UNIVERSITY OF CALIFORNIA,
IRVINE

# Framework for Analyzing Compound and Inter-Related Extremes

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Omid Mazdiyasni

Dissertation Committee:
Associate Professor Amir AghaKouchak, Chair
Professor Soroosh Sorooshian
Associate Professor Steven J. Davis

2018

# DEDICATION

To

My best friend and lovely wife,
Shirin

# TABLE OF CONTENTS

## LIST OF FIGURES

# ACKNOWLEDGMENTS

I would also like to thank my committee members: Professor Amir AghaKouchak, Professor Soroosh Sorooshian, Professor Kenneth Lamb, and Professor Steve Davis. Professor Davis inspired many great ideas in several of my papers, and was very helpful in providing informative feedback.
Professor AghaKouchak has been more than a PhD advisor to me during my time as a PhD student. He is an exceptional mentor, always giving and helping to the best of his ability, and providing his time and energy to help his students succeed. I am proud to call him my advising professor, research mentor, and close friend.
Professor Sorooshian is an exceptional professor and mentor. He was one of the main motivators and reasons I pursued my PhD, and I am forever grateful to him for encouraging me to continue on this path.
I would like to express my appreciation to my colleagues in the Center for Hydrometeorology and Remote Sensing, especially Ata, Matin, Pouya, Iman and Hassan. I would also like to thank Professor Foufoula-Georgiou and her research group member.
I would also like to thank Professor Lamb for initiating my interest in water resources and hydrology. Professor Lamb was my hydrology and hydraulics professor during my undergraduate studies, and inspired me to pursue topics in water resources throughout my graduate studies.
Two brilliant, insightful and hard working colleagues of mine, Mojtaba and Hamed deserve great credit in showing me how hard work pays off. They are great collaborators, and I hope to continue working with them and having a close professional and personal relationship with them.
I would like to thank Alireza who provided great feedback on his experiences as a PhD student and postdoc. He is a close friend of mine, and I cherished my time with him as a colleague.
Two extremely special people are my Italian friends Elisa and Alessandro. Ale has always had a positive attitude, and is very enjoyable to be around. I will especially cherish the times we played basketball together. I remember I was the first member of the research to meet Eli outside of our office in Engineering Tower. I enjoyed being colleagues with her, as well as building a friendship. Getting to know her and spending time with her have been especially rewarding, and I wish her success in her future career.
My close friend and colleague Pouya Vahmani has always shown interest in my work, and encouraged me during the tough times. I am grateful to him for his help during my PhD.
I would like to thank Mohsen Niknejad who I met during my time as a PhD student. He is one of my closest friends and a great colleague. We worked together on a patent, and it was always a pleasure working on and improving ideas with him. I hope to continue to work with him.
My colleague and best man, Dr. Mehran has truly helped me throughout my professional career and personal growth. I appreciate everything Dr. Mehran has done for me. I look forward to continuing to share ideas and work with him.
I would like to thank my mother, Mandana Pishdadi, my father, Parviz Mazdiyasni, and my brother, Arman Mazdiyasni for their constant encouragement and unlimited love. I appreciate everything they have done for me, and I am certain I would not be where I am today without their love and devotion. Thank you maman o baba.
Most of all, I would like to thank my beautiful, loving, caring, compassionate, fun, and encouraging wife, Shirin. Shirin is the greatest thing to have happened to me, and I cannot ever portray in words how she has helped me grow and succeed. Shirin is always a source of encouragement and compassion, and I hope I can be as good to her as she is to me. ❤

# Curriculum Vitae

**Omid Mazdiyasni, EIT**
University of California, Irvine
(949) 648-0672
omazdiya@uci.edu

**Education**

***Doctor of Philosophy***
University: University of California, Irvine
Major: Civil and Environmental Engineering, 2018

***Master of Science***
University: University of California, Irvine
Major: Civil and Environmental Engineering, 2015

***Bachelor of Science***
University: California State Polytechnic University, Pomona
Major: Civil Engineering, 2013

**Publications (Citations: 323):**

1. Chiang, Felicia, **Omid Mazdiyasni**, Amir AghaKouchak. "Amplified climate warming under drought conditions in observations and model simulations" *Science Advances* – In Press

2. Sadegh, Mojtaba, Hamed Moftakhari, Hoshin V. Gupta, Elisa Ragno**, Omid Mazdiyasni**, Brett Sanders, Richard Matthew, and Amir AghaKouchak. "Multi-hazard scenarios for analysis of compound extreme events." *Geophysical Research Letters*.

3. **Mazdiyasni, Omid**, et al. "Increasing probability of mortality during Indian heat waves." *Science Advances* 3.6 (2017): e1700066.

4. Moftakhari, H., AghaKouchak, A., Sanders, B. F., Matthew, R. A., **Mazdiyasni, O.** (2017). Translating Uncertain Sea Level Projections Into Infrastructure Impacts Using a Bayesian Framework. *Geophysical Research Letters*.

5. Moftakhari, Hamed, Amir AghaKouchak, Brett F Sanders, Richard A Matthew, **Omid Mazdiyasni**. "Translating Uncertain Sea Level Projections into Infrastructure Impacts Using a Bayesian Framework." *Geophysical Research Letters* (2017).

6. Safavi, Hamid R., Vahid Raghibi, **Omid Mazdiyasni**, Mohammad Mortazavi-Naeini. "A new hybrid drought-monitoring framework based on nonparametric standardized indicators." *Hydrology Research* (2017): nh2017266.

7. **Mazdiyasni, Omid**, and Amir AghaKouchak. "Substantial increase in concurrent droughts and heatwaves in the United States." *Proceedings of the National Academy of Sciences* 112.37 (2015): 11484-11489.

8. Mehran, Ali, **Omid Mazdiyasni**, and Amir AghaKouchak. "A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand." *Journal of Geophysical Research: Atmospheres* 120.15 (2015): 7520-7533.

9. AghaKouchak, Amir, LinYin Cheng, **Omid Mazdiyasni**, and Alireza Farahmand "Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought." *Geophysical Research Letters* 41.24 (2014): 8847-8852.

10. Golian, Saeed, **Omid Mazdiyasni**, and Amir AghaKouchak. "Trends in meteorological and agricultural droughts in Iran." *Theoretical and Applied Climatology* 119.3-4 (2014): 679-688.

**Patent**

1. AghaKouchak, A., Niknejad, M., **Mazdiyasni, O.**, & Momtaz, F. (2018). U.S. Patent Application No. 15/806,143.

**Under Review/In Revision**

1. **Mazdiyasni, Omid**, et al. "Empirical Teleconnections: A Data-Driven Approach for Improving Seasonal Forecasting" – *Science Advances*

2. **Mazdiyasni, Omid**, et al. "Heat wave Intensity Duration Frequency Curve: A Multivariate Approach for Risk and Attribution Analysis" – *Geophysical Research Letters*

3. Anjileli, Hassan, Hamed Moftakhari, **Omid Mazdiyasni**, Hamid Norouzi, Batool Ashraf, Alireza Farahmand, P. Bowler, M. Azarderakhsh, Travis E. Huxman, Amir AghaKouchak. "Analyzing high frequency soil respiration using a probabilistic model in a semi-arid Mediterranean climate – *Journal of Geophysical Research: Biogeosciences*

4. Sadegh, Mojtaba, Hamed Moftakhari, Amir AghaKouchak, Hoshin V. Gupta, Elisa Ragno, **Omid Mazdiyasni**, Richard Matthew, Brett Sanders. "Multi-Hazard scenarios for analysis of compound extreme events: Multivariate risk analysis, design and uncertainty assessment" – *Geophysical Research Letters*

5. Ehsan Raei, Mohammad Reza Nikoo, Amir AghaKouchak, Omid Mazdiyasni, and Mojtaba Sadegh. "GHWR: A Multi-Method Global Heatwave and Warm-Spell Record" – *Nature Scientific Data*

*6.* Alborzi, Aneseh; Mirchi, Ali; Moftakhari, Hamed; Mallakpour, Iman; Alian, Sara; Nazemi, Ali; Hassanzadeh, Elmira; **Mazdiyasni, Omid;** Ashraf, Samaneh; Madani, Kaveh; Norouzi, Hamidreza; Azarderakhsh, Marzi ; Mehran, Ali ; Sadegh, Mojtaba ; Castelletti, Andrea; AghaKouchak, Amir. "Climate-Informed Environmental Inflows to Revive a Drying Lake Facing Meteorological and Anthropogenic Droughts" – *Environmental Research Letters*

7. Ashraf, Samaneh, et al. "The Compounding Effects of Human Activities and Climatic Changes on Surface Water Availability in Western Iran" – *Geophysical Research Letters*

**Involved Research Proposals**

1. National Science Foundation I-Corps Proposal (Major Contributor – Prepared first draft) - Funded

2. Terrestrial Hydrology (NNH17ZDA001N-THP) – NASA proposal (Major Contributor) - Pending

3. MAPP - Advancing drought understanding, monitoring and prediction (NOAA Drought Task Force) - Declined

**Fellowships, Awards, and Certificates:**

1. International Institute for Applied Systems Analysis (IIASA) YSSP Fellowship

2. Competitive Edge Research Fellowship

3. JPL Center for Climate Sciences Summer School

4. Fundamentals of Engineering (FE)

5. Certificate of Copula for Hydrology and Climate Applications

**Noteworthy Presentations:**

1. GEWEX: Extremes and Water on the Edge Presentation - Heatwaves and Heat-Related Mortality in India (2018)

2. AGU Presentation – Heatwaves and Heat-Related Mortality in India (2017)

3. I-Corps Presentation – Weather Augmented Risk Determination System (2017)

4. AGU Presentation – Heatwaves and Heat-Related Mortality in India (2016)

5. IIASA Presentation –Increasing heatwave exposure in India (2016)

6. AGU Presentation – Changes in Concurrent Droughts and Heatwaves in the United States (2015)

7. AGU Chapman – Risk of Concurrent Climate Extremes: Insights from the 2014 California Drought (2015)

8. AGU Oral Presentation – Changes in Concurrent Droughts and Heatwaves in the United States (2014)

9. JPL Center for Climate Sciences Summer Presentation – Changes in Concurrent Droughts and Heatwaves in the United States (2014)

10. Competitive Edge Presentation – Assessing Vulnerability to Droughts (2013)

**Professional Memberships and Affiliations:**

1. UNESCO Regional Centre on Urban Water Management (RCUWM) for the West and Central Asia (2015)

2. American Geophysical Union (2013)

3. American Meteorological Society (2013)
4. American Society of Civil Engineers (2012)

**Journal Reviewer:**

1. Climatic Change
2. Journal of Geophysical Research
3. Journal of Hydrology
4. Sustainable Cities and Society
5. Journal of Hydrology

**Work Experience:**

**UNESCO Regional Centre on Urban Water Management (RCUWM) for the West and Central Asia: August 2015 – Present –** Reviewing and updating written requests and proposals for RCUWM. Consulting RCUWM director Dr. Chavoshian in project direction and description. Reviewing professional letters to other UNESCO affiliates.

**University of California, Irvine: June 2013 – Present –** Researching changes and impacts of droughts and heatwaves. Integrating unique statistical approaches to assess changes and impacts of droughts and heatwaves. Assisting advisor with courses and related work. Utilizing available research tools and consulting with experts in the field to successfully develop a model that portrays the vulnerability and resilience of certain regions to droughts.

**International Institute for Applied Systems Analysis (IIASA): May 2016 – September 2016 –** Determining changes in heatwave occurrences and exposure from 2010 - 2090. Analyzing the impacts on human population, and potential health hazards.

**Santa Ana Watershed Project Authority: June 2012 – June 2013-** Updating several engineering drawings per as-builds. Performing cost estimates, and providing supporting data to managers and clients. Creating budget sheets for manager's review. Creating monthly utility sheets. Recording data of existing pipe information.

**Southern California Edison: June 2011 – September 2011-** Researching data and validating record retention schedule to federal/State/Local Regulations and company policies. Developing the architecture and writing the proposal for an electronic version of the site EH&S compliance files. Assisting in performing Safety and Environmental facility assessments to verify compliance with company policies. Writing multiple articles for the company Safety Focus newsletter. Supporting analysts in data analysis on day to day projects. Attending monthly safety meetings with company management.

**Transportation and Energy Solutions: November 2009 – June 2011-** Preparing preliminary and final plans and specifications for city construction projects, including street projects and traffic signals. Spent over 1,000 hours doing field/intersection review, topographic data collection, fieldwork, equipment checks, and conduit checks. Designing Traffic Signal and Signing & Striping modifications to existing intersections. Performed cost estimates and budget analyses for different projects.

**Sumac Technologies: June 2009 – November 2009**- Managing merchant account and analysis performance data. Responsible for listing items on-line on over 20 different websites. In charge of monitoring the inventory and shipping hundreds of items daily. Customer service representative.

**Attended Meetings:**
GEWEX: Extremes and Water on the Edge (2018)
AGU Fall Meeting (2017)
NSF I-Corps Meeting (2017)
AGU Fall Meeting (2016)
AGU Fall Meeting (2015)
US-Iran Symposium on Climate Change: Impacts and Mitigation (2015)
AGU Chapman (2015)
AGU Fall Meeting (2014)
Western States Water Council meeting (2014), NASA JPL
Copula for Hydrology and Climate Application course (2014), IAHS
PMA Conference (2013)
Commodity Classics (2014)
JPL Center for Climate Sciences Summer School (September 2014)

# ABSTRACT OF THE DISSERTATION

Framework for Analyzing Compound and Inter-related Extremes

By

Omid Mazdiyasni

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2018

Professor Amir AghaKouchak, Chair

Extreme climatic events have significant impacts on society and the environment, especially when multiple hazards occur concurrently (e.g., drought and heat waves) or consecutively (wildfires and extreme temperatures). A large number of indicators have been developed to detect and study changes in extreme events across space and time. While the current climate extreme indicators provide useful information, most do not provide any information on compound/concurrent events. A compound event corresponds to a situation in which multiple (often interrelated) hazard drivers lead to an extreme outcome. Therefore, current univariate methods used for frequency analysis and risk assessment may underestimate the risk or occurrence probability of extreme events. After a comprehensive review of the existing methods, this study outlines frameworks for detecting, modeling, and analyzing inter-related events and processes including compound extremes.

# INTRODUCTION

Climate extremes can profoundly affect society, natural systems, and infrastructure [e.g., Handmer et al., 2012; Rocklöv and Forsberg, 2009; Easterling et al., 2000]. In many cases, extreme events are interconnected and their compounding effects intensify the impacts, as is the case with droughts and heatwaves [e.g., Aghakouchak et al., 2014, Mazdiyasni and AghaKouchak, 2015, Zscheischler & Seneviratne, 2017], coastal and pluvial flooding [e.g., Zheng et al., 2014; Wahl et al., 2015], and coastal and fluvial flooding [e.g., Moftakhari et al., 2017, lamb et al., 2010]. For instance, the 2014 California, 2012 central United States, 2010 Russian, 2003 European droughts and heat waves events are a few recent extreme events that were associated with a combination of extreme high temperatures and low precipitation. Another examples are the hurricanes/super-storms frequently hitting the east and gulf coast of United States (e.g. Katina 2005, Sandy 2012, and Harvey 2017) during which low atmospheric pressure events give rise to the coastal ocean water level and at the same time trigger intense rainfall in the coastal watershed.

Concurrent extreme events can have a multiplier effect, often having a strong dependence, and intensifying the risk to society and the environment, while the individual variables may not necessarily pose a significant or severe threat. Therefore, historical/projected changes in concurrent extremes cannot be investigated without accounting for interdependence between multiple variables or successive occurrences of the same variable where temporal dependencies exist. If the same event (such as a storm surge) affects multiple sites at the same time (or with a short time lag) univariate analysis approaches fail to take the intersite-dependencies into account for appropriate risk assessment and planning. Furthermore, large

impacts can also be driven by an unfortunate combination of non-extreme drivers, whose combined occurrence lead to system failure. This means, analyses based on one variable or event may not be sufficient to characterize and model extreme events that have multiple drivers with spatiotemporal dependencies.

All the above examples can be summarized under the notion of compound events. Compound events refer to a combination of multiple drivers and/or hazards that contribute to societal or environmental risks [Zscheischler et al., 2018]. IPCC report defines compound events as [Field et al., 2012]: "(1) two or more extreme events occurring simultaneously or successively; (2) combinations of extreme events with underlying conditions that amplify the impact of the events; (3) combinations of events that are not themselves extremes but lead to an extreme event or impact when combined."

Following this definition, rather than quantifying extremes in drivers or a range of hazardous impacts of one driver, a set or combination of drivers that result in a specific hazardous situation should be identified. Idier et al. [2013] first proposed the idea of an inverse approach to start from a threshold level (referred to as risk acceptability) to finally obtain the return period of this threshold. Such approaches would allow identification of all of the drivers of a critical impact. Defining a critical threshold level, however, is often challenging. For example, flood height is an indicator used to integrate the effects of storm surge, precipitation extremes, soil moisture and others. In contrast, for hot and dry extremes, typically no indicators exists that integrates the adverse effects of dry and hot conditions on specific systems. It is therefore often unclear how to weigh the importance of temperature against dryness [Mazdyasni &

AghaKouchak, 2015] and strong assumptions have to be made with respect to which part of the temperature-precipitation distribution is to be studied [Zscheischler & Seneviratne, 2017].

Current univariate indices and methods to detect extreme events can be mainly categorized into three types: (1) indices based on daily, monthly, and/or annual maxima/minima of a certain climate variable (i.e. temperature or precipitation); (2) indices based on duration of an extreme event (i.e. number of consecutive days in a year above a relative extreme threshold, such as temperatures above the 90th percentile of the long-term climatology), and (3) indices based on duration and/or frequency of an absolute extreme exceedance threshold (i.e. annual number of frost days (minimum temperature below 0 °C) or number of consecutive frost days per year). In the univariate case, extreme value analysis has been extensively used to study the magnitude of extreme events and the frequency of their occurrences based on either block maxima or peak over threshold approaches [e.g., Lang et al., 1999; Katz et al., 2002; Katz, 2010]. Many software packages have been developed for the implementation of these practices [Gilleland et al., 2013; Cheng et al., 2014; Gilleland and Katz, 2016 ].

However, modeling extreme events with multiple drivers (variables) is not as straightforward. Understanding the risks associated with compound extremes at present and estimating their future occurrences, requires more complicated, rigorous and robust methods. Such methods must be able to systematically take the correlation structures between variables into account and sufficiently deal with the inherently small number of extremes. In this study, we review statistical methods that may be used for modeling compound extremes. In the following chapters we first explore non-parametric empirical methods that have already proposed in the literature. Then moving onto the parametric category of methods we explain the theory and

implications of multivariate extreme approaches, copula analysis techniques, extremal dependence and conditional extreme methods, techniques for space-time extreme analysis, and finally the implication of Bayesian modeling.

# Empirical methods

Empirical (also known as nonparametric) methods have been widely used in univariate extreme value analysis [Citation]. Empirical methods are often used to determine and identify data behavior. Most empirical univariate methods can be extended for multivariate analysis and studying compound events. In the following, we review empirical methods that can be used for studying compound extremes in hydrology and climate.

**Empirical counting methods**

Simultaneous extremes can be studied by empirically counting the simultaneous occurrence of two or more events at different periods. This approach requires an empirical threshold for defining extremes of each variable. The exceedances (or non-exceedances) above (or below) the predefined threshold will be used for studying compound extremes. However, defining the term "*extreme*" and characterizing a multivariate threshold above (or below) which events are considered to be extreme is not a straightforward task (Zheng et al., 2014). The following four definitions for defining extreme events have been proposed in literature, which are not necessarily specific to compound events: (1) component-wise block maxima (Tawn et al., 1989) that consider the maximum values in each block, (2) threshold-exceedances (Resnick, 1987) that characterize an event as extreme if all of the contributing variables exceed a given

threshold, (3) point processes (Coles and Tawn, 1994) that transform data into radial and angular components and define the extremes as events beyond the radial threshold, and (4) conditional extreme events (Heffernan and Tawn, 2004), where the distribution of each variable is conditioned on threshold exceedances in other variables and vice versa.

[Beniston, 2009] investigated the trends in joint quantiles of temperature (below 25$^{th}$ percentile) and precipitation (above 75th percentile) extremes in Europe by counting joint extremes in precipitation and temperature records. [Fischer and Knutti, 2012] studied combined humidity and temperature extremes in the future by considering global wet-bulb temperatures during the hottest days (top 1%). [Hao et al., 2013] examined concurrent precipitation and temperature extremes globally based on the empirical occurrence of compound extremes. [Mazdiyasni and AghaKouchak, 2015] studied the changes in concurrent droughts and heatwaves from 1960 - 2010. Their results indicated that although there was a hiatus in rising temperatures, and drought occurrence did not have a statistically significant trend univariately, concurrent drought and heatwave events did experience a statistically significant increase across the United States. Figure 1 shows that for the heatwave threshold of 7 consecutive days above the 95th percentile of daily summer temperatures and a drought threshold of D0 or higher (more than 0.5 standard deviations below the mean), there is nearly a 100% increase in concurrent drought and heatwave events across the majority of the contiguous United States (CONUS). For extreme thresholds, it is difficult to obtain a large sample of compound extremes which may affect reliability of the analysis. In this approach, it is important to check the sample size for different thresholds to ensure the data is sufficient for

reliable statistical analysis [Beniston, 2009; Hao et al., 2013; Mazdiyasni and AghaKouchak, 2015].

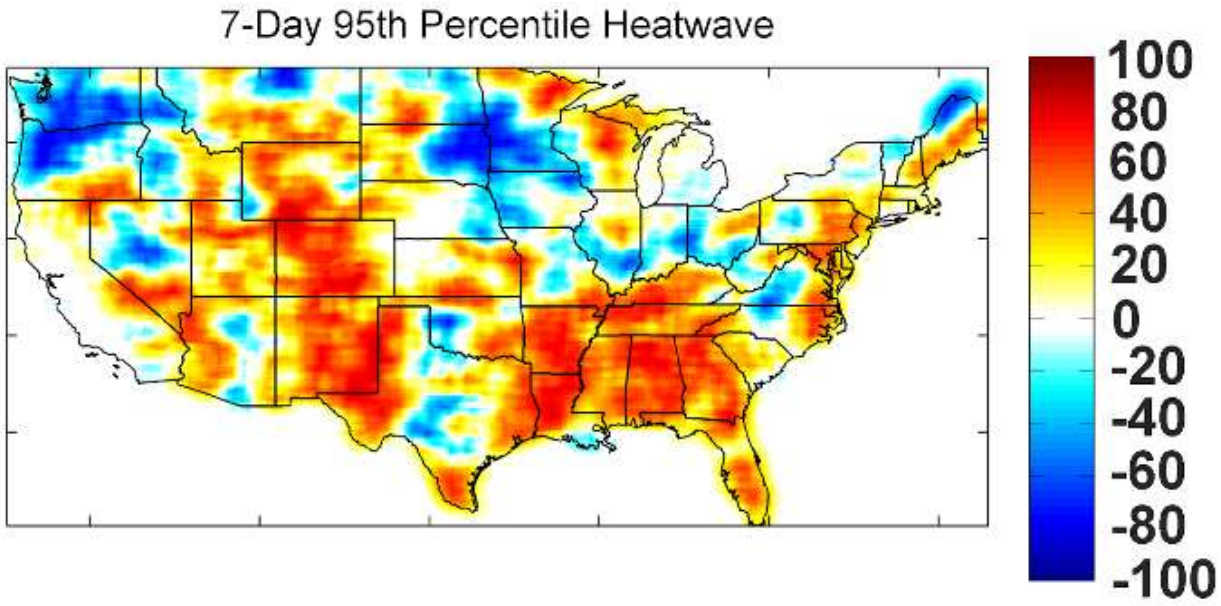



Figure 1: This figure shows the percent change in concurrent drought and heat waves during 1990 - 201 relative to 1960-1980. This figure shows the percent change of a 7-day heatwave above the 95th percentile of summer maximum temperatures occurring during 1-month D0 droughts (on the D0-D4 scale).

**Multivariate Index**

The characteristics of multiple or compound extremes can be summarized by an indicator, which can be defined as a combination of two or more variables. Then, this index is studied with univariate statistical methods. One such example is the Climate Extreme Index (CEI), which

incorporates the average of many types of extremes [Karl et al., 1996; Gleason et al., 2008] including drought, precipitation, and temperature extremes. [Gallant and Karoly, 2010] proposed a combined climate extreme index to investigate concurrent temperature and rainfall extremes. The heat index, defined by [Steadman, 1979; Steadman, 1984], is based on temperature and relative humidity and might also be considered an example when studying concurrent extreme [Perkins and Alexander, 2012]. Similarly, the Multivariate Standardized Drought Index (MSDI) is an integrated drought index that can be used to characterize meteorological and agricultural droughts (extremes) [Hao and AghaKouchak, 2013]. This index combines precipitation and soil moisture information to account for both the onset and persistence of drought.

**Structure variable**

The structure variable method is similar to the multivariate index method, but it focuses on the system response/behavior to a certain combination of underlying variables of interest. The structure variable can be defined [Coles, 2001]:

$$Z = \phi(Mx, My)$$

where, Φ is the response/behavior function and $M_x$ and $M_y$ are extremes of variables *X* and *Y* such as: max($M_x$, $M_y$), min ($M_x$,$M_y$), $M_x$+$M_y$ (sum of minima/maxima) and $M_x \times M_y$ (product of minima/maxima). The variable Z can then be modeled with the standard univariate extreme value distribution for inferences such as estimating the return level. However, with this method, the justification for the generalized extreme value (GEV) distribution is not strong, since other combinations of the variables *X* and *Y* may generate larger values from the function

Φ than that from the component maxima [Coles, 2001]. One example comparing structure variable and the bivariate/multivariate extremes based on the Fremantle and Port Pirie annual maxima sea-level data is given by [Coles, 2001]. Further limitations of this method are discussed in (Hawkes, 2008 and Neal et al, 2013).

**Influence diagram**

The influence diagram is a generalization of the Bayesian framework that addresses both probabilistic inference and decision making process. [Leonard et al., 2013] defined the compound event by extreme impacts and proposed the influence diagram as a general framework to define, map, analyze, model, and communicate the risk of such events. This approach formalizes the process of identifying the impact-dependence variables (and events) as well as the dependence structure between variables. It also describes the model structure necessary for risk evaluation.

# Multivariate Extreme Value Theory

Multivariate extreme value theory is used to parametrically model the probability of an extreme event (i.e. coastal flooding) due to multiple drivers (i.e. surge and wave height). Parametric methods are used to model the behavior of variables, allowing to infer values not observed in the dataset.

**Extreme value distribution and extreme dependence**

Consider $X_{1,\ldots},X_n$ to be independent and identically distributed (IID) random variables. Denote the maximum of the random variable as max($X_{1,\ldots},X_n$ ) with marginal distribution function *F*. We seek suitable parameters $a_n$ and $b_n$ such that as $n\rightarrow\infty$ so we have:

$$\Pr\left(\frac{\max(X_1,\ldots,X_n)-b_n}{a_n}\le x\right)=F^n(a_nx+b_n)\rightarrow G(x)$$

It can be shown that the unique limiting distribution *G* for the block maxima is the Generalized Extreme Value (GEV) distribution:

$$G(x)=\exp\left[-\left(1+\frac{\varsigma(x-u)}{\sigma}\right)^{-1/\varsigma}\right]_{\varsigma\neq 0}$$

where *u*, *σ,* and *ς* are the location, scale, and shape parameters.

An alternative method to modeling extremes is to focus on the extremes above a certain threshold *u* (peak over threshold or POT). In this case, the unique limiting distribution G' of the random variable exceeding the large threshold *u* is the Generalized Pareto Distribution (GPD):

$$G'(x)=1-\varsigma\left[1+\frac{\varsigma(x-u)}{\sigma}\right]^{-1/\varsigma}\quad x>u$$

Unlike the univariate extreme values modeling in which the asymptotic form of the distribution of the block maxima or the peak over threshold can be derived theoretically, there is no finite dimension class of multivariate extreme value distribution *G*.

In the multivariate case, the dependence structure centers the modeling of the extremes. For the random variables *X* and *Y* with the transformed marginal *U* and *V*, the dependence between the extremes can be expressed as [Coles et al., 1999]:

$$\chi = \lim_{u \to 1} P(X > u \mid Y > u)$$

After mathematical manipulation, we can obtain:

$$\chi(u) = 2 - \frac{\log P(U < u, V < u)}{\log P(U < u)}$$ , and $$\chi = \lim_{u \to 1} \chi(u)$$

where χ(u) ranges from 0 (perfectly independent) to 1 (perfectly dependent), inclusive.

When χ=0, it means that the measure is not able to provide the relative dependence information. To overcome such a limitation, a second dependence measure of extremes is defined as [Coles et al., 1999]:

$$\bar{\chi}(u) = \frac{2 \log P(U > u)}{\log P(U > u, V > u)} - 1 = \frac{2 \log(1-u)}{\log \bar{C}(u,u)} - 1$$ , and $$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u)$$

where -1≤ $\bar{\chi}(u)$ ≤1 and -1≤ $\bar{\chi}$ ≤1.

To summarize, the pair (χ , $\bar{\chi}$ ) provides the summary of the extremal dependence with (χ >0, $\bar{\chi}$ =1) signifying the asymptotic dependence, in which case χ determines the dependence strength, and (χ =0, $\bar{\chi}$ <1) signifies asymptotic dependence, in which case $\bar{\chi}$ determines the dependence strength [Coles et al., 1999].

**Max-stable process and parametric model**

A distribution G is called ***max-stable*** for every positive integer *k* if one can find vectors $\alpha_k$ and $\beta_k$ such that the following equation holds [Beirlant et al., 2006]:

$$Gk(\alpha_k x + \beta_k) = G(x)$$

The class of the max-stable distribution function coincides with the extreme value distribution function and is in its own domain of attraction. The max-stable process is the stochastic process that all finite dimensional distribution functions are multivariate extreme value (MEV) distribution functions, which can be viewed as the infinite-dimensional generalization of the extreme value distribution and has been commonly used for modeling spatial extremes with finite dimensions [Beirlant et al., 2006].

Many types of the parametric subfamily have been proposed to provide a simpler representation of the MEV distribution and cover a wide range of dependence at the same time [Bortot and Gaetan, 2013]. Some parametric models for extremes include the logistic distribution [Tawn, 1988], negative logistic distribution [*Joe*, 1990], Gaussian [Smith, 1990], bilogistic [Joe et al., 1992], polynomial [Nadarajah, 1999] and Dirichlet models [Coles and Tawn, 1991]. A variety of parametric models for bivariate and multivariate extreme value distributions have been reviewed by [Kotz et al., 2000] and [Beirlant et al., 2006]. Generally these models are limited to low dimensions. [Cooley et al., 2010] proposed a pairwise beta distribution that is flexible to describe the extremes of random vectors of dimensions greater than two.

**Component-wise maxima**

In this section, we try to define the simultaneous extremes with the method of block maxima, which is commonly used in univariate extreme value analysis. We restrict our discussion to the bivariate case. In this context, it differs from the univariate extreme since the variable of interest is a vector. Extension of the univariate extreme value theory is not immediately straightforward since there is no natural ordering in higher dimension [Barnett, 1976; Tawn, 1988].

Generally, multivariate extreme value analysis is based on component-wise ordering. Consider $Z_i(X_i, Y_i)$ $(1 \le i \le n)$ to be IID vector and define: $MaxZ_i(1 \le i \le n) = (M_{1n}, M_{2n})$, where $M_{1n}$=max $X_i$ and $M_{2n}$=max $Y_i$. Denote the joint distribution of the random vector $(X,Y)$ as $F$. We seek constraints $a_{n1}$, $b_{n1}$, $a_{n2}$ and $b_{n2}$ , such that:

$$\Pr\left(\frac{M_{x,n} - b_{n,1}}{a_{n,1}} \le x, \frac{M_{y,n} - b_{n,2}}{a_{n,2}} \le y\right) = F_n(a_{n,1} + b_{n,1}, a_{n,2} + b_{n,2}) \to G(x_1, x_2)$$

If this holds for suitable choices of the constraints $a_n$ and $b_n$, $G$ is said to follow a multivariate extreme value distribution and $F$ is in the domain of attraction of $G$. To isolate the dependence from marginal distributional features and for technical convenience, the marginal distribution is standardly described by the unit Frechet distribution $\Phi$:

$\Phi(y) = \exp(-y^{-1})$ for y>0

Consider the general case where the multivariate extreme value distribution (MEVD), denoted as $G$, can be expressed as:

$$G(x) = \exp\left\{-l\left[-\log G_1(z_1), \ldots, -\log G_n(x_n)\right]\right\}$$

where $l$ is the stable tail dependence function.

The Pickands dependence function has been commonly used to describe the dependence structure in the bivariate case, which can be defined as [Pickands, 1981; Beirlant et al., 2006]:

$A(t) = l(1-t, t)$ or $l(t, 1-t) = A(1-t)$

It can be viewed as the stable tail dependence function restricted to the unit simplex. The Pickands dependence function uniquely determines the stable tail dependence function as follows:

$$l(v_1, v_2) = (v_1 + v_2)A\left(\frac{v_1}{v_1 + v_2}\right)$$

A max-stable distribution G can be determined by the marginal G1 and G2 and also the Pickands dependence function A as [Beirlant et al., 2006]:

$$G(x1, x2) = \exp\left[\log(G_1(x_1)G(x_2))A\left(\frac{\log(G_1(x_1)}{\log(G_1(x_1)G(x_2))}\right)\right]$$

**Threshold exceedance**

For the block maxima method, data are "wasted" for estimation of the multivariate distribution function. In addition, the component-wise maxima may not correspond to actual observations. The exceedance threshold method can be used to define and model the multivariate extreme with a relatively larger sample. In the exceedance method, we assume that *F* is in the domain of attraction of a multivariate extreme value distribution function G, with a dependence structure described by the stable tail dependence function *l* [Beirlant et al., 2006].

With marginal tails estimated with the GPD or GEV distribution, a parametric form of the tail dependence function *l* can be assumed, which can be estimated with the point-process method or censored-likelihood method [Beirlant et al., 2006]. For the nonparametric estimation, the empirical version of the approximation is used [Beirlant et al., 2006]:

$$F(x) \approx 1 - l\left[-\log F_1(z_1), \ldots, -\log F_d(z_d)\right]$$

Another way to model exceedance is with the multivariate regular variation method, for which the basic idea is that the joint tail decays as a power function [Resnick, 2002; Cooley et al., 2010; Ballani and Schlather, 2011]. For the bivariate case with a random vector $Z=(Z_1,Z_2)$, it is said to vary regularly if there is a sequence $a_n$, $n=1,\ldots,\infty$ such that on the Borel subset of $C=[0,\infty]\backslash 0$:[Weller et al., 2012],

$$nP\left(\frac{Z}{a_n}\in .\right)\xrightarrow{v} v(.)$$

where $\xrightarrow{v}$ denote the vague convergence. The measure v(.) has the scaling property:

$$v(ta)=t^{-\alpha}v(A)$$

where α is the tail index. By defining $R=||Z||$ and $W=Z||Z||^{-1}$, where $||.||$ is any norm on C, we can get the polar coordinate representation of the regular variation condition:

$$nP\left(\frac{R}{a_n}>R, W\in B\right)\xrightarrow{v} r^{-\alpha}H(B)$$

where *H* is the spectral measure that characterizes the tail dependence. Note that the stable tail dependence function in the previous sections can be linked to the spectral measure via [Guillotte et al., 2011]:

$$l(s,t)=2\int_0^1 \max(ws,(1-w)t)dH(w)$$

# Copulas

Copulas are one of the most common methods for modeling the joint distribution of multiple variables because they have the advantage of describing the correlation structure independently of the marginal distribution. Based on Sklar's theorem, the *d* - dimensional joint distribution function *F* of the random vector $X_1, X_2,\ldots,X_d$ can be expressed with the copula *C* as [Sklar, 1959; Joe, 1997]:

$$F(x_1, x_2, ..x_d) = C(F1(x_1), F2(x_2), ..., Fd(x_d))$$

where $F_1(x_1),\ldots,F_d(x_d)$ are the marginal distributions. The copula *C* is unique when the marginal distributions are continuous.

A variety of copula families, such as the Archimedean copula, have been commonly used in hydrology and water resources for frequency analysis [Kao and Govindaraju, 2008; Vandenberghe et al., 2011] , precipitation simulation [Bárdossy and Pegram, 2009; AghaKouchak et al., 2010a], and geo-statistical interpolation [Bárdossy, 2006]. Copulas have also been widely used for coastal flood modeling and prediction for to compounding effects of coastal ocean water level and freshwater discharge (Bevacqua et al., 2017; Moftakhari et al., 2017), coastal water level and waves (Serafin and Ruggiero, 2014; Wahl et al., 2016), storm surge and precipitation (Wahl et al., 2015), and wave/surge characteristics (Corbella and Stretch, 2012, 2013; De Michele et al., 2007; Salvadori et al., 2014, 2015; Wahl et al., 2012).

Sadegh et. al, 2017 recently developed the Multivariate Copula Analysis Toolbox (MvCAT) for the comprehensive analysis of the dependence structure between multiple variables. MvCAT

uses 26 different copula functions, with model complexities ranging from one to three parameters, allowing for the simulation of both symmetric and asymmetric dependence structures. Model parameters in MvCAT are tuned against observed/empirical joint probabilities using either (1) a gradient-based ‘‘interior-point’’ optimization algorithm which estimates Hessian through a dense quasi-Newton approximation for a rapid search of the feasible space to derive the optimal parameter set; or (2) a state-of-the-art hybrid-evolution Markov Chain Monte Carlo (MCMC) simulation algorithm that not only estimates the global optimum parameter set. The second method also approximates the posterior distribution of copula parameters for uncertainty analysis in a Bayesian framework. The latter approach not only avoids converging to the local optima (i.e. it identifies the global optima), but it also characterizes uncertainties with regard to the length of record. In fact, [Sadegh et. al, 2017] showed that relatively short observation periods, among others, translate into large uncertainties in the design/prediction/modeling outcomes that often render the management/policy inaccurate or insufficient.

There are a variety of parametric copula models and some such as the extreme value copula might be more appropriate for handling extremes [Salvadori et al., 2007; Salvadori and De Michele, 2010]. The extreme copula satisfies this condition [Galambos, 1987; Joe, 1997; Gudendorf and Segers, 2010; Davison et al., 2012]:

$$C({u_1}^m,..,{u_d}^m) \to C^m(u_1,...,u_d)$$

In the bivariate case, the extreme value copula can be represented by the Pickands dependence function [Pickands, 1981]. The bivariate copula C is an extreme value copula if there exists a Pickands function A such that:

$$C(u,v) = \exp\left\{\log(uv)A\left(\frac{\log(v)}{\log(uv)}\right)\right\}$$

where u and v are the marginal probabilities.

# Conditional extremes modeling

The extremal dependence method in the previous section enables the modeling of the asymptotic independence, but it is subject to limitations such as difficulty in identifying a suitable parameterization *L.* Furthermore, under asymptotic independence, the simultaneously large, observed components become increasingly unlikely as the dimension increases [Bortot and Gaetan, 2013]. However, in meteorology cases, extreme phenomenon also occurs when one variable is extreme while the other is not. In this section, we will introduce the conditional extreme modeling that is capable of resolving this problem [Heffernan and Tawn, 2004] and [Heffernan and Resnick, 2007] .

The conditional extreme mode is a semi-parametric regression model of the form [Heffernan and Tawn, 2004; Heffernan and Resnick, 2007]:

$Y_{-i} = a_i(y_i) + b_{|i}(y_i)Z_{|i}$ for $Y_i > u_{yi}$

where $a_{/i}$ ($y_i$) and $b_{/i}$ ($y_i$) are the location and scale parameters to be estimated with aЄ(0,1) and bЄ(-∞,1). $Z_{/i}$ is a random variable independent of $Y$ for $Y_i > u_{yi.}$ As $Y$ increases, Z converges to a distribution $G$. The estimation of $a$, $b$ and $G$ is based on regression.

The advantage of this conditional extreme model is that it can be applied whether the variables are asymptotically dependent or asymptotically independent. In other words, it provides flexibility that many of the above-described methods do not. It can also be employed to estimate the probability for any extreme set.

For climate events, even if one is around the mean while another is extreme, it may be considered a concurrent extreme event. In this case, the conditional extreme model may be appropriate and can be applied. Examples of the development and application of the conditional extreme analysis can be found in [Keef et al., 2009; Jonathan et al., 2010; Das and Resnick, 2011; Jonathan et al., 2012; Keef et al., 2012; Gilleland et al., 2013].

## Spatial extremes

Spatial extremes modeling is relevant to studying compound extremes involving one variable at different locations (e.g., extreme rainfall or flooding at multiple locations) [Bacro and Gaetan, 2012; Cooley et al., 2012; Davison et al., 2012]. Most spatial extremes are based on max-stable distributions with block maxima data [Cooley et al., 2012]. For this problem, three models including the latent variables, the extreme copula, and the max-stable process have been introduced with a practical example of extreme rainfall at 51 locations in Switzerland [Davison et al., 2012]. In addition, [Gaume et al., 2013] studied extreme snowfall at different locations with the max-stable process.

## Space-Time extremes

Multivariate extremes in a space-time setting have recently attracted a great deal of attention [Davis and Mikosch, 2008; Kabluchko et al., 2009; Huser and Davison, 2012; Davis et al., 2013 (2011)]. [Huser and Davison, 2012] proposed a framework for the space-time modeling of extreme events based on the max-stable process for hourly precipitation. Since the finite dimensional distribution functions are intractable, standard maximum likelihood procedures do not work and pairwise likelihood methods based only on the pairwise density can be implemented [Davis et al., 2013]. Following the approach described by [Davis et al., 2013 (2011)], in which the max-stable process introduced in [Kabluchko et al., 2009] is extended to a space–time setting, [Davis et al., 2013] investigated the statistical inference for max-stable space–time processes. [Chailan et al., 2017] employed a semiparametric approach to investigate extreme wave climate in the space-time domain by exploiting max-stable processes.

## Conditional Extremes

We can analyze the impacts of a change in one variable on a covariate using a multivariate distribution function. [Madiyasni et al., 2017] utilized the copula functions to define the conditional probability of the impact of an event (i.e., heat-related mortality) given a covariate (i.e., heat wave) using the joint probability distribution (Grimaldi et al., 2016; Graler et al., 2013; Salvadori et al., 2013, Mazdiyasni et al., 2017).

To determine the conditional probabilities of Y > y (mortality rates exceeding a threshold) at different (X = x1, x2,….), (i.e. FY|X(Y>y|X), one can develop the conditional probability density function:

$$f_{Y|X}(y|x) = c[F_X(x), F_Y(y)] * f_Y(y)$$

where c is the probability density function (PDF) and FY(y) is the mortality marginal distribution. Once we choose a certain x (i.e. summer mean temperature) conditional PDF from Eq. 6, the probability of Y (the mortality rates) exceeding a particular threshold (y) is given by the area under the curve: fY|X(y|x). This allows calculating conditional PDF fY|X(y|x) for different values of x (e.g., summer mean temperatures=27 °C or heatwave days=6 in Figure XX).

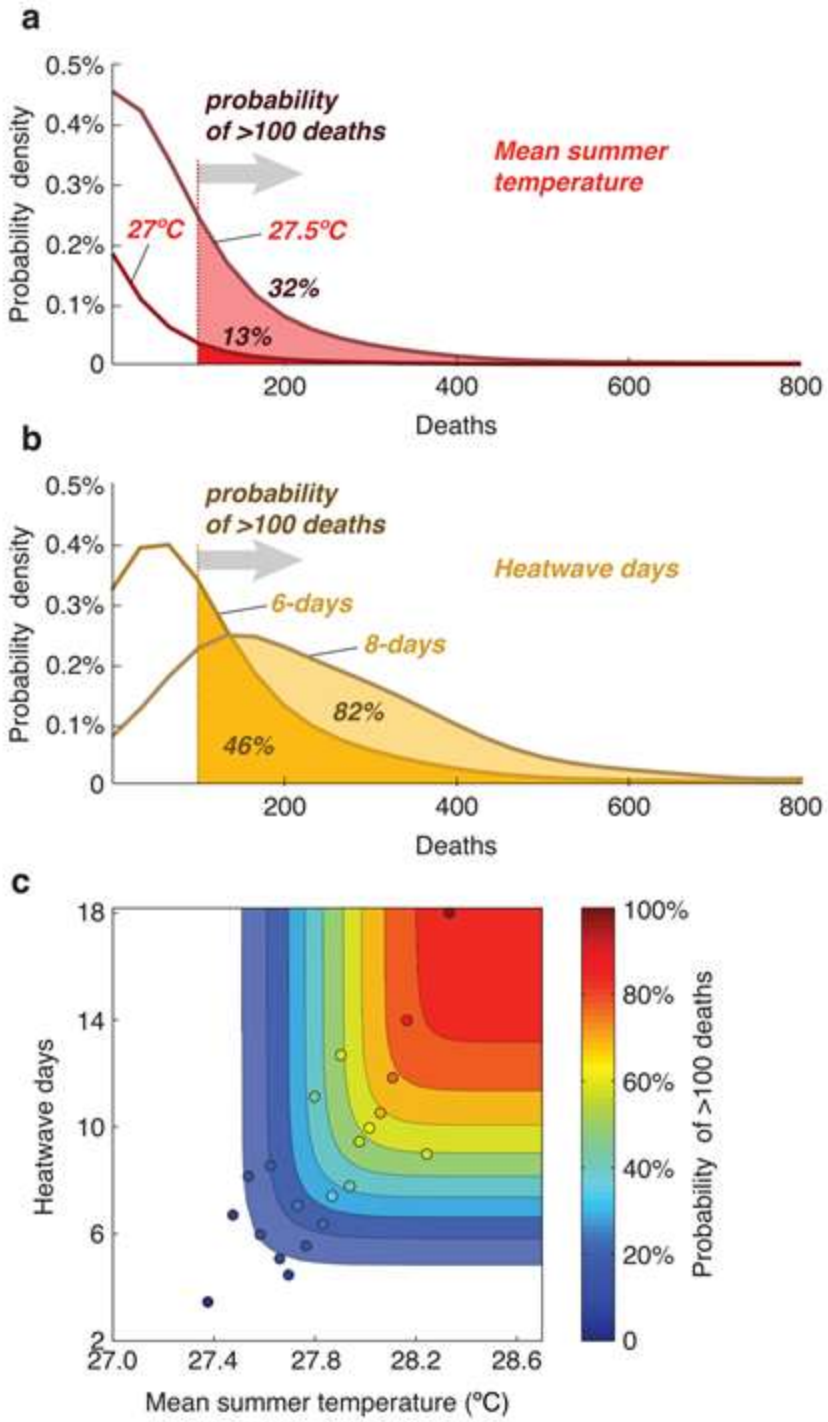



Figure 2: Probabilities of heatwave-caused mass-mortality events. Parametric conditional probability density functions for yearly mortality given certain thresholds for summer mean temperatures (a) and heatwave days (b). With 0.5 °C warmer mean temperatures or two more heatwave days per year, the probability of >100 heat-related deaths increase dramatically. The relationship between the two variables and probability of mass mortality events is shown in panel (c).

