems under intervention (Pearl, 2009a). Such a distinction between causality and association in hydrological systems has been pointed out by Klemes (1982), while discussing the relationships obtained by empirical analysis “The relationships initially discovered are of necessity simple …. They generally tell us what change in one observed quantity correspond to a change in another … they tell us what happens but do not derive the outcome from the dynamic mechanisms governing the process.”

Although experimental research, manipulation, and controlled testing provide a framework to understand causal processes, using such an approach in hydrometeorology is infeasible (e.g., manipulating global or regional climate), costly (e.g., experimental catchments), or inaccurate (e.g., using numerical models for controlled experiments). With these considerations in mind, causal inference from observations is an alternative avenue. Simply stated, the goal of all causal inference methods is to extract information regarding causal interactions among variables in a given system utilizing time series measurements with prior knowledge incorporated in the selection of variables. The last few decades witnessed significant advancements in theories as well as algorithms needed for causal inference from observational data sets. The earliest significant work in empirical causal inference was proposed by Granger (1969). In a seminal paper, Granger formulated a statistical method for causality which states that variable X has a causal effect on variable Y if variable X provides statistically significant information about future values of variable Y (Granger, 1969). The original framework of Granger causality (GC) has later been extended by introducing concepts of information flow resulting in Transfer Entropy (TE) as a measure of causality that is sensitive to both linear and nonlinear relationships (Schreiber, 2000). In addition to Granger framework, a fundamental work that influenced the field of causality was the introduction of probabilistic Graphical models (e.g., Bayesian Networks) and causal diagrams (Pearl, 1988, 1995, 2009b). Furthermore, almost simultaneously with the advances in causal discovery in the fields of statistics and machine learning, fundamental contributions in detecting causality from time series were conceived in the field of dynamical systems (Deyle & Sugihara, 2011; Sugihara et al., 2012; Sugihara & May, 1990). These contributions were built on the theories of time-delay embedding and reconstruction of attractors from time series (Takens, 1981).

The application of causal inference methods in hydrology and climate research has been gaining attention in recent years. Ruddell and Kumar (2009a, 2009b) adapted TE to characterize process networks of ecohydrological systems from observational data sets. Similarly, Sharma and Mehrotra (2014) developed an information theoretic measure to be used in system identification of natural systems. Recently, TE concepts have been adopted to develop process networks taking into account the partitioning of information into synergistic, unique, and redundant (Goodwell & Kumar, 2017a, 2017b). On the other hand, methods of causal detection based on time-delay embedding have been used to investigate the soil moisture—rainfall feedback (Wang et al., 2018). The outcomes of these studies are encouraging, and they highlight the potential of causal inference in improving identification of hydrometeorological systems from observational data sets. However, there exists a lack of comparative studies. Specifically, studies that compare the performance of different causal methods in the context of hydrometeorological systems; taking into consideration the challenges and practical limitations frequently encountered in such systems such as nonlinearity (e.g., threshold behavior), sample size, process and observational noise, and synchronization due to seasonality.

In view of the above discussion, the aim of the present study is three-fold. First, to present briefly four main causal inference methods: GC, TE, Graph-based casual methods (PC algorithm), and Convergent Cross Mapping (CCM) and discuss their theoretical underpinnings and assumptions. Second, to contrast the performance of these methods using synthetic data generated from a simple hydrological model yet representative of most features common to environmental systems and to investigate the impact of sample size and presence of noise on the performance of each method. Third, to use causal analysis in examining the significant environmental drivers of evapotranspiration in a shrubland region and to identify their relative contributions during summer and winter seasons.

2.2 Methods of Causal Inference

a. Granger Causality

Granger Causality (GC) was developed in the late 60’s by Clive Granger (Granger, 1969), and it is perhaps the first practical method to test for causality. GC is defined in both time and frequency domains, and it relies primarily on a fundamental assumption that the cause precedes the effect in time (i.e. two variables occurring at the same time step can’t be causally related). Although this assumption may appear to be trivial and intuitive, it has significant implications because causality is interpreted based on the time ordering of events; hence, the original method (Granger, 1969) is unable to detect contemporaneous causality. In addition, two secondary assumptions underlie GC. First, the cause provides useful information for predicting the effect at future time steps. Second, any variable in the system can be represented linearly by lagged values of system variables and an error term. That is, the system can be represented by a vector autoregressive (VAR) model. The latter condition implies that the underlying system is linear and stochastic, albeit some nonlinear processes can be modelled as VARs. It should be noted that the definition of causality in GC does not conform with more strict definitions of causality as in Pearl (2009b).

The assumptions underlying GC are encoded in its implementation which is conducted by constructing restricted and unrestricted regression models for each variable. Let's first consider a matrix of stationary time series observations ($\boldsymbol{\mathcal{H}}$) for a given system that consists of three variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ measured at times $t = 1, 2, \ldots, l$. Where l is the length of the time series, then $\boldsymbol{\mathcal{H}}$ is given as below:

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \boldsymbol{X}_3 & \cdots & \boldsymbol{X}_{l-1} & \boldsymbol{X}_l \\ \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \boldsymbol{Y}_3 & \cdots & \boldsymbol{Y}_{l-1} & \boldsymbol{Y}_l \\ \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 & \cdots & \boldsymbol{Z}_{l-1} & \boldsymbol{Z}_l \end{bmatrix} \tag{2.1}$$

To test the hypothesis that $\boldsymbol{Y}$ causes $\boldsymbol{X}$, written $\boldsymbol{Y} \Longrightarrow \boldsymbol{X}$, two p$^{\text{th}}$ order vector autoregressive (VAR) models are constructed as follows:

$$\boldsymbol{X}_t = \sum_{k=1}^{p} c_{xxk} \,.\, \boldsymbol{X}_{t-k} + \sum_{k=1}^{p} c_{xyk} \,.\, \boldsymbol{Y}_{t-k} + \sum_{k=1}^{p} c_{xzk} \,.\, \boldsymbol{Z}_{t-k} + \varepsilon_{xt} \tag{2.2}$$

$$\boldsymbol{X}_t = \sum_{k=1}^{p} \grave{c}_{xxk} \,.\boldsymbol{X}_{t-k} + \sum_{k=1}^{p} \grave{c}_{xzk} \,.\boldsymbol{Z}_{t-k} + \grave{\varepsilon}_{xt} \tag{2.3}$$

Where c_{xxk} and $\grave{c}_{xxk}$ are the regression coefficients of $\boldsymbol{X}_t$ regressed on $\boldsymbol{X}_{t-k}$ in the first and second model respectively. Similarly,c_{xzk} and $\acute{c}_{xzk}$ are the regression coefficients of $\boldsymbol{X}_t$ regressed on $\boldsymbol{Z}_{t-k}$ in the first and second model respectively. c_{xyk} is the regression coefficient of $\boldsymbol{X}_t$ regressed on $\boldsymbol{Y}_{t-k}$, while ε_{xt} and $\grave{\varepsilon}_{xt}$ are the residuals in the two models. The first model (equation 2.2) is unrestricted regression model that includes lags of variable $\boldsymbol{Y}$, while the second model (equation 2.3) is a restricted regression model. The Null hypothesis (H_0) that $\boldsymbol{Y}$ does not cause $\boldsymbol{X}$ is rejected if it is shown that the first model improves estimation compared to the second model. That is, the difference between the residuals of the two models is statistically significant according to an F-test at a given significance level (α).

b. Transfer Entropy

Transfer Entropy (TE) (Schreiber, 2000) can be understood as a nonparametric extension to the GC method; indeed, Barnett et al. (2009) showed that the two methods are equivalent for Gaussian processes. TE is based on conditional mutual information $\mathbb{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$, the amount of shared information between variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditioned on variable $\boldsymbol{Z}$, which is defined as follows:

$$\mathbb{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) = H(\boldsymbol{Y}|\boldsymbol{Z}) - H(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z})$$

$$= \int f_z(\boldsymbol{z}) \iint f_{x,y|z}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z}) \log \frac{f_{x,y|z}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z})}{f_{x|z}(\boldsymbol{x}|\boldsymbol{z}) \,.\, f_{y|z}(\boldsymbol{y}|\boldsymbol{z})} \, dx\, dy\, dz \tag{2.4}$$

Where H is Shannon entropy (Shannon, 1948), f denotes probability density function whereas $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are single realizations of the variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ respectively. $\mathbb{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$ measures the entropy reduction in variable $\boldsymbol{Y}$ when information of variable $\boldsymbol{X}$ is added conditioned on $\boldsymbol{Z}$. In the above formulation, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent conditioned on $\boldsymbol{Z}$, the numerator can be decomposed into $f_{x,y|z}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z}) = f_{x|z}(\boldsymbol{x}|\boldsymbol{z}) \,.\, f_{y|z}(\boldsymbol{y}|\boldsymbol{z})$ leading to $\mathbb{I}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) = 0$. Thus, no information is gained about variable $\boldsymbol{Y}$ if variable $\boldsymbol{X}$ is known.

Let's consider the same notations used earlier and denote by $\overline{\boldsymbol{X}}_{t-1}, \overline{\boldsymbol{Y}}_{t-1}$ and $\overline{\boldsymbol{Z}}_{t-1}$ the time series of variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ respectively as of time $t-1$; for example, $\overline{\boldsymbol{X}}_{t-1} = [\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \boldsymbol{X}_{t-3} \ldots]$. Similarly, $\overline{\boldsymbol{\mathcal{H}}}_{t-1}$ is the observations matrix as of time $t-1$. To evaluate the hypothesis $\boldsymbol{Y} \Longrightarrow \boldsymbol{X}$, TE is defined as follows:

$$TE(\boldsymbol{Y};\boldsymbol{X}) = \mathbb{I}(\boldsymbol{X};\overline{\boldsymbol{Y}}_{t-1}|\, \overline{\boldsymbol{\mathcal{H}}}_{t-1} \setminus \overline{\boldsymbol{Y}}_{t-1}) \tag{2.5}$$

$\overline{\boldsymbol{\mathcal{H}}}_{t-1} \setminus \overline{\boldsymbol{Y}}_{t-1}$ are all the values that belong to $\overline{\boldsymbol{\mathcal{H}}}_{t-1}$ and not to $\overline{\boldsymbol{Y}}_{t-1}$. Put simply, TE measures the shared information between $\boldsymbol{X}$ and the history of $\boldsymbol{Y}$ conditioned on the history of other observed variables. By contrasting the two definitions of GC and TE, one can comprehend the similarity of the two methods. Like GC, TE relies on the two assumptions that the cause precedes the effect, and the cause provides useful information for the prediction of the effect. However, unlike GC, TE makes no assumptions about the underlying structure of the data; therefore, it can detect both linear and nonlinear relationships. TE and other measures based on information theory have been used in several studies in hydrometeorological research (e.g. Ruddell and Kumar, 2009a; Runge et al., 2014; 2015; Sharma; 2000).

c. Graph-based Methods

In Graph theory, a directed graph representation describes the causal relationships among variables in a given system. The representation consists of nodes representing variables, and directed edges between nodes representing direct casual influences (Dechter, 2013). Kinship terminology (parent, child, descendants …etc.) is commonly used to describe relations in a causal graph. If the graph encodes probabilistic information (i.e. each directed edge in the graph is quantified by the conditional probability distribution of the child given its parent), the causal graph is called a Bayesian network (Pearl, 1988; Darwiche, 2009). Graph-based causal algorithms utilize a set of graphical rules that govern the retrieval of system causal graphs from nonexperimental data. These rules include, among others, the “d-separation” criterion (Pearl, 1988) and Causal Markov condition (Pearl & Verma, 1991; Kiiveri et al., 1984). See (Pearl, 1995) for an in-depth discussion of graphical rules.

The PC algorithm (Spirtes & Glymour, 1991) utilizes graphical rules to effectively recover causal relations among variables from observational data. Given time series of three variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$, lagged time series $\overline{\boldsymbol{X}}_{\boldsymbol{t}}, \overline{\boldsymbol{X}}_{\boldsymbol{t-1}}, \overline{\boldsymbol{X}}_{\boldsymbol{t-2}} \ldots \overline{\boldsymbol{X}}_{\boldsymbol{t-p}}$ are first constructed. Similarly, lagged time series of variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are constructed. The algorithm starts by considering a complete (fully connected) undirected graph containing all lagged time series variables as distinct variables. Then, the algorithm iterates over all variables as target variables and uses conditional independencies of increasing order to exclude variables that are independent of the target variable. The result of this process is a graph that is commonly sparser than the original fully connected graph, and then the algorithm utilizes graphical rules to direct the links.

d. Convergent Cross Mapping

Unlike the three methods described in previous subsections, Convergent Cross Mapping (CCM) rests on a different paradigm of causality; namely, the theory of time-delay embedding (Takens, 1981). Based on this theory and under certain conditions, the manifold of a chaotic dynamical system can be reconstructed using time-lagged observations of a single variable (state). Let's consider a chaotic dynamical system $\boldsymbol{\mathcal{M}}$ with the variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$. At any time t, the system is represented in the state space by the 3-dimensional point $\boldsymbol{\mathcal{M}_t} = \{\boldsymbol{X_t}, \boldsymbol{Y_t}, \boldsymbol{Z_t}\}$. For a sequence of observations of length l, the trajectory of $\boldsymbol{\mathcal{M}_t}$ for $t = \{1,2,3 \ldots .\, l\}$ in the state space constructs a manifold of the system $\boldsymbol{\mathcal{M}}$. Time-delay embedding theorem states that one can reconstruct the manifold using lagged time series of a single variable. For example, the trajectory of the point $\boldsymbol{\mathcal{M}_{xt}} = \{\boldsymbol{X_t}, \boldsymbol{X_{t-1}}, \boldsymbol{X_{t-2}}\}$ is called a shadow manifold, and it preserves certain topological properties of the original manifold. Similarly, shadow manifolds can be constructed from time series of variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$.

Based on Takens' theorem, Sugihara et al. (2012) developed CCM as a test for causality. To evaluate the hypothesis $\boldsymbol{Y} \Longrightarrow \boldsymbol{X}$, shadow manifolds from variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are constructed from the trajectory of the points $\boldsymbol{\mathcal{M}_{xt}}$ and $\boldsymbol{\mathcal{M}_{yt}}$ respectively. Where $\boldsymbol{\mathcal{M}_{xt}} = \{\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2} \ldots \boldsymbol{X}_{t-E}\}$ and $\boldsymbol{\mathcal{M}_{yt}} = \{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2} \ldots \boldsymbol{Y}_{t-E}\}$ for $t = \{1, 2 \ldots \grave{l}\,\}$. Here E is the dimension of the manifold, also known as the embedding dimension. $\grave{l}$ is the length of time series used to create the shadow manifolds which is a fraction of the total length of observations l. Next, the shadow manifold of variable $\boldsymbol{X}$ is used to identify the nearest neighbors of the point $\boldsymbol{\mathcal{M}_{xt}}$ for $t = \{\grave{l} + 1, \grave{l} + 2, \dots, l\}$ and their Euclidian distance from the point $\boldsymbol{\mathcal{M}_{xt}}$. Finally, the nearest neighbors and their distances are used to identify contemporaneous points on the shadow manifold of $\boldsymbol{Y}$; consequently, values of $\boldsymbol{Y_t}$ for $t = \{\grave{l} + 1, \grave{l} + 2, \dots, l\}$ are estimated. These estimated values are compared with the observed ones using a metric, often correlation coefficient. If the value of correlation coefficient is significant, the hypothesis $\boldsymbol{Y} \Longrightarrow \boldsymbol{X}$ is accepted.

2.3Examining the Performance of Causal Inference Methods in Recovering the Causal Structure of a Hydrological Bucket Model

a. Model Structure

In order to test the performance of the four causal methods in identification of causal relationships from time series, we resort to the use of a hydrological model. The purpose of using synthetic data generated from the model is that, unlike natural systems, the underlying causal relationships are well known. The model we use here, shown in Figure 2.1a, is a simple hydrological bucket model with four variables: rainfall $\boldsymbol{R}$, soil storage $\boldsymbol{S}$, interflow $\boldsymbol{I}$, and runoff $\boldsymbol{Q}$. This model simulates the process by which rainfall is transformed into runoff in a simplified manner. Essentially, the soil is represented as a bucket and the maximum amount of water it can store is defined by maximum storage (S_{max}). An outgoing flux of water from the bucket represents interflow $\boldsymbol{I}$, the lateral movement of water within the unsaturated zone. Whenever the bucket is full, rainfall will spill from the bucket and transform to runoff $\boldsymbol{Q}$.

In addition to the four variables $\boldsymbol{R}$, $\boldsymbol{S}$, $\boldsymbol{I}$ and $\boldsymbol{Q}$, the model has four parameters: the maximum soil storage S_{max}, and three parameters K_s, δ and ξ identifying the nonlinear storage-discharge relationship. The storage-discharge relationship employed here (equation 2.9) is nonlinear of concave type (i.e. $\boldsymbol{Q}$ as a function of $\boldsymbol{S}$ is a concave power law model). Botter et al. (2009) provides a detailed description of such relationships which has previously been employed in literature (e.g. Brutsaert & Nieber, 1977). In this relationship, δ is the soil moisture amount below which no interflow occurs, ξ is an exponent between 0.5 and 1 (Botter et al., 2009) and K_s characterizes how fast the bucket is depleted of water.

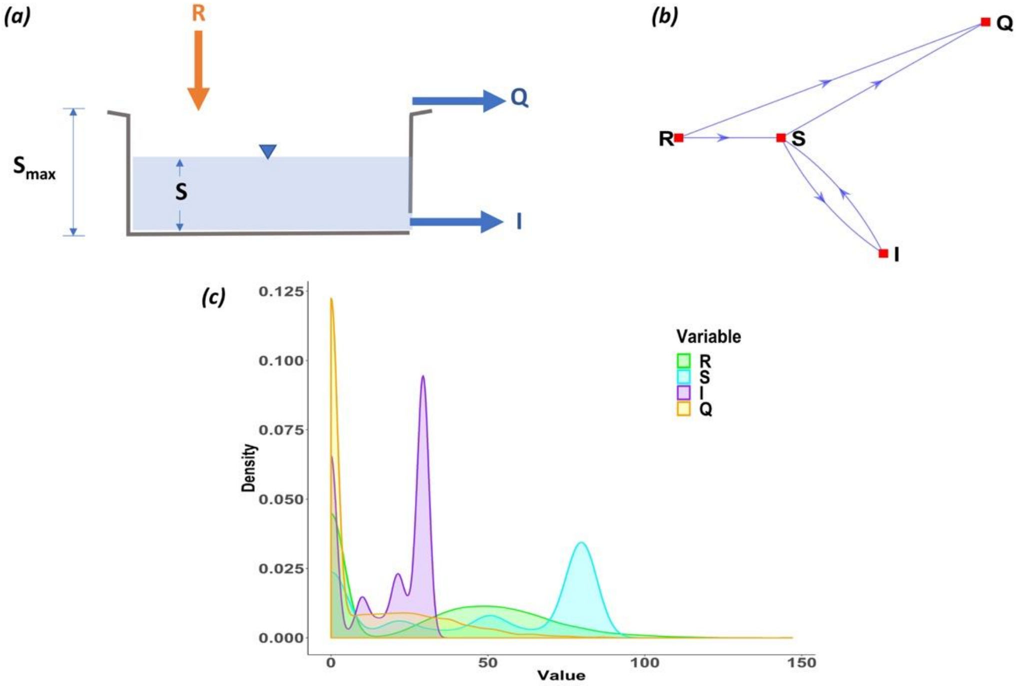


Figure 2.1 (a) *The 4-variables bucket model used in this study to generate synthetic data with the variables rainfall R, soil moisture S, interflow I and runoff Q;* S_{max} *is the maximum soil storage.* ***(b)*** *The causal graph of the same model in (a).* ***(c)*** *Marginal probability distributions of the four variables (R, S, I and Q). The probability distributions are estimated from a simulation of the model with a sample size of 10,000 and signal to noise ratio SNR of* 10^4.

The only forcing variable in the model, rainfall, is simulated as a stochastic process in two steps as described in Woolhiser (2011). First, the occurrence of rainfall is simulated with a binary variable $\grave{\boldsymbol{R}}$ (i.e. rain/no rain). The simulation is performed using a discrete-time first-order Markov chain model:

$$\tilde{p}_{i,j}(t) = Probability\left(\grave{\boldsymbol{R}}_t = j \mid \grave{\boldsymbol{R}}_{t-1} = i\right); \quad i,j = 0,1; \quad t > 1 \tag{2.6}$$

$\tilde{p}_{i,j}$ is the transition probability from the state i to the state j; the state is binary (1 for rain, 0 for no rain). In this study, we used the following transition probabilities for the model:

$$\begin{bmatrix} \tilde{p}_{0,0} = 0.8 & \tilde{p}_{0,1} = 0.2 \\ \tilde{p}_{1,0} = 0.2 & \tilde{p}_{1,1} = 0.8 \end{bmatrix} \tag{2.7}$$

The chosen transition probabilities are symmetric such that the likelihood of a rainy day followed by a non-rainy day ($\tilde{p}_{1,0}$) is equivalent to a non-rainy day followed by a rainy day ($\tilde{p}_{0,1}$). Similarly, $\tilde{p}_{0,0}$ and $\tilde{p}_{1,1}$ are equivalent. We opted to choose a high value for $\tilde{p}_{0,0}$ and $\tilde{p}_{1,1}$ in order to increase persistency in the model as this presents a challenge for causal methods to detect causal interactions. Second, the amount of rainfall $\boldsymbol{Y}$ is simulated using a beta distribution $\boldsymbol{Y} \sim \boldsymbol{Beta}(Beta_\alpha, Beta_\beta)$; the parameters $Beta_\alpha$ and $Beta_\beta$ are taken to be 9 [L/T] and 6 [L/T] respectively. Rainfall $\boldsymbol{R}$ is calculated as follows:

$$\boldsymbol{R}(t) = \grave{\boldsymbol{R}}(t).\boldsymbol{Y}(t) \tag{2.8}$$

The evolution of the model variables $\boldsymbol{S}$, $\boldsymbol{I}$ and $\boldsymbol{Q}$ is based on the continuity equation for conservation of mass, and it is described by the following set of equations:

$$\frac{d\,\boldsymbol{S}_t}{dt} = \boldsymbol{R}_{t-1} - \boldsymbol{I}_{t-1} + \boldsymbol{\eta}_s$$

$$\boldsymbol{I}_t = K_s * [\boldsymbol{S}_{t-1} - \delta]^{\xi} + \boldsymbol{\eta}_{\boldsymbol{I}} \tag{2.9}$$

$$Q_t = \begin{cases} \boldsymbol{S}_{t-1} + \boldsymbol{R}_{t-1} - S_{max} + \boldsymbol{\eta}_{\boldsymbol{Q}} & ;\ \boldsymbol{S}_t \geq S_{max} \\ \boldsymbol{\eta}_{\boldsymbol{Q}} & ;\ \boldsymbol{S}_t < S_{max} \end{cases}$$

Where $\boldsymbol{\eta}_{\boldsymbol{s}}$, $\boldsymbol{\eta}_{\boldsymbol{I}}$ and $\boldsymbol{\eta}_{\boldsymbol{Q}}$ are zero mean, constant variance and serially correlated noise (i.e. red noise). $\boldsymbol{\eta}_{\boldsymbol{s}}$ is calculated as follows:

$$\boldsymbol{\eta}_{\boldsymbol{s}(1)} = \boldsymbol{\omega}_1$$

$$\boldsymbol{\eta}_{\boldsymbol{s}(t+1)} = r.\boldsymbol{\eta}_{\boldsymbol{s}(t)} + (1 - r^2)^{\frac{1}{2}}.\boldsymbol{\omega}_{(t+1)} \qquad t > 1 \tag{2.10}$$

Where $r = 0.8$, and $\boldsymbol{\omega} \sim \boldsymbol{N}(0, \sigma^2)$ is a white Gaussian noise, normally distributed with a zero mean and variance σ^2. The variance σ^2 is calculated as follows:

$$\sigma^2 = \frac{\mathbb{E}[S^2]}{SNR} \tag{2.11}$$

Where $\mathbb{E}$ refers to the expected value of S, and SNR is the signal to noise ratio which is added to the model to examine the performance of causal methods under presence of noise. Other noise terms, η_I and η_Q, are calculated similarly by substituting $\mathbb{E}[I^2]$ and $\mathbb{E}[Q^2]$ respectively in equation 2.11. SNR is a measure of the noise level, and it takes values of 2, 3, 4, 5, 10, 20 and 10,000. We also express the strength of noise in decibels dB:

$$dB = 10 \log_{10} SNR \tag{2.12}$$

In causal theory, a very common way to represent causal interactions in a given system is to use directed graphs. Each node in the graph represents a variable or a sub-process in the system, while directed edges represent causal links with the arrow pointing towards the effect. Figure 2.1b shows the directed graph representation of the bucket model in equation 2.9. Note that each variable is a child of the arguments of its function. For example, $\boldsymbol{S}$ is a function of $\boldsymbol{R}$ and $\boldsymbol{I}$; therefore, the node $\boldsymbol{S}$ has the two parents $\boldsymbol{R}$ and $\boldsymbol{I}$. Since $\boldsymbol{R}$ is a forcing variable in the model, it has no parents in the causal graph of the model. In the remaining sections of this paper, we shall use graphical representations and kinship metaphors to discuss causal relations. Figure 2.1c shows the marginal probability distributions of the variables simulated by the model using process noise (dB [SNR] = 40 [10^4]). Clearly, all probability distributions have high density at zero. This is a key feature in many hydrological systems where considerable number of the fluxes observations are equal to zero.

b. Overall Performance

We use three metrics to evaluate the efficiency of causal inference methods in recovering the underlying structure of the hydrological model. These include the True Positives Rate (TPR), False Positives Rate (FPR) and the Detection Rate (DR). Let n_{sim} denote the number of simulations generated from the bucket model, then TPR is defined as the ratio of the number of correctly identified causal links by the algorithm to the actual number of causal links in the model, averaged across simulations.

$$TPR = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \frac{number\ of\ correctly\ identified\ causal\ links}{number\ of\ causal\ links} \tag{2.13}$$

Similarly, FPR is defined as the ratio of mistakenly identified links to the number of actual causal links in the model, averaged across simulations.

$$FPR = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \frac{number\ of\ mistakenly\ identified\ causal\ links}{number\ of causal\ links} \tag{2.14}$$

In equations 2.13 and 2.14, the denominator refers to the actual number of causal links in the model which is equal to 5; see Figure 2.1b. The metrics of TPR and FPR measure the efficiency of the algorithms in retrieving the whole causal structure of the system. However, the detection rate (DR) measures the efficiency of retrieving each causal link separately. It is defined as the percentage ratio of the number of times a causal link was detected by the algorithm to the total number of simulations.

$$DR\ (\%) = \frac{n_{detection}}{n_{sim}} * 100 \tag{2.15}$$

Where $n_{detection}$ refers to the number of times a causal link was either correctly or mistakenly identified by the algorithm.

Figure 2.2 shows the retrieved causal structure using each of the causal discovery algorithms (GC, TE, PC and CCM). The results in Figure 2.2 summarize the mean behavior of the algorithms across 100 simulations; that is, a causal link exists only if it has been identified by the algorithm in more than 50% of the simulations. By comparing the retrieved causal structure using GC (Figure 2.2a) and the true causal structure of the model (Figure 2.1b), all the true causal links were correctly identified by GC (blue links in Figure 2.2a). However, GC mistakenly identifies three causal links ($\boldsymbol{R} \Longrightarrow \boldsymbol{I}$, $\boldsymbol{Q} \Longrightarrow \boldsymbol{I}$ and $\boldsymbol{I} \Longrightarrow \boldsymbol{Q}$) (red links in Figure 2.2a). These false detections are attributed to the inability of the GC method to control for mediation and confounding of nonlinear relationships. For the link $\boldsymbol{R} \Longrightarrow \boldsymbol{I}$, the relationship between $\boldsymbol{R}$ and $\boldsymbol{I}$ is mediated by $\boldsymbol{S}$, that is, $\boldsymbol{R} \Longrightarrow \boldsymbol{S} \Rightarrow \boldsymbol{I}$. As for the relationship between $\boldsymbol{Q}$ and $\boldsymbol{I}$, the two variables share a common confounder which is $\boldsymbol{S}$, that is, $\boldsymbol{Q} \Leftarrow \boldsymbol{S} \Longrightarrow \boldsymbol{I}$. The three links would not be mistakenly identified if the GC method conditions properly on variable $\boldsymbol{S}$. Although GC conditions on variable $\boldsymbol{S}$, the nonlinearity of these relationships violate the assumption of linearity in GC method; therefore, resulting in false detection.

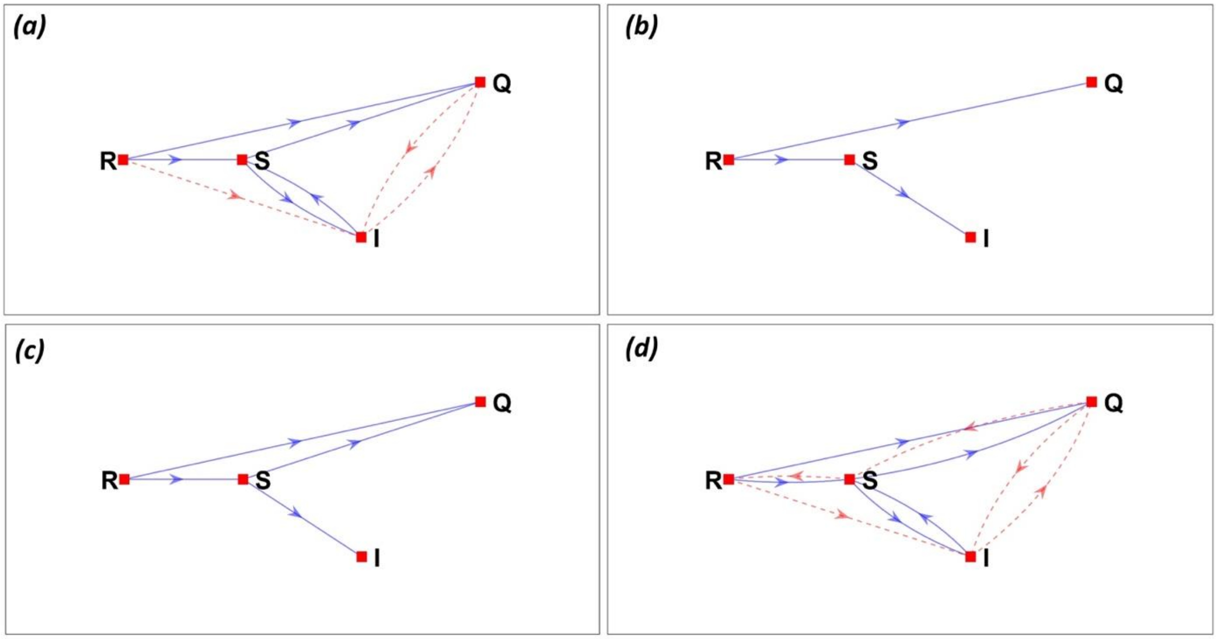


Figure 2.2 *Causal structure of the hydrological model in Figure 2.1a retrieved from 100 simulations of time series each with a length of 3000 using:* ***(a)*** *Granger Causality (GC),* ***(b)*** *Transfer Entropy (TE),* ***(c)*** *PC algorithm and* ***(d)*** *Convergent Cross Mapping (CCM). Blue links indicate causal links correctly identified by the algorithm (true positives). Red links indicate causal links falsely identified by the algorithm (false positives).*

By comparing the results of GC with the causal structures retrieved from TE (Figure 2.2b) and PC (Figure 2.2c), one can see that both TE and PC rule out the links mistakenly identified by GC. This is because both TE and PC are nonparametric methods; thus, they are able to detect nonlinear relationships. However, the two algorithms fail to detect the causal link $\boldsymbol{I} \Longrightarrow \boldsymbol{S}$; this causal link is in fact a feedback link such that variable $\boldsymbol{S}$ causes variable $\boldsymbol{I}$ which in turn feedback and impact variable $\boldsymbol{S}$. The reason behind under-detection of this link is that both algorithms accept the null hypothesis of the independence test $\mathbb{I}(\boldsymbol{S}; \boldsymbol{I} |\ \boldsymbol{R})$; that is, variable $\boldsymbol{S}$ is independent of variable $\boldsymbol{I}$ conditioned on variable $\boldsymbol{R}$. Specifically, variable $\boldsymbol{I}$ negatively impact variable $\boldsymbol{S}$ while variable $\boldsymbol{R}$ has a positive impact on variable $\boldsymbol{S}$, and both effects negate each other to maintain the mass balance

$\frac{\boldsymbol{dS}}{\boldsymbol{dt}} = \boldsymbol{R} - \boldsymbol{I}$. This type of relationships is a typical example for violations of the causal faithfulness assumption which states that "there are no precisely counterbalanced causal relationships in the system that would result in a probabilistic independence between two variables that are actually causally connected" (Andersen, 2013). For TE method, in addition to its inability to detect the causal link $\boldsymbol{I} \Longrightarrow \boldsymbol{S}$, the algorithm also fails to detect the link $\boldsymbol{S} \Longrightarrow \boldsymbol{Q}$. In evaluation of this causal link, the TE method examines the independence relationship $\mathbb{I}(\boldsymbol{S}; \boldsymbol{Q} | \{\overline{\boldsymbol{I}}_{\boldsymbol{t-1}}, \overline{\boldsymbol{R}}_{\boldsymbol{t-1}}, \overline{\boldsymbol{Q}}_{\boldsymbol{t-1}}\})$; that is, whether the variables $\boldsymbol{S}$ and $\boldsymbol{Q}$ are independent given the history of variables $\boldsymbol{I}, \boldsymbol{R}$ and $\boldsymbol{Q}$. Because the number of conditioning variables (i.e. $\overline{\boldsymbol{I}}_{\boldsymbol{t-1}}, \overline{\boldsymbol{R}}_{\boldsymbol{t-1}}, \overline{\boldsymbol{Q}}_{\boldsymbol{t-1}}$) is relatively large, and the relationship $\boldsymbol{S} \Longrightarrow \boldsymbol{Q}$ is a weak causal relationship, the detection rate of this link is low. Therefore, the link can only be detected for large sample size; see Figure 6 of Ombadi et al. (2020a).

Like GC, CCM mistakenly identifies the three links: $\boldsymbol{R} \Longrightarrow \boldsymbol{I}$, $\boldsymbol{Q} \Longrightarrow \boldsymbol{I}$ and $\boldsymbol{I} \Longrightarrow \boldsymbol{Q}$ (see Figure 2.2d). This is because CCM does not control for confounding and mediation. Additionally, CCM mistakenly identifies the relationship in the pairs ($\boldsymbol{S}, \boldsymbol{Q}$) and ($\boldsymbol{R}, \boldsymbol{S}$) as bidirectional causality rather than unidirectional. This points out to a limitation of CCM that when two variables are strongly coupled (synchronized), CCM identifies unidirectional causality as bidirectional. Sugihara et al. (2012) reported that in the case of extremely strong forcing, CCM will result in bidirectional causality between variables.

c. Impact of Sample Length

Here we analyze the sensitivity of each algorithm to changes in sample size; we perform the analysis over sample size of 100, 300, 500, 1000, 2000 and 3000. Each analysis is performed with a number of simulations (n_{sim} = 100). Figures 2.3a and 2.3b show the TPR and FPR respectively for each of the four algorithms. As can be seen in Figure 2.3a, TPR consistently increases with increasing sample length. At the limit of large sample length ($l = 3000$), both CCM and GC approach a TPR equal to 1 (i.e. all causal links are correctly detected by the algorithms). On the contrary, PC and TE approach a TPR < 0.8; this is because the feedback link $\boldsymbol{I} \Longrightarrow \boldsymbol{S}$ is not detected as illustrated in the previous section. The most important finding to note in Figure 2.3a is the insensitivity of CCM to sample length regarding the TPR. It shows that a sample size as small as 100 is sufficient for the CCM to identify all causal links in the model. On the other hand, GC, TE and PC show sensitivity to sample size; this is not surprising since the three methods are based on a probabilistic framework, and the statistical estimation improves as the sample size increases.

As for the false positives in Figure 2.3b, the results might seem counterintuitive as one would expect the FPR to consistently decrease with sample size. However, this is not necessarily the case when the false detection is not related to sample size. For example, both CCM and GC show increasing FPR with the increase in sample size. This is because the mistakenly identified links result from the limitations of the algorithms in controlling for confounding and mediation. These types of false detection increase as the sample size increases.

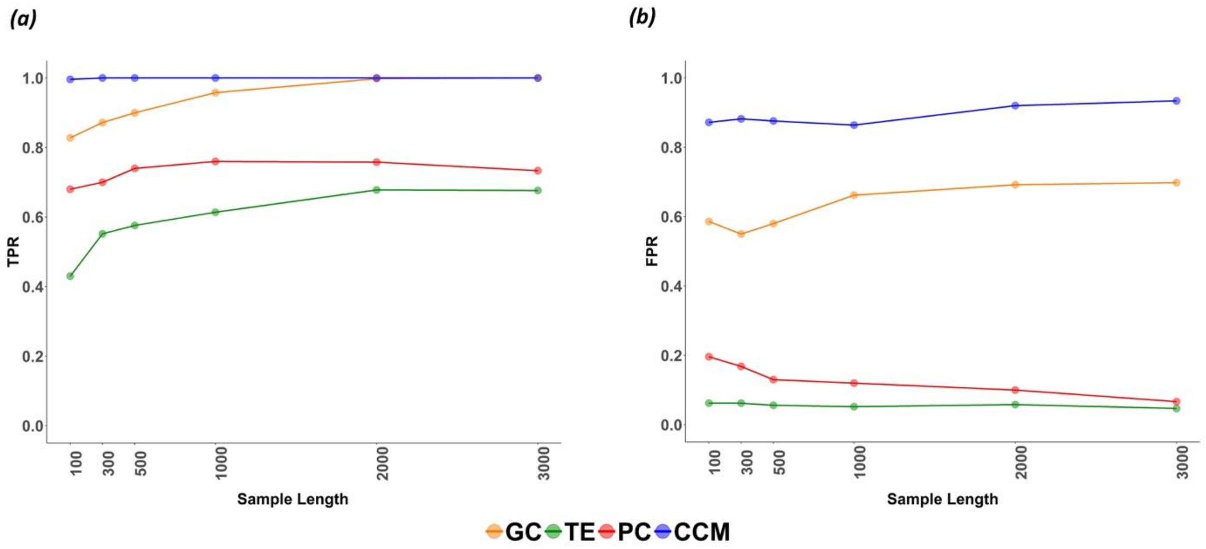


Figure 2.3 (a) *True positives rate (TPR) for each of the four causal algorithms (GC, TE, PC and CCM) for sample length of 100, 300, 500, 1000, 2000 and 3000; each averaged across 100 simulations.* ***(b)*** *same as in (a) but for the false positives rate (FPR).*

d. Presence of Noise

To assess the sensitivity of performance to presence of noise, we first analyze the sensitivity of performance to process noise. Unlike observational noise that is associated with errors in measurements, process noise means that there is a stochastic component in the underlying system. In the hydrologic model (equation 2.9), if the variance of $\boldsymbol{\eta_I}, \boldsymbol{\eta_s}$ and $\boldsymbol{\eta_Q}$ is zero, the model is completely deterministic and there is no process noise. As the variance takes values larger than zero, the model incorporates a stochastic component. In hydrometeorological systems, process noise can arise even in well-defined deterministic systems because of heterogeneity. For example, the rainfall-runoff process in catchments is not entirely deterministic as it has some stochastic component due to heterogeneity associated with land properties. Figures 2.4a and 2.4b show the TPR and FPR for dB[SNR] of 3[2], 4.8[3], 6[4], 7[5], 10[10], 13[20] and 40[10^4]; each is averaged across 40 simulations with a sample length of 1000. The model specifications of parameters and initial states are as shown in Table 2.1. As can be seen in Figure 2.4a, process noise has a minimal impact on the performance of the three methods (GC, TE and PC) with a slight increase in TPR as the level of noise is increased. This is because process noise is part of the dynamics, and variables remain stochastically coupled as the noise level increases. As a result, GC, TE and PC which assume the underlying system to be of stochastic nature, are able to maintain their performances as the noise increases. On the contrary, TPR of CCM decreases significantly for noise levels (dB < 4.8). The discernible decrease in performance of CCM in presence of process noise is expected since the method is based on an assumption of deterministic systems. However, the results also suggest that CCM can tolerate process noise up to 4.8 dB. Figure 2.4b shows that FPR for all the methods decreases as the noise level increases (i.e. lower dB). While this can be justified for the methods of GC, TE and PC due to their probabilistic framework, the results appear to be counterintuitive regarding CCM. However, the reason behind this is that as the process noise increases (lower dB), all variables in the system no longer contain dynamic information about each other. Consequently, the cross-mapping ability of CCM diminish leading to a decrease in both TPR and FPR.

Table 2.1 *Specifications of the model for the analysis in subsection b, c and d.*

Maximum soil storage (S_{max}) [$\mathbf{L}$]	80
Storage-discharge parameter 1 (K_s) [$\mathbf{1/T}$]	2.3
Storage-discharge parameter 2 (δ) [$\mathbf{L}$]	10
Storage-discharge parameter 3 (ξ)	0.6
Process noise (dB [SNR])	40 [10^4]
Initial soil storage ($\boldsymbol{S}_0$) [$\mathbf{L}$]	40

The second part of the analysis is examining the sensitivity of performance to presence of observational noise. For observational noise, time series are simulated from the model (equation 2.9) with a small process noise (SNR= 10^4; dB= 40). Then, observational noise is added after the time series are simulated from the model with noise levels in dB[SNR] of 3[2], 4.8[3], 6[4], 7[5], 10[10], 13[20] and 40[10^4]. This type of noise represents measurement error where the noise is not a result of the underlying system, but it is associated with the devices measuring the data. Figures 2.4c and 2.4d show TPR and FPR for different levels of noise each averaged across 40 simulations with a sample length of 1000. The model specifications of parameters and initial states are as shown in Table 2.1. Unlike in the case of process noise where the three methods of GC, TE and PC are insensitive to changes in noise level, the results here show that these methods in addition to CCM are all sensitive to presence of observational noise. Specifically, the performance of GC, TE and PC deteriorates as evidenced by a decrease in TPR (Figure 2.4c) and an increase in FPR (Figure 2.4d). As for CCM, both TPR and FPR decrease consistently with the increase in observational noise. This is because as the noise increases, causally related variables in the system no longer contain information signature of each other; thus, the efficiency of cross mapping degrades.

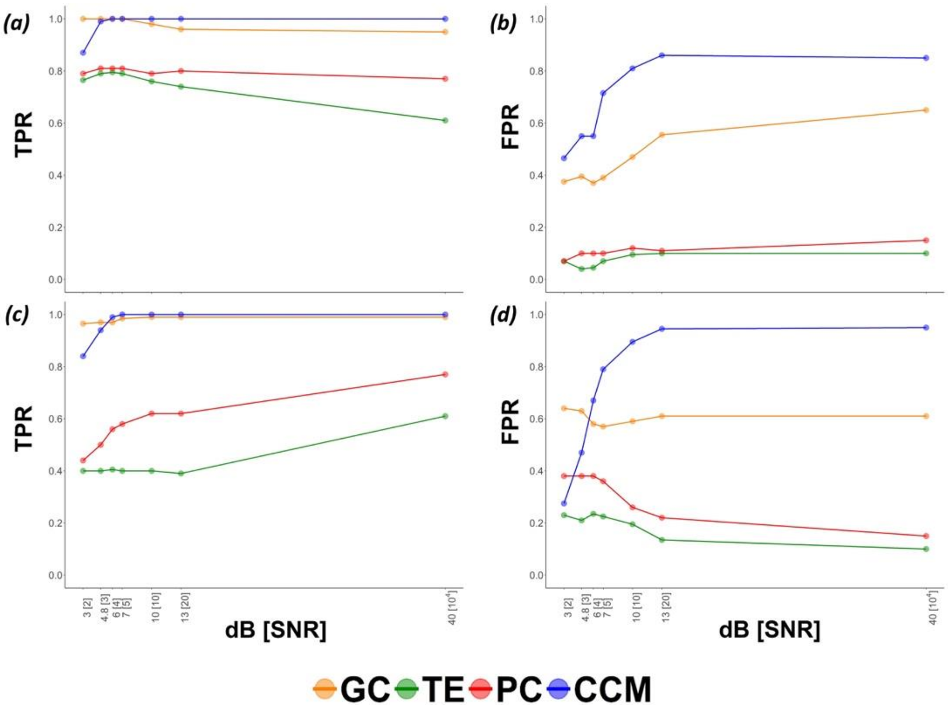


Figure 2.4 *Sensitivity of the causal algorithms (GC, TE, PC and CCM) to presence of noise with noise levels expressed as decibels (dB): 3, 4.8, 6, 7, 10, 13 and 40. Noise levels are also expressed in SNR: 2, 3, 4, 5, 10, 20 and 10^4. The plotted values are averaged across 40 simulations for each noise level.* ***(a)*** *True positives rate (TPR) under presence of process noise.* ***(b)*** *False positives rate (FPR) under presence of process noise.* ***(c)*** *True positives rate (TPR) under presence of observational noise.* ***(d)*** *False positives rate (FPR) under presence of observational noise.*

2.4 Causal Analysis of Environmental Drivers of Evapotranspiration

Evapotranspiration $\boldsymbol{ET}$ plays a central role in the Earth's water and energy cycles, and it is the primary process in the biosphere-atmosphere coupling. Several factors can potentially regulate evapotranspiration rate; these include net radiation $\boldsymbol{R_n}$, vapor pressure deficit $\boldsymbol{VPD}$, soil water content $\boldsymbol{SWC}$, air temperature $\boldsymbol{T_a}$, soil temperatures $\boldsymbol{T_s}$ and wind speed $\boldsymbol{WS}$. $\boldsymbol{R_n}$, $\boldsymbol{VPD}$ and $\boldsymbol{SWC}$ are considered as direct drivers of $\boldsymbol{ET}$, while other variables ($\boldsymbol{T_a}$, $\boldsymbol{T_s}$ and $\boldsymbol{WS}$) affect $\boldsymbol{ET}$ primarily through their regulation of canopy stomatal conductance. Several models with wide range of complexity exist to understand and simulate evapotranspiration process; however, modelling large-scale evapotranspiration remains a major source of uncertainty (Sivapalan et al., 2003; Mackay et al., 2007). Therefore, analysis of observational data is important to assess the significance of environmental controls on evapotranspiration and their seasonal and regional variations. Observational datasets were used by Vrugt et al. (2002) along with artificial neural networks to identify controlling factors of transpiration in a forested region, while Mackay et al. (2007) used observational data to quantify the differential impact of net radiation and vapor pressure deficit in regulating evapotranspiration in upland and wetland regions. In this section, we use the PC algorithm to identify the forcing environmental variables that control evapotranspiration rate and their relative contributions during summer and winter seasons in a shrubland region.

Observational dataset used in our analysis is obtained from Santa Rita Mesquite (US-SRM) FluxNet site. This site is located in southeastern Arizona (31.82°N, 110.87°W) at an elevation of 1118 m above sea level. The Koeppen climate classification of the site is Arid Steppe cold (BSk). The land cover is broadleaf vegetation shrublands, and it consists primarily of mesquite (Prosopis velutina) trees. Mean annual precipitation and temperature are 333 mm and 19°C respectively. Hourly time series were obtained by accumulating and averaging the native 30-minutes observations for the following variables: $\boldsymbol{ET}$, $\boldsymbol{R_n}$, $\boldsymbol{VPD}$, $\boldsymbol{SWC}$, $\boldsymbol{T_a}$, $\boldsymbol{T_s}$ and $\boldsymbol{WS}$. Table 2.2 shows the statistics of mean and standard deviation of each variable for the summer and winter seasons. Anomalies of hourly observations were calculated by subtracting the seasonal mean to remove effects of diurnal cycle. Time series were then tested for stationarity (monotonic trend) using Mann Kendall test; Table 2.2 shows p-values for each time series. When the null hypothesis of no trend was rejected at a significance level of 0.05, we removed a linear trend from the time series. The sample length for the summer season (JJA) and the winter season (DJF) is 24,288 and 23,832 observations respectively. They represent hourly observations at the US-SRM FluxNet site during the period 2004 -2014. We used the PC algorithm to infer the environmental drivers controlling evapotranspiration because, according to the results in the preceding section, it controls FPR (almost similar to that of TE) when the sample length is sufficient while achieving a higher TPR value than that of TE.

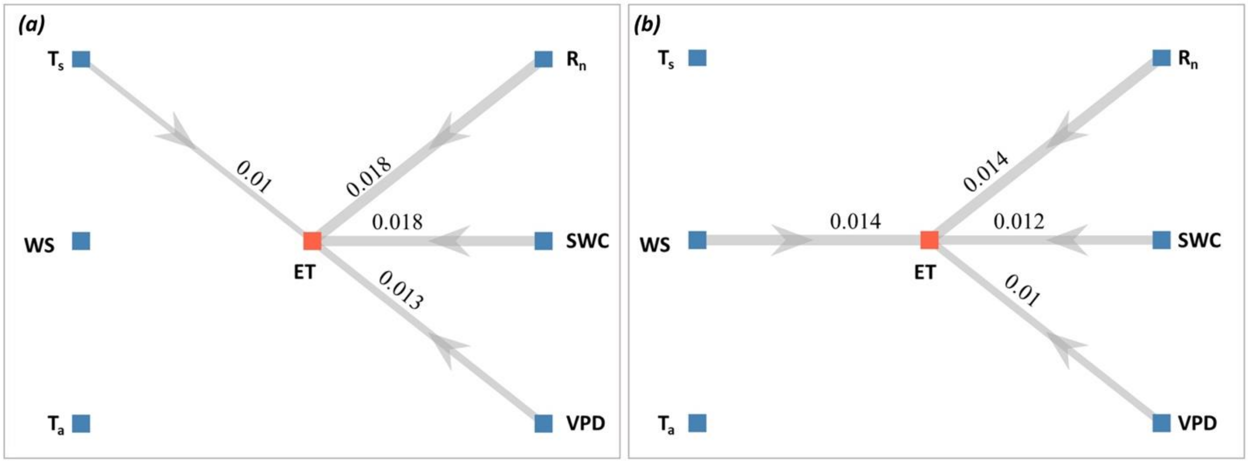


Figure 2.5 *Causal networks retrieved from the PC algorithm using hourly time series of the variables (ET, Rn, VPD, SWC, Ta, Ts and WS). Labels of the edges represent the strength of the causal links as determined by the values of conditional mutual information. The thickness of the edges between the nodes is proportional to the strength of the causal links.* ***(a)*** *Causal network of ET environmental drivers during the summer season (June, July and August).* ***(b)*** *Causal network of ET environmental drivers during the winter season (December, January and February).*

Figures 2.5a and 2.5b show the two causal networks obtained from the PC algorithm using hourly observations during summer and winter seasons. During the summer season (JJA) (Figure 2.5a), evapotranspiration rate is regulated by, in order of importance, net radiation and soil water content (equally important), vapor pressure deficit and soil temperature. The two remaining variables, wind speed and air temperature, are not causally related to evapotranspiration at a statistical significance level of 0.05. On the other hand, evapotranspiration during the winter season (DJF) (Figure 2.5b) is controlled by, in order of importance, net radiation and wind speed (equally important), soil water content and vapor pressure deficit. In order to understand the physics underlying these results, it is important to examine the dynamics of the vegetation cover on the site. Figure 2.6a shows the monthly variability in Gross Primary Production (GPP) averaged over the period (2004-2014). Clearly, GPP peaks during the summer because Mesquite trees which dominate the land cover at the site bloom and grow during the summer. On the contrary, GPP is very low during the winter. This means that during the winter season, bare soil evaporation is the predominant portion of evapotranspiration due to the limited vegetation cover. However, in the summer, transpiration represents a large portion of evapotranspiration. This provides an interpretation of the results that soil temperature is a causal factor only during the summer season because of its impact on regulating water uptake in plants, whereas it has no discernible impact on bare soil evaporation during the winter. Effect of soil temperature on water uptake and stomatal opening in plants was previously reported in Kramer (1940) and Feldhake & Boyer (1986) among others.

Table 2.2 *Mean and standard deviation for the hourly time series of the seven variables used in this analysis during summer (June, July and August) and winter (December, January and February). The p-vlaues of Mann Kendall trend test are also shown; the null hypothesis of no trend is rejected if the p-value is smaller than the significance level (0.05).*

	Summer (JJA)			Winter (DJF)		
	Mean	*Standard Deviation*	*p-value (Mann-Kendall)*	*Mean*	*Standard Deviation*	*p-value (Mann-Kendall)*
ET (mm)	0.07	0.09	0	0.02	0.03	0
T_a (°C)	27	4.7	0.01	10.5	5.6	0.01
WS (m/s)	2.29	1.22	0.85	2.51	1.52	0
T_s (°C)	32.2	6.2	0.07	12.54	4.7	0.31
VPD (hPa)	24.3	13.5	0.04	8.82	5.6	0.23
SWC (%)	4.56	2.77	0	5.82	2.15	0.07
R_n (W/m^2)	283.11	391.24	0.09	100	245.61	0.36

Figure 2.5b shows that wind speed plays a major role in controlling evapotranspiration in the winter season. Given that the primary effect of wind speed is to clear the air of humidity produced by evapotranspiration, it might be plausible that wind speed is not a significant causal factor during the summer because the advected air is humid. Advection of moisture during the summer at low levels of the atmosphere (geopotential heights greater than 800 mb) towards southwestern US is a key feature of the North American Monsoon, and it has been reported in several studies (e.g. Adams & Comrie, 1997). Furthermore, Figures 2.6b and 2.6c show the diurnal cycle of wind speed during summer and winter seasons respectively. Clearly, in the summer, maximum wind is in late afternoon (5 pm) lagged by several hours from the peak of evapotranspiration (noon). However, during winter season (Figure 2.6c), the lag time is shorter; thus, wind speed and evapotranspiration are nearly in-phase. Consequently, wind speed can clear the air of humidity and regulate evapotranspiration rate. It should also be noted that wind speed during the winter has larger variability (standard deviation = 1.52 m/s) compared to the summer (standard deviation = 1.22 m/s); See Table 2.2.

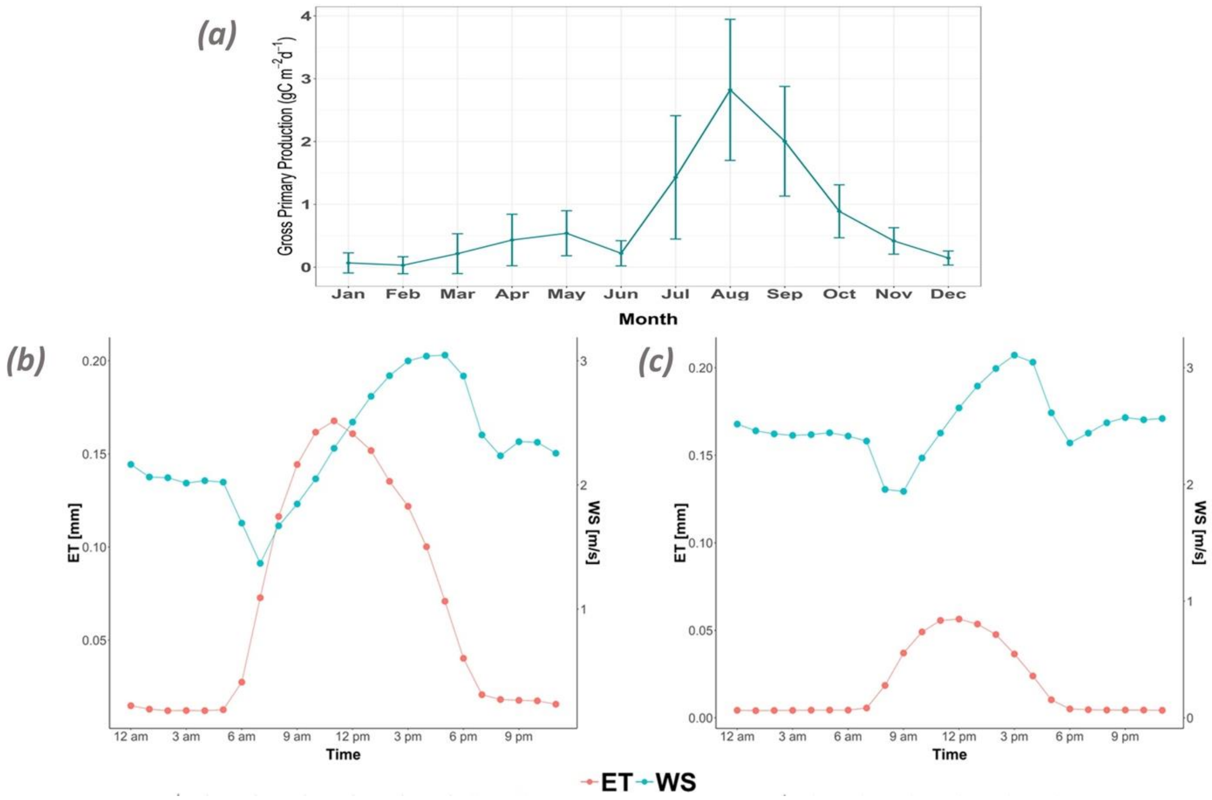


Figure 2.6 (a) *Monthly variability in Gross Primary Production (GPP) averaged over the period (2004-2014). Line plot shows the mean monthly GPP, while the vertical bars indicate the standard deviation of monthly GPP. GPP data used in this figure represents Gross Primary Production from Nighttime partitioning method.* ***(b)*** *Diurnal cycles of ET (red) and wind speed WS (cyan) during the summer season.* ***(c)*** *Diurnal cycles of ET (red) and wind speed WS (cyan) during the winter season.*

2.5 Conclusions

This study examined the efficiency of four causal inference methods (GC, TE, PC and CCM) in retrieving the causal structure from simulated time series of a hydrological model used as the “ground truth”. The hydrological model consists of four variables linked by relationships commonly found in hydrometeorological systems. Despite the low number of variables in the model, it represents a challenging task due to the large causal search space. Specifically, for each pair of variables (a, b), there are four possibilities: no causal relation, bidirectional causality, unidirectional causality from a to b, or unidirectional causality from b to a. This leads to $4^{\binom{4}{2}} =$

4096 possible causal structures to explain the time series simulated from the model. Based on the results of this study, it is found that several factors can guide the selection of a causal discovery algorithm as follows:

- **Sample length**: CCM is the least sensitive method to changes in sample length. The results demonstrate that a sample size as small as 100 is sufficient for CCM to identify all causal relationships in the model. On the contrary, the performance of GC, TE and PC improves as the sample length increases. This is attributed to the fact that they are based on a probabilistic framework; thus, statistical estimation improves as the sample size increases.

- **Nonlinearity:** among the four methods used in this study, GC is the only method that assumes a linear vector autoregressive (VAR) model for the underlying system. Despite this assumption, the results show that GC was able to detect nonlinear interactions; for example, the causal links $\boldsymbol{S} \Longrightarrow \boldsymbol{I}$ and $\boldsymbol{S} \Longrightarrow \boldsymbol{Q}$ (see Figure 2.2a). This may support the argument that many nonlinear processes can be modelled as VARs (Barnett & Seth, 2014). However, the results show that the assumption of modelling the system as a linear model while has no impact in detecting causal links, it leads to an increase in false positives.

- **Stochastic vs Deterministic systems:** While CCM assumes the underlying system to be deterministic, the three methods of GC, TE and PC are based on assumption of stochastic systems. In this study, the system is inherently deterministic, and the evolution of its variables is described through dynamical equations. However, we examined the performance of the algorithms by adding different levels of process noise which adds a stochastic component to the system. The results demonstrate that CCM can tolerate process noise up to 4.8 dB ($SNR = 3$).

- **Presence of counterbalanced relationships:** Attention must be paid when the system under study is expected to maintain counterbalanced interactions to fulfill physical laws such as conservation of mass, momentum and energy. These types of relationships are typical in hydrometeorological systems, and they represent a violation of the faithfulness assumption. Methods based on conditional independence, PC and TE, are unable to detect such relationships as evidenced by their inability to detect the causal link $\boldsymbol{I} \Longrightarrow \boldsymbol{S}$ (see Figures 2.2b and 2.2c).

- **Presence of observational noise:** As expected, presence of observational noise degrades the performance of all causal discovery methods. Specifically, the impact is more significant in the case of CCM in which the cross-mapping efficiency between variables diminish as the observational noise increases.

- **Confirmatory vs Exploratory studies:** The results indicate the existence of a tradeoff between TPR and FPR (see Figure 2.3). Therefore, If the purpose of a given study is exploratory, e.g. searching for climatic teleconnections of a certain phenomenon, then one might consider using GC or CCM due to their high TPR compared to TE and PC. On the contrary, if a study is confirmatory, e.g. selecting significant climatic teleconnections from a set of predefined teleconnections, then using TE and PC is more appropriate as they minimize the false detection.

A secondary aim of this study was to examine environmental drivers of evapotranspiration and their relative contributions during summer and winter seasons. The PC algorithm was applied as a causal discovery algorithm along with observational time series of the following variables: ET, R_n, VPD, SWC, T_a, T_s and WS. The results show that environmental drivers are dependent on season. While R_n, VPD and SWC are key drivers in regulating evapotranspiration in both seasons, the results demonstrate that T_s is a significant driver only in summer season, and WS controls evapotranspiration in winter season. The obtained results from causal analysis represents a hypothesis which can either be refuted or confirmed through further investigation. We provided an interpretation of the results based on the canopy seasonal dynamics and basic understanding of the evapotranspiration process. Hydrologic models commonly use a single relationship to estimate ET using a specific set of environmental drivers without prior information on which variables are dominant and significant in regulating ET. The results presented in this study highlight the importance of selecting ET models that are sensitive to the key drivers in each season. Similar causal analysis can be applied to investigate the differential impact of environmental drivers on evapotranspiration in sites across a range of climate conditions.

Chapter 3

3 Hydrologic Complexity: A Chaotic Dynamics Perspective

“This chapter is extracted from Ombadi et al. (2021a) with few edits incorporated for brevity and clarity”

3.1 Introduction

The movement and storage of water within hydrologic basins are regulated by a multitude of processes operating at various spatiotemporal scales; they include, among others, flow of water along hill slopes, subsurface flow through porous media, water uptake by plants and snowmelt. In addition to the inherent nonlinearity of these processes, their intricacy is further aggravated by the heterogeneity in landscape, drainage network, soil moisture, vegetation as well as the variability of precipitation and net radiation. The interplay among the aforementioned processes leads to emergent properties at the macroscale. Such properties essentially encapsulate the salient features of hydrological processes within the basin, and they provide a parsimonious description of basin behavior. The notion of basin complexity which refers to the number of dominant variables and the nonlinearity of processes governing watershed dynamics is often discussed as an element of a generic framework that seeks to characterize the essential signatures of watershed dynamics. Such a viewpoint, so-called “top-down” or “downward” approach, espoused by several hydrologists (Gupta et al., 2008; Klemes, 1983; McDonnell et al., 2007; Sivapalan et al., 2003) is considered to be useful to overcome the challenges facing watershed hydrology. The interest in basin complexity stems primarily from its usefulness in catchment classification as it facilitates and provides a framework for transferring models and prediction in ungauged basins. Additionally, it provides a basis for selection of models with appropriate complexity to resolve watershed dynamics. In this latter regard, basin complexity is closely tied to concepts of model parsimony, parameters identifiability and criteria of model selection (Bai et al., 2009; Beven, 2006; Butts et al., 2004).

Previous work on basin complexity has attempted to identify hydrologic complexity by extracting signatures from univariate or bivariate time series, either observed or simulated. For example, Jakeman and Hornberger (1993) interpreted basin complexity as the number of parameters in a model that can be supported by the information contained in rainfall and streamflow observations. Alternatively, Castillo et al. (2015) proposed an information-theoretic entropy measure that characterizes the spatial variability of soil moisture within the basin as a complexity index. They argued that soil moisture has first-order control over the basin behavior as it modulates the indirect effects of other variables. Recently, Pande and Moayeri (2018) used a statistical measure of complexity (Vapnik-Chervonenkis dimension) along with streamflow observations to identify the complexity of 412 watersheds across the Contiguous United States (CONUS). More related to the study presented here, Sivakumar and Singh (2012) and Vignesh et al. (2015) utilized theories of chaotic dynamical systems, namely time-delay embedding and phase space reconstruction, to identify the dimensionality of basins. The former examined basin complexity at 117 gauging stations in the western United States, and the latter identified basin dimensionality at 639 stations in CONUS. In the present study, we also rely on the theories of chaotic dynamical systems to identify complexity of hydrologic basins.

Chaotic dynamical systems are characterized by an apparently random and erratic behavior, despite being governed by nonlinear deterministic interactions. Although the concept of chaos dates back to the 1960's, it was only until the late 1980's that chaos theory was applied to the analysis of time series in environmental sciences. This can be ascribed to the development of the time-delay embedding theorem (Takens, 1981) which facilitated the reconstruction of system dynamics in multidimensional phase space using solely time series of a single variable. The idea that clues on the multivariate behavior of a system can be obtained from a one-dimensional time series has led to a surge in chaotic analysis of hydrologic time series (e.g. Elshorbagy et al., 2002; Lall et al., 1996; Porporato & Ridolfi, 1996; Rodriguez-Iturbe et al., 1989). This early research was guided by two motives: first, searching for the presence of chaotic determinism in system dynamics; second, utilizing chaotic determinism as a practical tool for prediction and estimation.

3.2 Data

The Model Parameter Estimation Project (MOPEX) dataset (Duan et al., 2006; Schaake et al., 2006) was assembled to provide hydrometeorological data from a large number of basins for parameterization of hydrologic and land surface models. The dataset provides high quality hydrometeorological observations of key variables (e.g. streamflow, precipitation, maximum and minimum daily temperature …etc.) as well as basin characteristics for 431 catchments in CONUS. MOPEX daily streamflow observations, the primary data used in this study, are a subset of USGS hydro-climatic data network (HCDN) streamflow observations. This dataset is suitable for climatic studies since it only consists of basins free from upstream flow regulations such as artificial diversions and storage. Here, we exclude stations with a record length less than 30 years; the remaining basins span a wide range of hydrologic, climatic and land surface characteristics. In addition to the streamflow observations, MOPEX dataset provides several climatic, topographic and geomorphologic characteristics of basins. Table 3.1 shows the basin characteristics used in this study for the purpose of investigating their control on dynamic complexity.

3.3 Methods

Let's consider an m-dimensional dynamical system; that is, the system dynamics arise from a set of *m* variables. This set is denoted by $\{X_k\}$, where $k = 1, 2, \dots m$. At any time t, the system can be represented in the phase space by the *m*-dimensional point:

$$\boldsymbol{x}(t) = \{x_1(t),\ x_2(t), \dots x_m(t)\} \tag{3.1}$$

Note that we use uppercase letter X_k to denote the entire time series of variable *k*, lowercase letter $x_k(t)$ to denote a single observation of X_k at time *t*, and bold lowercase letter $\boldsymbol{x}(t)$ to denote a single observation in the phase space at time *t*. In the case where only a single variable is observed, e.g. X_1, the theory of time-delay embedding (Takens, 1981; see also Packard et al., 1980) demonstrates that the topological features of the system manifold in the phase space is preserved if the system is represented by the time series of variable X_1 and its values shifted by a fixed delay time τ. For convenience, we shall remove the subscript in X_1 hereafter since only a single variable is considered; therefore, at any time t, the system is represented by the *E*-dimensional point:

$$\boldsymbol{x}(t) = \{x(t), x(t-\tau), x(t-2\tau), \dots x(t-(E-1)\tau)\} \tag{3.2}$$

In the above equation, E is the Embedding dimension. Note that E is not necessarily equal to the true dimension of the system m. Also, the delay time τ is a multiple of the sampling time step Δt; thus, the smallest value of τ is Δt. In the present study, the observed variable X is streamflow, whereas the remaining $m - 1$ variables are unobserved (these variables correspond to rainfall, soil moisture, air temperature …etc.). Therefore, we consider hereafter the phase space representation given by equation 3.2. Our two measures of dynamic complexity of hydrological basins are estimated from two methods generally categorized as nonlinear prediction methods. Essentially, nonlinear prediction entails that there exists a function in the E-dimensional phase space that describes the evolution of the system as follows:

$$\boldsymbol{x}(t+T) = f_T[\boldsymbol{x}(t)] \tag{3.3}$$

Where T is the prediction horizon (i.e. how far ahead) and f_T is a function that maps the phase space representation of the system at time t to the future state at time $t + T$. It is possible that an approximation of the function f_T can be computed from the samples to predict the future state of $\boldsymbol{x}(t)$ as follows:

$$\widehat{\boldsymbol{x}}(t+T) = \hat{f}_T\ [\boldsymbol{x}(t)] \tag{3.4}$$

The function $\widehat{f_T}$ is local (i.e. depends on the location of $\boldsymbol{x}(t)$ in the phase space) and nonlinear, hence the name "nonlinear prediction". Within the field of chaos theory, several methods have been proposed to approximate $\hat{f}_T$. Here, we consider two methods, namely the simplex nonlinear prediction method (Sugihara & May, 1990) and the sequential locally weighted global linear maps "S-map" (Sugihara, 1994). By means of these methods, two complexity measures can be identified; one is a proxy of the system dimensionality (i.e. active degrees of freedom) and the other is an indicator of the strength of nonlinearity. The two following subsections will briefly describe the two nonlinear prediction methods and the corresponding complexity indices.

a. Simplex method and identification of optimum embedding dimension

As discussed above, the objective of Simplex method (Sugihara & May, 1990) is to approximate the function $\hat{f}_T$ in equation 3.4. In doing so, the simplex method starts by constructing the manifold in the phase space using equation 3.2. Specifically, the manifold is constructed using the trajectory of the E-dimensional point $\boldsymbol{x}(t)$ for time series of length l. An example of such a manifold is shown in Figure 3.1a which is constructed from streamflow observations at gauging station (ID 01560000). Note that in this case, the phase space is 3 dimensional ($E = 3$). If $\boldsymbol{x}(l + T)$ is to be predicted, the E-dimensional point $\boldsymbol{x}(l)$ is located in the phase space, and then its $E + 1$ nearest neighbors are identified. Next, the locations of the $E + 1$ points on the phase space after T time steps are identified. Finally, $\boldsymbol{x}(l + T)$ is obtained by calculating a weighted average of these locations. The weights are exponential based on the distances of nearest neighbors from the point $\boldsymbol{x}(l)$. The intuition here is that the future system state can be predicted from the behavior of past values that are similar to the current state, with similarity being defined as closeness in the phase space.

The Simplex method can be used to estimate an optimum embedding dimension hereafter referred to as E_{opt}. This is done by iteratively selecting an E value ($E = 1, 2, 3$ ...) and using a subset of the observed time series to predict the remaining values. For each E value, the predicted and observed values are compared using correlation coefficient ρ. The value of E that achieves best prediction (maximum ρ) is the optimum embedding dimension. Simply, E_{opt} is the dimension that effectively unfolds system dynamics in the phase space as it eliminates singularities and trajectory crossings. It is, therefore, an indicator of the true system dimensionality m. Formally, Whitney embedding theorem (Whitney, 1944) gives the following inequality: $E_{opt} > 2m - 1$. This means that E_{opt} provides an upper bound on the true system dimensionality. However, because the time series we use in this study are finite and not noise-free, E_{opt} is only regarded to be proportional to m, not a strict upper bound on m. Figure 3.1b shows prediction accuracy for a range of E values using the same streamflow time series in Figure 3.1a. In this case, the optimum embedding dimension is 3.

b. S-map method and identification of nonlinearity index

The essential idea of the S-map method (Sugihara, 1994) for the prediction of $x(t+T)$ is to utilize a regression function that maps the *E+1* dimensional vector $\tilde{\boldsymbol{x}}(t)$ to the one dimensional point $x(t+T)$ (i.e. $\mathbb{R}^{E+1} \rightarrow \mathbb{R}^1$) using the following equation:

$$x(t+T) = \sum_{k=1}^{E+1} c_k \ \tilde{\boldsymbol{x}}_k(t) \tag{3.5}$$

Here, $\tilde{\boldsymbol{x}}(t)$ is equivalent to the *E*-dimensional point $\boldsymbol{x}(t)$ augmented by adding a constant term. Specifically, $\tilde{\boldsymbol{x}}_1(t) = 1$ and $\tilde{\boldsymbol{x}}_k(t)$ for $k = 2, 3, \dots E + 1$ is equivalent to the point $\boldsymbol{x}(t)$. The coefficient values c_k in equation 3.5 are estimated from the constructed phase space. In particular, if the phase space is constructed using a streamflow time series of length l, and the streamflow $x(l+T)$ is to be predicted, we can re-write equation 3.5 in matrix format with a slight modification as follows:

$$\boldsymbol{B}_{[n\times 1]} = \boldsymbol{A}_{[n\times(E+1)]} \times \boldsymbol{C}_{[(E+1)\times 1]} \tag{3.6}$$

Where:

$$\boldsymbol{B} = w(\|\boldsymbol{x}(t) - \boldsymbol{x}(l)\|) * x(t+T) \tag{3.7}$$

$$\boldsymbol{A} = w(\|\boldsymbol{x}(t) - \boldsymbol{x}(l)\|) * \widetilde{\boldsymbol{x}}(t) \tag{3.8}$$

$\boldsymbol{C}$ in equation 3.6 is the vector of coefficients c_k. Note that the number of rows in both $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted by n which is equal to the number of $\boldsymbol{x}(t)$ points that don't share coordinates with the point $\boldsymbol{x}(l)$. Equation 3.6 is modified from equation 3.5 by introducing a weighting factor w; this factor depends on the Euclidean distances in the phase space $\|\boldsymbol{x}(t) - \boldsymbol{x}(l)\|$, and it is calculated as follows:

$$w(d) = e^{-\theta d/\bar{d}} \tag{3.9}$$

Here, d is the distance $\|\boldsymbol{x}(t) - \boldsymbol{x}(l)\|$ and $\bar{d}$ is the average of distances for all values of t. The parameter θ takes values in the range $[0,\infty)$. If the value of θ is zero, $w(d) = 1$ regardless of the distance. In other words, all points in the phase space are assigned equal weights; this is equivalent to fitting a global linear map on all points $\boldsymbol{x}(t)$ in the phase space. On the contrary, if θ is large, points in the neighborhood of $\boldsymbol{x}(l)$ (i.e. smaller d) are assigned larger weights, and the weights decreases exponentially as a function of the distance. Put simply, as θ takes larger values, the function $\hat{f}_T$ becomes more local and nonlinear.

An indicator of the system nonlinearity can be estimated using S-map. Specifically, successive values of θ starting from zero are chosen, and prediction accuracy (ρ) for each value is calculated. The value that maximizes prediction which is denoted here by θ_{opt} is identified and considered to be a measure of the system nonlinearity. Values of θ_{opt} near zero indicates that the behavior of the system is linear whereas larger values of θ_{opt} means that more accurate prediction can be achieved by considering local neighborhood as opposed to the global phase space. This latter case clearly indicates the existence of state-dependent nonlinear behavior. It should be noted that S-map has previously been used to identify systems nonlinearity (Sugihara, 1994). Figure 3.1c shows prediction accuracy as a function of θ where the value of θ_{opt} is 3. The fundamental interpretation of θ_{opt} as a nonlinearity index is more precise than the general definition of nonlinear systems as systems in which output changes are not proportional to input changes. Nonlinearity in such a broad sense might stem from a variety of forms (e.g. polynomials of degrees higher than one, exponential …etc.); however, θ_{opt} is closely tied to state-dependent behavior. Essentially, it is a parameter of equation 3.4, and it is obtained by optimizing the function in equation 3.4 to yield best predictions. Therefore, a low value of θ_{opt} (e.g. $\theta_{opt} = 0$) means that a single function of $\hat{f}_T$ in equation 3.4 is valid to represent streamflow regardless of the basin dynamical state. On the contrary, a high value of θ_{opt} indicate that $\hat{f}_T$ is highly sensitive to the basin dynamical state.

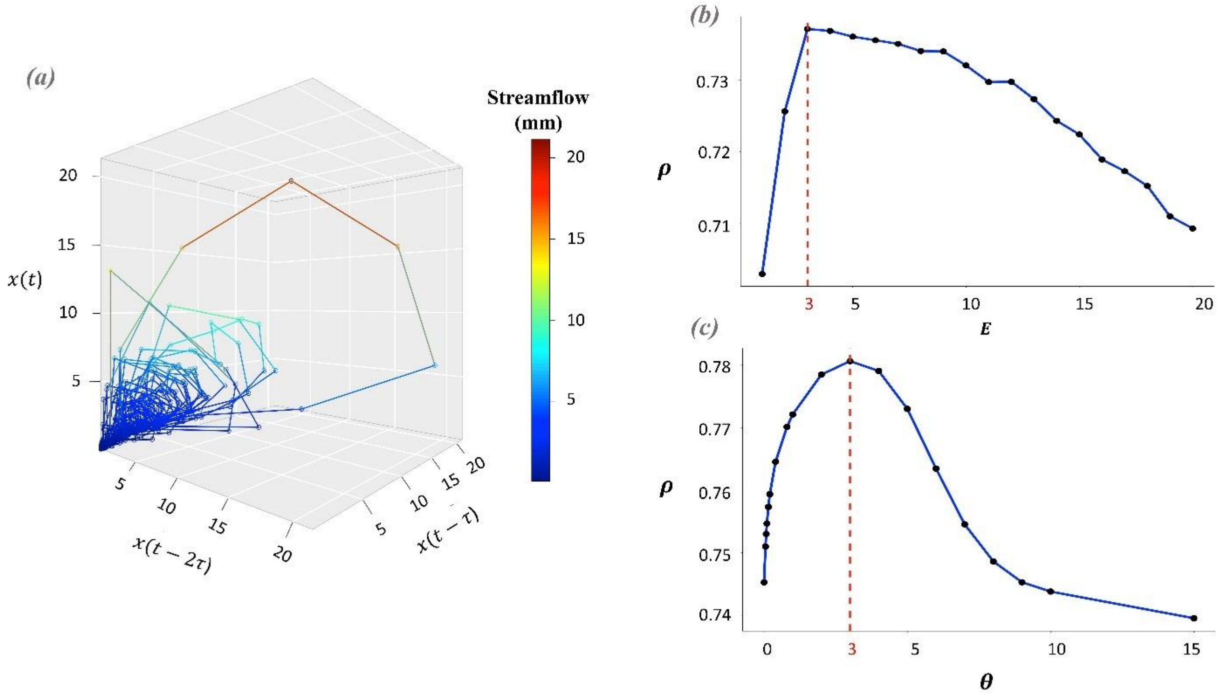


Figure 3.1(a) *The 3-dimensional phase space representation of streamflow time series from USGS gauging station (ID 01560000) located in Pennsylvania. Delay time* τ *is taken to be equal to the sampling time step of 1 day.* ***(b)*** *Prediction accuracy (correlation coefficient of observed and predicted streamflow) of the Simplex method for a range of E values; E=[1,2,3…20].* ***(c)*** *Prediction accuracy of the S-maps method for a range of* θ *values;* $\theta=[0,0.05,0.075,0.1,0.15,0.2,0.4,0.8,1,2,3,4,5,6,7,8,9,10,15]$*. In both (b) and (c), prediction is for 1 day ahead (T=1).*

c. Testing for statistical significance using surrogates

In order to obtain confidence on the values of complexity indices and to ensure that the process is indeed a form of chaotic determinism, we resort to the use of surrogates. The essential idea of testing for statistical significance using surrogates is to first define a null hypothesis on the underlying process that generated the observed time series. Surrogates are then generated in accordance with the null hypothesis, and a discriminating metric is computed from both the original time series and surrogates. Finally, the null hypothesis is rejected if the value of the metric is different from its distribution in the surrogate time series. A hierarchy of null hypotheses can be adopted for testing the significance of nonlinear dynamics; the simplest of which is to assume that the observed time series is the result of independent and identically distributed (i.i.d) random variable. However, under many circumstances, this hypothesis is insufficient, and a stricter null hypothesis is required. Several methods have been proposed in the literature to generate surrogate data consistent with different null hypotheses (e.g. Kugiumtzis, 2000; Schreiber & Schmitz, 1996; Theiler et al., 1992). Here, we use an algorithm that randomizes the phases of the Fourier transform of the original timeseries, and it preserves the mean, variance and autocorrelation (i.e. Fourier spectrum). This algorithm is described in Theiler et al. (1992); see also Osborne et al., 1986. Essentially, the original timeseries is transformed to the frequency domain, then the phases at all frequencies are randomized in a symmetrical manner for positive and negative frequencies. Finally, surrogate timeseries is obtained by transforming the phase randomized Fourier spectrum back to the time domain; see Theiler et al. (1992) for detailed description.

In the present study, we generate 100 surrogates for each timeseries of streamflow. Throughout our analysis, we compute any metric of interest for both the original timeseries and the 100 surrogates. There are three metrics that are considered in this study: (1) the accuracy of nonlinear prediction expressed as correlation coefficient ρ; (2) the optimum embedding dimension E_{opt}; and (3) the optimum nonlinearity index θ_{opt}. Let Q denote any of these three metrics computed for the original timeseries whereas μ_Q and σ_q are the mean and standard deviation of the metric computed from the 100 surrogates. The statistical significance is then computed as follows:

$$S = \frac{|Q - \mu_Q|}{\sigma_Q} \tag{3.10}$$

S indicates how distant the metric Q is from μ_Q in terms of units of standard deviation σ_Q. The distribution of Q in the surrogates is assumed to be gaussian, and a simple one-sided p-value can be computed to evaluate the statistical significance of any metric Q.

3.4 Results and Discussion

a. Presence of chaotic dynamics

Simplex and S-map methods were applied on each streamflow timeseries of the 413 MOPEX basins. Before we report the results on hydrologic complexity, an analysis is conducted to examine the presence of chaotic behavior in each streamflow timeseries. The analysis is performed by computing prediction accuracy (ρ) for both the original timeseries and surrogates. The statistical significance is then computed according to section 3.3; a basin is considered to exhibit chaotic dynamics if the prediction accuracy of the original timeseries is greater than that of the surrogate distribution with a p-value less than the statistical significance level α of 0.05. A statistically significant p-value means that the phase space representation provides useful information for prediction beyond what is contained in the autocorrelation. In other words, there exists an underlying dynamical attractor with coherent structure such that it can be exploited to achieve better prediction. On the contrary, if the basin behavior is not controlled by chaotic dynamics, the obtained prediction accuracy ρ of the original timeseries will be within the distribution of ρ in the surrogates. This latter case will result in a high p-value and the null hypothesis is accepted. It is worth mentioning that this analysis is not a conclusive test for the presence of chaotic dynamics which, as noted by Porporato and Ridolfi (1997), can only be achieved by applying a variety of techniques complementary to each other. It is nonetheless useful to pinpoint the existence of a weakest component of chaotic dynamics, and it allows us to account for autocorrelation effects in the timeseries.

Figure 3.2a shows a part of the 1-day ahead streamflow forecasts at USGS gauging station (ID 1631000) located in Virginia. We use 70% of the observations to construct the phase-space, and then validate the accuracy of nonlinear prediction in the remaining 30% of the record; Figure 3.2b shows the scatterplot of observed and forecasted streamflow with correlation coefficient ρ of 0.78. Figure 3.2c shows the histogram of forecast accuracy obtained for the 100 surrogates; their mean and standard deviation are 0.73 and 0.004 respectively. By assuming a gaussian distribution with these values for mean and standard deviation (shown in black in Figure 3.2c), a p-value can be computed which in this case is equal to 3×10^{-36}; therefore, the basin is considered to exhibit chaotic behavior.

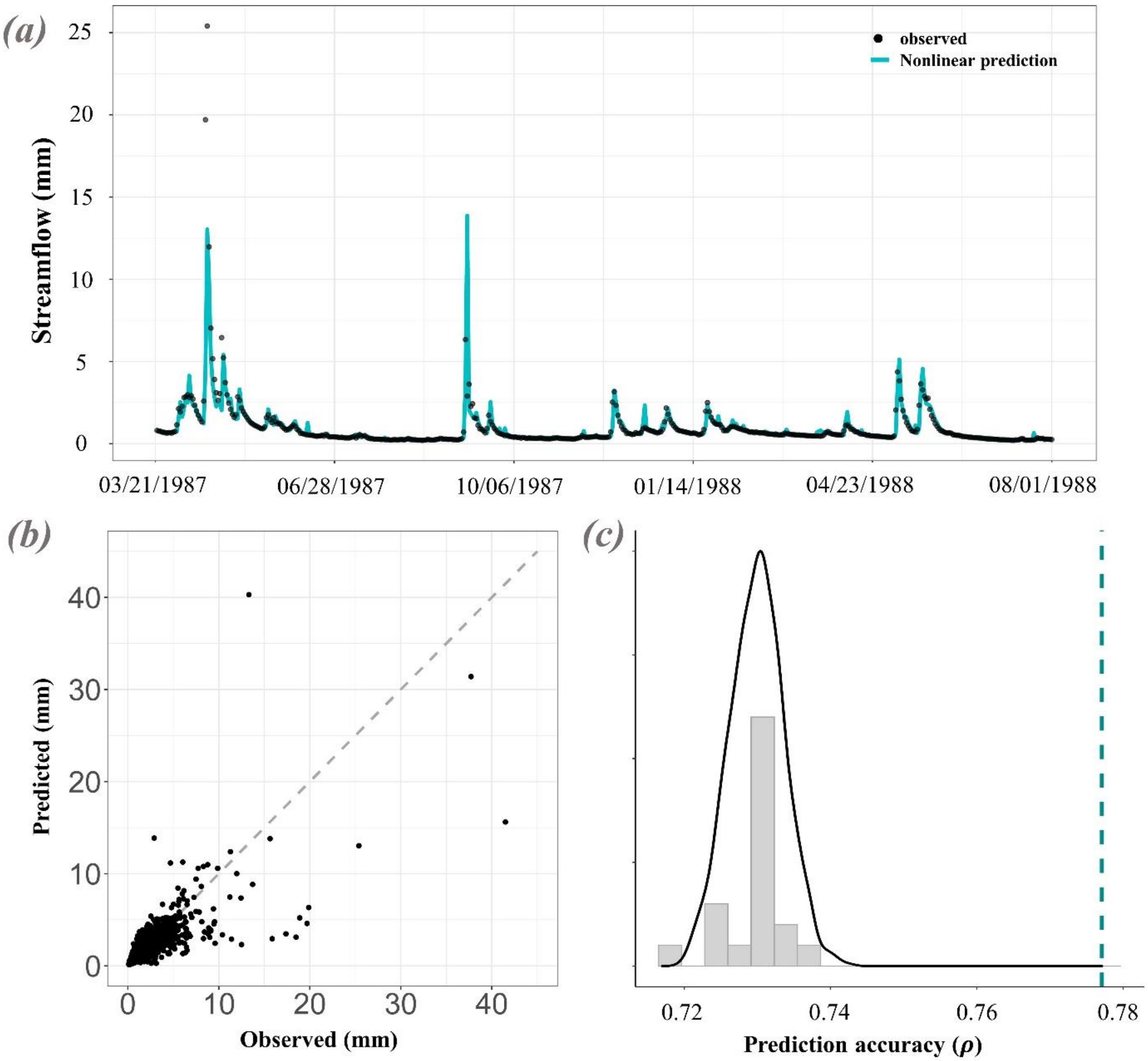


Figure 3.2 ***(a)*** *Daily observed streamflow (black points) and 1-day ahead forecasts (cyan line) for USGS gauging station (ID 1631000) located in Virginia for the period extending from late March 1987 to early August 1988.* ***(b)*** *scatterplot for observed and forecasted streamflow; correlation coefficient ρ=0.78; the optimum embedding dimension E_{opt}= 6.* ***(c)*** *The histogram of forecast accuracy obtained from the surrogate time series is shown in gray. A Gaussian distribution centered at the mean of histogram values ρ=0.73 with a standard deviation of 0.004 shown in black. The estimate of forecast accuracy for the original timeseries is shown as dashed cyan line ρ=0.78.*

b. Hydrologic complexity and its variability

Simplex and S-map were applied to the 268 basins that showed evidence of chaotic dynamics. In order to estimate E_{opt}, values of E in the range [1, 20] were iteratively selected to compute accuracy of prediction for a prediction horizon of 1 day ($T = 1$). As for the nonlinearity index θ_{opt}, θ was iteratively selected from the values {0, 0.05, 0.075, 0.1, 0.15, 0.2, 0.4, 0.8, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15}. In order to obtain statistical confidence on the values of E_{opt} and θ_{opt}, we computed their statistical significance according to the surrogate test. Because we seek to identify low-dimensional nonlinear chaotic behavior, any given value of E_{opt} is considered statistically significant if it is smaller than its surrogate distribution at a significance level of 0.05. Conversely, θ_{opt} is considered statically significant if its value is larger than its surrogate distribution. Figure 3.3 shows the joint and marginal distributions of E_{opt} and θ_{opt}; basins shown in dark blue are the ones where both complexity indices are statistically significant. Also, the histograms in Figure 3.3 show the marginal distribution of statistically significant E_{opt} and θ_{opt}. The following observations may be made: (1) There is a large variability in basin complexity with E_{opt} and θ_{opt} in the ranges 1 - 12 and 0 - 15 respectively. (2) A tradeoff between E_{opt} and θ_{opt} exists; consequently, very few basins exhibit both high number of active degrees of freedom and extreme nonlinearity. This is demonstrated by the negative slope of the joint density contour lines. (3) The dynamics in a large number of basins has a small number of active degrees of freedom. This can be seen from the positively skewed distribution of E_{opt} with a median ≈ 4 (black dashed line). (4) Similarly, the distribution of θ_{opt} is positively skewed with a long right tail and a median approximately equal to 4, suggesting that dynamics in most basins is moderately nonlinear, and very few basins exhibit highly nonlinear dynamics.

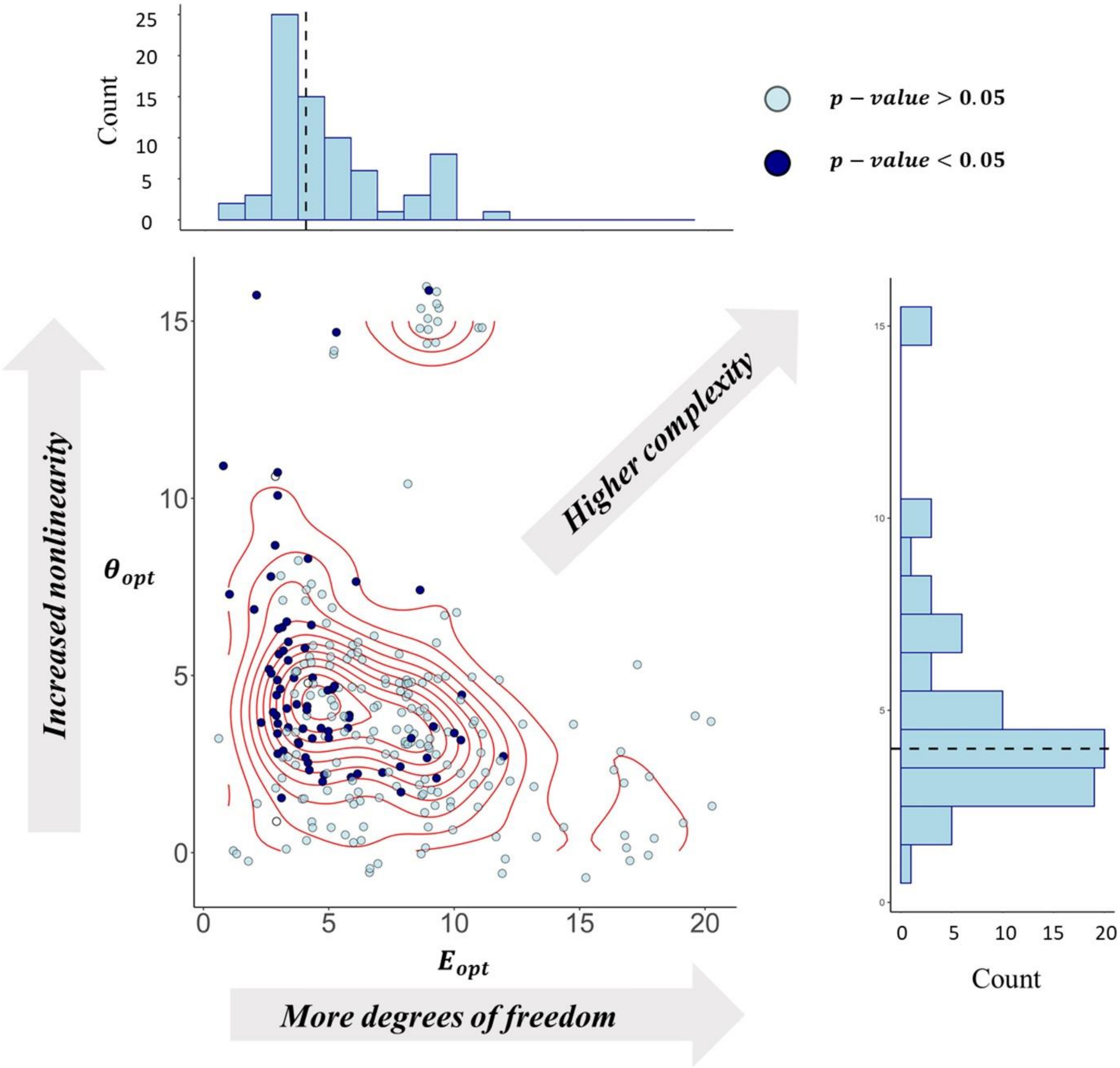


***Figure 3.3** Scatterplot of E_{opt} and θ_{opt} values for the 268 basins that were identified to exhibit chaotic dynamics; basins where both complexity indices are statistically significant at $\alpha = 0.05$ are shown in dark blue whereas other basins are shown in light blue. Red contour lines represent the joint bivariate distribution. Top panel shows the histogram of statistically significant E_{opt} values (black dashed line indicates the median). Right panel shows the histogram of statistically significant θ_{opt} values with black dashed line indicating the median. Basins with larger distance from the origin along the diagonal are more complex (high number of active degrees of freedom and increased nonlinearity).*

c. Impact of basin characteristics on hydrologic complexity

Here we examine the impact of basin characteristics on the value of the two complexity indices E_{opt} and θ_{opt}. The total number of basin characteristics is 15 which includes topographic, climatic and land surface properties. Table 3.1 shows the spearman correlation coefficient r_s between each of the 15 basin characteristics and the two complexity indices; only basins where both complexity indices are statistically significant were considered for this analysis. Clearly, E_{opt} is primarily related to topographic properties of the basin such as area, elevation, slope, stream length and drainage density (i.e. p-value of r_s < 0.05). The most significant relationship is with basin area (r_s = - 0.33) which is shown in Figure 3.4a; specifically, basins with larger areas tend to have lower dimensionality (i.e. low active degrees of freedom). This can be attributed to the fact that large basins are more effective in filtering out the random behavior induced by basin heterogeneities and rainfall variability. As a result, multiple degrees of freedom are subdued, and the dynamics reduce to very few dimensions. This is also in agreement with previous findings that streamflow forecasting in larger basins is more accurate than in small basins (Parajka et al., 2013; Zalenski et al., 2017) which might be attributed to the low number of active degrees of freedom in large basins. In particular, the smoothing effect of large basins has long been hypothesized as the reason of why hydrologic models tend to perform better in large basins (e.g. Nester et al., 2011); however, empirical observations that link the size of basin area to dynamics dimensionality remained to be elusive in most studies (e.g. Liu et al., 1998).

In addition to the impact of basin area on E_{opt}, there are also significant correlations with channel slope and stream length with r_s values of 0.32 and -0.29 respectively; see Figures 3.4b and 3.4c. The relationship regarding stream length can be viewed as a result of the strong covariance between stream length and basin area; particularly, larger basins will most likely be associated with longer stream length. Indeed, the Spearman correlation coefficient between basin area and stream length is 0.89. On the other hand, the positive correlation between E_{opt} and channel slope indicate that flashier basins have large dimensionality of dynamics. In other words, holding all other variables constant, an increase in channel slope means that the runoff response to rainfall is more direct, the basin is flashier, and the dimensionality of dynamics is large. If one is to relate high values of E_{opt} to low performance of models, then several studies have previously indicated that the difficulty of modelling catchments increases with flashiness (e.g. Lidén & Harlin, 2000; Yatheendradas et al., 2008) although the flashiness was attributed to climate properties and not to channel slope. It is noteworthy that Pande and Moayeri (2018) reported an opposite relationship that high slope is associated with lower complexity; however, they also highlighted that this relationship is only evident within a specific group of steep basins in their case study (i.e. it is not a general pattern).

Lastly, E_{opt} has significant correlations with elevation and drainage density D_d. These two relationships might be viewed as a result of covariation with basin area, stream length and channel slope. In particular, elevation and channel slope are correlated with r_s = 0.62; D_d on the other hand covaries with basin area as it is defined as the ratio of stream length to basin area. Moreover, the positive correlation between D_d and E_{opt} can be interpreted based on the fact that D_d is an indicator of the interaction between groundwater and surface water at the basin scale. More specifically, a higher value of D_d indicates less permeability of the surface and flashier response of the basin; thus, increasing the dynamic dimensionality. For instance, Le Moine (2008) shows that there is a significant correlation between D_d and high flows (i.e. flashier response). The upshot of the results regarding the impact of basin properties on E_{opt} is that low dimensional dynamics are present in large basins that filter out and smooth heterogeneities whereas basins with higher slope, smaller size and flashier behavior are governed by large degrees of freedom. It is worth mentioning that the relationship with basin size was observed in the work of Vignesh et al. (2015). Furthermore, other studies that reported the presence of low-dimensional chaotic dynamics in streamflow were mostly limited to large basins; for instance, the Aosta Valley basin with an area of 3313 km^2 (Porporato & Ridolfi, 1997). It is also important to note that there is a wide scatter in all relationships, and the values of E_{opt} are not fully explained by the size and slope of basins. This is, however, expected since the dynamics dimensionality arises from the interplay of several factors, and it is far more complex than to be explained by a single variable.

Table 3.1 *Relationship between 15 basin characteristics (topographic, climatic and land surface) and the two complexity indices E_{opt} and θ_{opt} expressed as spearman correlation coefficient (r_s); p-values of r_s are also shown. The values of r_s shown in bold font with an asterisk are statistically significant at 0.05 significance level. All values of r_s and p-value are rounded to two decimals.*

Basin Characteristic	E_{opt}		θ_{opt}	
	r_s	p-value	r_s	p-value
A	**-0.33 ***	0.00	0.02	0.89
Elevation	**0.17 ***	0.02	-0.04	0.77
Slope	**0.32 ***	0.00	-0.05	0.71
Stream length	**-0.29 ***	0.00	0.05	0.73
D_d	**0.24 ***	0.00	-0.04	0.78
DI	-0.01	0.93	-0.17	0.14
PSI	-0.02	0.79	-0.15	0.21
QSI	0.02	0.73	0.19	0.10
Snow fraction	0.04	0.48	-0.05	0.68
$Jan\ T_{min}$	-0.01	0.89	0.20	0.09
λ	0.02	0.73	-0.04	0.78
f_s	-0.05	0.57	0.08	0.65
Greenness	0.07	0.23	0.15	0.19
IGBP	-0.01	0.91	**-0.27 ***	0.02
Soil texture	0.07	0.28	-0.14	0.23

Unlike E_{opt} which is primarily modulated by topographic properties of the basin, presence of nonlinearity indicated by the value of θ_{opt} is only correlated at α = 0.05 with the type of vegetation cover (r_s = -0.27). Figure 3.4d shows the relationship between the nonlinearity index θ_{opt} and the dominating International Geosphere-Biosphere Programme (IGBP) land cover class. There are 17 IGBP land cover classes ranging from evergreen forests (classes 1 and 2) to built-up land and water bodies (classes 13 and 17). The IGBP land cover in the basins used for the analysis in this section are almost limited to classes 1 through 13. The gradient of IGBP classes from low (class 1) to high (class 13) represents the gradient in spatial and temporal extent of canopy cover from evergreen forests to built-up lands. Therefore, the negative correlation shown in Figure 3.4d indicates that basins with no or very limited canopy cover (i.e. classes 11 – 13) tend to behave more linearly whereas extensive vegetation is associated with more nonlinear dynamics.

It is plausible to interpret these results in terms of the complex impact of canopy cover on evapotranspiration, infiltration and interception storage. The nonlinearity indicated by the index θ_{opt} is particularly related to state-dependent behavior; essentially a high value of θ_{opt} indicates that the temporal evolution of streamflow is highly sensitive to the dynamical state of the basin. In light of this fundamental understanding, the relationship between θ_{opt} and vegetation cover can be explained through the sponge-effect hypothesis (Bruijnzeel et al., 2004). Generally, this hypothesis highlights the nonlinear state-dependent behavior of watersheds dominated by extensive vegetation cover (i.e. forests), and it postulates that the existence of forests reduces runoff peaks and annual runoff, but it enhances infiltration leading to increased streamflow during the dry season. There are several studies that examined this hypothesis using a paired catchment methodology, showed evidences of the sponge-effect hypothesis. The relationship with vegetation cover is also consistent with previous studies that linked higher basin complexity to vegetation with more storage capacity (Pande & Moayeri, 2018). Although that there is not an extensive literature on the explicit impact of vegetation cover on complexity of dynamics, it appears to be intuitive that more vegetation cover aggravates the nonlinearity of basins through its complex impact on evapotranspiration, infiltration and interception.

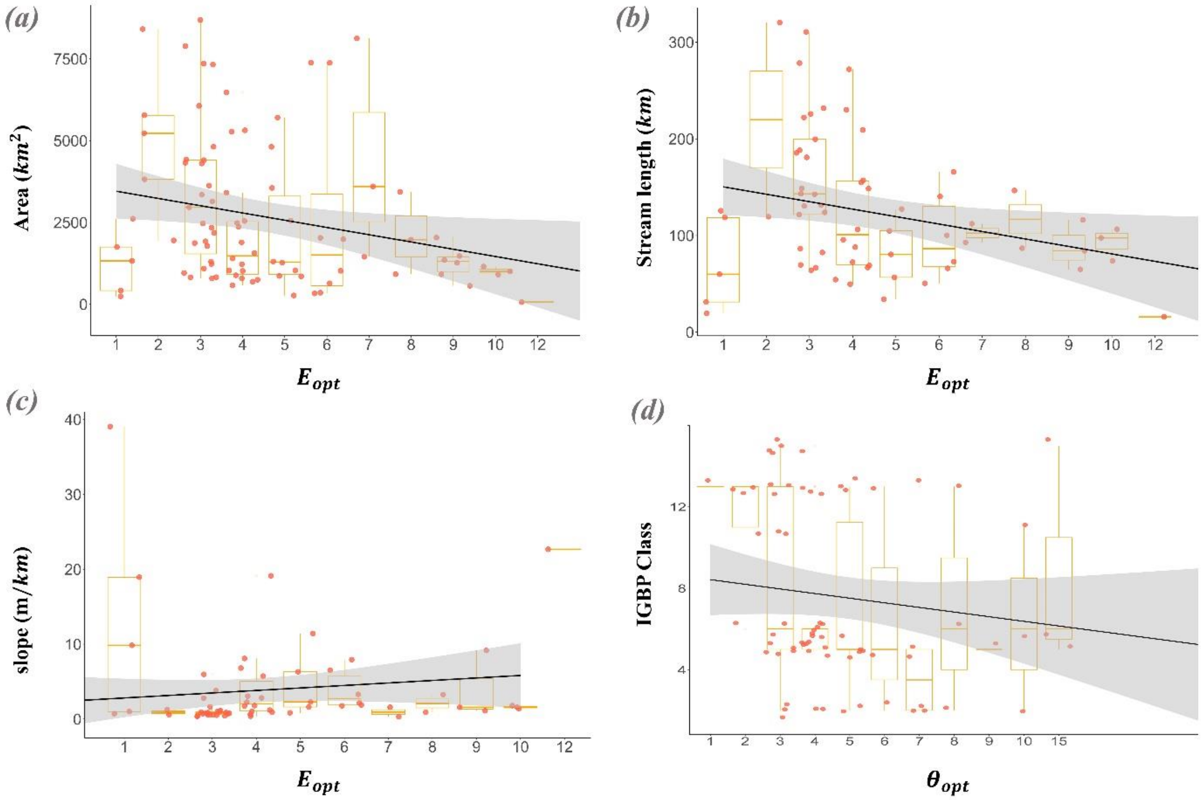


Figure 3.4 ***(a)*** *Boxplots for the values of basin area (km*^{2}*) corresponding to each* E_{opt} *group (1 to 12).* ***(b)*** *Boxplots for the values of stream length (km) corresponding to each* E_{opt} *group (1 to 12).* ***(c)*** *Boxplots for the values of main channel slope (m/km) for each* E_{opt} *group (1 to 12).* ***(d)*** *Boxplots for IGBP classes for each θ_opt group (1 to 15). In all plots, the black line is the least squares regression line with the shaded area showing the 95% confidence intervals.*

3.5 Conclusions

This study attempted to examine the qualitative character of watershed dynamics by applying techniques of chaotic dynamical systems, namely phase space reconstruction and nonlinear prediction. Their main advantage is that they are nonparametric and data-driven; thus, they are independent of structural assumptions in hydrologic models. While previous studies that used techniques of chaotic dynamical systems interpreted dynamic complexity in terms of a single measure indicating system dimensionality, we provided two measures of dynamic complexity. The first measure, E_{opt}, provides an estimation of system dimensionality (i.e. number of active degrees of freedom) whereas the second measure, θ_{opt}, is a proxy for the strength of dynamics nonlinearity. An analysis on 408 basins, from the MOPEX dataset, representing a wide range of climatic, topographic and land surface characteristics was conducted, and the main findings are as follows:

- Based on surrogate test to identify whether a component of deterministic chaos exists, it was found that 268 of the 408 basins exhibit chaotic dynamics. Further investigation revealed that snow fraction and rainfall seasonality (PSI) exert first order control on the presence of chaotic dynamics. Specifically, rainfed (i.e. low snow fraction) basins with lower PSI (i.e. more uniform rainfall) tend to exhibit chaotic behavior.