Figure 2 presents the results of a conditional probability density analysis of annual mortality given certain thresholds for summer mean temperatures and heatwaves. The shaded region represents the probability of mass heat-related mortality (i.e. heat-related deaths of more than 100 people), given different summer temperature values. For example, the Figure 3a shows that there is 13% probability that years with summer mean temperatures equal to 27 °C will have mass heat-related mortality. However, with an increase in summer mean temperatures of just 0.5 °C (to 27.5 °C), the probability of such levels of heat-related deaths jumps by a factor of 2.5 to 32%. Similarly, Figures 3b shows that the probability of heat-related mass mortality events increases from 46% to 82% (78% increase) where the average number of heatwave days across India shift from 6 days to 8 days, respectively. The substantial increase in mortality rates due to either a 0.5 °C increase in summer mean temperature or two more heatwave days suggests that future climate warming could have a relatively drastic human toll in India and similarly in developing tropical and subtropical countries.

# Discussion and concluding remarks

Extreme climate events have significant effects on society, infrastructure, the economy, and the environment. Many extreme events have multiple drivers such as droughts and heatwaves, and cannot be analyzed or detected using common univariate indices. These concurrent or compounding events often have significant ramifications due to their multivariate nature.

Compound events may be separated into three categories: (1) two or more extreme events occurring simultaneously or consecutively, (2) combinations of extreme events that exacerbate the impact of the events, (3) combination of events that are not extreme themselves, but have

extreme impacts when combined. However, analyzing and modeling such events is not as straightforward as univariate analyses. Robust statistical tools must be used to account for the interdependence between multiple drivers or successive occurrences of the same variable where temporal dependencies exist, to avoid significant biases.

Historical and projected changes in concurrent extremes cannot be investigated without accounting for interdependence between drivers. In this study, we review statistical methods that have been used to identify, analyze, and model compound extremes. We first explore empirical non-parametric methods, such as the counting method, multivariate indices, and structure variables. We then move onto theory and application of parametric methods, including copula analysis techniques, extremal dependence and conditional extreme methods, and techniques for space-time extreme analyses.

# Chapter 1

# Understanding Changes in Concurrent Extremes: Focus on Droughts and Heatwaves

The contents of Chapter 1 are published in the Proceedings of the National Academy of Sciences of the United States of America.

Citation: Mazdiyasni, Omid, and Amir AghaKouchak. "Substantial increase in concurrent droughts and heatwaves in the United States." *Proceedings of the National Academy of Sciences* 112.37 (2015): 11484-11489.

**A combination of climate events (e.g., low precipitation and high temperatures) may cause a significant impact on the ecosystem and society, though individual events involved may not be severe extremes themselves. Analyzing historical changes in concurrent climate extremes is critical to preparing for and mitigating the negative effects of climatic change and variability. This study focuses on the changes in concurrences of heatwaves and meteorological droughts from 1960 – 2010. Despite an apparent hiatus in rising temperature and no significant trend in droughts, I show a substantial increase in concurrent droughts and heatwaves across most parts of the United States, and a statistically significant shift in the distribution of concurrent extremes. While commonly used trend analysis methods do not show any trend in concurrent droughts and heatwaves, a unique statistical approach discussed in this study exhibits a statistically significant change in the distribution of the data.**

Heatwaves cause severe damage to society and the environment [*Easterling et al.*, 2000], with impacts on human health, air quality, and vegetation [*Ciais et al.*, 2005; *Vautard et al.*, 2005]. In 2003, for example, European countries faced an unprecedented heatwave, which in turn caused unusually high ozone concentrations **Error! Reference source not found.** and severe health problems, particularly in France where 15,000 extra deaths occurred **Error! Reference source not found.**[*Beniston*, 2004; *Seneviratne et al.*, 2012]. UNEP considers the European heatwave the world's most costly weather related disaster in 2003. Impacts were exacerbated because the region was in a drought[*Poumadere et al.*, 2005].

Heatwaves have a variety of direct, indirect, immediate, and delayed impacts, including higher water loss via evapotranspiration, lower yields of grains and other agricultural products[*Zaitchik et al.*, 2006], increased energy consumption, a decrease in efficiency of power plants[*Zamuda*

*et al.*, 2013], air pollution, and adverse effects on human health **Error! Reference source not found.Error! Reference source not found.**. Heatwaves have also contributed to an increase in the duration, size, and intensity of wildfires, causing economic losses and catastrophic environmental impacts **Error! Reference source not found.**.

Droughts also have pronounced impacts on society and the environment, such as significant reductions in gross primary productivity, leading to shortages in food production and increases in global food prices **Error! Reference source not found.**. The annual economic damage caused by droughts is estimated to be approximately $7 billion globally, with potential impacts on livestock, transportation by river, hydropower production, bioenergy, and energy consumption **Error! Reference source not found.**[*Greve et al.*, 2014; *Hao et al.*, 2014] [*Hao and AghaKouchak*, 2013] [*Wehner et al.*, 2011]**Error! Reference source not found.**.

Extreme climatic events can occur simultaneously, exacerbating environmental and societal impacts. Environmental hazards often result from a combination of climatic events **Error! Reference source not found.**[*AghaKouchak et al.*, 2014; *Seneviratne et al.*, 2006] over a range of spatial and temporal scales [*Hegerl et al.*, 2011; *Leonard et al.*, 2014]. A wildfire, for example, may occur on a hot, dry and windy day, although each of these individual conditions may not necessarily be extreme by themselves **Error! Reference source not found.**. In the Intergovernmental Panel on Climate Change (IPCC) special report on managing the risks of extreme events and disasters, the combination of multiple climate extreme events is termed a compound event **Error! Reference source not found.Error! Reference source not found.**. Most analyses of climate and weather extremes typically tend to focus on a single climatic condition; however, this univariate

approach may underestimate the effects of concurrent and compound extremes **Error! Reference source not found.**.

Sustained precipitation deficit in summer can be a contributory factor to hot summer days [*Mueller and Seneviratne*, 2012]. Heatwaves reduce the total energy transfer to the atmosphere, resulting in a decrease in convective precipitation **Error! Reference source not found.**. This in turn causes a soil-precipitation feedback loop that tends to extend or intensify drought conditions **Error! Reference source not found.**. The interaction between precipitation and temperature has been widely recognized in numerous studies[*Hao et al.*, 2013; *Shukla et al.*, 2015]. Heatwaves concurrent with droughts can intensify individual impacts of heatwaves or drought on society, the environment, and the global economy [*Perkins et al.*, 2012; *Shukla et al.*, 2015]. Studies suggest that changes in the relationship between precipitation and temperature may be more important than the changes in each of the variables individually **Error! Reference source not found.**[*Trenberth et al.*, 2014]. This study investigates changes in concurrent droughts and heatwaves in the United States using several different statistical techniques.

A heatwave is typically defined as a period of consecutive extremely hot days [*Perkins and Alexander*, 2013]**Error! Reference source not found.**, such as five consecutive days with temperature above the 90$^{th}$ percentile. Here, I use the 85$^{th}$, 90$^{th}$, and 95$^{th}$ percentiles of the warm season (May – October) temperature as extreme thresholds, and three heatwave durations (3, 5, and 7 days). A 5-day heatwave with a 90$^{th}$ percentile threshold is defined as five consecutive days with the maximum temperature exceeding the 90$^{th}$ percentile of the long-term climatology for that month. In this study, meteorological droughts are defined as precipitation deficits relative

to the climatology using the Standardized Precipitation Index (SPI)[*McKee et al.*, 1993]. Throughout this study, a drought is defined as an event that leads to SPI < -0.8 (approximately the 20$^{th}$ percentile precipitation). I use daily temperature and monthly precipitation information to identify historical droughts and heatwaves in the United States (see Data Section).

**Results**

We evaluated the changes in concurrent droughts and heatwaves during 1990-2010 relative to the baseline period 1960-1980. Figure 1 displays percent change in the occurrences of concurrent drought and heatwave events in each gridbox. Here, the percent change is based on the difference in the number of events in 1990-2010 relative to 1960-1980, divided by the total number of events. I present results for different durations (3-, 5-, and 7-day heatwaves) and extreme temperature thresholds (85$^{th}$, 90$^{th}$, and 95$^{th}$ percentiles). Figure 1 shows that the concurrence of all combinations of drought and heatwave intensities and durations have increased substantially in the south, southeast and parts of the western United States, and have decreased in parts of the Mid-West and northern United States. Notably, the longer and more severe (7-day 95$^{th}$ percentile) drought and heatwave concurrences have increased more than shorter, less severe concurrences (e.g., compare 7-day 95$^{th}$ percentile with the 3-day 85$^{th}$ percentile panels). This indicates that longer heatwaves (i.e., 7-day) have become more frequent in 1990-2010 compared to the shorter heatwaves (i.e., 3-day).

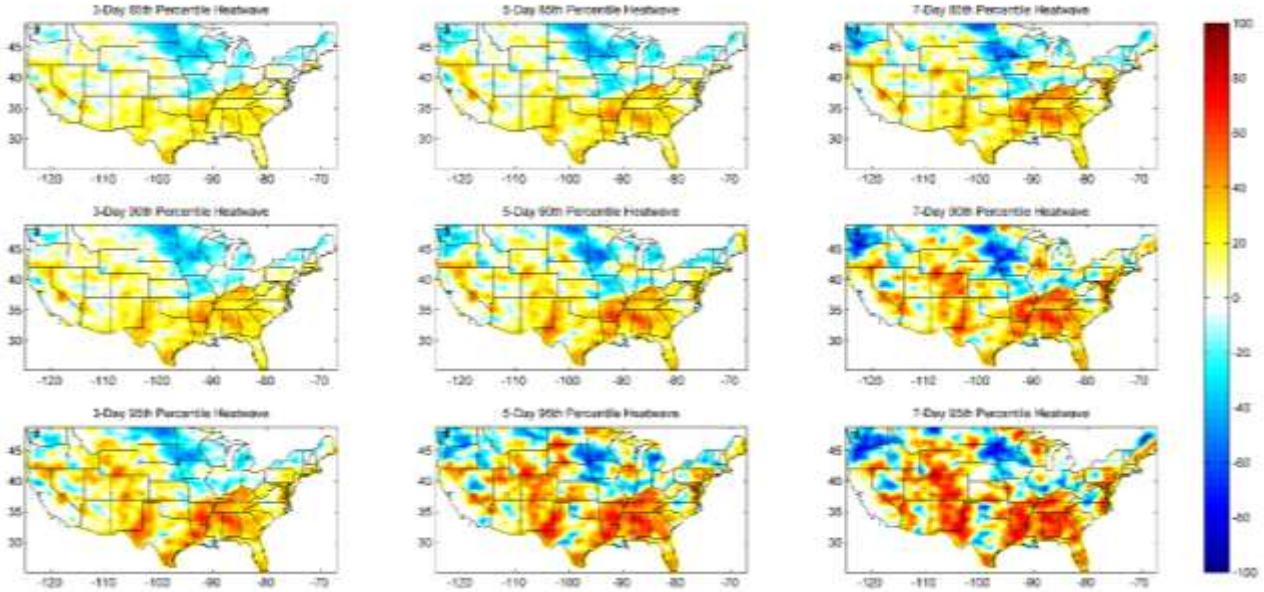


**Figure 1**| **Percent (%) change in concurrent droughts and heatwaves during 1990 – 2010 relative to 1960 – 1980 for each grid box**. The rows change in heatwave severity (85$^{th}$ percentile, 90$^{th}$ percentile, and 95$^{th}$ percentile), and the columns change in heatwave duration (3 d, 5 d, and 7 d).

Investigating the empirical cumulative distribution function (CDF) of the concurrent droughts and heatwaves reveals that a substantial change in extremes in from 1990 – 2010 relative to 1960 – 1980 (Figure 2). The x-axes represent the percent (%) of the contiguous United States in concurrent drought and heatwave. The y-axes show the corresponding cumulative probability. In each panel, the blue line is the CDF for the baseline period and the red line represents the CDF for 1990 – 2010. The CDF is based on data from the continental United States. As shown, for all intensities and durations during 1990 – 2010, the upper tail of the CDF has shifted to the right, indicating more extreme events in 1990-2010 relative to the baseline period (compare the red and blue lines in Figure 2). Notice that the shift is far more pronounced in the more extreme 7-day 95$^{th}$ percentile drought and heatwave concurrence as compared to other combinations. The two-sample Kolmogorov Smirnov test (Methods Section) confirms that the

CDFs of the concurrent droughts and heatwaves in the second period (1990-2010) are substantially different than those in the baseline period (1960-1980) at 0.05 significance level (95% confidence) for all heatwave durations and intensities except for 3-day, 85$^{th}$ percentile heatwaves (Table 1).

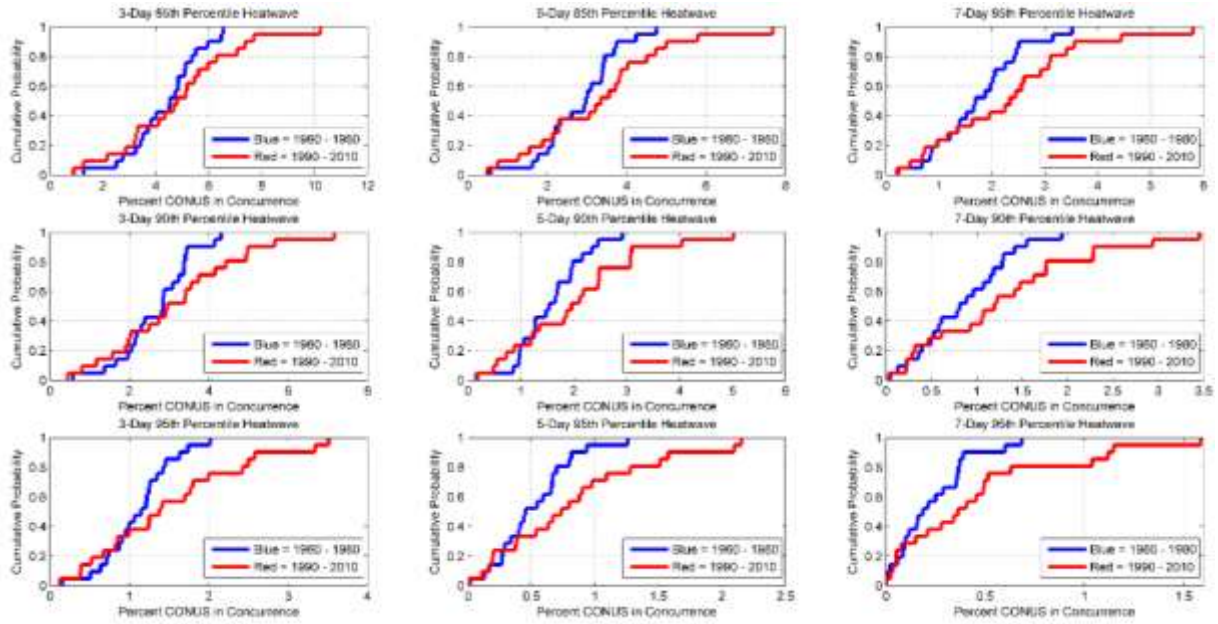



**Figure 2**| **The empirical CDF of drought and heatwave concurrences from 1960 to 1980 (blue) and 1990 to 2010 (red)**. The x axes represent the percent (%) of CONUS in concurrent drought and heatwave (See Appendix 1 Data Sources and Processing for more information on percent of CONUS). The rows change in heatwave severity (85$^{th}$ percentile, 90$^{th}$ percentile, and 95$^{th}$ percentile), and the columns change in heatwave duration (3 d, 5 d, and 7 d).

Past studies focused on changes in drought trends report conflicting results [*Damberg and AghaKouchak*, 2014; *Sheffield et al.*, 2012] **Error! Reference source not found.**[*Trenberth et al.*, 2014]. Here, I investigate the percent of the continental United States in concurrent droughts and heatwaves for different durations and intensities from 1960 – 2010 (Figure 3; see also Figure S1 in Appendix 1). For the 90$^{th}$ percentile threshold, the percent of the country in drought and heatwave can range between 6% (7-day heatwave) to 9.6% (3-day heatwave) – see the boxplots of the percent CONUS in concurrent drought and heatwave for all durations and

severities in Figure S4 in Appendix 1. While the CDFs clearly indicate changes in concurrent droughts and heatwaves, the commonly used Mann-Kendall trend test (see Methods Section) does not show a statistically significant trend (95% confidence) in the fraction of the continental United States under concurrent drought and heatwave (Table S1 in Appendix 1 provides the test statistics results). This can be attributed to limitations of statistical trend tests discussed in previous studies [*Clarke*, 2010] or lack of sensitive tools for change detection.

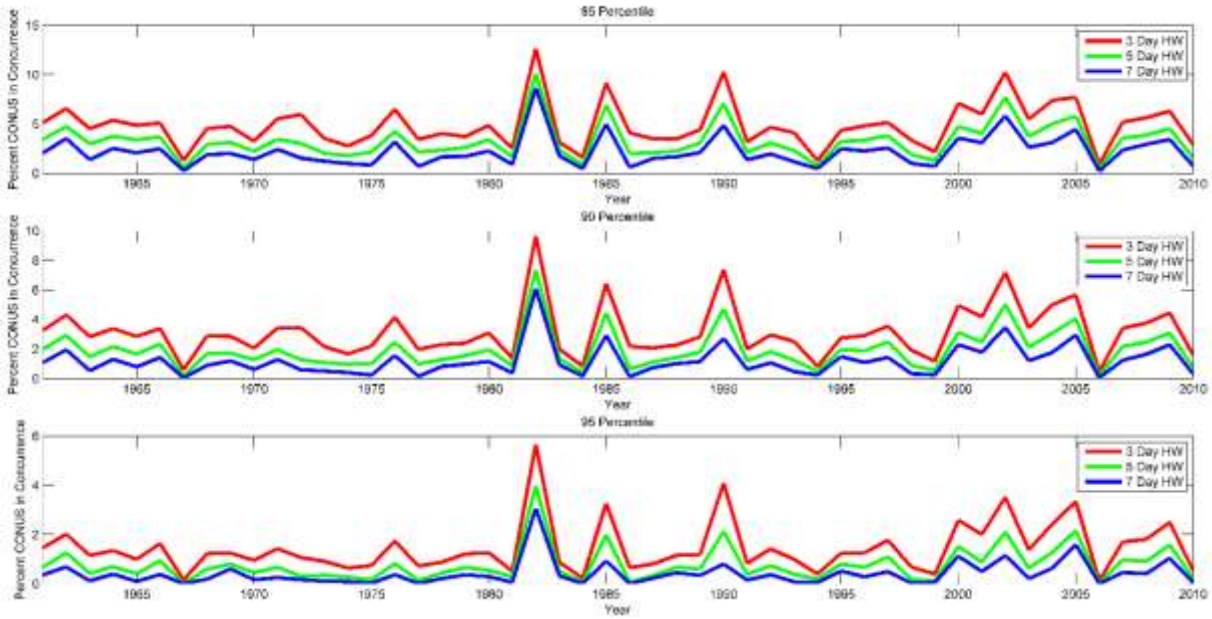



**Figure 3**| **Percent (%) of CONUS in concurrent drought and heatwave**. Percent (%) of CONUS in concurrent drought and heatwave from 1960 to 2010.

Here, I explore an approach based on the Cramér–von Mises change point detection test statistic (see Methods Section and Appendix 1) to investigate changes in concurrent droughts and heatwaves. I argue that this method is more sensitive to potential changes in time series and is well suited to investigate climate time series. This method, primarily used in economics and finance, evaluates different periods of data and determines statistically significant changes throughout time series. Figure 4 shows the Cramér–von Mises statistics for drought and

heatwave concurrences during 1960 –2010. The y-axes indicate a dimensionless measure of divergence between the empirical distributions of data before and after any given year as a continuous function (see Methods Section). For all plots, the maximum divergence occurs between 1998 and 1999, indicating substantial departure of the drought and heatwave CDFs before and after the red line in Figure 4. This information cannot be achieved from the commonly used trend analysis method or distribution change evaluation approaches. Recent reports suggest an apparent hiatus or so-called pause in global warming since 1999-2000.

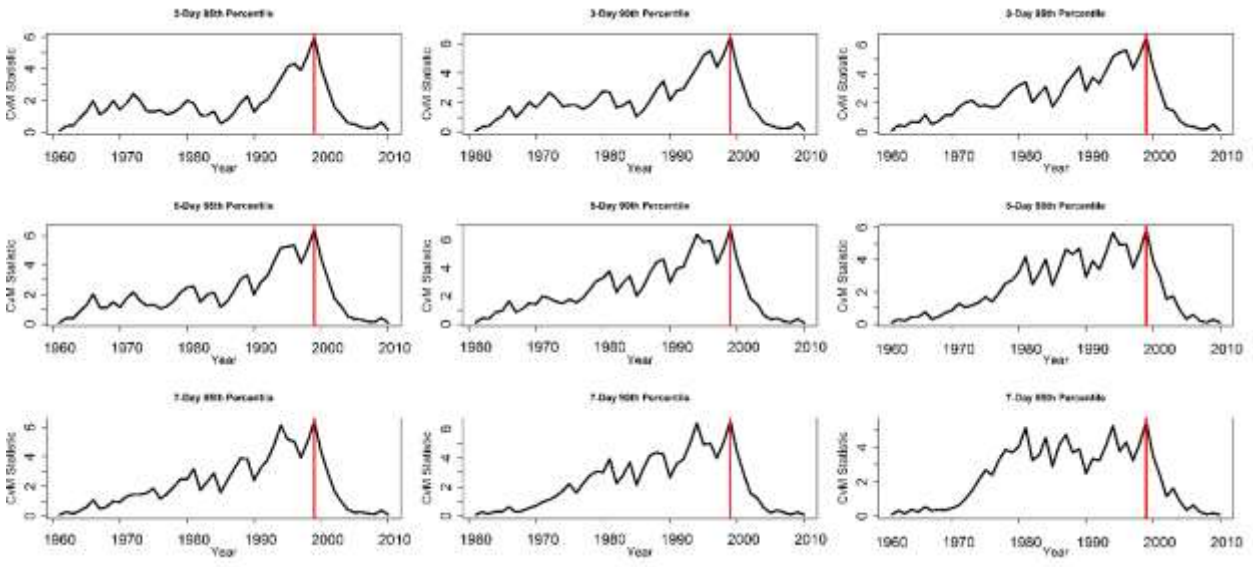


**Figure 4**| **The Cramer-von Mises (CvM) change point statistic from 1960 - 2010**. The rows change in heatwave severity (85$^{th}$ percentile, 90$^{th}$ percentile, and 95$^{th}$ percentile), and the columns change in heatwave duration (3 d, 5 d, and 7 d). The red lines indicate the point of maximum divergence between the distributions of concurrent drought and heatwave events.

Possible explanations include a long-lasting solar energy output minimum, low stratospheric water vapor, an increase in early 21$^{st}$ century volcanic activity, and a more frequent La Niña phase since the major El Niño event of 1997-1998[*Held*, 2013]. However, analyses show no pause in the occurrence of hot extremes over land since 1997 [*Seneviratne et al.*, 2014] or even

the mean global temperature[*Karl et al.*, 2015]. Rather, during the hiatus, exceedances of 30 extreme warm days per year have increased[*Seneviratne et al.*, 2014]. The results in Figures 2 and 4 indicate a statistically significant (at the 0.05 significance level) change in concurrent drought and heatwave events across many regions. Figure 2 indicates more extreme drought and heatwave concurrences in the latter two decades. This is consistent with the increase in extreme warm days during this period [*Seneviratne et al.*, 2014]**Error! Reference source not found.**. However, this conclusion cannot be reached using the commonly used statistical trend analysis techniques (e.g., Mann-Kendall trend test) used in hydrology and climate literature. Unlike the Mann-Kendall trend test, which investigates monotonic changes in the ranks of variables over time, the Cramér–von Mises test focuses on changes in the distributions of subsamples of the data. Typically, climatologists evaluate a certain period against a baseline. The Cramér–von Mises test is a flexible approach that allows investigators to examine different subsamples (e.g., projected and baseline periods) for potential distributional changes. The methodology outlined in this paper shows statistical changes in extremes beyond those achieved with commonly used methods.

**Data**

Both precipitation and temperature data sets are from the observation-based forcings developed for the North American Land Data Assimilation System (NLDAS) Variable Infiltration Capacity simulations over the contiguous United States [*Livneh et al.*, 2013; *Maurer et al.*, 2002]. Daily temperature and monthly precipitation data with a spatial resolution of 1/8° are used for detecting droughts and heatwaves (see Appendix 1 for more information on the data).

**Methods**

Here, the two-sample Kolmogorov Smirnov (KS) test assesses differences between the CDFs of the concurrent drought and heatwave events. KS is a nonparametric test that can evaluate two distribution functions (two-sample) based on the distance between their empirical distribution functions. The null hypothesis is that the two distribution functions are drawn from the same distribution at a certain significance level (here, $\alpha = 0.05$). I use the two-sample KS test to compare different types of droughts and heatwaves (e.g., 3-day 85th percentile, 5-day 90th percentile) in 1990-2010 relative to 1960-1980. The test indicates whether the data from the two periods come from the same distribution at a 0.05 significance level.

The Mann-Kendall (MK) trend test [*M Kendall*, 1975] **Error! Reference source not found.** here assesses the presence of a statistically significant (0.05 significance level) trend in the time series of the fraction of CONUS in concurrent drought and heatwave. The MK test is a nonparametric approach based on the empirical ranks of time series widely used in hydrology and climatology.

We use a framework based on the Cramér–von Mises change point detection to evaluate temporal changes in the concurrent drought and heatwave events [*Holmes et al.*, 2013] [*Bücher and Kojadinovic*, 2016; *Cramér*, 1928; *Kojadinovic*, 2013; *Sharkey and Killick*, 2014]. This approach detects changes in the empirical cumulative distribution functions by comparing two subsamples ($\hat{F}_S(x)$ and $\hat{F}_T(x)$) of the original time series:

$$\hat{F}_S(x) = \frac{1}{\tau}\sum_{i=1}^{\tau} I(X_i \leq x)$$

$$\hat{F}_T(x) = \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} I(X_i \leq x)$$

where $\hat{F}_S(x)$ and $\hat{F}_T(x)$ are the empirical CDF of the two subsamples, *I* is the indicator function, *n* denotes sample size, and the terms $\frac{1}{\tau}$ and $\frac{1}{n-\tau}$ are adjustment factors for the length of each subsample. The test measures the divergence between the empirical distributions as:

$$W_{\tau,n} = \int_{-\infty}^{\infty} \left|\hat{F}_S - \hat{F}_T\right|^2 dF_t(x)$$

where $W_{\tau,n}$ can be computed as the square of the mean distance between the empirical distributions [*Bücher and Kojadinovic*, 2016; *Sharkey and Killick*, 2014]:

$$W_{\tau,n} = \sum_{i=1}^{n} \left|\hat{F}_S(X_i) - \hat{F}_T(X_i)\right|^2$$

Larger divergence values, *W*, indicate greater changes in the cumulative distributions. Here, the null hypothesis is that there is no change in the data over time and the two subsamples come from the same distribution. The null hypothesis is rejected if at an unspecified point $\tau$, $\hat{F}_S(x)$ and $\hat{F}_T(x)$ come from statistically different distributions. Since, I do not have any prior information on the position of $\tau$ in the time series, the test involves computing $W_{\tau,n}$ for all $1 < \tau < n$ [*Ross and Adams*, 2012] however, for different values of $\tau$, the variance of the two subsamples will be different. For this reason, the $W_{\tau,n}$ statistics are adjusted so that both periods exhibit equal mean and variance for all values of $\tau$ [*Maurer et al.*, 2002] (see Appendix 1 for more). The methods in this study should be applied to independent and identically distributed time series. Appendix 1 provide more information on the sampling approach and temporal autocorrelation of the data (see Figure S5 in Appendix 1 and the corresponding discussion).

## Appendix 1

### Data Sources and Processing

We used daily temperature and monthly precipitation data from a unique long-term and observation-based data set that provides forcings for land-surface models [*Livneh et al.*, 2013; *Maurer et al.*, 2002]. In this data set, daily precipitation is based on data from more than 10,000 precipitation gauges available from the National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer Program (COOP). Using the synergraphic mapping system (SYMAP) approach[*Shepard*, 1984; *Widmann and Bretherton*, 2000], the precipitation observations are gridded to 1/8$^{o}$ resolution. Monthly precipitation fields are daily precipitation accumulations. Maximum daily temperature data are also from the NOAA's COOP stations and regridded using the same approach.

Information about the NOAA COOP data can be found here:

http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/cooperative-observer-network-coop

The 1/8$^{o}$ gridded daily temperature data are available:

http://hydro.engr.scu.edu/files/gridded_obs/daily/ncfiles/

The 1/8$^{o}$ gridded monthly precipitation data are available:

http://hydro.engr.scu.edu/files/gridded_obs/monthly/ncfiles/

This study focuses on warm season (May – October) temperature extremes (i.e., 85$^{th}$, 90$^{th}$, and 95$^{th}$ percentiles of the maximum temperature) and the corresponding monthly precipitation. I derive the percent (%) of the contiguous United States (CONUS) in concurrent drought and

heatwave by summing the number of concurrent events in the warm season normalized by CONUS's area (sum of the pixels in concurrent drought and heatwave in each year's May-October, divided by the total number of pixels in CONUS). More than one heatwave can occur during a warm season and by using this approach, I can account for multiple concurrent drought and heatwave events in any given pixel.

**Results and Discussion**

Figure 1 in the main paper shows percent change in concurrent drought and heatwave events during 1990-2010 relative to 1960-1980. I present the results for 1-month SPI<-0.8 (approximately, $20^{th}$ percentile precipitation), different heatwave durations (3-, 5-, and 7-day) and extreme temperature thresholds ($85^{th}$, $90^{th}$, and $95^{th}$ percentiles). To show consistent patterns of percent change in concurrent drought and heatwave events, I show similar figures for SPI<-0.5 (Figure S1) and 3-month SPI<-0.8 (Figure S2). The general patterns remain similar to Figure 1 for different drought definitions. For better visualization, the percent change values are smoothed using a 3×3-pixel running mean.

*Figure S1: Percent (%) change in concurrent droughts (SPI<-0.5) and heatwaves during 1990-2010 relative to 1960-1980 for each gridbox. The rows change in heatwave severity (85th percentile, 90th percentile, and 95th percentile) while the columns change in heatwave duration (3-day, 5-day, and 7-day).*

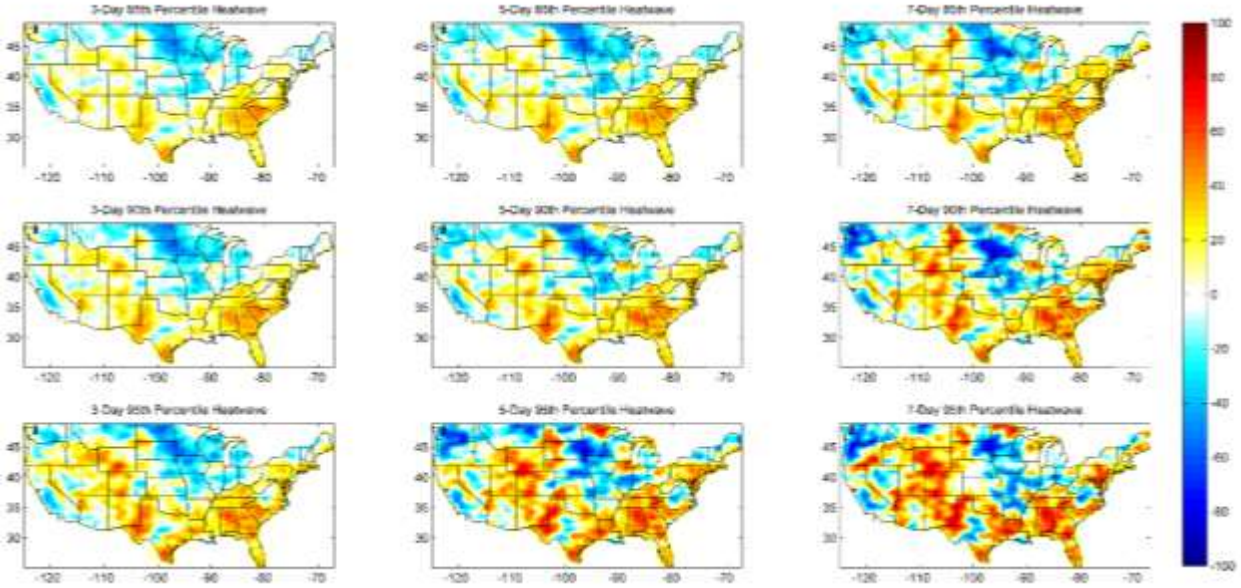


*Figure S2: Percent (%) change in concurrent droughts (3-month SPI<-0.8) and heatwaves during 1990-2010 relative to 1960-1980 for each gridbox. The rows change in heatwave severity (85th percentile, 90th percentile, and 95th percentile) while the columns change in heatwave duration (3-day, 5-day, and 7-day).*

Figure 3 in the main paper displays the percent of the Continental United States (CONUS) in concurrent drought and heatwave for different duration and intensities from 1960 – 2010. For a better visualization, Figure S3 shows the percent (%) of CONUS in concurrent drought and heatwave from 1960-1980 (bottom x-axes) relative to 1990-2010 (top x-axes) for three different heatwave durations and severity thresholds. The results show more concurrent drought and heatwaves during 1990-2010 (top x-axes), consistent with the findings of Figure 2. However, the Mann-Kendall Trend test does not indicate any statistically significant change in the concurrent drought and heatwave time series in Figure 3.

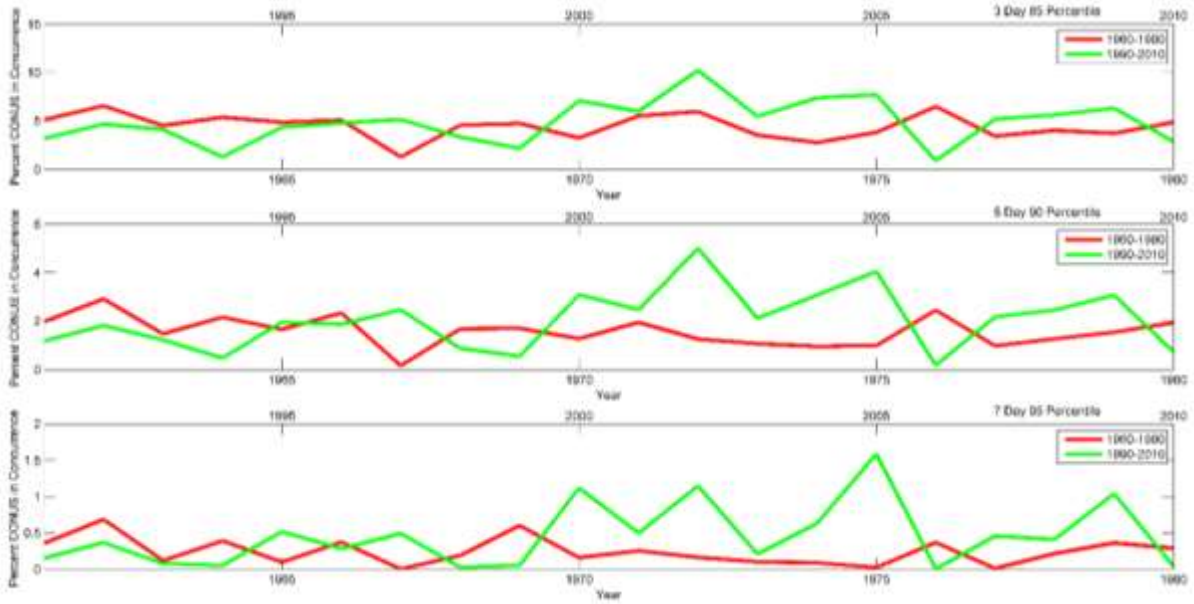



*Figure S3: Percent (%) of the contiguous United States (CONUS) in concurrent drought and heatwave from 1960-1980 (bottom x-axes) relative to 1990-2010 (top x-axes) for three different heatwave durations and severity thresholds.*

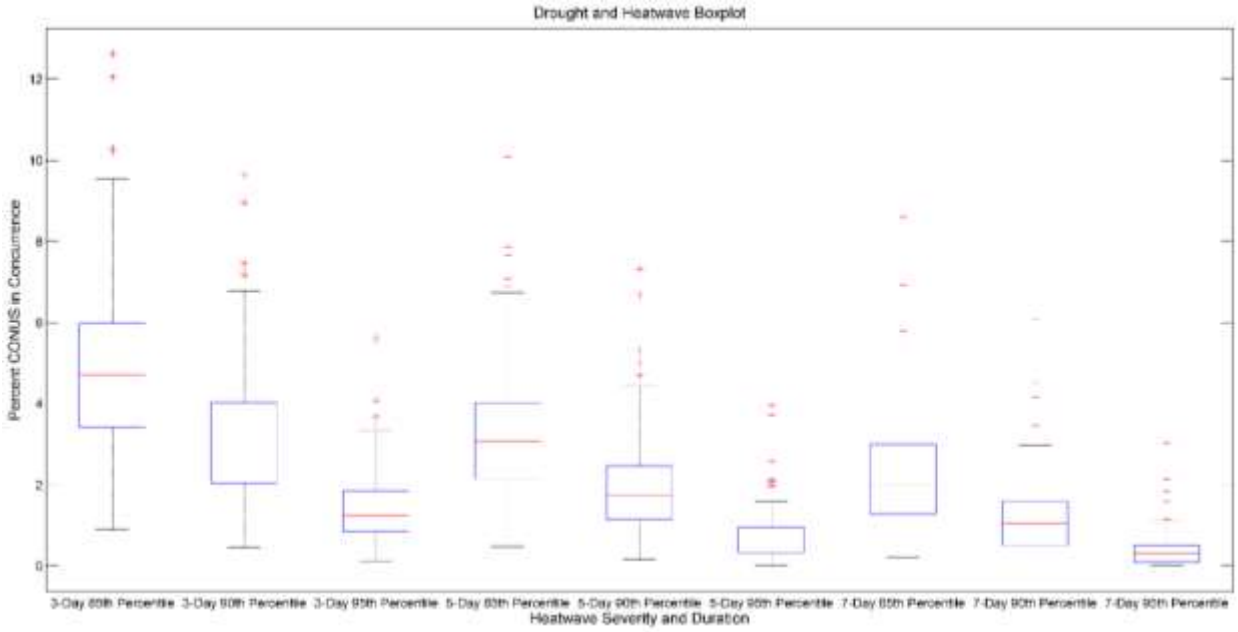



*Figure S4: Boxplots of the percent (%) of the contiguous United States (CONUS) in concurrent drought and heatwave from 1960 to 2010 (based on data presented in Figure 3) for three different heatwave durations (3-,5-,7-day) and severity thresholds (85$^{th}$, 90$^{th}$, 95$^{th}$ percentiles). The central marks (red lines) are the medians and the edges of the boxes are the 25th and 75th percentiles. The whiskers show the most extreme values not identified as outliers.*

Both droughts and heatwave events exhibit dependence across space (see Figure 1). However, the statistical tests in this paper require independent and identically distributed time series[*Clarke*, 2010]. Basically, multiple pixels that belong to one event should not be counted as multiple independent occurrences. To address this issue, at any given time step, I extracted one value: the fraction of CONUS in concurrent drought and heatwave. In other words, I identify all concurrent droughts and heatwaves. Then, instead of using the count of the events (or pixels), I evaluate the fraction of the country. As shown in Figure S5, the values extracted at

each time step, used in Figures 2-4, do not exhibit any statistically significant temporal correlation at 0.05 significance level.

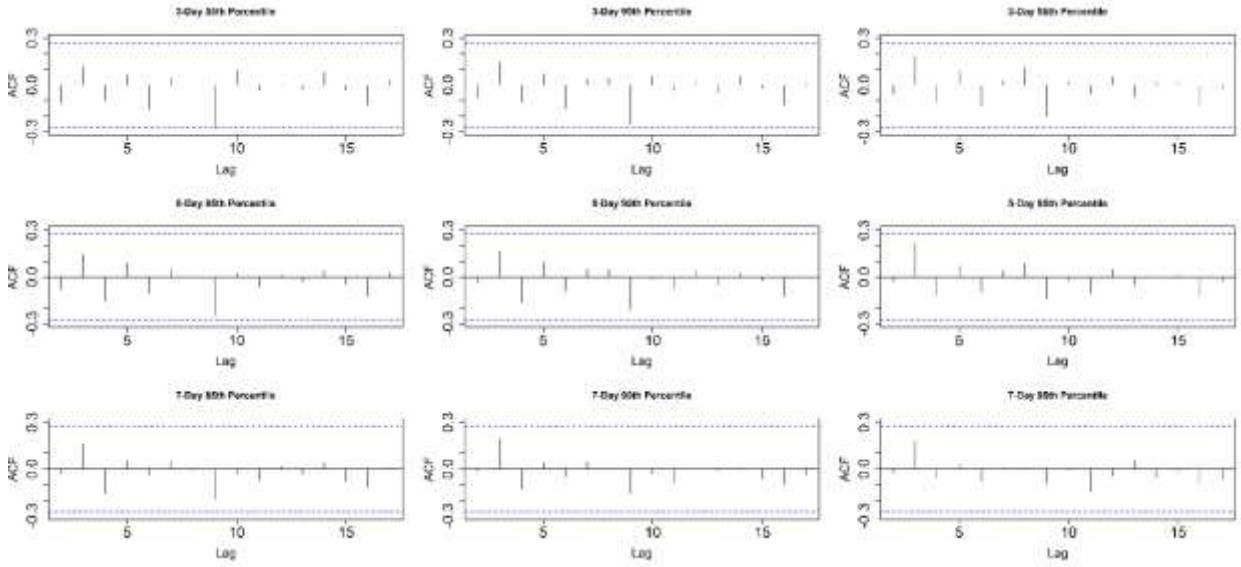



*Figure S5: Autocorrelation Function (ACF) of the time series used in Figure 3. The dashed lines represent the 95% confidence intervals (0.05 significance level).*

| Drought and Heatwave | H | p-value |
|---|---|---|
| 3-Day, 85 Percentile | 0 | 0.530908 |
| 5-day, 85 Percentile | 1 | 0.00108 |
| 7-day, 85 Percentile | 1 | 1.57E-05 |
| 3-day, 90 Percentile | 1 | 3.10E-06 |
| 5-day, 90 Percentile | 1 | 3.10E-06 |
| 7-day, 90 Percentile | 1 | 3.10E-06 |
| 3-day, 95 Percentile | 1 | 1.68E-09 |
| 5-day, 95 Percentile | 1 | 1.68E-09 |
| 7-day, 95 Percentile | 1 | 1.68E-09 |

*Table S1: Trends in drought and heatwave concurrence from 1960-2010 based on the Mann-Kendall trend test. A p-value > 0.05 indicates that the null-hypothesis of no trend cannot be rejected at a significance level of 0.05 (95% confidence). Column 2 shows the corresponding p-values, and column 3 shows the trend line slope for the best linear fit.*

**Method**

In the Cramér–von Mises change point detection method, the $W_{\tau,n}$ statistics are adjusted to ensure the two subsamples ($\hat{F}_S(x)$ and $\hat{F}_T(x)$) have equal mean and variance for all values of $\tau$ [*Clarke*, 2010]. This can be achieved using a standardization approach as [*Anderson*, 1962]:

$$\sigma_{\mathrm{W}_{\tau,\mathrm{n}}} = \sqrt{\frac{(n+1)\left(\left(1-\frac{3}{4}\,\tau\right)n^2 + (1-\tau)n - \tau\right)}{45n^2(n-\tau)}}$$

$$\mu_{\mathrm{W}_{\tau,\mathrm{n}}} = \frac{(n+1)}{6n}$$

where $\mu_{\mathrm{W}_{\tau,\mathrm{n}}}$ and $\sigma_{\mathrm{W}_{\tau,\mathrm{n}}}$ are the mean and standard deviation of $\mathrm{W}_{\tau,\mathrm{n}}$ . The so-called maximized test statistic can then be defined as [45]:

$$\mathrm{W}_{\mathrm{n}} = \max_{\tau} \frac{\mathrm{W}_{\tau,\mathrm{n}} - \mu_{\mathrm{W}_{\tau,\mathrm{n}}}}{\sigma_{\mathrm{W}_{\tau,\mathrm{n}}}}$$

This method of standardizing $\mathrm{W}_{\tau,\mathrm{n}}$ and maximizing $\mathrm{W}_{\mathrm{n}}$ for all values of $\tau$ gives a single test statistic for significance analysis and hypothesis testing. A bootstrap sampling approach outlined inreferences [*Anderson*, 1962; *Bücher and Kojadinovic*, 2016; *Bücher et al.*, 2014; *Gombay and Horváth*, 1999; *Hawkins and Deng*, 2010]can determine the p-value and hypothesis testing. The approximate p-value is computed using a multiplier approach as:

$$\frac{(0.5 + \sum_{i=1}^{N} 1_{\{W_i \geq W\}})}{N+1}$$

where $W$ and $W_i$ are the test statistic and a multiplier replication, respectively. In a bootstrap process with N=1000 replications, if the result falls in the $\alpha$-tail ($\alpha$=0.05) of the N simulations, the test is statistically significant at a 0.05 significance level. The bootstrapping approach is known as the sequential block bootstrapping procedure, and can be used for both independent and serially dependent variables [*Bücher and Kojadinovic*, 2016; *Politis and White*, 2004]. Given

that my data is independent and identically distributed (Figure S5), I use a unit block length, which leads to independent and identically distributed multiplier sequences. In addition to the above approximate p-value, the so-called two-sample Cramer-von Mises test (CMTest2) can compare two different periods (data records). This test evaluates whether independent samples from two periods (here, 1960-1980 and 1990-2010) are drawn from the same population[*Anderson*, 1962; *Anderson and Darling*, 1952]. In this test, the p-value can be derived similar to the two-sample Kolmogorov Smirnov test. Table S2 summarizes the p-values of CMTest2 for drought and heatwaves during 1990-2010 relative to 1960-1980. The table confirms a statistically significant change in the distribution of concurrent drought and heatwaves at 0.05 significance level (95% confidence) for all heatwave durations and intensities except for 3-day, 85$^{th}$ percentile heatwaves. For more information on the procedure and statistical significance analysis and bootstrapping approach, refer to the "cpm" (http://cran.r-project.org/src/contrib/Archive/cpm/) and "npcp" (http://cran.r-project.org/web/packages/npcp/index.html) R packages.

| Drought and Heatwave | H | p-value |
|---|---|---|
| 3-Day, 85 Percentile | 0 | 0.9095 |
| 5-day, 85 Percentile | 0 | 0.871 |
| 7-day, 85 Percentile | 0 | 0.5919 |
| 3-day, 90 Percentile | 0 | 0.7576 |
| 5-day, 90 Percentile | 0 | 0.6492 |
| 7-day, 90 Percentile | 0 | 0.4355 |
| 3-day, 95 Percentile | 0 | 0.6728 |
| 5-day, 95 Percentile | 0 | 0.4452 |
| 7-day, 95 Percentile | 0 | 0.4452 |

*Table S2: p-values of the Cramer von-Mises test for concurrent drought and heatwaves in 1990-2010 relative to 1960-1980. A p-value < 0.05 indicates that the null-hypothesis of no change can be rejected at a 0.05 significance level (95% confidence).*

# Chapter 2a

## Generalized Conditional Probability Approach for Studying Inter-related Extremes: Heat wave impact on mortality

The contents of Chapter 2a are published in Science Advances.

Citation: Mazdiyasni, Omid, et al. "Increasing probability of mortality during Indian heat waves." *Science advances* 3.6 (2017): e1700066.