- Values of complexity measures for the 268 basins show a wide range of variability with E_{opt} and θ_{opt} in the range of (1 - 12) and (0 - 15) respectively. However, the distribution of both measures is positively skewed with 50% of basins having a value of E_{opt} and θ_{opt} below 4. This means that the dynamics in most basins is moderately nonlinear and have a low number of active degrees of freedom.

- The dynamic dimensionality (number of active degrees of freedom), as estimated by E_{opt}, is found to be primarily related to basin area. Specifically, larger basins tend to have lower dimensionality. This is attributed to the impact of large basins in filtering out heterogeneities within the basin. E_{opt} is also positively correlated with channel slope indicating that flashier basins have large dimensionality.

- Unlike dynamic dimensionality, the nonlinearity of dynamics is found to be regulated by the type of vegetation cover in the basin. Specifically, basins with year-round, extensive vegetation cover are associated with nonlinear dynamics.

- Preliminary results show that grouping of basins in terms of the value of E_{opt} and θ_{opt} provides insights on prediction accuracy of daily streamflow for prediction horizons of 1 day up to 2 weeks. In particular, it is shown that low-dimensional, nonlinear dynamics lead to higher prediction accuracy.

The results presented here have implications and potential relevance to several applications as it provides a basis for catchment classification based on similarity of dynamic complexity. This is particularly useful because the approach we used is data-driven, and it is not limited by our ability to represent and model hydrologic processes; this property is often preferred in catchment classification systems (Beven, 2011; Wagener et al., 2007). The findings that relate basin characteristics to the dynamic complexity measures presented in this study can be useful in obtaining a proxy for the complexity of system dynamics in ungauged basins. Consequently, this allows making informative decisions regarding extrapolation of hydrologic model parameters to ungauged basins and selection of reference streamflow gauges (Archfield & Vogel, 2010). Arguably, this is more advantageous compared to commonly used approaches of spatial proximity to extrapolate parameters to ungauged basins. Moreover, since the measures presented in this study estimate both dynamic dimensionality and nonlinearity separately, they potentially have implications in model selection. Specifically, to guide the selection of models with appropriate complexity, both in terms of strength of nonlinearity and number of active degrees of freedom, that is well suited to resolve watershed dynamics. Finally, we also note that the results presented here may have implications and guide the application of data-driven (e.g. deep learning) approaches to streamflow forecasting.

Chapter 4

4 Multivariate State-space Forecasting of Streamflow

4.1 Introduction

A major focus of surface hydrology revolves around streamflow forecasting in the short-range to mid-range with lead times of hours to seasons respectively. Over the years, an enormous body of research was placed on developing techniques for reliable streamflow forecasting. These techniques range from sophisticated, physically based models that attempt to explicitly characterize the heterogenous hydrologic processes within the watershed to purely data-driven, model-free techniques.

Data-driven techniques for streamflow forecasting attempt either to learn statistical patterns in time series or to model the dependence between streamflow and climatic teleconnections such as the El Nino Southern Oscillation (ENSO) (e.g. Piechota et al., 1998; Hidalgo-Muñoz et al., 2015). In both cases, a variety of empirical models have previously been used; these include regression analysis (e.g. Kothyari & Garde, 1991), Markov Chain (Yapo et al., 1993), independent component analysis (Westra et al., 2008), model trees (Solomatine & Dulal, 2003), artificial neural networks (ANNs) (e.g. Hsu et al., 1995; 2002; Minns & Hall, 1996; Thirumalaiah & Deo, 1998; Tokar & Johnson, 1999; Birikundavyi et al., 2002; Moradkhani et al., 2004) and phase space nonlinear prediction (e.g. Porporato & Ridolfi, 1997; Sivakumar et al., 2002). The forecasting method in the present study belongs to this last class of empirical models commonly referred to as phase space nonlinear prediction or empirical dynamical modelling (EDM).

Broadly speaking, EDM utilizes a multi-dimensional phase (state) space representation of historical observations to obtain forecasts; the coordinates of this phase space are represented by lagged time series of streamflow. This reconstruction of phase space is theoretically grounded based on Takens theorem (Takens, 1981; see also Packard et al., 1980). The central idea in making forecasts from a given system state using this phase space representation is to utilize our knowledge on the dynamical evolution of historical system states that are similar to the current one. Therefore, EDM, in spirit, is an "analogues" approach with a unique definition of similarity based on the closeness of system states in the phase space. Although less popular than ANNs, this approach received its fair share of attention in streamflow forecasting. For instance, it has been used by Porporato & Ridolfi (1997) to provide 12-hour and 24-hour lead time streamflow forecasts in the Aosta Valley river basin. Their results demonstrated the utility of EDM in providing reliable short-range streamflow forecasts with correlation coefficients > 0.95. Sivakumar et al. (2002) compared EDM and a multilayer perceptron artificial neural network for daily streamflow forecasting at 1-day and 7-day lead times in the Chao Phraya River basin in Thailand. While both approaches achieved reasonable forecast accuracy (correlation coefficient > 0.85), it was found that EDM outperforms the neural network model especially at long lead time (7-days). Additionally, a handful of other studies demonstrated the benefits attained by EDM for streamflow forecasting (Jayawardena & Lai, 1994; Jayawardena & Gurung, 2000; Sivakumar et al., 2001).

In spite of the encouraging results reported by previous studies in using EDM for streamflow forecasting, a major limitation of EDM is that it utilizes only a one-dimensional time series and leaves aside information on other variables that could potentially improve forecasting. While it seems that adding time series of variables such as precipitation, evaporation and temperature to the phase space is an intuitive and straightforward extension to conventional EDM, such an approach has not been adopted in streamflow forecasting. One reason behind this is the lack of theoretical principles for multivariate embedding, unlike univariate embedding which is rooted in Takens theorem. Another reason is the difficulty that lies in the selection of the optimum embedding coordinates; for instance, to construct a 5-dimensional multivariate embedding from a set of four variables and a maximum lag time of 10, there are 575,757 distinct possible embeddings. However, the significant advancements in the field of dynamical systems in recent years provide a remedy for both issues. Regarding the former, Takens theorem was recently extended to general multivariate embeddings by Deyle and Sugihara (2011). As for the latter, we present in this study an efficient algorithm that is guaranteed to converge to the optimum embedding coordinates. This algorithm builds on recent developments in causal inference; specifically, the method of convergent cross mapping (Sugihara et al., 2012).

To this end, the objective of this study is to present an efficient algorithm that extends univariate EDM to multivariate EDM, hereafter referred to as m-EDM, and examines its utility for daily streamflow forecasting. Furthermore, we compare the performance of m-EDM to state-of-the-art recurrent neural networks; specifically, Long short-term memory (LSTM). LSTMs (Hochreiter & Schmidhuber,1997) are particularly adept at learning statistical patterns in time series that have long lag time dependencies; thus, they are considered to be state-of-the-art ANNs in modelling sequences (e.g. speeches, handwriting and time series), and they have recently been applied in hydrology for various tasks (e.g. Fang et al., 2017; Kratzert et al., 2018). LSTMs should, therefore, be a reasonable model against which to compare the utility of m-EDM. m-EDM allows us to leverage information from other variables that could potentially improve forecasting. In the present study, time series of precipitation ($\boldsymbol{P}$), potential evapotranspiration ($\boldsymbol{PET}$), daily maximum temperature ($\boldsymbol{T_{max}}$) and streamflow ($\boldsymbol{Q}$) are used for m-EDM. The experiment is conducted on a set of 9 hydrologic basins from the Model Parameter Estimation Project (MOPEX) dataset. This set of basins is selected such that it encompasses basins with wide variability in their sizes and climatic properties.

4.2Data

a. General Characteristics

A set of nine catchments in the contiguous United State (CONUS) were selected for this study; see Figure 4.1. These catchments are part of the Model Parameter Estimation Project (MOPEX) dataset (Schaake et al., 2006; Duan et al., 2006) which was developed for the purpose of providing sufficient, high-quality observations for parameterization of hydrological models. The 9 catchments were selected on two bases. First, representing variability in catchment size; thus, catchments were selected from three categories: small (size < 200 km^2), medium (500 km^2 < size < 1000 km^2) and large (size > 4000 km^2). The thresholds selected for grouping catchments into these three categories are rather subjective and depends on the range of catchment size available in the MOPEX dataset; for instance, the smallest catchment size in MOPEX dataset is 67 km^2. Second, for each of the three categories, three catchments are selected to represent the spectrum of climatic properties. Precisely, we use the Dryness Index (*DI*), defined as the ratio of mean annual potential evapotranspiration to mean annual precipitation, as a basis for selection. Accordingly, for each category of catchment size, three basins are selected that are: relatively humid ($DI \leq 0.55$), temperate (0.6 < DI < 0.75) and arid (DI > 1.4). Figure 4.1 shows the location of the USGS gauging stations at the 9 selected catchments. Basic attributes of the catchments are shown in Table 4.1. Apart from representing variability in catchment size and DI, the catchments span a wide range of elevation (283 – 1195 m).

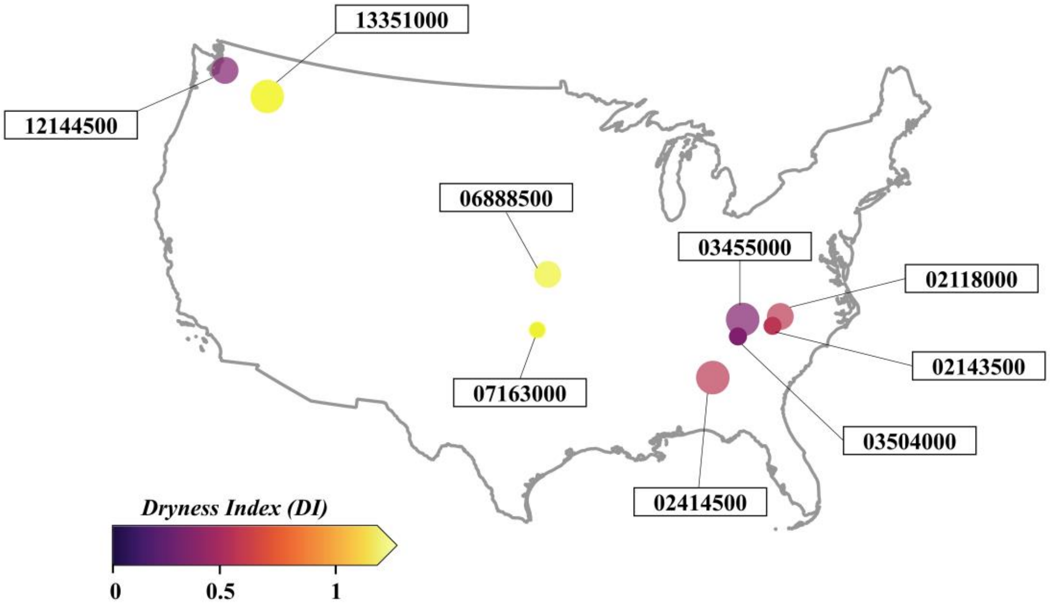


Figure 4.1 *Location of USGS gauging stations at the 9 catchments selected for this study. The size of symbols is proportional to catchment size with their values shown in Table 4.1. Symbols color indicates the value of the Dryness Index (DI) in the basin. Labels show the USGS gauging station IDs.*

Table 4.1 *The 9 MOPEX catchments selected for this study and their attributes.*

USGS gauge ID	***Site name***	***Area*** (km^2)	***Dryness Index (DI)***	***Elevation (m a.s.l)***	***Annual Precipitation (mm)***	***Record length (years)***
07163000	Council Creek near Stillwater, OK	80	1.50	-	892	46
03504000	Nantahala River near Rainbow Springs, NC	135	0.39	1195	1960	55
02143500	Indian Creek near Laboratory, NC	179	0.75	290	1219	52
02118000	South Yadkin River near Mocksville, NC	793	0.73	317	1195	55
06888500	Mill Creek near Paxico, KS	818	1.41	399	875	50
12144500	Snoqualmie River Near Snoqualmie, WA	971	0.26	1006	2513	45
02414500	Tallapoosa River at Wadley, AL	4338	0.71	283	1376	54
03455000	French Broad River near Newport, TN	4812	0.55	-	1404	55
13351000	Palouse River at Hooper, WA	6475	1.96	735	525	52

b. Sample Split for Calibration and Validation

Table 4.1 shows that the time series length at the 9 catchments is equal to, or longer than, 45 years. For each catchment, we utilize 70% of the record length for calibration, and the remaining 30% for validation. These correspond to a minimum of 31.5 and 13.5 years for calibration and validation respectively. Therefore, they are expected to encompass streamflow variability at interannual and multi-year time scales. In the case of m-EDM, the calibration sample (70% of the total record) is further divided into two parts. The first (60% of the total record) is used to build the forecasting model whereas the second (10% of the total record) is used to select an optimum model among different competing models.

4.3Methods

a. Problem Statement

Our goal in this study is to use time series of three predictor variables: precipitation ($\boldsymbol{P}$), potential evapotranspiration ($\boldsymbol{PET}$) and daily maximum temperature ($\boldsymbol{T_{max}}$) as well as streamflow time series ($\boldsymbol{Q}$) to predict streamflow T time steps ahead. Mathematically, this is formulated as follows:

$$Q_{t+T} = \boldsymbol{g}[\, Q_t, Q_{t-1} \ldots Q_{t-\tau^*}, P_t \ \ldots P_{t-\tau^*}, PET_t \ \ldots PET_{t-\tau^*}, Tmax_t \ldots Tmax_{t-\tau^*}] \tag{4.1}$$

Where T is the prediction horizon (i.e. how far ahead) and τ^* is the maximum lag time to be considered as an input. Note that we use the notation of uppercase letters (not bold) to refer to single values; for instance, Q_{t-1} denotes a single streamflow observation at time $(t-1)$. The function $\boldsymbol{g}$ maps the inputs available at time t to streamflow at time $(t+T)$. The approach presented in this study of m-EDM as well as the LSTM are approximations of the function $\boldsymbol{g}$. The two following subsections discusses in detail the m-EDM approach and the design of LSTM.

b. Multivariate Empirical Dynamical Modelling (m-EDM) Framework

Identifying the optimum embedding dimension

The first step in any empirical dynamical modelling method is to identify the optimum embedding dimension (E_{opt}); that is, the number of phase space coordinates. Commonly, E_{opt} is defined by iteratively selecting an E value from a set of values $\{1, 2, \ldots\ E_{max}\}$, and computing the forecast accuracy of each embedding. E_{opt} is then defined as the embedding dimension that yields the best forecast accuracy. In this study, we use univariate embedding of streamflow timeseries (i.e. lagged time series of streamflow) and a set of E values in the range {1, 2, … 20} to define the optimum embedding dimension. Figure 4.2a shows a 3-dimensional embedding of streamflow at catchment (ID 02118000); this is constructed from streamflow time series $\boldsymbol{Q}_t$ and its values lagged by time steps of 1 and 2 (i.e. $\boldsymbol{Q}_{t-1}$ and $\boldsymbol{Q}_{t-2}$) using the first section of calibration data (60% of the total record). Note that uppercase bold letters refer to entire time series, not a single value. For each value of E in the range {1, 2, … 20}, the phase space of streamflow is reconstructed. Next, forecasts are made using the phase space for the second section of calibration data (10% of the total record), and their accuracy (correlation coefficient) is computed. Figure 4.2b shows the forecasts accuracy as a function of E in the range {1, 2, … 20}; the accuracy increases significantly as the value of E increases from 1 to 4, and it decreases beyond a value of E = 4. Thus, the optimum embedding dimension (E_{opt}) for this catchment is equal to 4. The forecasts from the phase space are carried out using the Simplex method which is analogous to the K-nearest neighbor approach. For detailed information on the Simplex method for phase space forecasting, interested readers should refer to Sugihara and May (1990). This first step in m-EDM framework is equivalent to the conventional univariate embedding that was previously used in several studies for streamflow forecasting.

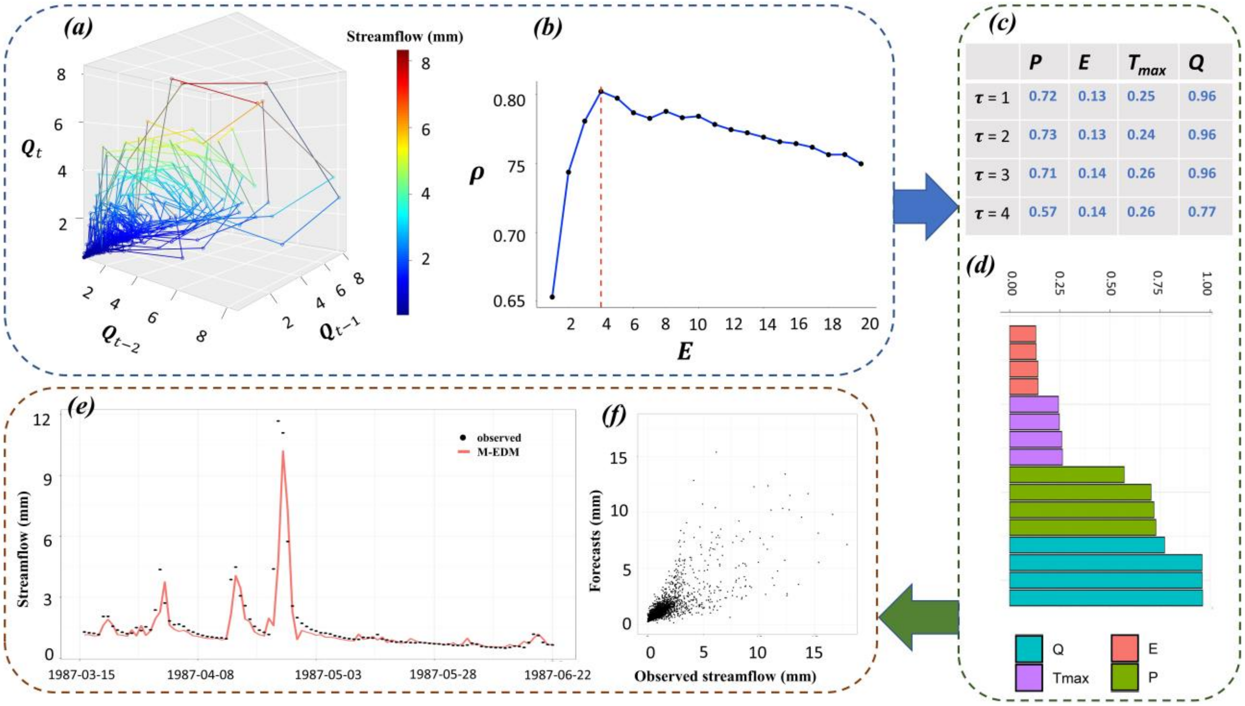


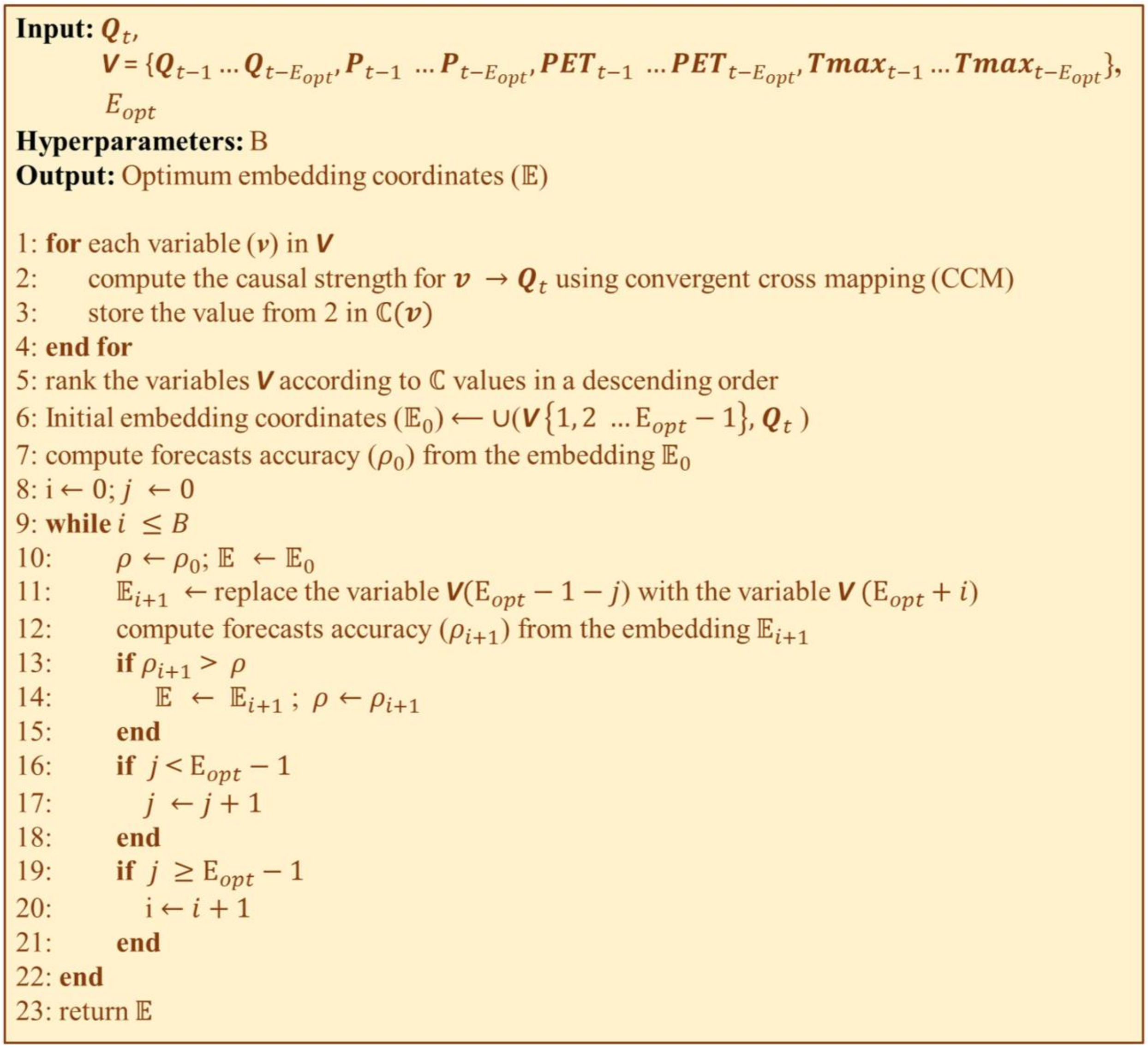
Figure 4.2 *(a) A 3-dimensional univariate embedding of the streamflow time series at catchment (ID 02118000). The coordinates of the embedding are the streamflow timeseries $\boldsymbol{Q}_t$ and its values lagged by time steps of 1 and 2 (i.e. $\boldsymbol{Q}_{t-1}$ and $\boldsymbol{Q}_{t-2}$). (b) The forecast accuracy, indicated by correlation coefficient, as a function of embedding dimension (E) for the same time series in (a). (c) The causal strength of each predictor variable with respect to the target variable $\boldsymbol{Q}_t$; these values are calculated using the method of convergent cross mapping. (d) The predictor variables ranked based on their causal strength values in (c). (e) A segment of observed (black dots) and predicted (red line) streamflow. Forecasts are made using a multivariate embedding that resulted from the algorithm. (f) Scatter plot of observed and forecasted streamflow values; correlation coefficient is 0.78.*

Identifying the embedding coordinates

After identifying the optimum embedding dimension, the embedding coordinates must be selected from a large number of possible combinations. Here, we assume that the maximum time lag to be considered for coordinates is equivalent to E_{opt} in order to minimize the number of possible combinations. Despite of this assumption, the number of possible combinations is enormous. For the time series in Figures 4.2a and 4.2b (E_{opt} = 4), the number of possible combinations is $\binom{(4*4)-1}{(4-1)} = \binom{15}{3} = \frac{15!}{3!\,(15-3)!} = 455$. It is very expensive, computational-wise, and impractical to iterate over all possible embeddings. This becomes almost impossible as the value of E_{opt} increases because the number of possible combinations grows super-exponentially as a function of E_{opt}.

Here, we circumvent the aforementioned issue by resorting to recent advances in causal inference. Specifically, we use the method of convergent cross mapping (CCM; Sugihara et al., 2012) to compute the causal strength of each variable in the set $\{\boldsymbol{Q}_{t-1} \dots \boldsymbol{Q}_{t-E_{opt}}, \boldsymbol{P}_{t-1} \dots \boldsymbol{P}_{t-E_{opt}}, \boldsymbol{PET}_{t-1} \dots \boldsymbol{PET}_{t-E_{opt}}, \boldsymbol{Tmax}_{t-1} \dots \boldsymbol{Tmax}_{t-E_{opt}}\}$ with respect to the target variable $\boldsymbol{Q}_t$. CCM is an efficient method that utilizes concepts of phase space reconstruction to establish causal relations; See Sugihara et al. (2012) for detailed information on CCM and Ombadi et al. (2020) for information on the use of CCM for reconstructing causal relations in hydrometeorological systems. Figure 4.2c shows the causal strength of each lagged variable for the time series in Figures 4.2a and 4.2b. These values can be interpreted as causal relations of the predictor variables: $\boldsymbol{P}$, $\boldsymbol{PET}$ and $\boldsymbol{T_{max}}$. On the contrary, the values for $\boldsymbol{Q}_{t-1} \dots \boldsymbol{Q}_{t-E_{opt}}$ do not represent causal relations per se; however, they are the result of autocorrelation in the time series of streamflow. Next, these variables are ranked according to their causal strength (Figure 4.2d). Our algorithm, shown in Figure 4.3, starts by taking the highest (E_{opt} – 1) variables as an initial guess of embedding coordinates. It only considers (E_{opt} – 1) coordinates because one embedding coordinate is already defined which is $\boldsymbol{Q}_t$. This initial guess of embedding coordinates will yield the best forecasts if the variables are independent of each other. That is, each variable provides unique information for the prediction of variable $\boldsymbol{Q}_t$; however, this seldom occurs. In practice, the initial guess of embedding coordinates will most likely result in lagged time series of streamflow as it is the case in Figure 4.2d. Consequently, the embedding coordinates are not independent of each other, and their information is redundant.

In order to converge to the optimum embedding coordinates, our algorithm employs a heuristic that is asymptotically guaranteed to reach convergence; Figure 4.3 shows a pseudocode of the algorithm. The algorithm starts from the initial guess of embedding coordinates ($\mathbb{E}_0$) to compute forecasts. Next, at each iteration, the algorithm replaces the variable with lower causal strength with one of the variables that were not originally considered in the initial embedding. In all iterations, the replacement is guided using the values of causal strength. That is, the variable with the lower causal strength is removed, and the one with the highest causal strength from the set of variables that were originally not included in the initial embedding is added. At each iteration, the algorithm computes the forecast accuracy, and it updates the embedding coordinates if a forecast accuracy larger than the current one is achieved. This procedure continues for a maximum number of iterations (B); B is a hyperparameter that is specified for the algorithm and is set to be 100 in this study. Clearly, if B is set to be equal to the possible number of embedding combinations, then the algorithm will yield the optimum embedding coordinate. Furthermore, because the algorithm only updates the embedding when a better forecasts accuracy is attained, it is not possible for the accuracy to decrease with iterations. Therefore, asymptotic convergence of the algorithm is guaranteed.



```
Input: $\boldsymbol{Q}_t$,
    $\boldsymbol{V} = \{\boldsymbol{Q}_{t-1} \ldots \boldsymbol{Q}_{t-E_{opt}}, \boldsymbol{P}_{t-1} \ldots \boldsymbol{P}_{t-E_{opt}}, \boldsymbol{PET}_{t-1} \ldots \boldsymbol{PET}_{t-E_{opt}}, \boldsymbol{Tmax}_{t-1} \ldots \boldsymbol{Tmax}_{t-E_{opt}}\}$,
    $E_{opt}$
Hyperparameters: B
Output: Optimum embedding coordinates ($\mathbb{E}$)

1: for each variable ($\boldsymbol{v}$) in $\boldsymbol{V}$
2:     compute the causal strength for $\boldsymbol{v} \rightarrow \boldsymbol{Q}_t$ using convergent cross mapping (CCM)
3:     store the value from 2 in $\mathbb{C}(\boldsymbol{v})$
4: end for
5: rank the variables $\boldsymbol{V}$ according to $\mathbb{C}$ values in a descending order
6: Initial embedding coordinates ($\mathbb{E}_0$) $\longleftarrow \cup(\boldsymbol{V}\{1, 2 \ldots \mathrm{E}_{opt} - 1\}, \boldsymbol{Q}_t)$
7: compute forecasts accuracy ($\rho_0$) from the embedding $\mathbb{E}_0$
8: $\mathrm{i} \leftarrow 0; j \leftarrow 0$
9: while $i \leq B$
10:     $\rho \leftarrow \rho_0$; $\mathbb{E} \leftarrow \mathbb{E}_0$
11:     $\mathbb{E}_{i+1}$ $\leftarrow$ replace the variable $\boldsymbol{V}(\mathrm{E}_{opt} - 1 - j)$ with the variable $\boldsymbol{V}(\mathrm{E}_{opt} + i)$
12:     compute forecasts accuracy ($\rho_{i+1}$) from the embedding $\mathbb{E}_{i+1}$
13:     if $\rho_{i+1} > \rho$
14:         $\mathbb{E} \leftarrow \mathbb{E}_{i+1}$ ; $\rho \leftarrow \rho_{i+1}$
15:     end
16:     if $j < \mathrm{E}_{opt} - 1$
17:         $j \leftarrow j + 1$
18:     end
19:     if $j \geq \mathrm{E}_{opt} - 1$
20:         $\mathrm{i} \leftarrow i + 1$
21:     end
22: end
23: return $\mathbb{E}$
```

Figure 4.3 *Pseudocode shows the algorithm used in this study to select an embedding coordinate given an input vector of variables and their lagged time series by time lag (τ) of 1, 2 …*E_{opt}*. Where* E_{opt} *is the optimum embedding dimension. The algorithm has one hyperparameter, B, which sets the maximum number of iterations for the algorithm.*

The Forecast Model

The final step in the m-EDM approach is to use the embedding coordinates obtained from the algorithm in Figure 4.3 to compute streamflow forecasts. The previous sections use only the first part of the calibration data (60% of the total record) to build the model, and the remaining calibration data (10% of the total record) to select between competing models. However, in this final step, the entire calibration data (70% of the total record) is used to build the model, and the validation data (30% of the total record) is used to compute forecasts accuracy. Forecasts are made from the phase space using the method of sequential locally weighted global linear maps "S-map" (Sugihara, 1994). As its name states, this method approximates the phase space by fitting maps in the phase space that are weighted locally. This is achieved mathematically by solving a singular value decomposition (SVD) problem; see Sugihara (1994) for detailed information on the S-map method. The forecasts of the time series shown in Figures 4.2a and 4.2b and their accuracy made using S-map are shown in Figures 4.2e and 4.2f.

c. Long short-term memory (LSTM)

LSTM networks are a specific type of recurrent neural networks that was developed by Hochreiter and Schmidhuber (1997), and its unique architectural design enables it to model temporal sequences that have long lag time dependencies. They have been applied successfully in several hydrometeorological applications including short-term precipitation forecast (Akbari Asanjan et al., 2018), prediction of satellite-based soil moisture (Fang et al., 2017) and rainfall-runoff modelling (Kratzert et al., 2018). Several variants of LSTMs exist; for instance, the variants introduced in Gers et al. (1999), Gers and Schmidhuber (2000) and Cho et al. (2014). The LSTM used in this study is shown in Figure 4.4, and it is described in equations 4.1 to 4.6.

The framework we use here consists of two LSTM layers followed by a dense output layer. Figure 4.4a shows the architectural design of the network; specifically, it shows the two LSTMs unrolled through time. The first LSTM layer takes an observations matrix of four variables ($\boldsymbol{P}$,

$\boldsymbol{PET}$, $\boldsymbol{T_{max}}$ and $\boldsymbol{Q}$) as input; the number of rows in the matrix is equivalent to the batch size. Meanwhile, the second LSTM layer takes the output of the first layer as an input. Finally, the last output from the second LSTM layer is mapped using a dense layer to streamflow vector $\boldsymbol{Y}$. The internals of the LSTM shown in Figure 4.4c consists of three gates: input gate ($\boldsymbol{i_t}$), forget gate ($\boldsymbol{f_t}$) and output gate ($\boldsymbol{o_t}$). Each of these gates takes an input ($\boldsymbol{x_t}$) and a hidden state from the previous time step ($\boldsymbol{h_{t-1}}$) according to the following equations:

$$\boldsymbol{i_t} = \sigma(\boldsymbol{W_i}\, \boldsymbol{x_t} + \boldsymbol{U_i}\, \boldsymbol{h_{t-1}} + \boldsymbol{b_i}) \tag{4.2}$$

$$\boldsymbol{f_t} = \sigma(\boldsymbol{W_f}\, \boldsymbol{x_t} + \boldsymbol{U_f}\, \boldsymbol{h_{t-1}} + \boldsymbol{b_f}) \tag{4.3}$$

$$\boldsymbol{o_t} = \sigma(\boldsymbol{W_o}\, \boldsymbol{x_t} + \boldsymbol{U_o}\, \boldsymbol{h_{t-1}} + \boldsymbol{b_o}) \tag{4.4}$$

Where $\boldsymbol{W}$ and $\boldsymbol{U}$ are the input and hidden weight matrices respectively, and $\boldsymbol{b}$ is the bias vector for each gate. Whereas $\sigma(.)$ is the sigmoid function. Next, the values of the input and forget gates are used to update the memory cell ($\boldsymbol{c_t}$) of the LSTM according to the following equation:

$$\boldsymbol{c_t} = \boldsymbol{f_t} \odot \boldsymbol{c_{t-1}} + \boldsymbol{i_t} \odot\, tanh(\boldsymbol{W_c}\, \boldsymbol{x_t} + \boldsymbol{U_c}\, \boldsymbol{h_{t-1}} + \boldsymbol{b_c}) \tag{4.5}$$

Where $\boldsymbol{W}$, $\boldsymbol{U}$ and $\boldsymbol{b}$ are as defined before, $\odot$ is the inner product, and $tanh(.)$ is the hyperbolic tangent function. Finally, the memory cell value is combined with the output gate to compute a new hidden state as follows:

$$\boldsymbol{h_t} = \tanh(\boldsymbol{c_t}) \odot \boldsymbol{o_t} \tag{4.6}$$

The procedure in equations (4.2) to (4.6) is carried out at the first and second LSTM layers. Then, the final hidden layer of the second LSTM is used as an input to a dense neural network layer with linear activation unit, see Figure 4.4c, as follows:

$$\boldsymbol{Y} = \boldsymbol{W_d}\, \boldsymbol{h_{t-\tau^*}} + \boldsymbol{b_d} \tag{4.7}$$

Where $\boldsymbol{W}$ and $\boldsymbol{b}$ are the weight matrix and bias vector respectively. $\boldsymbol{Y}$ is the vector of streamflow forecasts that correspond to the input batch.

The remaining hyper-parameters of the LSTM are defined as follows: the number of units = 32, batch size = 730 (i.e. 2 years), maximum lag time (τ^*) = 30, and number of epochs = 20. These values were the result of testing different values of each parameter in forecasting streamflow at the catchment (USGS ID 02118000). In training, the batch is drawn at random without replacement from the calibration data; and each epoch covers the entire calibration data. The objective function used to train the LSTM network is the mean squared error.

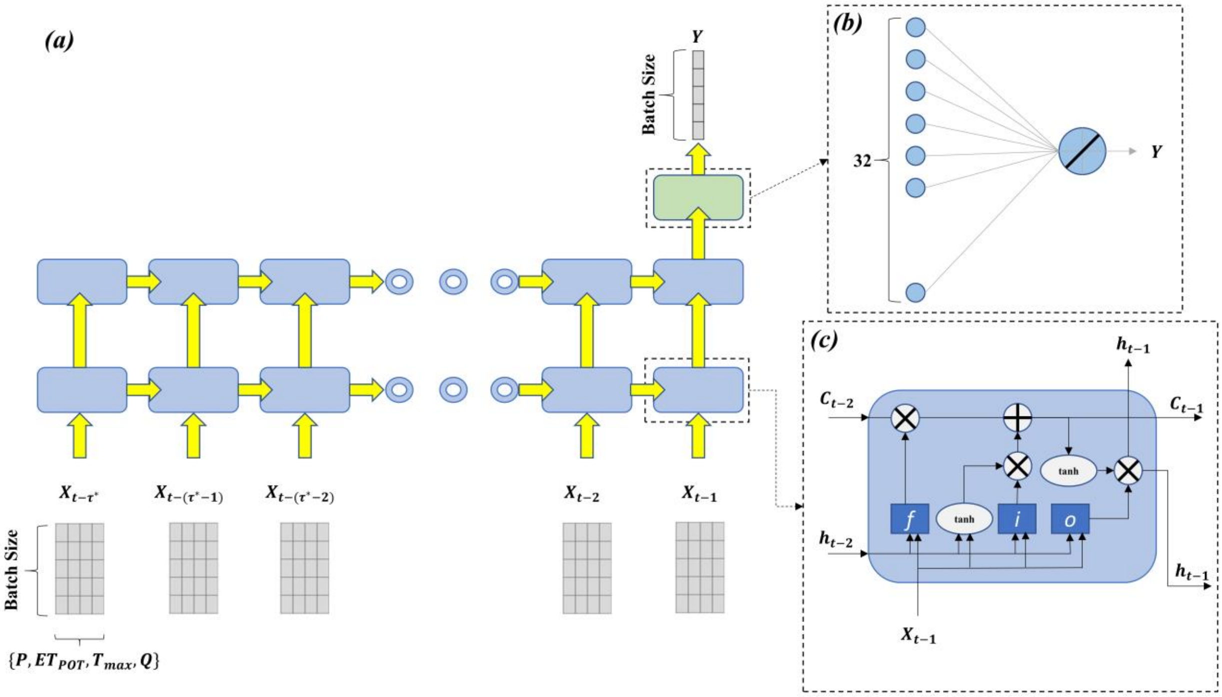


Figure 4.4 *(a) Architecture of the LSTM framework used in this study which consists of two LSTM layers (blue rectangles connected with yellow arrows), and a dense feedforward neural network layer (green rectangle). The two LSTMs are unrolled in time for convenient representation. The input matrices for each time step are shown in gray. (b) A dense layer with 32 input nodes and a linear activation unit to transform the last output of the second LSTM to streamflow. (c) The internal structure of the LSTM consists of three gates: input gate ($\boldsymbol{i}$), forget gate ($\boldsymbol{f}$) and output gate ($\boldsymbol{o}$) as well as memory cell ($\boldsymbol{c}$) and hidden state ($\boldsymbol{h}$). See equations and main text for detailed description.*

4.4 Results and Discussion

a. Overall Performance

In the following subsections, we present the results of streamflow forecasts obtained by the m-EDM approach and compare them to LSTM forecasts, for prediction horizon of 1 day ($T = 1$ day). In order to evaluate streamflow forecasts, we use three metrics: correlation coefficient (corr), Nash-Sutcliffe efficiency (NSE) and root mean squared error (RMSE). Denoting the streamflow forecasts at time *t* by Y_t, these metrics are calculated as follows

$$\rho(Q,Y) = \frac{mean[\,(Q_t - \bar{Q})\,(Y_t - \bar{Y})]}{\sigma_Q\,\sigma_Y} \tag{4.8}$$

$$RMSE(Q,Y) = \sqrt{\frac{\sum_{t=1}^{n}(Q_t - Y_t)^2}{n}} \tag{4.9}$$

$$NSE\,(Q,Y) = 1 - \frac{\sum_{t=1}^{n}(Q_t - Y_t)^2}{\sum_{t=1}^{n}(Q_t - \bar{Q}_t)^2} \tag{4.10}$$

Where $\bar{Q}$ and $\bar{Y}$ are the mean values of observations and forecasts respectively. Similarly, σ_Q and σ_Y are the standard deviations of observations and forecasts respectively. n is the length of time series.

Figure 4.5 shows the 1-day ahead streamflow forecasts obtained from the m-EDM and LSTM methods for catchment (ID 03504000); see Table 4.1. Figure 4.5a shows a segment of the streamflow time series from mid-April to late July of the hydrological year 1987. Overall, the two methods appear to capture the observed streamflow patterns adequately; in particular, the recession limb from mid-April to early May. However, discrepancies between forecasts and observations are apparent in the fluctuations of streamflow in the period of late June to early July (i.e. the sudden spikes in streamflow). Figures 4.5b and 4.5c show scatterplots of observed streamflow against forecasts obtained from m-EDM and LSTM for the entire calibration period. Clearly, the scatter of points away from the 45° degrees line is more significant at high values of streamflow. This is specifically apparent in the LSTM forecasts (Figure 4.5c). Overall, ρ of forecasted streamflow from m-EDM is 0.91 compared to 0.85 of LSTM. Table 4.2 shows the values of ρ as well as *NSE* and *RMSE* at the 9 catchments.

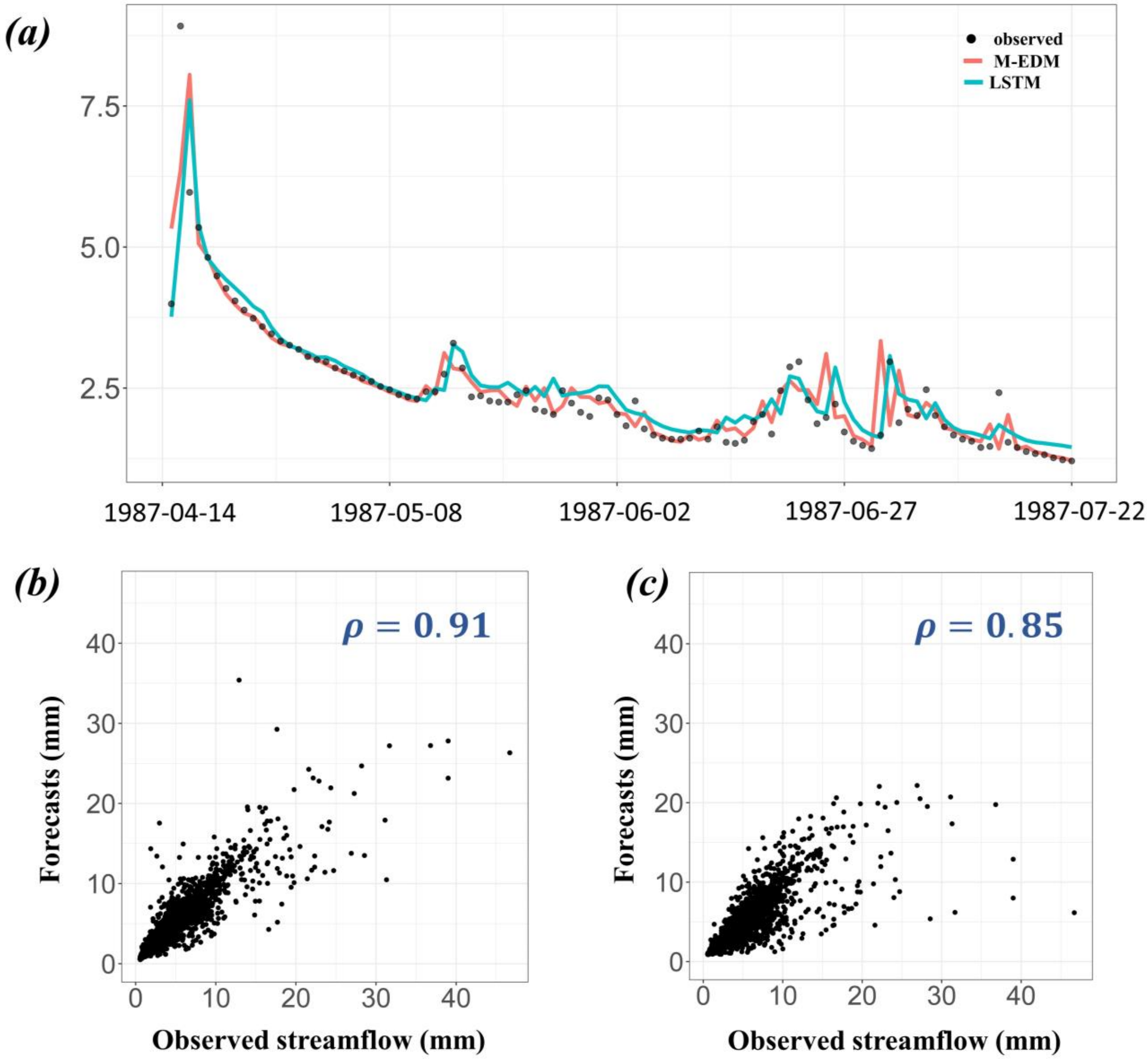


Figure 4.5 *(a). Segment of streamflow time series at catchment (ID 03504000): observed (black dots), forecasted from m-EDM (red) and forecasted from LSTM (cyan). The forecast time series shown are for one day ahead ($T = 1$). (b) Scatter plot of observed streamflow (x-axis) and forecasts of m-EDM (y-axis). (c) Same as in (b), but for forecasts of LSTM.*

Figure 4.6 shows boxplots of ρ and *NSE* (at different thresholds) for streamflow forecasts of m-EDM and LSTM at the 9 catchments. By examining the boxplots of ρ and *NSE* (entire record), one immediately infers that the accuracy of forecasts is nearly equivalent for both methods with a ρ median of 0.865 and 0.869, and an *NSE* median of 0.749 and 0.749 for m-EDM and LSTM respectively. LSTM appears to provide better overall performance as evident by higher values of ρ and *NSE* at the 25th percentile (i.e. bottom line of the boxplots). However, the utility of the m-EDM approach is clearly shown when we examine the forecast accuracy for streamflow observations above certain thresholds. Figure 4.6 shows *NSE* values for streamflow observations above 75th, 85th and 95th percentiles. Remarkably, m-EDM shows a median of 0.87, 0.87 and 0.81 compared to LSTM medians of 0.52, 0.44 and 0.17 for the three thresholds respectively. This last result motivates us to examine the performance of m-EDM in forecasting extreme streamflow as well as low flows which will be shown in the next subsection.

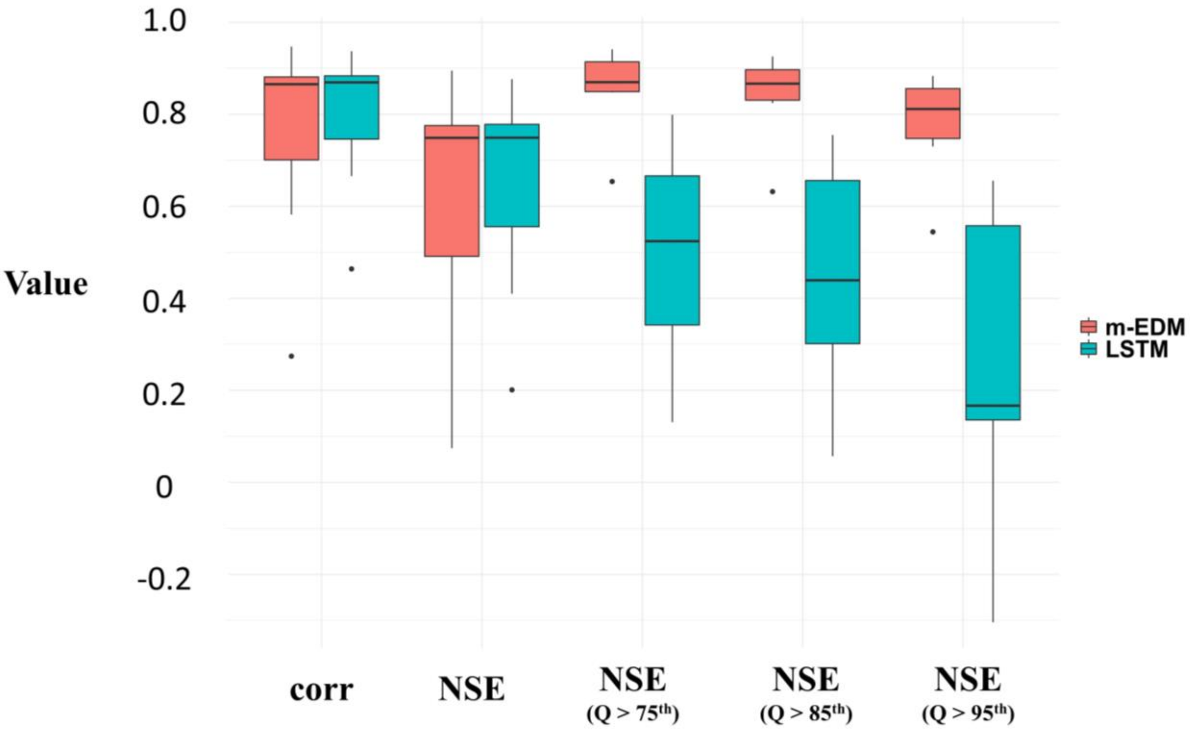


Figure 4.6 *Boxplots show the forecast (1 day ahead) accuracy for m-EDM (red) and LSTM (cyan) in terms of correlation coefficient (ρ), Nash-Sutcliffe efficiency (NSE), and NSE for streamflow above thresholds of 75th , 85th and 95th percentiles.*

Table 4. 2 *The statistics of streamflow forecast accuracy at the 9 catchments, namely correlation coefficient (ρ), Nash-Sutcliffe efficiency (NSE) and root mean squared error (RMSE).*

Site ID	***ρ***		***NSE***		***RMSE (mm)***	
	m-EDM	***LSTM***	***m-EDM***	***LSTM***	***m-EDM***	***LSTM***
02118000	0.78	**0.93**	0.61	**0.86**	0.78	**0.48**
02143500	0.70	**0.75**	0.49	**0.56**	1.28	**1.2**
02414500	0.87	0.87	0.75	0.75	**0.88**	0.93
03455000	0.88	0.88	0.78	0.78	0.57	0.57
03504000	**0.91**	0.85	**0.84**	0.72	**1.20**	1.63
06888500	0.58	**0.67**	0.32	**0.41**	1.97	**1.83**
07163000	0.27	**0.46**	0.07	**0.20**	**2.11**	3.10
12144500	**0.88**	0.87	**0.78**	0.75	**3.25**	3.44
13351000	**0.95**	0.94	**0.90**	0.88	**0.12**	0.14

b. Performance in extreme and low flows

Here, we evaluate the performance of m-EDM in forecasting extreme and low streamflow events. The two measures chosen for these events are annual maximum series of streamflow and annual minimum 7-days streamflow. The former is commonly used in frequency analysis of streamflow to develop flow duration curves, whereas the latter is a commonly used measure for low flows that is used in determining the waste load in streams, and safe withdrawal of water from streams. Both time series are extracted from each hydrological year (Oct 1st – Sep 30th). The minimum 7-day streamflow in a given year is identified by first computing a running average with a window of 7 days, then extracting the minimum value.

Figure 4.7a shows boxplots of the RMSE for annual maximum series of streamflow. The median of RMSE for m-EDM is 8.4 mm, and it is lower than that of LSTM (10.5 mm). Figure 4.7b shows boxplots of the RMSE for annual minimum 7-days flow; the RMSE for m-EDM (0.03 mm) is significantly lower than that of LSTM (0.14 mm). These results are consistent with the observations in the previous subsection, and they corroborate the finding that m-EDM is superior to LSTM in forecasting extreme and low flows. We speculate that the reason behind this is that LSTM, like other ANNs, is trained by minimizing an objective function; here, the objective function is mean squared error. Thus, LSTM will minimize the objective function at the expense of extreme events at the tail of the distribution (i.e. there are few extreme events, and their total contribution to the objective function is insignificant). Similarly, low flow events will be overlooked in the optimization because their values are small, and their combined contribution to the objective function is negligible. On the contrary, the framework of m-EDM is not based on minimizing an objective function which is an advantage of m-EDM over regression techniques, including neural networks (e.g. LSTM).