**Rising global temperatures are causing increases in the frequency and severity of extreme climatic events such as floods, droughts, and heatwaves. Here, I analyze changes in summer temperatures, the frequency, severity and duration of heatwaves, and heat-related mortality in India between 1960 and 2009, using data from the India Meteorological Department. Mean temperatures across India have risen by more than 0.5 °C over this period, with statistically significant increases in heatwaves. Using a novel probabilistic model, I further show that the increase in summer mean temperatures in India over this period corresponds to a 146% increase in the probability of heat-related mortality events of more than 100 people. In turn, my results suggest that future climate warming will lead to substantial increases in heat-related mortality, particularly in developing, low-latitude countries such as India where heatwaves will become more frequent and populations are especially vulnerable to these extreme temperatures. My findings indicate that even moderate increases in mean temperatures may cause great increases in heat-related mortality, and support efforts of governments and international organizations to build-up the resilience of these vulnerable regions to more and more severe heatwaves.**

Global mean temperatures are expected to increase by as much as 5.5 °C by the end of this century[*Stocker et al.*, 2013], which is in turn expected to increase the intensity of heatwaves around the world[*Hansen et al.*, 2012; *G A Meehl and Tebaldi*, 2004; *Shi et al.*, 2015], with the largest relative effects on summer temperatures in developing regions such as Africa, South America, the Middle East and south Asia[*Diffenbaugh and Scherer*, 2011]. The impacts of such heatwaves on human and natural systems include decreased air quality, diminished crop yields,

increased energy consumption, increased evapotranspiration, intensification of droughts, and—perhaps most concerning of all—direct effects on human health[*Basu*, 2009; *Mazdiyasni and AghaKouchak*, 2015]. Heat stress during periods of high temperatures may also exacerbate health problems such as cardiovascular and respiratory disease and cause life-threatening crises[*Hajat et al.*, 2007; *Wilker et al.*, 2012]. Certain segments of the population such as the young, elderly, and poor may therefore be especially susceptible to these health impacts due to existing health conditions as well as lack of basic resources such as clean drinking water, shelter, access to air conditioning and healthcare[*Bouchama and Knochel*, 2002]. Populations without central air conditioning tend to have higher heat-related mortality rates[*O'Neill et al.*, 2005].

In light of geographical patterns of warming and vulnerable populations, here I present an analysis of a half-century (1960-2009) of temperature, heatwaves, and related mortality in India. Previous studies report that between 1971 and 2007, there was an increase of over 0.5 °C in mean temperatures across India[*Kothawale et al.*, 2010], and the projected annual spatial warming in India will be between 2.2 – 5.5 °C by the end of the 21$^{st}$ Century, with higher projections over north, central, and west India[*Dholakia et al.*, 2015; *Kumar et al.*, 2013]. Based on data from the World Bank, of the 1.24 billion people living in India in 2011 (18% of global population), an estimated 23.6% earned <$1.25 per day and ~25% did not have any access to electricity, making them especially vulnerable to the impacts of heatwaves[*Murari et al.*, 2015]. This vulnerability has been made clear by events in recent years: heatwaves in 2010 killed over 1300 people in the city of Ahmedabad alone, prompting the start of efforts to develop coordinated Heat Action Plans[*Knowlton et al.*, 2014]. However, these efforts remain limited

and localized, and in 2013 and 2015 the country experienced another bout of intense heatwaves that killed over 1500 and 2500 people across the country, respectively. Since then, there have been several more deadly heatwaves, including the most intense Indian heatwave in recorded history in May of 2016 when maximum temperatures in Jaisalmer reached 52.4 °C.

Heatwaves are usually described as successive hot days[*G A Meehl and Tebaldi*, 2004; *Perkins and Alexander*, 2013], and are often defined on a percentile basis[*Mazdiyasni and AghaKouchak*, 2015]. Here, I consider heatwaves to be three or more consecutive days of temperatures above the 85$^{th}$ percentile of the hottest month for each specific location. Figure S1 shows the heatwave threshold values across India (85$^{th}$ percentile of hottest month's mean temperature °C). For each warm season in India (April-September) from 1960 to 2009, I assess four different heatwave properties: (1) accumulated heatwave intensity, (2) annual heatwave count, (3) mean heatwave duration, and (4) heatwave days. The annual heatwave count and mean duration are simply the number of heatwave events that occur each year and their average duration in days, respectively. Heatwave days is the product of heatwave count and heatwave duration, and represents the number of days under heatwave condition. I evaluate accumulated heatwave intensity as the cumulative cooling degree-day (CDD), or the sum of the daily mean temperature during a heatwave subtracted by 22 °C, over the entire heatwave event [i.e. (daily temperature °C – 22 °C) x duration in days]. I perform my heatwave analyses based on summer mean temperatures (i.e. the average of mean daily temperatures during the summer), because I believe this to be a better indicator of accumulated heat stress. However, I also provide analyses based on summer maximum temperatures in the Appendix 2 section. These analyses are performed using 1° x 1° daily temperature records from the India

Meteorological Department. Finally, I use the results from these retrospective analyses to develop a conditional probabilistic model of the relationships among summer mean temperatures, heatwave days, and heat-related mortality that I apply to estimate the probability distribution of heat-related mortality related to mean climate warming in the future. Further details of my analytic approach are provided in the Methods section.

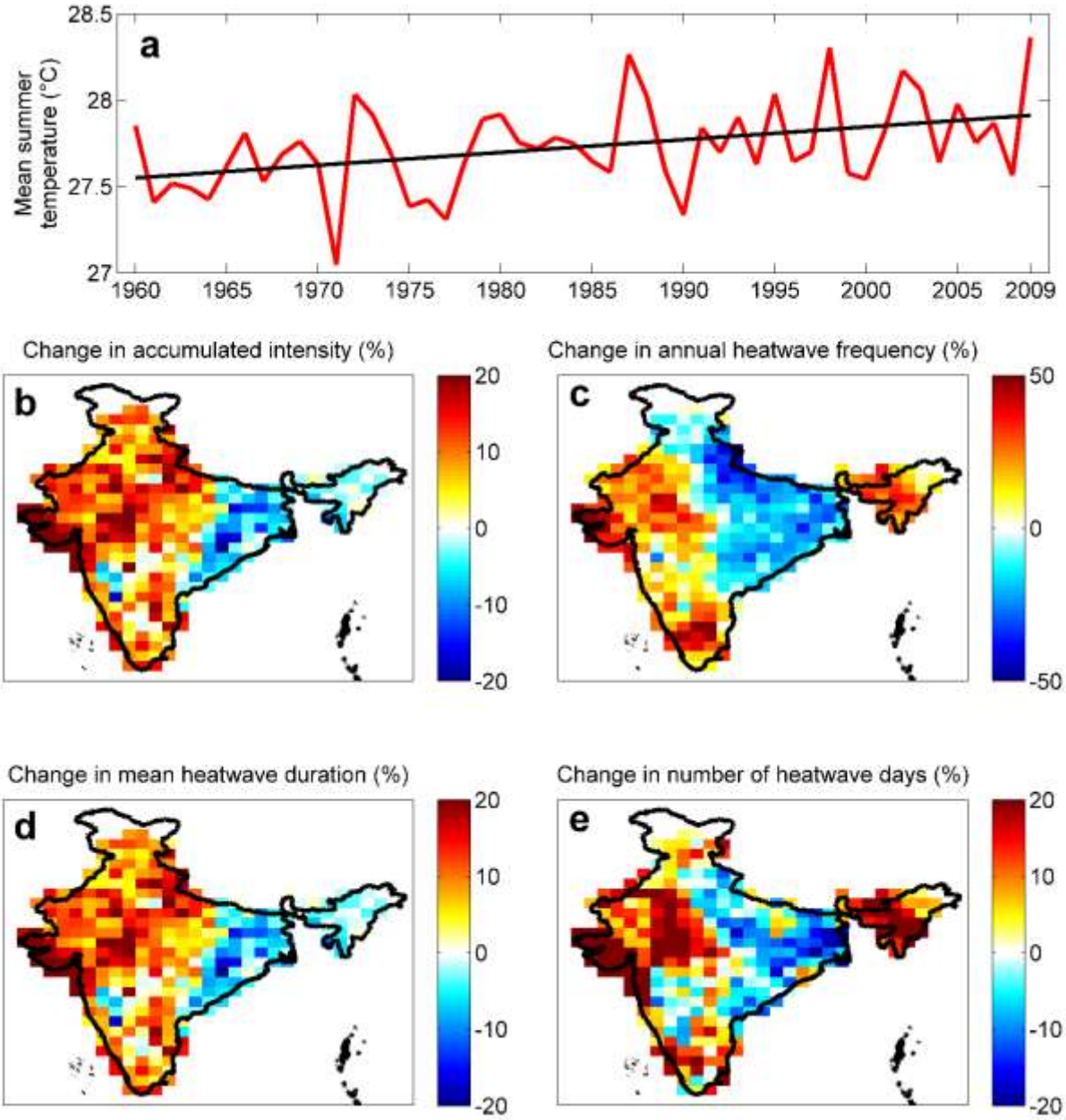


**Figure 1**| **Temperature and heatwave increases in India 1960-2009**. Summer mean temperatures in India have increased 1960-2009 as indicated by the Mann-Kendall Trend Test (**a**). The (**b**) accumulated heatwave intensity, (**c**) number of heatwave events, (**d**) heatwave duration, and (**e**) heatwave days during the latter period (1985-2009) has also increased over most areas of India relative to the preceding period of 1960-1984.

Figure 1a shows that summer mean temperatures have increased substantially from 1960 to 2009. The time series exhibits a statistically significant (95% confidence interval) upward trend confirmed using the Mann-Kendall trend test. The accumulated intensity, count, duration, and heatwave days of Indian heatwaves have also increased over the analyzed time period over most of the country, and especially in northern, southern, and western parts of India (Figure 1b-e).

The red shading that dominates most of the maps in Figure 1 indicates that the observed increases are widespread and strong: southern and western India experienced 50% more heatwave events during the period 1985-2009 than during the previous 25-year period (calculated by dividing the difference in the number of events from 1985-2009 relative to 1960-1984 by the total number of events). Similarly, heatwave days and the mean duration of heatwaves have increased by approximately 25% in the majority of India. Appendix 2 Figure S2 shows the same analysis for heatwaves calculated by using summer maximum temperatures. Appendix 2 Figure S3 shows mean values for each heatwave characteristic from 1960 – 1984 and 1985 – 2009, separately. Appendix 2 Figure S4 shows the areas where there was a statistically significant trend confirmed by the Mann-Kendall Trend Test.

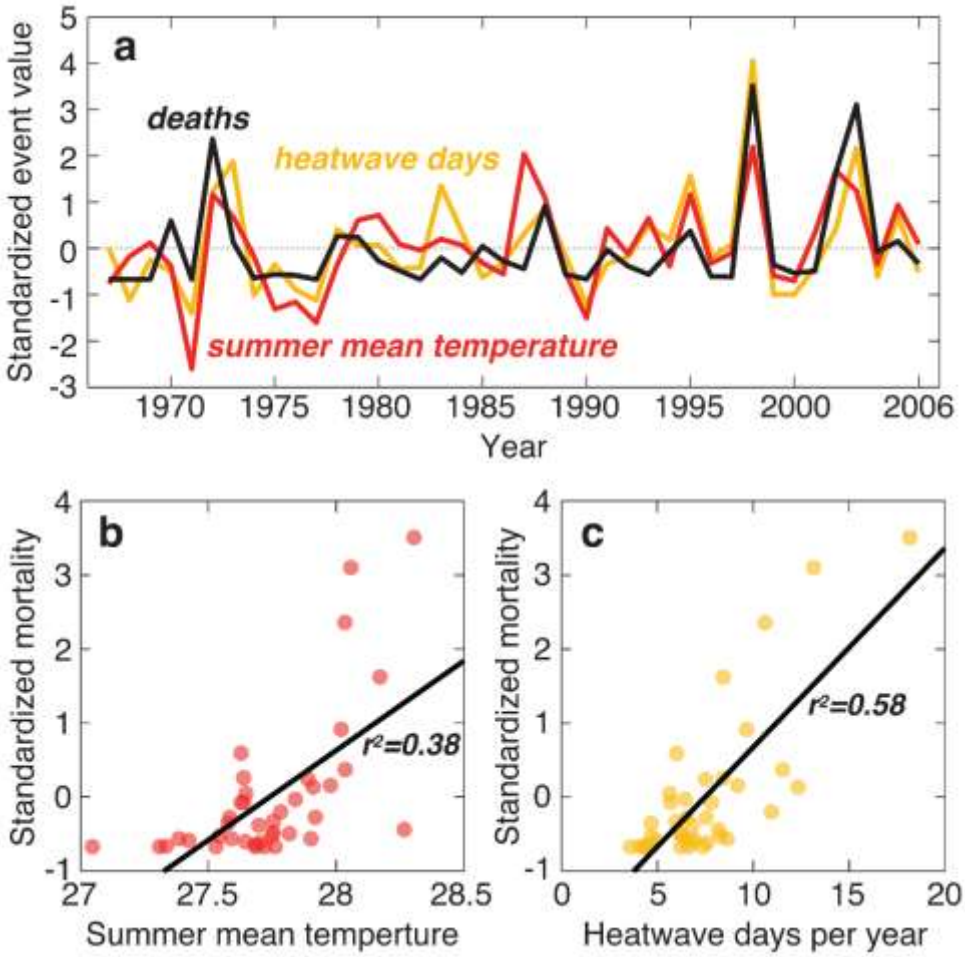



**Figure 2 | Standardized number of heatwave days, summer mean temperatures, and heat-related mortality**. Standardized trends show the correspondence among the three variables. In years where heatwaves days (yellow) and summer mean temperature (red) are above-average, heat-related deaths also spike upwards.

Figure 2a shows the relationships among standardized values of summer mean temperatures, heatwave days, and annual heat-related mortality occurring over the period 1967-2007 (the period for which reliable mortality data were available; see Appendix 2 for details). Although high summer mean temperatures often correspond to spikes in deaths, the correlation of temperatures to deaths is weaker (Pearson's linear correlation=63%, $r^2$=0.38; Fig. 2b) than the correlation with the number of heatwave days each year (Pearson's linear correlation=77%, $r^2$=0.58; Fig. 2c), especially in the years when there were high mortality rates.

In an effort to understand the underlying mechanisms of heatwave mortality, I further explored its relationship with population and income-levels in India. Figure 3 shows that the relationship between population-weighted heatwave days and mortality rates is only slightly better than that between mortality and summer mean temperatures (Pearson's linear correlation 67%, $r^2$=0.44; Fig. 3b). However, the correlation between income-weighted heatwave days and mortality rates is better (Pearson's linear correlation 77%, $r^2$=0.58; Fig. 3c).

Based on these correlations, I infer that the relationship between income and human health is stronger than that of physical conditions and health, perhaps as the result of access to air conditioning or medical care. It is known that some highly populated regions have low income per capita (e.g., northern India) and also many rural, low populated regions also have low income per capita (i.e. central and eastern India), which I show in Figure S5.

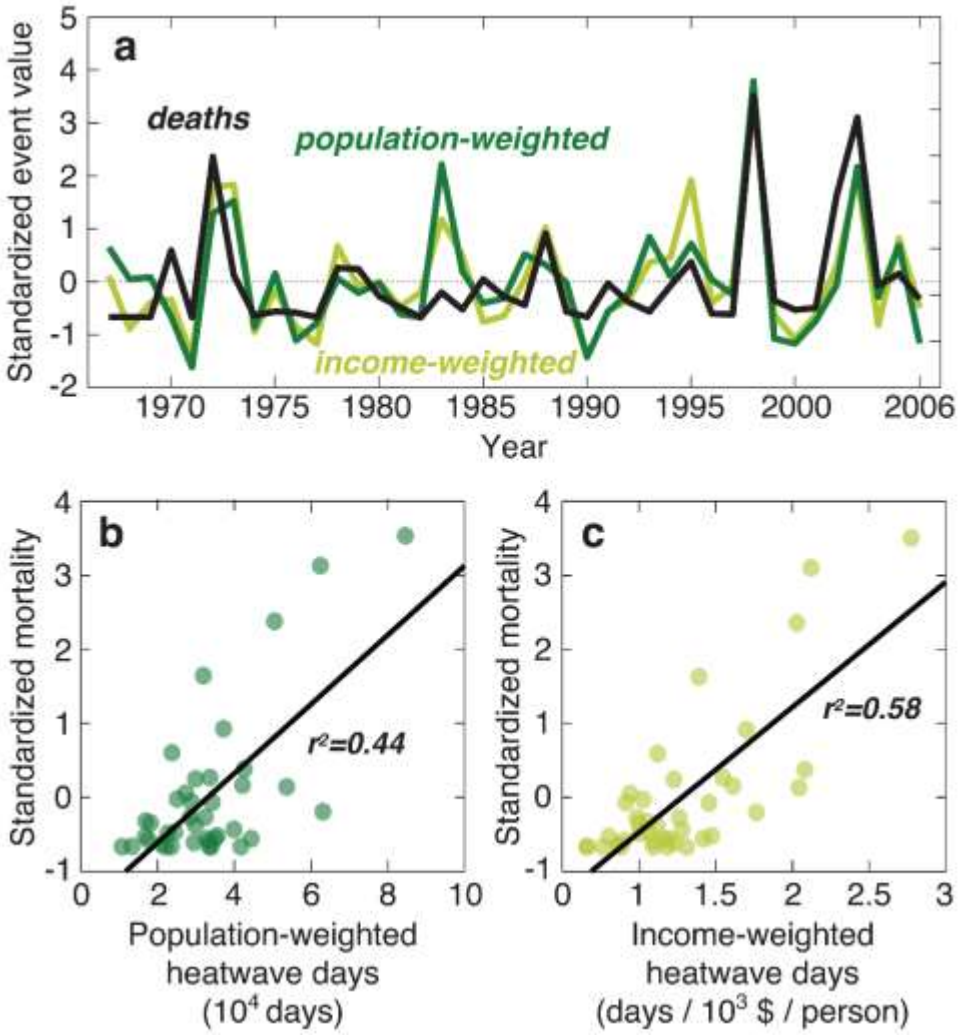



**Figure 3 | Standardized population-weighted heatwave days, income-weighted heatwave days, and heat-related mortality**. Standardized trends show the correspondence among the three variables. In most years where income-weighted heatwaves (red) and population-weighted heatwaves (yellow) are above-average, heat-related deaths also increase dramatically.

Figure 2a highlights several years—1972, 1988, 1998, and 2003—in which there were more than ten heatwave days on average across India, with corresponding spikes in heat-related mass-mortality of between 650 and 1500 people. However, there are a few years, such as 1973, 1983, 1984, and 1995, in which there were an above-average number of income-weighted heatwave days, but a low number of deaths. A possible explanation is that the areas where these latter heatwaves occurred tended to be less populous and/or wealthier regions (see

Appendix 2 Figure S5). Such observations reinforce previous work that has highlighted poverty as a significant factor in climate-induced mortality, such as heatwave deaths[*Hamoudi and Sachs*, 1999].

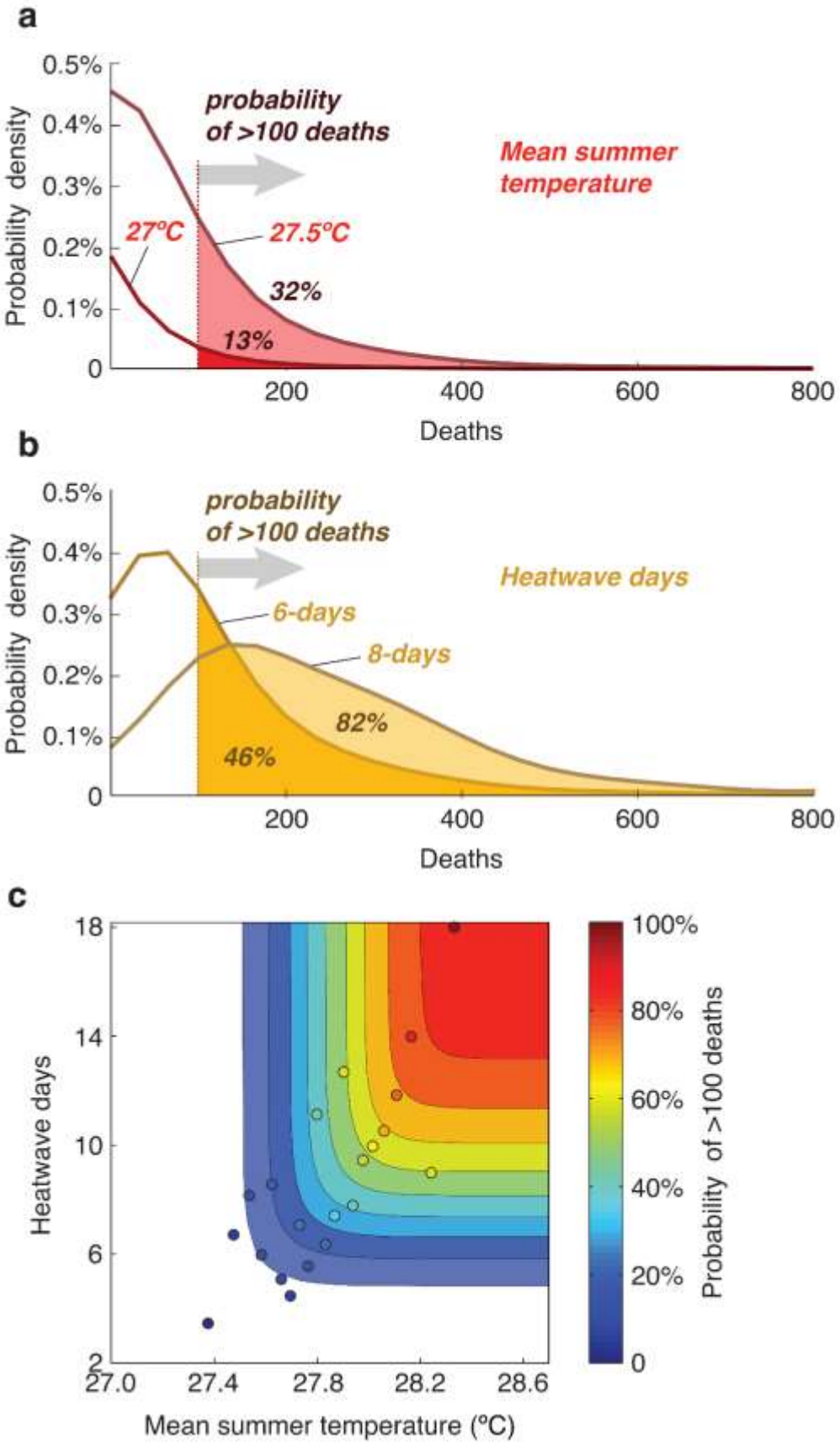

a
0.5%
0.4%
0.3%
0.2%
0.1%
0
Probability density
probability
of >100 deaths
Mean summer
temperature
27°C
27.5°C
32%
13%
200
400
600
800
Deaths
b
0.5%
0.4%
0.3%
0.2%
0.1%
0
Probability density
probability
of >100 deaths
Heatwave days
6-days
8-days
82%
46%
200
400
600
800
Deaths
c
18
14
10
6
2
Heatwave days
27.0
27.4
27.8
28.2
28.6
Mean summer temperature (°C)
100%
80%
60%
40%
20%
0
Probability of >100 deaths

**Figure 4 | Probabilities of heatwave-caused mass-mortality events**. Parametric conditional probability density functions for yearly mortality given certain thresholds for summer mean temperatures (**a**) and heatwave days (**b**). With 0.5 °C warmer mean temperatures or 2 more heatwave days per year, the probability of >100 heat-related deaths increase dramatically. The relationship between the two variables and probability of mass mortality events is shown in panel **c**.

Figure 4 presents the results of a conditional probability density analysis (see Methods) of annual mortality given certain thresholds for summer mean temperatures and heatwave days. The shaded region represents the probability of mass heat-related mortality (i.e. heat-related deaths of more than 100 people), given different summer temperature values. For example, the Figure 4a shows that there is 13% probability that years with summer mean temperatures equal to 27 °C will have mass heat-related mortality. However, with an increase in summer mean temperatures of just 0.5 °C (to 27.5 °C), the probability of such levels of heat-related deaths jumps by a factor of 2.5 to 32%. Figure S6 shows a similar relationship with summer maximum temperatures. Similarly, Figures 3b shows that the probability of heat-related mass mortality events increases from 46% to 82% (78% increase) where the average number of heatwave days across India shift from 6 days to 8 days, respectively. The substantial increase in mortality rates due to either a 0.5 °C increase in summer mean temperature or 2 more heatwave days suggests that future climate warming could have a relatively drastic human toll in India and similarly in developing tropical and subtropical countries. Meanwhile, some experts expect India's temperature to rise 2.2 – 5.5 degrees Celsius[*Dholakia et al.*, 2015; *Kumar et al.*, 2013] .

By almost all measures, heatwaves have increased dramatically across India over the past half-century and with them the incidence of heat-related mortality. Projected increases in global mean temperatures under a range of climate change scenarios can be expected to extend these

trends. Although India is particularly susceptible to heatwaves given its geography and current state of human development, there are many countries that are similarly vulnerable to the extreme heat events in the ever-warming world. My results suggest that even moderate and practically unavoidable increases in mean temperatures, such as 0.5 °C, may lead to large increases in heat-related mortality unless measures are taken to substantially improve the resilience of vulnerable populations.

## Methods

#### Temperature and Mortality Data

Daily temperature data based on 395 weather stations and interpolated at 1° x 1°spatial resolution was obtained from the India Meteorological Department[*Srivastava et al.*, 2009]. Mortality data was also obtained from the India Meteorological Department and based on mortality data from annual reports which compiled information from newspaper and other sources about mortality during specific extreme heat events[*IMD*, 2009] .

#### Statistical Methods

This paper uses the Kolmogorov-Smirnov test to test changes in distribution functions of heatwaves in different periods. I use the two-sample Kolmogorov-Smirnov (KS) test to analyze the differences between the cumulative distribution functions (CDFs) for the number of heatwaves. This study compares the different types of heatwaves that occurred in 1985 – 2009 relative to those in 1960 – 1984. The KS test is a nonparametric test that evaluates whether there is a statistically significant change between two distributions by calculating the largest distance between their empirical distributions. The null hypothesis is that the data sets come

from the same distribution at a certain confidence interval (95% confidence interval in this study).

The KS test determines changes in empirical distribution functions by comparing pre-change and post-change samples, defined as:

$\hat{F}_S(x) = \frac{1}{\tau}\sum_{i=1}^{\tau} I(X_i \leq x)$;

$\hat{F}_T(x) = \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} I(X_i \leq x)$.

where $\hat{F}_S(x)$ and $\hat{F}_T(x)$ are the empirical CDF of the two subsamples, *I* is the indicator function, *n* denotes sample size, and the terms $\frac{1}{\tau}$ and $\frac{1}{n-\tau}$ are adjustment factors for the length of each subsample. The KS test statistic is the maximum difference between two empirical distributions:

$D_{\tau,n} = \sup_x |\hat{F}_S(x) - \hat{F}_T(x)|$,

Larger divergence values ($D_{\tau,n}$ and $W_{\tau,n}$) represent greater changes in the cumulative distributions[*Ross and Adams*, 2012; *Sharkey and Killick*, 2014].

The Mann-Kendall (MK) trend test analyzes whether there is a statistically significant trend (95% confidence interval) in the number of heatwaves per year time series. The MK test is a nonparametric test that uses the empirical ranks of time series, and is widely used in hydrology and climatology[*M G Kendall*, 1948].

In this paper, I use R-squared measure to determine how close the mortality and heatwave characteristics' data are to the respective fitted regression lines. The R-squared statistic measures the proportion of variance in the dependent variable that is predictable from the independent variable. Lower R-squared values depict the dependent variable (mortality) cannot

be predicted from the independent variable (heatwave days), and higher values portray the dependent variable can be predicted with little to no error from the independent variable[*Wright*, 1921].

**Calculation of Conditional Probabilities**

To derive the conditional probabilities presented in Figure 4, this paper utilizes the multivariate copula functions[*Gräler et al.*, 2013; *Grimaldi et al.*, 2016; *Salvadori et al.*, 2011; *Salvadori et al.*, 2013] to find the joint probability distribution of mortality and summer mean temperatures across India. I fit the Frank copula and t-copula families to the summer mean temperatures and mortality, and heatwave days and mortality data, respectively, because they have the highest statistically significant (95% confidence interval) maximum likelihood out of all the copula families. Maximum Likelihood values and p-values for five major copula families with respect to summer mean temperatures and heatwave days relative to mortality are in the Appendix 2 Table S1. A copula function is defined as the multivariate distribution function[*Nelsen*, 1999; *Salvadori and De Michele*, 2007]:

$$\mathrm{F}_{\mathrm{X}_1 \ldots \mathrm{X}_n}(\mathrm{x}_1, \ldots, \mathrm{x}_i, \ldots, \mathrm{x}_n) = \mathrm{C}\left[\mathrm{F}_{\mathrm{X}_1}(\mathrm{x}_1), \ldots, \mathrm{F}_{\mathrm{X}_i}(\mathrm{x}_i), \ldots, \mathrm{F}_{\mathrm{X}_n}(\mathrm{x}_n)\right] = \mathrm{C}(\mathrm{U}_1, \ldots, \mathrm{U}_i, \ldots, \mathrm{U}_n) \quad (1)$$

where C is the cumulative distribution function (CDF) of the copula and $\mathrm{F}_{\mathrm{X}_i}(\mathrm{x}_i)$ is the non-exceedance probability of $\mathrm{x}_i$ (marginal distribution). Here I use the bivariate form to estimate the joint probability distribution of mortality rates and summer mean temperatures as well as mortality rates and heatwave days in India:

$$\mathrm{F}_{\mathrm{XY}}(\mathrm{x}, \mathrm{y}) = \mathrm{C}[\mathrm{F}_{\mathrm{X}}(\mathrm{x}), \mathrm{F}_{\mathrm{Y}}(\mathrm{y})] \quad (2)$$

To determine the conditional probabilities of mortality rates exceeding a threshold (Y > y) at different summer mean temperatures (X = $x_1$, $x_2$,….), i.e. $\mathrm{F_{Y|X}(Y > y \mid X)}$, I develop the conditional probability density function[*Madadgar and Moradkhani*, 2013]:

$$\mathrm{f_{Y|X}(y \mid x) = c[F_X(x), F_Y(y)] \,.\, f_Y(y)} \quad (3)$$

where c is the probability density function (PDF) and $\mathrm{f_Y(y)}$ is the mortality marginal distribution. Once I choose a certain summer mean temperature conditional PDF from Eq. 3, the probability of the mortality rates (Y) exceeding a particular threshold (y) is given by the area under the curve: $f_{Y|X}(y \mid x)$. This allows calculating conditional PDF $\mathrm{f_{Y|X}(y \mid x)}$ for different values of x (e.g., summer mean temperatures=27 °C or heatwave days=6 in Figure 4).

**Appendix 2a**

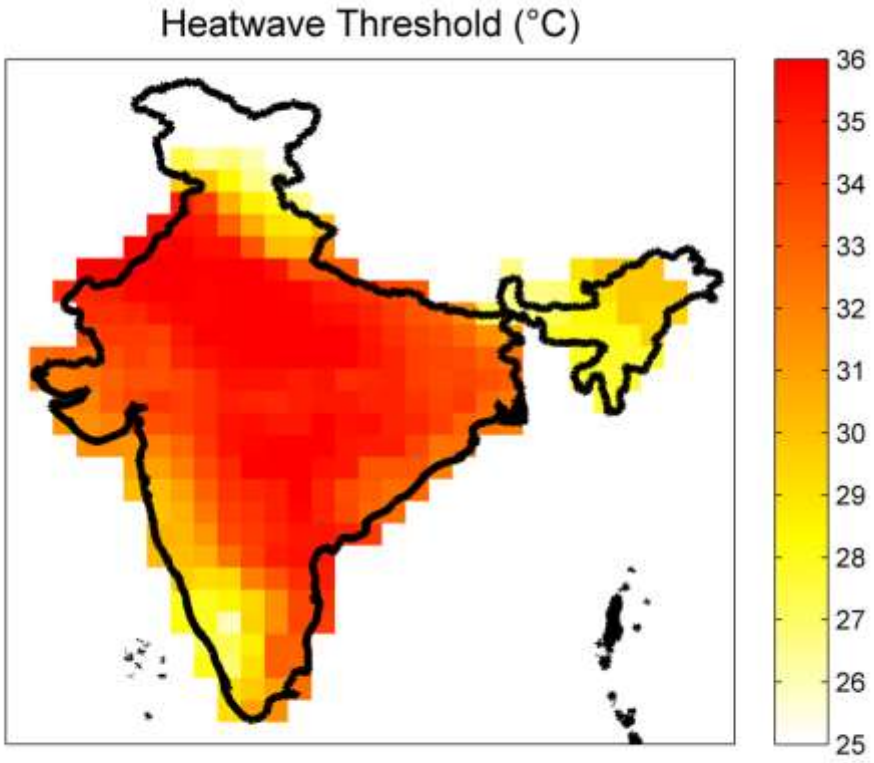



Figure S1 shows the heatwave thresholds across India. I calculate these thresholds by obtaining the 85$^{th}$ percentile of the daily mean temperatures of the hottest month in each grixbox. I show that the threshold of heatwaves are extremely high (>36 °C) for the majority of India.

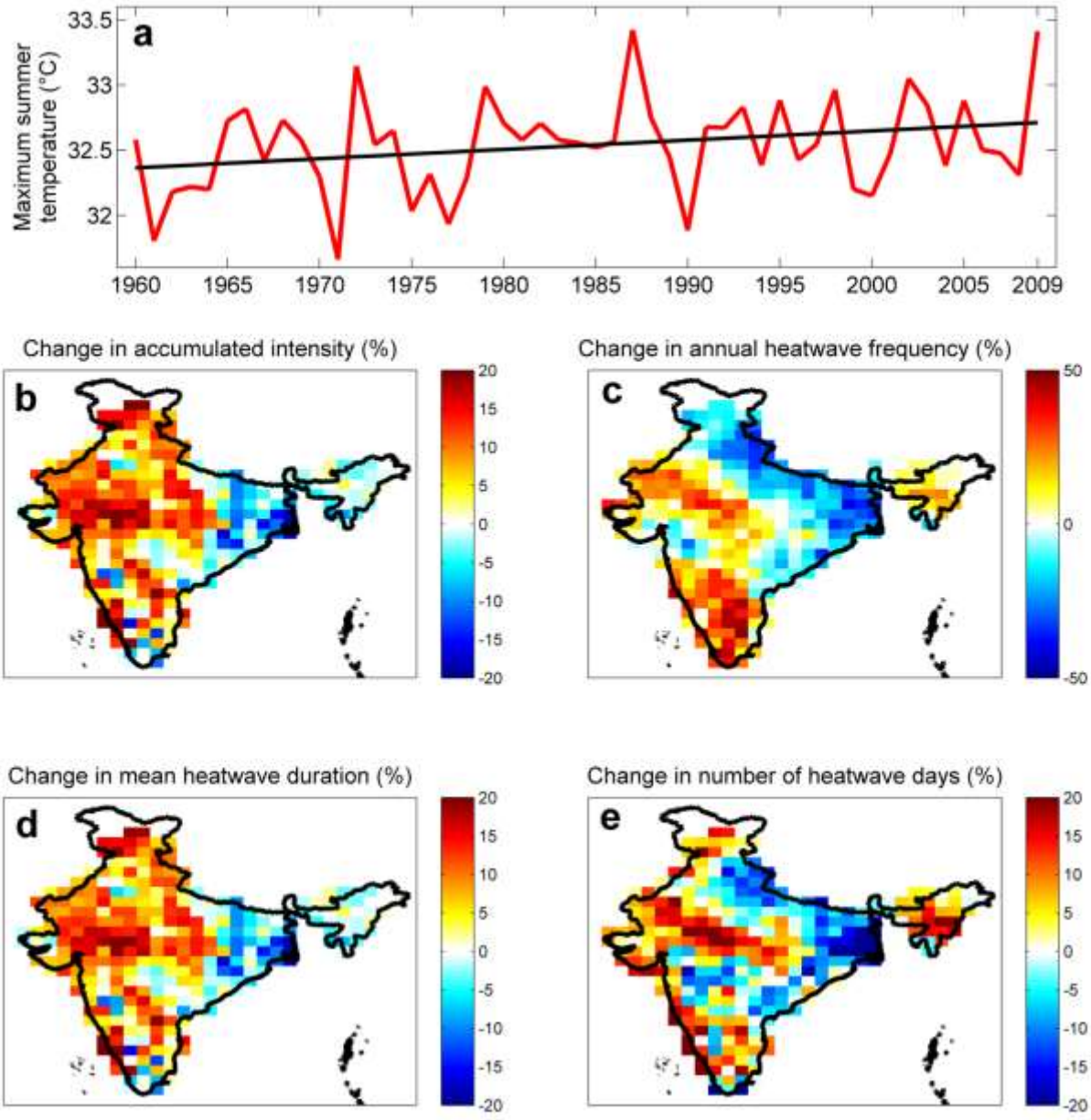



Figure S2a shows that summer mean temperatures have increased substantially from 1960 –to 2009. The time series exhibits a statistically significant (p=0.05) upward trend confirmed using the Mann-Kendall trend test.

In turn, the accumulated intensity, count, duration, and heatwave days of Indian heatwaves have also increased over the analyzed time period over most of the country, and especially in northern, southern, and western parts of India (Figure S2b).

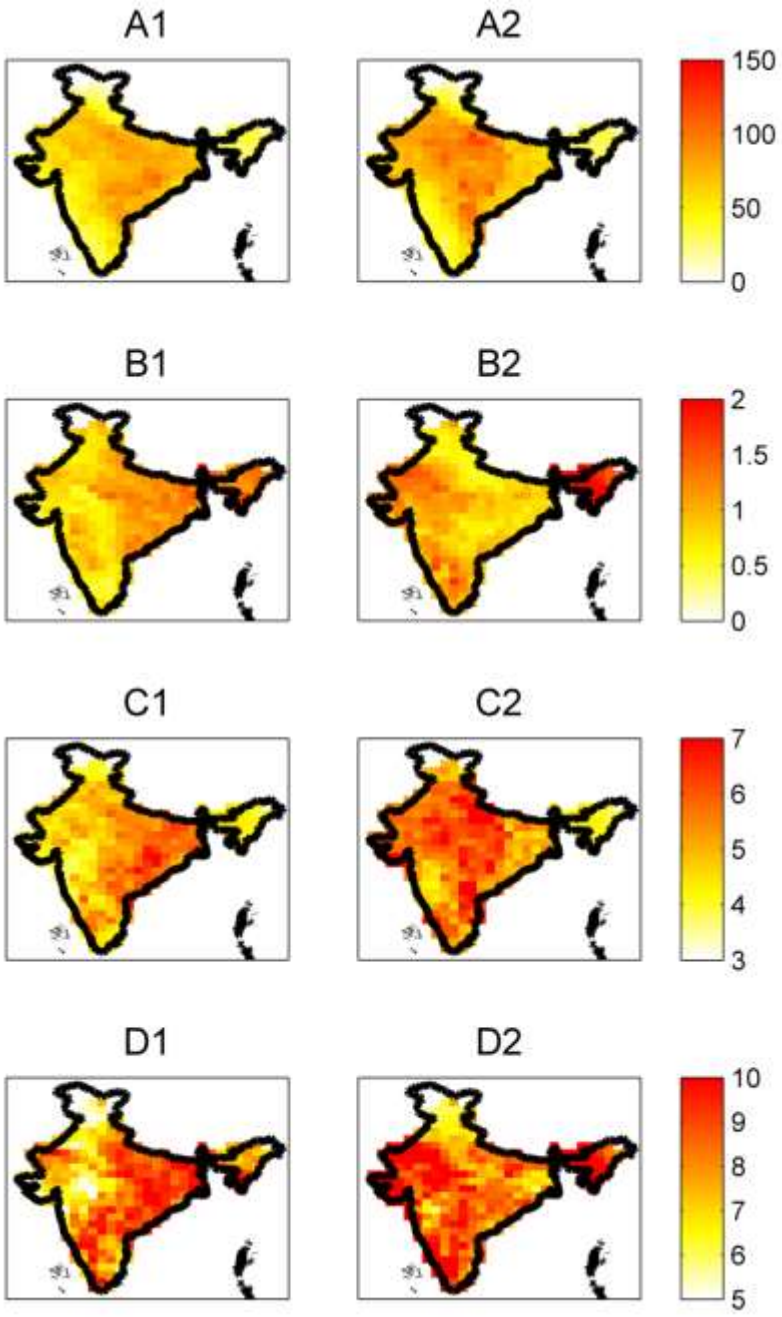


Figure S3 shows the mean heatwave characteristic value for two 25-year periods (1960 – 1984 and 1985 – 2009). Column 1 represents mean heatwave values for 1960 – 1984 and column 2 represents mean values for 1985 – 2009. Row ‘A’ represents accumulated intensity (°C), row ‘B’ represents count (Number of Events), row ‘C’ represents duration (days), and row ‘D’ represents heatwave days (days). For example, B1 represents the mean heatwave count during 1960 – 1984, and C2 represents the mean heatwave duration during 1985 – 2009.

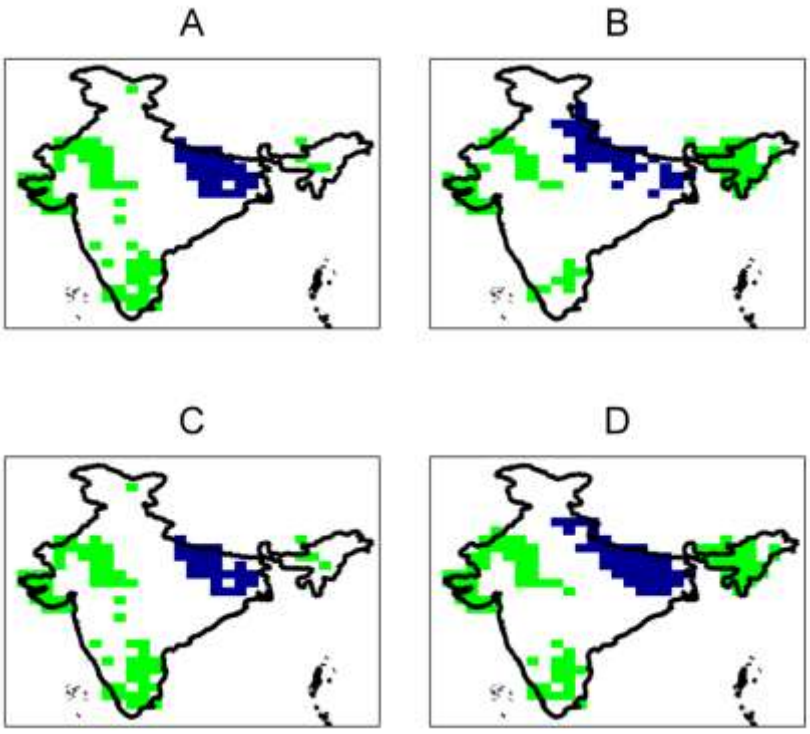



Figure S4 shows the trend in the accumulated intensity, count, duration, and heatwave days of Indian heatwaves' distribution functions from 1960 – 2009 based on the Mann-Kendall trend test. Subplot 'A' represents accumulated intensity, subplot 'B' represents count, subplot 'C' represents duration, and subplot 'D' represents heatwave days. The green pixels show locations where there is a statistically significant (significance level 0.05) positive trend, and the blue pixels show locations where there is a statistically significant negative trend. South and western India show that there is a significant increase in heatwave duration, frequency, and severity.

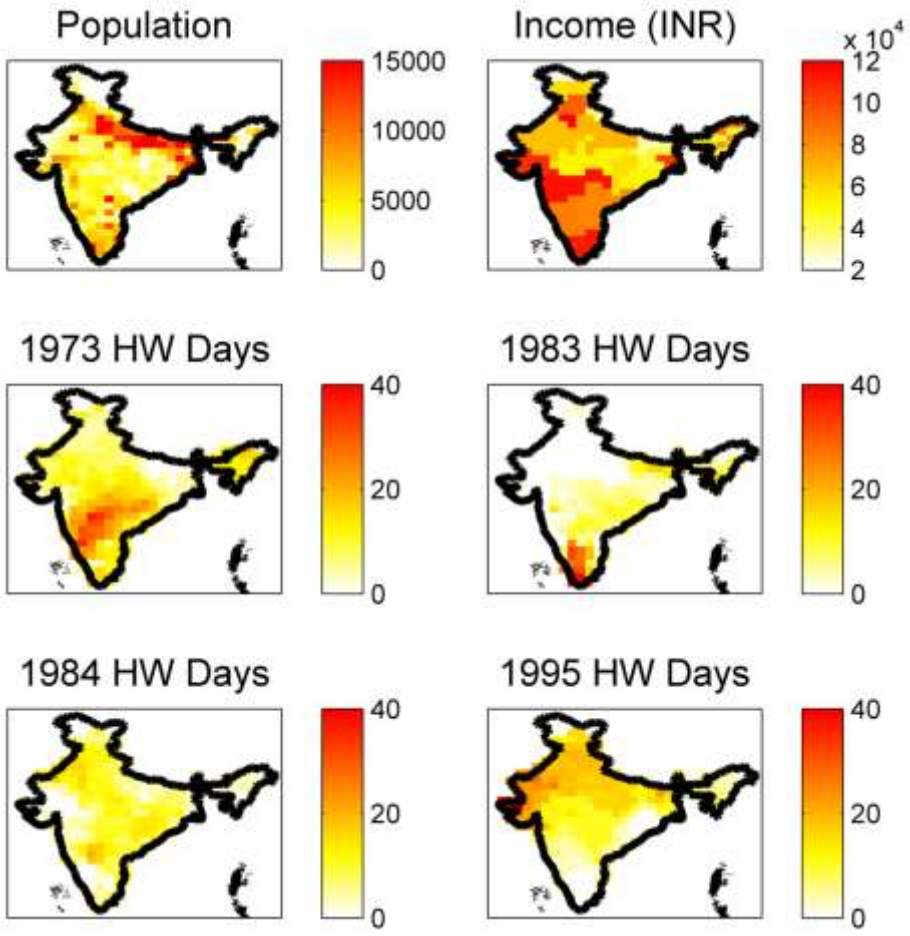



Figure S5 shows the population and income spatial distribution in India, and the number of heatwave days that occurred in 1973, 1983, 1984, and 1995. This figure shows that although many heatwaves occurred during those years, they occurred in less populous and/or wealthier regions, and therefore caused a low heat-related mortality rate.