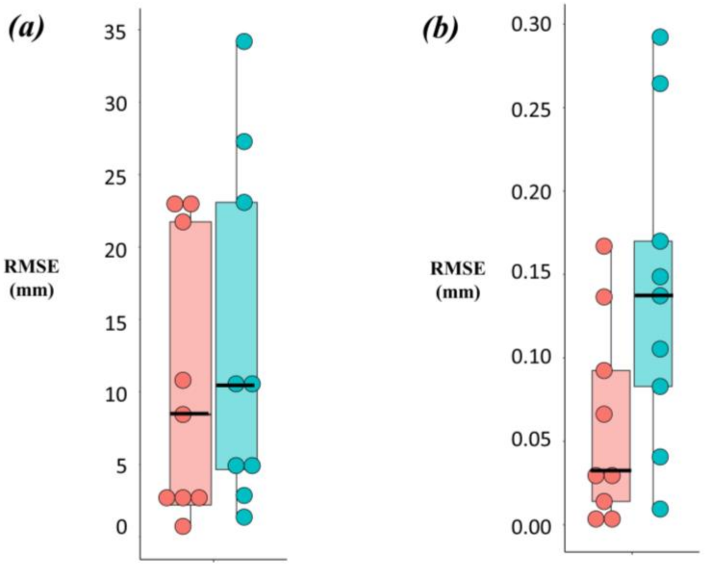


Figure 4.7 (a) *Boxplots of root mean squared error (RMSE) for annual maximum series of streamflow forecasted from m-EDM (red) and LSTM (cyan).* ***(b)*** *Same as in (a), but for annual time series of minimum 7-day streamflow. In both (a) and (b), circles show the values of RMSE for each catchment.*

4.5 Concluding Remarks

This study presented a novel algorithm that builds on recent developments in causal inference to extend univariate empirical dynamical modelling to the multivariate setting. Consequently, this allows us to leverage information from variables such as precipitation, potential evapotranspiration and daily maximum temperature to obtain streamflow forecasts. Convergence of the algorithm to the optimum embedding coordinates is guaranteed asymptotically. Next, we examined the performance of the algorithm in streamflow forecasts at 9 catchments in the contiguous United States that span a wide range of basin size and climate. The results are compared against the forecasts obtained from a Long short-term memory (LSTM) network. Our rationale behind the selection of LSTM as a benchmark model is the remarkable strength of LSTM in forecasting sequences such as time series which has been shown in several studies including a handful in hydrometeorology.

It is shown that while both approaches achieve an adequate and comparable performance with a median *NSE* of 0.75, m-EDM is significantly superior in forecasting extreme and low flow events. This is demonstrated by a median *NSE* (for events above 75th percentile) of 0.87 compared to 0.52 form LSTM. Moreover, the *RMSE* values of annual maximum streamflow (8.4 mm) and minimum 7-days flow (0.03 mm) obtained from m-EDM are significantly lower than that of LSTM, 10.5 mm and 0.14 mm respectively. We attribute the reamrkable differences between the two methods in forecasting extreme and low flow events to the inherent learning process in each method. On one hand, the training of LSTM is carried out by minimizing an objective function which in this study is mean squared error. As a result, the algorithm will be tailored to capture the mean behaviour, paying little attention to extreme events at the tail of the distribution and low flow events due to their minimal contribution to the objective function. On the other hand, the learning process in m-EDM is not based on minimizing an objective function. In fact, the m-EDM approach utilizes a non-uniform and state-dependnt functions to transform inputs into streamflow forecasts. This dichotomy in training m-EDM and LSTM is not limited to these two particular methods but extends to their general classes of empirical models, namely phase space prediction and regression methods respectively. It is plausible that the objective function in training LSTM can be manipulated to favour extreme or low flow events; perhaps, by implementation of regularization techinques. This will, however, be at the expense of decreasing overall forecasts accuracy. On the contrary, m-EDM does not suffer from this tradeoff which is a main advantage of the algorithm.

Although we only presented daily streamflow forecasts in this study, the m-EDM approach can seamlessly be applied at other temporal scales (e.g. hourly and sub-daily). The asuumption we adopted here that the maximum lag time is euqialvent to the embeding dimension will most likely need to be relaxed. Specifically, an educated guess of the maximum lag time based on expert knowledge might be preferable.

Chapter 5

5 Information-theoretic Diagnosis of Infrared Brightness Temperature and Precipitation Relationship

“This chapter is extracted from Ombadi et al. (2021b) with few edits incorporated for brevity and clarity”

5.1 Introduction

Precipitation is a vital component of the hydrologic cycle, and its accurate measurement across spatiotemporal scales provides indispensable information for management of water resources, design of resilient infrastructure and monitoring the functioning of ecosystems. In recent decades, satellite-based precipitation estimation proved to be effective and complementary to conventional precipitation measurement techniques of in-situ gauges and radars. Satellites provide near-global coverage over both land and oceans at high spatiotemporal resolutions, and they are not prone to instrumentation damage resulting from extreme weather conditions — an issue that limits the effectiveness of in-situ rainfall gauges. They are also not susceptible to issues of mountain blockage that often limits the spatial coverage of radars. Both geostationary Earth orbiting (GEO) and low Earth orbiting (LEO) satellites provide useful information for the estimation of precipitation rates. The former provides imagery in multiple spectral bands within the visible and Infrared (IR) wavelengths that can be used to infer cloud-top characteristics whereas the latter deploy passive microwave (PMW) sensors to provide information on the atmosphere column. While both sources are used to derive precipitation, IR imagery due to its high temporal resolution (5 – 30 mins) is the primary input data used by several operational algorithms such as the PERSIANN algorithms (Hsu et al., 1997; Sorooshian et al, 2000; Nguyen et al., 2018) and Hydro-Estimator (Scofield & Kuligowski, 2003). Moreover, IR data is often used in tandem with PMW information to provide precipitation rates by algorithms such as the Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis (TMPA; Huffman et al., 2007). The essential idea of using IR imagery to estimate precipitation rates relies on the empirical evidences that clouds with cold brightness temperature are strongly correlated with precipitation amount; similarly, the spatial extent of cold clouds is associated with that of surface precipitation. These empirical findings reflect the complex and indirect physical relationship between surface precipitation rate and the characteristics of cloud tops.

Broadly speaking, IR-based precipitation estimation algorithms can be categorized into two classes. First, algorithms that relate precipitation rate at a given spatial grid to the corresponding IR brightness temperature at the same grid. The prototypical algorithm of this class is the Geostationary Operational Environmental Satellite (GOES) Precipitation Index (GPI; Arkin & Meisner, 1987). GPI assigns a constant precipitation rate of 3 $mm\ h^{-1}$ for all spatial grids with brightness temperature lower than $235\ K$. Other variants of the GPI algorithm have also been proposed, and they include the Adjusted GPI algorithm (Adler et al., 1994) and the Threshold-Matched Precipitation Index (TMPI; Huffman et al., 2001). The former holds the $235\ K$ threshold constant while allowing the precipitation rate to vary in space and time, whereas the latter allows both the brightness temperature threshold and the precipitation rate to vary locally by month. Second, there are algorithms that attempt to extract information from IR brightness temperature at neighboring spatial grids to better characterize the properties of clouds. These include among others the method proposed by Griffith et al. (1978), the PERSIANN algorithm (Hsu et al., 1997) and the PERSIANN – Cloud Classification System (PERSIANN-CCS; Hong et al., 2004). The latter algorithm classifies clouds into different groups and fits a distinct exponential function for each group to describe the relationship between IR brightness temperature and precipitation.

Several qualitative statements were previously made about the advantages and limitations of using IR brightness temperature to estimate precipitation. Most notable is the strong (low) association between IR data and convective (non-convective) precipitation — an assertion supported by both empirical evidences (e.g. Arkin & Meisner, 1987) and physical reasoning. This is because IR brightness temperature is strongly correlated with the height of clouds, and hence, the depth and strength of convection; on the contrary, it is not clear how IR relates to non-convective precipitation. Another limitation of using IR imagery to estimate precipitation is the inability of IR brightness temperature to satisfactorily resolve regional precipitation patterns caused by complex terrain (Arkin & Meisner, 1987; Dinko et al., 2008; Nguyen et al., 2020). Moreover, IR imagery is generally not efficient in distinguishing between non-precipitating cold cirrus clouds and convective clouds (Hong et al., 2007) although some empirical methods were relatively successful in separating convective and cirrus clouds (e.g. Adler & Negri, 1988). Apart from these limitations, IR data suffers from issues related to the viewing geometry of geosynchronous satellites at higher, extratropical latitudes (Wark et al., 1962; Huffman et al., 2001). It is, therefore, important to carry out studies with the aim of investigating the significance of the foregoing statements and developing novel hypotheses on the inherent relationship between IR and precipitation. Such studies can potentially pinpoint spatial domains and temporal scales at which IR imagery is not an appropriate proxy for estimating precipitation and hence must be supplemented with ancillary data sources to obtain reliable precipitation estimates. Additionally, diagnosis studies can highlight regions at which dependence between IR and precipitation is stronger than currently reported by IR-based operational algorithms, indicating the potential for further development of these algorithms.

Although some of the early studies related to Infrared-based rainfall estimation (e.g. Stout et al., 1979; Arkin, 1979; Richards & Arkin, 1981) were mostly concerned with diagnosing the relationship between IR brightness temperature and precipitation at different temporal and spatial scales, the focus since then has been shifted to address the more practical question of acquiring reliable precipitation estimates from IR imagery. As a result, little attention, if any, has been directed towards further diagnosis of the IR and precipitation inherent relationship despite the numerous benefits that may be gained from such an investigation. Most of the early results obtained on the relationship between Infrared brightness temperature and precipitation were based on correlation analysis of data from the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE; Hudlow, 1979) collected from a period of 3 months over a small spatial domain in the Atlantic Ocean. It should also be emphasized that the common practice in recent years of evaluating satellite-based precipitation estimates with respect to measurements of in-situ rainfall gauges is not an alternative to diagnosis studies. This is because the outcomes of evaluation studies can only be interpreted within the limited context of the specific algorithm used to obtain the precipitation estimates. In other words, it is unknown whether the observed errors in satellite estimates of precipitation should be attributed to the inherent information content of IR imagery, the assumptions and adequacy of the algorithm used to obtain the estimates or a combination of both.

To this end, the objective of this study is to diagnose the inherent relationship between IR imagery and precipitation using state-of-the-art information-theoretic techniques along with extensive dataset of more than 1.3×10^9 pairs of IR and precipitation observations. Specifically, we address the following questions: (1) What are the regional patterns of dependence between IR brightness temperature and precipitation? (2) Which factors exert a first-order impact on these regional patterns? (3) How do the dependence relationships change across the hierarchy of time and space scales? The novelty of the present study lies in the following. First, we use an information-theoretic measure capable of detecting a wide range of dependence relationships; thus, it is suitable to investigate the nonlinear, space-time varying relationship between IR and precipitation that was previously explored using only linear measures (i.e. correlation). Second, our approach is free of specific assumptions on IR brightness temperature thresholds commonly adopted in IR-based precipitation estimation algorithms; therefore, the outcomes of the study are generalizable and of benefit to all algorithms that utilize IR brightness temperature to estimate precipitation. Finally, the study area we consider here, namely the contiguous United States (CONUS), encompasses a wide range of precipitation regimes and to our knowledge is the largest study area to be examined for such analysis.

5.2Data and Study Area

a. NCEP STAGE IV Precipitation

STAGE IV precipitation dataset provides hourly multi-sensor observations at a spatial resolution of 4 km. It is produced by the National Center for Environmental Prediction (NCEP) at the National Oceanic and Atmospheric Administration (NOAA), and it combines observations from automated rain gauges with those obtained from the national network of Weather Surveillance Radar – 1988 Doppler (WSR-88D). Specifically, the precipitation estimates obtained by radars are bias adjusted using observations from approximately 3000 rain gauges (Lin and Mitchell, 2005). In this study, we use STAGE IV hourly precipitation observations for the entire year of 2010. The rationale behind the selection of the year 2010 is that it represents a normal year of precipitation activity in CONUS. That is, it is neither a year of strong El Niño nor is it a part of the multiyear drought that hit California in the years 2011 – 2015 (Seager et al., 2015; Luo et al., 2017). More specifically, annual precipitation of the year 2010 represents the 55th percentile of the distribution of annual precipitation for the period (2002 – 2019). STAGE IV dataset covers the entire contiguous United States at the daily scale; however, the hourly dataset used in the present study exhibits missing observations in the Pacific Northwest of CONUS and parts of Nevada (the unshaded region within CONUS in Figure 5.1a). It is worthwhile to mention that STAGE IV precipitation observations are free of any satellite-based rainfall estimates; thus, they are independent of Infrared brightness temperature data.

b. Infrared (IR) Data

The IR brightness temperature dataset used in the present study is provided by NCEP (Janowiak et al., 2001). This dataset merges Infrared data from five geostationary satellites: GMS, Meteosat-7, Meteosat-5, GOES-8 and GOES-10 to provide uniform observations at a spatial resolution of 4 km and half hourly temporal resolution. The peak frequency of IR channels for the five satellites ranges between 10.7 and 11.5 microns. Specifically, the peak frequencies in microns for each satellite are as follows: GMS (11.0), Meteosat-7 (11.5), Meteosat-5 (11.5), GOES-8 (10.7) and GOES-10 (10.7). The biases resulting from differences in the peak frequency of IR channels are corrected using a calibration strategy (Janowiak et al., 2001). Moreover, an adjustment algorithm is utilized to correct for viweing geometry effects. This is because the IR brightness temperature at spatial grids far from satellite nadir will be colder than if they were directly at the nadir. In particular, the correction methods of Joyce et al. (2001) and Joyce and Arkin (1997) are employed to correct for viewing angle effects. This dataset is available from November 1998 at quasi-global coverage (60°N - 60°S), and it is the primary IR data used by operational satellite-based precipitaiton estimation algorithms. In the present study, we only use IR data at the same spatial extent of STAGE IV precipitation data (gray shaded region in Figure 5.1a).

c. Data Pre-processing

Both STAGE IV precipitation and IR datasets are available at spatial resolution of 0.04° x 0.04°. In this study, we examine the dependence between the two variables at different spatial resolutions (0.04° x 0.04°, 0.1° x 0.1°, 0.25° x 0.25° and 1° x 1°); therefore, whenever it is needed, we resample the dataset from its native spatial resolution to the target one by averaging values of all the fine grids contained within a coarse grid. The resampling throughout this study is exclusively an upscaling (i.e. fine to coarse spatial resolution). Another pre-processing step is needed in order to match the temporal resolution of the two datasets. Specifically, STAGE IV dataset is available at an hourly temporal resolution reporting an accumulated amount, whereas IR data has 30-minutes native resolution representing instantaneous observations. Therefore, we average the two instantaneous IR observations within each hour and use the resulting value in direct correspondence to that of STAGE IV precipitation at the same hour.

d. Study Area

In the present study, we consider the part of CONUS covered by STAGE IV hourly precipitation dataset (gray shaded region in Figure 5.1a) as the primary study area. In order to thoroughly investigate the impact of regional precipitation regimes on the dependence between IR brightness temperature and precipitation, we divide CONUS into 8 distinct climatic divisions. These climatic divisions, indicated by numbers in Figure 5.1a, are named as follows: (1) Northeast, (2) Upper Midwest, (3) Ohio Valley, (4) Southeast, (5) South, (6) Northern Rockies and Plains, (7) West and (8) Monsoonal Southwest. The first six divisions conform with the US climate regions defined by the National Centers for Environmental Information (Karl & Koss, 1984). However, the 7th and 8th climate divisions in this study, namely West and Monsoonal Southwest, deviate from the conventional division. Specifically, we separate the states of Arizona and New Mexico from the remaining states in Southwestern CONUS and define them as the 8th climate division. This is because precipitation in these two states is mostly dominated by the North American Monsoon, and it differs from that of the neighboring states. Figure 5.1a, 5.1b and 5.1c show the boundaries of the eight climatic divisions, mean precipitation of the year 2010 (mm/day) and number of no-rain days in 2010, respectively. It is clear from Figures 5.1b and 5.1c that there is an east to west gradient in annual precipitation amounts and number of no-rain days. For instance, some regions within climate division 1 show more than 150 rainy days per year, whereas the semi-arid part of climate division 7 averages less than 30 rainy days per year. Both figures also show the effect of topography in precipitation amounts and number of rainy days; for instance, the bands of higher annual precipitation and number of rainy days along the Sierra Nevada and Rocky Mountains.

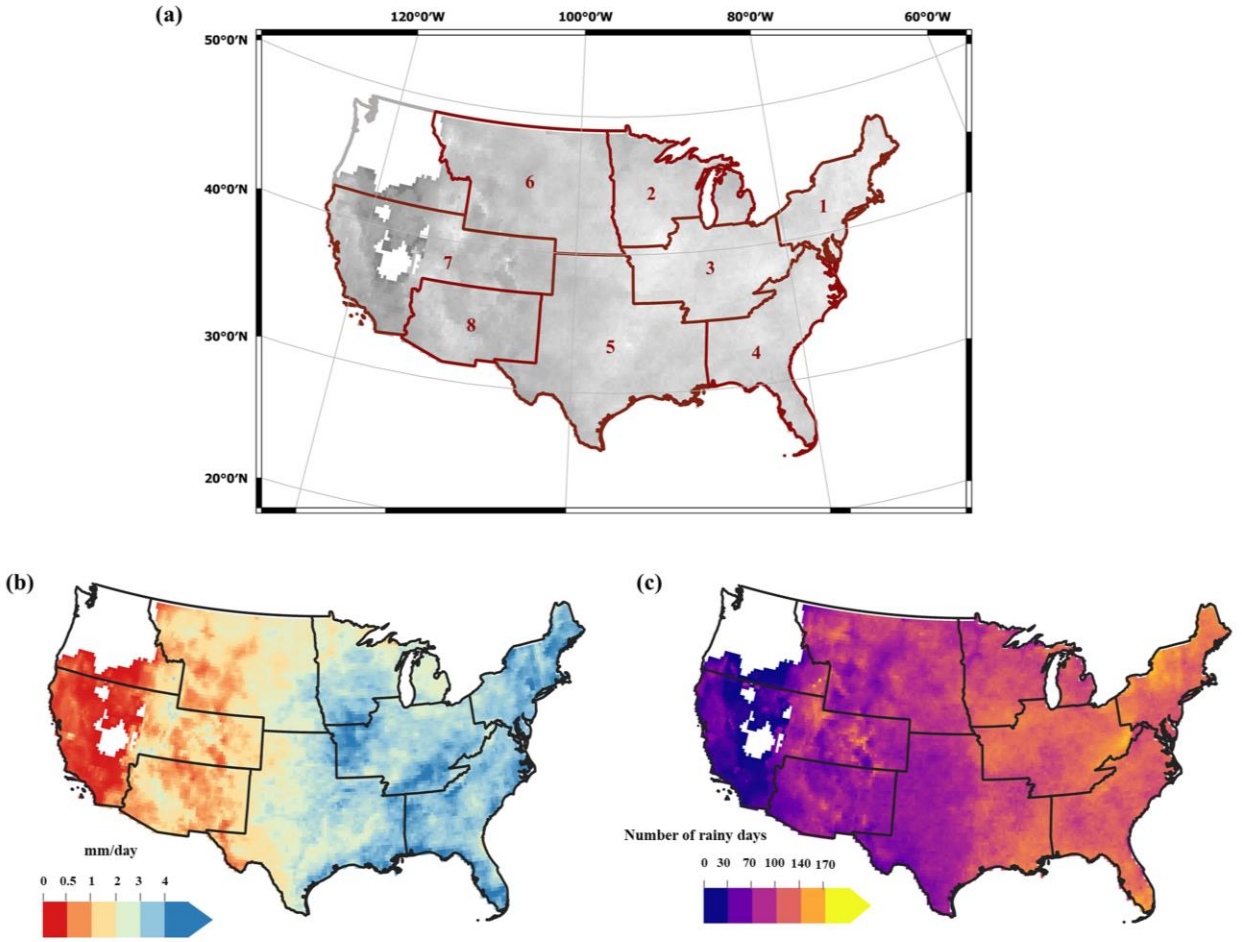


Figure 5.1 (a) *The contiguous United States (CONUS) considered in the present study is divided into 8 climatic divisions: (1) Northeast, (2) Upper Midwest, (3) Ohio Valley, (4) Southeast, (5) South, (6) Northern Rockies and Plains, (7) West and (8) Monsoonal Southwest. The gray shading indicates the spatial extent*

of the hourly Stage IV dataset used in the present study. ***(b)*** *Annual precipitation of the year 2010 computed from STAGE IV data and expressed in units of mm/day.* ***(c)*** *Number of rainy days in the year 2010 computed from STAGE IV data; a threshold of 1 mm is used to filter no-rain days.*

5.3Methods

In order to satisfactorily characterize the dependence between IR brightness temperature and precipitation, a measure of dependence with the following desirable properties is needed. First, it must be able to detect a wide range of relationships — both functional and non-functional, linear and nonlinear. Second, its value should be bounded from both ends (i.e. upper and lower bounds) to facilitate interpretability across time and space scales. The first property is important because the relationship between IR and precipitation is not entirely linear and certainly exhibits some nonlinearities, whereas the second property allows us to compare the dependence relationships at different regions and distinct seasons. While all information-theoretic measures satisfy the first property to some extent, the second one is difficult to achieve with the traditional information-theoretic measures. Fortunately, however, the maximal information coefficient (MIC) proposed by Reshef et al. (2011) satisfy both properties in addition to several other desirable ones. It was, therefore, selected in the present study to examine the dependence relationship between IR and precipitation. In the following two subsections, we first discuss briefly the fundamentals of information theory and then describe the MIC measure.

a. Information Theory

The early work of Claude Shannon (1948) laid the foundation of information theory which has been applied since then in numerous fields — more recently in environmental and hydrological sciences (e.g. Singh, 1997; Sharma, 2000; Ombadi et al., 2020). The most fundamental concept of information theory is the entropy (H) which quantifies the uncertainty in the distribution of a random variable. In particular, the entropy of a continuous random variable X can be estimated as follows:

$$H(X) = \int f_x(x) \log_b \frac{1}{f_x(x)} \, dx \tag{5.1}$$

Here, $f_x(x)$ is the marginal distribution of the variable X, and b is the base of the logarithm which is taken to be 2. If one is interested to quantify the dependence between two random variables X and Y, the entropy in equation 5.1 can be extended to define a dependence measure known as mutual information (MI) defined as follows:

$$MI(X, Y) = \iint f_{xy}(x, y) \; log \frac{f_{xy}(x, y)}{f_x(x) f_y(y)} \, dx \, dy \tag{5.2}$$

In equation 5.2, f_{xy} is the joint distribution of variables X and Y, whereas f_x and f_y are the marginal distributions of variables X and Y respectively. Mutual information possesses several useful traits, and thus it is often used as a measure of dependence in a wide range of applications. Specifically, MI is free of assumptions on the underlying functional form of dependence; thus, it is capable of characterizing both linear and nonlinear dependence. Moreover, MI has a lower bound of zero which indicates independence of random variables. This can easily be inferred from equation 5.2; specifically, if the two random variables X and Y are independent, then the numerator of the log can be decomposed into $f_{xy}(x, y) = f_x(x) \, f_y(y)$ which leads to $MI = 0$. Despite these useful properties, MI is not adequate to characterize the dependence between IR and precipitation.

This is because MI is only lower-bounded, and thus it is not possible to contrast its values across spatial extents. In the following subsection, we will describe an information-theoretic measure that share the same roots of MI, but it overcomes the foregoing limitations.

b. *Maximal Information Coefficient (MIC)*

The fundamental idea underlying the Maximal Information Coefficient (MIC) (Reshef et al., 2011) is that for any bivariate dependence relationship, there exists a grid to partition the data such that the dependency is fully captured. Thus, the procedure to obtain MIC is initialized by partitioning the bivariate data into grids of different size. Let *G* denotes a grid defined by (g_X, g_y) where g_X and g_y refer to the number of bins (partitions) in *x* and *y* coordinates, respectively. The possible integer values of g_X and g_y are defined based on the sample size *n*; in particular, only the integer values of g_X and g_y satisfying the inequality $(g_x \,.\, g_y) < n^{0.6}$ are considered. For a given grid *G*, the mutual information (*MI*) of equation 5.2 is calculated over all possible layouts of the grid, and the maximum value denoted by $\max\{MI_G\}$ is identified. This step is repeated over all possible grids of *G*, and a matrix that consists of the elements $\max\{MI_G\}$ is constructed. The values of these elements are normalized by dividing each value by the corresponding value of $\log(\min\{g_x, g_y\})$. This last step normalizes the variation resulting from the different resolution of the grids, and it ensures that the values of all elements are bounded between 0 and 1. Finally, MIC is obtained as the maximum value across all elements of the matrix. Using the foregoing notations, MIC is formally defined as follows:

$$\mathrm{MIC} = \max_{(g_x \cdot g_y < n^{0.6})} \frac{MI_G}{\log(\min\{g_x, g_y\})} \tag{5.3}$$

MIC satisfies the two desirable properties mentioned earlier as it was proven to be capable of detecting association in linear, nonlinear and periodic functions as well as non-functional dependence (Reshef et al., 2011). Its value is bounded between 0 and 1 and is approximately equivalent to the coefficient of determination $\boldsymbol{R^2}$ only for functional relationships. These two properties allow us to diagnose the dependence between IR brightness temperature and precipitation across different regions with clear interpretability of MIC. In addition to these properties, MIC also accounts for sample length because the possible resolutions of grids *G* given by g_X and g_y are a function of sample length *n*. This last point ensures that we can utilize MIC values to compare relationships at different temporal scales (i.e. different sample length) without any loss of interpretability.

In order to illustrate the robustness of MIC in measuring the dependence between IR and precipitation, we generate a synthetic time series using the following equation:

$$P = 4\exp\left(\frac{-(T_B - 215)}{20}\right) \tag{5.4}$$

Where P and T_B refer to precipitation ($mm\ hr^{-1}$) and infrared brightness temperature (K), respectively. This equation describes an idealized exponential relationship of the type often assumed in algorithms estimating precipitation from IR data (e.g. Hong et al., 2004; Nguyen et al., 2020). It should be noted that the coefficients used in this equation are only used for illustration purposes, and that the value of such coefficients often depends on the region and season of precipitation, for instance, Nguyen et al. (2020) show how such curves are adjusted for different regions and seasons to obtain precipitation estimates from IR brightness temperature. The top panel of Figure 5.2 shows scatterplots for data generated from equation 5.4 with length n of 1000. The leftmost figure in the top panel shows data free of noise, whereas the two figures to the right show the data diluted with gaussian noise with zero mean and a standard deviation $\sigma = c * \sigma_p$. Where σ_p is the standard deviation of the noise-free precipitation data generated from equation 5.4, and c is a factor with values of 0.1 and 0.4. The bottom panel shows the same relationships as the top one except that the sample length n is taken to be 500. For each scatterplot, the value of MIC is computed and shown in Figure 5.2. Clearly, the results demonstrate that for noise-free exponential relationship, MIC equals 1 which highlight its generality to detect nonlinear relationships. As the noise level increases with $c = 0.1$ and 0.4, MIC decreases with values of 0.92 and 0.64, respectively for the top panel. Finally, by comparing MIC values for the top and bottom panels, it can be seen that the differences are not substantial. This highlights that MIC accounts for differences in sample length, and it provides an approximately equal values for the same relationships even when the sample length is halved.

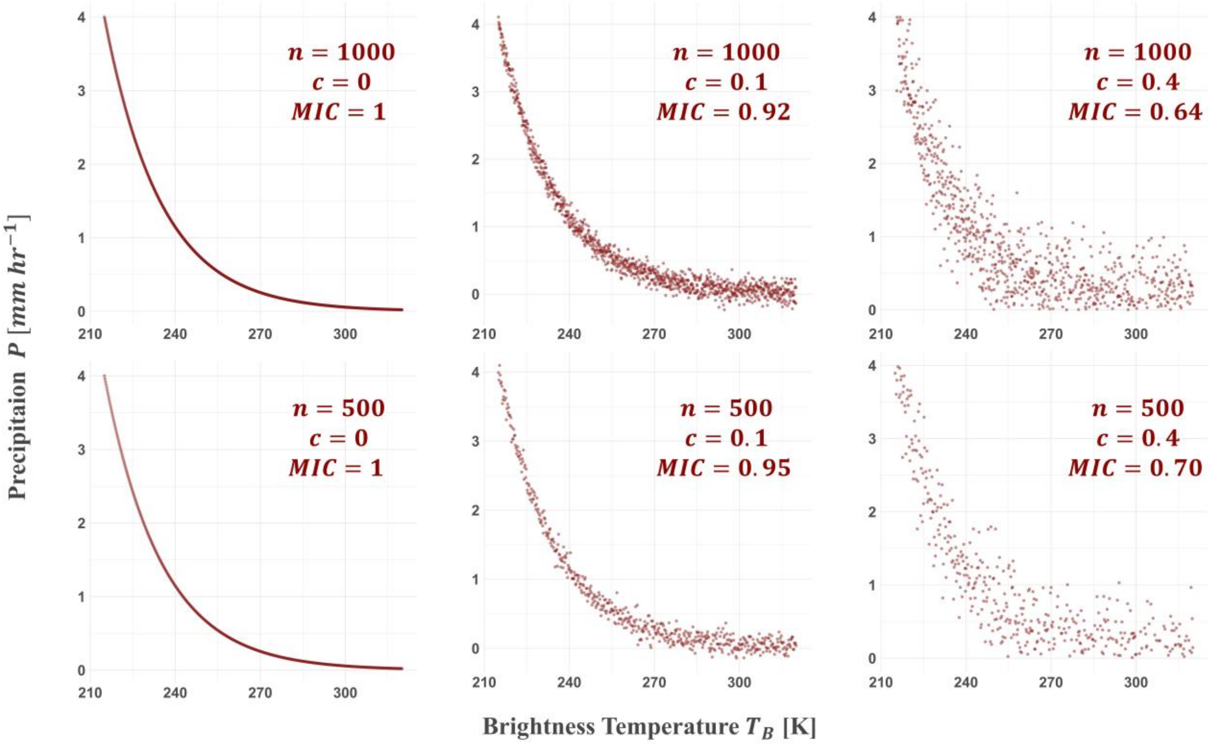


Figure 5.2 *Scatterplots for data generated from equation 5.4. The top panel shows data with sample length* ($n = 1000$) *and noise levels* ($c = 0, 0.1$ *and 0.4) from left to right. The bottom panel shows data with sample length* ($n = 500$) *and noise levels* ($c = 0, 0.1$ *and 0.4) from left to right. MIC values are also shown for each scatterplot.*

5.4 Results and Discussion

a. General and Regional Patterns

We start our analysis by exploring the regional patterns of dependence between IR brightness temperature and precipitation with both temporal and spatial resolution of the data being held constant. Specifically, the two datasets are considered at their highest temporal resolution of 1-hour and the native spatial resolution of 0.04° x 0.04°. The implicit assumption throughout our analysis is that information on precipitation at any given spatial grid is only carried by IR brightness temperature at the same grid. In this study, the issue of spatial aggregation itself is under investigation; therefore, the foregoing assumption is not only justified, but it is even necessary in order to assess the impact of spatial aggregation on the relationship between IR and precipitation.

Figure 5.3 shows the spatial distribution of MIC values for the entire year of 2010 as well as the four seasons: March, April and May (MAM); June, July and August (JJA); September, October and November (SON) and December, January and February (DJF). In general, the values of MIC ranges from a minimum of 0.03 to a maximum of 0.37. At the annual scale, Figure 5.3a shows that parts of the Upper Midwest and Ohio Valley (i.e. climate divisions 2 and 3 respectively) have higher MIC values with respect to other climatic divisions. By comparing Figure 5.3a to Figures 5.3b, 5.3c, 5.3d and 5.3e, it can be shown that there is only moderate variability in the spatial patterns of dependence at the annual scale compared to that of seasonal scales. This is because the combination of all hourly observations during the year obscures the differences resulting from distinct precipitation regimes. It is, thus, important to examine the spatial patterns of dependence for each season separately in order to obtain a more detailed view. Figures 5.3b to 5.3e show the values of MIC calculated for each season separately. Generally, there are prominent spatial features of dependence in each season as well as clear distinction between seasonal patterns of dependence. Figures 5.3b and 5.3c show that IR is highly associated with precipitation in the eastern two thirds of CONUS during MAM and exclusively in the Upper Midwest and Northern Rockies during JJA. This can be related to the strong association between IR brightness temperature and convective precipitation (Arkin & Meisner, 1987). In particular, mesoscale convective systems (MCS) are the dominant precipitation regime over parts of CONUS east of the continental divide, and they are responsible for more than half of annual precipitation and the majority of warm season (May – August) precipitation (Haberlie & Ashley, 2019). The strong dependence between IR and precipitation over parts of CONUS east of the continental divide during MAM and JJA is consistent with previous empirical findings; for instance, Arkin and Meisner (1987) showed that warm season monthly precipitation and IR are strongly correlated over this region with correlation coefficients exceeding 0.75.

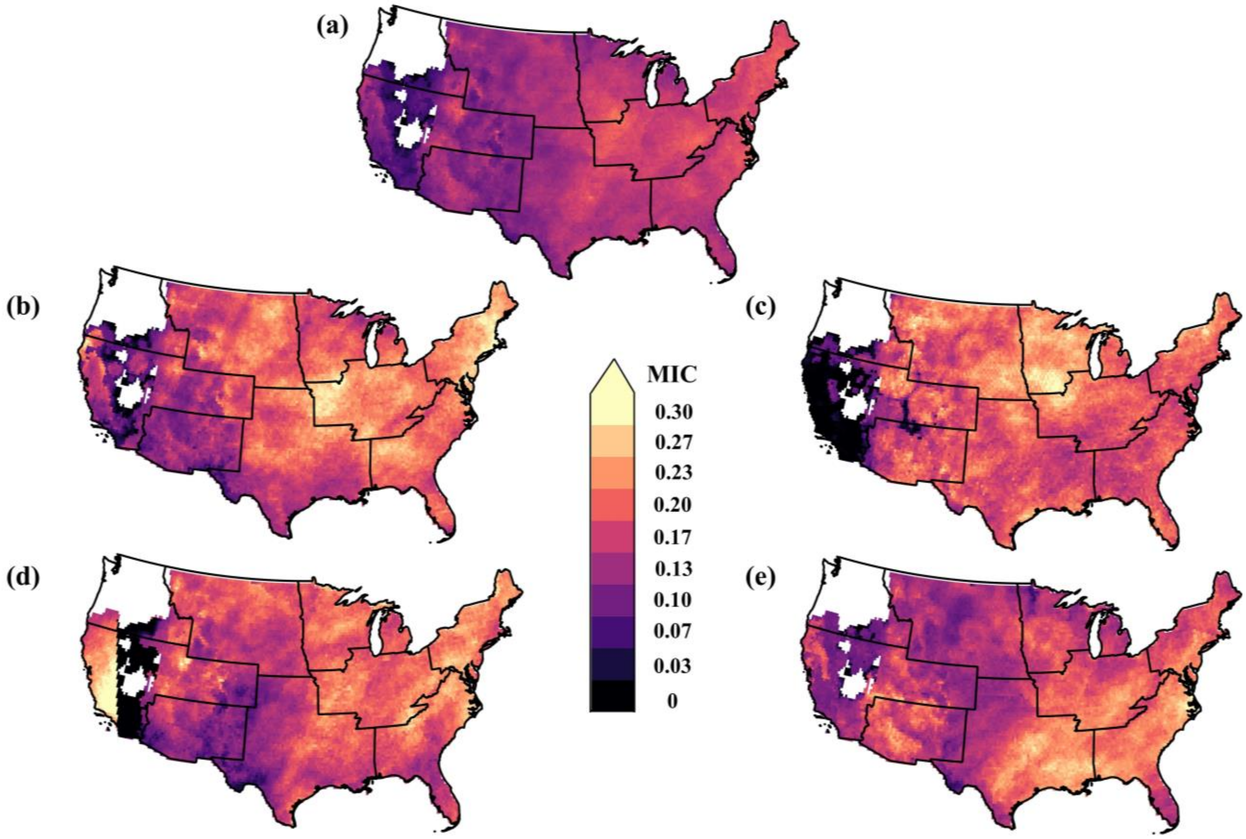


Figure 5.3 Spatial patterns of Maximal Information Coefficient (MIC) computed from IR brightness temperature and Stage IV precipitation, at spatial grids of 0.04° x 0.04° and hourly time resolution, for: **(a)** the entire year of 2010; **(b)** March, April and May (MAM); **(c)** June, July and August (JJA); **(d)** September, October and November (SON); and **(e)** December, January and February (DJF).

Figures 5.3d and 5.3e show the values of MIC for the two seasons of SON and DJF respectively. A main feature during both seasons is that high values of MIC are mostly observed over the Northeast and Southeast (i.e. climate divisions 1 and 4 respectively). Contrary to warm season, precipitation in the Midwest (climate divisions 2, 3, 5 and 6) during SON and DJF is not strongly associated with IR brightness temperature mainly because of the lack of convective storms. The high values of MIC observed in Northeast, and to a lesser extent in Southeast, during SON might reflect the activity of Atlantic hurricanes during the season. This is because the type of clouds associated with hurricanes are mostly of the cumulus and cumulonimbus type, and they share similarities with those of MCS (Houze, 2014). Figure 5.4 shows the mean values of MIC averaged over all spatial grids within each of the 8 climatic divisions of CONUS. These results further highlight the foregoing observations; for instance, the two divisions of Upper Midwest and Northern Rockies and Plains have the highest MIC values during JJA due to the activity of convective precipitation. As for SON and DJF, it is the Northeast and Southeast respectively that show the highest value of MIC.

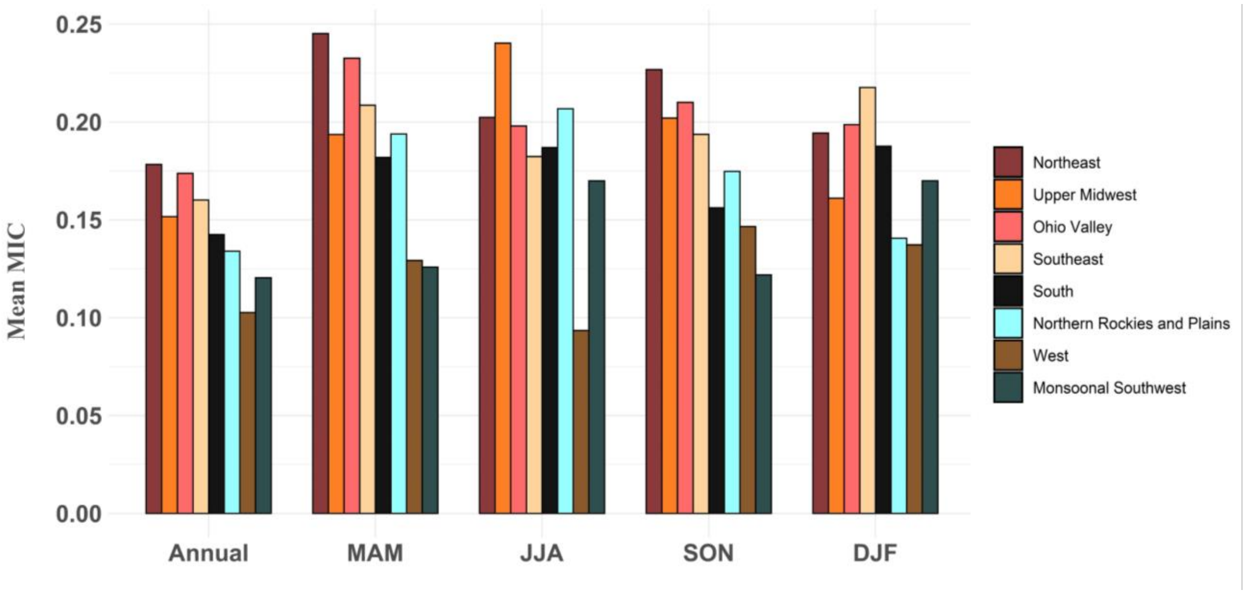


Figure 5.4 *Mean values of MIC for the entire year of 2010 (Annual), MAM, JJA, SON and DJF averaged over the 8 climatic divisions of contiguous United States (CONUS) considered in this study.*

Interestingly, Western United States (climate division 7) shows considerable seasonal variability in the values of MIC ranging from extremely low association during JJA (Figure 5.3c) to high MIC during SON (Figure 5.3d). In particular, the mean values of MIC for JJA and SON are 0.09 and 0.15 respectively; see Figure 5.4. The mean value during SON, however, obscures the spatial patterns shown in Figure 5.3d since it averages the high MIC values over western part of climate division 7 with the extremely low values over the arid part. In order to infer the potential reason behind the pronounced seasonal variability of dependence between IR and precipitation over the West, it is important to point out a crucial aspect of the dependence relationships captured by MIC in the present study. MIC encapsulates two types of dependence: first, the association between IR brightness temperature and the amount of precipitation; second, the adequacy of using IR brightness temperature to identify no rain instances. The latter reflects the ability of identifying no rain instances using only IR brightness temperature. If we exclude the area of climate division 7 covered by mountain ranges (Rocky and Sierra Nevada), the remaining part is mostly arid with extremely few precipitation occurrences. For instance, Southern California averages only 35 days of rain throughout the year (see Figure 5.1c). It is, therefore, plausible to hypothesize that the low MIC values during JJA, MAM and to a lesser extent DJF as well as the pronounced seasonal variability of MIC is caused by the low number of precipitation occurrences. This hypothesis is examined in Figure 5.5a which shows the bivariate relationship between MIC and the number of no rain hourly instances for each spatial grid over CONUS during the entire year of 2010. Clearly, there is a coherent relationship whereby MIC is low for spatial grids that have high number of no rain instances, and it increases consistently as the number of hourly no rain instances decreases. Precisely, the correlation between MIC and the number of no-rain observations is quite significant with a value of -0.73 (p-value $< 2.2 \times 10^{-16}$). This observation provides supporting evidences to the foregoing hypothesis and explains MIC patterns in Western United States. Put simply, Figure 5.5a indicates that, on average, the information content of IR brightness temperature is lower for regions with higher frequency of no-rain instances.

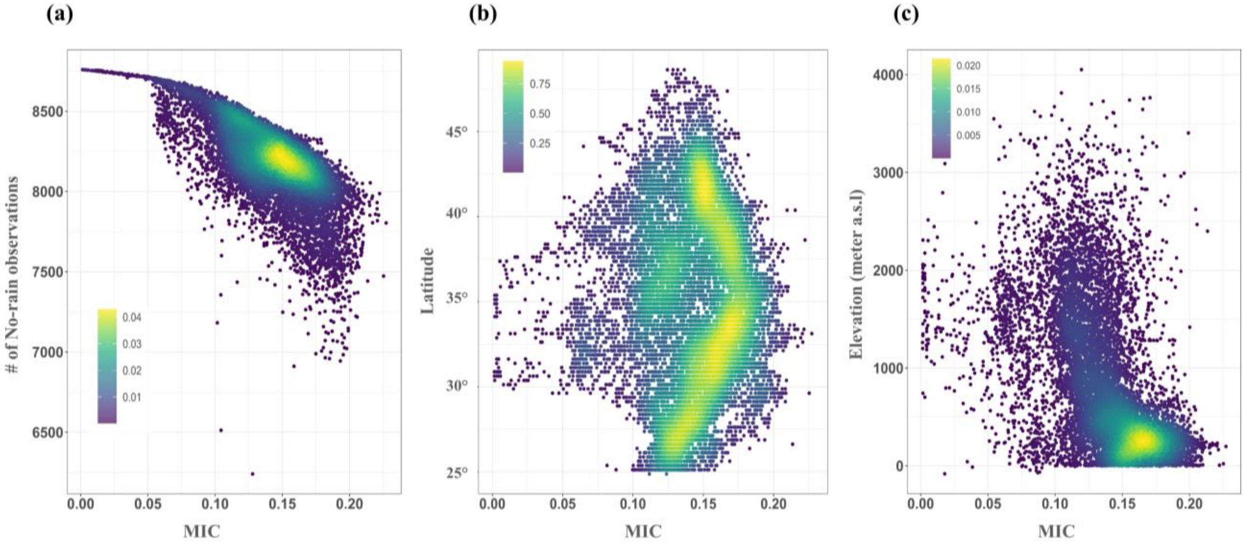


Figure 5.5 *Density plots for the relationship of MIC (horizontal axis) with:* ***(a)*** *number of hourly no-rain observations,* ***(b)*** *latitude and* ***(c)*** *elevation in meters above sea level. The plots show all 0.04° x 0.04° spatial grids within the spatial extent of Stage IV. The color coding represents the density of points in the bivariate distirbution with brighter colors indicating higher density.*

The relationship shown in Figure 5.5a is quite robust for the entire spatial domain of CONUS, and it is perhaps the result of complex interplay of the inefficiency of IR data in separating convective clouds form non-precipitating cirrus clouds and the effects of ground temperature in detecting clear skies. In order to investigate this relationship more thoroughly, we narrow down our analysis to climate division 5 (South). Because the frequency of occurrence of cirrus clouds over this region is extremely low (Sassen et al., 2008), it is feasible to assess separately the impact of ground temperature on the relationship between IR information content and number of no-rain observations. Figure 5.6 shows scatterplots of this relationship during the warm season (JJA) and the cold season (DJF) along with least square regression lines. Clearly, the relationship is stronger during the cold season (DJF) with a correlation coefficient of -0.89 (p-value < 2.2 x 10^{-16}) compared to a warm season correlation of -0.3 (p-value < 2.2 x 10^{-16}). This indicates that, over this region, IR brightness temperature is more effective in identifying no-rain instances in the warm season because of high ground temperatures; however, this effectiveness is reduced in the cold season, leading to the strong relationship between no-rain observations and MIC (red regression line in Figure 5.6). These results highlight that while the strong relationship between no-rain observations and IR information content is robust at the annual scale across CONUS, there is significant variability in this relationship both regionally and seasonally.

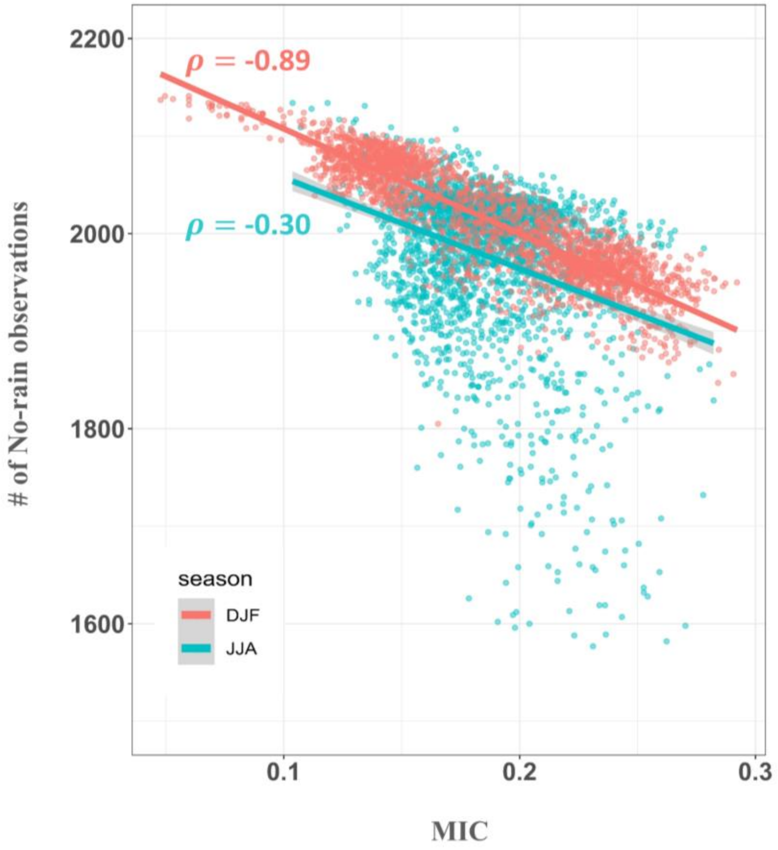


Figure 5.6 *Scatterplots of the relationship between MIC and number of hourly no-rain observations for JJA (cyan) and DJF (red). The plots show all 0.04° x 0.04° spatial grids within the spatial extent of climate division 5 (South). Least squares regression lines are drawn in cyan and red for JJA and DJF, respectivley, with the values of correlation coefficients shown in the plot.*

Apart from the regional patterns observed in Figures 5.3 and 5.4, there might be variability in MIC at small spatial scales resulting from complex terrain and orographic rainfall. Figure 5.5c shows the bivariate relationship between annual MIC and the elevation at each spatial grid in CONUS; the value of Pearson correlation coefficient is -0.53 (p-value $< 2.2 \times 10^{-16}$). As expected, higher values of MIC are associated with regions of low elevation because of the absence of orographic rainfall in these regions. In other words, IR brightness temperature is less informative in estimating precipitation at high elevations where orographic rainfall dominates. This affirms the numerous observations reported by previous studies on the difficulty of estimating orographic rainfall over complex terrain using IR brightness temperature. For instance, Ombadi et al. (2018) and Miao et al. (2015) showed that PERSIANN-CDR, an IR-based precipitation dataset, tends to underestimate precipitation at high elevation regions. Similar observations were also reported for another precipitation dataset that utilizes IR brightness temperature and PMW, namely TRMM 3B43 (Hashemi et al., 2017). Finally, we also assess the potential impact of the satellite viewing angle (i.e. the accuracy of IR data far from satellite nadir) on the MIC values. Figure 5.5b shows that there is no relationship between MIC and latitude since high values of MIC seem to exist both at low and high latitudes. Specifically, Pearson correlation coefficient between MIC and latitude is -0.01 (p-value = 0.25). It is, thus, plausible to conclude that the corrections employed in the IR dataset (Janowiak et al. 2001) mitigated the biases in the data at spatial grids far from satellite nadir.

b. Temporal Aggregation

We now turn our analysis to examine the impact of temporal aggregation on the dependence relationship between IR brightness temperature and precipitation. While the foregoing analysis was performed at the native temporal resolution of 1 hr, here we analyze the data across a hierarchy of temporal scales (Δt) given in hours by $\Delta t = \{1, 3, 6, 12, 24\}$. The IR brightness temperature data at each time scale is obtained by averaging the data from the native temporal resolution, whereas STAGE IV precipitation data is obtained by accumulating the total amount of precipitation from the native resolution. The MIC values are then computed at each spatial grid (0.04° x 0.04°) for the entire year of 2010 as well as for each individual season of MAM, JJA, SON and DJF. In the following discussion, the mean value of MIC is calculated for each temporal aggregation and season by taking the arithmetic average of all the corresponding values of MIC over CONUS.

Figure 5.7a shows the mean values of MIC computed from the four seasons as a function of temporal aggregation. Several interesting observations can be made from the figure. First, it appears that regardless of seasons, the dependence between IR and precipitation becomes stronger as the temporal aggregation increases. This is somewhat expected since the effect of aggregation in time will average the random errors present at fine temporal scales. Second, the rate of improvement in MIC gained by temporal aggregation seems to be more significant at first (e.g. from 1 to 3 hr), but it decreases consistently as the temporal aggregation increases. This behavior of diminishing returns is rather interesting and will be discussed in detail in the following paragraph. Third, across all temporal aggregations, MIC is higher in JJA (orange line) followed by MAM (black line). These observations reflect the patterns we discussed in section 4.1; specifically, IR is strongly correlated with precipitation over most parts of CONUS during JJA and MAM due to the high activity of convective storms.

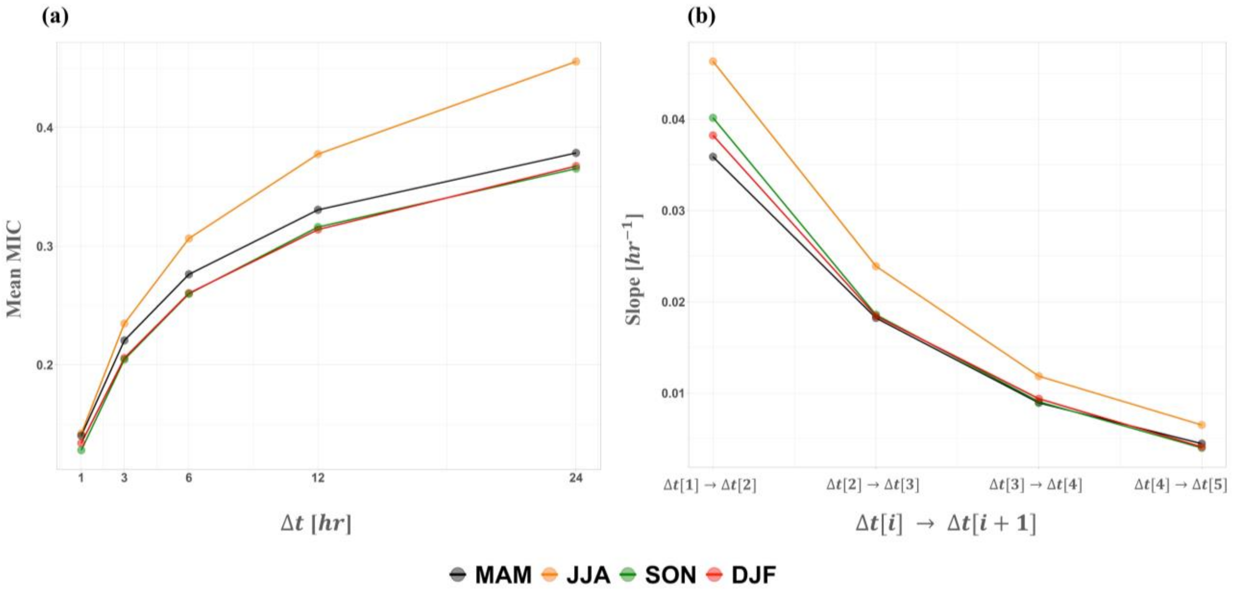


Figure 5.7 (a) *Mean values of MIC for the four seasons (MAM, JJA, SON and DJF) as a function of temporal aggregations in hours given by* $\Delta t = \{1, 3, 6, 12, 24\}$*.* ***(b)*** *The slope of the curves in (a) computed between each two sucessive temporal aggregations. MIC is considered dimensionless; thus, the slopes are shown in units of* hr^{-1}*.*

The interesting behavior of diminishing returns observed in Figure 5.7a is worthy of further investigation. Thus, Figure 5.7b shows the rate of improvement obtained through temporal aggregations $\Delta t[i] \rightarrow \Delta t[i+1]$ for $i = 1, 2, 3$ and 4. In other words, the values in the vertical axis of Figure 5.7b are the slopes of the curves in Figure 5.7a with units of hr^{-1} because MIC is unitless. Figure 5.7b shows that the rate of improvement in MIC as a result of temporal aggregation decreases persistently as we move toward coarse temporal scales. For instance, MIC during JJA increases by 0.044 for each hour as we aggregate from 1 to 3 hr; however, this rate drops to 0.007 as we aggregate from 12 to 24 hr. This pattern is consistent for the remaining seasons, namely MAM, SON and DJF. More interestingly, Figure 5.7b shows that the rate of improvement resulting from temporal aggregation is higher during JJA (orange line) than any other season. Given that the dominant precipitation regime during JJA in most parts of CONUS is convective, the results of

Figure 5.7 do not only indicate that IR is strongly associated with convective precipitation but also highlight that IR and precipitation dependence increases more significantly with temporal aggregation for this specific type of precipitation.