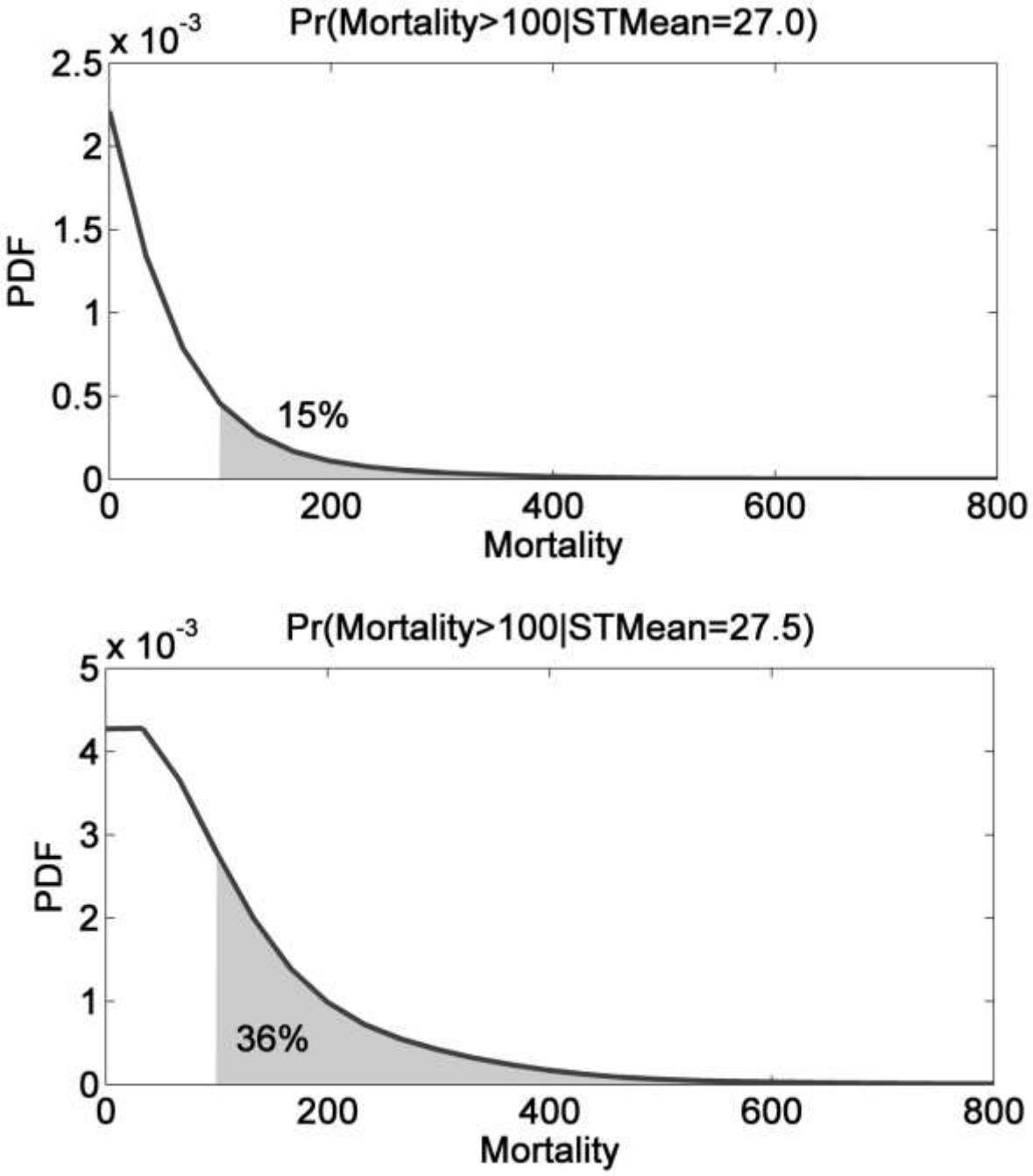



Figure S6 shows the results of a conditional probability density analysis of mortality given certain thresholds for summer maximum temperatures. This figure shows that there is 15% probability that years with summer maximum temperatures equal to 27° C will have mass heat-

related mortality. However, with an increase of only 0.5 °C in summer maximum temperatures, the probability of mass heat-related mortality jumps by a factor of 2.4.

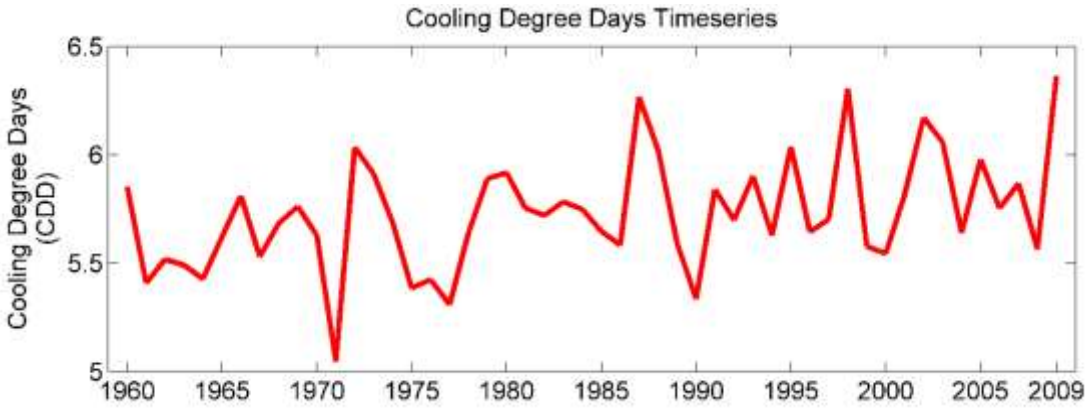



Figure S7 shows that cooling degree days have increased substantially from 1960 –to 2009. The time series exhibits a statistically significant (p=0.05) upward trend confirmed using the Mann-Kendall trend test.

Table S1: Maximum Likelihood and p-values for mean summer mean temperature (MST Mean)/mortality and heatwave days (HW Days)/mortality for different copula families. Columns 2 and 4 show Maximum Likelihood values and columns 3 and 5 show their corresponding p-values. The copula family with the highest maximum likelihood value with a p-value below 0.05 was chosen to be used for Figure 4.

| Copula Family | MST Mean MLV | MST Mean p-val | HW Days MLV | HW Days p-val |
|---|---|---|---|---|
| Gumbel | 11.6 | 0.080 | 14.2 | 0.149 |
| Clayton | 6.5 | 0.001 | 10.0 | 0.004 |
| Frank | 7.9 | 0.004 | 10.6 | 0.016 |
| Normal | 9.8 | 0.018 | 13.5 | 0.090 |
| t | 10.0 | 0.027 | 13.5 | 0.102 |

Table S2: Values of mortality, heatwave days, and summer mean temperatures. I obtained the datasets from the India Meteorological Department. The mortality data is from 1967 – 2006, while the heatwave days and summer mean temperature data is from 1960 – 2009.

| Year | Mortality | HW Days | Summer Mean Temperatures |
|---|---|---|---|
| **1960** | | 9.050739 | 27.8521519 |
| **1961** | | 5.754764 | 27.4080983 |
| **1962** | | 5.930383 | 27.51762347 |
| **1963** | | 4.150327 | 27.49037859 |
| **1964** | | 6.845197 | 27.42619596 |
| **1965** | | 6.960335 | 27.62023659 |
| **1966** | | 7.707924 | 27.80948661 |
| **1967** | 0 | 7.401298 | 27.52937173 |
| **1968** | 0 | 4.352241 | 27.68388849 |
| **1969** | 0 | 6.658364 | 27.76147249 |
| **1970** | 500 | 6.013894 | 27.62789409 |
| **1971** | 0 | 3.595238 | 27.04596265 |
| **1972** | 1200 | 10.61896 | 28.03449524 |
| **1973** | 317 | 12.34712 | 27.9105063 |
| **1974** | 11 | 4.621805 | 27.68548921 |
| **1975** | 43 | 6.398347 | 27.3842219 |
| **1976** | 34 | 4.960204 | 27.42377627 |
| **1977** | 0 | 4.33396 | 27.30842847 |
| **1978** | 368 | 8.435231 | 27.63775698 |

| | | | |
|---|---|---|---|
| **1979** | 361 | 7.516383 | 27.88949626 |
| **1980** | 156 | 7.535881 | 27.91779152 |
| **1981** | 75 | 6.17398 | 27.75313802 |
| **1982** | 0 | 6.224471 | 27.71969241 |
| **1983** | 185 | 10.95383 | 27.78297578 |
| **1984** | 58 | 8.298676 | 27.74795401 |
| **1985** | 284 | 5.653175 | 27.64714428 |
| **1986** | 156 | 6.512237 | 27.58309827 |
| **1987** | 91 | 8.188689 | 28.26588249 |
| **1988** | 627 | 9.663396 | 28.01920181 |
| **1989** | 42 | 6.89859 | 27.59048017 |
| **1990** | 2 | 4.103175 | 27.33463379 |
| **1991** | 252 | 6.455109 | 27.84021383 |
| **1992** | 114 | 6.826083 | 27.69787413 |
| **1993** | 42 | 8.6519 | 27.90152599 |
| **1994** | 234 | 7.829858 | 27.62782151 |
| **1995** | 413 | 11.53829 | 28.03556385 |
| **1996** | 26 | 6.818267 | 27.64455423 |
| **1997** | 21 | 7.587665 | 27.70257507 |
| **1998** | 1655 | 18.17089 | 28.30480507 |
| **1999** | 126 | 4.680534 | 27.57711712 |
| **2000** | 55 | 4.672401 | 27.54351233 |
| **2001** | 70 | 6.206297 | 27.81637593 |
| **2002** | 909 | 8.443608 | 28.17109168 |
| **2003** | 1494 | 13.15832 | 28.05730882 |
| **2004** | 237 | 5.69694 | 27.63898922 |

| | | | |
|---|---|---|---|
| **2005** | 326 | 9.186835 | 27.97736915 |
| **2006** | 135 | 5.959325 | 27.75310397 |
| **2007** | | 7.791435 | 27.86753166 |
| **2008** | | 4.220098 | 27.56406295 |
| **2009** | | 9.273163 | 28.36352389 |

In Figure 3, the annual number of heatwaves and annual mortality rates are standardized for better visualization using the following formula:

$$Standardized = \frac{x - \bar{x}}{\sigma}$$

where $\bar{x}$ is the dataset mean and $\sigma$ is the dataset standard deviation.

# Chapter 2b

## Generalized Conditional Probability Approach for Studying Inter-related Extremes: Tropopause Level Pressure impact on California Precipitation

**Sea surface temperatures and teleconnection patterns such as El Nino/La Nina are considered the main culprits behind major California droughts. However, the underlying relationship between sea surface temperatures (SSTs) and precipitation anomalies is relatively weak. In 2015-2016 the most extreme El Nino did not lead to a wet season as expected, which triggered a series of studies on this topic. Here I show that tropopause level pressure in a region in the northeastern Pacific Ocean (dubbed the PARS-NEP region) plays a major role in whether California will experience a wet or dry year and often dominates the role of SST-based teleconnections. My results indicate that pressure in the PARS-NEP region Granger-Causes precipitation in California during the wet season. I show that when pressure in the PARS-NEP region is in the lower (upper) tertile, 85% of wet seasons across California have a positive (negative) precipitation anomaly. The observed relationship between PARS-NEP and California precipitation is stronger than all the commonly used SST-based climatic indictors frequently used for understanding causes of droughts.**

Meteorological droughts are by far one of the most expensive natural disasters, and can lead to pervasive impacts, including energy, water, and food deficiencies, as well as wildfires [*Wilhite*, 2005]. Many regions are subject to water stress due to high populations and relatively low annual rainfall, but proactive drought management is possible with reliable seasonal precipitation prediction systems. Through satellite observations, reanalysis products, and macro-scale hydrologic models, drought monitoring capabilities have been rapidly growing. However, drought prediction at a seasonal and semi-annual scale remains a great challenge due to limited precipitation forecast skills [*M Hoerling et al.*, 2014; *S D Schubert et al.*, 2008], which may be a result of an inadequate understanding behind the physics of rainfall events.

Most studies that focus on the cause of California droughts relate droughts to sea surface temperatures and climatic teleconnections [*Bradley et al.*, 1987; *DeFlorio et al.*, 2013; *M P Hoerling et al.*, 1997; *Kurtzman and Scanlon*, 2007; *McCABE and Dettinger*, 1999; *Rasmusson and Wallace*, 1983; *Ropelewski and Halpert*, 1986; *Trenberth*, 1984; 1997]. Although teleconnections provide valuable information on California precipitation, they are not the only factors in determining dry or wet conditions. While the Pacific Decadal Oscillation (PDO), Multivariate ENSO Index (MEI), and Atlantic Multi-decadal Oscillation (AMO) are related to California precipitation, they exhibit a weak relationship and low predictability [*S Schubert et al.*, 2016].

Current models often result into a strong response in precipitation simulation when there is a sizable anomaly in sea surface temperatures in the equatorial Pacific Ocean [*Kam et al.*, 2014]. I believe that a key limitation in determining drought conditions is too much emphasis on climatic teleconnections, and a lack of understanding on the impacts of other climatic variables affecting drought. For example, the third most extreme El Nino event on record occurred in the winter of 2015-2016, forcing extreme rainfall forecasts in California; however, California remained mostly dry [*Kintisch*, 2016].

Recent studies highlight the inverse correlation of geopotential height on precipitation anomalies in California [*Swain et al.*, 2016; *Swain et al.*, 2014; *Teng and Branstator*, 2016]. Here I show that there is a strong relationship between California precipitation and tropopause level pressure in a region in the northeastern Pacific Ocean, dubbed the PARS-NEP (Pressure of Atmosphere in tRopopauSe-North-Eastern Pacific) Region. Using causality assessment techniques, I explore the role PARS-NEP in wet and dry conditions in California.

## Data and Methods

### Data

We obtained the global tropopause level pressure dataset from the Earth System Research Laboratory's NCEP/NCAR Reanalysis 1 project. The data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/. The dataset's temporal resolution is monthly, and the spatial resolution is 2.5 x 2.5 degrees. I obtained the monthly precipitation dataset from the National Oceanic and Atmospheric Administration's (NOAA) National Climate Data Center (NCDC).

### Statistical Analysis

We determine PARSI by standardizing the 6-month moving average of monthly pressure in the PARS-NEP region, to compare with wet season precipitation. For purposes of comparability, I perform a similar moving average with the other teleconnections.

We similarly standardize the mean monthly values of well-known teleconnections (Multivariate Enso Index (MEI), Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), Pacific/North American teleconnection pattern (PNA), and Arctic Oscillation (AO)). I define the conditional probability as

$$\Pr(Precip_{wet} > P_T | var_m > var_T) = \frac{\Pr(Precip_{wet} > P_T \cap var_m > var_T)}{\Pr(var_m > var_T)} m = \{5 - 11\}$$

and

$$\Pr(Precip_{wet} < P_T | var_m < var_T) = \frac{\Pr(Precip_{wet} < P_T \cap var_m < var_T)}{\Pr(var_m < var_T)} m = \{5 - 11\}$$

where $Precip_{wet}$ is the sum of precipitation during the wet season (NDJFMA), $P_T$ is the threshold for precipitation, *var* is the standardized mean monthly values of my variable, $var_T$ is the threshold for *var*, and *m* is the first month of the 6 month moving average.

**Results**

In my study, I compare prediction capabilities of the Pressure of Atmosphere in tRopopauSe Index (PARSI) to commonly used teleconnections and climate indices (i.e., Multivariate Enso Index (MEI), Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), Pacific/North American teleconnection pattern (PNA), and Arctic Oscillation (AO)). I calculate PARSI by standardizing the 6-month moving average of the mean monthly pressure in the PARS-NEP region, shown in Figure 1a.

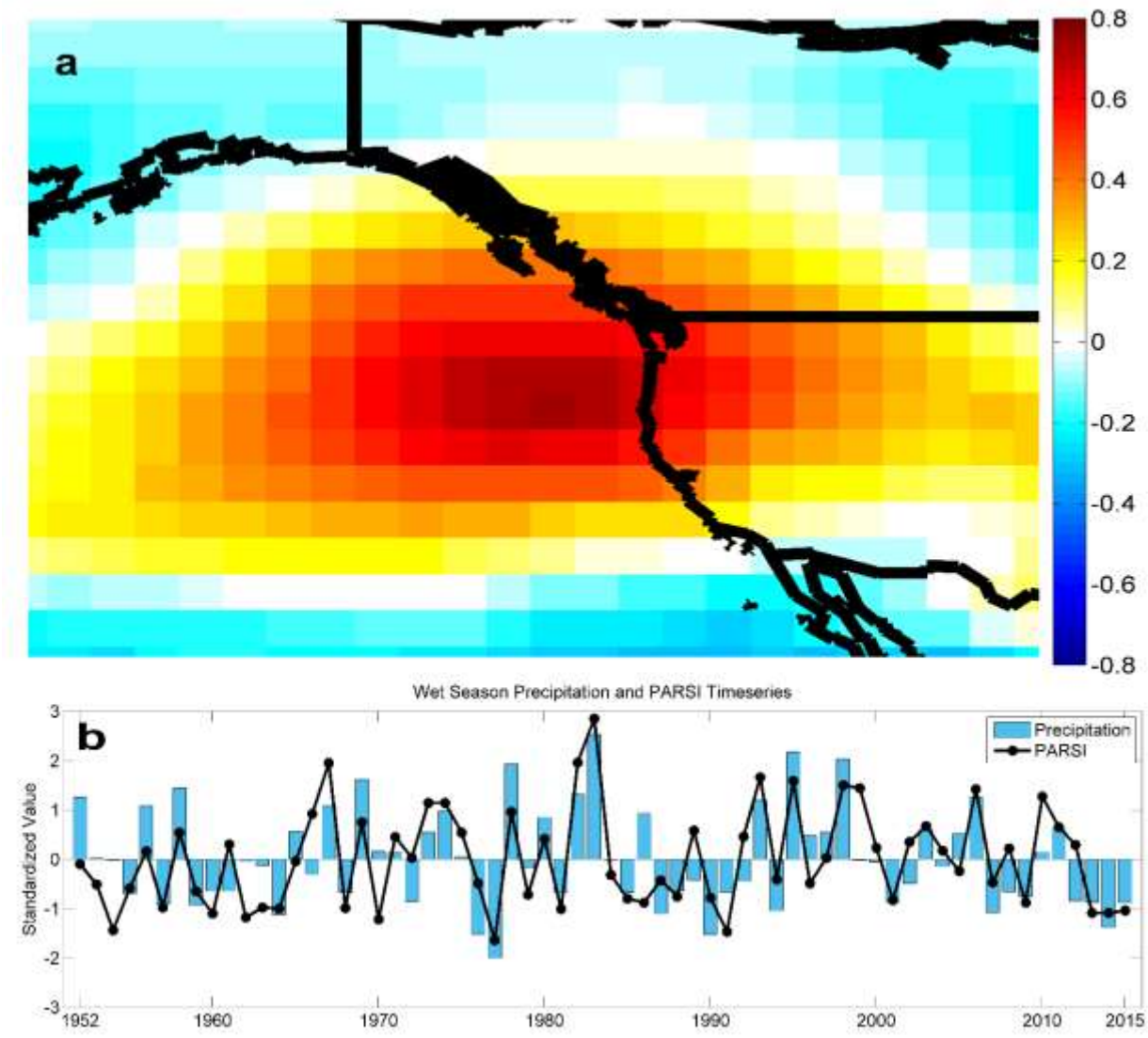



**Figure 1| Precipitation and PARSI timeseries and PARS-NEP location**. The location of the PARS-NEP Region is between latitude: 42.5 – 47.5, longitude: -124.6 - -134.7 (a). California wet season precipitation closely follows PARSI (b) with a correlation coefficient of 0.72.

Standardization transforms the datasets to have a mean of zero, and a standard deviation of one. The reason I standardize the datasets is to improve comparability between the variables; that all variables transform to being on the same scale, both in terms of mean and variance. Therefore, $PARSI_{Apr}$ is the standardized mean monthly pressure of November – April and $PARSI_{Oct}$ is the standardized mean monthly pressure of May – October. Throughout this paper, I compare averages of all variables across a six-month period for purposes of comparability, so

ENSO $SOI_{Apr}$ is the standardized average ENSO SOI of November – April and ENSO $SOI_{Oct}$ is the standardized average ENSO SOI of May – October. My results show a significantly greater correlation coefficient (+0.72) between $PARSI_{Apr}$ and California NDJFMA precipitation than with any other climate index or teleconnection with any 6-month window (highest correlation: ENSO $SOI_{Feb}$, (SONDJF): -0.31). Figures S2 and S3 compare the relationship of ENSO $SOI_{Feb}$ and $PARSI_{Apr}$ with California NDJFMA precipitation. I also show that $PARSI_{Apr}$ statistically significantly Granger Causes [*Granger*, 1969] precipitation throughout the wet period (significance level = 0.05), as can be seen in Table S1. Figure 1b shows standardized average precipitation strongly follows $PARSI_{Apr}$ during the wet season, especially during extreme precipitation events, such as 1977, 1978, 1982, 1983, 1995, 1998, and the 2013 – 2015 drought years.

Figure 2 shows that PARSI exhibits significantly higher predictability relative to the commonly used climate indices. Here, I define predictability by using a simple empirical conditional probability approach. I assess the wet and dry signal of NDJFMA based on the sign of PARSI, MEI, SOI, etc. (i.e., *Pr*(SPI<-0.5 | PARSI<-0.5), *Pr*(SPI>0.5 | PARSI>0.5), *Pr*(SPI<-0.5 | MEI<-0.5), *Pr*(SPI>0.5 | MEI>0.5), etc.) using observed precipitation and climate indices from 1951 – 2015. The results indicate that PARSI exhibits higher predictability relative to the other climate indices, and can potentially serve as a skillful predictor of precipitation in California.

Figure 2 shows the predictability of dry NDJFMA conditioned under different climate indices (MEI, SOI, PDO, PNA, and AO) relative to PARSI. I changed the sign of the variables with negative correlations so that all variables had a positive correlation. I did this to improve comparability between the different variables. Figure 2a shows that the probability of below average precipitation is 80% when PARSI is below -0.5 standard deviations from its mean, and

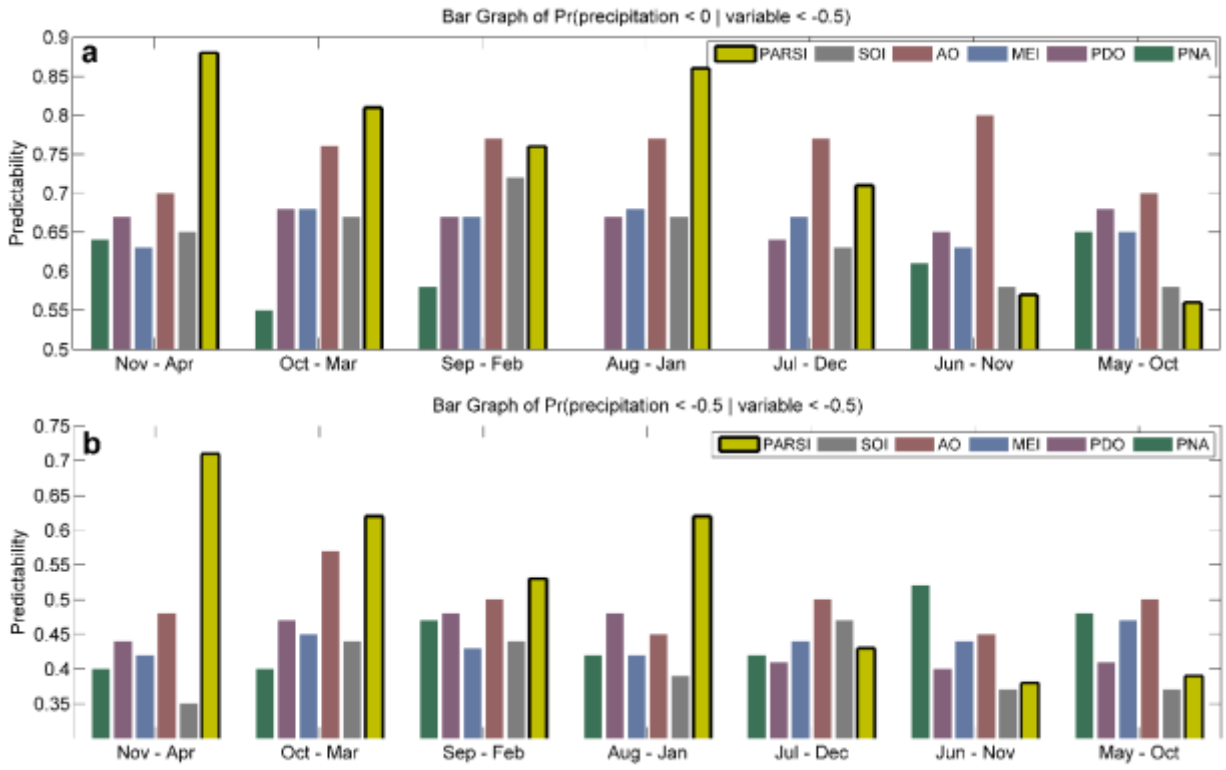



**Figure 2 | Probability of dry NDJFMA conditioned under different climate indices (PARSI, MEI, SOI, PDO, PNA, and AO) relative to PARSI**. Probability of below average precipitation conditioned on PARSI and climate indices being in the lower tertile (a), and probability of precipitation in the lower tertile conditioned on PARSI and climate indices being in the lower tertile (b) for 6 month periods beginning from May - November.

Figure 2b shows the probability of precipitation being in the lower tertile is 71% when PARSI is below -0.5 standard deviations from its mean. Figure S1 shows predictability of wet NDJFMA conditioned under PARSI and the different climate indices, and shows that PARSI has a significantly higher predictability compared with the climate indices (MEI, SOI, PDO, PNA, and AO).

Figure 3 portrays the timeseries of November – February California precipitation and pressure in the PARS-NEP region. This figure shows that the PARSI values strongly correspond with the 2012-2016 California drought. However, in 2017, PARSI shows a significant positive anomaly, corresponding to the third wettest NDJF in my record. The majority of precipitation events in

California occur between November and April, mainly due to the wintertime polar jet stream and prevailing North Pacific storm track [*Swain et al.*, 2016]. I argue that PARSI has a direct relationship with wet season California precipitation because the high pressure in the PARS-NEP region diverts the higher latitude atmospheric rivers from California. When there is low pressure in the PARS-NEP region, the polar jet streams continue on their path in higher latitudes, across Oregon, Washington, and British Columbia.

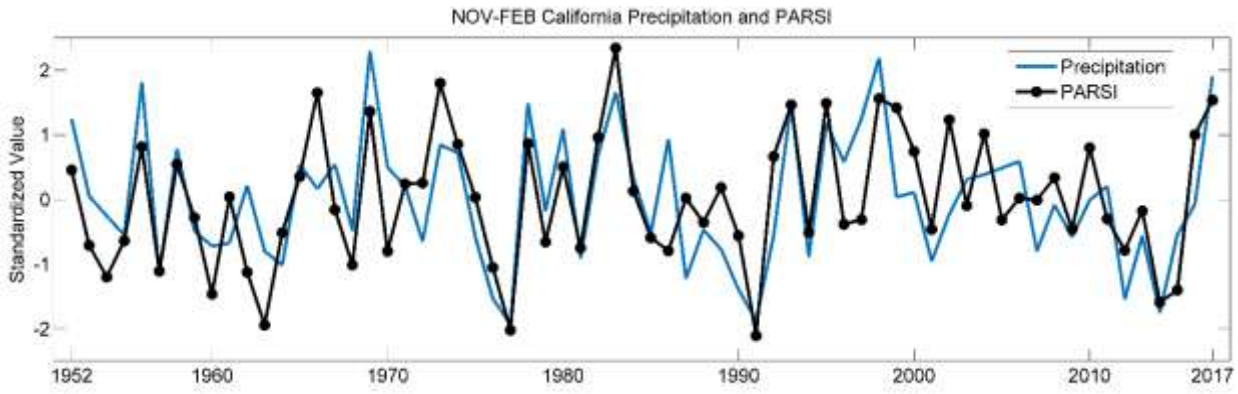



**Figure 3 | Probability of dry NDJFMA conditioned under different climate indices (PARSI, MEI, SOI, PDO, PNA, and AO) relative to PARSI**. Probability of below average precipitation conditioned on PARSI and climate indices being in the lower tertile (a), and probability of precipitation in the lower tertile conditioned on PARSI and climate indices being in the lower tertile (b) for 6-month periods beginning from May - November.

SSTs and teleconnection patterns are currently known as the main cause of California droughts and wet periods. Although they are a major force in affecting precipitation in California, SSTs portray a weak statistical relationship, while PARSI dominates the role in explaining precipitation anomalies. It is important for climate scientists and policymakers to make informed decisions by including all variables that may cause precipitation anomalies, such as pressure in the PARS-NEP region.

**Discussion and Conclusions**

Although sea surface temperatures and teleconnection patterns play an important role in producing precipitation in California, my results show that pressure anomalies in the PARS-NEP region plays an even greater role and can dominate the role of established teleconnections. I show that pressure in the PARS-NEP region has a strong relationship (correlation: 0.72) with California precipitation, especially during extreme dry and wet years, such as 1983, 1998, and the 2013-2015 drought. The fact that California remained mostly dry in 2015-2016 during the third most extreme El Nino event may be explained by a below average PARSI. I show that over 85 percent of wet-season precipitation events were above (below) average when PARSI was in the upper (lower) tertile. My results indicate that PARSI can dominate precipitation and can affect the impacts of well-known teleconnections such as ENSO. My results can provide explanations as to why some extreme teleconnection events do not necessarily lead the expected impacts (e.g., wet or dry condition) on California precipitation. I believe that by incorporating PARS-NEP information can improve explaining precipitation anomalies in California.

## Appendix 2b

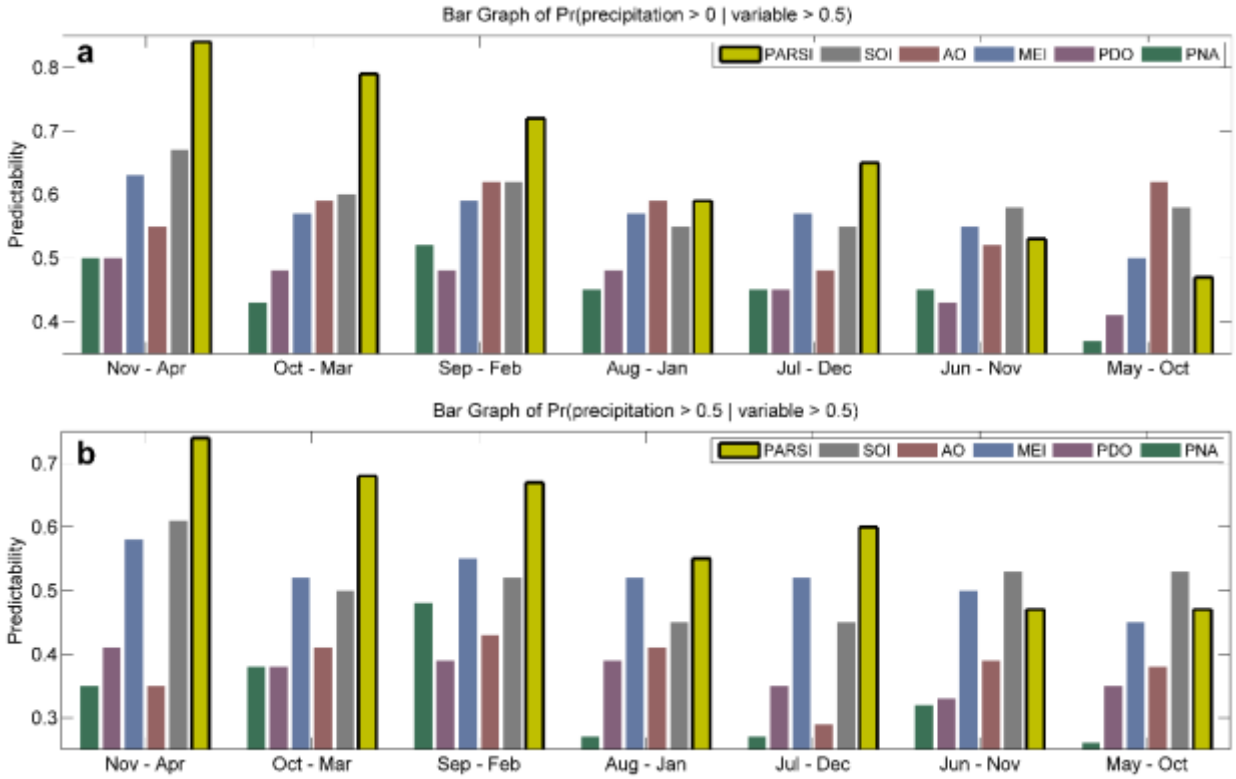



**Figure S1 | Probability of wet NDJFMA conditioned under different climate indices (PARSI, MEI, SOI, PDO, PNA, and AO) relative to PARSI.** Probability of above average precipitation conditioned on PARSI and climate indices being in the upper tertile (a), and probability of precipitation in the upper tertile conditioned on PARSI and climate indices being in the upper tertile (b) for 6 month periods beginning from May – November.

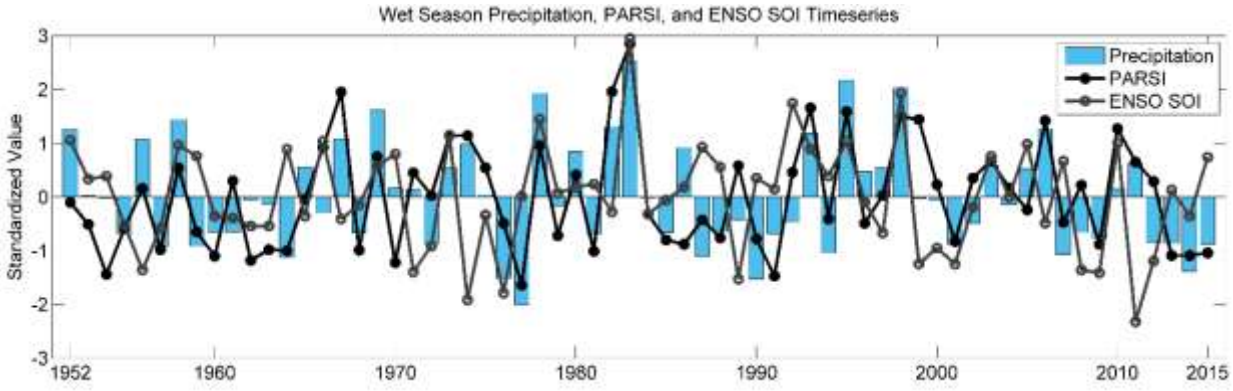
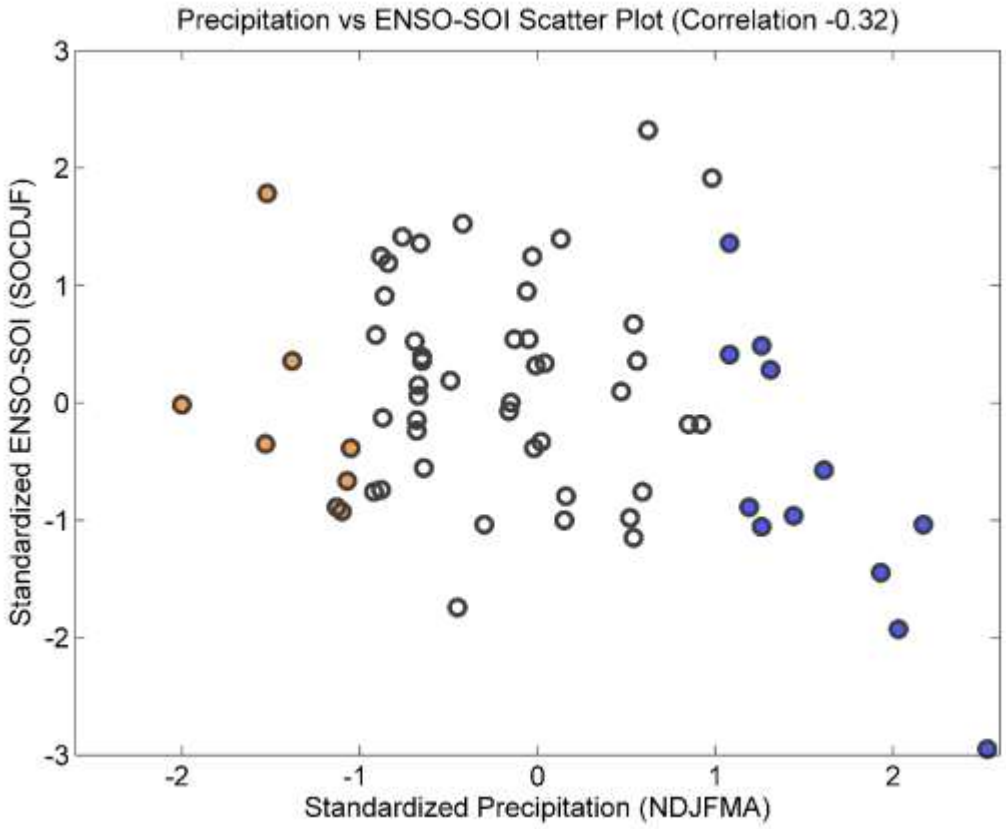



**Figure S2 | PARSI (NDJFMA) and ENSO SOI (SONDJF) timeseries imposed on wet season (NDJFMA) precipitation.** The ENSO SOI timeseries has the highest correlation with California precipitation during the wet season compared to the other teleconnections with any lead time. I switched the sign of SOI for better visualization. The correlation of wet season precipitation with (a) ENSO SOI is -0.31 and (b) PARSI is 0.72.


Precipitation vs ENSO-SOI Scatter Plot (Correlation -0.32)
Standardized ENSO-SOI (SOCDJF)
Standardized Precipitation (NDJFMA)
3
2
1
0
-1
-2
-3
-2
-1
0
1
2

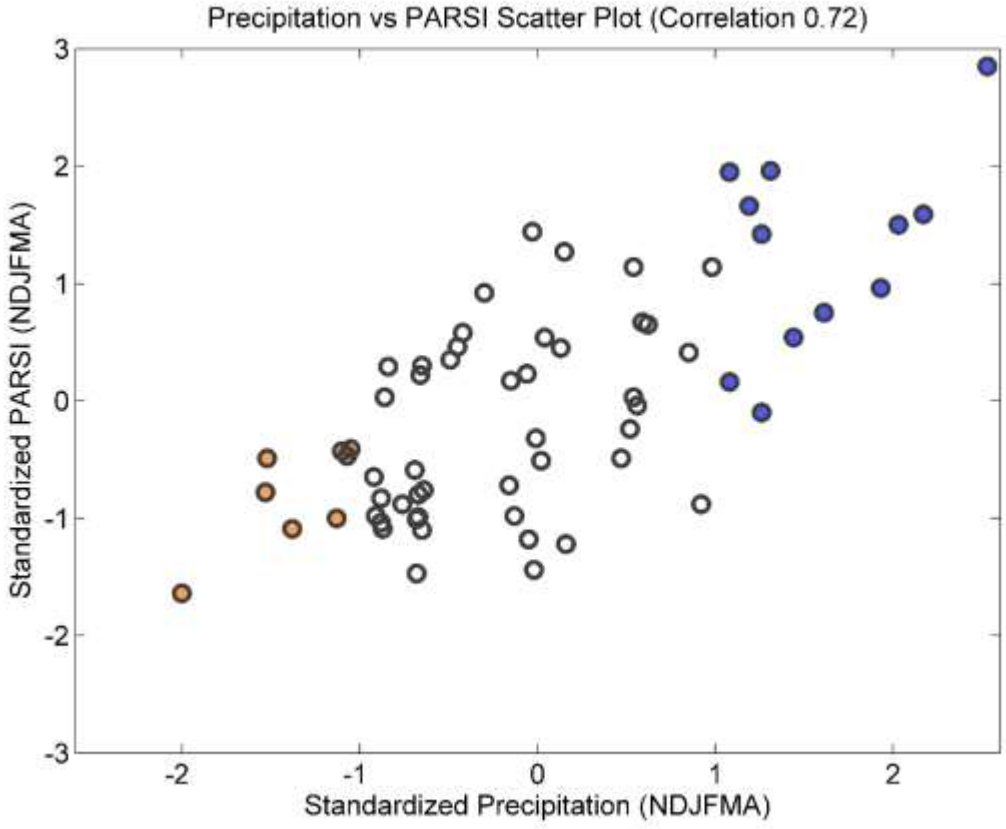

Precipitation vs PARSI Scatter Plot (Correlation 0.72)
Standardized PARSI (NDJFMA)
Standardized Precipitation (NDJFMA)
3
2
1
0
-1
-2
-3
-2
-1
0
1
2

**Figure S3 | PARSI (NDJFMA) and ENSO SOI (SONDJF) scatter plot against wet season (NDJFMA) precipitation.** The ENSO SOI timeseries has the highest correlation with California precipitation during the wet season compared to the other teleconnections with any lead time. The correlation of wet season precipitation with (a) ENSO SOI is -0.31 and (b) PARSI is 0.72.

Table S1: F-Statistic, F-Distribution Critical Value, and p-value for November – April PARSI Granger Causing California precipitation during the wet period (NDJFMA). If the F-Statistic is greater than the F-Distribution Critical Value, I reject the null hypothesis that y does not Granger Cause x. I show that the F statistic is greater than the F-Distribution Critical Value, and thus wet season PARSI does Granger Cause wet season precipitation. Furthermore, I show that the p-value is below the significance level (0.05), also indicating a statistically significant granger causality.

| F-Statistic | F-Distribution Critical Value | p-value |
|---|---|---|
| **66.6418** | 2.7911 | 2.8e-11 |

We standardize the climatic indices for better visualization and comparability using the following formula:

$$Standardized = \frac{x - \bar{x}}{\sigma}$$

where $\bar{x}$ is the dataset mean and $\sigma$ is the dataset standard deviation.

# Chapter 3

# Empirical Teleconnections: Compounding Oceanic-Atmospheric Variables impact on Precipitation

**Atmospheric teleconnections such as the El Niño Southern Oscillation (ENSO) are thought to be important drivers of dry and wet periods around the globe. Yet, even well-studied teleconnections like ENSO have weak statistical relationships with dry/wet periods in most regions and offer limited predictive skill. Here, I demonstrate an exhaustive, bottom-up, brute-force search algorithm for discovering predictive relationships between specific ocean-land regions. Based on 32 years of reanalysis data (1983-2015), I discover linkages that are distinct from known teleconnections, but which better predict precipitation with 4-12 months lead-time. I further evaluate my empirical approach by reversing the analysis, demonstrating that the predictive skill I observe is not mere coincidence; the results have strong spatial coherence. The teleconnections revealed here may have immediate practical value in anticipating seasonal weather events, and also help to focus future research on the physical mechanisms underlying the planet's strongest and most important teleconnections.**

## Introduction

Meteorological droughts are especially costly and disruptive natural disasters that often lead to crop failures, water shortages, wildfires, and increased energy use[*Harto et al.*, 2012; *Saini and Westgate*, 1999; *Westerling and Swetnam*, 2003; *Wilhite*, 2005] . Periods of above average precipitation can be equally destructive, often causing damage to crops, flooding, and landslides. Efforts to avoid these negative impacts depend upon reliable seasonal predictions of precipitation that enable water managers and policy makers to plan and prepare.

Although satellite observations, reanalysis products, and hydrological models have improved in recent years (and now support advanced drought monitoring), seasonal forecasts of precipitation have little skill [*Cayan and Roads*, 1984; *Cayan et al.*, 1999; *Cayan et al.*, 2009; *Martin Hoerling and Kumar*, 2003; *Martin Hoerling et al.*, 2013; *M P Hoerling et al.*, 1997; *Kirtman et al.*, 2014; *Tebaldi and Knutti*, 2007; *White et al.*, 2003]. This is because well-known teleconnections such as El Niño Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), Pacific/North American pattern (PNA), and Arctic Oscillation (AO)—which are the primary basis for such seasonal forecasts[*Bradley et al.*, 1987; *DeFlorio et al.*, 2013; *M P Hoerling et al.*, 1997; *Khalil et al.*, 2007; *Kurtzman and Scanlon*, 2007; *McCABE and Dettinger*, 1999; *Rasmusson and Wallace*, 1983; *Ropelewski and Halpert*, 1986; *Steinschneider and Lall*, 2016; *Trenberth*, 1984; 1997; *Wolter and Timlin*, 1998; 2011]—display weak predictive capabilities[*S Schubert et al.*, 2016]. For example, my results show that ENSO only predicts approximately 40% of wet and dry years across California, while other teleconnections show even weaker relationships.

Here, I present results from an exhaustive, brute-force search algorithm that leads to better predictions of seasonal precipitation (specifically, whether a rainy-season will be dry, wet, or normal). Details of my methodology are in the *Methods* section. In short, my search algorithm determines all possible combinations of a predictor variable (e.g., sea surface temperature) with different lead times (e.g., 4-month to 12-month) which result in a certain outcome (e.g., dry or wet conditions). As demonstrated here, the search first defines the rainy season for ten regions of interest based on a long-term climatology (using monthly data from the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data

Record) then categorizes each year in the record of each location as either wet, dry, or normal (defined as precipitation in the upper 33$^{rd}$ or lower 33$^{rd}$ percentile or between, respectively), and finally evaluates sea surface temperatures (SSTs) at different lead times in all grid cells to identify the best correlations with rainy-season precipitation in each region of interest. The result is a standardized index that forecasts seasonal precipitation based only on SSTs that consistently outperforms models which rely on established teleconnections. In this paper, I refer to the empirical teleconnection as MAZDAK. I use a brute force methodology to determine the empirical teleconnection, using data from the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (See Methods Section)

**Results**

Our results indicate that using the proposed bottom up approach greatly improves seasonal precipitation prediction. In this paper, I refer to the empirically derived teleconnection indicator as MAZDAK. Figure 1a shows a bar graph depicting how well MAZDAK predicts California (hereafter, MAZDAK-CA) rainy season precipitation (Nov – Apr), compared to several well-known teleconnections, such as ENSO-MEI, ENSO-SOI, PDO, PNA, and AO for the historical period 1983 - 2015. I inversed the signs of teleconnections that exhibit negative relationship with precipitation for better visualization and easier comparison. This figure shows that MAZDAK-CA captures wet-, normal-, and dry-years significantly better than commonly used teleconnections, with a much lower false alarm ratio. The MAZDAK-CA index (using only SSTs) was able to capture 80% of dry and wet hits, with below 10% false alarm ratios, with a 7-month lead-time (Feb – Apr of previous year). I note that the existing dynamical and statistical models

do not offer much predictive skill beyond 3 months [*Kirtman et al.*, 2014; *Tebaldi and Knutti*, 2007]. This index shows considerably better predictability with respect to the other teleconnections with the same lead-time. For example, of the teleconnections I used for comparison, SOI has the best predictability, with 60% dry-, and 30% wet-hit rate, and 30% dry and wet false alarm ratio.