There are very few studies that have explicitly examined the impact of temporal aggregation on the dependence between IR and precipitation. Most notable is the work of Richards and Arkin (1981) who examined the impact of temporal averaging on the relationship between IR and precipitation. Their study was limited to 3 months of hourly data obtained from the GATE experiment over a small spatial domain in the eastern Atlantic Ocean, and they found that temporal averaging has negligible impact on the IR-precipitation relationship. Their analysis was based on spatial grids of 2.5° which is significantly coarser than the 0.04° we used for the analysis in Figure 5.7. It is, therefore, plausible that the discrepancy of the results stems from the difference in spatial resolutions. In other words, it seems that at large spatial scales of approximately 2.5°, the temporal averaging is no longer important and does not lead to improvement of dependence between IR and precipitation. We speculate that the interesting observations inferred from Figure 5.7, namely the diminishing returns behavior, is closely associated with the so-called life-cycle effect (Griffith et al., 1978; Stout et al., 1979; Richards & Arkin, 1981). This effect describes the behavior of convective clouds whereby most of the precipitation occurs during the growing and mature stages of clouds, whereas the dissipating stage produces little or no rain resulting in time lagged dependence between IR brightness temperature and precipitation. These time lag effects are almost entirely subdued when aggregating at fine temporal scales (e.g. 1 to 3 hr) because the life cycle of convective storms is typically less than 2 hours. Therefore, further aggregation at larger time scales will not be equally effective in improving the dependence between IR and precipitation.

c. *Spatial Aggregation*

In this section, we examine briefly the impact of spatial aggregation on the dependence between IR brightness temperature and precipitation. Five different spatial scales, namely 0.04°, 0.1°, 0.25°, 0.5° and 1° are used to examine the impact of spatial averaging. Figure 5.8 shows the mean values of MIC averaged over the study area for the four seasons of MAM, JJA, SON and DJF. This mean value of MIC is calculated by taking the arithmetic average of all MIC values for grids within the domain of STAGE IV observations (gray shaded region in Figure 5.1a). The results show that the impact of spatial averaging is almost identical to that of temporal averaging. Specifically, it can be seen from Figure 5.8a that spatial aggregation results in stronger dependence between IR and precipitation. Also, the dependence is stronger during JJA (orange line), followed by MAM (black line). Moreover, Figure 5.8b shows that the impact of spatial aggregation on IR and precipitation dependence exhibits a diminishing returns behavior. Specifically, the rate of increase in dependence obtained from spatial aggregation during JJA is 0.29 per degree when aggregating from 0.1° to 0.25°; however, this value decreases to only 0.075 per degree as we aggregate from 0.5° to 1.0°. Similar to temporal aggregation, this can be attributed to the life-cycle effect of clouds because the different stages of a storm (growing, mature and dissipation) exhibits changes in the size of clouds; thus, spatial averaging will mitigate these differences. Spatial aggregation is also shown to be more effective in improving the dependence between IR and precipitation during JJA. This is consistent with the results in Figure 5.7, and it indicates that both temporal and spatial averaging have significant impact on the dependence between IR and precipitation with respect to convective precipitation.

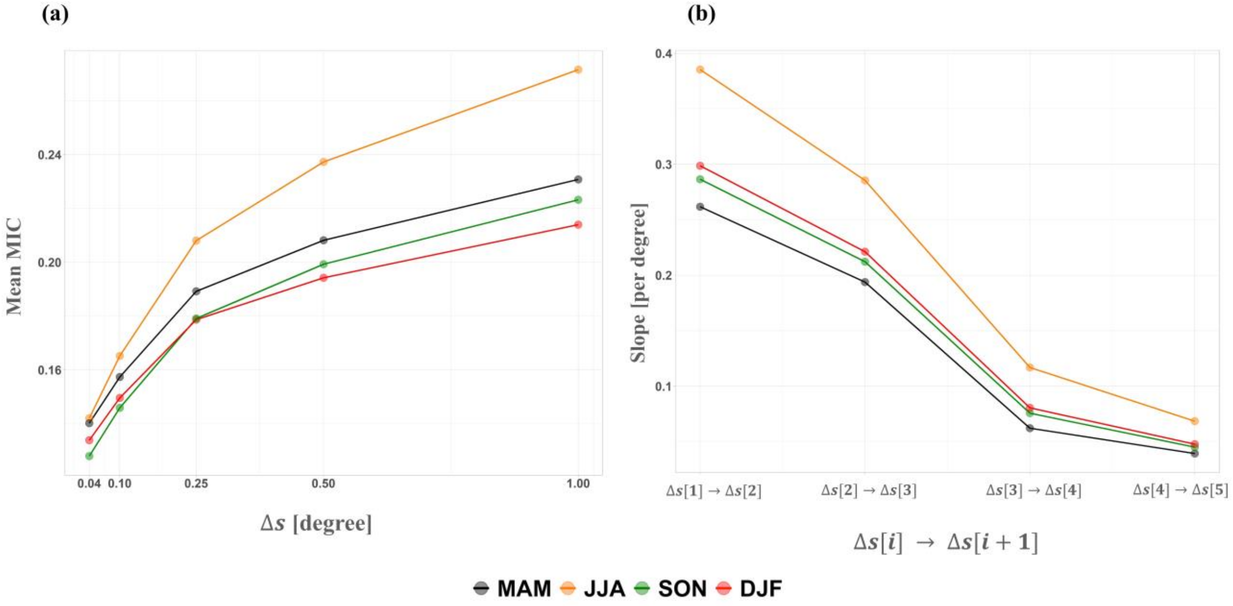


Figure 5.8 (a) *Mean values of MIC for the four seasons (MAM, JJA, SON and DJF) as a function of spatial aggregations given by* $\Delta s = \{0.04^\circ, 0.1^\circ, 0.25^\circ, 0.5^\circ, 1^\circ\}$. ***(b)*** *The slope of the curves in (a) computed between each two sucessive spatial aggregations. MIC is considered dimensionless; thus, the slopes have units of 1/degree.*

d. Spatiotemporal Aggregation

The previous two subsections examined the impact of spatial and temporal aggregation on the dependence between IR and precipitation by changing time or spatial scale while holding the other constant. However, it is important to examine the impact of bivariate (spatiotemporal) aggregation in order to acquire a complete picture on the variability of dependence across time and space scales. Figure 5.9 shows contour maps of MIC values averaged over CONUS and computed for the four seasons across spatial scales of 0.1°, 0.25°, 0.5° and 1°, and temporal scales of 1, 3, 6, 12 and 24 hours. Generally, Figure 5.9 reflects the results in Figures 5.7 and 5.8 with increasing MIC values across the main diagonal (i.e. consistent increase in MIC as we move rightward and upward in the horizontal and vertical axes respectively). The negative slope of contour lines is a reflection of the relationships shown in Figures 5.7 and 5.8. The values of MIC during JJA (Figure 5.9b) are higher than other seasons which reinforces the observations in the previous two subsections; namely, the dependence between IR and precipitation over CONUS is stronger in JJA across all time and space scales due to the high activity of convective storms. Furthermore, the diminishing returns behavior of increased dependence as a result of spatiotemporal aggregation can be inferred from the steep gradient of contours at fine spatiotemporal scales (bottom left corner in Figures 5.9a, 5.9b, 5.9c and 5.9d) and the dispersed contours at coarse scales (upper right corner in Figures 5.9a, 5.9b, 5.9c and 5.9d). Also, the gradient of MIC contours is more significant in JJA (Figure 5.9b) which reflects the results shown in Figures 5.7b and 5.8b.

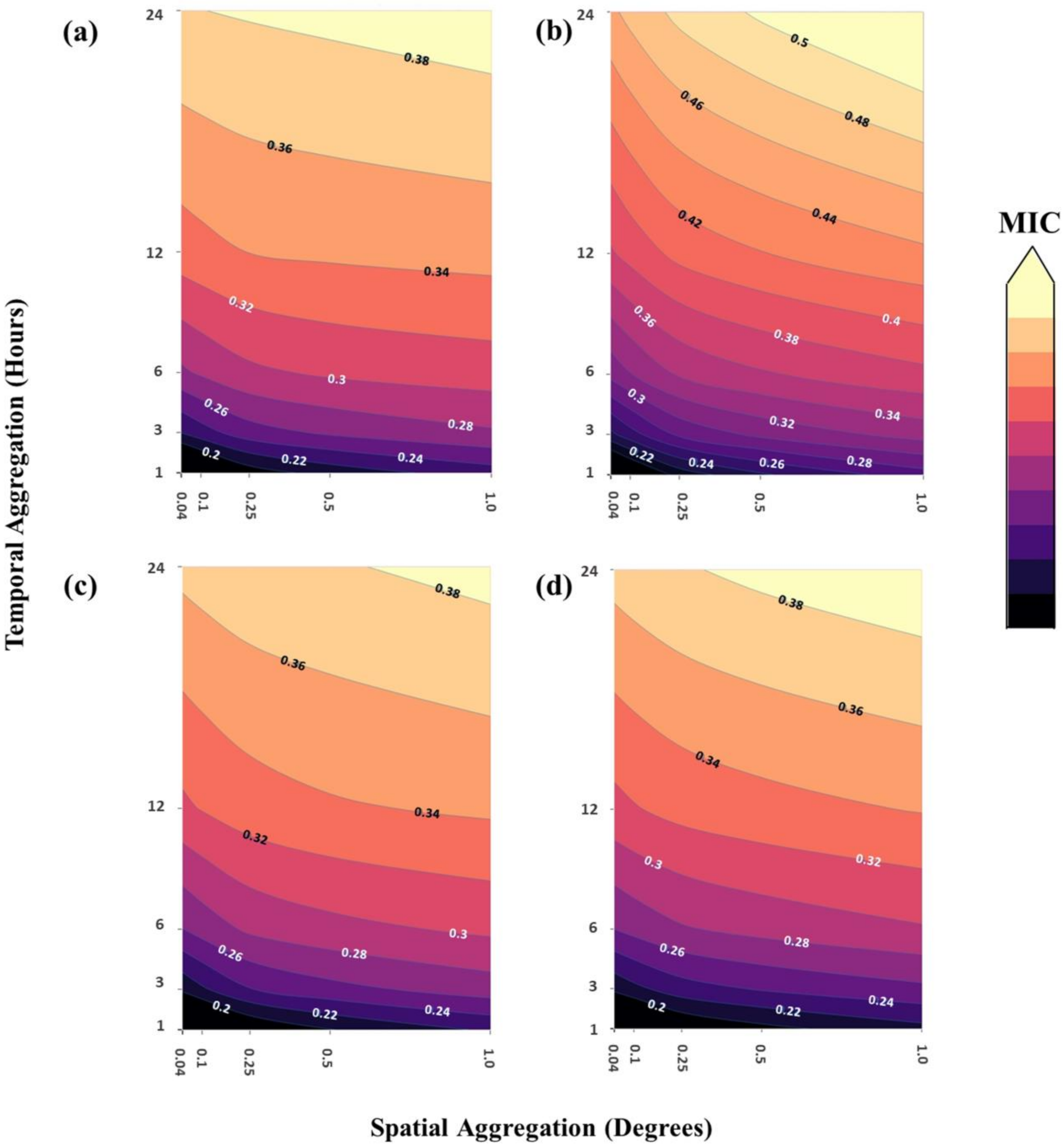


Figure 5.9 Mean values of MIC as a function of both spatial aggregation (degrees) and temporal aggregation (hours) for: ***(a)*** *MAM,* ***(b)*** *JJA,* ***(c)*** *SON and* ***(d)*** *DJF. The surfaces are estimated from values of MIC at spatial resolutions of 0.04°, 0.1°, 0.25°, 0.5° and 1° and temporal resolutions of 1, 3, 6, 12 and 24 hours. Bright colours indicate higher values of MIC. For improved visulaization, the colorbar is adjusted for each panel seperately with the values of MIC shown as contour labels.*

5.5Diagnosis of IR-based precipitation estimation algorithms

One of the practical applications of the analysis presented in this study is to utilize information on the inherent dependence between IR and precipitation, as measured by MIC, to diagnose IR-based precipitation estimation algorithms. Here, we show an example of diagnosing the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) algorithm. PERSIANN is an IR-based precipitation estimation algorithm developed by Hsu et al. (1997), and it represents the backbone of the PERSIANN family of satellite-based precipitation datasets (Nguyen et al., 2018). Specifically, PERSIANN is an artificial neural network (ANN) model that operates in two steps: first, infrared imagery (10.2 – 11.2 μm) are clustered and mapped to a hidden layer; second, the data is mapped from the hidden layer to the output space of rainfall rate. The ANN model is trained (i.e. estimation of parameters) by using information from Passive microwave observations.

The fundamental idea in diagnosing PERSIANN is to first examine its accuracy in estimating precipitation using a benchmark dataset. Then, the accuracy of the algorithm can be compared to the inherent dependence between IR and precipitation. This comparison will allow us to identify regions where the algorithm does not perform as accurate as indicated by the inherent dependence. Similarly, we can identify regions where the algorithm is not responsible for bad performance (i.e. regions where the inherent dependence is extremely low). Figures 5.10a and 5.10d show the value of Pearson correlation coefficient between PERSIANN and Stage IV hourly observations during the 2010 MAM and DJF seasons, respectively. Clearly, PERSIANN performs poorly during MAM over southwestern United States, namely the states of Arizona, Utah, Nevada and California. During DJF season, PERSIANN conspicuously underperforms over Northern

Rockies and plains (climate division 6) and parts of California and Nevada. Diagnosis of the PERSIANN algorithm will allow us to identify whether the poor performance of PERSIANN over these regions should be attributed to the algorithm or the inherent dependence between IR and precipitation. In order to compare the accuracy of PERSIANN estimates (as measured by correlation coefficient) and the inherent dependence (as measured by MIC), we need to scale the values of correlation and MIC. Let ρ denotes the correlation coefficient between PERSIANN and Stage IV, then the scaled correlation coefficient $\bar{\rho}$ is given by:

$$\bar{\rho} = \frac{\rho - \min\{\rho\}}{\max\{\rho\} - \min\{\rho\}} \tag{5.5}$$

Where $\min\{\rho\}$ and $\max\{\rho\}$ are the minimum and maximum values of correlation, respectively, over all spatial grids within the CONUS domain. Similarly, the scaled values of MIC can be obtained using the following equation:

$$\overline{MIC} = \frac{MIC - \min\{MIC\}}{\max\{MIC\} - \min\{MIC\}} \tag{5.6}$$

Where $\min\{MIC\}$ and $\max\{MIC\}$ are the minimum and maximum values of MIC, respectively, over all spatial grids within the CONUS domain. The values obtained from equations 5.5 and 5.6 fall within the range of 0 to 1 which allows us to account for the difference in units between the two measures of ρ and MIC.

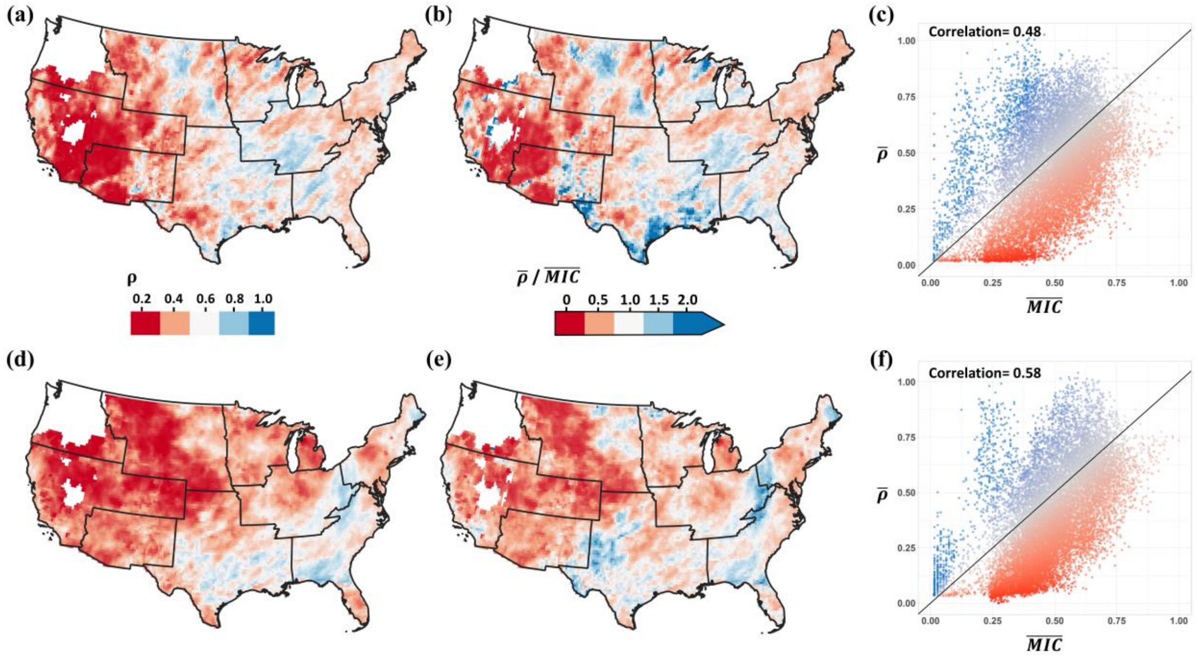


Figure 5.10 (a) *Pearson correlation coefficient (ρ) between hourly precipitaiton observations of PERSIANN and Stage IV during the months of March, April and May (MAM) of the year 2010.* ***(b)*** *The value of $\bar{\rho}/\overline{MIC}$ for hourly precipitation during the 2010 MAM season.* ***(c)*** *Scatterplot for the values of $\overline{MIC}$ and $\bar{\rho}$ for all spatial grids (0.25° × 0.25°) within the CONUS domain; the color coding corresponds to that of the map in (b). Figures* ***(d)***, ***(e)*** *and* ***(f)*** *in the lower panel are the same as the ones in the upper panel, but they correspond to the months of December, January and February (DJF) of the year 2010.*

Figures 5.10c and 5.10f show scatterplots of all spatial grids within the study spatial domain; the horizontal axis shows the value of $\overline{MIC}$, whereas the vertical axis shows the value of $\bar{\rho}$. By assuming a linear relationship between $\overline{MIC}$ and $\bar{\rho}$, all the points that fall above the one-to-one line (blue points) indicate that the accuracy of the PERSIANN algorithm ($\bar{\rho}$) is higher than what would be expected from the inherent dependence ($\overline{MIC}$). Conversely, the points that fall below the one-to-one line (red points) indicate that the performance of the PERSIANN algorithm is not as accurate as that indicated by the inherent dependence. This latter case highlights regions where the algorithm is not efficiently utilizing the entire information content of IR imagery. Figures 5.10b and 5.10e show spatial maps of the values of $\bar{\rho}$ divided by $\overline{MIC}$ with the color coding of the maps corresponding to that of the scatterplots. Grids shown in red color in the two maps are regions where the PERSIANN algorithm underutilizes the information content provided by IR imagery. For instance, Figures 5.10b and 5.10c pinpoint that the PERSIANN algorithm underutilizes the IR information content over Arizona and Southern California in MAM season and over Northern Rockies and Plains in DJF season. Furthermore, by comparing the two maps in Figures 5.10d and 5.10e, one can infer that the poor performance of PERSIANN over western US (dark red grids in Figure 5.10d) is attributed to the inherent information content of IR imagery, because the corresponding value of $\frac{\bar{\rho}}{MIC}$ shown in Figure 5.10e over the region is close to 1. This result is in agreement with recent findings made by Nguyen et al. (2020) which highlight that IR-based precipitation estimation over western US can be improved by incorporating ancillary data in the form of precipitation climatology to support IR data.

Similar analysis to the foregoing one may provide insights into the performance of IR-based precipitation estimation algorithms. However, it should be noted that this analysis is based primarily on two assumptions. First, there is a linear relationship between the skill of the algorithm, expressed as correlation coefficient, and the value of MIC. Second, correlation coefficient and MIC adequately represents the rain estimation skill and the IR information content, respectively. Therefore, the results obtained from such analysis should be interpreted within the context of these assumptions. For instance, the second assumption does not account for other measures of rain estimation skill (e.g. root mean squared error, bias …etc.). Such measures are important to be considered in order to obtain a complete view of the performance of any precipitation estimation algorithm. Overall, the foregoing analysis can be useful in obtaining insights on the performance of any IR-based precipitation estimation algorithm; however, caution must be exercised when interpreting the results.

5.6Conclusions

This study was motivated by the following questions: (1) What are the regional patterns of dependence between IR brightness temperature and precipitation? (2) Which factors exert a first-order impact on these regional patterns? (3) How do the dependence relationships change across the hierarchy of time and space scales? The findings of our analysis can be classified into two categories: first, results that corroborate assertions and validate hypotheses previously reported in the literature; second, new observations that, to the best of our knowledge, were not reported in previous studies. Regarding the former, we show in this study that spatial patterns of dependence between IR and precipitation are mostly contingent upon the type of precipitation. Specifically, the dependence is stronger at spatial domains (e.g. parts of CONUS east of the continental divide) and time periods (e.g. the warm season) at which convective precipitation is dominant. This observation affirms the well-known, strong association between IR and convective precipitation. Also, our results corroborate the numerous statements reported in literature on the inadequacy of IR brightness temperature for estimating warm, orographic rainfall. Moreover, our results provide quantitative evidences to support the intuition that both temporal and spatial averaging lead to improved dependence between IR and precipitation. Finally, we also showed that effects of geostationary satellites viewing geometry are well accounted for in the IR dataset produced by NCEP which is the main input used by IR-based precipitation estimation algorithms.

More interestingly, the analysis presented in this study highlights several observations that, to the best of our knowledge, were not previously reported in the literature. These observations can be summarized as follows:

- **I.** The information content of Infrared brightness temperature is low for regions with higher frequency of no rain instances. However, this relationship exhibits variability regionally and seasonally with more significance during the cold season.
- **II.** Although temporal aggregation leads to stronger dependence between IR and precipitation, the relationship shows a diminishing returns behavior whereby temporal aggregation at fine time scales is much more effective than at coarse scales.
- **III.** The strong association between IR and precipitation resulting from temporal aggregation is dependent on seasons, and hence the precipitation regime. In particular, it is shown that temporal aggregation is much more effective in JJA than any other season, most likely due to the dominance of convective precipitation during the season.
- **IV.** The effect of spatial aggregation leads to strong dependence between IR and precipitation with a diminishing returns behavior. This is hypothesized to be the result of the life-cycle effect.

The results presented here have several far-reaching consequences. These observations are quite useful in guiding the development of operational algorithms that utilize IR for estimating precipitation. For instance, the first of the abovementioned observations indicates that ancillary variable such as information provided by satellite water vapor channels (5.2 μm – 8.5 μm) or passive microwave (PMW) observations might be needed specifically to support IR data in identifying no rain instances. We also note that the findings of this study have important implications for the development of multi-source precipitation datasets. The development of such datasets often requires decisions on the weighting factors of different sources (e.g. IR-based precipitation, radars, in-situ …etc.). Thus, the findings presented here can provide an objective scheme that assign higher weights to IR-based precipitation when the inherent dependence is stronger, and vice versa.

Most importantly, the analysis presented in section 5 exemplifies the potential of using information-theoretic analysis of dependence to diagnose satellite-based precipitation estimation algorithms. Specifically, such an analysis facilitates identifying which parts of the errors are attributed to the algorithm as opposed to the inherent information content of IR. Thus, the analysis in section 5 provides a methodology to diagnose IR-based precipitation estimation algorithms over distinct spatial and temporal scales. Furthermore, the analysis can seamlessly be extended to diagnose algorithms for estimating precipitation from satellite PMW observations.

Part II: Hydroclimatic Applications of Satellite-based Precipitation

Chapter 6

6 Developing Intensity-Duration-Frequency (IDF) Curves from Satellite-based Precipitation

“This chapter is extracted from Ombadi et al. (2018) with few edits incorporated for brevity and clarity”

6.1 Introduction

Engineering design of infrastructure requires information about runoff magnitudes for which the structures will be designed to withstand during their lifetime. In order to estimate these magnitudes, Intensity (Depth) Duration Frequency – IDF (DDF) - curves are the typical input to hydrological models used by hydrologists and civil engineers for design purposes. They represent a mathematical relationship between frequency, duration and intensity (depth) of rainfall events. Their accuracy is contingent upon input data quality and statistical inference methods. The concept of the IDF dates back to the efforts of Bernard (1932) and since then many studies have focused on improving statistical inference methods used in IDF development. Most notable are the studies of Hosking and Wallis (1997) of introducing L-moments estimation (see also Hosking, 1990), the use of probability weighted moments estimation (Greenwood et al., 1979; Landwehr et al., 1979), parametric formulation of IDF relationships (Koutsoyiannis et al., 1998) and implementing regionalization methods such as the Index Flood method (Dalrymple, 1960; Hosking and Wallis, 1993; Wallis, 1982). Today, atlases of IDF curves have already been developed for several developed countries; an example of such efforts is NOAA Atlas 14 developed by the National Weather Service (NWS) at National Oceanic and Atmospheric Administration (NOAA) (Bonnin et al., 2006, 2011; Perica et al., 2011, 2013a, 2013b) which succeeded NOAA Atlas 2 developed in 1973.

Despite the aforementioned methodological advancements in IDF formulation, construction of IDF curves for most countries around the world remains a major challenge. This is mainly because of the limited availability of long rainfall records with adequate spatial distribution to reflect temporal variation and spatial heterogeneity of precipitation. As has been stated earlier, the accuracy of IDF curves is dependent on both input data quality and statistical inference methods. While considerable research focus has been given to the latter, only a handful of studies examined the former. Some of these studies investigated the use of alternative sources of rainfall measurements such as radar (Eldardiry et al., 2015; Marra et al., 2017; Marra & Morin, 2015; Overeem et al., 2008), satellite-based precipitation or downscaled global climate model's simulations of precipitation (DeGaetano & Castellano, 2017). Regarding the use of satellite-based precipitation, Endreny and Imbeah (2009) utilized Tropical Rainfall Measuring Mission (TRMM) rainfall dataset in combination with rainfall data from ground gauges to construct IDF curves over Ghana. Similarly, Awadallah et al. (2011) investigated the use of TRMM and ground-based rainfall data to develop IDF curves over a region in Northwestern Angola. Recently, Gado et al. (2017) used the PERSIANN-CDR dataset to develop IDF curves in ungauged sites by combining ground gauge data from neighboring sites in two basins in Colorado and California. Meanwhile, Marra et al. (2017) used CMORPH data to develop IDF curves over the eastern Mediterranean and compared them to IDF derived from radar data. Overall, these studies highlighted the benefit of using satellite-retrieved precipitation as an alternative source, particularly in partially gauged sites.

However, several reasons limit the adequacy of these studies and the extension of their application to other regions. Firstly, the methods used in most of these studies strongly rely on the partial availability of rainfall datasets with sufficient record length from ground gauges in the site of interest or in their neighboring sites which is not satisfied in many regions. Secondly, they approached the use of satellite-based precipitation in IDF development from a case-study perspective and focused on small scale regions, therefore, it is uncertain whether these methods will provide adequate results in regions with different climatic and precipitation regimes. Finally, and most importantly, the results of these studies were either not evaluated or evaluated in a small-scale domain. In other words, it is unknown whether these results underestimate or overestimate IDF curves that would ideally be derived from a dense network of rainfall gauges data.

In light of the aforementioned issues associated with previous studies on the use of satellite-based precipitation to develop IDF curves, this study provides a methodological framework for developing IDF curves from satellite-based precipitation. This is achieved by, firstly, considering and analyzing the systematic error component (i.e. bias) in extreme satellite-retrieved precipitation, secondly, considering the necessary transformation of satellite-based precipitation from an area averaged to point rainfall and finally, the application of commonly used regionalization methods to derive IDF curves. The area-to-point transformation implied in this framework is based on previous research studies that focused on the reverse transformation (i.e. point-to-area). The framework is then applied to develop IDF curves of durations 1, 2 and 3 days over the contiguous United States (CONUS). While this research was motivated by the potential of using satellite-based precipitation to develop IDF curves in data-scarce regions, CONUS has been chosen as a testbed because of the availability of rigorous IDF estimates from ground gauges provided by NOAA Atlas 14. Therefore, IDF curves derived from satellite-based precipitation are evaluated and compared to NOAA Atlas 14 to assess the performance of the framework.

6.2 Data and Case Study

a. PERSIANN-CDR

The satellite-based precipitation dataset used in this study to derive IDF curves is Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR) (Ashouri et al., 2015). This dataset has daily temporal resolution, spatial resolution of (0.25° x 0.25°) and near global coverage (60° S - 60° N) for the period (1983 – delayed present). PERSIANN-CDR is based on infrared (IR) imagery from geostationary satellites and it is a unique dataset because of its relatively long record (1983 – delayed present) compared to other satellite-based precipitation products. Other satellite-based precipitation products that can potentially be utilized for the development of IDF curves include the Climate Prediction Center morphing method (CMORPH) (Joyce et al., 2004) and Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) (Huffman et al., 2007). Both products have high temporal resolution of 3 hours and a spatial resolution of (0.25° x 0.25°). CMORPH record covers the period (2002 – present) meanwhile TMPA is available for the period (1998 – 2015). In this study, we opted to select PERSIANN-CDR primarily because of its relatively long record.

Although in this study IDF curves have been solely derived from PERSIANN-CDR, two secondary datasets were used. Firstly, CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS, hereafter referred to as CPC, has been used to estimate parameters of the bias adjustment model. Secondly, the NOAA Atlas 14 dataset has been used as a basis for the evaluation of IDF curves derived from satellite-based precipitation.

b. CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS

CPC dataset was developed by NOAA's Climate Prediction Center. It covers the period (1948 – Present) and has a similar spatial resolution to PERSIANN-CDR (0.25° x 0.25°). However, in this study, only CPC record in the period (1983 – Present) (i.e. same time coverage as PERSIANN-CDR) has been used. CPC dataset was produced from a dense gauge network over the CONUS with approximately 8500 stations and a mean station-to-station distance of ~30km (Chen et al., 2008). The interpolation algorithm used to develop the products is the Optimal Interpolation (OI) method (Gandin, 1965); this method proved to be reliable and provides results with high correlation in several studies (Chen et al., 2008; see also Bussières & Hogg, 1989; Creutin & Obled, 1982).

c. NOAA Atlas 14

NOAA Atlas 14 dataset was developed by NOAA's National Weather Service (NWS) (Bonnin et al., 2006, 2011; Perica et al., 2011, 2013a, 2013b) and it is not yet available for the states of Texas, Oregon, Washington, Idaho, Montana and Wyoming. NOAA Atlas 14 over CONUS is divided into 5 geographic regions as shown in Figure 6.1; these geographic regions have been adopted in this study for evaluation purposes since they represent, to some extent, regions with distinct climatic and precipitation regimes. NOAA Atlas 14 is derived from a dense network of rainfall gauges with an average record length range of (54 – 68) years (Bonnin et al., 2006, 2011; Perica et al., 2011, 2013a, 2013b).

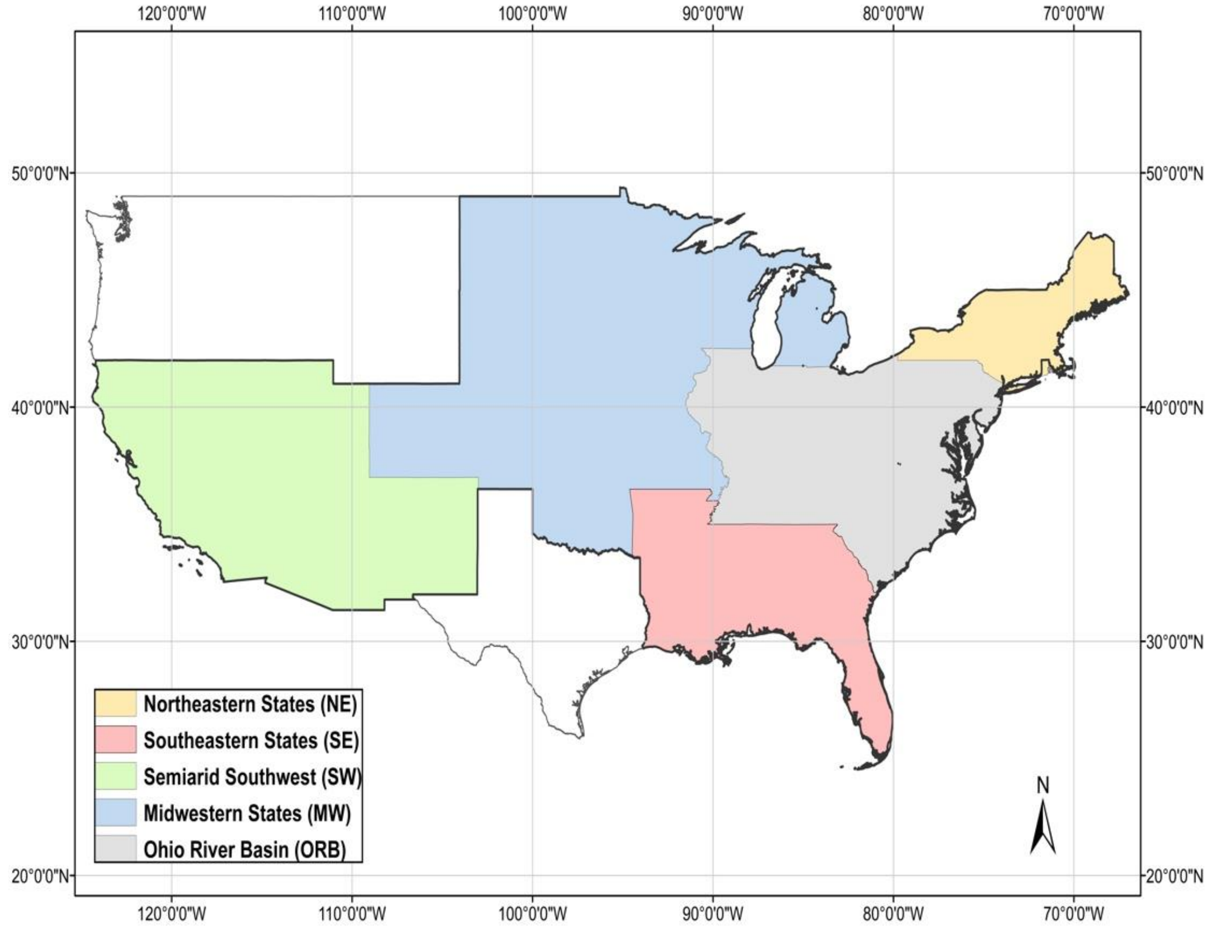


Figure 6.1 *Geographic regions of CONUS according to NOAA-Atlas 14 volumes (Volume 1 and 6: Semiarid Southwest and California, Volume 2: Ohio River Basin and Surrounding States, Volume 8: Midwestern States, Volume 9: Southeastern States and Volume 10: Northeastern States). Updating IDF curves for Texas is in progress while updated IDF curves are unavailable for the states of Washington, Oregon, Idaho, Montana and Wyoming.*

6.3 Methodology

a. Bias in Satellite-based Extreme Precipitation

In recent decades, a multitude of studies have been devoted to the evaluation of satellite-retrieved precipitation (e.g. AghaKouchak et al., 2011; Behrangi et al., 2011; Dinko et al., 2008; Ebert et al., 2007; Sorooshian et al., 2000). While these evaluation studies differ from each other in many aspects, such as geographic location over which the evaluation is performed, temporal scale (e.g. daily, monthly) and evaluation metrics, the consensus is that satellite-based precipitation exhibits errors, both random and systematic. Moreover, satellite-based precipitation products have lower skill in detecting heavy rainfall (Mehran & AghaKouchak, 2014). Therefore, it is necessary to examine errors in satellite-based precipitation prior to their use in IDF development.

As far as IDF studies are concerned, only extreme rainfall events, defined as events higher than the 99th percentile of the distribution of rainfall totals accumulated over a specific duration, are of importance. This is because both approaches commonly used to sample extreme events, namely Annual Maximum Series (AMS) and Partial Duration Series (PDS), contain rainfall values that are typically higher than the 99th percentile. Hence, in this study, analysis of errors in PERSIANN-CDR is carried out as follows. Firstly, AMS is extracted from both ground-based precipitation (CPC) and satellite-based precipitation (PERSIANN-CDR) datasets for each grid (0.25° x 0.25°); the AMS length is 33 years and it is extracted from the period of hydrological years (1984 – 2016). Secondly, an adjustment factor (ζ) is defined as the ratio of ground-based (CPC) to satellite-based precipitation (PERSIANN-CDR), that is:

$$\zeta_{(x,y,k)} = \frac{R_{G\ (x,y,k)}}{R_{S\ (x,y,k)}} \tag{6.1}$$

Where: $\zeta_{(x,y,k)}$ is the adjustment factor for the k^th event in the AMS at location (x,y), $R_{G\ (x,y,k)}$ is the k^th ground-based rainfall event in the AMS at location (x,y) and $R_{S\ (x,y,k)}$ is the k^th Satellite-based rainfall event in the AMS at location (x,y). Next, at each grid location the average adjustment factor $\bar{\zeta}_{(x,y)}$ of values in equation 6.1 is calculated; this factor represents the systematic error (i.e. bias) in extreme satellite-based precipitation.

Figure 6.2 shows the relationship between elevation and $\bar{\zeta}_{(x,y)}$. It can be clearly seen that the bias is significantly correlated with elevation as indicated by Pearson's correlation coefficient value of 0.54. This indicates that 29% (0.54^2) of the variability in the bias can be explained linearly by elevation. Hence, it can be concluded that, in general, satellite-based precipitation (PERSAINN-CDR) tends to have higher bias, particularly underestimation bias, in high altitude regions. Presence of this relationship in PERSIANN-CDR as well as other satellite-based precipitation products has been observed in previous studies (e.g. Miao, 2015; Hashemi, 2017). This is due to the fact that warm orographic rainfall over high altitude regions poses a challenge to satellite-based precipitation retrieval algorithms based on IR imagery (Dinko et al., 2008).

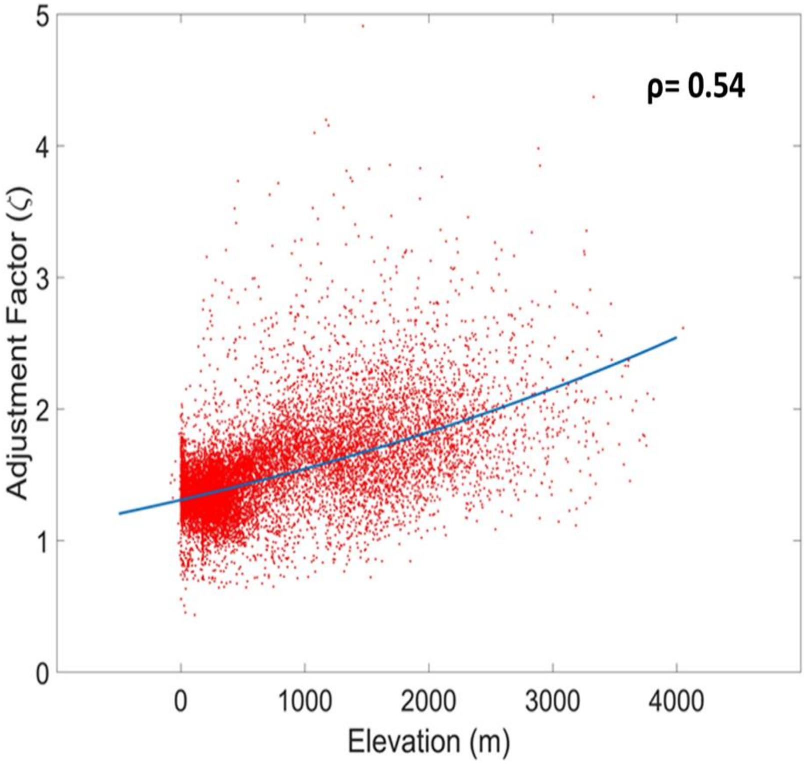


Figure 6.2 *Relationship between Elevation (meters) and the adjustment factor defined in equation (6.1) for annual maximum series of 1 day. Red dots represent observations while the blue line represent the adjustment model calculated from equation (6.2) with the values of parameters given in Table 6.1.*

b. Bias Adjustment Model

Based on the previous analysis, the following model is proposed as a new approach to adjust extreme satellite-retrieved precipitation; the model utilizes elevation as the only explanatory variable.

$$\bar{\zeta}_{(x,y)} = \alpha * e^{\beta E_{(x,y)}} \tag{6.2}$$

Where E is elevation in meters; α and β are parameters; and $\bar{\zeta}_{(x,y)}$ is defined as before.

Figure 6.2 (blue curve) shows the estimated adjustment factors based on the model for 1-day annual maximum series. Estimation of the model parameters can be carried out in a simple manner by recognizing that the model can be solved analytically by linearization. This can be performed by taking the natural logarithm of both sides in equation 6.2, then solving for the values of the parameters $ln\ (\alpha)$ and β using ordinary least squares solution. It should be noted that the parameters α and β are estimated for each duration of interest (i.e. 1-day, 2-days and 3-days) separately. Table 6.1 lists the values of parameters and their 95% confidence intervals for each duration of interest.

Table 6.1 *Values of the bias adjustment model parameters for annual maximum series of durations 1, 2 and 3 days. Values in blue are the 95% confidence intervals for the parameters estimates.*

	1 day	**2 days**	**3 days**
$\boldsymbol{\alpha}$	1.308 (1.3019 – 1.3142)	1.2557 (1.2497 – 1.2616)	1.2180 (1.2123 – 1.2238)
$\boldsymbol{\beta}$	1.6627×10^{-4} $(1.6186 – 1.7068) \times 10^{-4}$	1.7978×10^{-4} $(1.7531 – 1.8424) \times 10^{-4}$	1.7865×10^{-4} $(1.7418 – 1.8313) \times 10^{-4}$

c. *Transformation of Areal Rainfall to Point Rainfall*

An important issue to be considered when developing IDF curves from satellite-based precipitation, is the areal nature of the data, since all products of satellite-based precipitation estimate the average precipitation depth over a grid area, which is in the case of PERSIANN-CDR is (25 km x 25 km = 625 km^2). Areal rainfall distribution has both a lower mean and variance compared to the distribution of point rainfall; this follows directly from the fact that the former is an averaged random process of the latter. It is widely stated in the literature that in general the difference between the areal and point distributions increases with decrease in the total rainfall depth (Eagleson, 1970; Rodriguez-Iturbe & Mejia, 1974a). This is mainly because events that produce low amounts of rainfall tend to be more localized. This relationship is significantly present in satellite-based precipitation as shown in Figure 6.3. It can be clearly seen that high quantiles of rainfall depths correspond to low quantiles of bias (i.e. systematic error) with a Pearson's correlation coefficient value of - 0.38.

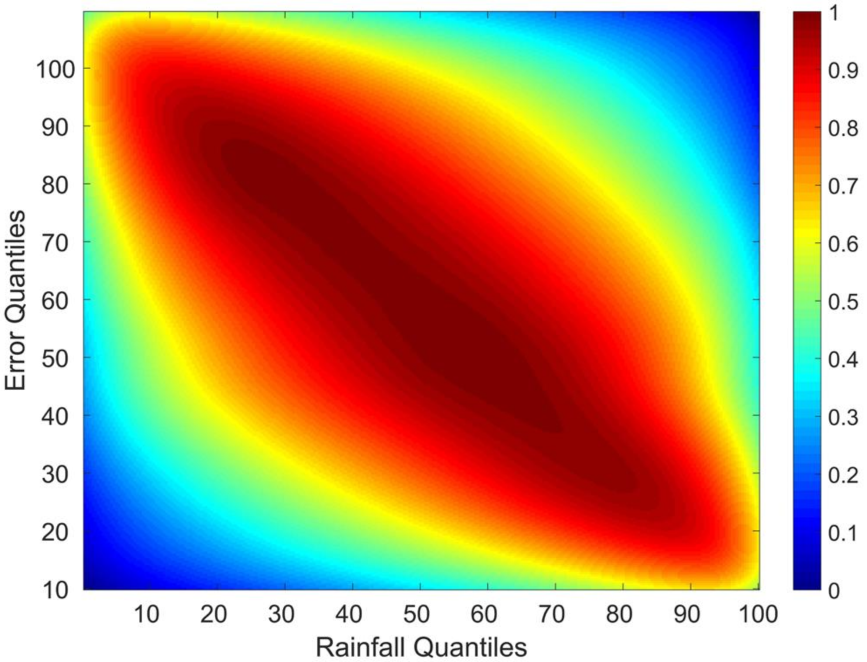


Figure 6.3 *Density plot for the joint distribution of 1-day annual maximum series rainfall quantiles and error quantiles; quantiles are calculated using the Weibull plotting position. Data used to plot the density include all grids over the CONUS.*

Considerable research attention has been assigned to the development of methods that transform point IDF curves to areal IDF curves; such a transformation requires reduction factors commonly known as Areal Reduction Factors (ARFs). Methods of developing ARFs fall into two categories: firstly, empirical methods which utilize rainfall time series data from gauge network in a specific region to develop relationships between point and area-averaged rainfall (e.g. U.S. Weather Bureau, 1957; 1958), secondly, theory-based methods which are based on the stochastic representation of rainfall fields in space and time. In this study, we adopt a theory-based approach to derive ARFs which was proposed by Sivapalan and Bloschl (1998) and is based on the spatial correlation structure of the rainfall field. It should be noted that contrary to the common use of ARFs, we are interested in transforming areal to point rainfall; thus, we will use the reciprocals of ARFs. The methodology consists of firstly assuming an isotropic correlogram (i.e. spatial correlation structure) of point rainfall of the following exponential form:

$$\rho(r) = \exp\,(-r/\lambda) \tag{6.3}$$

Where ρ is correlation, r is the Euclidean distance between two points and λ is a parameter that specifies the decay in correlation. To estimate the parameter λ, equation 6.3 has to be fitted to preserve the mean observed correlation at a distance known as the characteristic distance $\bar{\bar{r}}_A$ (Rodriguez-Iturbe & Mejia, 1974a, 1974b); this distance is a function of the shape and size of the area under consideration. The characteristic distance ($\bar{\bar{r}}_A$) is defined as the mean distance between two randomly chosen points in the region of interest and its distribution was provided by Ghosh (1951). Matern (1960) used the distribution to compute the ratio of the characteristic to the maximum distances for unit areas with standard shapes (e.g. square, circle …etc.). The following result was found for a square unit area (A):

`

$$\bar{\bar{r}}_A = 0.7374 * \text{diagonal (A)} \tag{6.4}$$

Applying this result on the grids of PERSIANN-CDR (25 km x 25km) will result in a characteristic distance of 26.07 km. However, because the distances between the grids centers for which equation (6.3) can be computed can only take multiples of 25 km (i.e. 25, 50, 75 …etc.), we have taken $\bar{\bar{r}}_A$ to be 25 km. Then, the average observed cross-correlation between the annual maximum series at distances of 25 km was calculated for each of the geographic sections shown in Figure 6.1. Finally, equation 6.3 is fitted to the values of observed correlations to estimate the value of parameter λ. Additionally, in order to evaluate the bias resulted from assigning a value of 25 km instead of 26.07 km to the characteristic distance, the sensitivity of the parameter λ to changes in the characteristic distance have been investigated. The results (not shown here) demonstrate that the sensitivity is different in each geographic region depending on the precipitation mechanism. However, the average sensitivity is in the order of 7.5% for a change of 25 km in the characteristic distance and it increases consistently with more significant changes in the characteristic distance. Therefore, the bias in the parameter λ resulted from assigning a value of 25 km instead of 26.07 km is on average significantly less than 7.5%.

After estimating the parameter λ, the variance reduction factor κ^2, defined as the expectation of the correlation between any two random points within the region under consideration, can be calculated according to the following equation (Rodriguez-Iturbe & Mejía, 1974a; Sivapalan & Bloschl, 1998):

$$\kappa^2 = E[\rho\ (x_1 - x_2)] \tag{6.5}$$

Furthermore, Rodriguez-Iturbe and Mejía (1974a) showed that equation 6.5 can be simplified by integrating the product of the probability density function of variable r and the correlation function according to the following equation:

$$\kappa^2 = \int_0^{diag(A)} \rho\,(r)\;\; f_R(r).\; dr \tag{6.6}$$

Where ρ and r is defined as above and $f_R(r)$ is the probability density function of the random variable *r*. For a square area with side length *a* (e.g. in the case of PERSIANN-CDR, *a* = 25 km), Ghosh (1951) has derived the distribution of r (i.e. $f_R(r)$).

$$f_R(r) = \begin{cases} \dfrac{4r}{a^4}\left(\dfrac{1}{2}\pi a^2 - 2ar + \dfrac{1}{2}r^2\right), & 0 \le r \le a \\ \dfrac{4r}{a^4}\left\{\sin^{-1}\dfrac{a}{r} - a^2\cos^{-1}\dfrac{a}{r} + 2a\sqrt{r^2 - a^2} - \dfrac{1}{2}(r^2 + 2a^2)\right\}, & a \le r \le \sqrt{2a} \end{cases} \tag{6.7}$$

The final step is to use the variance reduction factor estimated from equation 6.6 to adjust the parameters of the GEV probability distribution that will be fitted to the data according to equations 6.8 and 6.9. These equations have been derived by Sivapalan and Blöschl (1998) by matching the parameters of areal and point extreme rainfall distributions in the particular case of zero area. See (Sivapalan and Blöschl, 1998) for detailed derivation of equations 6.8 and 6.9.

$$\mu_p = \frac{\mu_A}{\kappa^2\,(0.39 + 0.61\kappa^{-1.6})} \tag{6.8}$$

$$\alpha_p = \frac{\alpha_A\,(1 - 0.17\ln(\kappa^{-2}))}{\kappa^2} \tag{6.9}$$

Where μ_p and μ_A are the point and areal GEV distribution location parameters respectively, similarly α_p and α_A are the point and areal scale parameters respectively.

This theory-based approach to derive area-to-point transformation factors has been validated in Sivapalan and Blöschl (1998). The validation was performed by comparing the ARFs derived by this method to ARFs observed in actual storms. In this study, we investigated the validity of the methodology by examining an extreme rainfall event over Texas on August 27, 2017 associated with hurricane Harvey. Total 24-hr rainfall was obtained from NCEP Stage IV multi-sensor (i.e. radar and gauges) precipitation data, then the observed ARFs were calculated (red line with markers in Figure 6.4). Next, using 1-day IDF estimates for that region reported in Cleveland et al. (2015), the observed ARFs were matched through the selection of appropriate correlation length λ. Figure 6.4 demonstrates that the two enveloping curves for the observed ARFs correspond to correlation lengths of 120 km and 160 km. Since the correlation length reflects information about the rainfall generating mechanism, these large values of correlation length are consistent with the large synoptic scale event that produced this storm. It should be noted that this is an approximate validation since the observed ARFs are storm-centered (i.e. specific to storm) meanwhile the simulated ARFs are fixed-area ARFs.