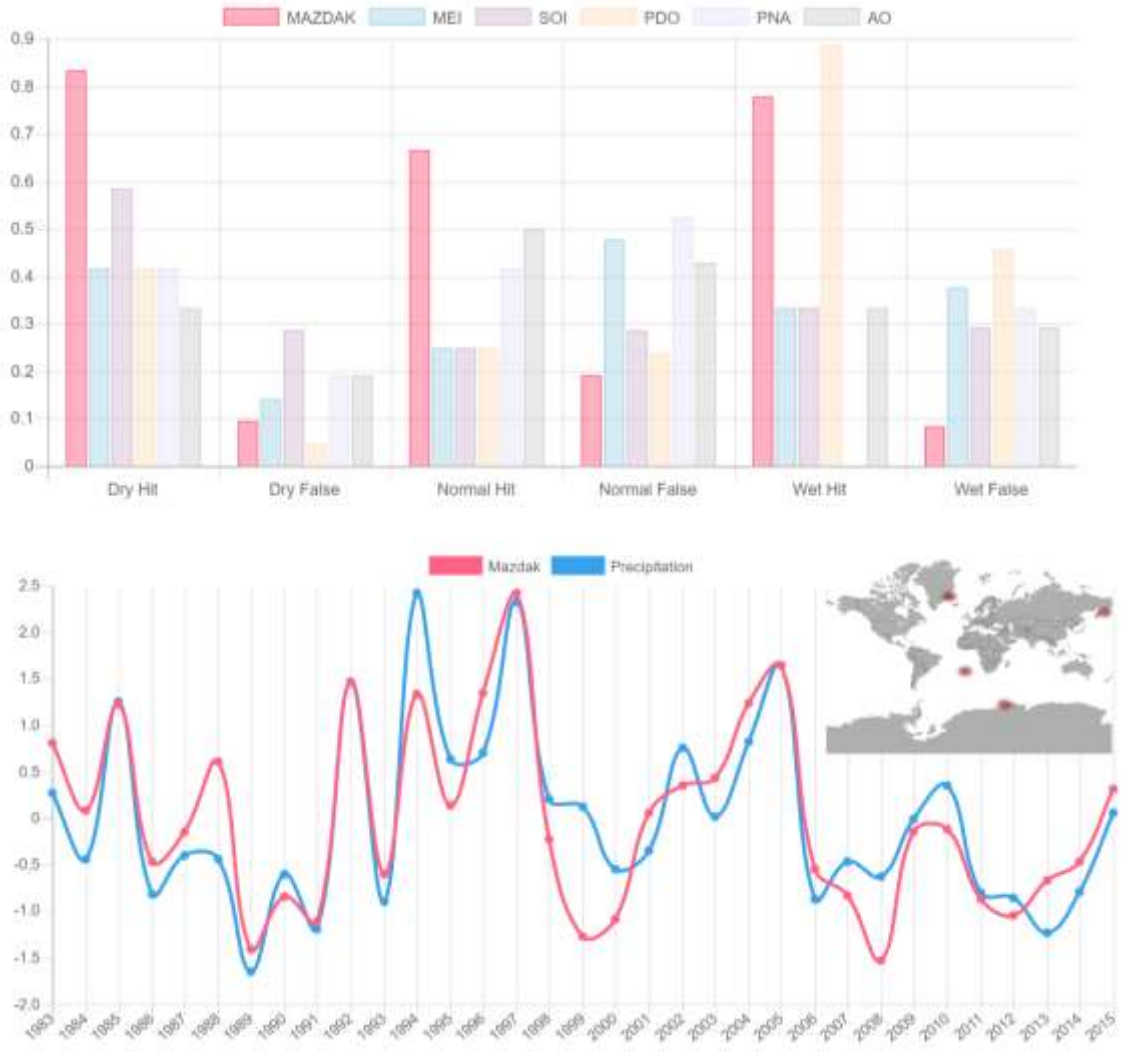



**Figure 1**| **California Precipitation, MAZDAK-CA, and the existing teleconnections**. The bar graph of MAZDAK-CA hit and false alarm rates compared to five well-known teleconnections (a). California wet season precipitation closely follows MAZDAK-CA (b) with a correlation coefficient of 0.86. The locations that provide predictive information are highlighted with red dots in the global map in panel b.

The empirical teleconnection indicates that sea surface temperatures from four seemingly independent locations (see the red dots in the global map in Figure 1b) provide predictive information for California. I note that the predictive points are not located in places where the well-known teleconnections are defined. The output of the model exhibits a strong interdependence with California's wet season precipitation with a correlation coefficient of 0.86 (see Figure). In this figure, I have standardized precipitation values for better comparability with MAZDAK. The model captures California's significant wet and dry years, such as the 1997/1998 precipitation event and the 2011 – 2014 drought (Figure 1b). To examine the robustness of the approach I have conducted similar analysis in nine other locations (political divisions) from around the world with different climate conditions, including: Iran, Hubei (China), Liaoning (China), Ethiopia, New South Wales (Australia), Goias (Brazil), Greece, Odisha (India), Stockholm (Sweden). Figures S1-S30 show the same charts as Figure 1 for the other nine locations. Figures S1-S30 show the top three combinations of SSTs that produce the highest correlation value. These charts similarly show the MAZDAK has significantly higher correlation and hit rate values than the well-known teleconnections. To show that MAZDAK provides predictive information beyond the period I used to determine the SST locations (calibration period), I determine the SST locations using only part of the precipitation data (1983 – 2010). I then extended the timeseries through 2015 to show how well MAZDAK captured precipitation outside of the calibration period. Figures S31 – S60 show the bar graph, line graph, and SST location map of all 10 locations, using precipitation data 1983 – 2010 to determine the locations. The results indicate there is only a slight decrease in correlation of MAZDAK with wet-season precipitation, (e.g., in the case of California, MAZDAK-CA changes from 0.86 to 0.83).


**Figure 2**| **MAZDAK hit rate radar chart**. The radar chart portrays a summary of hit rates of MAZDAK for ten locations across the globe. Dry hit rates (a) and wet hit rates (b) are relatively high. The red line represents the probability (33%) that you will obtain a hit based on long-term climatology. MAZDAK improves the hit ratio for nine of the ten locations.

Figures S61 – S90 show the bar graph and line graph for all 10 locations, using the method of leave one out for calibration. This means that there was no information used to determine the

SST locations for each target year. This is a commonly used metric to determine the reliability and robustness of wet-season precipitation prediction, as there is no precipitation information used to determine the MAZDAK value of each year. The results of leave one out shows that MAZDAK-CA can capture over 70% of dry periods with 15% false alarm, and over 75% of wet periods, with 20% false alarm ratio. Figure 2 shows the summary of dry and wet hits for the ten political divisions I tested, using the method of leave one out. The red line represents the probability (33%) that you will obtain a hit based on persistence-based climatology. MAZDAK produces a dry hit ratio of over 50% in eight of the ten locations and a wet hit ratio of over 60% in seven of the ten locations. This shows that I can create indices based on sea surface temperatures that have significantly improved predictability than current teleconnections as well as persistence based predictions, through MAZDAK. I show that although this method may not be universal for all political divisions (i.e. Goias, Brazil), it does improve precipitation predictability in most regions.

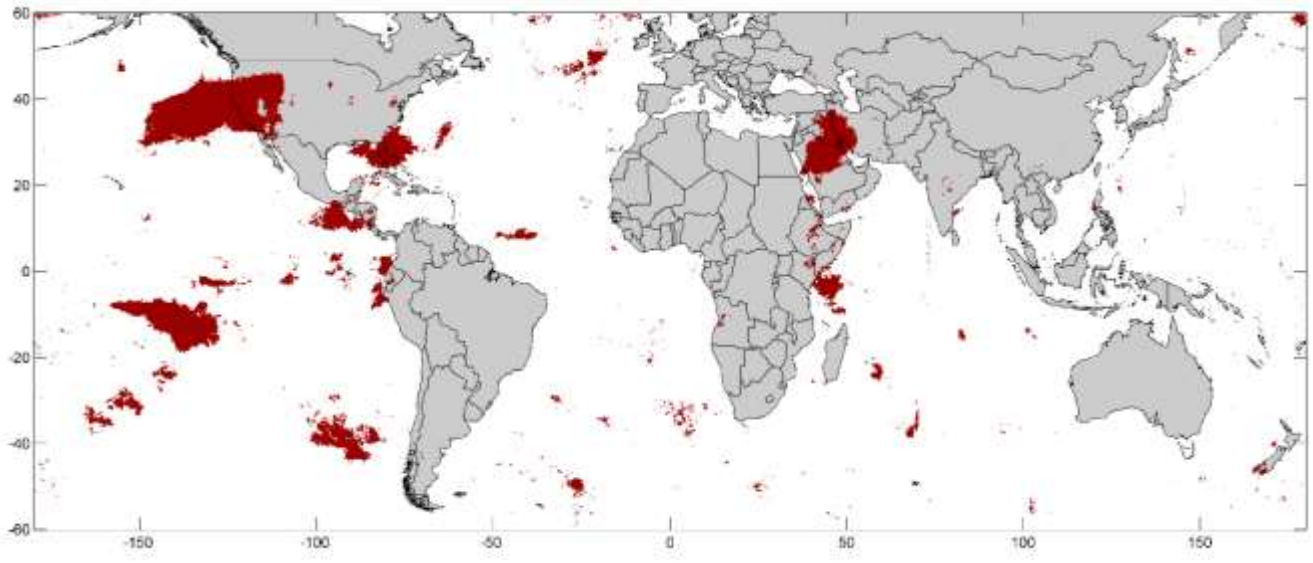


**Figure 3**| **MAZDAK-CA significant correlation map**. This map shows the grids that MAZDAK-CA has a statistically significant (significance level = 0.01) correlation with global precipitation. I show that MAZDAK-CA has a statistically significant correlation with precipitation across California, as well as parts of the ENSO region and the Middle East.

We acknowledge that a strong correlation between two variables does not necessarily prove causality. I attempted to show causality using a backward analysis approach. Here I determine the precipitation grid boxes that have a statistically significant correlation, for example, with the SST locations that provide predictive information for California. In other words, I determine the predictability of MAZDAK across all grid cells and determine the grid boxes that precipitation has a statistically significant correlation with the MAZDAK index. If the area that shows a significant correlation with the predictive points show some consistent spatial pattern, I can infer that the relationship is related to physical processes across coherent geographic regions. Figure 3 shows the regions where wet season precipitation has a statistically significant (significance level = 0.01) correlation with the locations that provide predictive information. I show that MAZDAK-CA has a statistically significant correlation with precipitation across California and southwestern US, demonstrating a striking coherent spatial pattern. This indicates that the observed relationship is not random and involves a data-driven teleconnection that cannot be explained with the well-known teleconnection, moisture transport patterns, etc. It is interesting to note that there is also a significant relationship with MAZDAK-CA and the ENSO region in the Pacific Ocean. This figure portrays that the relationship between MAZDAK-CA and California wet-season precipitation is not a mere coincidence, but rather that there is a strong correspondence to the entire region. Figures S91 – S100 portray similar maps for all the political divisions I tested in this study, and show that the regions in

question have a statistically significant correlation (significance level = 0.01) with their corresponding MAZDAK Index.

**Discussion**

We hypothesize that there are an infinite number of teleconnections (combinations of sea surface temperatures (SST), sea level pressure, ocean current velocity, near-surface ocean temperatures, wind speed, and other oceanic-atmospheric variables) that may serve as predictor variables to droughts and wet periods in many regions across the globe, with several months lead-time. Relationships between precipitation events and their drivers may be hidden and go unseen due to their significant distances in time and space. By using an empirical approach, I can determine the relationship between precipitation events at a specific location, and the oceanic-atmospheric variables across each gridbox around the world. This bottom-up approach identifies the drivers through the resulting impacts, rather than attributing impacts to oceanic-atmospheric patterns (i.e. ENSO) that may or may not be a significant driver. This complete, brute-force search method can highlight teleconnection patterns associated with wet-season precipitation for certain regions and may provide an improvement to current seasonal precipitation prediction models.

In this paper, I show the results of empirically derived teleconnections using only sea surface temperature; however, other oceanic-atmospheric variables may also provide information for seasonal precipitation prediction. To demonstrate potential applications to other variables, Figures S101-S109 show the locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 4 – 12 month lead. I show that ocean current velocity in the mid-Pacific Ocean have a statistically significant relationship with

California precipitation for lead times of 6 – 8 months, which can be incorporated into the empirically derived teleconnection. However, it is also possible for some variables to have no predictive power, as is the case for tropopause level pressure portrayed in Figures S110-S118. These figures shows that tropopause level pressure does not show a strong pattern of high correlation with California wet-season precipitation. This is expected as pressure is a more short-term weather phenomenon. While it is significantly related to local precipitation, it does not offer long-term memory and seasonal predictive power. This means that the results of the search for different variables is consistent with the expected persistence. For different case study or basins, it is important to investigate all possible climatic variables' predictive capabilities using the proposed brute-force methodology, and combine their predictive powers to create a robust model.

This paper suggests that seasonal prediction of climatic events, such as wet-season precipitation can be advanced using data driven models that do not appear to have a physical explanation (at least not at this time). Further research is required to investigate the physical mechanisms behind the relationship between the identified points and the target variable (e.g., seasonal precipitation). In fact, I attempted to explore this by investigating not only sea surface temperatures, but also ocean currents and tropopause level pressure patterns (see Figures 101-118). However, the results did not point us to a specific physical explanation of the predictive skill of the model. I are hoping that future work through future work by scientists with different background (e.g., climatology, oceanography, statistics) lead to a better understanding of the underlying physics of the predictive power of the model.

**Methods**

**Data.** I obtained the sea surface temperature datsaset from the NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST) V2, provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA, from their website at http://www.esrl.noaa.gov/psd/. The sea surface temperature dataset has a temporal resolution of one month, and a spatial resolution of 1° x 1° [*Reynolds et al.*, 2002].

We use the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR) obtained from the Center for Hydrometeorology and Remote Sensing at the University of California, Irvine. PERSIANN-CDR is a newly developed satellite-based precipitation product that covers more than three decades (January 1, 1983 – present) of daily precipitation estimations at 0.25° x 0.25° spatial resolution for the 60°S – 60°N latitude band [*Ashouri et al.*, 2015]. The data can be obtained from: http://chrsdata.eng.uci.edu.

We obtained the global tropopause level pressure dataset from the Earth System Research Laboratory's NCEP/NCAR Reanalysis 1 project. The data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/. The dataset's temporal resolution is monthly, and the spatial resolution is 2.5 x 2.5 degrees. I obtained the monthly precipitation dataset from the National Oceanic and Atmospheric Administration's (NOAA) National Climate Data Center (NCDC) [*Kalnay et al.*, 1996].

We used zonal ocean current velocities in my evaluation of the relationship between ocean currents and precipitation (available for download at ftp://podaac-ftp.jpl.nasa.gov/). The velocity estimates are calculated by Earth Space Research (ESR), using a number of satellites

and in situ observations available through NASA and NOAA. The dataset has a spatial resolution of 1 degree and a temporal resolution of 5 days, which I averaged to the monthly scale.

**Statistical Analysis.** I determine the empirical teleconnection by finding the two to five locations where the 3-month average of SSTs have the highest correlation with each political division's wet-season (6 months), using a brute-force (exhaustive) search algorithm [*Korf*, 1985]. I put a temporal constraint where the SSTs must have a lead-time of between 4 to 12 months. Once I have the SST locations, I determine $MAZDAK_R$ (Raw Empirical Teleconnection):

$$MAZDAK_R = (\pm\, SST1) + \ (\pm SST2) + \ (\pm 1\, SST3) \ + \ (\pm 1\, SST4) \ + \ (\pm 1\, SST5)$$

where SST1, SST2, SST3, SST4, and SST5 are the 3-month averaged SST values. I multiply the SST value by +1 if the correlation with wet-season precipitation is positive, and by -1 if the correlation is negative. I use between 2 and 5 locations to find the empirical teleconnection. The number of points is determined based on which combination of SSTs has the highest correlation. Then I standardize $Tele_R$ to find Tele (Empirical Teleconnection Index):

$$MAZDAK = \frac{MAZDAK_R - mean(MAZDAK_R)}{stdev(MAZDAK_R)}$$

We similarly standardize the mean monthly values of well-known teleconnections (Multivariate Enso Index (MEI), Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), Pacific/North American teleconnection pattern (PNA), and Arctic Oscillation (AO)). I define the conditional probability of wet hit as

$$\Pr(Precip_{wet} > P_T | var_m > var_T) = \frac{\Pr(Precip_{wet} > P_T \cap var_{wet-m} > var_T)}{\Pr(var_{wet-m} > var_T)} m = 4 - 12$$

dry hit as

$$\Pr(Precip_{wet} < P_T | var_m < var_T) = \frac{\Pr(Precip_{wet} < P_T \cap var_{wet-m} < var_T)}{\Pr(var_{wet-m} < var_T)} m = 4 - 12$$

where $Precip_{wet}$ is the sum of precipitation during the wet season (e.g., NDJFMA), $P_T$ is the threshold for (wet or dry) precipitation, *var* is the standardized mean monthly values of my variable, $var_T$ is the threshold for (wet or dry) *var*, and *m* the best 3 months to calculate the empirical teleconnection with a lead-time of 4 – 12 months from the first month of the wet season.

# Appendix 3

**Section 1:**
In this section the results are presented based on analysis of record over 1983-2015 to find the locations that best estimate the precipitation.

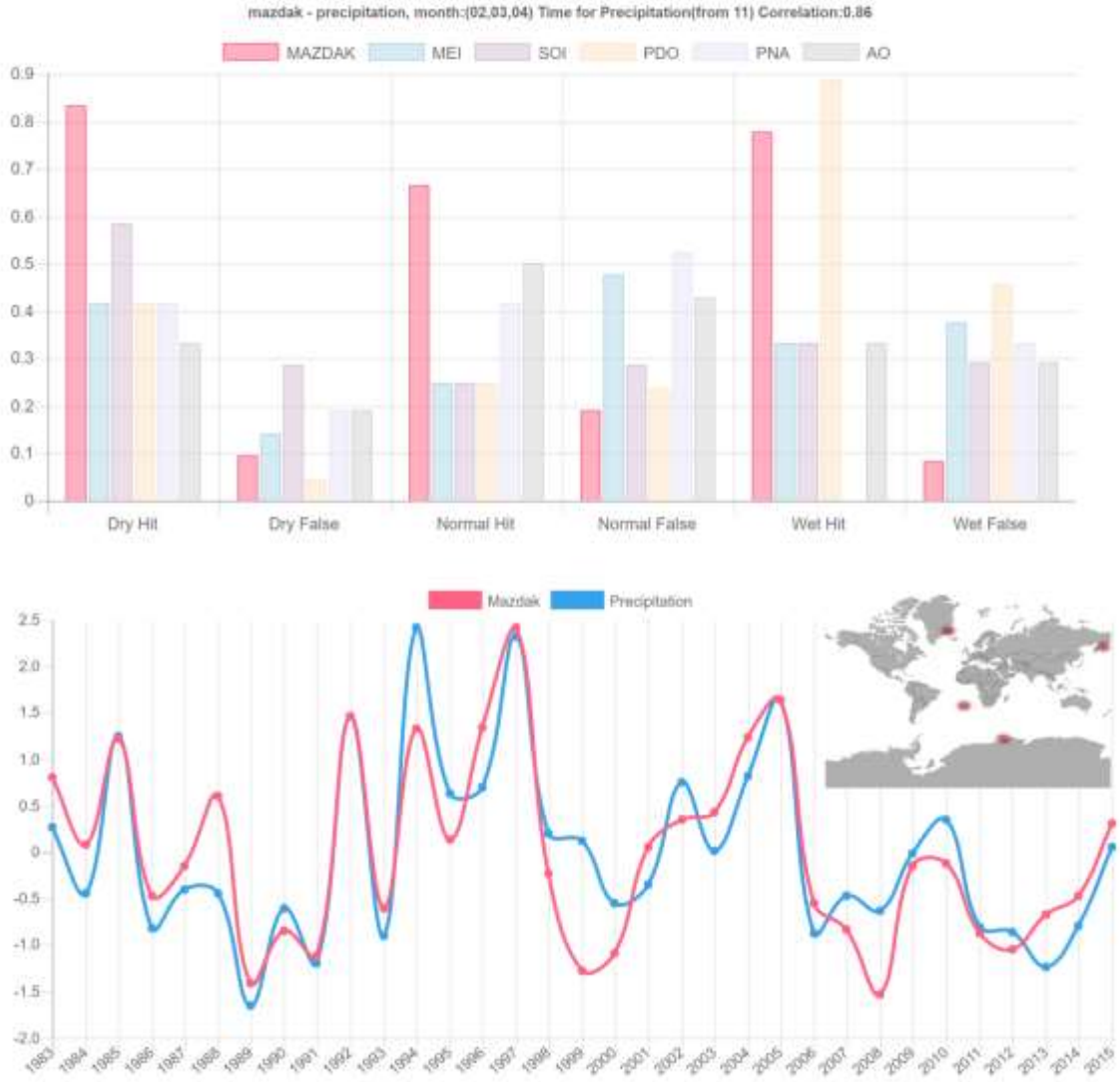



Figure S1| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

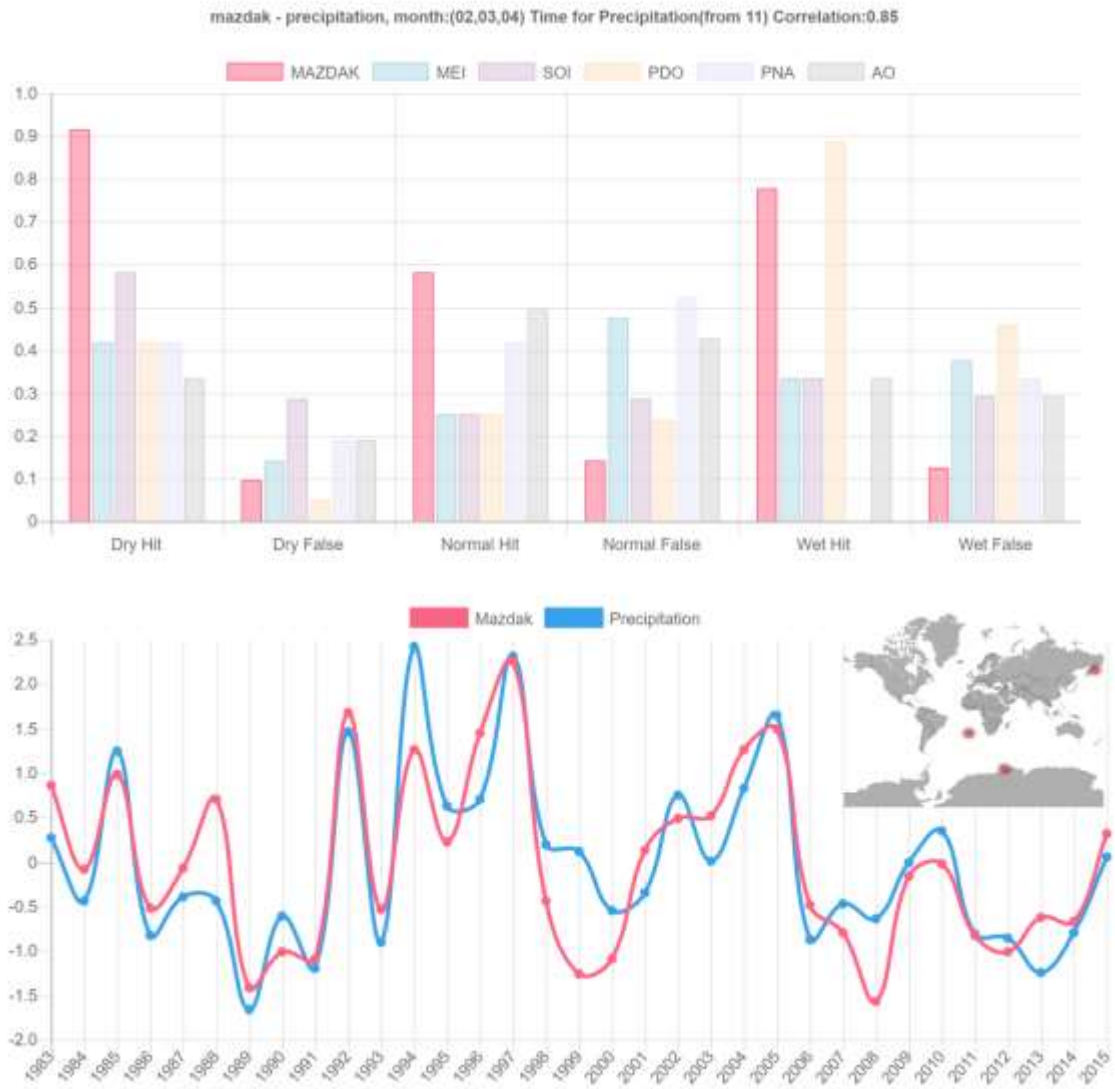



Figure S2| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

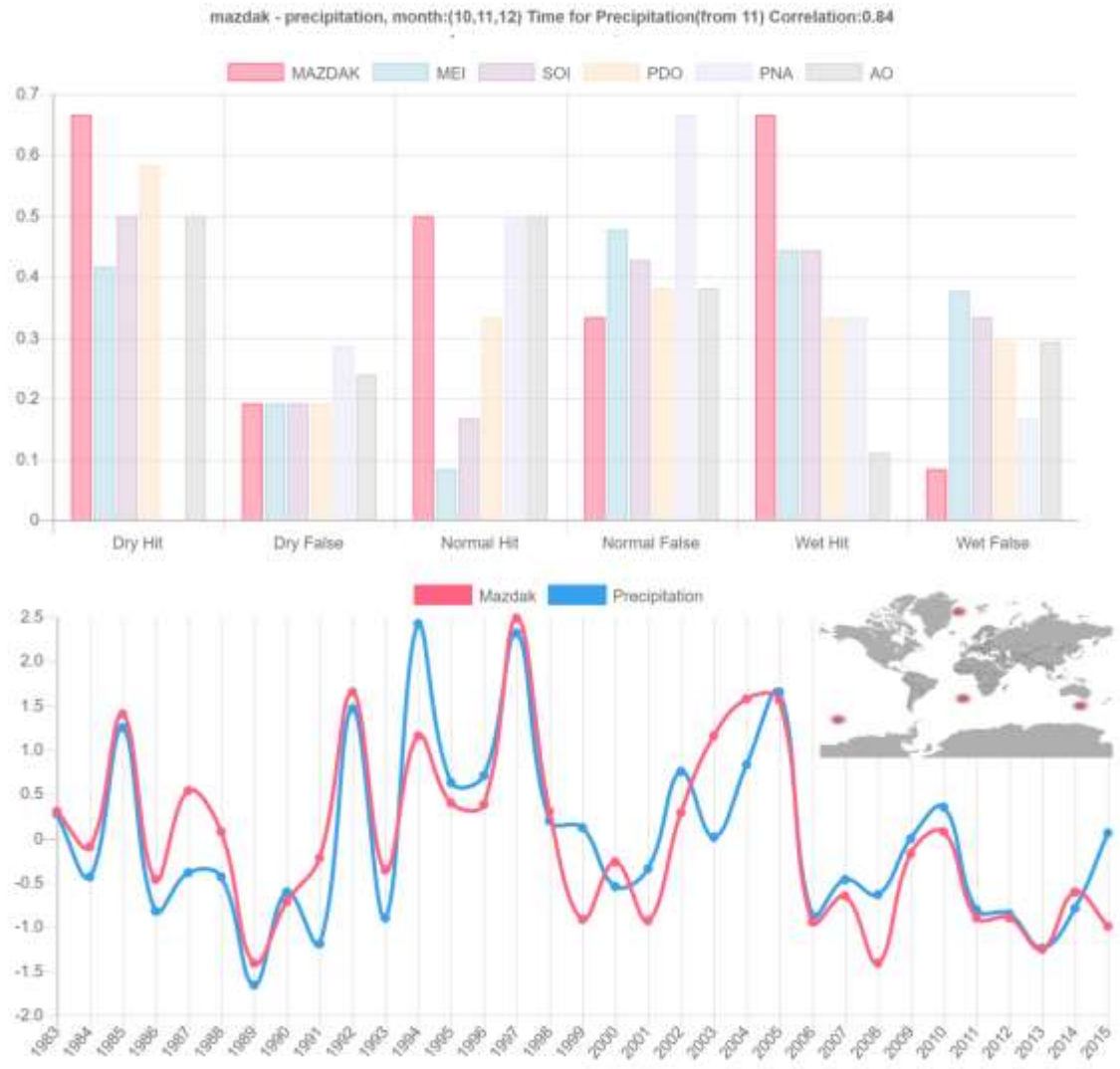



Figure S3| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

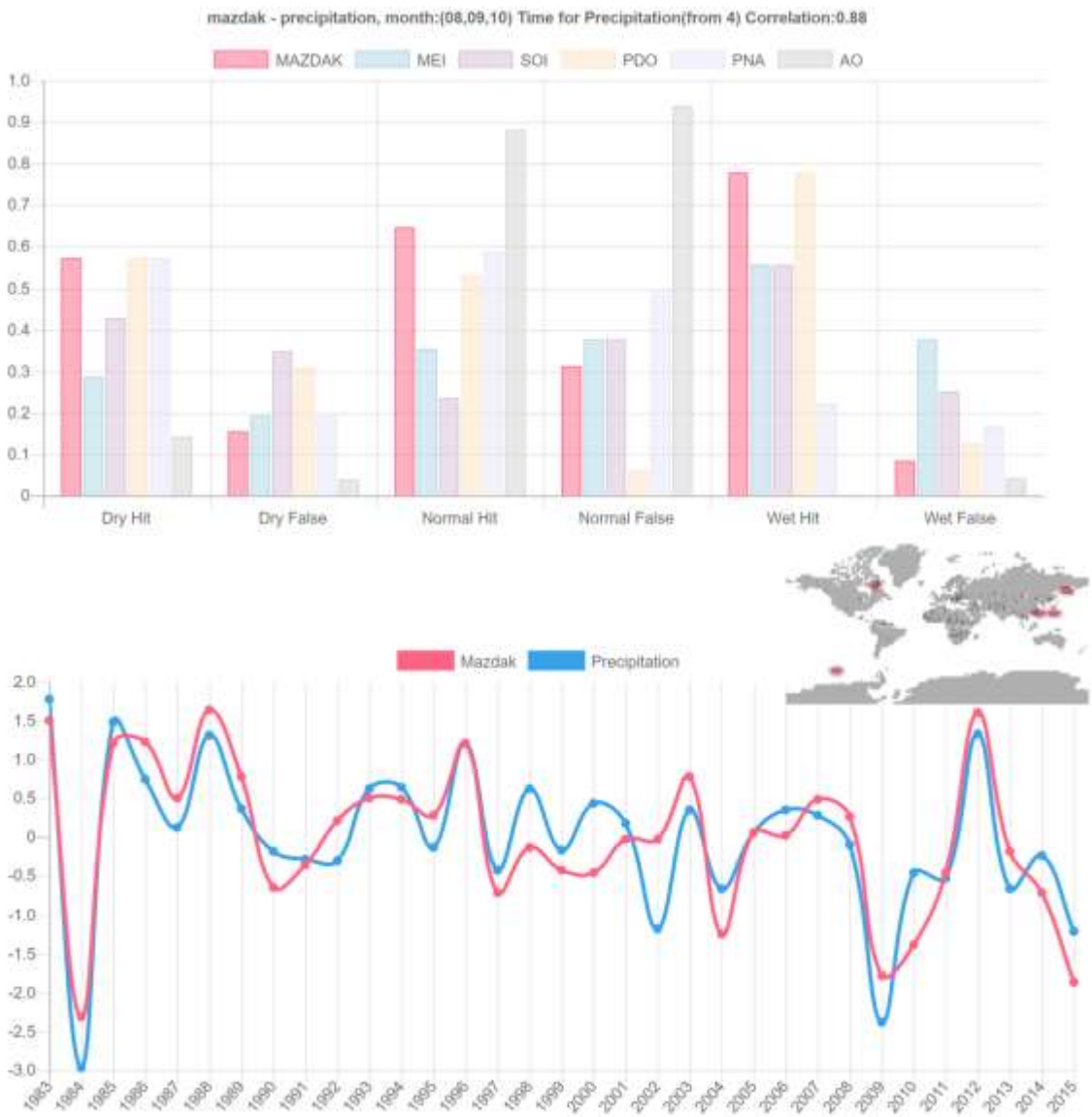



Figure S4| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

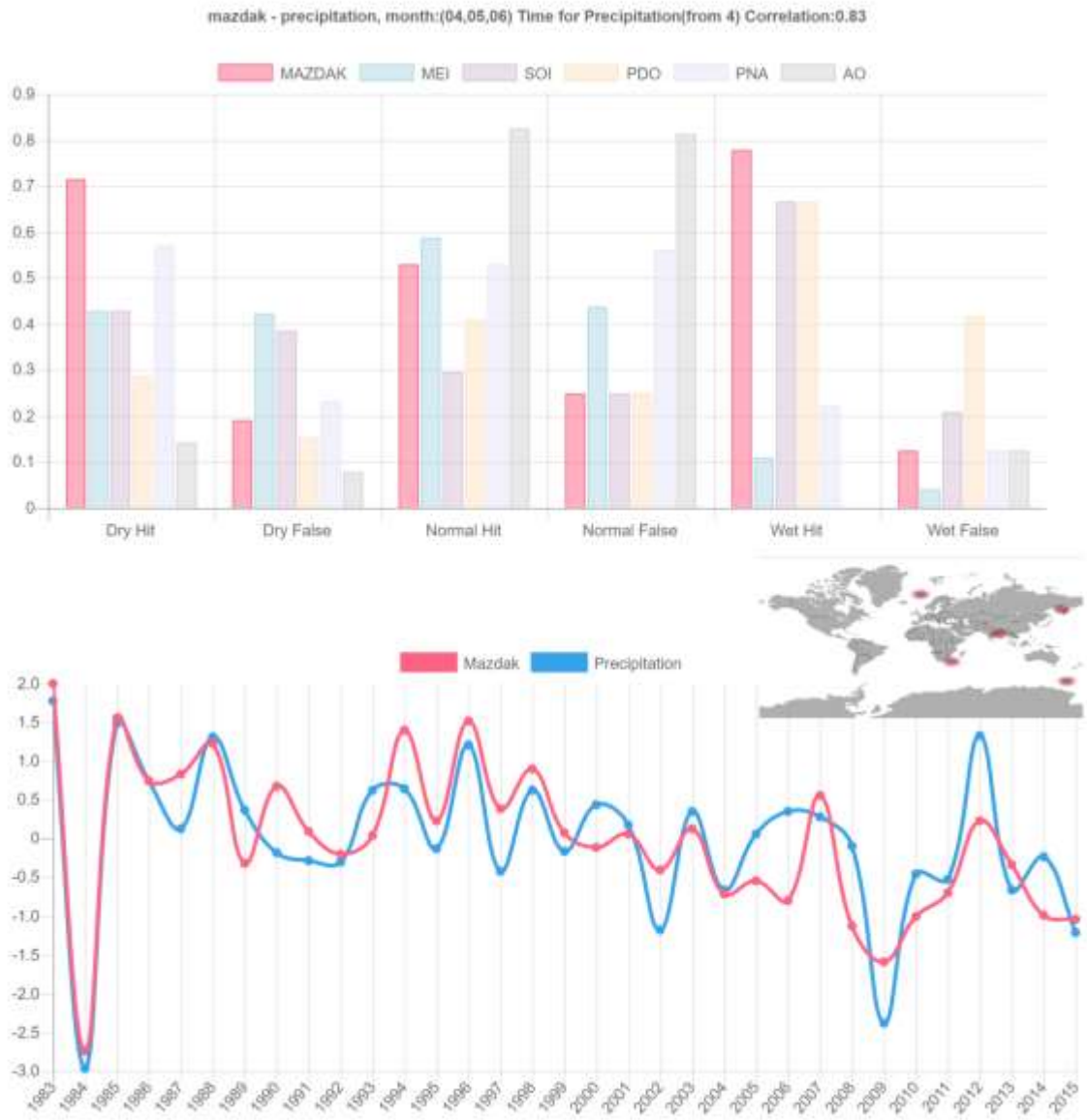



Figure S5| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the Second highest correlation coefficient.

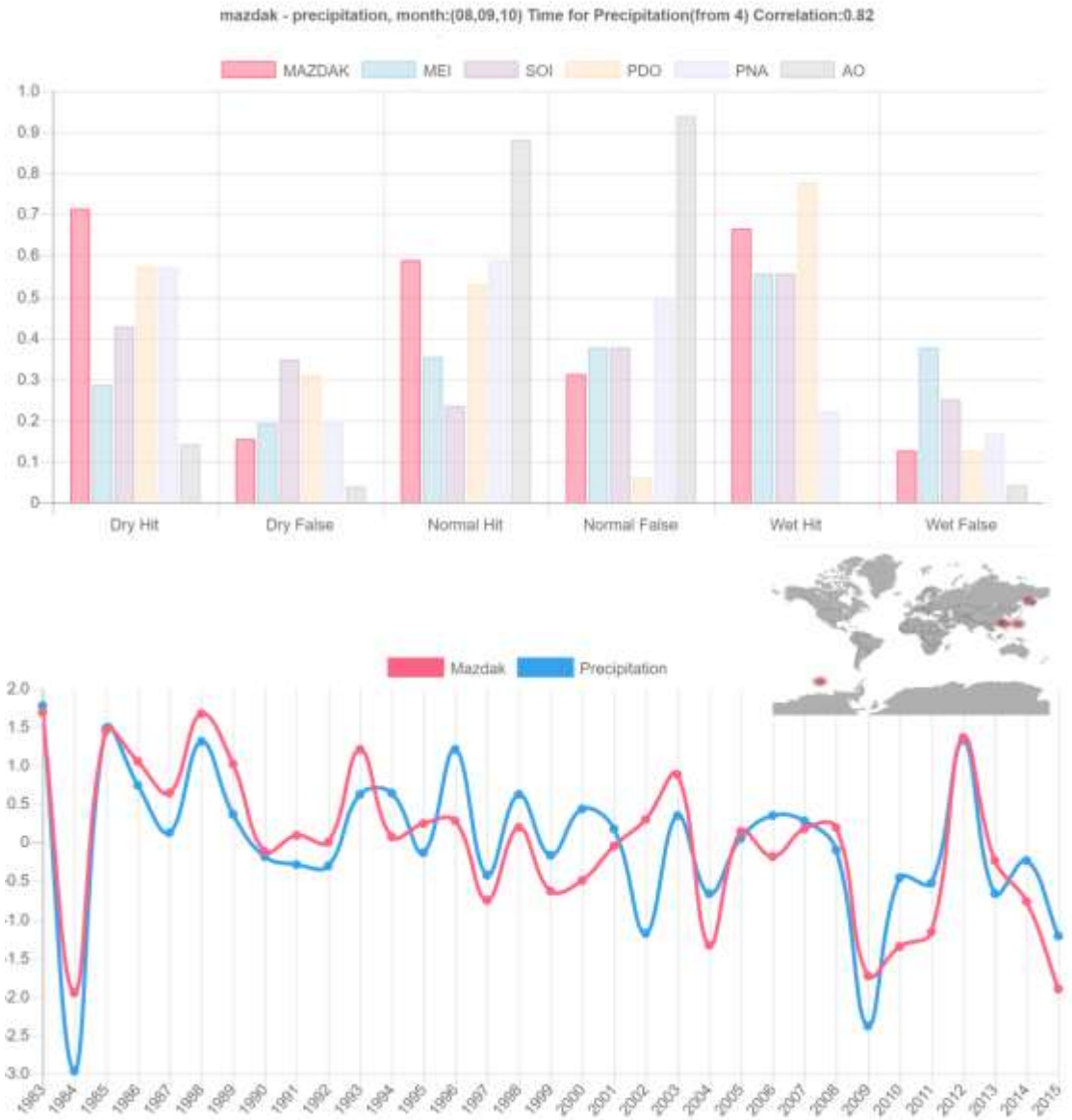



Figure S6| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

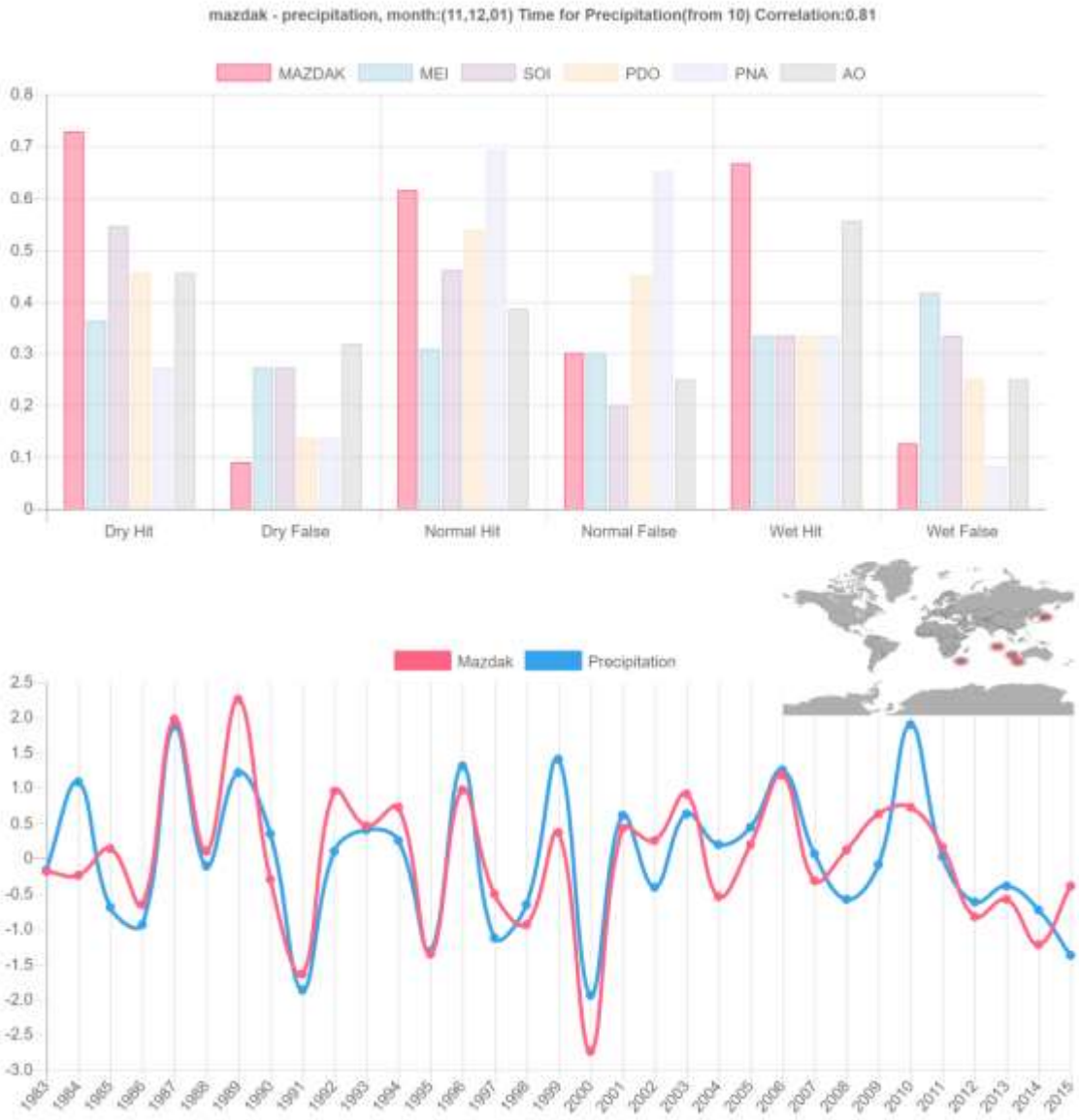



Figure S7| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

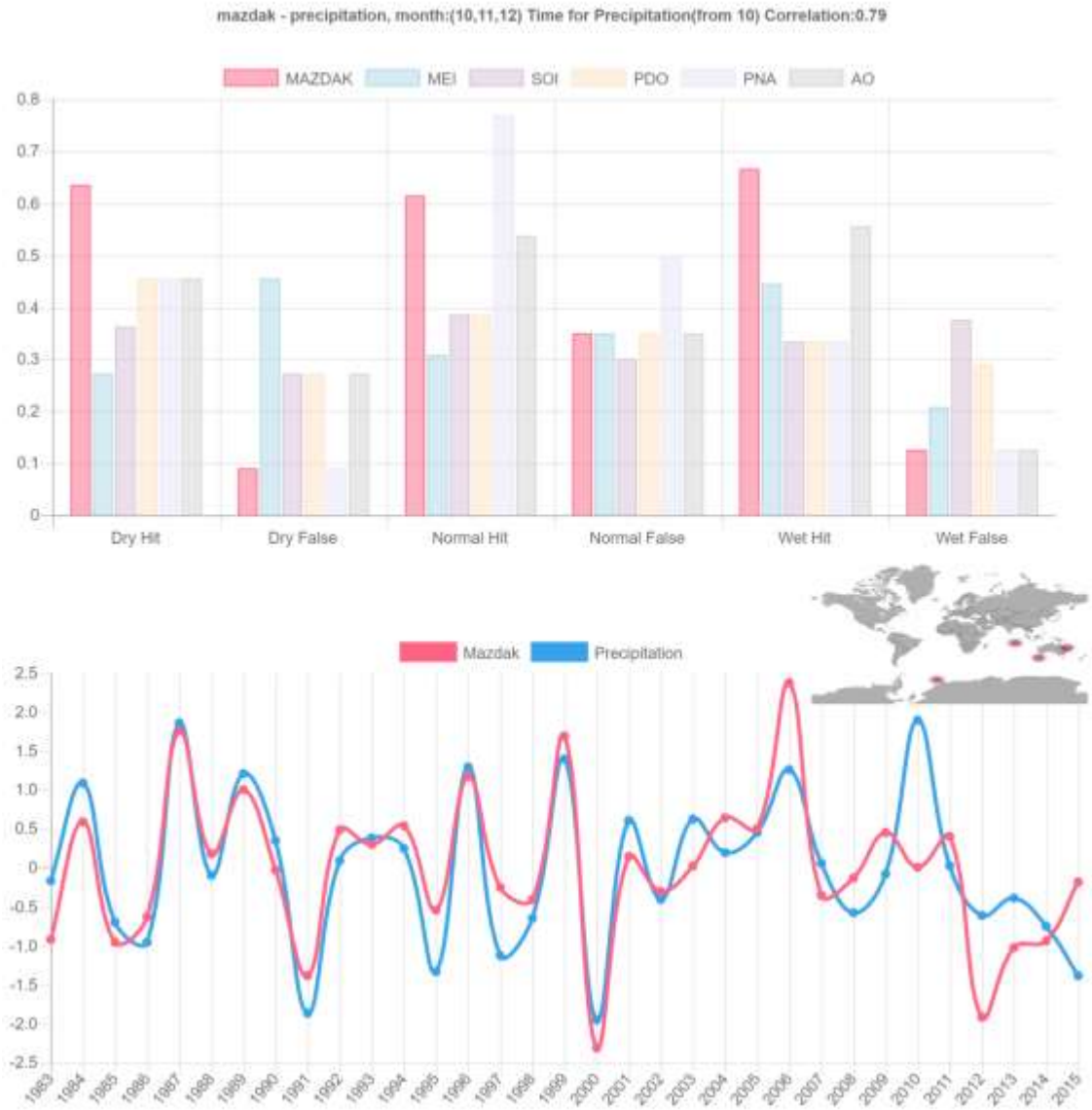



Figure S8| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

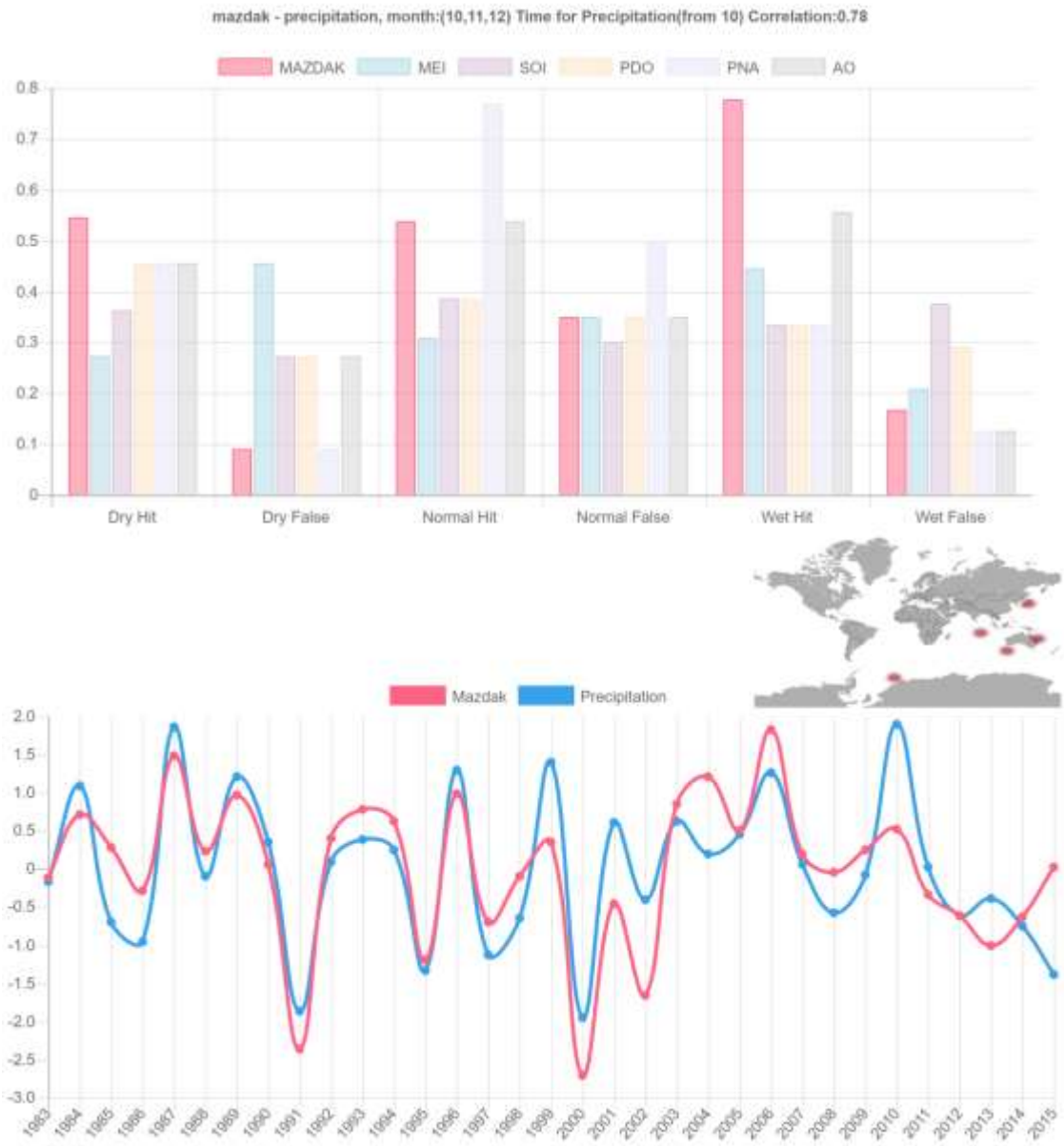



Figure S9| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

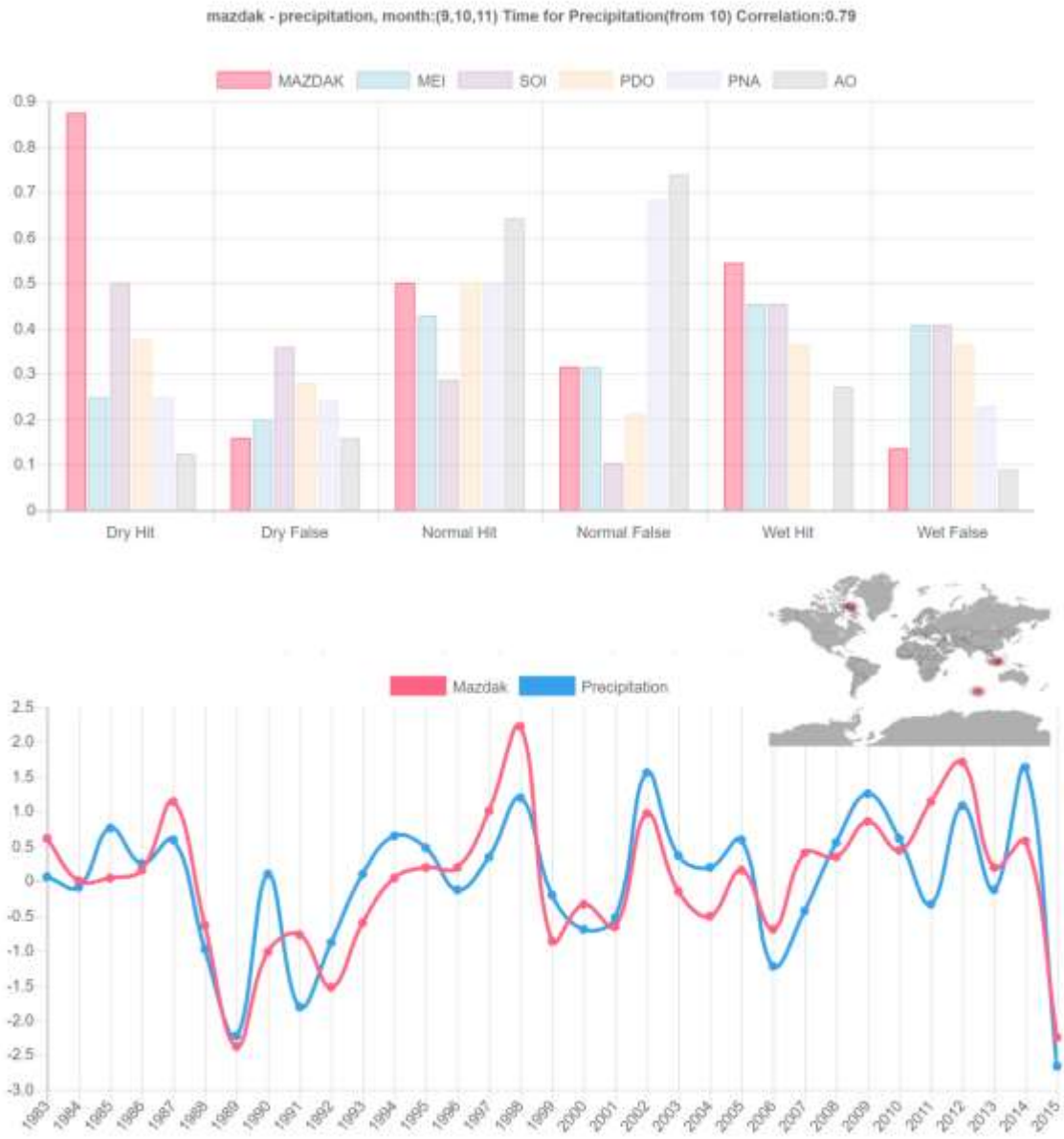



Figure S10| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

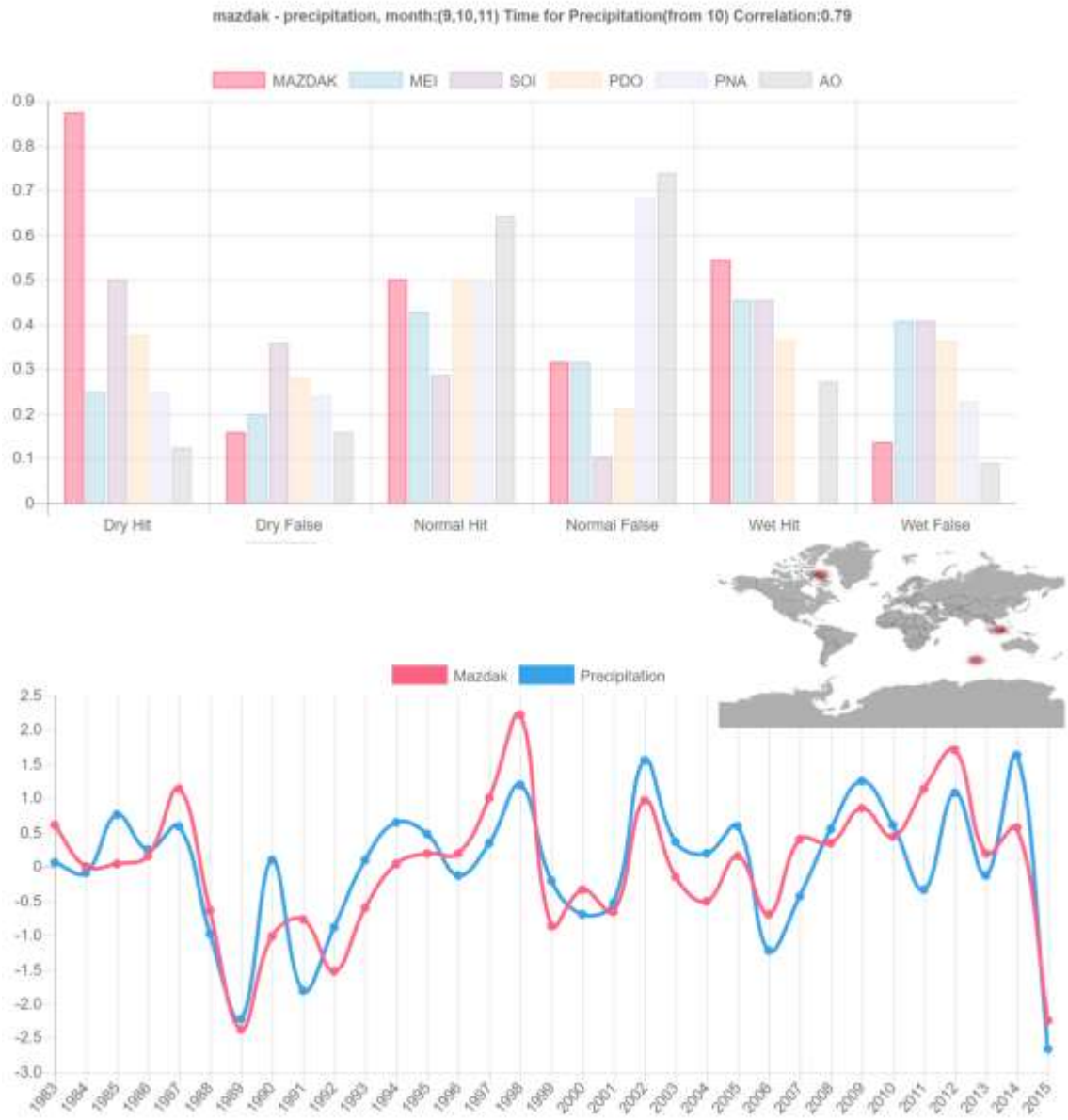



Figure S11| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

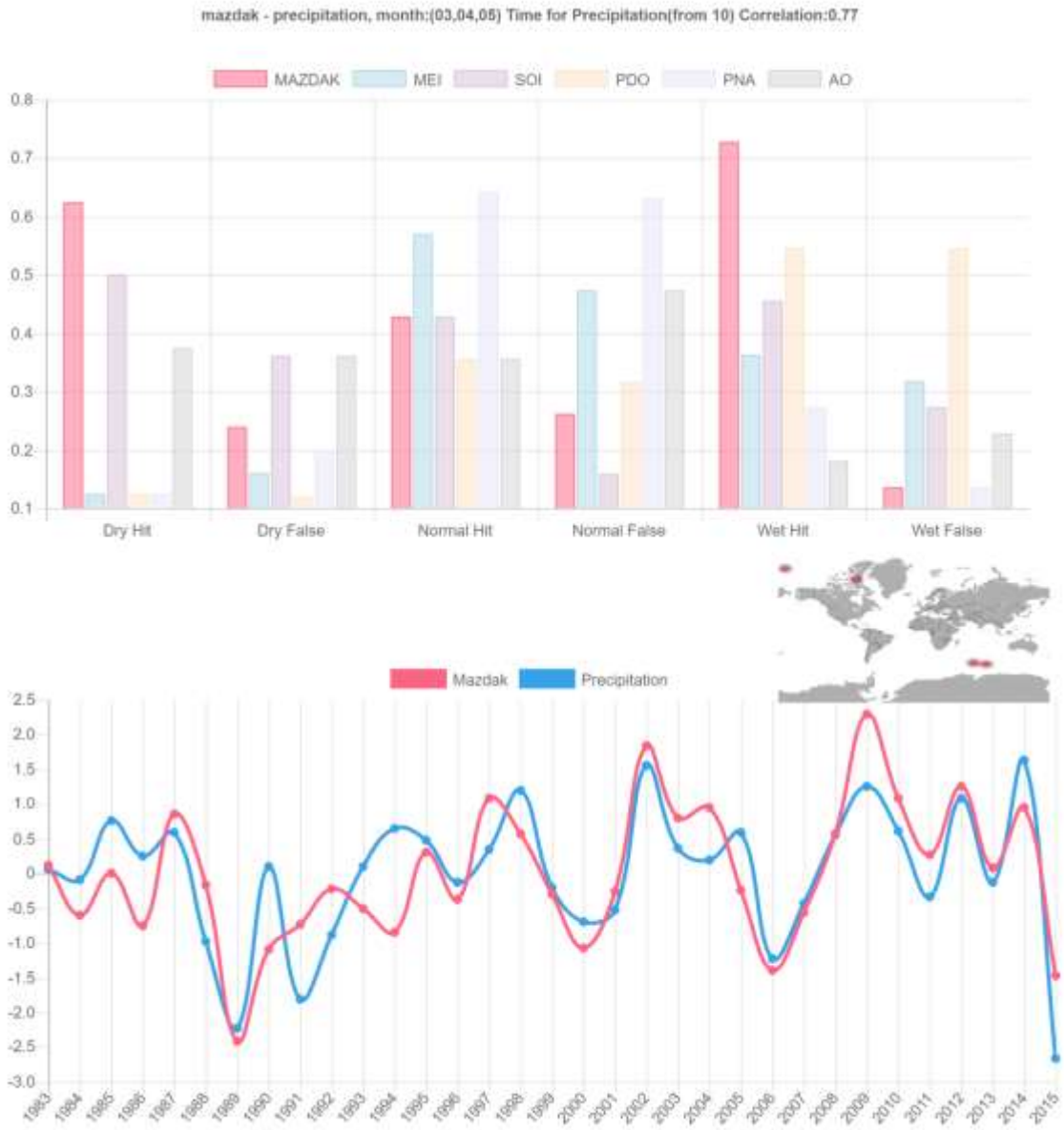



Figure S12| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

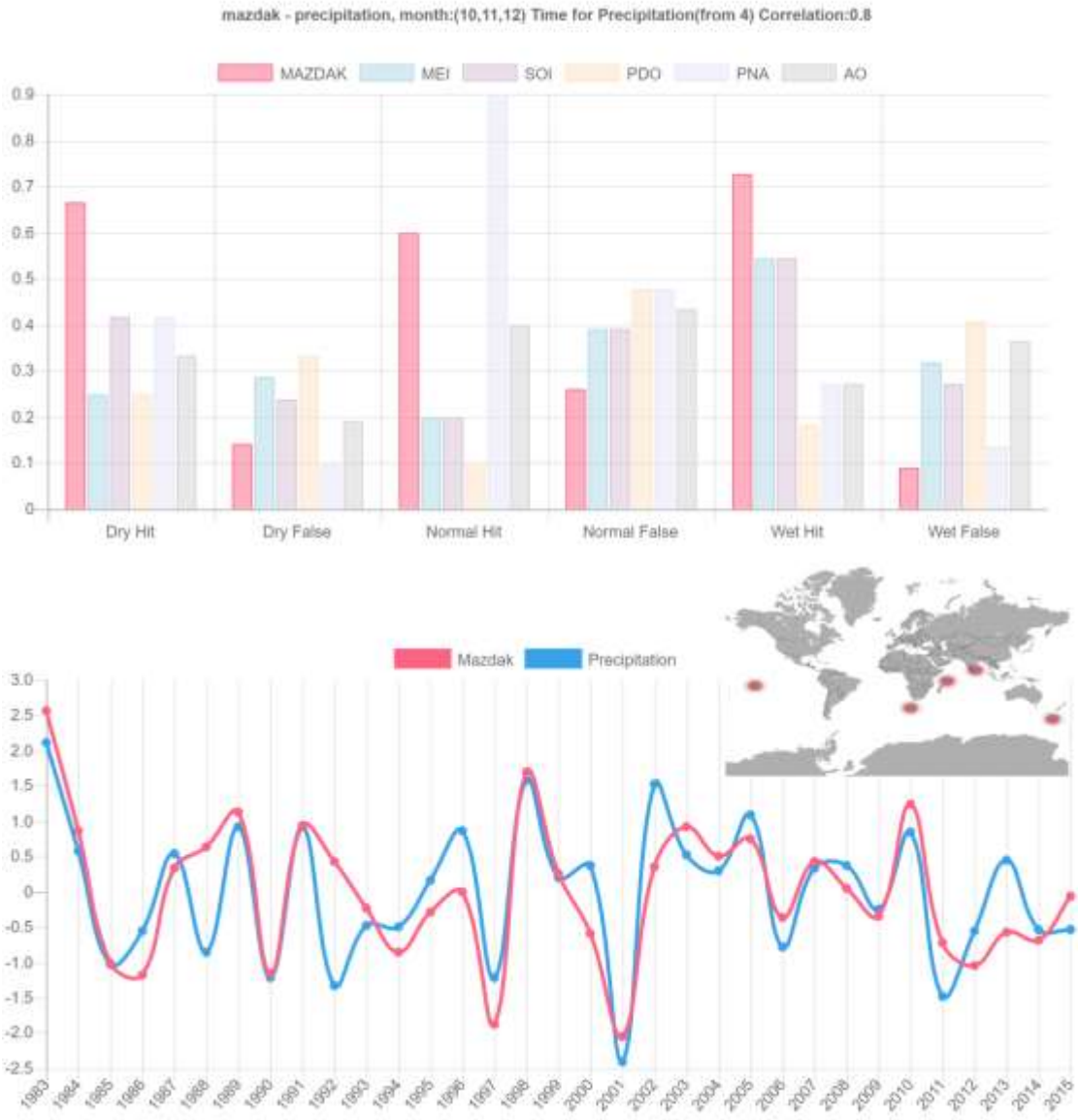



Figure S13| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

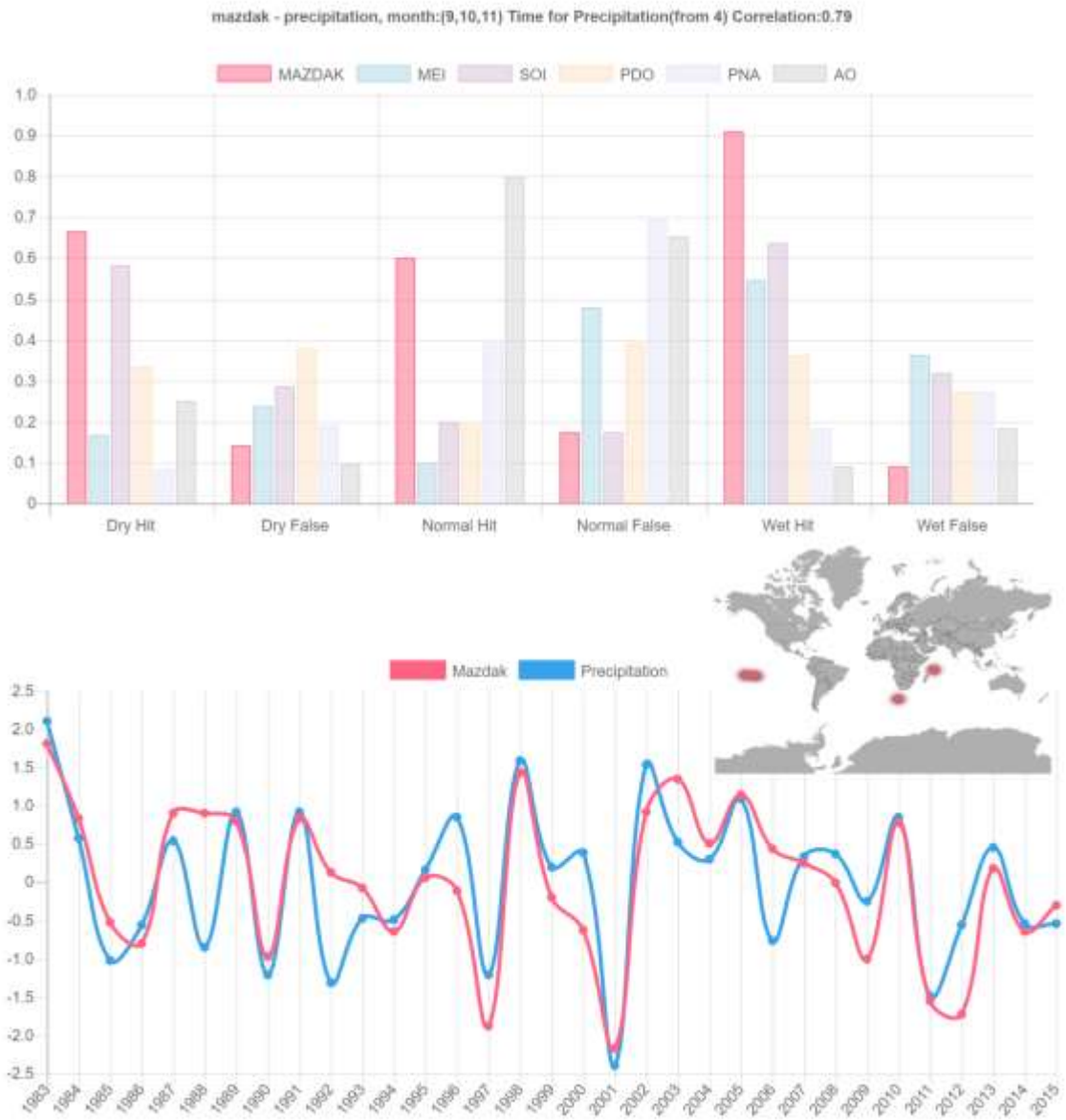



Figure S14| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

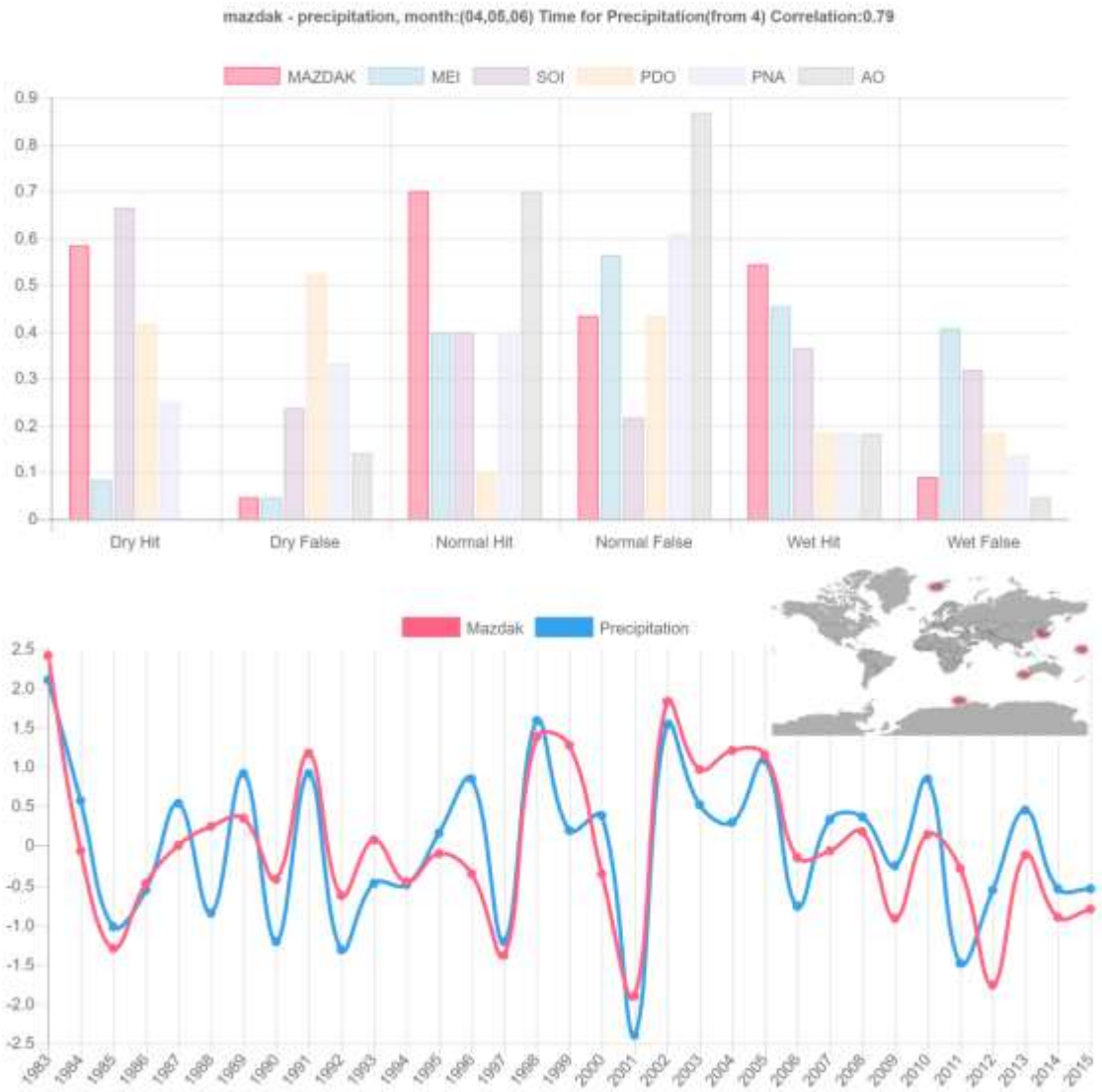



Figure S15| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

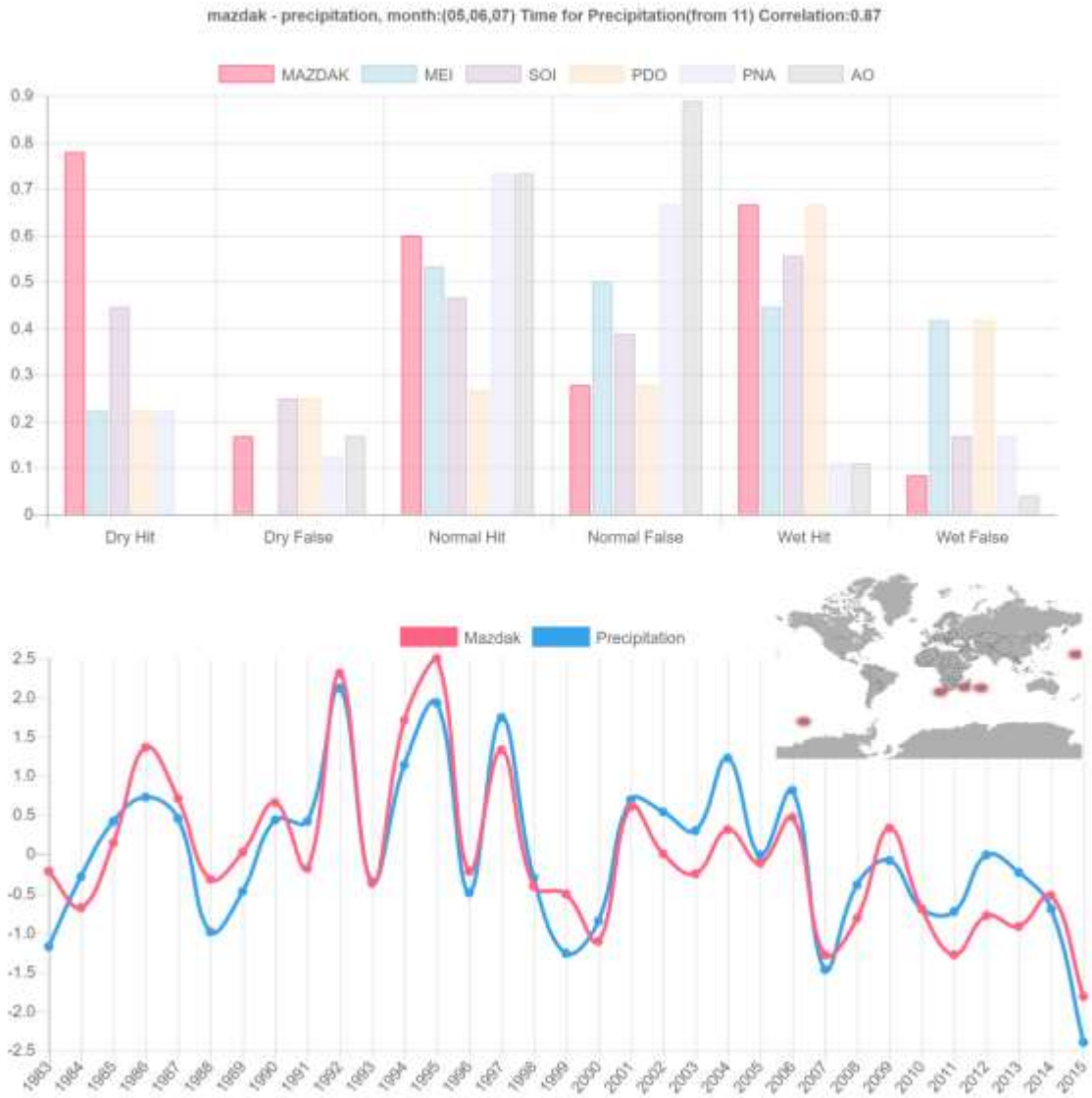



Figure S16| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

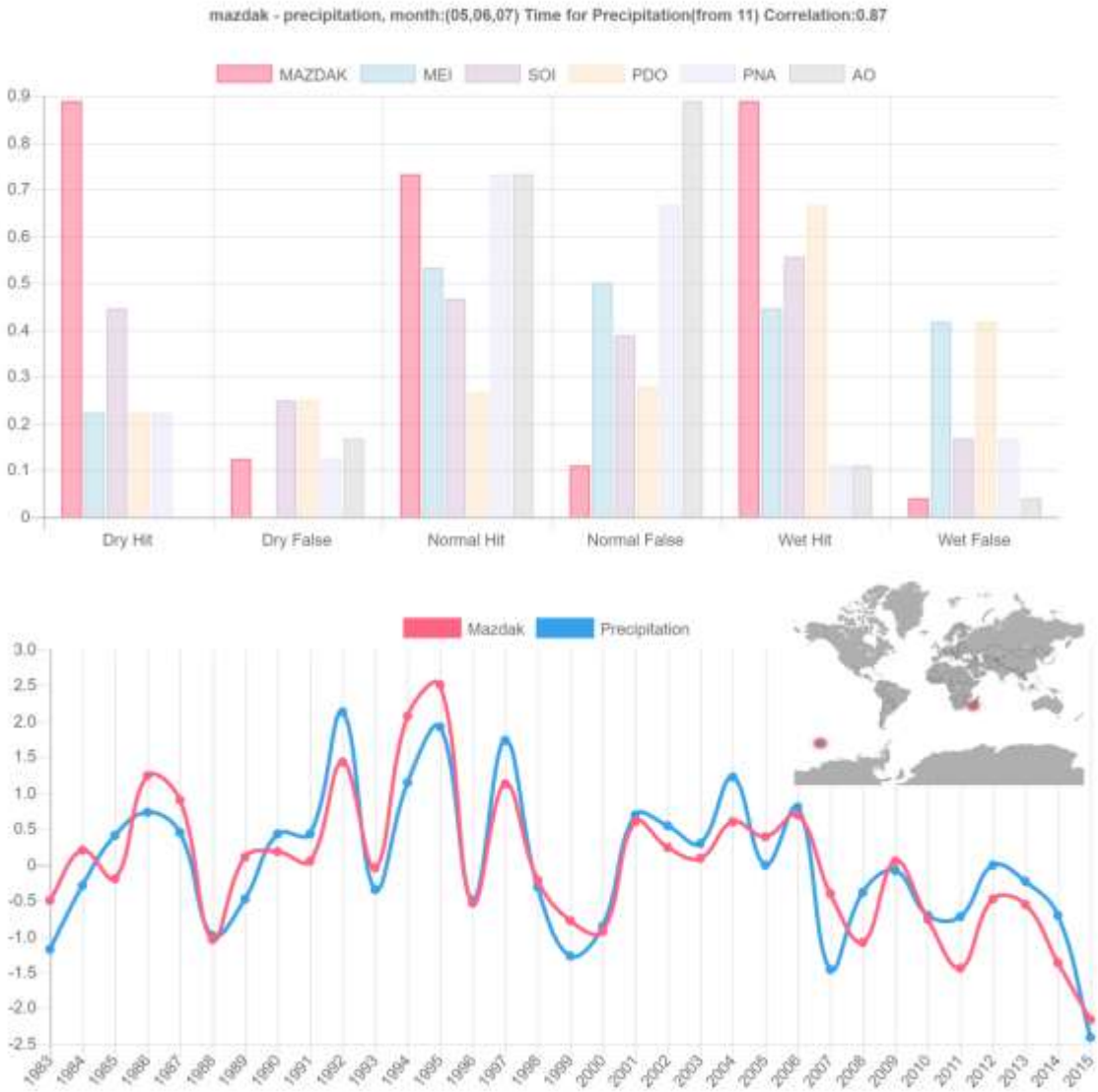



Figure S17| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

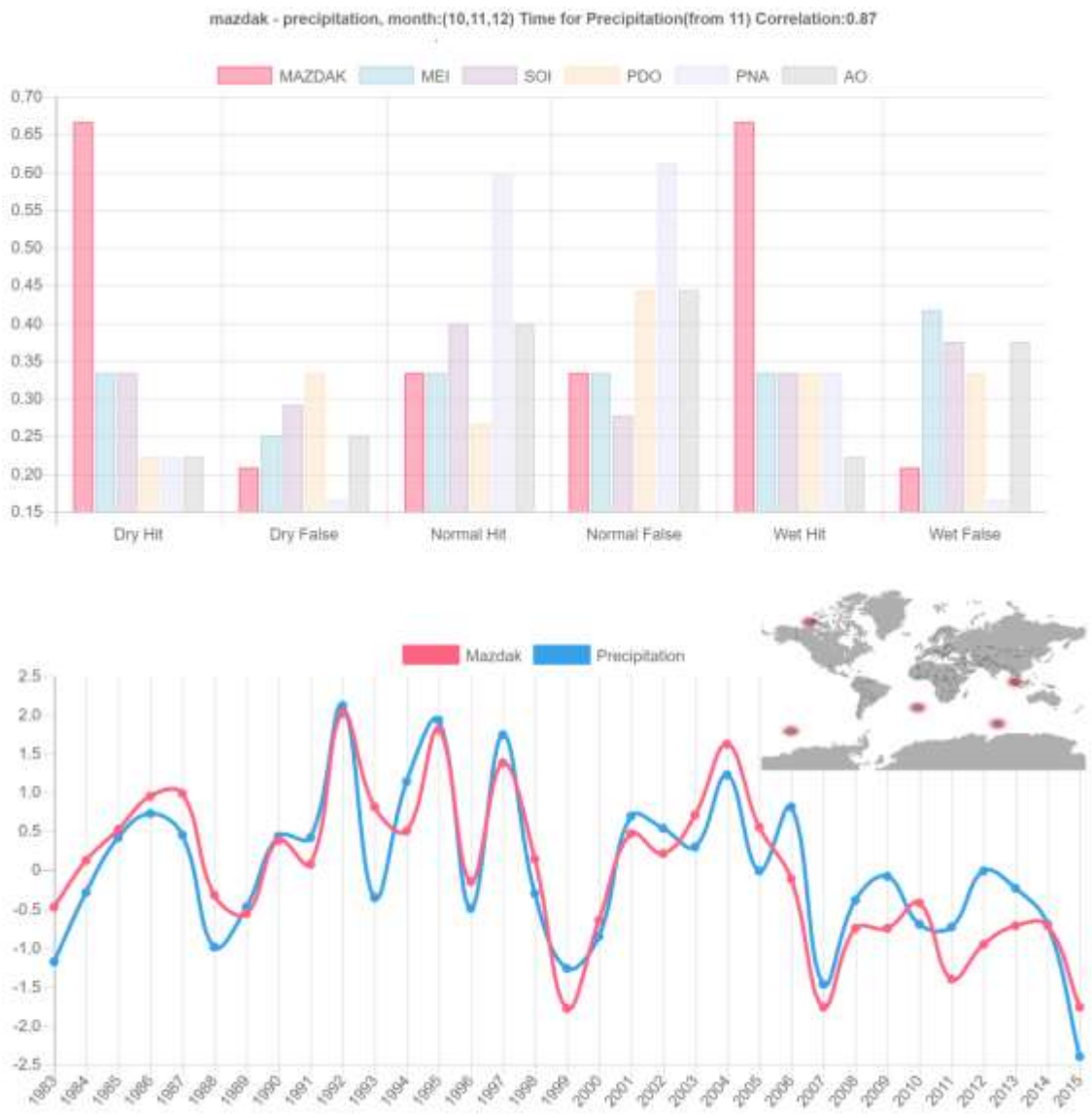



Figure S18| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

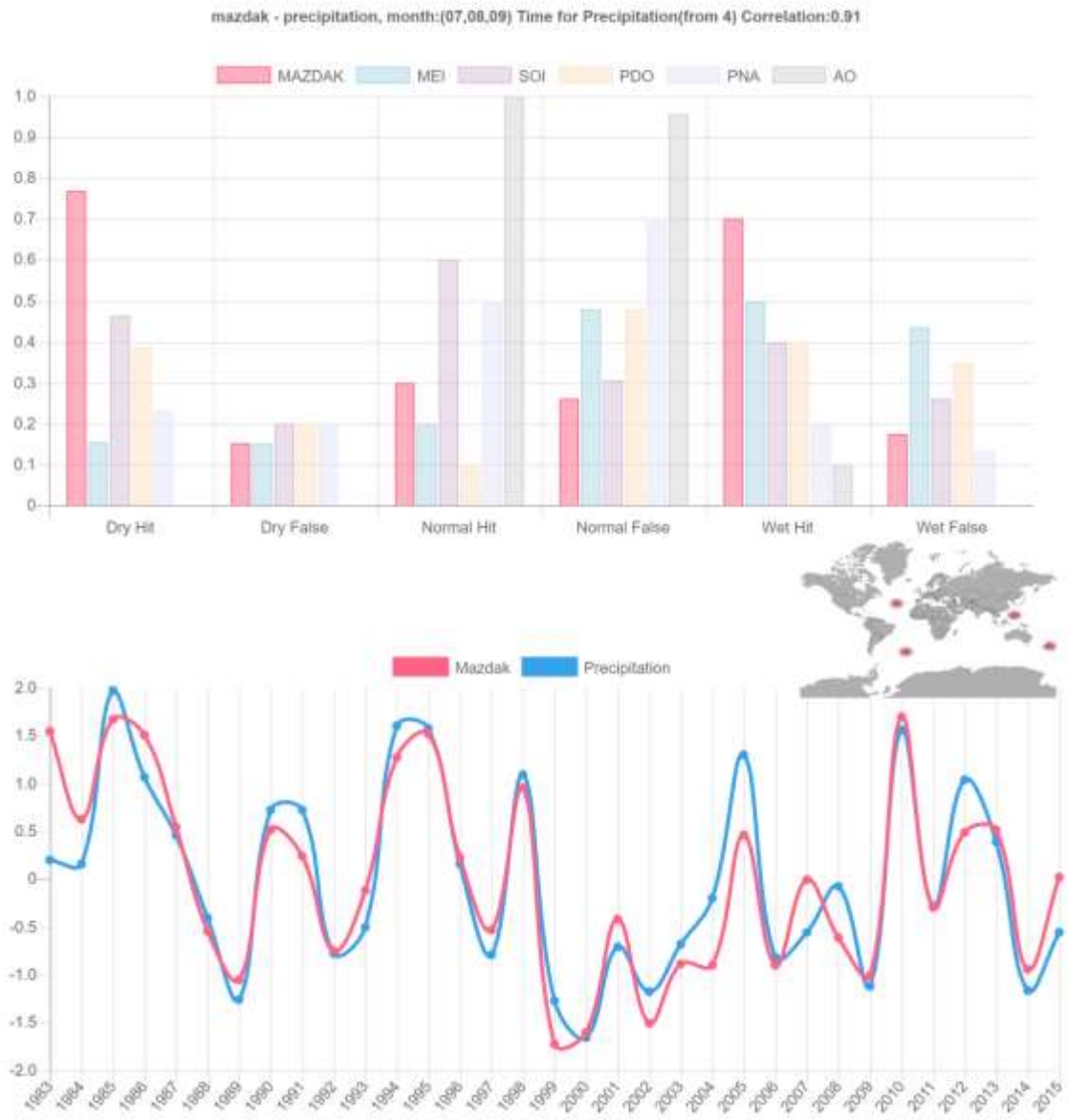



Figure S19| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

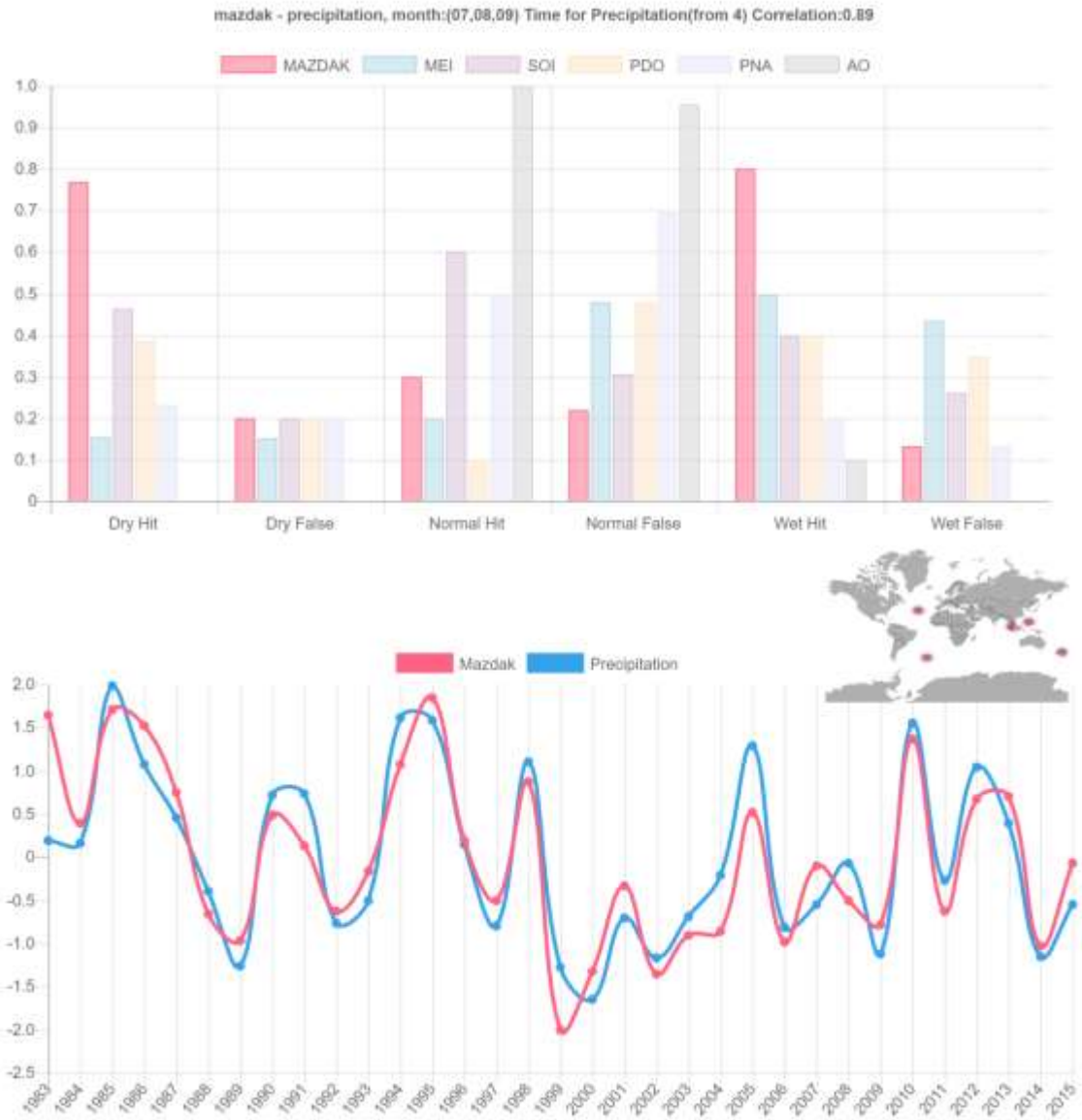



Figure S20| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S21| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

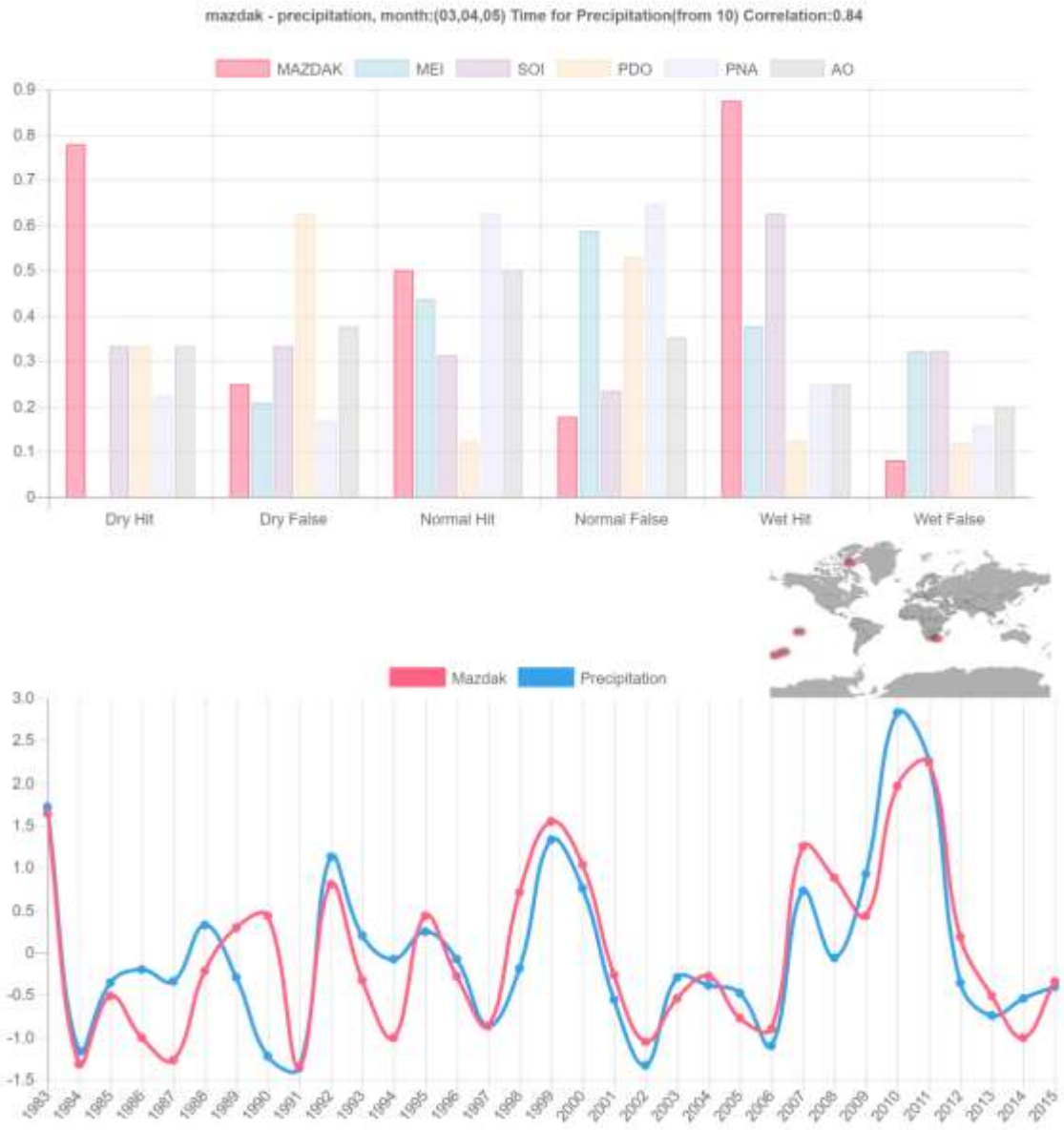



Figure S22| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S23| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