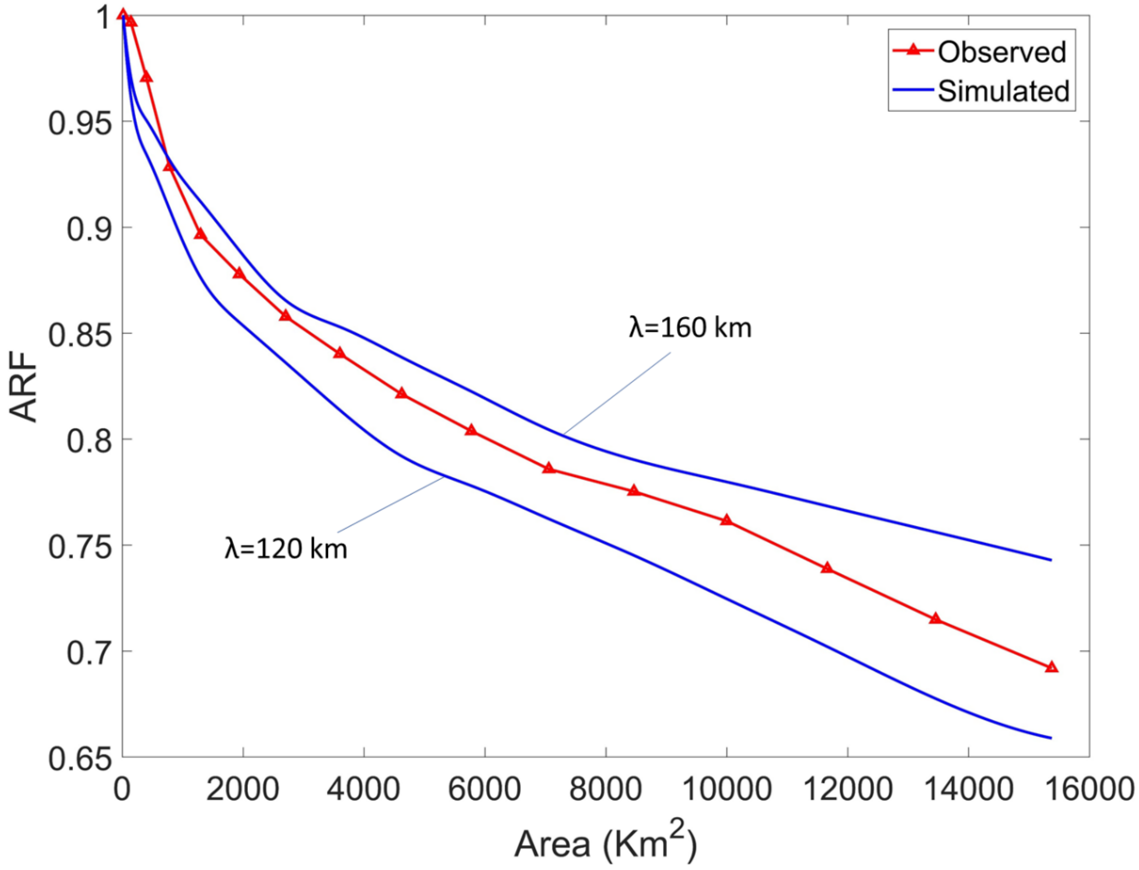


Figure 6.4 *Observed Area Reduction Factors (ARFs) (Red line with markers) and estimated ARFs by the proposed methodology (Blue lines) for an extreme rainfall event on August 27, 2018 over Texas associated with hurricane Harvey.*

d. *Developing IDF Curves*

After adjusting the bias in the annual maximum series extracted from PERSIANN-CDR using the model described earlier, the process of developing IDF curves is carried out in several steps illustrated in Figure 6.5. Firstly, regionalization is applied to improve the statistical inference by increasing the number of samples. This is achieved by creating homogenous regions using the k-means algorithm to cluster grids. This step starts with input data to the algorithm that constitute latitude, longitude, elevation and mean annual precipitation; these data to a certain extent define different climatic divisions. Next, the output clusters from the k-means algorithm are tested statistically for homogeneity using the method described in Hosking and Wallis (1993). In this method the within-cluster variation in L-CV (i.e. the ratio of second to first L-moments) is compared with what would be expected by simulations from a general probability distribution; in this study the Wakeby distribution (Houghton, 1978; see also Hosking & Wallis, 1997) was used. If clusters are not satisfactory according to the homogeneity check, clustering is repeated with increasing the number of groups. It should be noted that clustering might be different for each duration of interest (e.g. 1 day, 2 days …etc.) since it depends on L-CV values of each AMS.

Following the identification of homogenous regions, the AMS at each grid is normalized by dividing it by the mean AMS value. Then, the AMSs in each homogenous region are combined and fitted to a Generalized Extreme Value (GEV) distribution. The choice of the GEV distribution to model the extreme rainfall process was validated using the Kolmogorov–Smirnov test (Massey, 1951); results showed that GEV is an adequate distribution to represent the annual maximum series. The location and scale parameters of the distribution are then adjusted to account for the transformation of areal to point rainfall using the approach described in section 3.3.

Finally, precipitation quantiles corresponding to return periods (2, 5, 10, 25, 50 and 100) years are calculated using the index flood procedure (Hosking & Wallis, 1997). In this approach, the quantiles for each homogenous region, also known as the regional growth factors, are estimated. Next, to account for normalization, the quantiles in each grid cell are calculated by multiplying the mean AMS value at the cell by the growth factor according to the following equation:

$$q_{(x,y)} = \mu_{(x,y)} * \hat{q} \tag{6.10}$$

Where $q_{(x,y)}$ is the quantile at grid (x,y), $\mu_{(x,y)}$ is the mean of AMS at grid (x,y) and $\hat{q}$ is the regional growth factor for the homogenous region of interest.

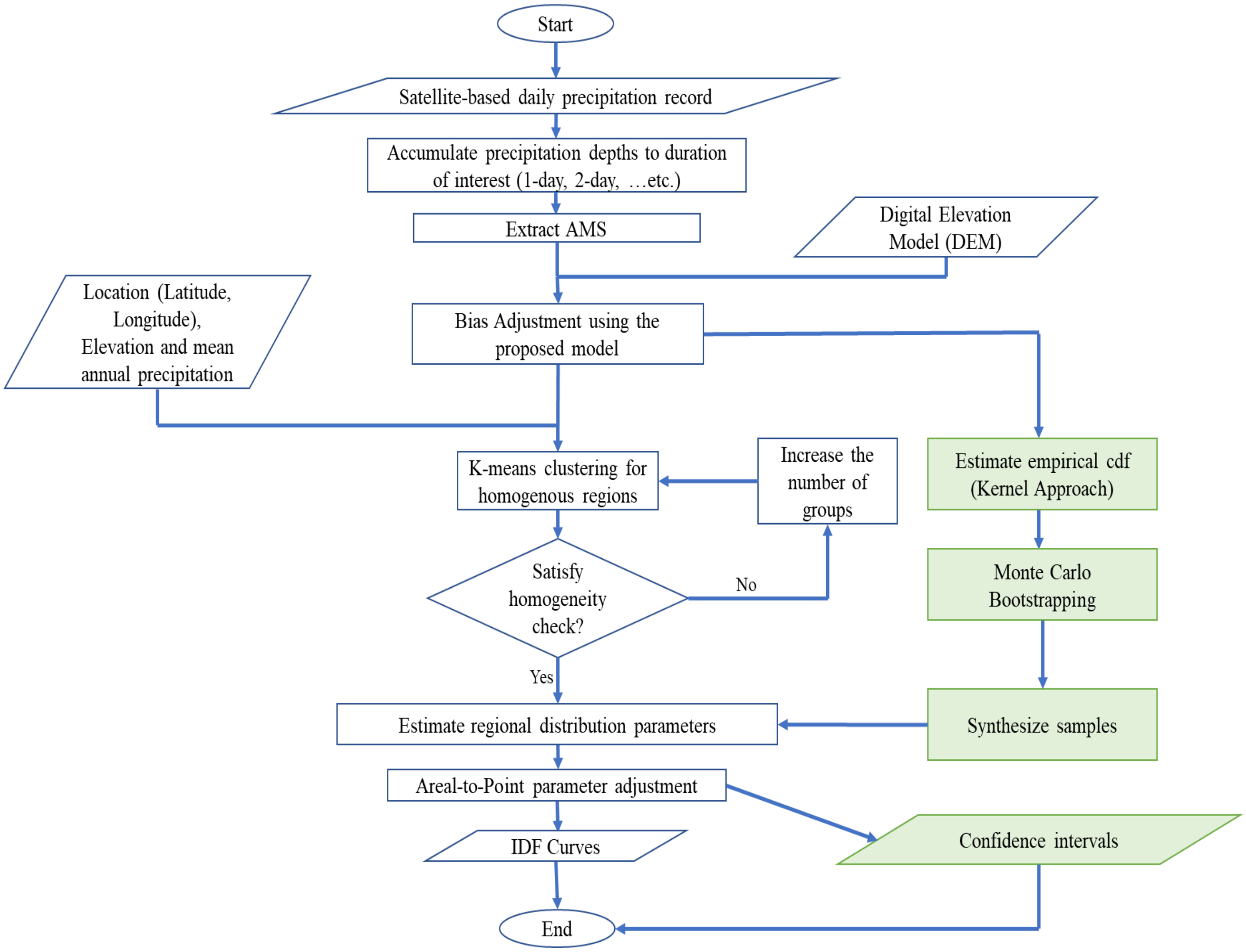


Figure 6.5 *Flowchart illustrates the process of developing intensity-duration-frequency (IDF) curves from satellite-based precipitation. Processes in green illustrate generally the process of estimating confidence intervals.*

e. Estimation of Confidence Intervals

Confidence intervals are estimated using Monte Carlo bootstrapping, the method consists of three steps. Firstly, the at-site empirical cumulative distribution function (cdf) is estimated at each grid cell center using Kernel density estimation (Parzen, 1962; Rosenblatt, 1956). Secondly, samples of AMS are extracted from the empirical distribution with the same length of record as the original AMS. The sampling is performed by drawing a uniform random variable in the range

(0,1), then the empirical cdf is used to estimate the corresponding quantile. It should be noted that Monte Carlo sampling is implemented 1000 times to approximate the asymptotic properties of the population distribution. In the final step, the quantiles are estimated using the method described in section 3.4, then the 5th and 95th percentiles are computed from the data to obtain the 90% confidence interval.

6.4 Results and Discussion

a. Bias Adjustment

Figure 6.6 illustrates the impact of the bias adjustment at a high-altitude location (a) and a low-altitude location (b). Clearly, the results suggest that: Firstly, PERSIANN-CDR before adjustment and CPC (red dots) follow an identical distribution since the quantiles lie almost perfectly on a straight line. Secondly, the bias in the case of high-altitude regions (Figure 6.6a) is more significant than the bias in low-altitude regions (Figure 6.6b). This provides further demonstration to the analysis presented earlier regarding the significant correlation between elevation and bias. Finally, the bias adjustment model removes a sizeable portion of the systematic error as can be seen from the close alignment of the quantiles after adjustment (blue dots) with the 45° line (gray dotted line). However, it should be noted that the remaining bias, illustrated by the blue dots falling below the 45° line, will be accounted for by the areal to point transformation. Furthermore, it can be discerned from Figure 6.6 that the bias adjustment results in an overestimation for the largest event in the AMS. This is primarily because the average bias adjustment factor estimated for all values in the AMS is higher than the actual bias in the highest AMS value; this result is consistent with the analysis shown in Figure 6.3.

Although the bias adjustment model presented in this study is effective in removing bias, it can be seen from Figure 6.2 that for a given elevation, there is a range of values for the adjustment factor. In other words, the elevation is not a satisfactory and/or sufficient explanatory variable in some locations. Results of further analysis (not shown here) demonstrate that the multiplicative bias in the adjusted AMS from PERSIANN-CDR is considerable over the California Central Valley, northern parts of California, Oregon and Washington. In particular, the bias over these regions is mostly an overestimation bias. This analysis highlights that while the model is effective in removing bias over most regions in CONUS, it has limitations regarding the adjustment of overestimation bias.

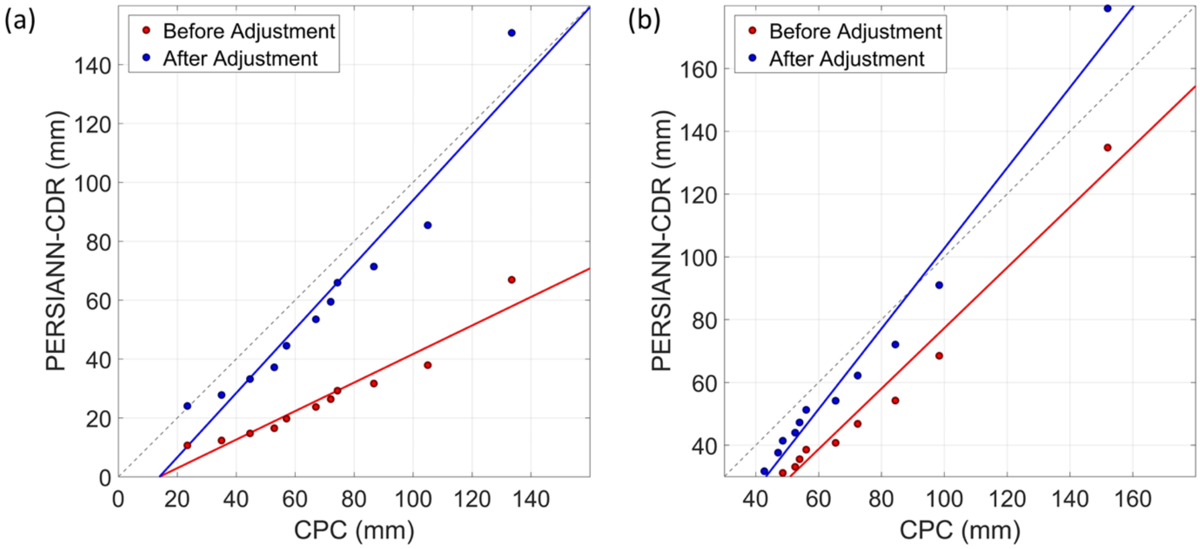


Figure 6.6 *Q-Q plot comparing the quantiles of AMS extracted from CPC (horizontal axis) and AMS extracted from PERSIANN-CDR (vertical axis) at* ***(a)*** *(37.625° N, 119.375° W), California, altitude= 3272m and* ***(b)*** *(37.625° N, 78.125° W), Virginia , altitude= 96 m. Red dots and line represent the quantiles before adjustment and its linear fit. Similarly, Blue dots and line represent the quantiles after adjustment and its linear fit. Gray dotted line represents the equality line (x=y).*

b. Areal to Point Rainfall Transformation

Figure 6.7a shows the contribution of area-to-point transformation in reducing the relative error of IDF estimates compared to that of the bias correction. Clearly, the bias adjustment is the prime factor in improving IDF estimates, however, areal-to-point transformation plays a considerable role in reducing the relative error of IDF estimates. A decreasing trend for the contribution of area-to-point transformation as the duration of IDF increases can be discerned from Figure 6.7a. Further evidence to support this conclusion is demonstrated in Figure 6.7b which shows the relationship between the transformation factor, duration and return period. The inverse relationship of the transformation factor and duration is consistent with previous studies (e.g. Mineo et al., 2018; Asquith & Famiglietti, 2000) and it is justified by rainfall behavior since short-duration events are primarily associated with small areal extent and convective rainfall, meanwhile, long-duration events are distributed over a large area (Mineo et al., 2018; Sivapalan & Blöschl, 1998). On the other hand, the transformation factors increase with return period as shown in Figure 6.7b. This relationship shows that the transformation method is not independent of return period and it is consistent with previous studies (e.g. Veneziano & Langousis, 2005). This is because the transformation is applied to both the location and scale parameters of the distribution. Sivapalan and Blöschl, (1998) showed that this transformation method results in a decrease of the coefficient of variation as the area increases unlike transformation methods that assume independence of return period resulting in a constant coefficient of variation.

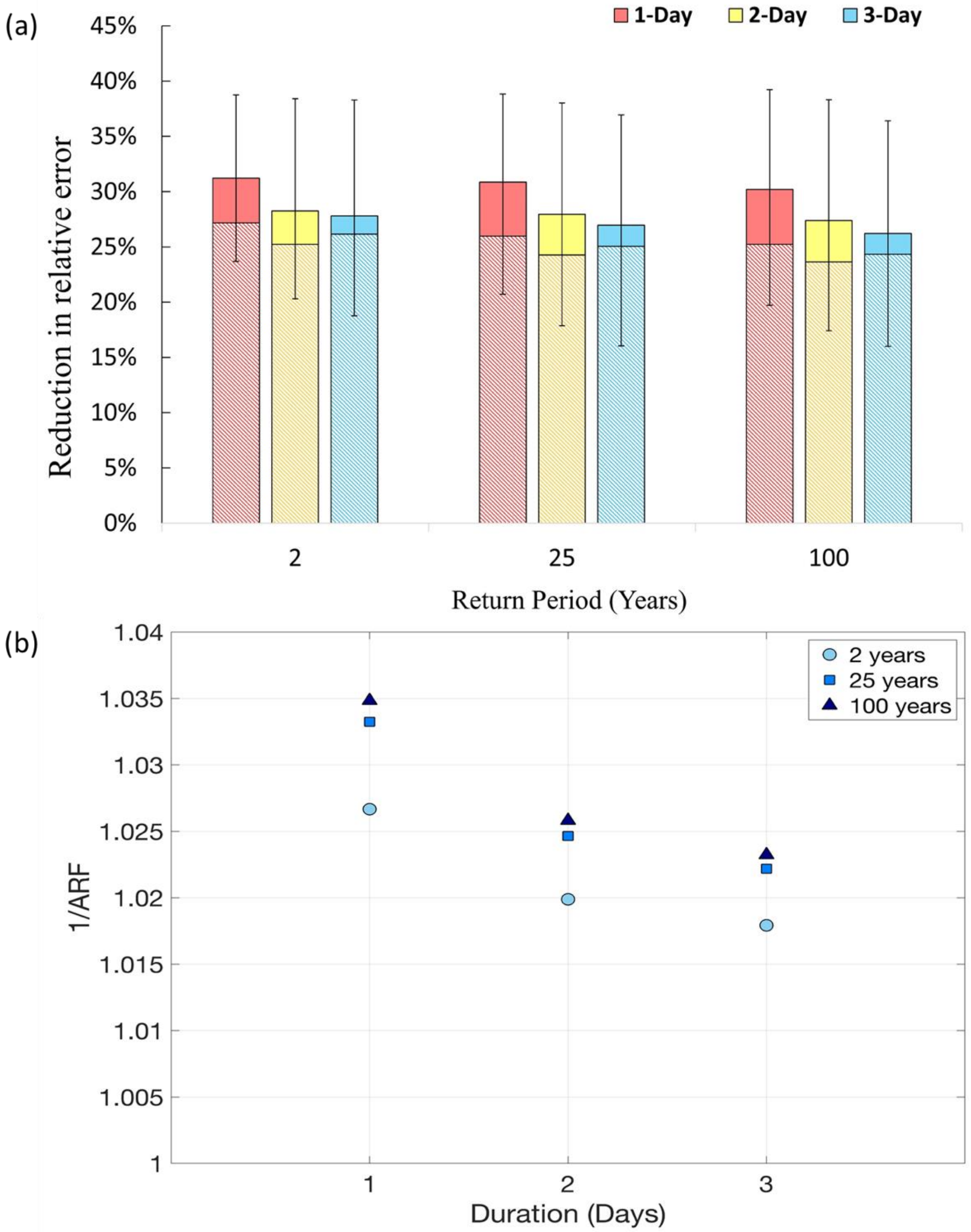


Figure 6.7 (a) *Reduction in mean relative error of IDF curves derived from PERSIANN-CDR compared to NOAA Atlas 14 for durations of 1,2 and 3 days and return periods of 2, 25 and 100 years. Solid color bars represent the contribution of area-to-point transformation and hatched bars represent the contribution of bias adjustment. Black bars represent the standard deviation of the reduction in relative error.* ***(b)****Relationship between the mean values of the inverse Area Reduction Factor (e.g. 1/ARF), durations of 1,2 and 3 days and return periods of 2, 25 and 100 years.*

c. IDF Curves Evaluation

IDF curves derived from PERSIANN-CDR are evaluated against NOAA-Atlas 14 precipitation frequency estimates. The evaluation is performed over the contiguous United States (CONUS) except the states of Washington, Oregon, Idaho, Montana, Wyoming and Texas because of unavailability of NOAA-Atlas 14 estimates in these states as shown in Figure 6.1. The evaluation is performed for IDF with durations 1, 2 and 3 days and return periods 2, 5, 10, 25, 50 and 100 years. The main metric used for evaluation of IDF estimates derived form PERSIANN-CDR is the percentage relative error which is defined as follows:

$$Relative\ Error\ (\%) = \left(\frac{IDF_{PERSIANN-CDR} - IDF_{NOAA-Atlas14}}{IDF_{NOAA-Atlas14}} \right) * 100\ \% \qquad (6.11)$$

This is an adequate performance metric since it is normalized and therefore not sensitive to the absolute values of rainfall. This allows us to examine the performance of IDF estimates over the whole spatial domain regardless of variations in climate

Figure 6.8 shows the relative error of IDF curves over the whole spatial domain of NOAA Atlas 14 (see Figure 6.1) for durations 1, 2 and 3 days and return periods 2, 5, 10, 25, 50 and 100 years. While the errors are considerable for 1-day duration with the median errors in the range (-17% – -22%) for return periods (2 – 100 years), the errors are less significant in longer durations. For example, in the case of 2 days IDF, the median errors range is (-6% – -12%) meanwhile for 3 days IDF, the median errors range is (-3% – -8%). This trend of improved performance with larger durations is due to the increased accuracy of satellite-based precipitation over long time scales as well as the temporal mismatch comparing remotely sensed and gauged rainfall over short periods.

It should also be noted that the errors are more pronounced in high return periods, and this is attributed to the relatively short record of PERSIANN-CDR (~30 years) compared to the length of record used to derive NOAA Atlas 14 which on average ranges from 54 to 68 years (Bonnin et al., 2006, 2011; Perica et al., 2011, 2013a, 2013b). Overall, IDF relationships derived from PERSIANN-CDR tend to underestimate the amount of precipitation, however, the errors are not significant in durations of 2 days and larger.

Since the previous analysis only reveals information about the aggregate performance over the whole NOAA Atlas 14 spatial domain, it is important to examine the accuracy of IDF curves over different geographic regions. Therefore, IDF curves have been evaluated separately over each of the geographic sections shown in Figure 6.1. While the average relative errors over all geographic regions are comparable and do not indicate large differences as shown in Figure 6.9a, the percentages of IDF curves that lie within the confidence interval of NOAA Atlas 14 clearly highlight that the accuracy of IDF curves derived from PERSIANN-CDR varies significantly. As can be seen from Figure 6.9b, the accuracy is higher over the Northeastern States since 77%, 86% and 84% of 1-day, 2-days and 3-days IDF curves respectively lie within the 90% confidence interval. It is followed by the Southeastern States where approximately 43%, 79% and 86% of 1-day, 2-days and 3-days IDF curves lie within the confidence interval. The poorest accuracy is observed over the Southwestern States where only 20% of 1-day IDF curves lie within the confidence interval.

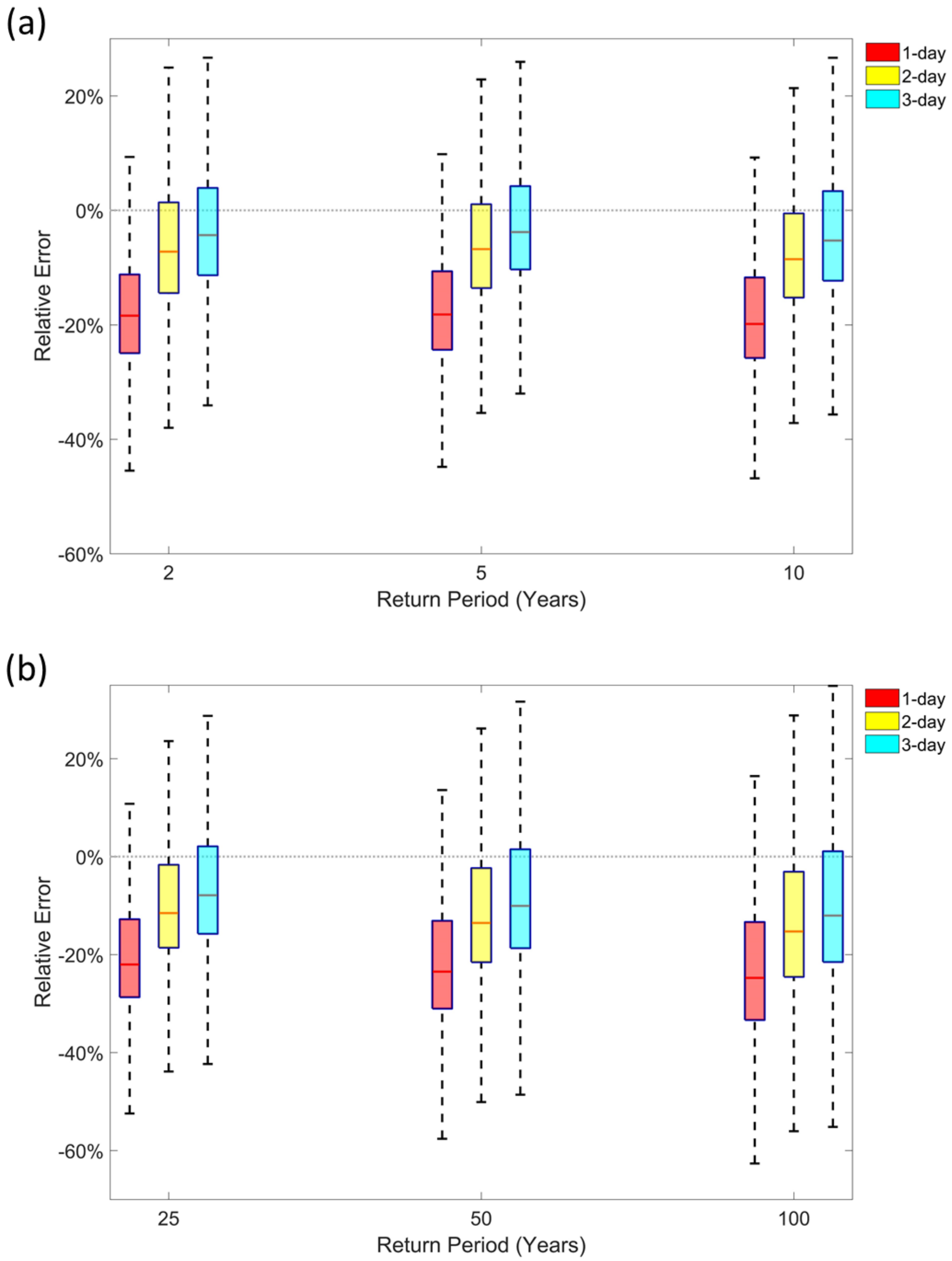


Figure 6.8 *Boxplots of satellite-based IDF relative error for durations of (1, 2 and 3) days.* ***(a)*** *Return periods of 2, 5 and 10 years.* ***(b)*** *Return periods of 25, 50 and 100 years. Thick lines inside boxes indicate the median value, boxes indicate the interquartile range and dashed lines indicate the range.*

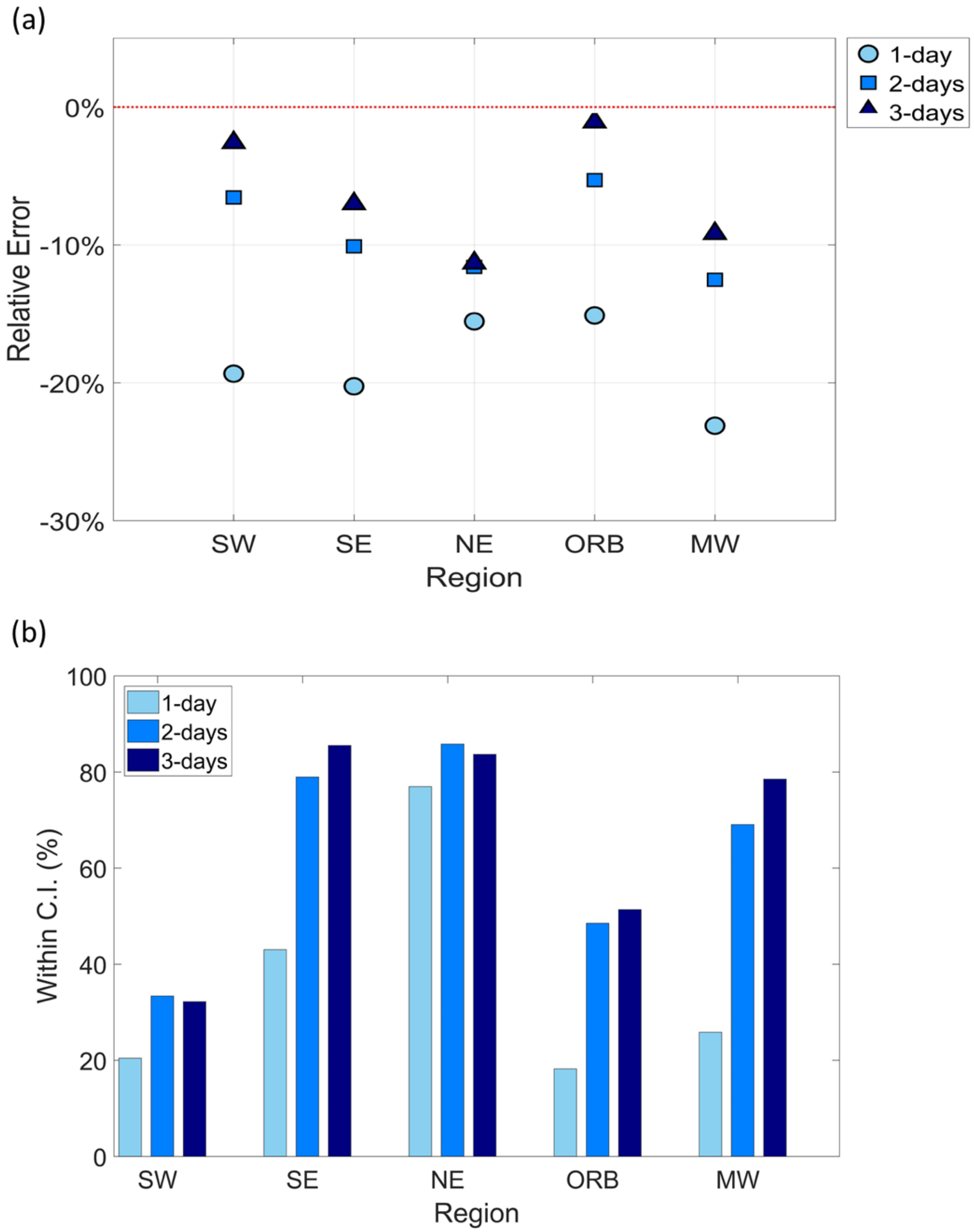


Figure 6.9 (a) *Average relative error in satellite-based IDF with return period of 25 years for the five CONUS divisions defined in Figure 6.1 (SW: Semiarid Southwest, SE: Southeastern states, NE: Northeastern states, ORB: Ohio River Basin and MW: Midwestern States).* ***(b)*** *Percentage of satellite-based IDF estimates that falls within the 90% confidence intervals, return period is 25 years.*

In order to understand the sources of observed errors in PERSIANN-CDR IDF curves, we compare IDF curves from the original PERSIANN-CDR (i.e. without adjustment and area-to-point transformation) and from the CPC record (1984 -2015); the results are shown in Figure 6.10. By comparing IDF curves derived from CPC (black dotted lines) and NOAA Atlas 14 (black lines), it can be clearly seen that IDFs from CPC exhibit underestimation errors. Since the data used to derive NOAA Atlas 14 is identical to CPC data, the observed differences are primarily due to the length of record as we have used CPC record of approximately 30 years long. This highlights that while the observed errors can potentially be attributed to several sources such as the difference in spatial scale, it is important to consider the relatively short length of record as the main source of underestimation.

An important point to be concluded from Figure 6.10 is that the bias adjustment and the area-to-point transformation are important, and they improve the results significantly. This can be clearly seen by comparing the original PERSIANN-CDR IDFs (red dotted line) and IDFs derived after adjustment and transformation (red lines). For example, in both Figures 6.10a and 6.10b, IDFs derived from adjusted PERSIANN-CDR lie within the confidence interval of NOAA Atlas 14 meanwhile IDFs before adjustment are considerably underestimated and lie out of the confidence interval. Figure 6.10c shows an example of an IDF where bias adjustment and area to point transformation improves the results yet not sufficiently as the obtained IDF (red line) lies outside the confidence interval. On the other hand, it can be seen from Figure 6.10d that IDF curves from the original PERSIANN CDR (red dotted line), that is, without adjustment and area-to-point transformation, are overestimating. Therefore, the framework used in this study to adjust the bias and account for the areal nature of satellite-retrieved precipitation exacerbates the errors leading to an increased overestimation as shown by the red line in Figure 6.10d. This highlights that while the adjustments embedded in the methodology are essential for the development of accurate IDF curves, special attention should be paid in regions where satellite-based precipitation products show peculiar performance such as the case over the Central Valley of California.

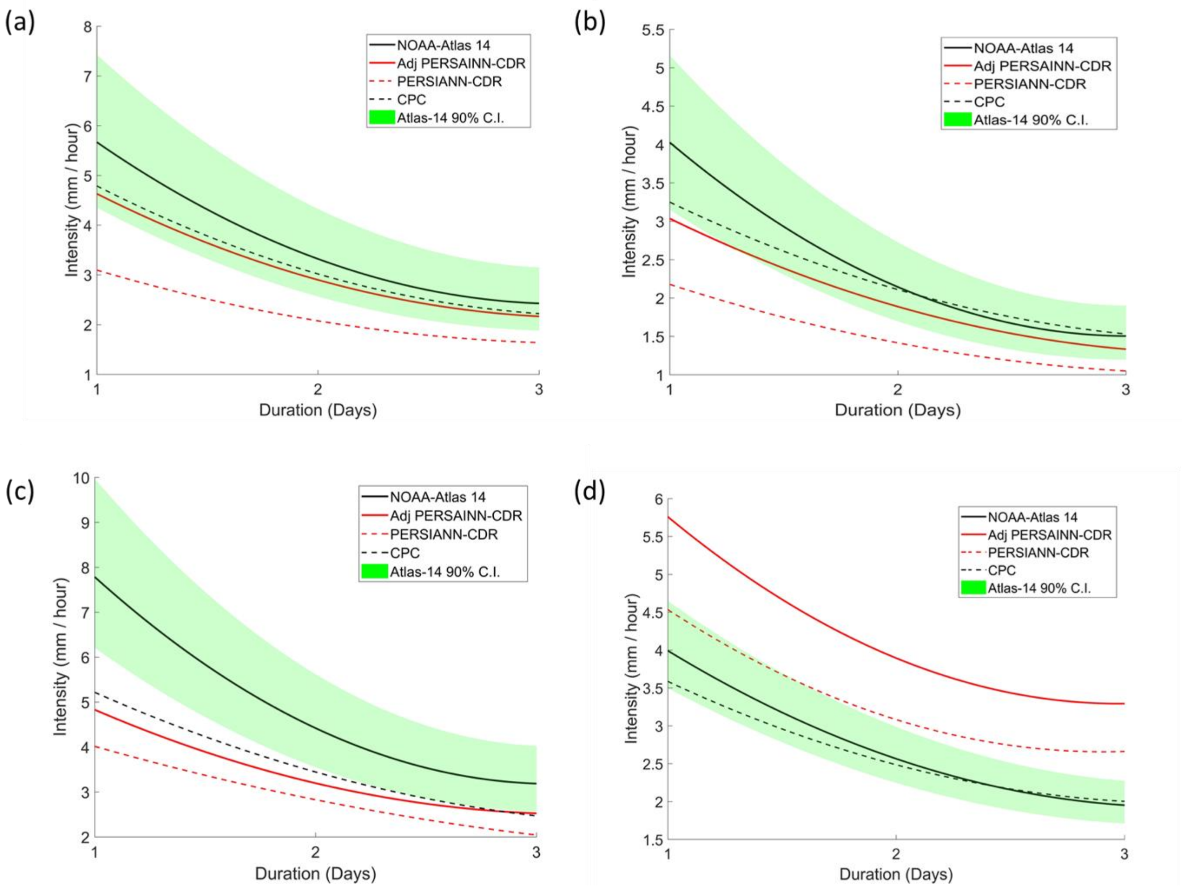


Figure 6.10 *IDF curves for durations 1, 2 and 3 days and return period of 25 years. IDF curves from original PERSIANN-CDR, adjusted PERSIANN-CDR, CPC and NOAA Atlas 14 are plotted along with confidence intervals of NOAA Atlas 14.* ***(a)*** *(42.875°N, 72.125°W), Vermont.* ***(b)*** *(46.375°N, 101.375°W), North Dakota.* ***(c)*** *(27.375°N, 80.625°W), Florida.* ***(d)*** *(39.625°N, 122.125°W), Northern California.*

d. Uncertainty and Impact of Regionalization

The issue of uncertainty in satellite-based IDF curves is more difficult than when ground measurements are used to develop these curves. This is because there are several components of uncertainty to be considered. Firstly, uncertainty arises from the estimation process since satellites do not measure precipitation directly but rather utilize other information as a proxy for rainfall rate. Secondly, there are uncertainties induced by the methodological framework proposed in this study; these include the bias adjustment model and the transformation from areal to point rainfall. Finally, the commonly considered source of uncertainty, estimation of distribution parameters.

In this section, we only discuss uncertainty that arises from the estimation of distribution parameters. Confidence intervals of IDF estimates are computed using Monte Carlo bootstrapping as described earlier. Firstly, we highlight the importance of regionalization in constraining the uncertainty to narrower limits. Figure 6.11a shows the coefficient of variation for the distribution of quantiles corresponding to 2, 5, 10, 25, 50 and 100 years for both cases of using regionalization and at-site (i.e. no regionalization) estimation. The distribution is obtained by extracting 1000 samples, then estimating the quantiles. Meanwhile, the coefficient of variation (i.e. the ratio of the standard deviation to the mean) is used to assess uncertainty since it is a normalized measure and hence allows us to examine all regions regardless of variation in their climate. It can be seen that for lower quantiles such as those corresponding to 2 and 5 years return period, the impact of regionalization is barely noticeable. However, as higher quantiles are considered, the differences in the coefficient of variation are more significant with regionalization leading to lower coefficients of variation. This indicates the importance of regionalization in reducing the uncertainty, particularly for quantiles in the tail of the distribution.

Furthermore, regionalization in the case of satellite-based precipitation is more effective in reducing uncertainty since the amount of available data is immense. To illustrate, an arbitrary homogenous area of 6250 km^2 that might be covered with 3 ground gauges with average record length of 50 years will generate (3*50 = 150 samples), while on the other hand, PERSIANN-CDR will provide (10 grids * 30 = 300 samples). The increased sample size will result in a decrease of the uncertainty range. As it can be seen from Figure 6.11b, uncertainty ranges in the case of IDFs derived from PERSIANN-CDR are smaller than those of NOAA Atlas 14. This highlights that implementing regionalization in the case of satellite-based precipitation is more effective.

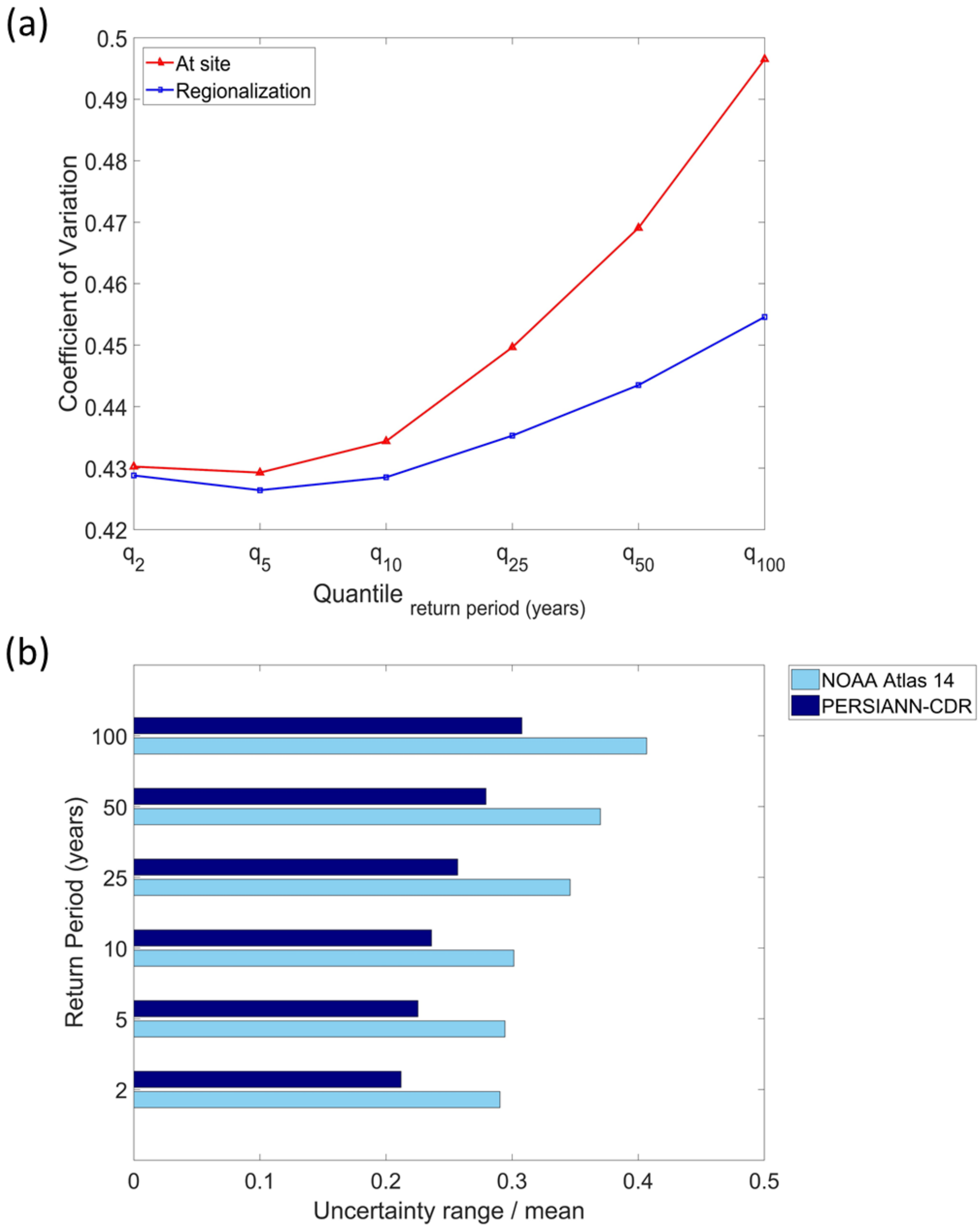


Figure 6.11 (a) *Coefficient of variation (CV) in the distribution of quantiles of 1-day IDF with return periods (2, 5, 10, 25, 50 and 100) years. Blue and red curves represent the values of CV for the case of regionalization and at-site estimation respectively.* ***(b)*** *Uncertainty ranges normalized by the mean for quantiles corresponding to (2, 5, 10, 25, 50 and 100) years. Values are averaged over the whole spatial domain of NOAA Atlas 14 shown in Figure 6.1.*

6.5 Conclusions

The goal of this study is to contribute and advocate for the development of methods that facilitate the use of satellite-retrieved precipitation in developing IDF curves. This is of particular interest to developing countries where existing networks of ground gauges do not provide sufficient spatial coverage or record length to develop IDF curves. Given the continuous advancement in remote sensing and the retrieval of precipitation from satellites, it is worthy of attention to dedicate more research efforts towards the development of methods that ensure the incorporation of satellite-based precipitation in the design, operation and planning of infrastructure. This study has attempted to examine this issue from a methodological point of view by considering and accounting for the characteristics of satellite-based precipitation. The methodology used in this study is different from previously reported studies on the use of satellite-based precipitation in the development of IDF curves which approached this issue from a case study perspective. The ultimate aim of this study is to contribute in the development of general methodologies that can provide adequate results in the absence of in-situ rainfall measurements.

While the main motivation for this research is the potential use of satellite-based precipitation to construct IDF curves for developing countries, it is important at this early stage of methodological research to examine the methods by evaluating them in regions with extensive networks of ground gauges with sufficient length of record. This has been the rationale behind the selection of CONUS as a testbed to evaluate the proposed methodology. It is important to emphasize that the methods proposed in this study are neither tailored to a specific region nor to a specific satellite-based precipitation product. Furthermore, we emphasize that estimating the adjustment model parameters is the only step in the proposed framework that requires the availability of ground-based measurements. The question then arises, “Are these parameter estimates sufficiently robust such that they can be applied in other regions?”. The answer is twofold. Firstly, it is expected that these estimates are robust over most regions since data from all grids over CONUS, which represent a variety of climatic and precipitation regimes, have been used in the estimation process. Secondly, as has been shown in this study, the model has limitations in adjusting overestimation bias over specific locations such as the Central Valley of California. However, further studies are sorely needed to explore the bias in PERSIANN-CDR as well as other satellite-based precipitation products over different regions. It is also important to note that in regions with partial coverage of ground rainfall gauges, information from ground gauges may be incorporated to validate the bias adjustment model presented in this study which will lead to improved performance. We also acknowledge that the bias in other satellite-based precipitation products does not necessarily follow the same characteristics observed in PERSIANN-CDR. For example, Endreny and Imbeah (2009) reported that bias in TRMM rainfall depths over Ghana is primarily overestimation bias meanwhile in this study, PERSIANN-CDR mainly exhibits underestimation bias. Thus, an extensive analysis of bias is required in other satellite-based products prior to their use in IDF development.

Overall, the results of this study highlight the potential of using satellite-based precipitation as an alternative source to the commonly used ground-based measurements in developing IDF curves. Through comparison with NOAA Atlas 14 estimates, which have been used as a benchmark, we found that the median relative errors in satellite-based IDFs over CONUS are in the range of (-17% - -22%), (-6% – -12%) and (-3% – -8%) for 1-day, 2-days and 3-days IDFs respectively. Furthermore, a significant percentage of satellite-based IDF curves fall within the confidence interval of NOAA Atlas 14 for most geographic sections of CONUS with the best results over the Northeastern States with 77%, 86% and 84% of 1-day, 2-days and 3-days IDFs within the confidence interval. These promising results corroborate findings reported in Gado et al. (2017) which demonstrated that the use of satellite-based data with bias adjustment from local gauges provides accurate quantile estimates. The increase in IDF error with increase in the return period can be attributed to uncertainty associated with the short length of record; this relationship is consistent with uncertainty analysis of IDF derived from remotely-sensed observations (Marra et al., 2017). We also highlight that IDFs derived from PERSIANN-CDR in this study over the Central Valley of California exhibit higher errors since the original product is overestimating in this region. This emphasizes the importance of considering any peculiar performance of satellite-based precipitation over specific regions prior to the development of IDF curves. It also pinpoints that elevation is not a satisfactory and/or sufficient explanatory variable in some locations to adjust bias in extreme satellite-based precipitation.

Finally, there are several important questions regarding the use of satellite-based precipitation in IDF development that remain unanswered and in need of further investigation. Firstly, quantifying the different sources of uncertainty in satellite-based IDFs that arise from the estimation of rainfall rates, bias adjustment, transformation of areal to point rainfall and the estimation of distribution parameters. In this study, we only dealt with the uncertainty in the estimation of distribution parameters, however, other sources of uncertainty should not be ignored. A possible approach to deal with uncertainty from the estimation process is to consider several satellite-based precipitation products in an ensemble approach which will provide uncertainty limits for the random error component. With regard to bias adjustment, it might be beneficial to estimate the parameters of the adjustment model using Bayesian regression to provide uncertainty bounds to the parameter estimates. Secondly, further research is needed to investigate the impact of the liquid/frozen precipitation partitioning since satellite-based precipitation provides estimates of the total precipitation while in the development of IDF curves usually only liquid precipitation is considered. This might only be of significance in regions that receive considerable amounts of frozen forms of precipitation (i.e. snow, ice and hail) during extreme precipitation events. Finally, as the results of this study have shown that regionalization of IDFs derived from satellite-based estimates is more effective in reducing the uncertainty in distribution parameters due to the availability of more information, it is important to develop regionalization methods that can exploit the information content of satellite-based precipitation datasets more efficiently.

Chapter 7

7 Bayesian Model Averaging of Precipitation Projections in the Nile River Basin

“This chapter is extracted from Ombadi et al. (2021c) with few edits incorporated for brevity and clarity”

7.1 Introduction

The Nile river basin constitutes approximately 10% of the African continent (Swain 2008) extending across eleven countries. A total population of 462 million in these countries is growing at an annual growth rate of 2.5%, faster than the average global growth rate estimated at 1.1%. Consequently, the population of these countries is projected to reach 836 million (81% increase) by the year 2050 (The World Bank 2018; 2020). A key challenge, therefore, that face these countries is to sustain the burgeoning food and energy demand of this growing population. Water lies at the heart of natural resources that play a pivotal role in securing this demand. Therefore, assessment of climate change impacts on precipitation is important due to its direct effect on water availability in headwaters countries as well as its impact on the Nile streamflow yield which is the main source of water for riparian countries, namely Sudan and Egypt.

Several studies have been devoted to the assessment of climate change impacts on precipitation in the Nile River basin and its headwaters basins (e.g. Conway 1996, Yates and

Strzepek 1996, 1998, Kim and Kaluarachchi 2009, Elshamy et al 2009, Bhattacharjee and Zaitchik 2015, Fenta Mekonnen and Disse 2018). Earlier studies found that general circulation models (GCMs) frequently show disagreement in the sign of change of annual precipitation projections. For instance, Conway (1996) used 3 GCMs to assess climate change impact on precipitation in the Blue Nile and Lake Victoria sub-basins; results showed that percentage change in precipitation ranges from -1.9% to 7.4% in the two sub-basins. More recently, Kim and Kaluarachchi (2009) showed that mean annual precipitation in the upper Blue Nile sub-basin is projected to increase by 11% based on a weighted average of 6 GCMs outcomes. On the contrary, Elshamy (2009) reported the outcomes of 17 GCMs and showed that projected change in mean annual precipitation in the upper Blue Nile sub-basin ranges from -15% to +14% with more models reporting a decrease in precipitation. These results, among others, emphasize that there is a wide uncertainty and inter-model differences in precipitation projections, and they indicate that a consensus on how climate change will impact water resources in the Nile basin is yet to be reached.

Two different approaches are commonly adopted to treat uncertainty of GCMs. At one end of the spectrum is the ensemble mean which overlooks historical performance of the models and assigns equal weights to all models. At the other end, there is an approach that selects a number of best performing models and discards the remaining ones. The former is less accurate at regional scales and in cases where there is a spread in model projections (Schaller et al 2011) whereas the latter is highly dependent on the specific metrics chosen for performance evaluation (Schaller et al 2011, Bhattacharjee and Zaitchik 2015). Between these two extremes lies the approach of model averaging in which models are neither weighted equally nor some of them are discarded entirely. Specifically, model averaging methods take advantage of retrospective analysis of GCMs simulations benchmarked against observations, and they assign weights to models according to their performance. A major issue, however, that lessen the effectiveness of such methods is the dearth of quality controlled, dense gauge precipitation observations in the Nile basin. Here, we surmount this issue by resorting to high spatial resolution and long record of historical observations provided from Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Climate Data Record (PERSIANN-CDR; Ashouri et al. 2015). PERSIANN-CDR is a high spatial resolution satellite-based dataset that is bias adjusted using gauge observations at the monthly scale; thus, providing a unique dataset for retrospective analysis of GCMs. To this end, the focus of the present study is to first evaluate the performance of 20 GCMs from the 6th Coupled Model Intercomparison Project (CMIP6) against PERSIANN-CDR over the Nile basin. Next, a model averaging approach, namely Bayesian Model Averaging (BMA; Raftery et al., 2005) is implemented to derive probability distributions of precipitation projections in the Nile basin for the future period (2015 – 2100).

7.2 Data and Study Area

a. CMIP6

Many climate models participating in CMIP6 have reported their simulations for the different CMIP6 experiments. In the present study, two experiments are of concern: historical and the Shared Socioeconomic Pathways (SSP) SSP5-8.5. The historical experiment provides GCMs simulations for the period (1850-2014), and it is intended to be used for assessment of model performance in simulating historical observations. Here, we only use data from the period 1983 onward to be consistent with the available record of observed precipitation from PERSIANN-

CDR. SSP5-8.5 is the future scenario that corresponds to high greenhouse gas emissions, and it is the equivalent to RCP8.5 “business as usual” scenario in CMIP5. Currently, a set of 20 models have reported their simulations for both historical and SSP5-8.5 experiments. These models have been used in this study to examine climate change impact on precipitation in the Nile basin, and they are shown in Table A.9. For each model, we only consider the first ensemble member for future projections under SSP5-8.5. Also, we consider dataset at monthly temporal resolution for both historical and SSP5-8.5.

b. PERSIANN-CDR

PERSIANN-CDR (Ashouri et al. 2015, see also Nguyen et al. 2018) is a satellite-based precipitation dataset based on infrared (IR) imagery. It has near-global coverage (60°S - 60°N) with a spatial resolution of 0.25° x 0.25° and a daily temporal resolution. PERSIANN-CDR is suitable for climatic studies because of its long record of +37 years (1983 – delayed present). It is particularly advantageous because it is bias adjusted using Global Precipitation Climatology Project (GPCP) monthly 2.5° x 2.5° precipitation data (Adler et al. 2018). Therefore, it maintains monthly precipitation at 2.5° x 2.5° that is consistent with GPCP while capturing spatial rainfall variability at higher spatial resolution. This last point emphasizes that PERSIANN-CDR has sufficient credibility to be used as a baseline dataset for evaluation of CMIP6 GCMs. PERSIANN-CDR is widely used for a range of hydrologic and hydroclimatic studies (e.g. Ombadi et al. 2018, Nguyen et al. 2020), and it has previously been used for evaluation of GCMs (Nguyen et al. 2017). Here, we use PERSIANN-CDR at monthly temporal accumulations.

c. *Study Area*

In this study, we consider the entire Nile basin for our analysis (shown in Figure 7.1; gray lines). The analysis is performed at the grid scale (1° x 1°) due to the wide variability of climate and precipitation regimes in the Nile basin. This variability is clearly shown in Figure 7.1 with the south-to-north gradient in precipitation which represents the variability in climate from tropical humid in the south to hyper arid in the north. Throughout this study, we carry out the analysis at the grid scale and then aggregate the results at the entire Nile basin as well as its headwaters basins, namely the Blue Nile and Upper White Nile basins (gray lines in Figure 7.1). We focus on these two sub-basins due to their significant contribution to the Nile streamflow yield.