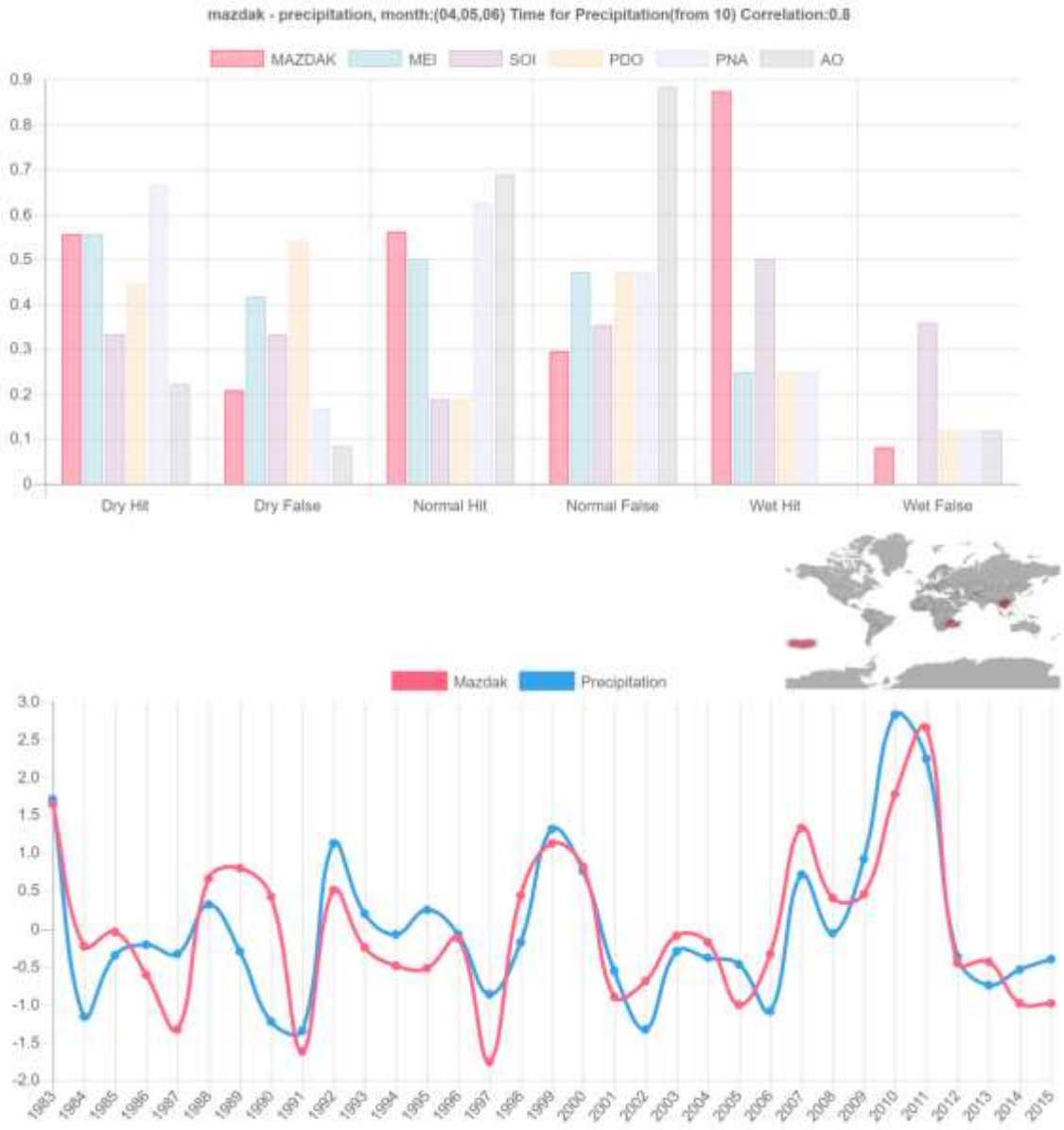



Figure S24| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

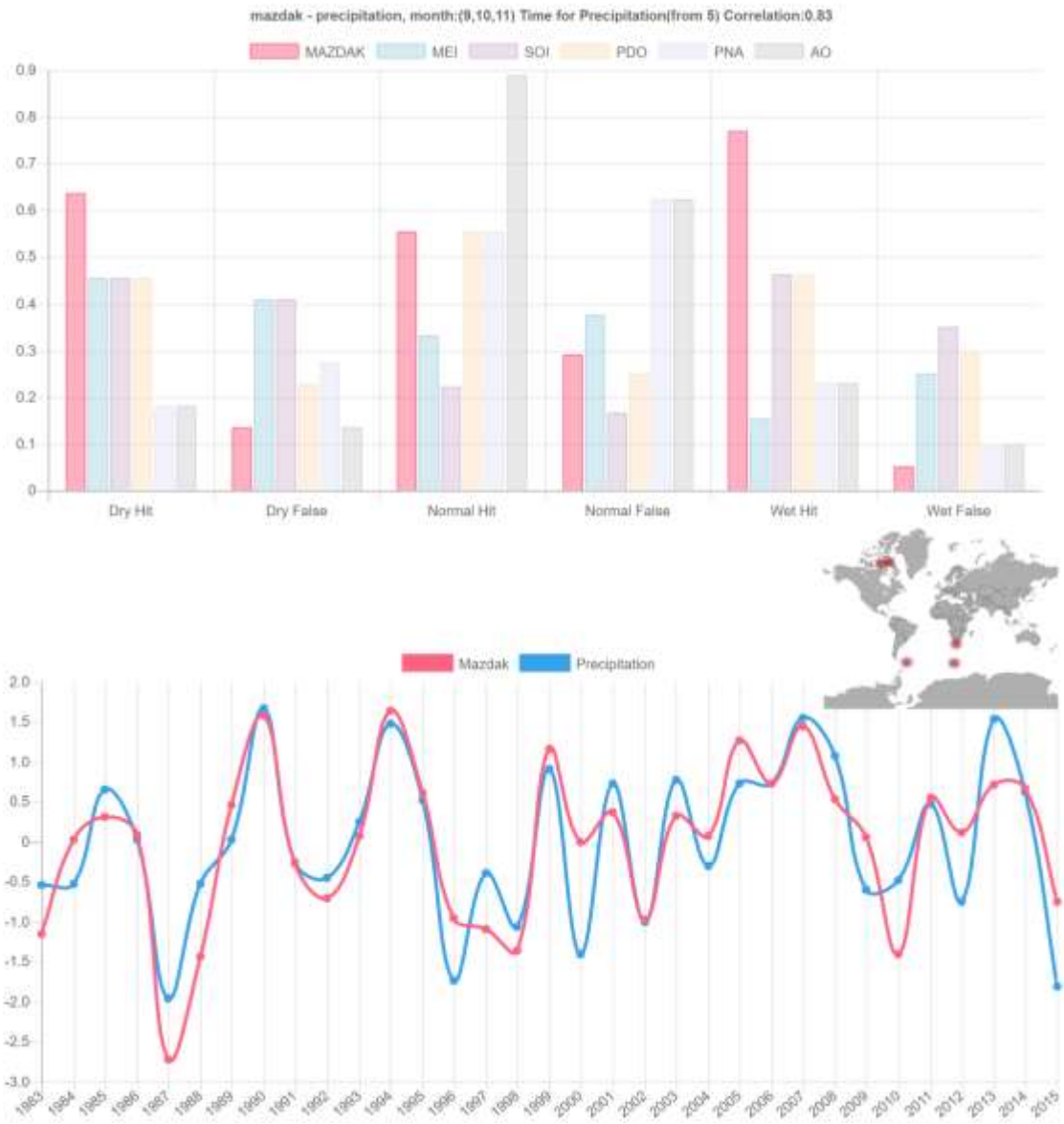



Figure S25| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

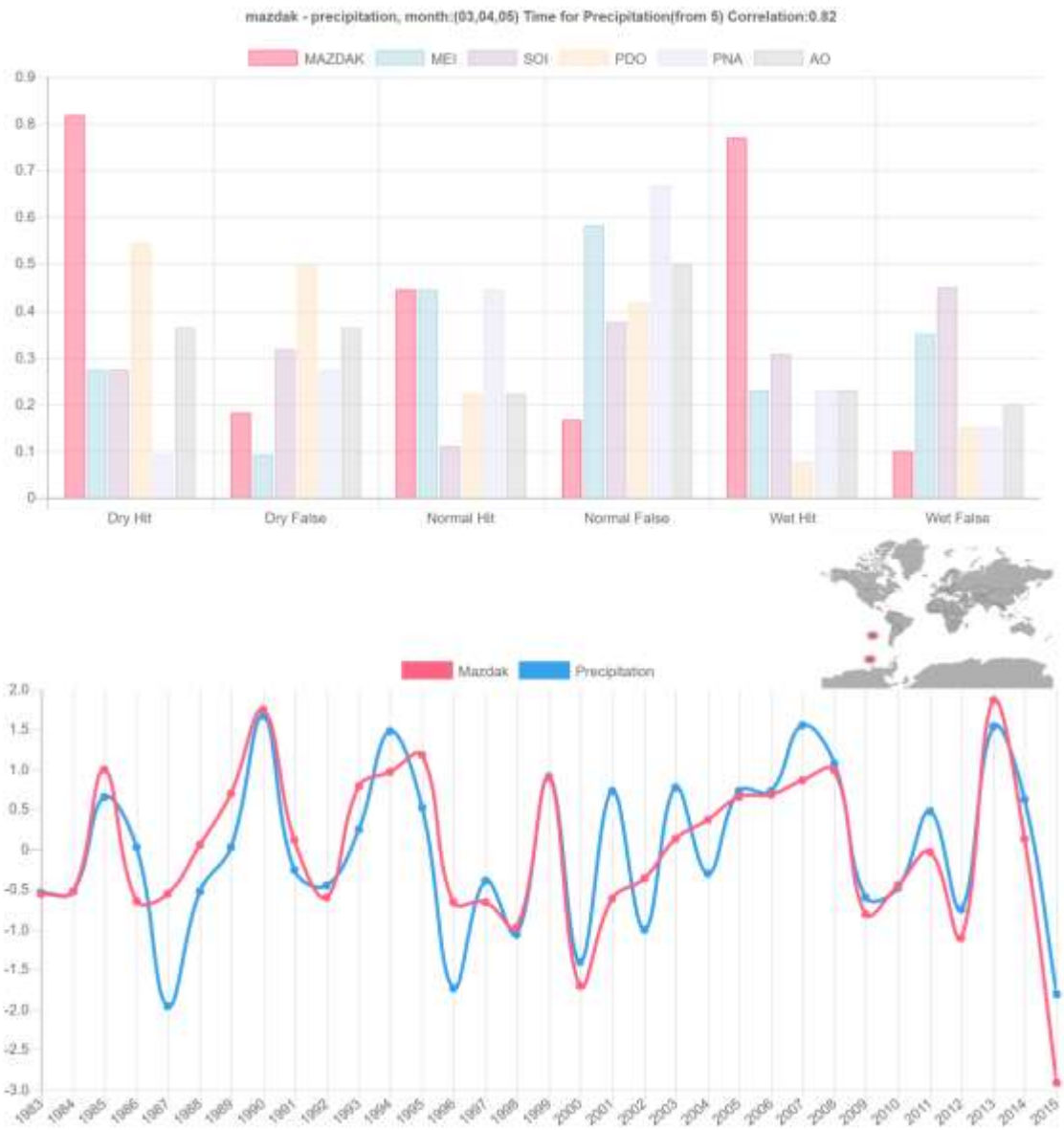



Figure S26| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

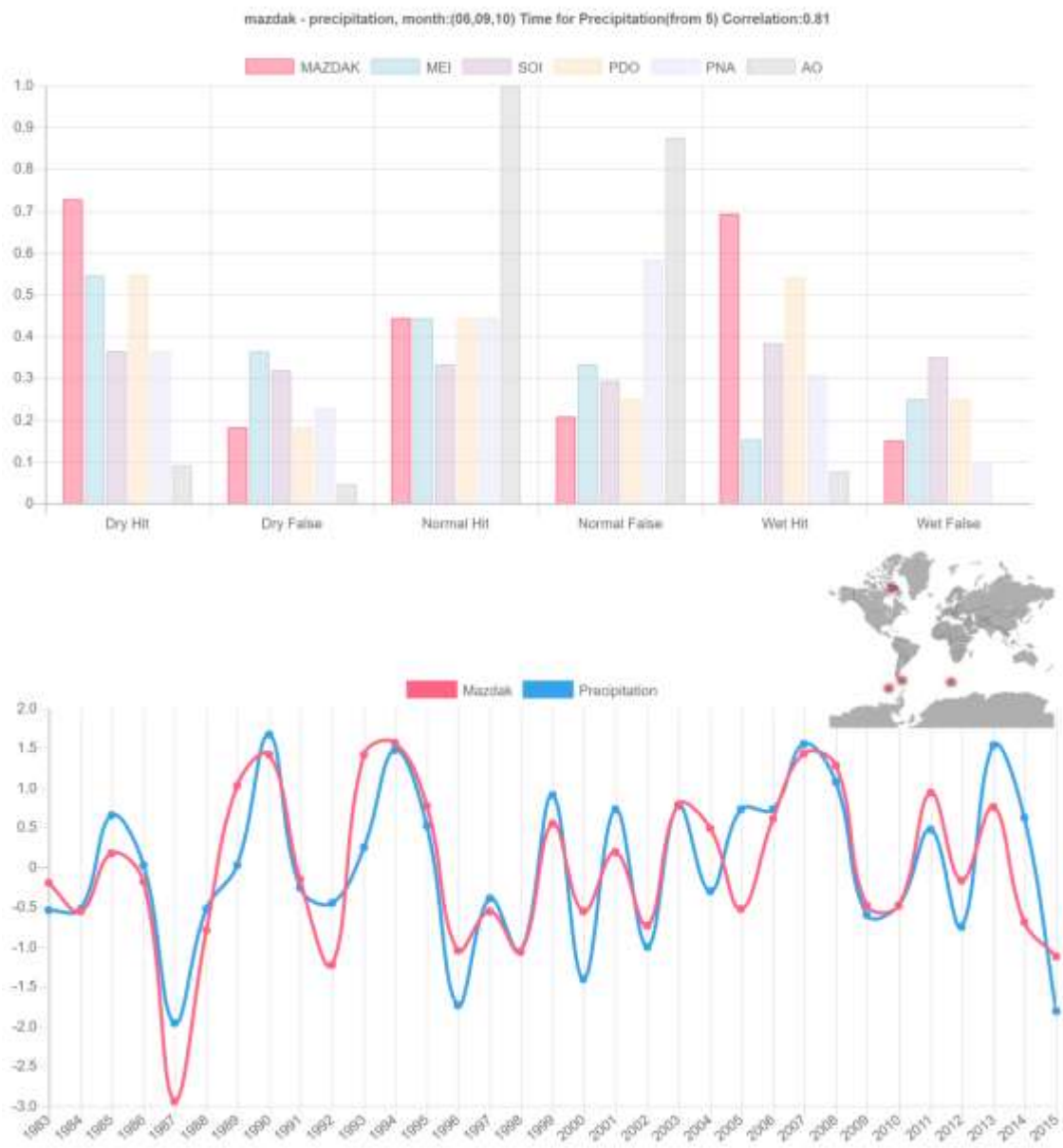



Figure S27| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

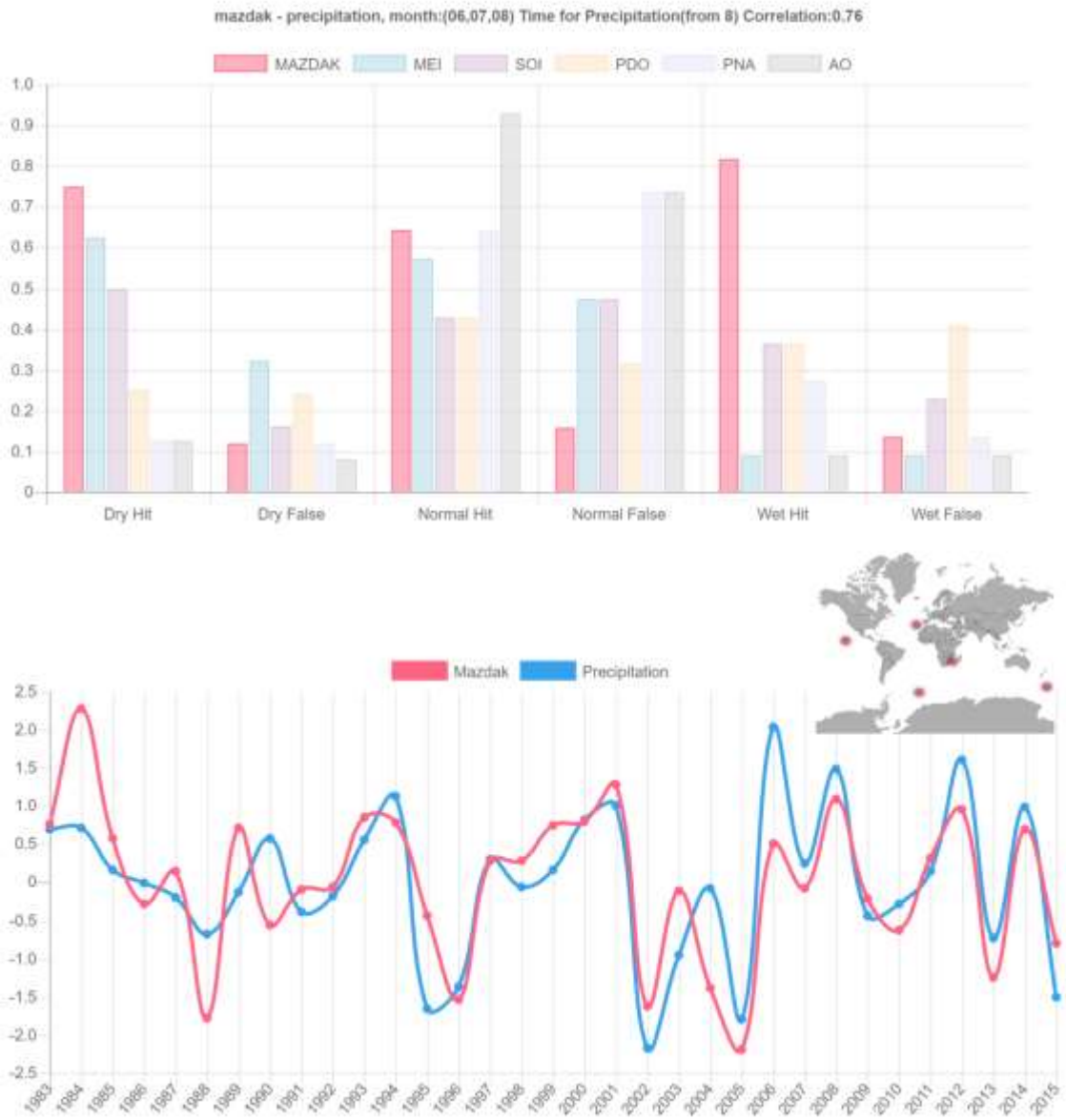



Figure S28| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

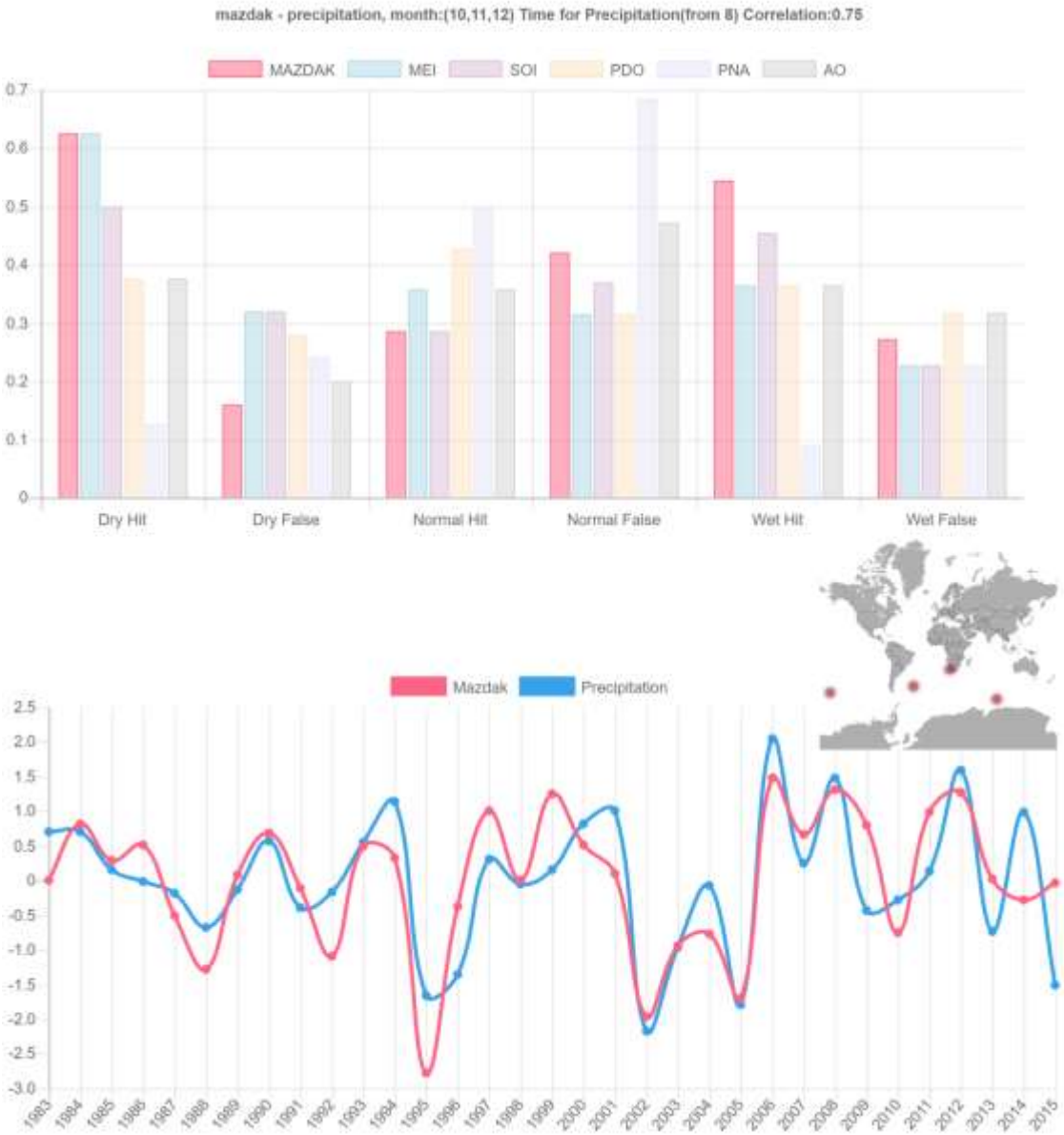



Figure S29| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

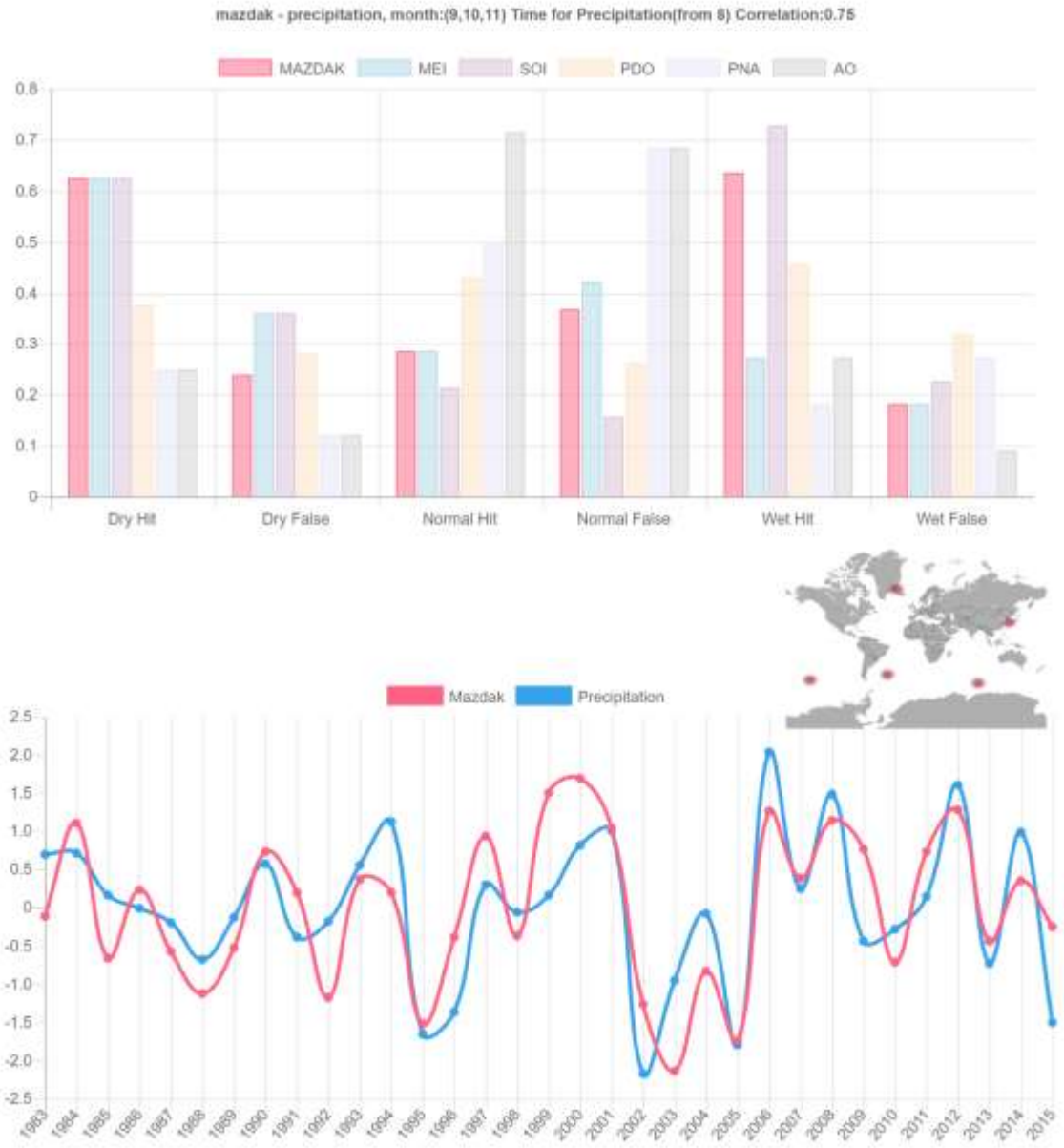



Figure S30| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

**Section 2:**
In this section the results are presented based on analysis of record over 1983-2010 to find the locations that best estimate the precipitation.

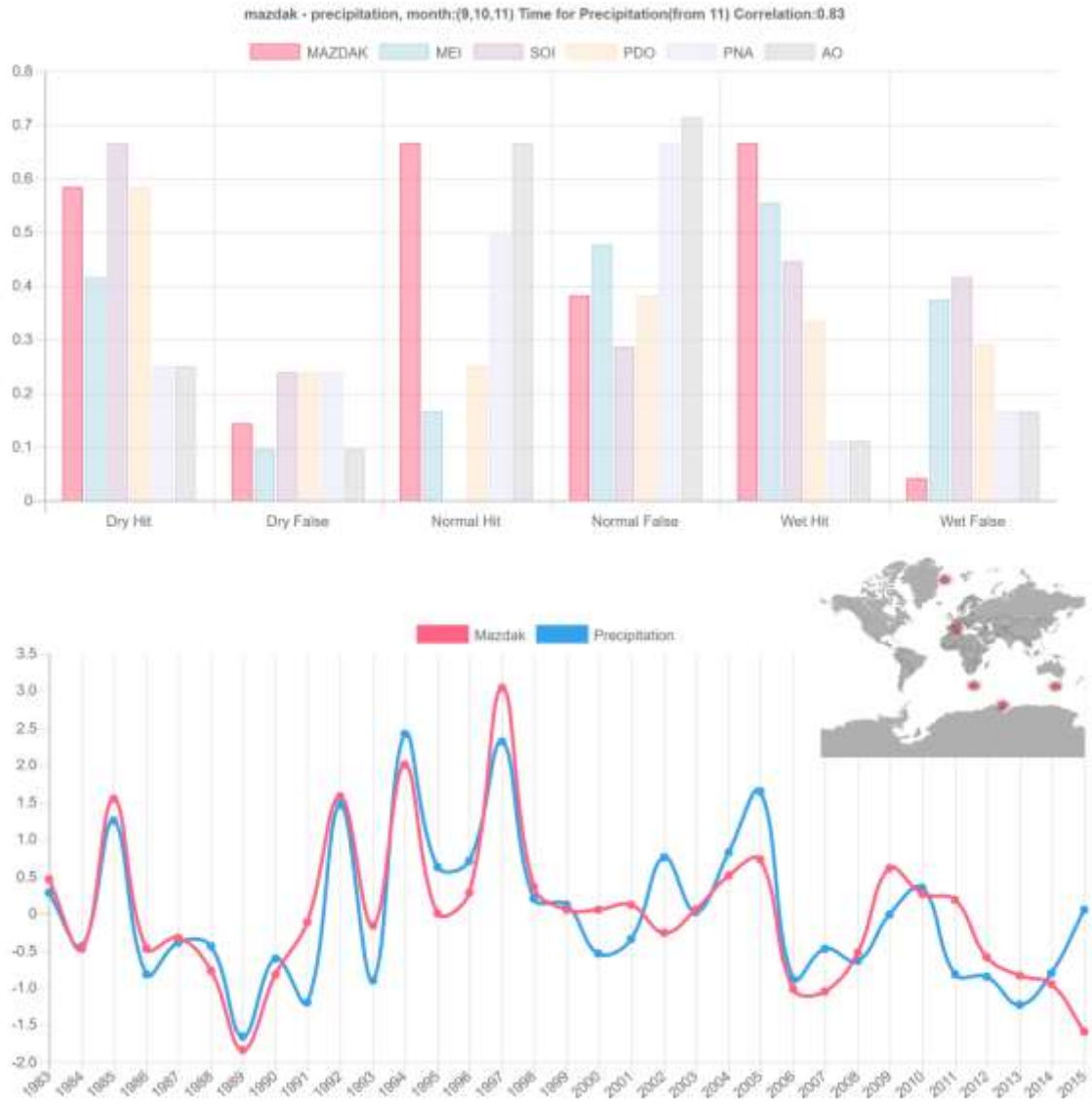



Figure S31| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

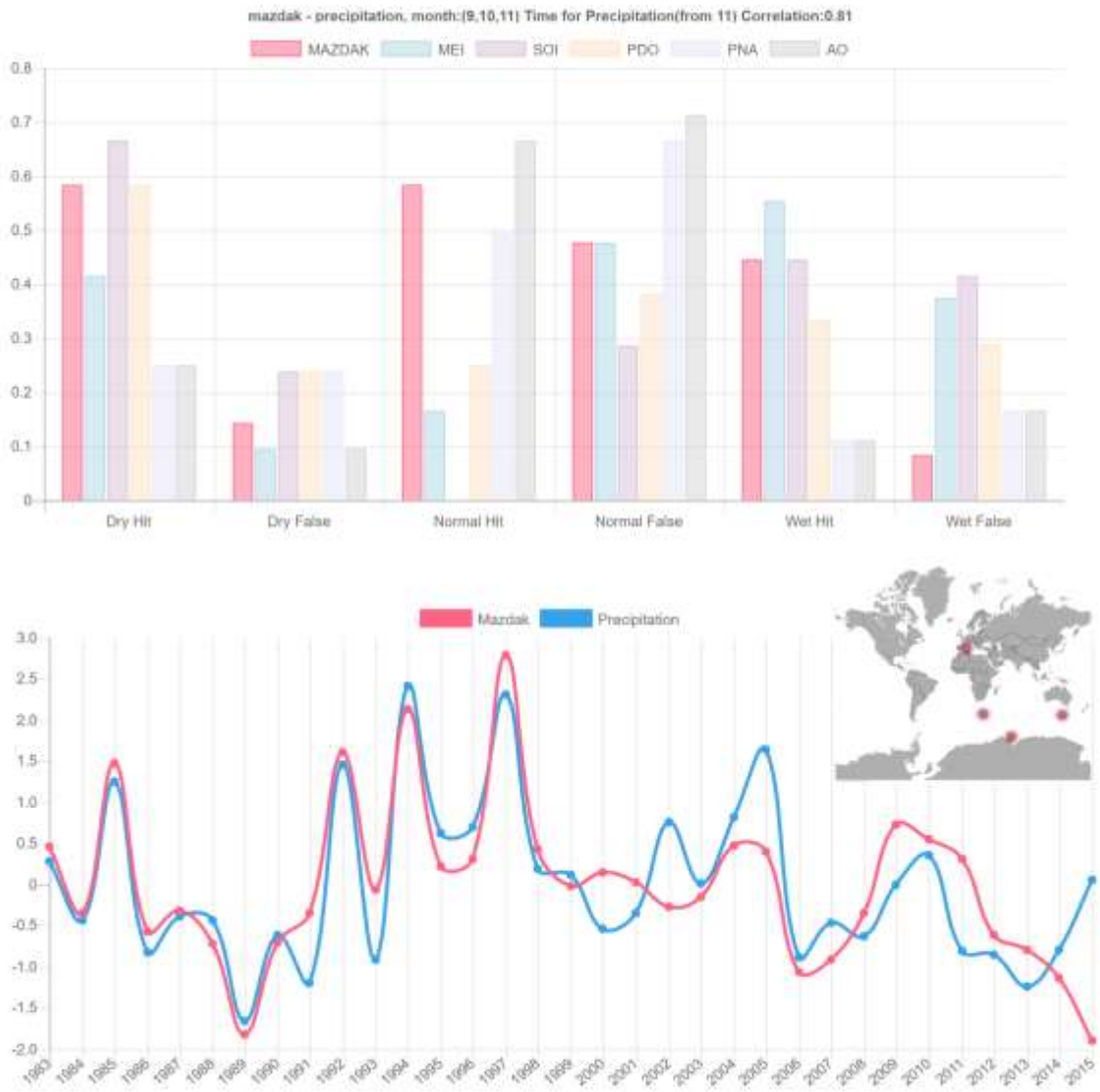



Figure S32| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

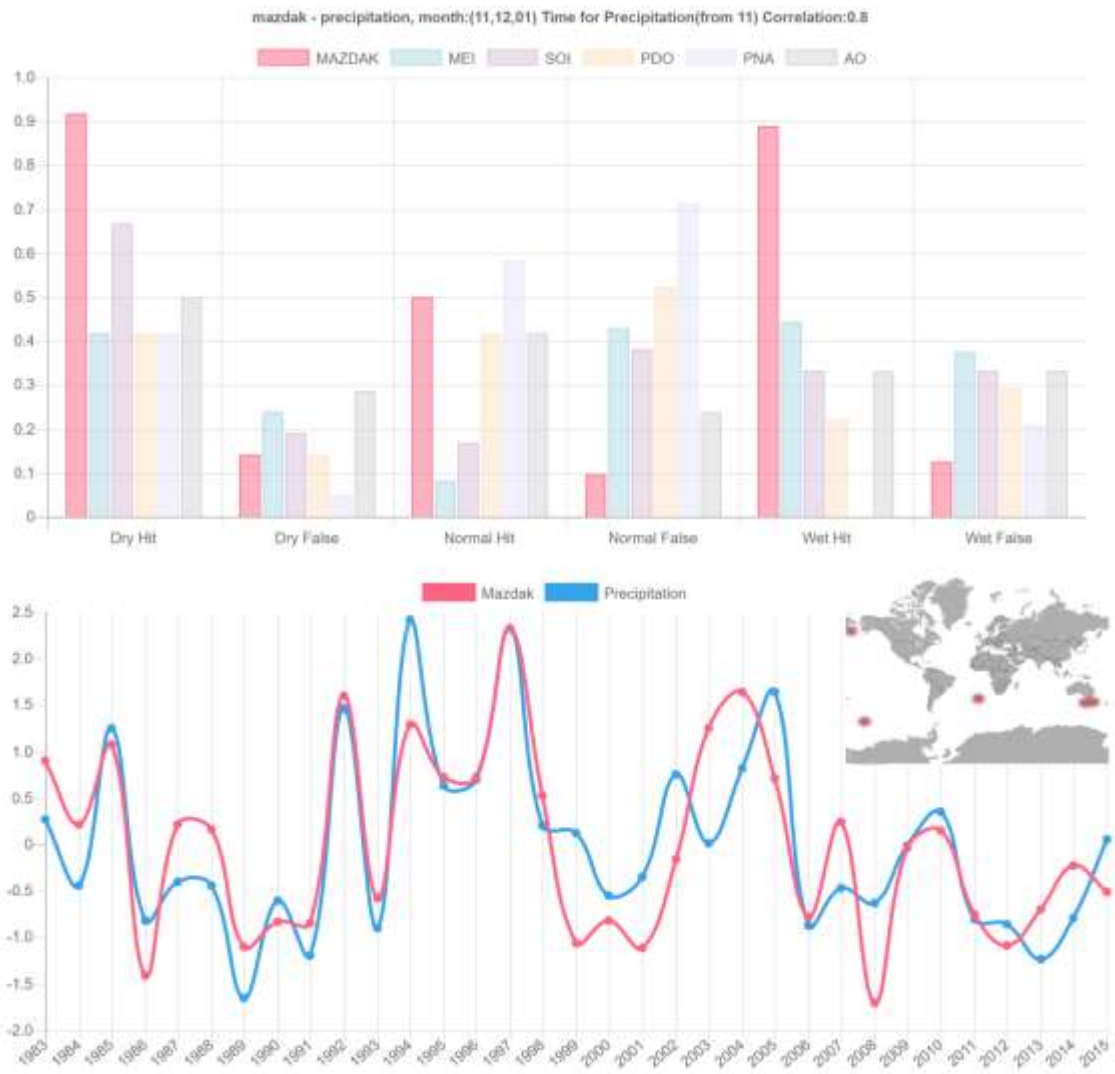



Figure S33| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

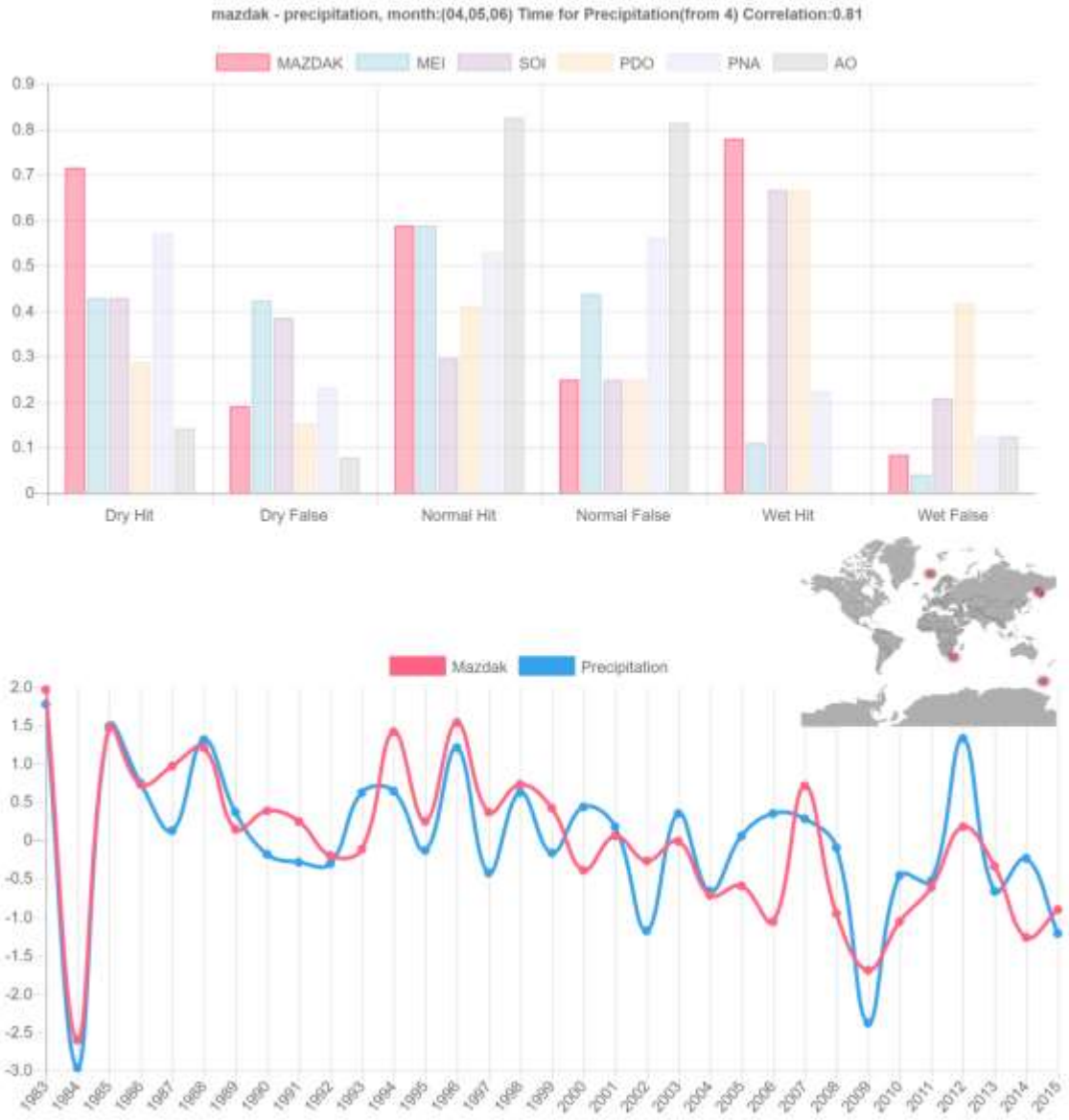



Figure S34| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

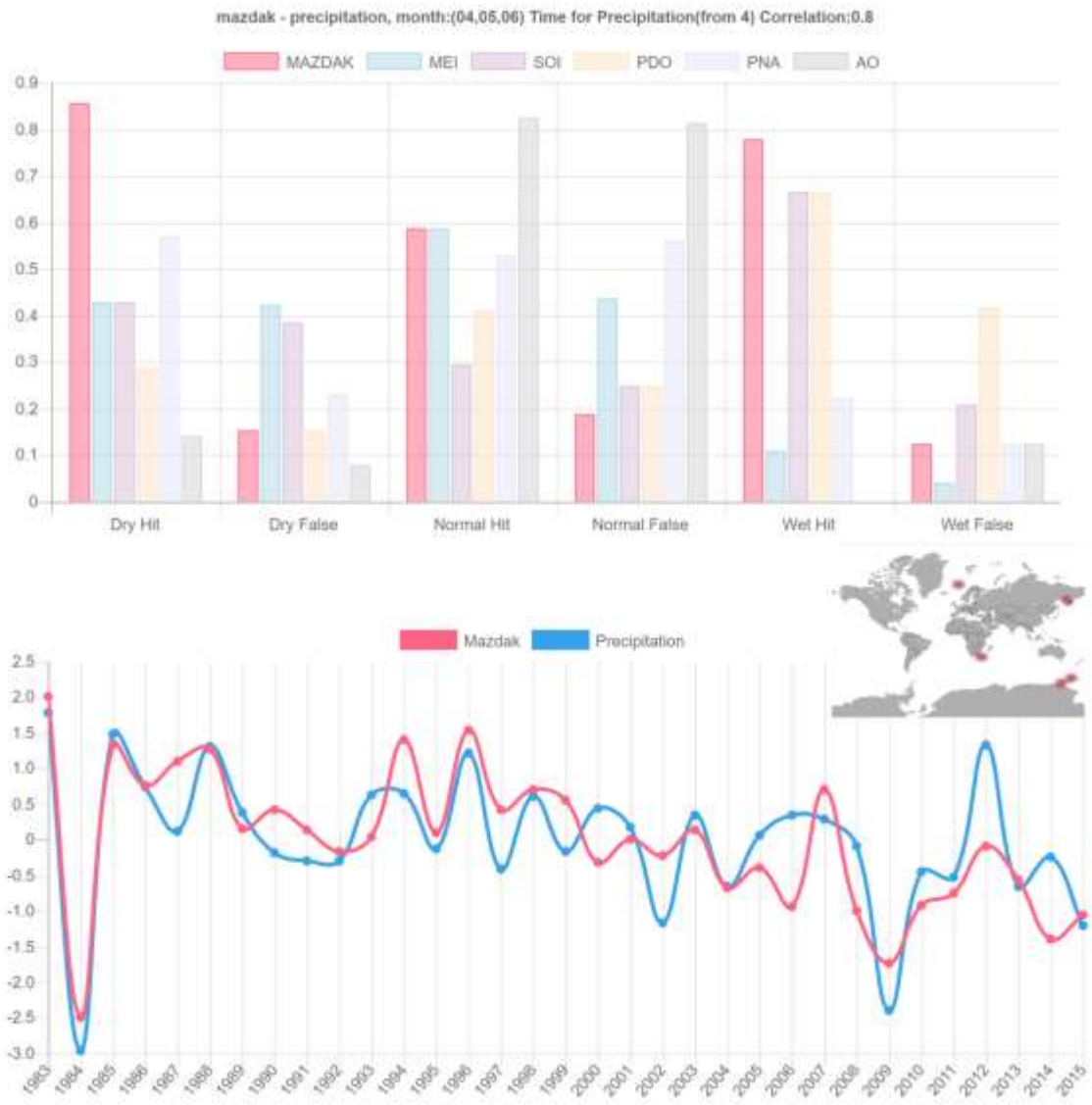



Figure S35| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the Second highest correlation coefficient.