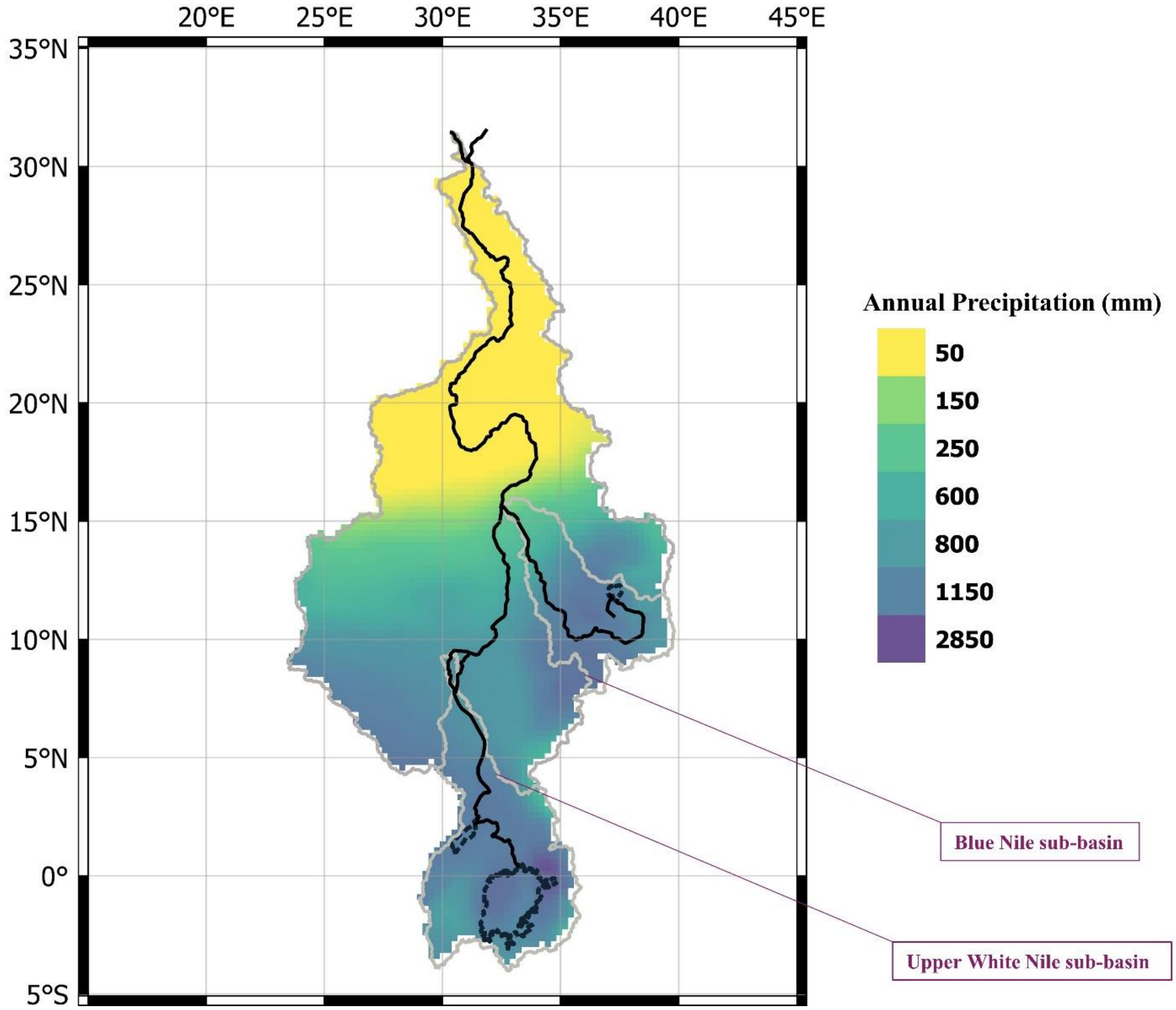


Figure 7.1 *The Nile river basin and its headwaters basins: the Blue Nile and Upper White Nile sub-basins (gray line). The Nile river and its tributaries are shown in solid balck line. Mean annual precipitation is computed from PERSIANN-CDR for the period (1983-2014).*

7.3 Methods

a. Bias Adjustment

CMIP6 model simulations and PERSIANN-CDR data were first re-gridded to a common spatial resolution of (1° x 1°) using bilinear interpolation. Bias adjustment coefficients were then calculated for each grid from the historical simulations (1983-2014) according to the following linear model:

$$\boldsymbol{y}^{\boldsymbol{H}} = a + b * \boldsymbol{f}_{\boldsymbol{k}}^{H} \tag{7.1}$$

Where $\boldsymbol{y}^{\boldsymbol{H}}$ is PERSIANN-CDR annual (or monthly) precipitation time series at a given grid for the period (1983-2014), and $\boldsymbol{f}_{\boldsymbol{k}}^{\boldsymbol{H}}$ is the corresponding annual (or monthly) precipitation from the $\boldsymbol{k}^{\text{th}}$ GCM model; the superscript $\boldsymbol{H}$ refers to "historical" whereas a and b are the bias adjustment coefficients.

b. Bayesian Model Averaging (BMA)

General description

Bayesian model averaging (BMA; Raftery et al. 2005, see also Duan et al. 2007, Ajami et al. 2007) aims to reduce multi-model uncertainty by assigning weights to all models, with the weights representing posterior probabilities of the models given historical observations. BMA has previously been used to derive probability distributions of continental precipitation and temperature projections from a CMIP3 multi-model ensemble (Duan and Philips 2010). The BMA predictive distribution is a weighted sum of conditional probability distributions of individual models. Let’s consider the same notations used earlier and denote by $\boldsymbol{f_k}$ annual (or monthly) precipitation projections of the $\boldsymbol{k}^{\text{th}}$ model. BMA yields the following predictive model:

$$p(\boldsymbol{y}|\boldsymbol{f_1},\boldsymbol{f_2},\dots,\boldsymbol{f_K}) = \sum_{k=1}^{K} w_k\, p_k(\boldsymbol{y}|\boldsymbol{f_k}) \tag{7.2}$$

Where $\boldsymbol{y}$ is the sought-after precipitation projections. The left-hand side represents the probability density function (pdf) of the BMA model which is equal to a weighted sum of the individual conditional pdfs of models 1, 2, …, K. As noted earlier, the weights w_k represent posterior probabilities of models conditioned on historical observations; thus, they sum to 1. The pdfs p_k for k = 1, 2, …, K are commonly assumed to be normal distributions which is the case in the present study. The weights w_k are estimated by maximizing the log-likelihood function of the pdf in the left-hand side using historical observations. Put simply, $\boldsymbol{y^H}$ and $\boldsymbol{f_k^H}$ are substituted for $\boldsymbol{y}$ and $\boldsymbol{f_k}$ respectively in equation 7.2 in order to estimate w_k. Several techniques such as the expectation-maximization (EM) algorithm (Dempster et al. 1977) can be used to converge to a local maximum of the log-likelihood function. Here, we use a Differential Evolution – Markov Chain (DE-MC) algorithm (Ter Braak 2006) to find the optimum values of w_k.

DE-MC Algorithm and estimation of BMA weights

The DE-MC algorithm (Ter Braak 2006) combines the genetic algorithm variant of Differential Evolution (Storn and Price 1997) and the sampling techniques of Markov Chain Monte Carlo (MCMC) for optimization over real parameter space. The essential idea of DE-MC is that a large number of *N* chains, each consists of a randomly sampled vector of parameters, are run in parallel. The *N* chains exchange information with each other according to the simple rules of DE, and they are updated sequentially to provide a new candidate solution at each step. This process continues for a number of *T* steps which is often referred to as the number of generations, in connection with the natural selection basis of genetic algorithms. Both *N* and *T* are hyperparameters specified for the algorithm, and they are chosen to be 500 and 10,000 respectively in the present study. The choice of hyperparameters can be judged tentatively by tracking the convergence rate of the algorithm across generations, and the chosen values of hyperparameters were found to be optimal in this study.

At each generation of *T*, the algorithm evaluates the fitness of the *N* population members using an objective function. In the present study, the objective function is obtained from equation 7.2 modified by substituting $\boldsymbol{y^H}$ and $\boldsymbol{f_k^H}$ for $\boldsymbol{y}$ and $\boldsymbol{f_k}$ respectively. Specifically, the conditional probability distribution $p_k(\boldsymbol{y^H}|\boldsymbol{f_k^H})$ of the **k**th model is assumed to be a normal distribution of the following form:

$$p_k\left(\boldsymbol{y^H}\middle|\boldsymbol{f_k^H}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\,\sigma^{-2}\,\left(\boldsymbol{y^H}(t) - \boldsymbol{f_k^H}(t)\right)^2\right] \tag{7.3}$$

Where σ^2 is the variance of the normal distribution and considered as a parameter to be optimized by the DE-MC, whereas $\boldsymbol{y^H}(t)$ and $\boldsymbol{f_k^H}(t)$ are the annual historical PERSIANN-CDR observations and simulations of the **k**th model respectively for the year *t*. For numerical stability, the objective function is considered as a log likelihood of the following form:

$$\mathcal{L}\left(\boldsymbol{y}^{\boldsymbol{H}} | \boldsymbol{f}_{\boldsymbol{1}}^{\boldsymbol{H}}, \boldsymbol{f}_{\boldsymbol{2}}^{\boldsymbol{H}}, \ldots, \boldsymbol{f}_{\boldsymbol{K}}^{\boldsymbol{H}}\right) = \sum_{t=1}^{n} \log\left\{\sum_{k=1}^{K} w_k\, p_k(\boldsymbol{y}^{\boldsymbol{H}} | \boldsymbol{f}_{\boldsymbol{k}}^{\boldsymbol{H}})\right\} \tag{7.4}$$

Where $\mathcal{L}$ refers to the log likelihood of the distribution and *n* is the number of years in the historical period (1983 – 2014). The log likelihood in equation 7.4 is used as an objective function by the DE-MC algorithm to evaluate the fitness of the *N* population members at each generation. Each candidate solution consists of the vector of parameters w_k in addition to the variance σ^2. It should be noted that all the foregoing steps are implemented for each spatial grid in the study area separately.

Computation of BMA uncertainty range

One of the main advantages of the BMA approach is offering a posterior distribution instead of a single expected value (i.e. mean). Consequently, this distribution can be sampled to obtain uncertainty ranges with regard to precipitation projections. Following the estimation of the parameters w_k and σ^2 as illustrated in the previous section, the posterior distribution of precipitation projections is given by equation 7.2 after substituting the obtained values of parameters. In order to sample this posterior distribution, we follow two simple steps: first, a number in the range 1, 2, … K is drawn randomly with probabilities w_1, w_2 … w_K; second, a random sample is drawn from the distribution $p_k(\boldsymbol{y}|\boldsymbol{f_k})$. These two steps are repeated 1000 times to approximate the posterior distribution at each spatial grid.

c. Evaluation Metrics

In the present study, we first evaluate the performance of CMIP6 GCMs in simulating annual precipitation for the historical period (1983 – 2014). This is carried out using metrics of bias, relative bias, seasonality and spatial correlations, coefficient of variability and rank histogram. Bias in this study is calculated as the difference given by mean annual precipitation of CMIP6 GCM minus mean annual precipitation of PERSIANN-CDR. Thus, positive and negative values indicate overestimation and underestimation respectively. Relative bias is obtained by normalizing the bias; specifically, dividing by mean annual precipitation of PERSIANN-CDR.

Seasonality correlation is computed as Pearson correlation coefficient between the climatology of monthly CMIP6 GCMs and PERSIANN-CDR precipitation whereas the spatial correlation is computed as Pearson correlation coefficient between mean annual precipitation of CMIP6 GCMs and PERSIANN-CDR for all spatial grids within a basin. The coefficient of variation for CMIP6 GCMs and PERSIANN-CDR is computed as the ratio of the standard deviation of annual precipitation to its mean. Finally, the rank histogram (Hamill 2001) is an efficient and convenient method to assess the reliability of ensemble forecasts. Its fundamental idea is to keep track of the rank of observed precipitation with respect to forecasts of ensemble members at each time step of forecast; these ranks are then used to construct the histogram. If the ensemble range effectively captures uncertainty, then the histogram is expected to be uniform. On the other hand, if the ensemble range is too narrow (wide), the rank histogram will be concentrated near the ends (center). It should also be emphasized that in the evaluation of historical model performance as well as in future projections, we also examine the performance of ensemble mean of the 20 CMIP6 GCMs. This ensemble mean is an arithmetic average of the 20 GCMs, and it should not be confused with the BMA mean. The latter is the mean of the posterior distribution obtained from Bayesian model averaging and is only used in future projections.

7.4 Results and Discussion

a. Evaluation of CMIP6 GCMs for the recent past (1983-2014)

Bias in annual precipitation

We first examine the performance of the different GCMs in simulating the mean value of annual precipitation for the baseline period (1983 – 2014). Figure 7.2 shows the bias in spatially averaged annual precipitation over the Nile, Blue Nile and Upper White Nile basins for each GCM as well as the ensemble mean with respect to PERSIANN-CDR. There is a clear spread between the models with a bias range of (-430 – 389 mm), (-619 – 661 mm) and (-738 – 791 mm) in the Nile, Blue Nile and Upper White Nile basins respectively; see Table 7.1. These biases are significant since they represent up to 64%, 61% and 64% of mean annual precipitation in the three basins respectively. Although the ensemble mean reduces the biases, it fails to outperform the best performing model in the three basins.

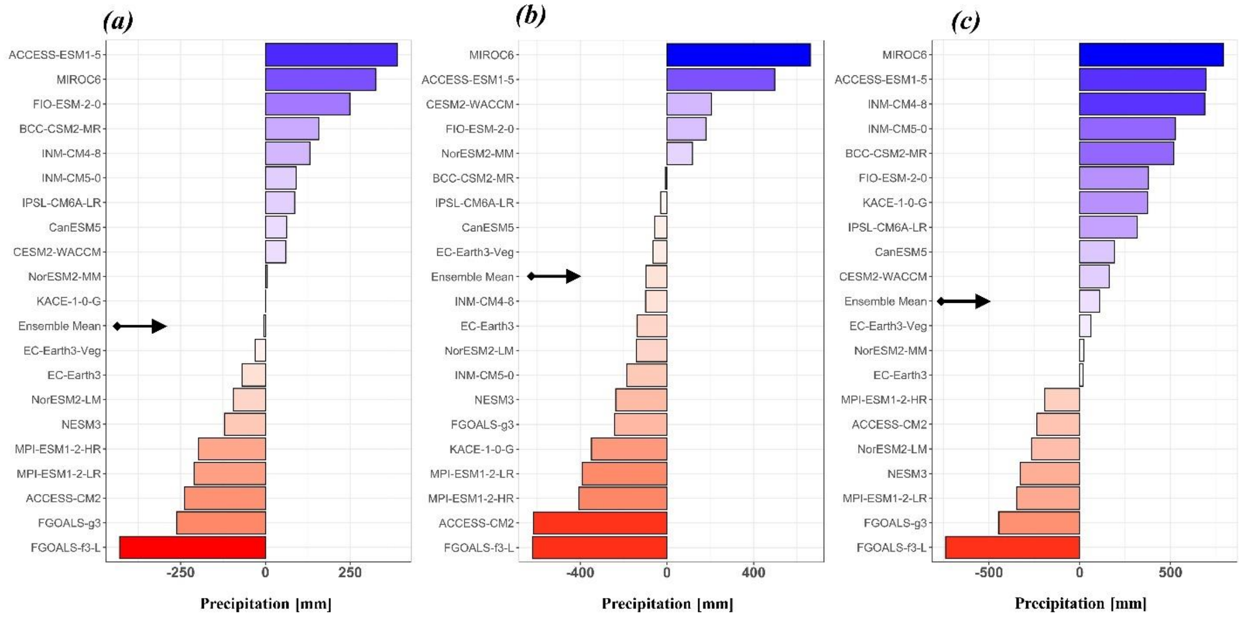


Figure 7.2 *Bar chart shows the bias in annual precipitation of the 20 CMIP6 GCMs and the ensemble mean with respect to PERSIANN-CDR in the period (1983 – 2014). Annual precipitation is spatially*

averaged over (a) the entire Nile basin, (b) the Blue Nile basin and (c) the Upper White Nile basin. Black arrows point to ensemble mean.

Figure 7.3 shows the biases proportional to mean annual precipitation (i.e. relative bias) of the 20 GCMs in addition to the ensemble mean for each grid (1° x 1°) in the Nile basin. Apart from inter-model differences in bias, Figure 7.3 shows that there is a considerable spatial variability in bias within individual models. The values of relative bias over large areas of the basin exceed ± 0.3 (stapled grids in Figure 7.3) which underscore the importance of bias adjustment of GCM outputs prior to evaluation of future projections. In addition to examining the ability of GCMs in simulating the amount of total precipitation in the basins, it is important to investigate their accuracy in simulating the spatial patterns of precipitation. Table 7.1 shows the spatial correlation coefficient of the 20 GCMs and the ensemble mean against PERSIANN-CDR. This reflects how well each model represents the spatial variability of annual precipitation within the Nile basin and its two headwaters basins. Clearly, all the models fairly represent the spatial variability of annual precipitation within the Nile basin as evidenced by correlation coefficients greater than 0.8. Furthermore, the ability of the models to represent spatial variability within the Blue Nile basin is quite reasonable with a minimum correlation coefficient of 0.58. However, the correlation of spatial variability within the Upper White Nile basin is drastically lower, with many models showing a negative correlation, and a maximum correlation coefficient of only 0.49. This highlights that while the GCMs performance in terms of bias is comparable in the Nile basin and its headwaters basins, the GCMs specifically underperform in the Upper White Nile basin with regard to representation of spatial variability. We speculate that the lower performance of GCMs in simulating the spatial variability of precipitation in the Upper White Nile basin is due to the complex rainfall regime in this region. In addition to mechanisms such as monsoonal winds that modulate rainfall in East Africa (i.e. Blue Nile basin), this region is also affected by the interplay of several factors including the Intertropical Convergence Zone (ITCZ), the El Niño-Southern Oscillation (ENSO) and Quasi-biennial Oscillation (QBO) among others (Nicholson and Yin 2002, Dezfuli 2017).

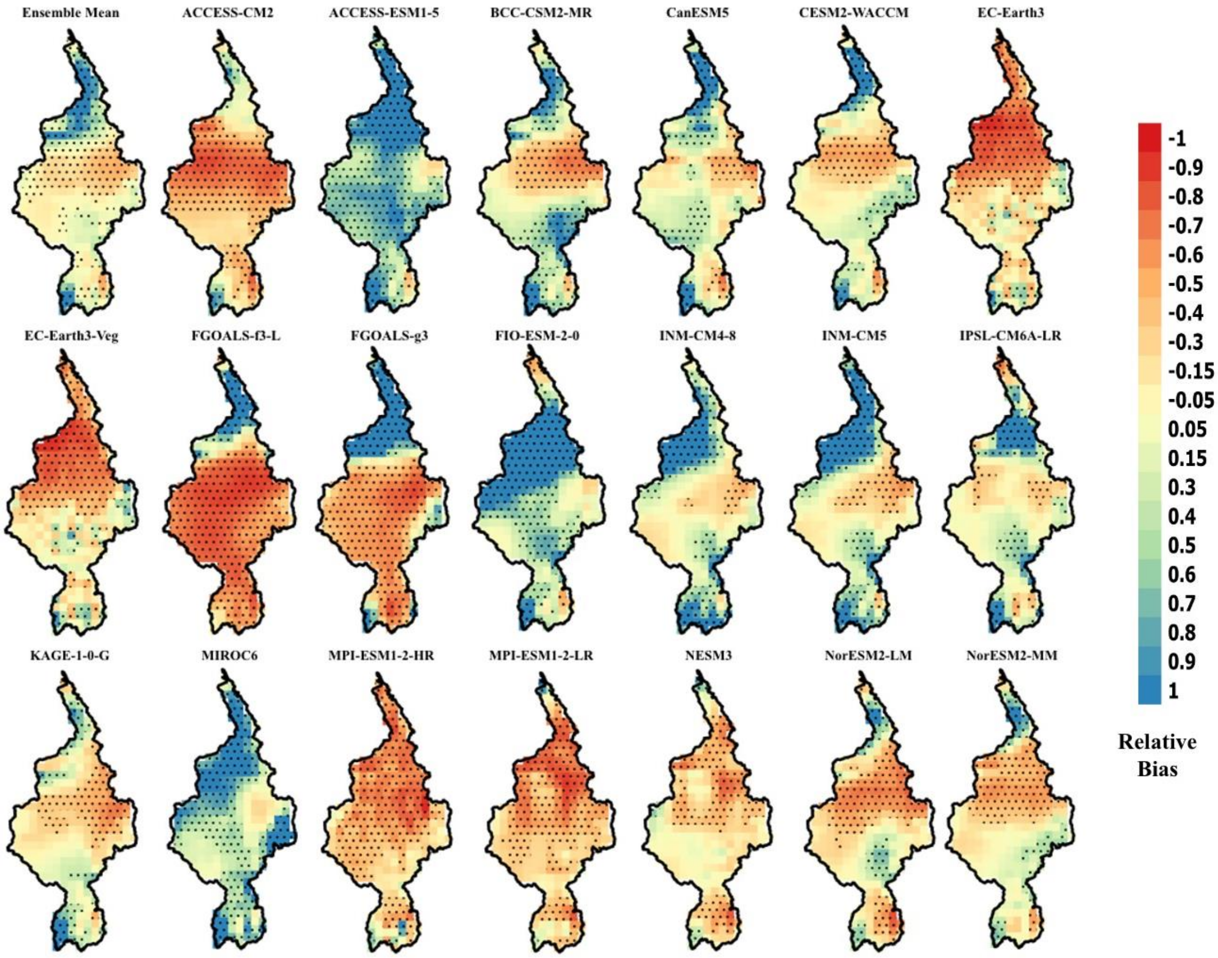


Figure 7.3 *Maps show the relative bias of annual precipitation during the baseline period (1983 – 2014) for each model of the 20 CMIP6 GCMs and the ensemble mean benchmarked against PERSIANN-CDR. Relative bias is calculated as the absolute bias (annual precipitation GCM – annual precipitation PERSIANN-CDR) normalized by annual precipitation PERSIANN-CDR. Blue and red colors show overestimation and underestimation bias respectively. Stapled grids indicate values of relative bias > 0.3 or < -0.3.*

Table 7.1 *Evaluation of CMIP6 GCMs precipitation against PERSIANN-CDR over the entire Nile, Blue Nile (B Nile) and Upper White Nile (W Nile) basins.*

Model	**Bias (mm)**			**Spatial correlation**			**Seasonality correlation**		
	Nile	**B Nile**	**W Nile**	**Nile**	**B Nile**	**W Nile**	**Nile**	**B Nile**	**W Nile**
ACCESS-CM2	-239	-615	-238	0.81	0.82	†	0.94	0.97	0.55
ACCESS-ESM1-5	389	497	696	0.89	0.9	†	0.91	0.94	0.55
BCC-CSM2-MR	157	**-6** *	517	0.87	0.88	†	0.96	**0.99** *	0.75
CanESM5	63	-56	190	0.83	0.91	†	0.93	0.96	0.55
CESM2-WACCM	61	205	163	0.91	0.9	0.11	0.91	0.96	0.68
Earth3	-31	-65	61	0.9	0.79	**0.49** *	0.88	**0.99** *	0.5
Earth3-Veg	-68	-138	**18** *	0.89	0.78	0.47	0.89	**0.99** *	0.54
FGOALS-f3-L	-430	-619	-738	0.85	0.87	0.09	0.87	0.97	0.5
FGOALS-g3	-261	-242	-446	0.68	0.58	†	0.81	0.95	0.54
FIO-ESM-2-0	249	179	377	0.9	0.93	0	0.81	0.72	0.59
INM-CM4-8	131	-98	688	0.86	0.83	0.27	0.88	0.95	0.72
INM-CM5-0	90	-184	525	0.87	0.86	0.12	0.87	0.9	0.77
IPSL-CM6A-LR	87	-30	315	0.86	0.85	†	0.9	0.94	0.63
KAGE-1-0-G	**1***	-348	374	0.85	0.8	0.08	0.95	**0.99** *	0.52
MIROC6	326	661	791	0.87	0.73	0.18	**0.98** *	0.98	0.71
MPI-ESM1-2-HR	-198	-405	-193	0.85	0.84	0.22	0.9	0.93	0.76
MPI-ESM1-2-LR	-210	-391	-349	0.91	**0.95** *	0.19	0.95	0.96	0.7
NESM3	-121	-235	-328	0.89	**0.95** *	†	0.96	0.97	0.76
NorESM2-LM	-94	-140	-266	0.84	0.9	†	0.95	0.97	0.68
NorESM2-MM	5	118	21	0.92	0.9	0.27	**0.98** *	**0.99** *	**0.93** *
Ensemble Mean	-5	-96	109	**0.92** *	0.89	0.08	0.96	**0.99** *	0.71

***** The best performing model according to the metric under consideration. † correlation coefficient is negative.

Interannual variability and uncertainty

Figure 7.4a shows the annual precipitation coefficient of variation (ratio of standard deviation to mean) for the 20 GCMs and PERSIANN-CDR. Clearly, all models severely underestimate the interannual variability in the Nile basin and its headwaters basins. Specifically, the average coefficient of variation for the 20 GCMs is 4 to 7 times less than that of PERSIANN-CDR. Consequently, the bias adjusted ensemble of GCMs is under-dispersive which entails that the ensemble does not represent the true uncertainty in annual precipitation. This is demonstrated in Figure 7.4b which shows the rank histogram of PERSIANN-CDR with respect to the bias-adjusted GCMs ensemble for the period (1983 – 2014). If the ensemble truly captures the variability of annual precipitation, the rank histogram in the bins (2 - 19) should contain 19/21, or 90.5%, of PERSIANN-CDR values. Instead, the ensemble only contains 25% of PERSIANN-CDR observations. These results highlight that using bias adjusted GCMs will lead to underestimation in the uncertainty of precipitation projections. It will be shown later how this issue can be resolved using the Bayesian model averaging approach.

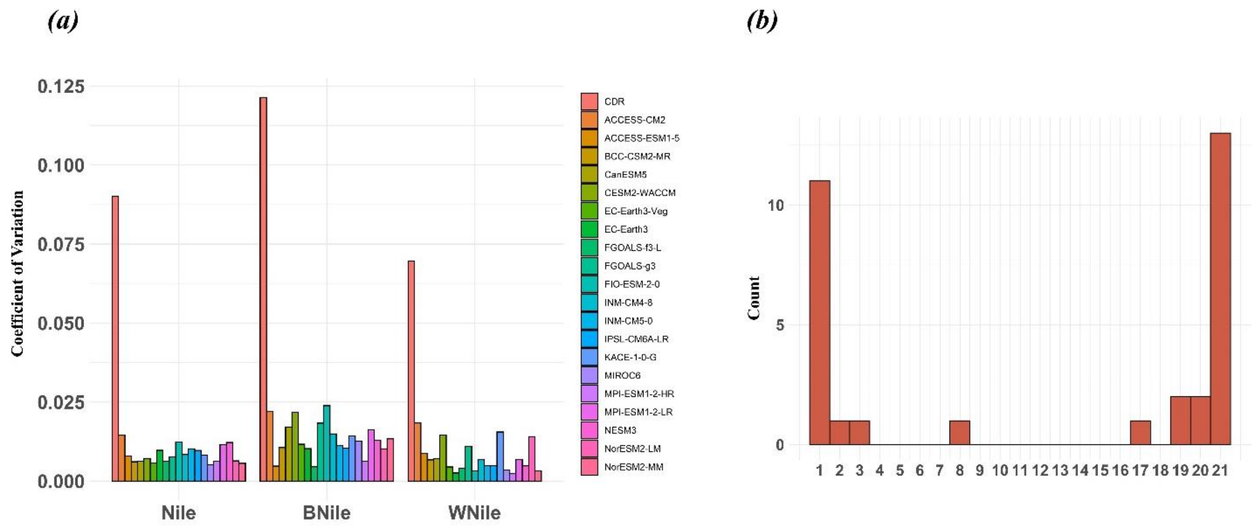


Figure 7.4 (a) *Coefficient of variation of PERSIANN-CDR and the 20 GCMs annual precipitation for the period (1983-2014).* ***(b)*** *Rank histogram of PERSIANN-CDR annual precipitation for the period (1983-2014) with respect to the 20 GCMs.*

Seasonal cycle

Here we evaluate the performance of the GCMs in capturing the seasonal cycle of precipitation. This is particularly important from the standpoint of assessing the hydrological impacts of climate projections such as variability in the Nile river flow and reservoir operations. Figure 7.5 shows the observed climatology monthly precipitation (red line) as well as simulations of the 20 GCMs (black dashed lines) and their ensemble mean (solid black line). The two headwaters basins are characterized by distinct precipitation regimes; see Figures 7.5b and 7.5c. Specifically, precipitation in the Blue Nile sub-basin is monsoonal with pronounced seasonality (July – September) meanwhile Upper White Nile sub-basin experiences two rainy seasons (March – May, October – December) (Conway 2005). Specifically, the seasonal rainfall pattern in the Upper White Nile basin follows the seasonal migration of the Intertropical Convergence Zone (ITCZ) which leads to a bimodal seasonal cycle (Kizza et al 2012). The seasonal cycle over the entire Nile basin, thereby, is a reflection of the cycles at the two headwaters basins; specifically, there is a major peak in (July – September) and a less pronounced one around (April – May). Despite overestimation and underestimation bias, the GCMs adequately capture the seasonal variability in precipitation. This is particularly apparent in the Nile and Blue Nile basins with the ensemble mean showing a correlation coefficient of 0.96 and 0.99 respectively in capturing the seasonal cycle; see Table 7.1. On the contrary, the GCMs are less capable of capturing the seasonal cycle over the Upper White Nile basin with a correlation coefficient of 0.71 for ensemble mean; in addition, the ensemble mean overestimates the second rainy season (October – December).

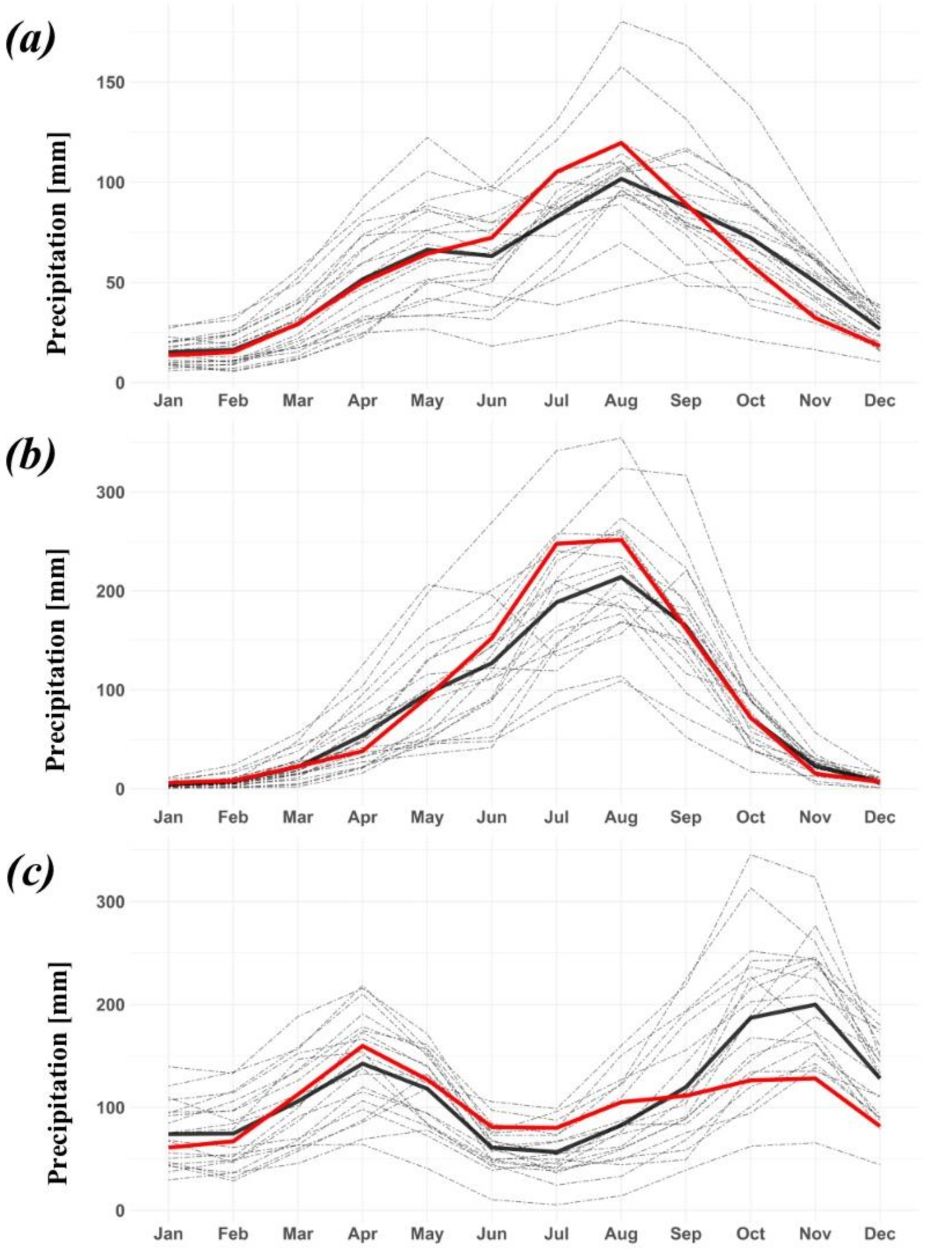


Figure 7.5 *Climatology of monthly precipitation for the period (1983 – 2014) spatially averaged over:* ***(a)*** *the entire Nile basin,* ***(b)*** *the Blue Nile basin, and* ***(c)*** *the Upper White Nile basin. The 20 CMIP6 GCMs are shown in thin black dashed lines. The ensemble mean is shown in solid black line whereas the observed precipitation from PERSIANN-CDR is shown in red.*

Overall, there are numerous observations to be drawn from the retrospective analysis of GCMs simulations; however, two key findings are particularly worthy of consideration. First, the notion of a best performing model is very sensitive to the specific metric used for evaluation as well as the spatial domain of analysis. Table 2 shows the best performing model with respect to each metric (in bold font and an asterisk). Clearly, a different "best performing" model can be selected according to each metric and spatial domain. For instance, KAGE-1-0-G is the best performing model in terms of bias in annual precipitation over the entire Nile basin (bias= 1 mm) whereas NorESM2-MM is the best performing model in capturing the seasonal cycle of precipitation in the three basins. Second, although the ensemble mean provides adequate performance, it does not outperform all individual models as clearly shown in Table 7.1. This pinpoints that the ensemble mean is sensitive to ensemble members at the end of the performance spectrum. It also underlines that analysis of future projections can benefit from advanced model averaging schemes that consider retrospective model performance to provide a superior estimate to that of individual models.

b. Precipitation Projections for the period (2015-2100)

BMA mean precipitation projections

Here, we analyze mean precipitation projections obtained from the BMA model for the period (2015 - 2100) with respect to PERSIANN-CDR for the baseline period (1983 – 2014). Annual precipitation series of the 20 GCMs for the period (2015 – 2100) were first bias adjusted using the coefficients estimated from equation 7.1. Next, the BMA weights and their corresponding BMA precipitation projections were computed. These calculations were performed at the grid scale as opposed to the entire spatial domain due to the wide climatic variability and the different precipitation regimes in the Nile basin.

Figure 7.6a shows the projected changes in mean annual precipitation spatially averaged over the entire Nile basin from the 20 GCMs, ensemble mean and BMA. There is a spread in model projections with 14 models indicating an increase in mean annual precipitation and 6 models showing a decrease. Overall, percentage change in mean annual precipitation ranges from -1.7 % to 3.2 %. The BMA shows a statistically insignificant increase of 1.34 % (p-value = 0.2) (see Figure 7.6a and Table 2) compared to 0.82% from the ensemble mean. Figure 7.6b shows the percentage change in mean annual precipitation projected from BMA for the period (2015 – 2100) with respect to the baseline period (1983 – 2014). Clearly, there is spatial variability both in the sign and magnitude of change. A slight decrease in precipitation is observed in southern regions (the Upper White Nile sub-basin) whereas the eastern regions (Blue Nile sub-basin) show both an increase and a decrease in precipitation. The statistically significant changes in precipitation, at a significance level of 0.05, are observed over the riparian arid regions (stapled grids in Figure 7.6b) which have almost no impact on Nile streamflow. Specifically, there is a significant increase in precipitation in Northern Sudan (15°N - 20°N), and a precipitation decrease to the northward. Furthermore, Figure 7.6c shows the number of models that agree in the sign of change with BMA projections out of the 20 GCMs used in this study. It appears that spatial grids in which the projected BMA change is significant (stapled grids in Figure 7.6b) are grids in which a large number of the 20 GCMs agree in the sign of change.

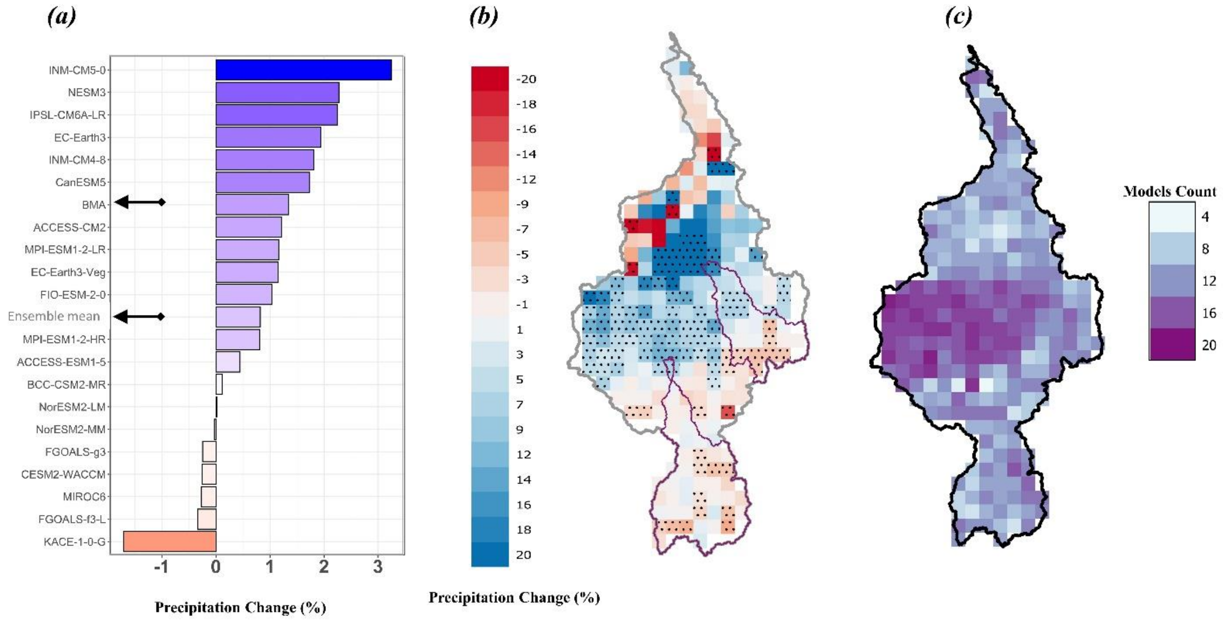


Figure 7.6 (a) *Percentage change of spatially averaged annual precipitation projected from 20 bias adjusted GCMs, ensemble mean and BMA model. Spatial averaging is carried out over the entire Nile river basin. Black arrows point to ensemble mean and BMA.* ***(b)*** *Percentage change of annual precipitation projected from the BMA model for the period (2015 – 2100) with respect to the baseline period (1983 – 2014) at spatial grids of 1 °x 1 °. Stapled grids indicate a statistically significant change at α= 0.05.* ***(c)*** *The number of models that agree on the sign of change of the BMA model out of the 20 GCMs used in this study.*

It is important to narrow the analysis down to regional scales of unified precipitation regimes. Here we focus on headwaters basins, namely the Blue Nile and Upper White Nile sub-basins (see Figure 7.1). These basins are characterized by distinct precipitation regimes as shown in Figure 7.5. Figure 7.7a shows the decadal moving average of percentage change in projected annual precipitation at the Blue Nile sub-basin. Inter-model differences are clearly present with a range of -5% to 5% (dashed thin black lines). BMA and ensemble mean are nearly equivalent, and they show no noticeable change in precipitation. Precisely, BMA shows a change of 0.03%, not statistically significant with p-value of 0.49 (see Table 7.2). At the Upper White Nile sub-basin

(Figure 7.7b), BMA deviates from ensemble mean, and it indicates a decrease of -1.65% in mean annual precipitation, p-value of 0.09 (see Table 7.2).

In addition to precipitation projections of the BMA and ensemble mean, Figure 7.7 and Table 7.2 shows the projected change in precipitation in each basin from a selected subset of 3 models. The selection criterion is to identify the 3 models with the least bias in the historical period (1983 – 2014); see Figure 7.1 and Table 7.1. In each basin, a subset of 3 models is selected, and its mean is calculated. Table 7.2 shows that the estimate of the best 3 models is consistently opposite in sign to the estimate of BMA and ensemble mean. However, their projected changes are small (< 1%) and statistically insignificant at $\alpha = 0.05$. We also examined precipitation projections for the rainy seasons in the Nile headwaters basins due to their impact on the variability of the Nile streamflow. The results are shown in Table 7.3, and they don't show a statistically significant trend, whether decreasing or increasing. Of particular importance is the (June – August) rainy season in the Blue Nile basin since it contributes 60% of the annual Nile flow. Table 7.4 shows that the projected change is statistically insignificant with a decrease of -0.09% (p-value = 0.49).

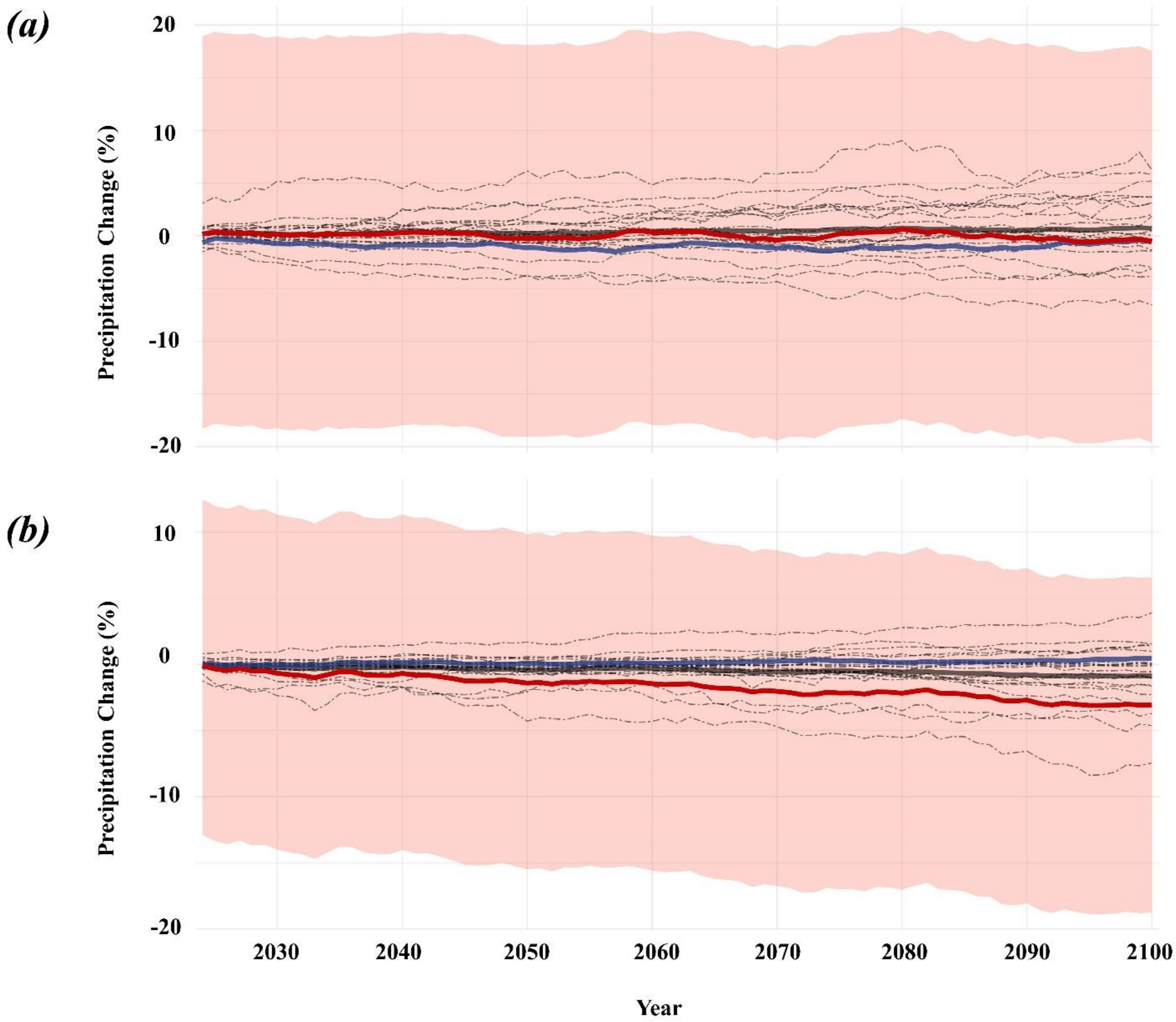


Figure 7.7 *10-years moving averages of percentage change in projected annual precipitaiton for the period (2015 – 2100) with respect to the baseline period (1983 – 2014). The horizontal axis shows the year at the end of the 10-years time window. Dashed thin black lines, thick black, red and blue lines indicate projections of the 20 GCMs, ensemble mean, BMA and "best 3 models" respectively. The pink shaded area represents 90% uncertainty bounds of the BMA model.* ***(a)*** *Spatially averged over the Blue Nile sub-basin.* ***(b)*** *Spatially averaged over the Upper White Nile sub-basin.*

Table 7.2 *Projected changes in mean annual precipitation in the Nile, Blue Nile and Upper White Nile basins. In parentheses are the p-values of the projected changes.*

Basin	Ensemble mean	Best 3 models	BMA	BMA 90% Confidence Interval	
				Lower (%)	Upper (%)
Nile	0.82 % (0.3)	-0.19% (0.45)	1.34 % (0.2)	-3.2	5.4
Blue Nile	0.43% (0.42)	-0.92% (0.33)	0.03% (0.49)	-11.2	16.3
Upper White Nile	-0.45% (0.36)	0.17% (0.44)	-1.65% (0.09)	-9.0	5.9

Uncertainty in BMA precipitation projections

As discussed earlier, the bias adjusted GCMs ensemble is under-dispersive; thus, it underestimates the uncertainty of precipitation. The BMA approach provides a remedy to this problem because it accounts for two types of variability. Specifically, the BMA total variability is decomposed into two components: between and within variability (Raftery et al. 2005). The former considers the spread of ensemble members whereas the latter accounts for the variability within the individual members. This is clearly shown in Figure 7.7 which shows the BMA 90% confidence interval (shaded pink area). While the spread of models (black dashed lines) is limited to a range of approximately (-5% — 5%) in the two basins, the BMA 90% confidence interval extends to approximately ±20%. This extended uncertainty is the result of the BMA approach consideration of the within variability that is not accounted for in the multi-model ensemble.

Figure 7.8 shows the distributions of the BMA precipitation projections for the period (2015 – 2100) expressed as a percentage change with respect to the baseline period (1983 – 2014). The distributions also show the mean (black dashed line) and the 90% confidence interval limits (red dashed lines). The mean values are the same as those shown in Table 7.2. Figure 7.8a shows the distribution for the Nile basin; the 90% interval range is (-3.2% — 5.4%) with a width of 8.6%. This shows that the probability of an increase in precipitation is higher than that of a decrease. On the contrary, Figure 7.8c shows that the probability of a decrease in rainfall at the Upper White Nile basin is higher with a 90% confidence interval range of (-9% — 5.9%) with a width of 14.9%. As for the Blue Nile basin, the uncertainty range is wider; specifically, (-11.2% — 16.3%) with a width of 27.5%. Besides the wide range of uncertainty in the Blue Nile basin, Figure 7.8c shows that the distribution is more centered around 0%; thus, there is also increased uncertainty in the sign of change in precipitation projections.

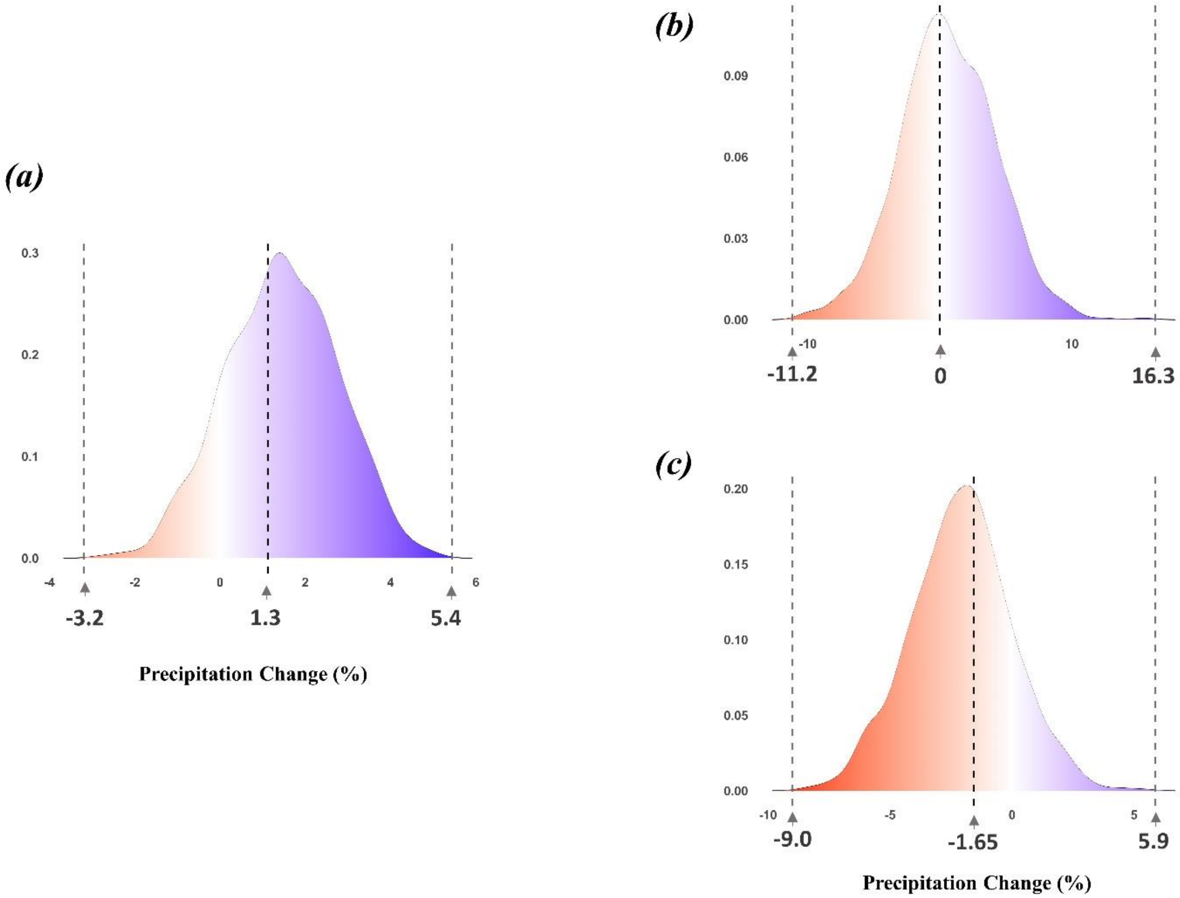


Figure 7.8 *The distirbution of BMA precipitation projections expressed as the mean percentage change with respect to baseline period (1983 – 2014). The distribution mean and 90% confidence bounds are shown in black and gray dashed lines respectively. Precipitation is spatially averaged over:* ***(a)*** *the Nile basin,* ***(b)*** *the Blue Nile basin and* ***(c)*** *the Upper White Nile basin. Fill colors of red and blue indicate decrease and increase respectively whereas white color indicates no change.*

Table 7.3 BMA projected changes in seasonal precipitation in the Nile, Blue Nile and Upper White Nile basins.

Basin	June - August		October - December		March - May	
	Change in mean (%)	p-value	Change in mean (%)	p-value	Change in mean (%)	p-value
Blue Nile	-0.09	0.49	-0.83	0.45	0.18	0.49
Upper White Nile	0.37	0.48	0.47	0.45	-0.09	0.49

7.5 Conclusions

This study examined the performance of 20 CMIP6 GCMs in simulating precipitation for the period (1983 – 2014) over the Nile basin, and then used a Bayesian model averaging scheme to derive precipitation projections for the period (2015 – 2100). The main findings of retrospective analysis are as follows:

- The bias in most GCMs simulations is significant (up to 64% of mean annual precipitation) which consequently pinpoints the importance of bias adjustment prior to analysis of precipitation projections. In addition, the spatial patterns of bias vary considerably within individual models both in the sign and value.

- Although all models fairly represent spatial patterns and seasonal cycle of precipitation over most regions in the Nile basin, the results show that the performance of models is less accurate at the Upper White Nile basin.

- Selection of a “best performing model” is highly dependent on the specific metric chosen as a criterion. Moreover, the results show that the ensemble mean usually does not outperform all individual models.

- All models severely underestimate the interannual variability as represented by the coefficient of variation. As a result, the ensemble range underestimates the uncertainty of precipitation.

Bayesian model averaging show that projected changes in precipitation varies spatially across the Nile basin with clear regional patterns; in particular, a mild decrease of -1.65% in the Upper White Nile sub-basin, almost no change (0.03%) in the Blue Nile sub-basin, and significant changes (both increasing and decreasing) in the arid riparian Nile basin. Regarding the Blue Nile sub-basin, our results are similar to those reported by Elshamy et al. (2009) which showed no change in annual precipitation based on 17 CMIP3 GCMs. However, they are at odds with results in Kim and Kaluarachchi (2009), and Fenta Mekonnen and Disse (2018) which showed an increase of 11% and (2.1% — 43.8%) respectively. Generally, it is not possible to make a conclusive judgement on which study, among previous studies and including the present one, has more credibility because they differ significantly in the models, climate scenarios, future time period and geographical regions. Nonetheless, we argue that a strict and more cautious approach compared to previous ones has been adopted in this study. Lastly, the BMA probability distributions show that the probability of a decrease in annual precipitation is more likely in the Upper White Nile basin. Moreover, the uncertainty in annual precipitation projections over the Blue Nile basin is higher both in terms of values and sign of change.

Part III: Conclusions

Chapter 8

8 Conclusions and Future Directions

8.1 Summary of findings

As outlined in section 1.2, the results of the studies presented in the dissertation have been reported in several peer-reviewed publications. The main findings of each study can be summarized as follows:

- **Chapter 2:** The experiment that was carried out to explore the performance of causal inference methods in recovering the causal structure of hydrometeorological systems revealed several insights into the advantages and disadvantages of each method. These insights are summarized in section 2.5, and they can potentially serve the hydrometeorological community as a guideline to the suitability of each causal inference method, and the selection of the most appropriate method to a given research problem. Furthermore, causal analysis of observational datasets with the aim of investigating the differential impact of environmental variables in regulating evapotranspiration resulted in several key findings. First, the results show that environmental drivers of evapotranspiration and their relative strength vary considerably from one season to another. Second, the results highlight the potential of causal inference methods in the unambiguous detection of interactions regarding the evapotranspiration process. These findings indicate that causal inference methods can potentially be used with observational datasets across a large number of FLUXNET sites to examine the variability in drivers of evapotranspiration. Such an analysis will aid in the selection of models that are tailored and only sensitive to the dominant variables deriving evapotranspiration; thus, improving our simulation and predictive capacity of evapotranspiration mechanisms.

- **Chapter 3:** This study utilized methods rooted in the theory of chaotic dynamical systems to examine the properties of dynamical behavior in hydrologic basins across the contiguous United States. The main findings of this study are the coherent relationships that were discovered between the properties of dynamical behavior and the physical characteristics of hydrologic basins. These include the statistically significant relationships between basin size and dynamic dimensionality, and the impact of vegetation cover on the presence of nonlinearity. These relationships indicate the possibility of formulating laws that govern the complexity of dynamical behavior in a hydrologic basin using information related to its physical characteristics. The implications of such findings are of immense importance to extrapolation of parameters and selection of rainfall-runoff models in ungauged hydrologic basins — an issue that is deemed to be one the grand challenges facing contemporary hydrology.

- **Chapter 4:** The results of this study highlight the potential of using state-space forecasting in providing daily streamflow forecasts with an accuracy higher than that provided by state-of-the-art deep learning models. While the results presented in this chapter is not intended to be a comprehensive comparison of state-space forecasting and deep learning models, the findings highlight that the paradigm of using the state-space to make forecasts of streamflow overcome some of the limitations that are present in deep learning models — e.g. bias of deep learning models toward the mean behavior, and lower forecast skill in capturing the extremes. These results also indicate the potential of integrating state-space models with deep learning to leverage the strength of both approaches.

- **Chapter 5:** The results presented in this study demonstrate the potential of utilizing information-theoretic measures in revealing patterns of association in complex relationships such as the nonlinear relationship between infrared brightness temperature and precipitation. We report several interesting findings, some of which reaffirm observations that was previously reported in the literature, in addition to few novel observations. These latter observations include the relationship between the number of no-rain instances and the information content of IR imagery on precipitation as well as the diminishing returns behavior in the dependence between IR and precipitation that results from an increase in spatial and temporal scales. We also presented a case study of using this analysis to diagnose operational algorithms for estimating precipitation from IR brightness temperature. We argue that this approach of diagnosing the inherent dependence is quite useful in attributing the errors in satellite-based precipitation estimates to either the information content of IR imagery or the assumptions embedded in the algorithm used for estimation.

- **Chapter 6:** The results of this study show that by developing a framework to account for the inherent biases in satellite-based precipitation, one can obtain IDF curves estimates within an error range of (17% — 22%), (6% — 12%) and (3% — 8%) for 1-day, 2-days and 3-days IDFs, respectively, and return periods in the range of (2 — 100 years). More specifically, the framework we developed in this study account for two types of biases. First, the underestimation bias that was found to be well correlated with elevation in the estimates of PERSIANN-CDR. Second, the underestimation resulting from the areal estimate of satellites as opposed to the point estimate. This framework can potentially be applied in ungauged regions depending on the adequacy of the bias adjustment coefficients estimated over the contiguous United States (CONUS). We argue that these estimates might be representative of a wide range of climatic conditions because they were estimated using data from the entire spatial domain of CONUS which encompasses a wide range of climatic conditions. However, we also emphasize the importance of carrying out similar analysis using observational datasets from other regions to check and revise the estimates of coefficients in the bias adjustment model presented in the study.

- **Chapter 7:** In this study, we used Bayesian Model Averaging (BMA) to constrain future projections of precipitation. We argue that this is a more informative approach to adopt than the commonly used approach of ensemble mean. We utilized a long historical record of satellite-based precipitation observations from PERSIANN-CDR to examine future changes in precipitation over the Nile river basin in a warmer climate. We found that projected change in the upper White Nile river basin indicate a slight decrease of precipitation (statistically significant at the 0.1 significance level), whereas projected change in the Blue Nile river basin is highly uncertain both in magnitude and sign of change. These results are of interest and relevance to the current intense geopolitical debate on the availability of water resources in the Nile river basin amongst the completion and filling of the mega Ethiopian dam — the GERD.

8.2 Future Directions

Data-driven methods used in the first part of the dissertation possess a great potential in quantifying feedback mechanisms, unraveling the cause and effect relationships, and confronting models with observations. This potential stems primarily from the fact that these data-driven methods are tailored to detect multivariate, nonlinear interactions; contrary to traditional methods of data analysis in hydrometeorological systems that deals mostly with analysis of univariate and bivariate time series. We argue that discovering multivariate, nonlinear interactions that underly complex hydrometeorological systems is an important aspect to consider in expanding the frontiers of hydrologic sciences. This is of particular importance to address the challenges facing contemporary hydrologic research such as improving our understanding of scale-dependent behavior of hydrologic systems, unraveling patterns of emergent behavior in hydrologic systems, and extrapolating empirical knowledge to ungauged hydrologic basins. Each of these challenges require the formulation of macroscopic laws that describe under general conditions the behavior and response of hydrologic basins as well as the control of basin's physical properties on its dynamic behavior.

The application of exploratory data-driven methods in hydrologic research is nothing but opportune due to the unprecedented availability of hydrometeorological data records in the form of in-situ and remotely sensed observations, field campaigns, climatological reconstructions, reanalysis datasets, and model simulations. We argue that more research along the lines of the studies presented in the first part of the dissertation, especially chapter 2, is sorely needed to advance the use of causal inference, nonlinear dynamics, and information theory in analysis of hydrometeorological systems. Furthermore, we anticipate that in the foreseeable future, researchers in the field of hydrology and climate will benefit significantly from the use of such data-driven methods.

While the studies presented here uses observations solely to detect patterns of interaction, we realize the potential of applying such data-driven exploratory methods in tandem with models. In such a case, there is a significant potential to diagnose model structural errors (epistemic errors) benefiting from juxtaposition of observations and models. We also point out that in an era where deep learning is exuberantly applied in hydrologic sciences, there is a potential of using exploratory data-driven methods to support interpretability of deep learning models and imposing physical constraints (e.g. conservation of mass, energy and momentum) on deep learning models to control their out-of-sample performance (e.g. projections of future climate).

Bibliography

Adams, D. K., & Comrie, A. C. (1997). The north American monsoon. *Bulletin of the American Meteorological Society*, **78**(10), 2197-2214.

Adler, R. F., & Negri, A. J. (1988). A satellite infrared technique to estimate tropical convective and stratiform rainfall. Journal of Applied Meteorology and Climatology, 27(1), 30-51.

Adler, R. F., Huffman, G. J., & Keehn, P. R. (1994). Global tropical rain estimates from microwave-adjusted geosynchronous IR data. Remote Sensing Reviews, 11(1-4), 125-152.

AghaKouchack, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, 116, D02115.

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., & Peng, Q. (2018). Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. *Journal of Geophysical Research: Atmospheres*, **123**(22), 12-543.

Andersen, H. (2013). When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, **80**(5), 672-683.

Arkin, P. A. (1979). The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array. Monthly Weather Review, 107(10), 1382-1387.

Arkin, P. A., & Meisner, B. N. (1987). The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84. Monthly Weather Review, 115(1), 51-74.

Ashouri, H., Hsu, K., Sorooshian, S., & Braithwaite D. K. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, 96 (1), 69 – 83.

Asquith, W. H., & Famiglietti, J. S. (2000). Precipitation areal-reduction factor estimation using an annual-maxima centered approach. Journal of Hydrology, 230(1-2), 55-69.

Awadallah, G. A., ElGamal, M., ElMostafa, A., & ElBadry, H. (2011). Developing Intensity-Duration-Frequency Curves in Scarce Data Region: An Approach using Regional Analysis and Satellite Data. *Scientific Research Engineering*, 3, 215 – 226.

Bai, Y., Wagener, T., & Reed, P. (2009). A top-down framework for watershed model evaluation and selection under uncertainty. *Environmental Modelling & Software*, **24**(8), 901-916.

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of neuroscience methods*, **223**, 50-68.

Barnett, L., Barrett, A. B., & Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical review letters*, **103**(23), 238701.

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, **438**(7066), 303.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & Roo, A. D. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences, 21(1), 589-615.

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K., & Sorooshian, S. (2011). Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology*, 397 (3-4), 225 – 237.

Bernard, M. M. (1932). Formulas for rainfall intensities of long durations. *Transactions of the American Society of Civil Engineers*, 96 (1), 592 – 606.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, **320**(1-2), 18-36.

Birikundavyi, S., Labib, R., Trung, H. T., & Rousselle, J. (2002). Performance of neural networks in daily streamflow forecasting. *Journal of Hydrologic Engineering*, **7**(5), 392-398.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., ... & Stumpp, C. (2019). Twenty-three Unsolved Problems in Hydrology (UPH)–a community perspective. Hydrological Sciences Journal, 64(10), 1141 – 1158.

Bonhommeau, S., Dubroca, L., Le Pape, O., Barde, J., Kaplan, D. M., Chassot, E., & Nieblas, A. E., 2013. Eating up the world's food web and the human trophic level. P. Natl. Acad. Sci. USA. 110(51), 20617-20620.

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2006). NOAA Atlas 14 Volume 2 Version 3.0, Precipitation-Frequency Atlas of the United States. Silver Spring, MD: NOAA, National Weather Service.

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2011). NOAA Atlas 14 Volume 1 Version 5.0, Precipitation-Frequency Atlas of the United States. Silver Spring, MD: NOAA, National Weather Service.

Botter, G., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2009). Nonlinear storage-discharge relations and catchment streamflow regimes. *Water resources research*, **45**(10).

Bruijnzeel, L. A. (2004). Hydrological functions of tropical forests: not seeing the soil for the trees?. *Agriculture, ecosystems & environment*, **104**(1), 185-228.

Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, **13**(3), 637-643.

Bussieres, N., & Hogg, W. (1989). The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. *Atmosphere-Ocean*, 27, 521-541.

Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of hydrology*, **298**(1-4), 242-266.

Castillo, A., Castelli, F., & Entekhabi, D. (2015). An entropy-based measure of hydrologic complexity and its applications. *Water resources research*, **51**(7), 5145-5160.

Charney, J. G. (1975). Dynamics of deserts and drought in the Sahel. *Quarterly Journal of the Royal Meteorological Society*, **101**(428), 193-202.

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. Journal of Geophysical Research: Atmospheres, 113.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Cleveland, T. G., Herrmann, G. R., Tay, C. C., Neale, C. M., Schwarz, M. R., & Asquith, W. H. (2015). New Rainfall Coefficients–Including Tools for Estimation of Intensity and Hyetographs in Texas. Lubbock, TX: Texas Tech Center for Multidisciplinary Research in Transportation (TechMRT).

Creutin, J. D., & Obled, C. (1982). Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Water resources research*, 18, 413-431.

Dalrymple, T. (1960). Flood-frequency analyses, Manual of Hydrology: Part 3. US Geological Survey Water Supply Paper, 1543 – A. Washington, DC: US Government Printing Office.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. Cambridge University Press.

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, ***111***(2), 157-170.

de Noblet-Ducoudré, N., Boisier, J. P., Pitman, A., Bonan, G. B., Brovkin, V., Cruz, F., ... & Van Der Molen, M. K. (2012). Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: results from the first set of LUCID experiments. *Journal of Climate*, **25**(9), 3261-3281.

Dechter, R. (2013). Reasoning with probabilistic and deterministic graphical models: Exact algorithms. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **7**(3), 1-191.

DeGaetano, A. T. & Castellano, C. M. (2017). Future projections of extreme precipitation intensity-duration-frequency curves for climate adaptation planning in New York State. *Climate Services*, 5, 23-25.

Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS One*, **6**(3), e18295.

Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS One*, **6**(3).

Dinko, T., Chidzambwa, S., Ceccato, P., Connor, S. J., & Ropelewski C. F. (2008). Validation of high-resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing*, 29 (14), 4097 – 4110. doi: 10.1080/01431160701772526

Dooge, J. C. (1986). Looking for hydrologic laws. Water Resources Research, 22(9S), 46S-58S.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., ... & Hogue, T. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, **320**(1-2), 3-17.

Eagleson, P. S. (1970). Dynamic Hydrology. New York: McGraw-Hill.

Ebert, E.E., Janowiak, J. E., & Kidd, C. (2007). Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. *Bulletin of the American Meteorological Society*, 88, 47–64.

Eldardiry, H., Habib, E., & Zhang, Y. (2015). On the use of radar-based quantitative precipitation estimates for precipitation frequency analysis. Journal of Hydrology, 531, 441-453.

Elshorbagy, A., Simonovic, S. P., & Panu, U. S. (2002). Estimation of missing streamflow data using principles of chaos theory. *Journal of Hydrology*, **255**(1-4), 123-133.

Eltahir, E. A. (1998). A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water resources research*, **34**(4), 765-776.

Endreny, T. A. & Imbeah, N. (2009). Generating robust rainfall intensity–duration–frequency estimates with short-record satellite data. *Journal of Hydrology*, 371 (1-4), 182 – 191.

Entekhabi, D., Rodriguez-Iturbe, I., & Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology*, **184**(1-2), 3-17.

Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophysical Research Letters*, **44**(21), 11-030.

Feldhake, C. M., & Boyer, D. G. (1986). Effect of soil temperature on evapotranspiration by C3 and C4 grasses. *Agricultural and forest meteorology*, **37**(4), 309-318.

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*.

Findell, K. L., Berg, A., Gentine, P., Krasting, J. P., Lintner, B. R., Malyshev, S., ... & Shevliakova, E. (2017). The impact of anthropogenic land use and land cover change on regional climate extremes. *Nature communications*, **8**(1), 989.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific data. 2(1), 1-21.

Gado, T. A., Hsu, K., & Sorooshian, S. (2017). Rainfall frequency analysis for ungauged sites using satellite precipitation products. *Journal of Hydrology*, 554, 646 – 655.

Gandin, L. S. (1965). Objective analysis of meteorological fields. *Israel Program for Scientific Translations*, 242.

Gauch, M., Mai, J., & Lin, J. (2019). The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. arXiv preprint arXiv:1911.07249.

Gers, F. A., & Schmidhuber, J. (2000, July). Recurrent nets that time and count. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium (Vol. 3, pp. 189-194). IEEE.

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.

Ghosh, B. (1951). Random distances within a rectangle and between two rectangles. Bulletin of Calcutta Mathematical Society, 43, 17-24.

Goodwell, A. E., & Kumar, P. (2017a). Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts. *Water Resources Research*, **53**(7), 5899-5919.

Goodwell, A. E., & Kumar, P. (2017b). Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, **53**(7), 5920-5942.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 15 (5), 1049 – 1054.

Griffith, C. G., Woodley, W. L., Grube, P. G., Martin, D. W., Stout, J., & Sikdar, D. N., 1978. Rain estimation from geosynchronous satellite imagery—Visible and infrared studies. Mon. Weather Rev. 106(8), 1153-1171.

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes: An International Journal*, **22**(18), 3802-3813.

Haberlie, A. M., & Ashley, W. S., 2019. A radar-based climatology of mesoscale convective systems in the United States. J. Climate. 32(5), 1591-1606.

Hashemi, H., Nordin, M., Lakshmi, V., Huffman, G. J., and Knight, R. (2017). Bias Correction of Long-Term Satellite Monthly Precipitation Product (TRMM 3B43) over the Conterminous United States. *Journal of Hydrometeorology*, 18, 2491–2509.

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of climate*, **19**(21), 5686-5699.

Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Castro-Díez, Y., Argüeso, D., & Esteban-Parra, M. J. (2015). Long-range seasonal streamflow forecasting over the I berian P eninsula using large-scale atmospheric and oceanic information. *Water Resources Research*, **51**(5), 3543-3567.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, **9**(8), 1735-1780.

Hong, Y., Adler, R. F., Negri, A., & Huffman, G. J., 2007. Flood and landslide applications of near real-time satellite rainfall products. Nat. Hazards. 43(2), 285-294.

Hosking, J. R. M. & Wallis J. R. (1997). Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge, UK: Cambridge University Press.

Hosking, J. R. M. & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. Water Resources Research, 29 (2), 271 – 281.

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B*, 52 (1), 105 – 124.

Houghton, J. C. (1978). Birth of a parent: The Wakeby Distribution for modeling flood flows. Water Resources Research, 14, 1105-1109.

Houze Jr, R. A. (2014). Clouds in Hurricanes In *Cloud dynamics*, 53, 405-437. Academic press.

Hsu, K. L., Gao, X., Sorooshian, S., & Gupta, H. V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteorol. 36(9), 1176-1190.

Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, **31**(10), 2517-2530.

Hsu, K. L., Gupta, H. V., Gao, X., Sorooshian, S., & Imam, B. (2002). Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. *Water Resources Research*, **38**(12), 38-1.

Hudlow, M. D., & Patterson, V. L. (1979). GATE radar rainfall atlas. National Oceanic and Atmospheric Administration, Environmental Data and Information Service.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., ... & Susskind, J. (2001). Global precipitation at one-degree daily resolution from multisatellite observations. Journal of hydrometeorology, 2(1), 36-50.

Huffman, J. G., Bolvin, D. T., Nelkin, E. J. & Wolf, D. B. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. Journal of Hydrometeorology, 8, 38-55.

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model?. *Water resources research*, **29**(8), 2637-2649.

Janowiak, J. E., Joyce, R. J., & Yarosh, Y. (2001). A real-time global half-hourly pixel-resolution infrared dataset and its applications. Bulletin of the American Meteorological Society, 82(2), 205-218.

Jayawardena, A. W., & Gurung, A. B. (2000). Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach. *Journal of Hydrology*, **228**(3-4), 242-264.

Jayawardena, A. W., & Lai, F. (1994). Analysis and prediction of chaos in rainfall and stream flow time series. *Journal of hydrology*, **153**(1-4), 23-52.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. (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.

Joyce, R., & Arkin, P. A. (1997). Improved estimates of tropical and subtropical precipitation using the GOES precipitation index. Journal of Atmospheric and Oceanic Technology, 14(5), 997-1011.

Joyce, R., Janowiak, J., & Huffman, G. (2001). Latitudinally and seasonally dependent zenith-angle corrections for geostationary satellite IR brightness temperatures. Journal of Applied Meteorology, 40(4), 689-703.

Karl, T., & Koss, W. J. (1984). Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983.

Kiiveri, H., Speed, T. P., & Carlin, J. B. (1984). Recursive causal models. *Journal of the Australian Mathematical Society*, **36**(1), 30-52.

Klemes, V. (1982). Empirical and causal models in hydrology. In *Scientific Basis of Water-Resources Management*, 95-104, National Academy Press.

Klemeš, V. (1983). Conceptualization and scale in hydrology. Journal of hydrology, 65(1-3), 1-23.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., ... & Liu, P. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, **305**(5687), 1138-1140.

Kothyari, U. C., & Garde, R. J. (1991). Annual runoff estimation for catchments in India. *Journal of Water Resources Planning and Management*, **117**(1), 1-10.

Koutsoyiannis, D., Kozonis, D., & Manetas A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. Journal of Hydrology, 206 (1-2), 118-135.

Krajewski, W. F., Ghimire, G. R., & Quintero, F. (2020). Streamflow Forecasting without Models. *Journal of Hydrometeorology*, **21**(8), 1689-1704.

Kramer, P. J. (1940). Root resistance as a cause of decreased water absorption by plants at low temperatures. *Plant physiology*, **15**(1), 63.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology & Earth System Science*, **22**(11), 6005-6022.

Kugiumtzis, D. (2000). Surrogate data test for nonlinearity including nonmonotonic transforms. *Physical Review E*, **62**(1), R25.

Laio, F., Porporato, A., Revelli, R., & Ridolfi, L. (2003). A comparison of nonlinear flood forecasting methods. *Water Resources Research*, **39**(5).

Lall, U., Sangoyomi, T., & Abarbanel, H. D. (1996). Nonlinear dynamics of the Great Salt Lake: Nonparametric short-term forecasting. *Water Resources Research*, **32**(4), 975-985.

Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. Water Resources Research, 15 (5), 1055 – 1064.

Le Moine, N. (2008). Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit? (Doctoral dissertation, Doctorat Géosciences et Ressources Naturelles, Université Pierre et Marie Curie Paris VI).

Lidén, R., & Harlin, J. (2000). Analysis of conceptual rainfall–runoff modelling performance in different climates. *Journal of hydrology*, **238**(3-4), 231-247.

Lin, Y., & Mitchell, K. E. (2005, January). 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications. In Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA (Vol. 10).

Liu, Q., Islam, S., Rodriguez-Iturbe, I., & Le, Y. (1998). Phase-space analysis of daily streamflow: characterization and prediction. *Advances in water resources*, **21**(6), 463-475.

Liu, Q., Islam, S., Rodriguez-Iturbe, I., & Le, Y. (1998). Phase-space analysis of daily streamflow: characterization and prediction. *Advances in water resources*, **21**(6), 463-475.

Luo, L., Apps, D., Arcand, S., Xu, H., Pan, M., & Hoerling, M. (2017). Contribution of temperature and precipitation anomalies to the California drought during 2012–2015. Geophysical Research Letters, 44(7), 3184-3192.

Mackay, D. S., Ewers, B. E., Cook, B. D., & Davis, K. J. (2007). Environmental drivers of evapotranspiration in a shrub wetland and an upland forest in northern Wisconsin. *Water Resources Research*, **43**(3).

Marra, F. & Morin, E. (2015). Use of radar QPE for the derivation of Intensity-Duration-Frequency curves in a range of climatic regimes. Journal of Hydrology, 531, 427-440.

Marra, F., Morin, E., Peleg, N., Mei, Y., & Anagnostou, E. N. (2017). Intensity–duration–frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean. Hydrology and Earth System Sciences, 21(5), 2389-2404.

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association, 46(253), 68-78.

Matérn, B. (1960). Spatial variation. In Meddelanden fran Statens Skogsforskningsinstitute, 49 (5).

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., ... & Weiler, M. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resources Research, 43(7).

Mehran, A. & AghaKouchak, A. (2014). Capabilities of Satellite Precipitation datasets to Estimate Heavy Precipitation Rates at Different Temporal Accumulations. Hydrological Processes, 28, 2262-2270.

Miao, C., Ashouri, H., Hsu, K., Sorooshian S., & Duan, Q. (2015). Evaluation of the PERSIANN-CDR Daily Rainfall Estimates in Capturing the Behavior of Extreme Precipitation Events over China. *Journal of Hydrometeorogy*, 16, 1387–1396.

Milliman, J. D., Farnsworth, K. L., Jones, P. D., Xu, K. H., & Smith, L. C. (2008). Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. *Global and planetary change*, **62**(3-4), 187-194.

Mineo, C., Ridolfi, E., Napolitano, F., & Russo, F. (2018). The areal reduction factor: A new analytical expression for the Lazio Region in central Italy. Journal of Hydrology, 560, 471-479.

Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological sciences journal*, **41**(3), 399-417.

Moradkhani, H., Hsu, K. L., Gupta, H. V., & Sorooshian, S. (2004). Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *Journal of Hydrology*, **295**(1-4), 246-262.

Nester, T., Kirnbauer, R., Gutknecht, D., & Blöschl, G. (2011). Climate and catchment controls on the performance of regional flood simulations. *Journal of Hydrology*, **402**(3-4), 340-356.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., ... & Thorstensen, A. R., 2018. The PERSIANN family of global satellite precipitation data: A review and evaluation of products. Hydrol. Earth Syst. Sci. 22(11), 5801-5816.

Nguyen, P., Shearer, E. J., Ombadi, M., Gorooh, V. A., Hsu, K., Sorooshian, S., ... & Ralph, M., 2020. PERSIANN Dynamic Infrared–Rain Rate Model (PDIR) for High-Resolution, Real-Time Satellite Precipitation Estimation. Bull. Am. Meteorol. Soc. 101(3), E286-E302.

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2018). Developing intensity-duration-frequency (IDF) curves from satellite-based precipitation: Methodology and evaluation. Water Resources Research, 54(10), 7752-7766.

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7), e2020WR027251.

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021a). Complexity of Hydrologic Basins: A Chaotic Dynamics Perspective. *Journal of Hydrology*, 126222.

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021b). How much information on precipitation is contained in satellite infrared imagery?. Atmospheric Research, 105578.

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2021c). Retrospective Analysis and Bayesian Model Averaging of CMIP6 Precipitation in the Nile River Basin. Journal of Hydrometeorology, 22(1), 217-229.

Osborne, A. R., Kirwan Jr, A. D., Provenzale, A., & Bergamasco, L. (1986). A search for chaotic behavior in large and mesoscale motions in the Pacific Ocean. *Physica D: Nonlinear Phenomena*, **23**(1-3), 75-83.

Overeem, A., Buishand, A. & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. Journal of Hydrology, 348, 124-134.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical review letters*, **45**(9), 712.

Pande, S., & Moayeri, M. (2018). Hydrological interpretation of a statistical measure of basin complexity. *Water Resources Research*, **54**(10), 7403-7416.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins--Part 1: Runoff hydrograph studies. *Hydrology & Earth System Sciences Discussions*, **10**(1).

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. The Annals of Mathematical Statistics, 33, 1065-1076.

Pearl, J. & Verma, T. (1991). A theory of inferred causation. In Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, San Mateo, CA.

Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4), 669-688.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics surveys*, **3**, 96-146.

Pearl, J. (2009b). *Causality*. Cambridge university press.

Perica, S., Deborah, M., Pavlovich, S., Roy, I., St. Laurent, M., Trypaluk, C., et al. (2013a). NOAA Atlas 14 Volume 8 Version 2.0, Precipitation-Frequency Atlas of the United States, Midwestern States. Silver Spring, MD: NOAA, National Weather Service.

Perica, S., Deborah, M., Pavlovich, S., Roy, I., St. Laurent, M., Trypaluk, C., et al. (2013b). NOAA Atlas 14 Volume 9 Version 2.0, Precipitation-Frequency Atlas of the United States, Southeastern States. Silver Spring, MD: NOAA, National Weather Service.

Perica, S., Dietz, S., Heim, S., Hiner, L., Maitaraia, K., Martin, D., et al. (2011). NOAA Atlas 14 Volume 6 Version 2.0, Precipitation-Frequency Atlas of the United States, California. Silver Spring, MD: NOAA, National Weather Service.

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., ... & Friedlingstein, P. (2010). The impacts of climate change on water resources and agriculture in China. *Nature*, **467**(7311), 43.

Piechota, T. C., Chiew, F. H., Dracup, J. A., & McMahon, T. A. (1998). Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation. *Water resources research*, **34**(11), 3035-3044.

Porporato, A., & Ridolfi, L. (1996). Clues to the existence of deterministic chaos in river flow. *International Journal of modern physics B*, **10**(15), 1821-1862.

Porporato, A., & Ridolfi, L. (1997). Nonlinear analysis of river flow time sequences. *Water Resources Research*, **33**(6), 1353-1367.

Porporato, A., & Ridolfi, L. (2001). Multivariate nonlinear prediction of river flows. Journal of Hydrology, 248(1-4), 109-122.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., ... & Sabeti, P. C. (2011). Detecting novel associations in large data sets. science, 334(6062), 1518-1524.

Richards, F., & Arkin, P. (1981). On the relationship between satellite-observed cloud cover and precipitation. Monthly Weather Review, 109(5), 1081-1093.

Rodriguez-Iturbe, I. & Mejía, J. M. (1974). On the transformation of point rainfall to areal rainfall. Water Resources Research, 10(4), 729–735.

Rodriguez-Iturbe, I., Febres De Power, B., Sharifi, M. B., & Georgakakos, K. P. (1989). Chaos in rainfall. *Water Resources Research*, **25**(7), 1667-1675.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics, 27, 832-837.

Ruddell, B. L., & Kumar, P. (2009a). Ecohydrologic process networks: 1. Identification. *Water Resources Research*, **45**(3).

Ruddell, B. L., & Kumar, P. (2009b). Ecohydrologic process networks: 2. Analysis and characterization. *Water Resources Research*, **45**(3).

Runge, J., Petoukhov, V., & Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate*, **27**(2), 720-739.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., ... & Kurths, J. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications*, **6**, 8502.

Sassen, K., Wang, Z., & Liu, D. (2008). Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol lidar and infrared pathfinder satellite observations (CALIPSO) measurements. Journal of Geophysical Research: Atmospheres, 113(D8).

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. Journal of statistical Physics, 65(3-4), 579-616.

Schaake, J., Cong, S., & Duan, Q. (2006). US MOPEX data set (No. UCRL-JRNL-221228). Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).

Schaake, J., Cong, S., & Duan, Q. (2006). US MOPEX data set (No. UCRL-JRNL-221228). Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, **85**(2), 461.

Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical review letters*, **77**(4), 635.

Scofield, R. A., & Kuligowski, R. J. (2003). Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. Weather and Forecasting, 18(6), 1037-1051.

Seager, R., Hoerling, M., Schubert, S., Wang, H., Lyon, B., Kumar, A., ... & Henderson, N. (2015). Causes of the 2011–14 California drought. Journal of Climate, 28(18), 6997-7024.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, **27**(3), 379-423.

Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model. *Journal of Hydrology*, **239**(1-4), 249-258.

Sharma, A., & Mehrotra, R. (2014). An information theoretic alternative to model a natural system using observational information alone. *Water Resources Research*, **50**(1), 650-660.

Singh, V. P. (1997). The use of entropy in hydrology and water resources. Hydrological processes, 11(6), 587-626.

Sivakumar, B., & Singh, V. P. (2012). Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrology and Earth System Science,* **16**, *4119-4131*.

Sivakumar, B., Jayawardena, A. W., & Fernando, T. M. K. G. (2002). River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *Journal of hydrology*, **265**(1-4), 225-245.

Sivapalan, M. & Blöschl, G. (1998). Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. Journal of Hydrology, 204, 150-167.

Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological processes*, **17**(11), 2101-2111.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., ... & Oki, T. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological sciences journal*, **48**(6), 857-880.

Solomatine, D. P., & Dulal, K. N. (2003). Model trees as an alternative to neural networks in rainfall—runoff modelling. *Hydrological Sciences Journal*, **48**(3), 399-411.

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

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

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, **9**(1), 62-72.

Stout, J. S., Martin, D. W., & Sikdar, D. N. (1979). Estimating GATE rainfall with geosynchronous satellite images. Monthly Weather Review, 107(5), 585-598.

Sugihara, G. (1994). Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, **348**(1688), 477-495.

Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**(6268), 734.

Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *science*, **338**(6106), 496-500.

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980* (pp. 366-381). Springer, Berlin, Heidelberg.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, **58**(1-4), 77-94.

Thirumalaiah, K., & Deo, M. C. (1998). River stage forecasting using artificial neural networks. *Journal of Hydrologic Engineering*, **3**(1), 26-32.

Tokar, A. S., & Johnson, P. A. (1999). Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*, **4**(3), 232-239.

U.S. Weather Bureau. (1957). Rainfall Intensity-Frequency regime, 1, The Ohio valley, Tech. Pap. 29. Washington, DC: US Department of Commerce.

U.S. Weather Bureau. (1958). Rainfall Intensity-Frequency regime, 2, Southeastern United States, Tech. Pap. 29. Washington, DC: US Department of Commerce.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins-Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, **17**(6), 2263.

Vignesh, R., Jothiprakash, V., & Sivakumar, B. (2015). Streamflow variability and classification using false nearest neighbor method. Journal of Hydrology, 531, 706-715.

Vrugt, J. A., Bouten, W., Dekker, S. C., & Musters, P. A. (2002). Transpiration dynamics of an Austrian Pine stand and its forest floor: identifying controlling conditions using artificial neural networks. *Advances in water resources*, **25**(3), 293-303.

Wallis, J. R. (1982). Hydrologic problems associated with oilshale development. In S. Rinaldi (Eds.), Environmental Systems and Management: Proceedings of the IFIP WG 7.1 Working Conference on Environmental Systems Analysis and Management (pp. 85 – 102). Amsterdam, Netherlands: Elsevier Science Ltd.

Wang, E., Zhang, Y., Luo, J., Chiew, F. H., & Wang, Q. J. (2011). Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. *Water Resources Research*, **47**(5).

Wang, H., Masters, S., Edwards, M. A., Falkinham III, J. O., & Pruden, A. (2014). Effect of disinfectant, water age, and pipe materials on bacterial and eukaryotic community structure in drinking water biofilm. Environmental science & technology, 48(3), 1426-1435.

Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z., & Duan, W. (2018). Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. *Scientific reports*, **8**(1), 12171.

Wark, D. Q., G. Yamamoto, and J. Lienesch, 1962. Infrared flux and surface temperature determinations from TIROS radiometer measurements. Meteor. Sat. Lab. Rep. 10, Weather Bureau, Washington, D.C., 84 pp. [Available from NOAA/NESDIS, Room 703, 5200 Auth Rd., Camp Springs, MD 20746.]

Westra, S., Sharma, A., Brown, C., & Lall, U. (2008). Multivariate streamflow forecasting using independent component analysis. *Water Resources Research*, **44**(2).

Whitney, H. (1944). The self-intersections of a smooth n-manifold in 2n-space. Annals of Mathematics, 220-246.

Woolhiser, D. A. (2011). Precipitation: stochastic properties. In *Encyclopedia of Water Science, Second Edition. Taylor and Francis, New York*, 906-908.

Yapo, P., Sorooshian, S., & Gupta, V. (1993). A Markov chain flow model for flood forecasting. *Water resources research*, **29**(7), 2427-2436.

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., & Stewart, A. (2008). Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resources Research, 44(5).

Ye, H., & Sugihara, G. (2016). Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science*, **353**(6302), 922-925. https://doi.org/10.1126/science.aag0863

Zalenski, G., Krajewski, W. F., Quintero, F., Restrepo, P., & Buan, S. (2017). Analysis of national weather service stage forecast errors. Weather and Forecasting, 32(4), 1441-1465.

Appendix

Table A.1 Table of datasets used in the dissertation.

Dataset	***Spatial Coverage or Location used in the study***	***Dissertation Chapter***	***Published Article***	***Link to dataset repository***
FLUXNET	US Santa Rita, Arizona FLUXNET site (US-SRM)	Chapter 2	Ombadi et al. (2020a)	**[https]**
Hydrologic Bucket Model Simulations	-	Chapter 2	Ombadi et al. (2020a)	**[https]**
The Model Parameter Estimation Project (MOPEX)	All Sites	Chapter 3 and Chapter 4	Ombadi et al. (2021a)	**[https]**
NCEP STAGE IV	Contiguous United States (CONUS)	Chapter 5	Ombadi et al. (2021b)	**[https]**
NCEP IR Data	CONUS	Chapter 5	Ombadi et al. (2021b)	**[https]**
PERSIANN-CDR	CONUS	Chapter 6	Ombadi et al. (2018)	**[https]**
CPC Unified Gauge-Based Analysis of Daily Precipitation	CONUS	Chapter 6	Ombadi et al. (2018)	**[https]**
NOAA Atlas 14	CONUS	Chapter 6	Ombadi et al. (2018)	**[https]**
Digital Elevation Model	CONUS	Chapter 6	Ombadi et al. (2018)	**[https]**
CMIP6 (20 models; historical and SSP585)	The Nile River Basin	Chapter 7	Ombadi et al. (2021c)	**[https]**
PERSIANN-CDR	The Nile River Basin	Chapter 7	Ombadi et al. (2021c)	**[https]**

Table A.2 Table of datasets generated in the dissertation.

Dataset	*Spatial Coverage or Location used in the study*	*Dissertation Chapter*	*Published Article*	*Link to dataset repository*
Hydrologic Bucket Model Simulations	-	Chapter 2	Ombadi et al. (2020a)	**[https]**
Daily IDF curves for return periods of 2, 5, 10, 25, 50 and 100 years.	CONUS	Chapter 6	Ombadi et al. (2018)	**[https]**
Complexity Indices for MOPEX hydrologic basins	CONUS	Chapter 3	Ombadi et al. (2021a)	**[https]**

Table A.3 List of symbols and notations used in chapter 2.

Symbol	Definition
$\boldsymbol{X} \Longrightarrow \boldsymbol{Y}$	Variable $\boldsymbol{X}$ causes $\boldsymbol{Y}$; also used to denote causal links in the graph
l	Length of time series
$\boldsymbol{\mathcal{H}}$	Matrix of time series observations (size = number of variables $*\ l$)
$\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$	System variables
$\boldsymbol{X_t}$, $\boldsymbol{Y_t}$ and $\boldsymbol{Z_t}$	A single observation at time t for the variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
c_{xxk}	Regression coefficient of $\boldsymbol{X_t}$ regressed on $\boldsymbol{X_{t-k}}$ in the unrestricted model
$\grave{c}_{xxk}$	Regression coefficient of $\boldsymbol{X_t}$ regressed on $\boldsymbol{X_{t-k}}$ in the restricted model
c_{xzk}	Regression coefficient of $\boldsymbol{X_t}$ regressed on $\boldsymbol{Z_{t-k}}$ in the unrestricted model
$\grave{c}_{xzk}$	Regression coefficient of $\boldsymbol{X_t}$ regressed on $\boldsymbol{Z_{t-k}}$ in the restricted model
c_{xyk}	Regression coefficient of $\boldsymbol{X_t}$ regressed on $\boldsymbol{Y_{t-k}}$
ε_{xt} and $\grave{\varepsilon}_{xt}$	Residuals in the unrestricted and restricted regression models, respectively
α	Significance level
p	Model order
$\boldsymbol{H}(\boldsymbol{Y} \mid \boldsymbol{Z})$	Entropy of variable $\boldsymbol{Y}$ given information about variable $\boldsymbol{Z}$
$\mathbb{I}(\boldsymbol{X};\boldsymbol{Y} \mid \boldsymbol{Z})$	Conditional mutual information between variables ($\boldsymbol{X}$, $\boldsymbol{Y}$) conditioned on $\boldsymbol{Z}$
f	Probability density function
$\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$	Single realizations of the variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
$\overline{\boldsymbol{\mathcal{H}}}_{t-1}$	Observations matrix as of time $t-1$
$\overline{\boldsymbol{X}}_{\boldsymbol{t-1}}$, $\overline{\boldsymbol{Y}}_{\boldsymbol{t-1}}$ **and** $\overline{\boldsymbol{Z}}_{\boldsymbol{t-1}}$	Time series of variables $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ respectively as of time $t-1$
$\boldsymbol{ne}(\overline{\boldsymbol{X}}_{\boldsymbol{t}})$	Set of nodes connected to $\overline{\boldsymbol{X}}_{\boldsymbol{t}}$ in a causal graph
$\mathbb{S}$	A subset of the nodes $\boldsymbol{ne}(\overline{\boldsymbol{X}}_{\boldsymbol{t}})$
$\boldsymbol{\mathcal{M}}$	A chaotic dynamical system
$\boldsymbol{\mathcal{M}_t}$	Representation of system $\boldsymbol{\mathcal{M}}$ in the phase space at time **t**
$\boldsymbol{\mathcal{M}_{xt}}$	Representation of system $\boldsymbol{\mathcal{M}}$ at time t in the shadow manifold of $\boldsymbol{X}$
$\boldsymbol{\mathcal{M}_{yt}}$	Representation of system $\boldsymbol{\mathcal{M}}$ at time t in the shadow manifold of $\boldsymbol{Y}$
$\boldsymbol{E}$	Embedding dimension in the CCM algorithm
$\boldsymbol{R}$	Rainfall
$\boldsymbol{S}$	Soil moisture
$\boldsymbol{I}$	Interflow
$\boldsymbol{Q}$	Runoff
S_{max}	Maximum soil storage
K_s	Storage-discharge parameter 1
δ	Storage-discharge parameter 2
ξ	Storage-discharge parameter 3
$\tilde{p}_{i,j}$	Transition probability from the state i to the state j
$Beta_{\alpha}, Beta_{\beta}$	Beta distribution parameters
$\grave{\boldsymbol{R}}$	A binary variable indicating whether the day is rainy or not
$\boldsymbol{Y}$	Simulated rainfall amount
$\boldsymbol{\eta_s}$, $\boldsymbol{\eta_I}$ and $\boldsymbol{\eta_Q}$	Red noise for the hydrological model equations.
r	Autocorrelation parameter of the red noise
$\boldsymbol{\omega}$	White gaussian noise with variance σ^2
$\boldsymbol{ET}$	Evapotranspiration
$\boldsymbol{R_n}$	Net radiation
$\boldsymbol{VPD}$	Vapor pressure deficit
$\boldsymbol{SWC}$	Soil water content
$\boldsymbol{T_a}$	Air temperature
$\boldsymbol{T_s}$	Soil temperature

Table A.4 List of symbols and notations used in chapter 3.

Symbol	Definition
m	Number of variables in a dynamical system.
X_k	Time series of the variable *K*.
$x_k(t)$	Single observation of the variable *K* at time *t*.
$\boldsymbol{x}(t)$	Single observation in the phase space at time *t*.
E	Embedding dimension
τ	Delay time
T	Prediction horizon
f_T	Function that the dynamical system state from current time to *T* steps ahead
$\widehat{f_T}$	An approximation of the function f_T
l	Length of time series
E_{opt}	Optimum embedding dimension
ρ	Correlation coefficient
$\widetilde{\boldsymbol{x}}(t)$	An augmentation of the vector $\boldsymbol{x}(t)$
c	Coefficients in the S-map equation
$\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$	Matrix representation of the S-map method
w	A weighting factor function
d	Euclidean distance in the phase space
$\bar{d}$	Average of Euclidean distances in the phase space
θ	Nonlinearity index
θ_{opt}	Optimum nonlinearity index
Q	A discriminating metric for a chaotic dynamical system
μ_Q	Mean value of the discriminating metric computed from surrogates
σ_Q	Standard deviation value of the discriminating metric computed from surrogates
r_s	Spearman correlation coefficient
D_d	Drainage density
DI	Dryness Index
PSI	Precipitation seasonality index
QSI	Streamflow seasonality index
λ	Synchroneity of precipitation and potential evapotranspiration
f_s	Soil infiltration capacity

Table A.5 List of symbols and notations used in chapter 4**.**

Symbol	Definition
n	Length of time series
$\boldsymbol{P}$	Precipitation time series [n ×1]
$\boldsymbol{PET}$	Potential Evapotranspiration time series [n ×1]
$\boldsymbol{T}$	Daily maximum temperature time series [n ×1]
$\boldsymbol{Q}$	Streamflow time series [n ×1]
t	Time
P_t, PET_t, T_t and Q_t	A single observation at time $\boldsymbol{t}$ for the variables $\boldsymbol{P}$, $\boldsymbol{PET}$, $\boldsymbol{T}$ and $\boldsymbol{Q}$
Y_t	Forecasted streamflow at time $\boldsymbol{t}$
τ	Lag time
τ^*	Maximum lag time
DI	Dryness Index
$\boldsymbol{g}$	General forecast function
E	Embedding Dimension
E_{opt}	Optimum Embedding Dimension
$\mathbb{E}_0$	Initial embedding coordinates used in the algorithm
τ_p	Prediction horizon
B	Maximum number of iterations in selection of embedding coordinates
ρ	Pearson correlation coefficient
NSE	Nash-Sutcliffe efficiency
$RMSE$	Root mean squared error
$\boldsymbol{W}$	LSTM input weight matrices
$\boldsymbol{U}$	LSTM hidden weight matrices
$\boldsymbol{b}$	LSTM bias vector
$\boldsymbol{i}$	LSTM input gate
$\boldsymbol{f}$	LSTM forget gate
$\boldsymbol{o}$	LSTM output gate
$\boldsymbol{c}$	LSTM cell state
$\boldsymbol{h}$	LSTM hidden state
$\sigma(.)$	Sigmoid function
$tanh(.)$	Hyperbolic tangent function

Table A.6 List of symbols and notations used in chapter 5.

Symbol	Definition
$H(X)$	Shannon Entropy of a random variable X
$f_x(x)$	Marginal distribution of the random variable X
$f_y(y)$	Marginal distribution of the random variable Y
$f_{xy}(x,y)$	Joint distribution of the two random variables X and Y
$MI(X,Y)$	Mutual information of the two random variables X and Y
MIC	Maximal Information Coefficient
G	A grid defined for the computation of MIC.
g_X	Number of bins (partitions) in the horizontal axis.
g_y	Number of bins (partitions) in the vertical axis.
MI_G	Maximal Information Coefficient computed from the grid G.
n	Sample size.
P	Precipitation
T_B	Infrared (IR) Brightness Temperature
c	Coefficient in the range [0, 1] used to include gaussian noise.
$\Delta\mathrm{t}$	Temporal scale of aggregation.
Δs	Spatial scale of aggregation.
ρ	Pearson correlation coefficient
$\bar{\rho}$	Normalized Pearson correlation coefficient
$\overline{MIC}$	Normalized Maximal Information Coefficient

Table A.7 List of symbols and notations used in chapter 6.

Symbol	Definition
ζ	Adjustment factor for a single event in the annual maximum series of precipitation
$R_{G\,(x,y,k)}$	Kth event of the annual maximum series of ground-based rainfall at location (x, y)
$R_{S\,(x,y,k)}$	Kth event of the annual maximum series of satellite-based rainfall at location (x, y)
$\bar{\zeta}_{(x,y)}$	Mean value of adjustment factor for location (x,y)
α and β	Parameters of the bias adjustment model
$E_{(x,y)}$	Elevation in meters of the grid (x, y)
ρ	Pearson correlation coefficient
r	Euclidean distance between two points
λ	Parameter that specifies decay in correlation with distance
$\bar{\bar{r}}_A$	The characteristic distance
A	Square area used in the computation of $\bar{\bar{r}}_A$
κ^2	Variance reduction factor
$f_R(r)$	Probability density function of the random variable r
μ_p	Location parameter of GEV for point rainfall
α_p	Scale parameter of GEV for point rainfall
μ_A	Location parameter of GEV for areal rainfall
α_A	Scale parameter of GEV for areal rainfall
$q_{(x,y)}$	Quantile of extreme precipitation at location (x, y)
$\mu_{(x,y)}$	Mean value of annual maximum series at location (x, y)
$\hat{q}$	Regional growth factor for homogenous regions
$IDF_{PERSIANN-CDR}$	IDF curves estimated from PERSIANN-CDR
$IDF_{NOAA-Atlas14}$	IDF curves estimated from NOAA Atlas 14

Table A.8 List of symbols and notations used in chapter 7.

Symbol	Definition
$\boldsymbol{y}^{\boldsymbol{H}}$	PERSIANN-CDR annual (or monthly) precipitation time series
$\boldsymbol{f}_{\boldsymbol{k}}^{\boldsymbol{H}}$	Annual (or monthly) precipitation time series from the $\boldsymbol{k}$th GCM model
a and b	Bias adjustment coefficients
$\boldsymbol{y}$	Precipitation projections
α and β	Parameters of the bias adjustment model
$p_k(\boldsymbol{y}\|\boldsymbol{f_k})$	Conditional probability distribution of y given simulations of the model K
w_k	Weights of the Bayesian Model Averaging (BMA)
σ^2	Variance of the normal distribution assumed for conditional pdfs of p_k
$\mathcal{L}$	Log likelihood function of the BMA model
n	Number of years
T	Number of generations in the optimization algorithm
N	Number of chains in the optimization algorithm

Table A.9 CMIP6 models, and their spatial resolution, used in the study presented in chapter 7.

Model	Institute	Resolution (Lat° x Lon°)
Earth3	EC-Earth-Consortium, Europe	0.702 x 0.703 *
Earth3-Veg	EC-Earth-Consortium, Europe	0.702 x 0.703 *
MPI-ESM1-2-HR	Max Planck Institute for Meteorology, Germany	0.935 x 0.9375 *
CESM2-WACCM	National Center for Atmospheric Research (NCAR), USA	0.942 x 1.25
FIO-ESM-2-0	First Institute of Oceanography-Qingdao National Laboratory for Marine Science and Technology (FIO-QLNM), China	0.942 x 1.25
NorESM2-MM	NorESM Climate modeling Consortium (NCC), Norway	0.942 x 1.25
FGOALS-f3-L	Chinese Academy of Sciences, China	1 x 1.25
BCC-CSM2-MR	Beijing Climate Center, China	1.121 x 1.125 *
MIROC6	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, The University of Tokyo, National Institute for Environmental Studies, and RIKEN Center for Computational Science (MIROC), Japan	1.4 x 1.406 *
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation-Australian Research Council Centre of Excellence for Climate System Science (CSIRO-ARCCSS), Australia	1.25 x 1.875
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Australia	1.25 x 1.875
KAGE-1-0-G	National Institute of Meteorological Sciences/Korea Meteorological Administration (NIMS-KMA), Republic of Korea	1.25 x 1.875
INM-CM4-8	Institute for Numerical Mathematics, Russia	1.5 x 2
INM-CM5-0	Institute for Numerical Mathematics, Russia	1.5 x 2
IPSL-CM6A-LR	Institut Pierre Simon Laplace, France	1.268 x 2.5
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Germany	1.865 x 1.875 *
NESM3	Nanjing University of Information Science and Technology, China	1.865 x 1.875 *
FGOALS-g3	Chinese Academy of Sciences, China	2.279 x 2 *
NorESM2-LM	NorESM Climate modeling Consortium (NCC), Norway	1.895 x 2.5
CanESM5	Canadian Centre for Climate Modelling and Analysis, Canada	2.789 x 2.813 *

***** Approximate resolution since the native resolution is not in regular grids.