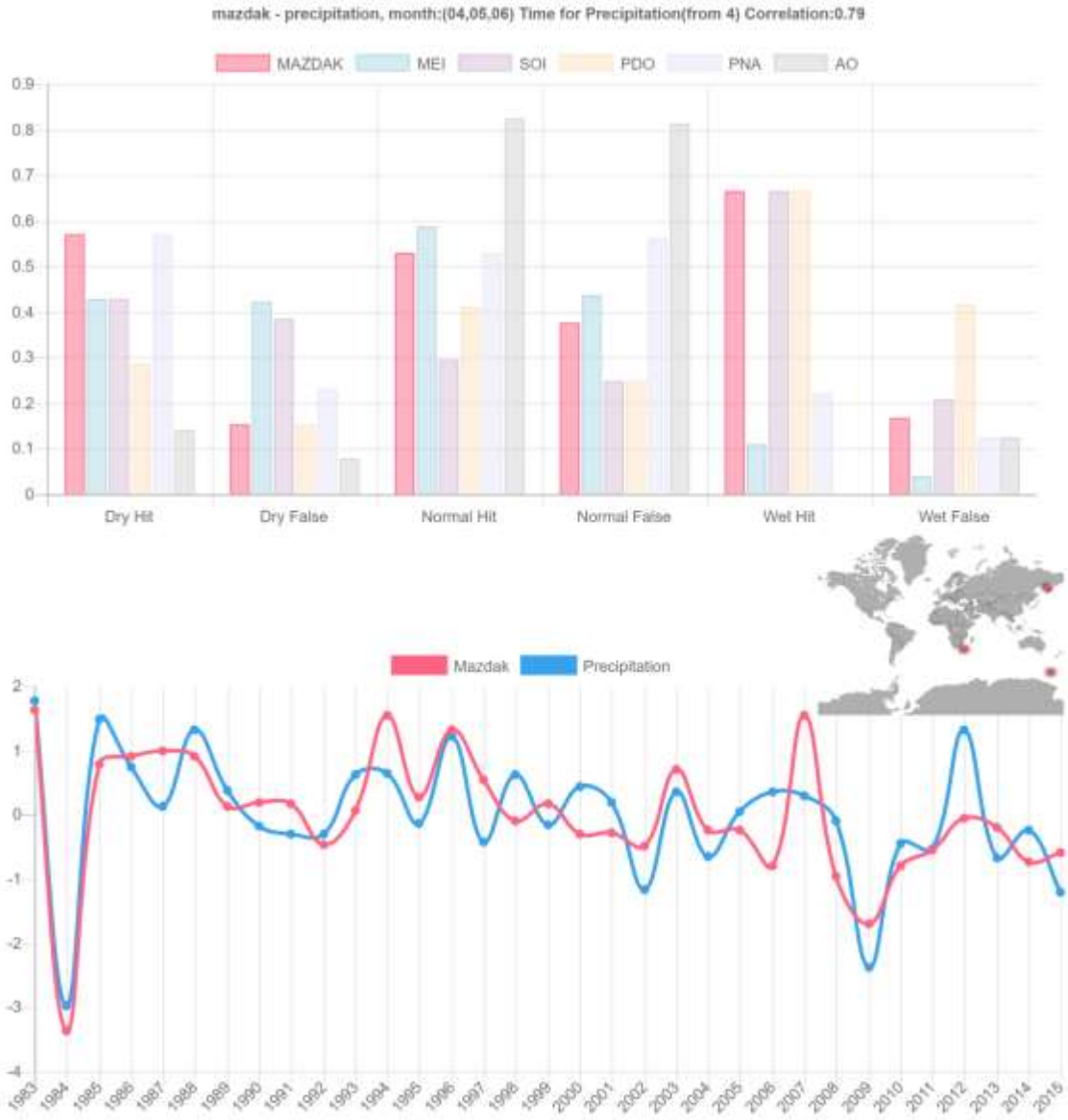



Figure S36| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S37| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S38| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

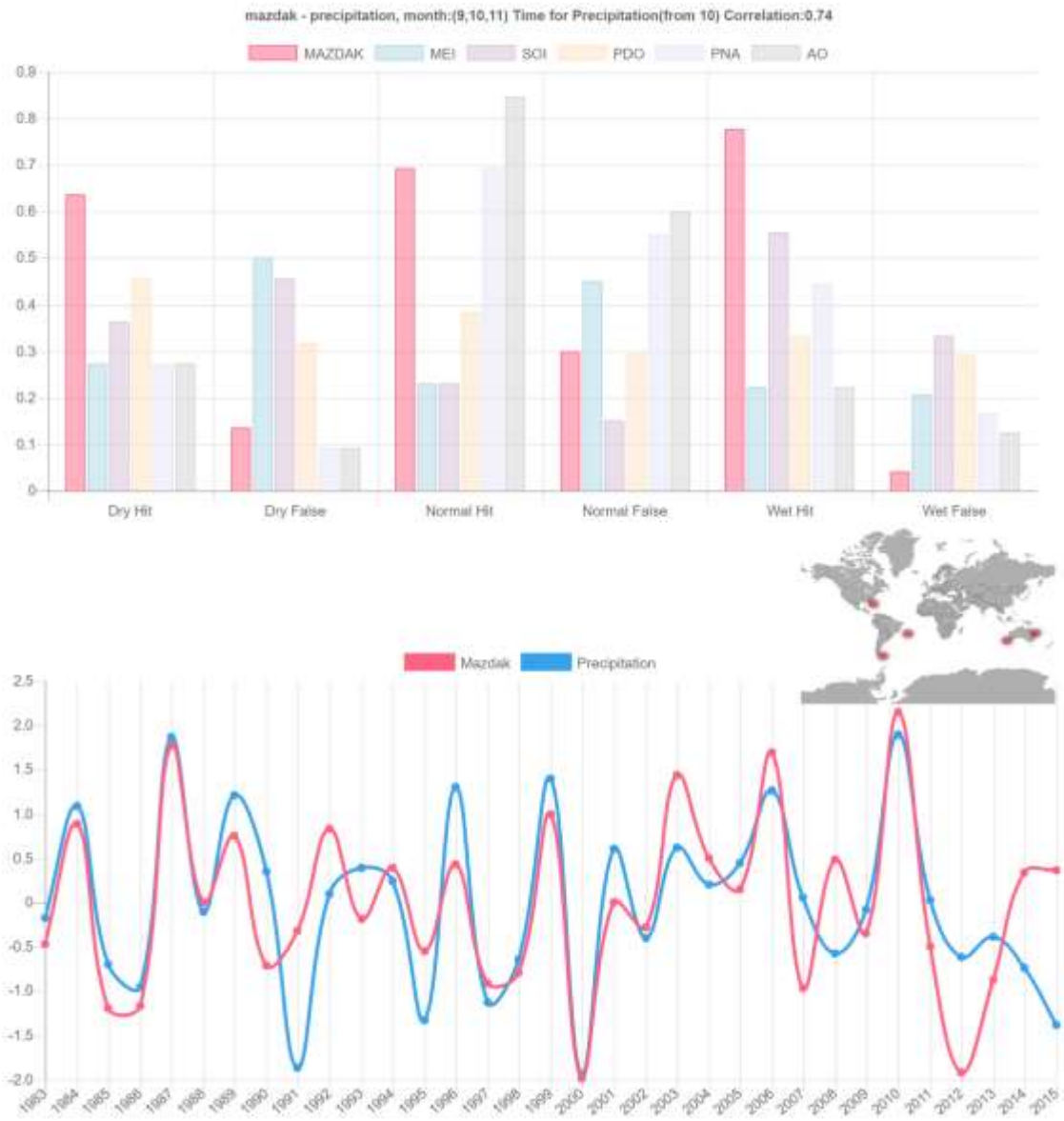



Figure S39| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S40| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

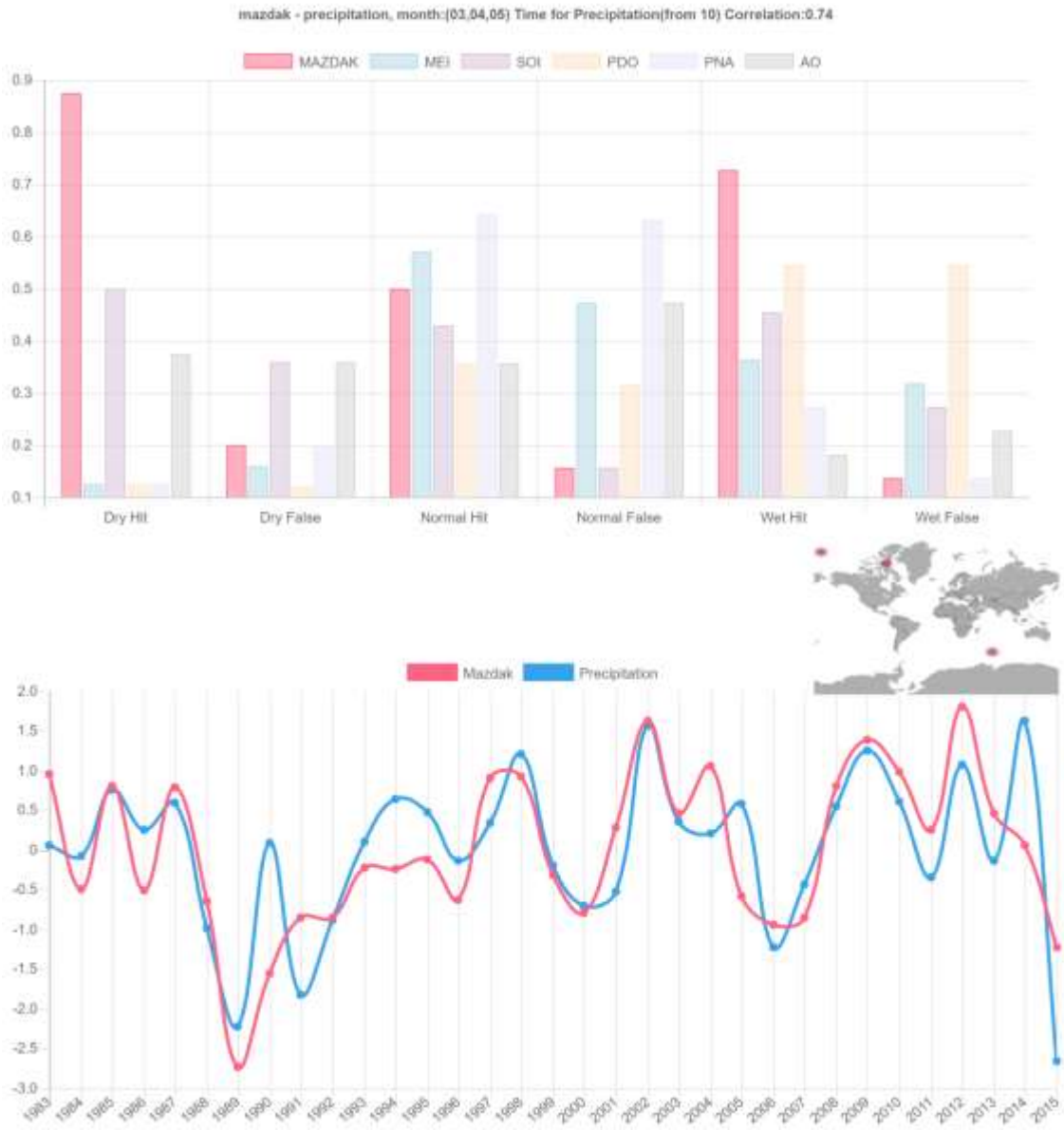



Figure S41| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S42| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

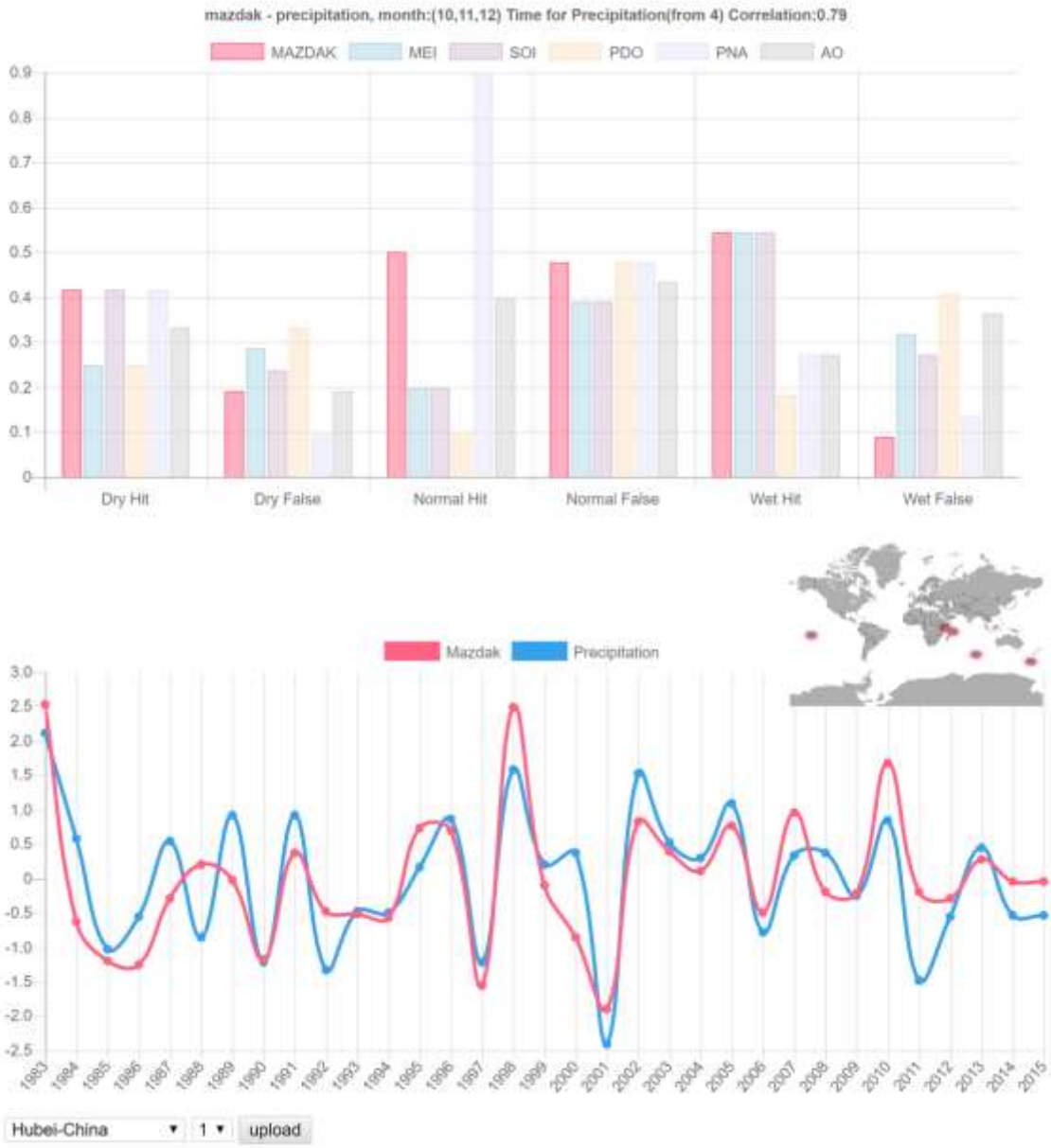



Figure S43| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

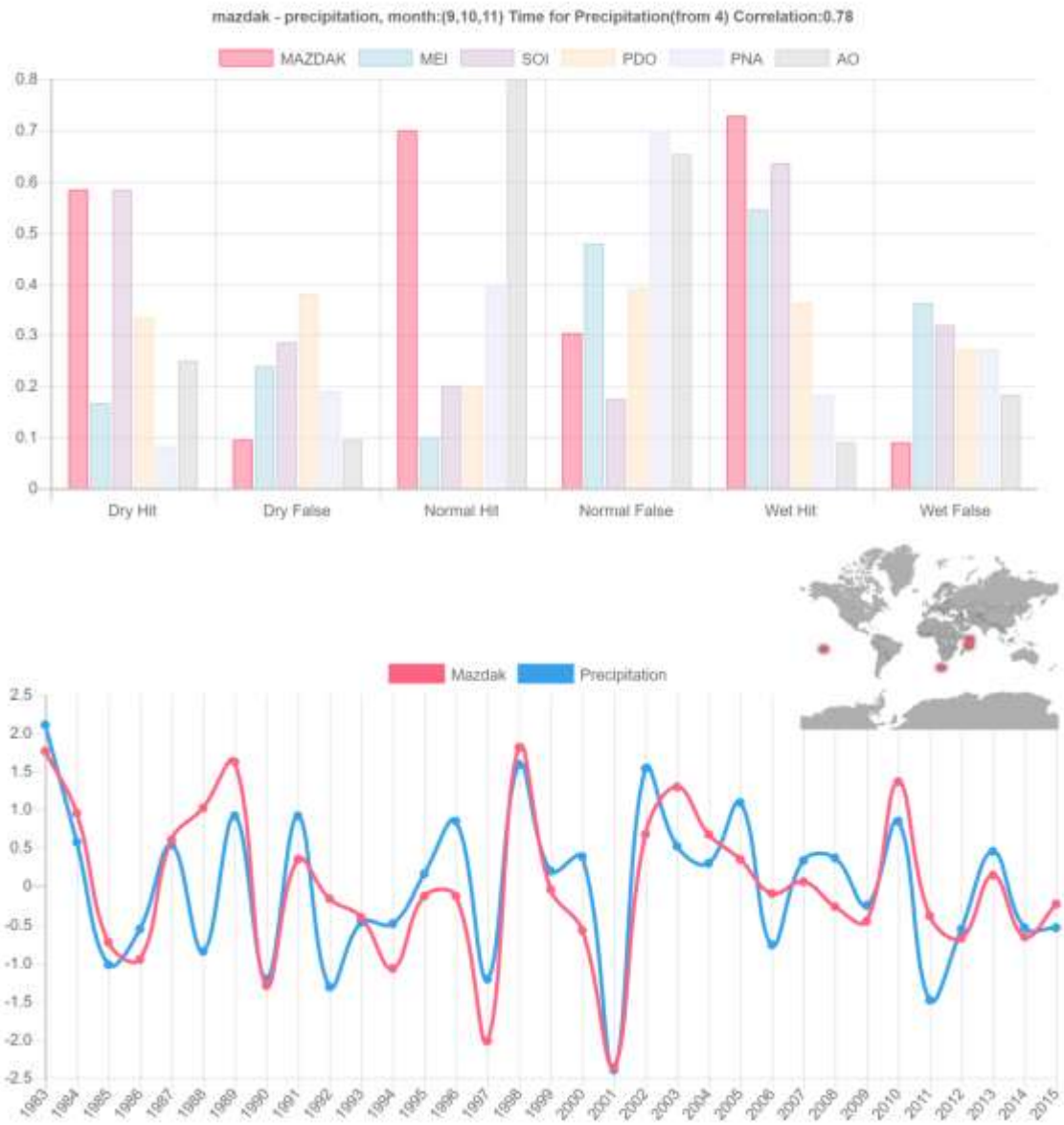



Figure S44| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

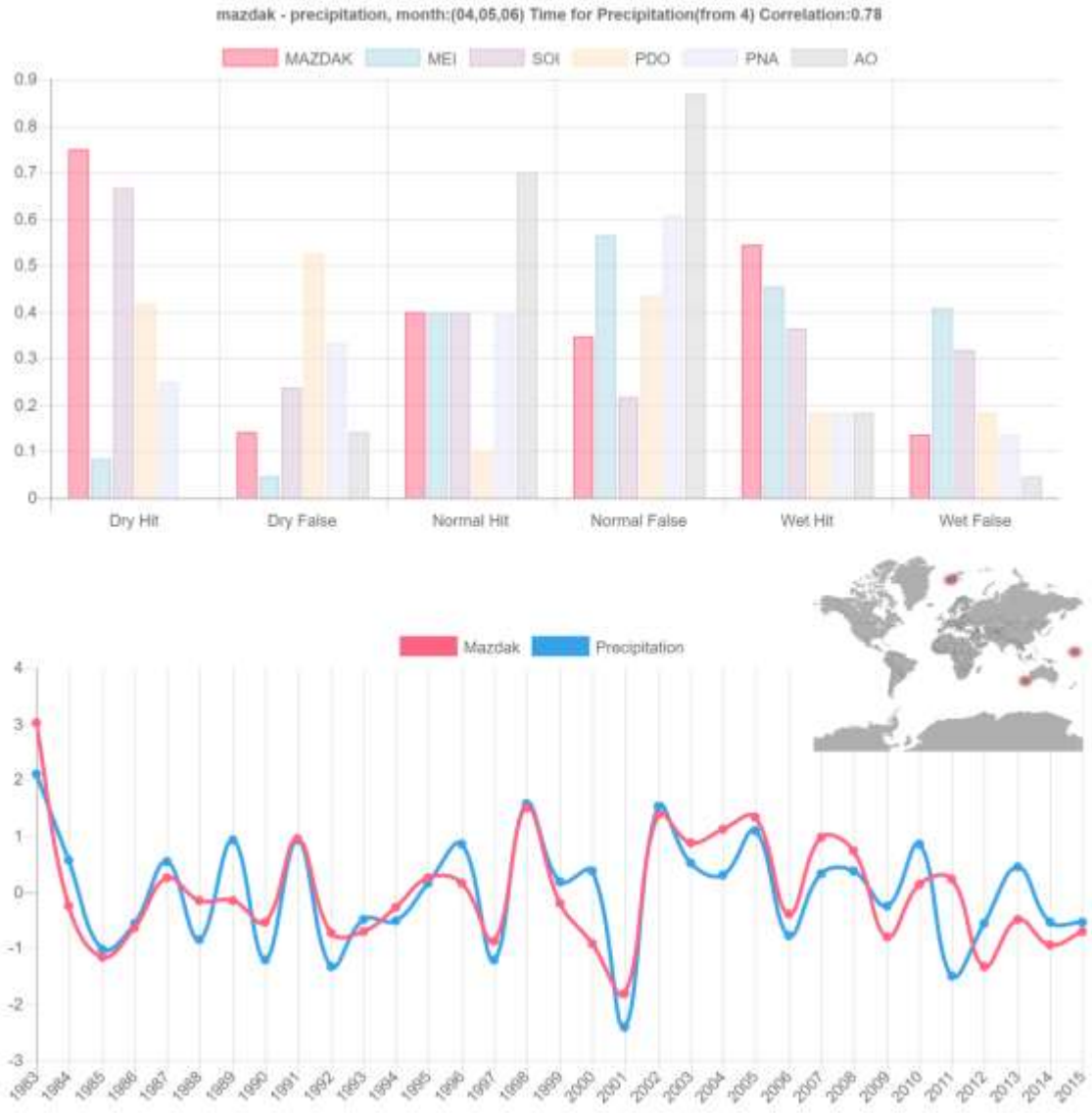



Figure S45| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

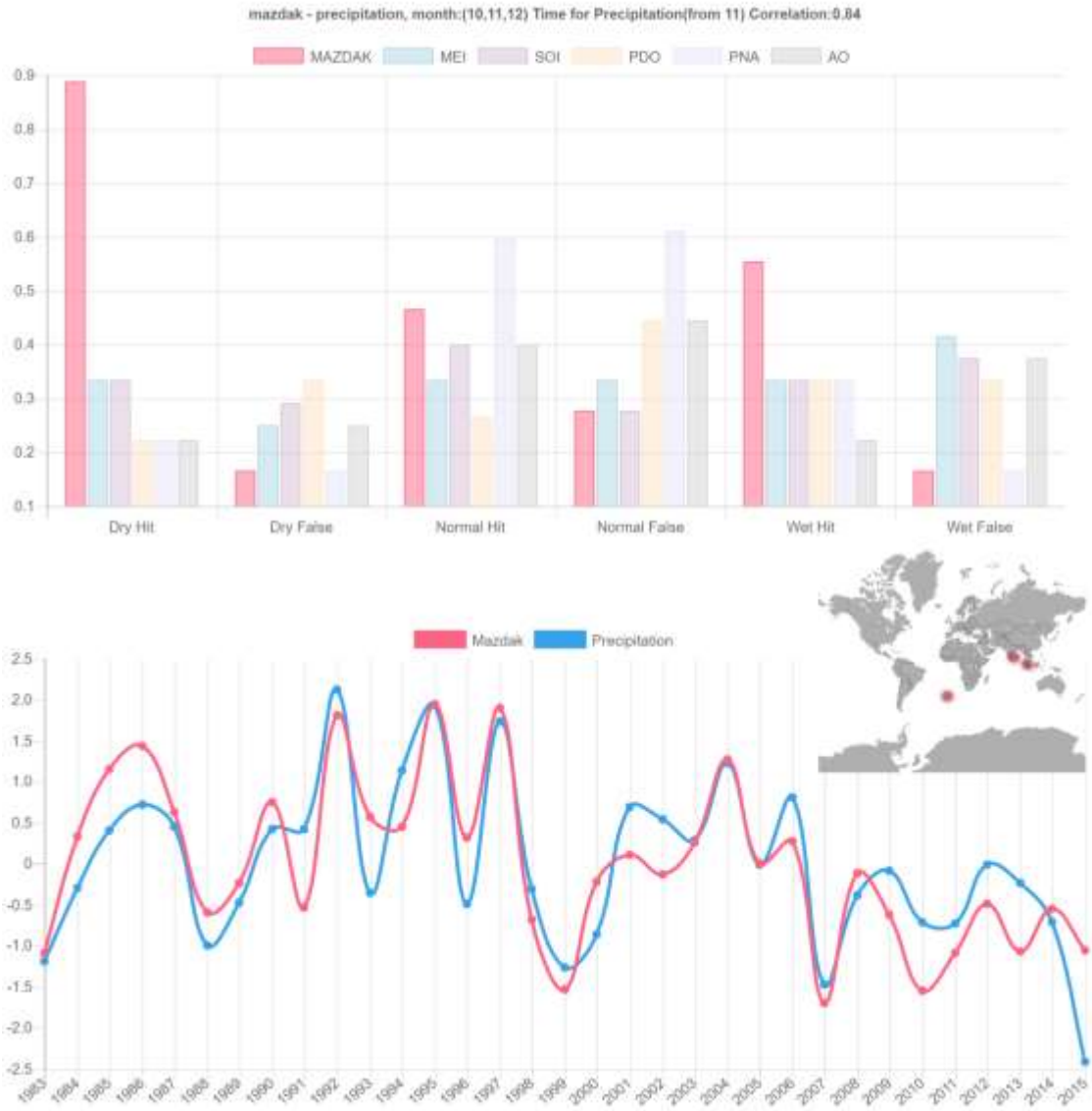



Figure S46| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

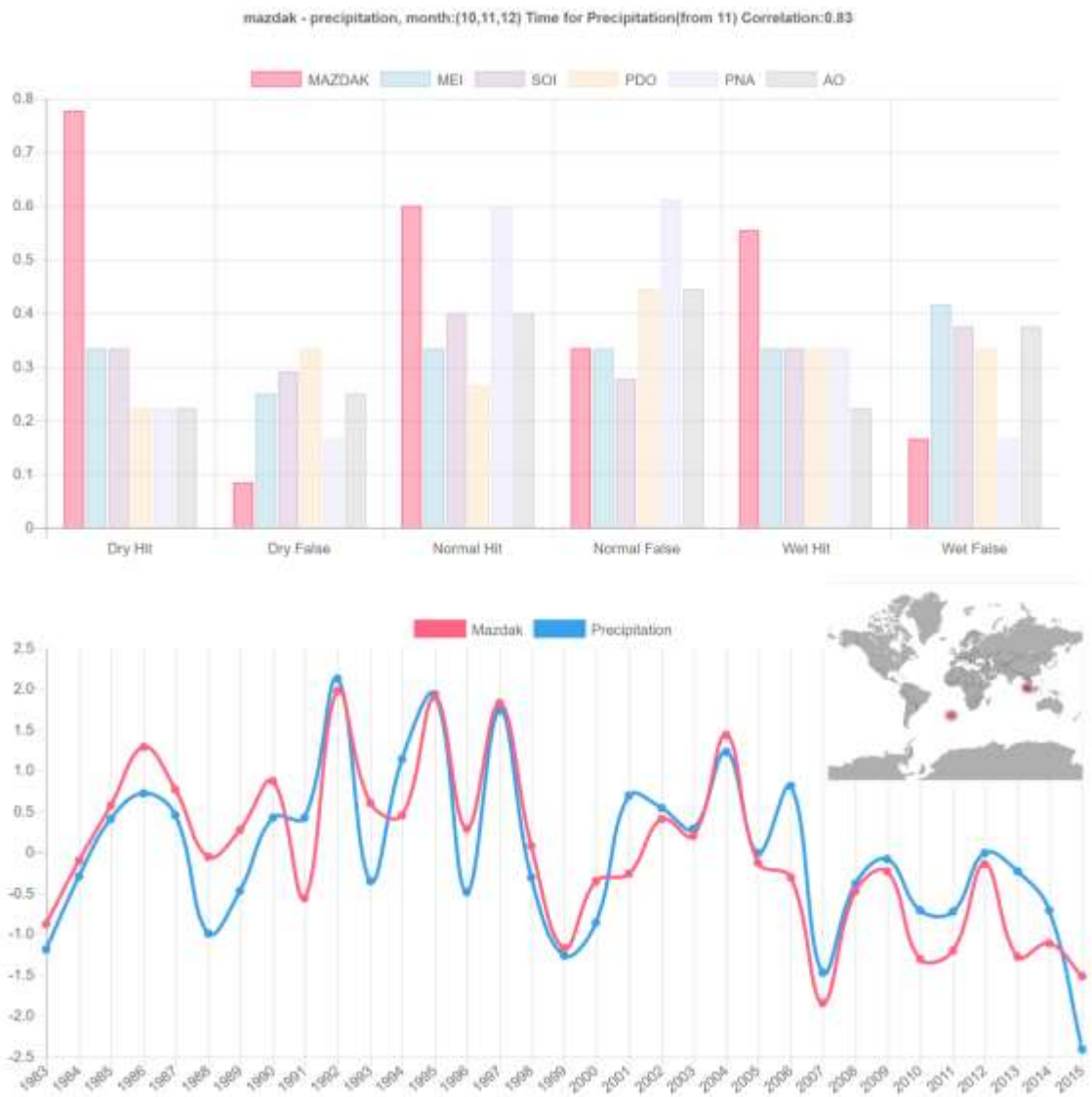



Figure S47| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

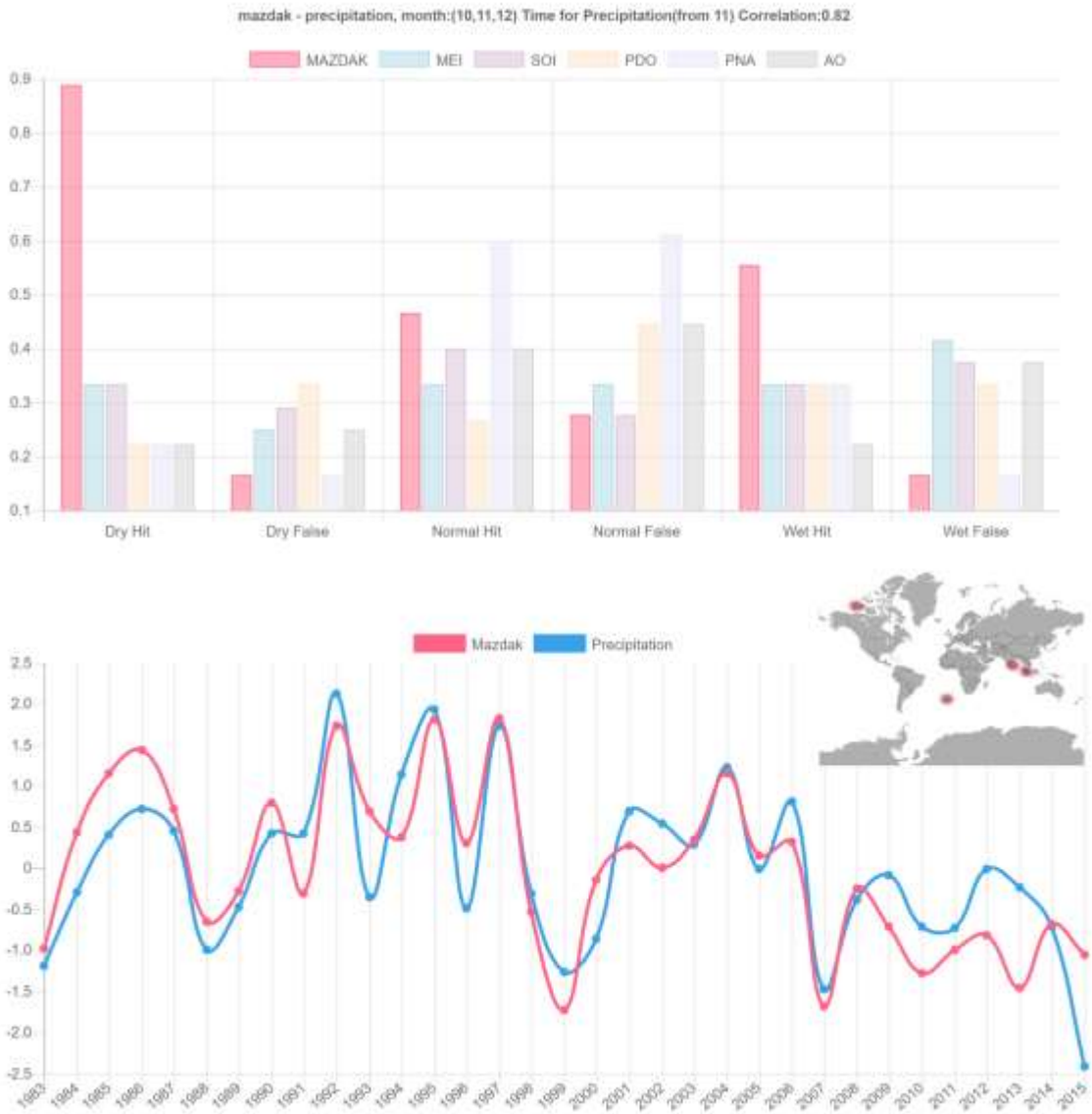



Figure S48| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

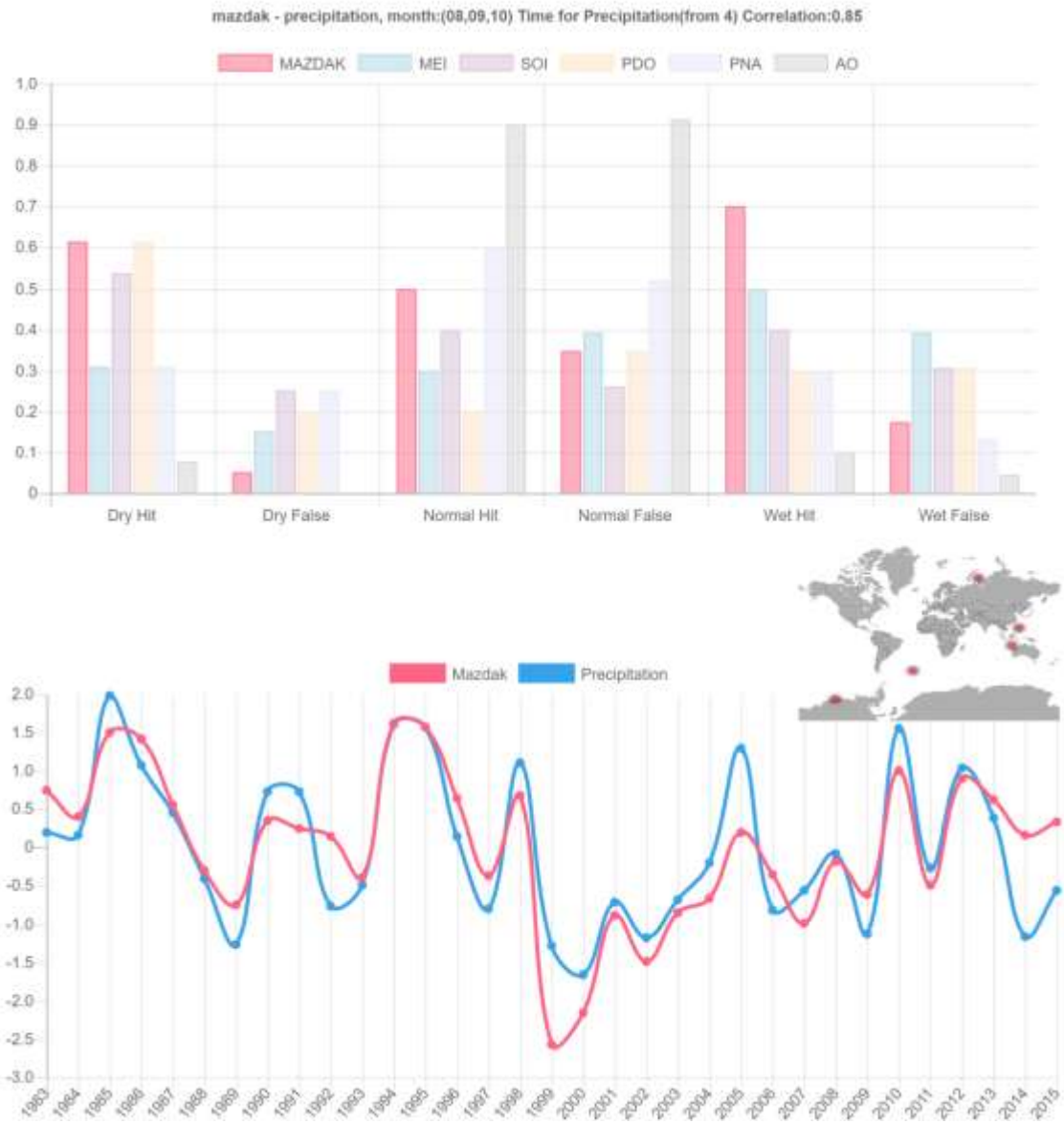


Figure S49| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

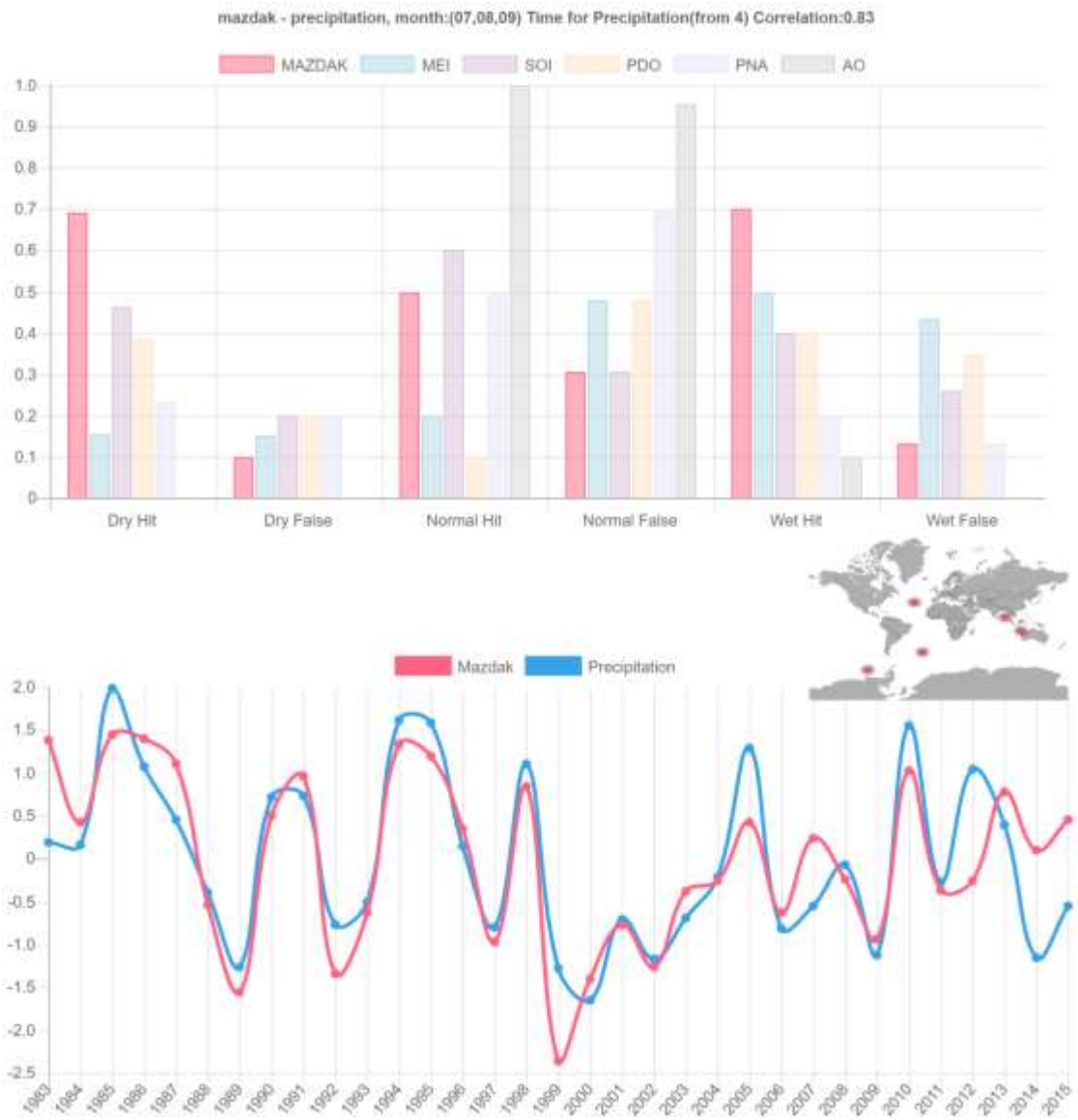



Figure S50| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

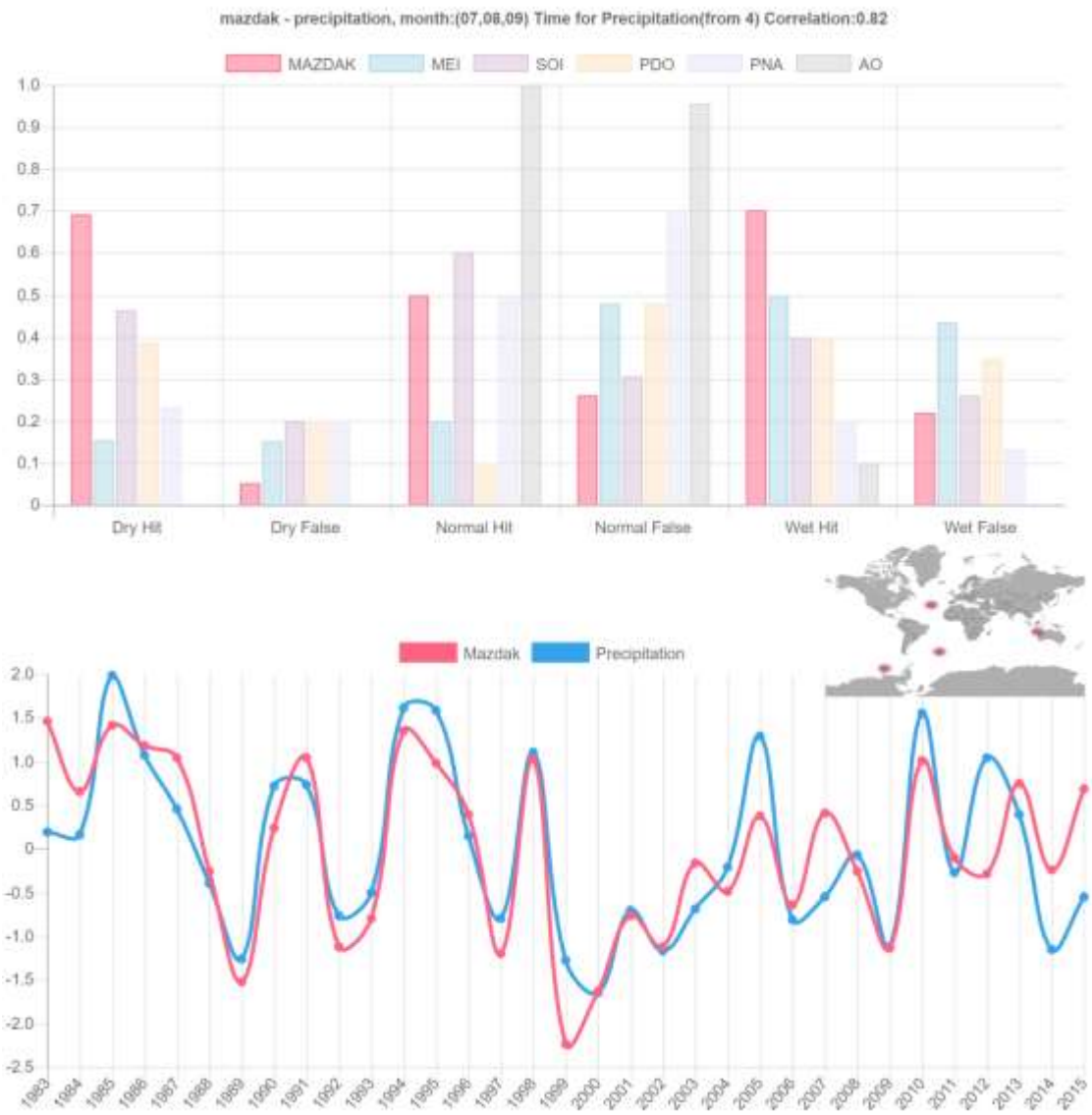



Figure S51| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

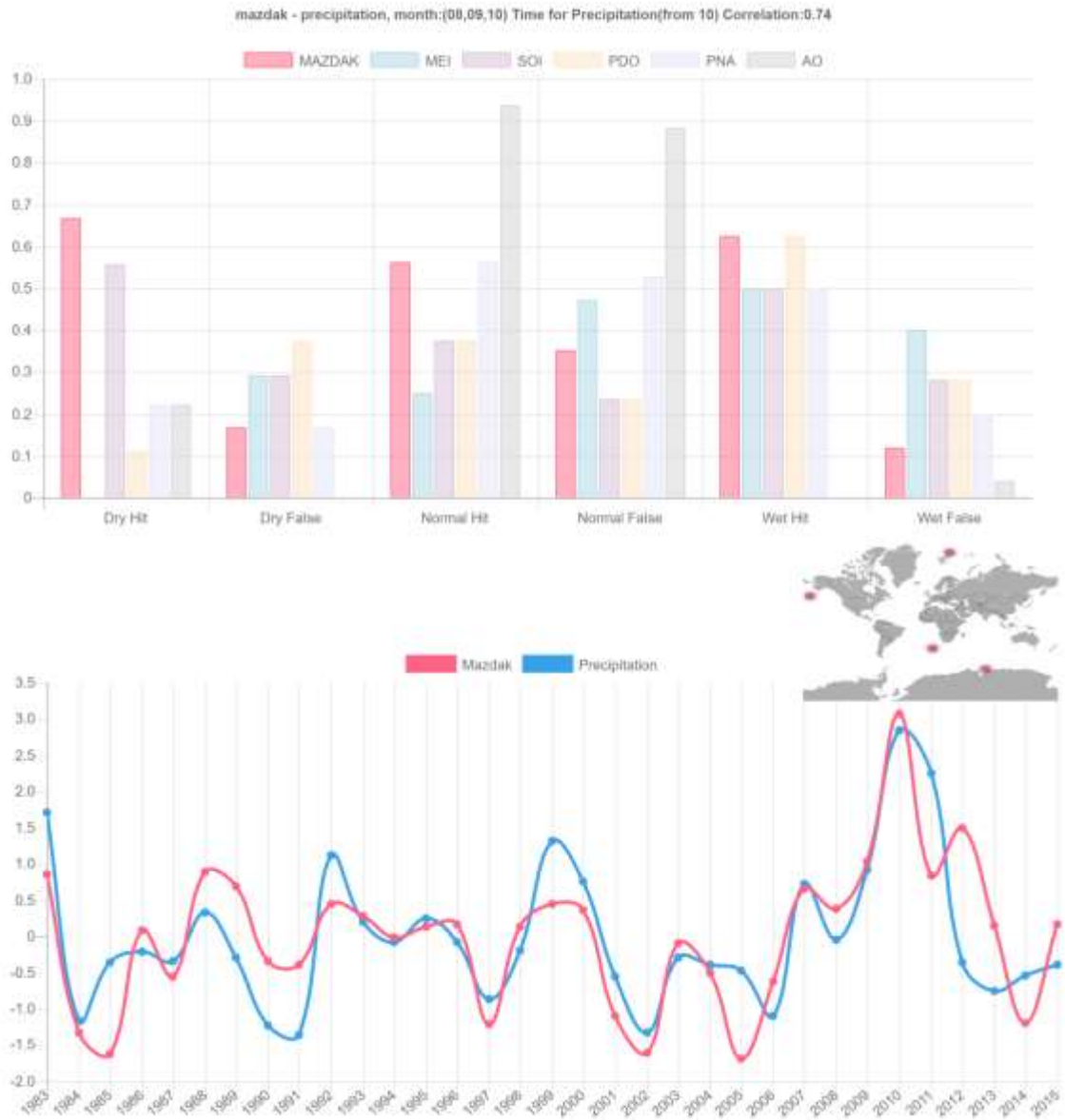



Figure S52| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

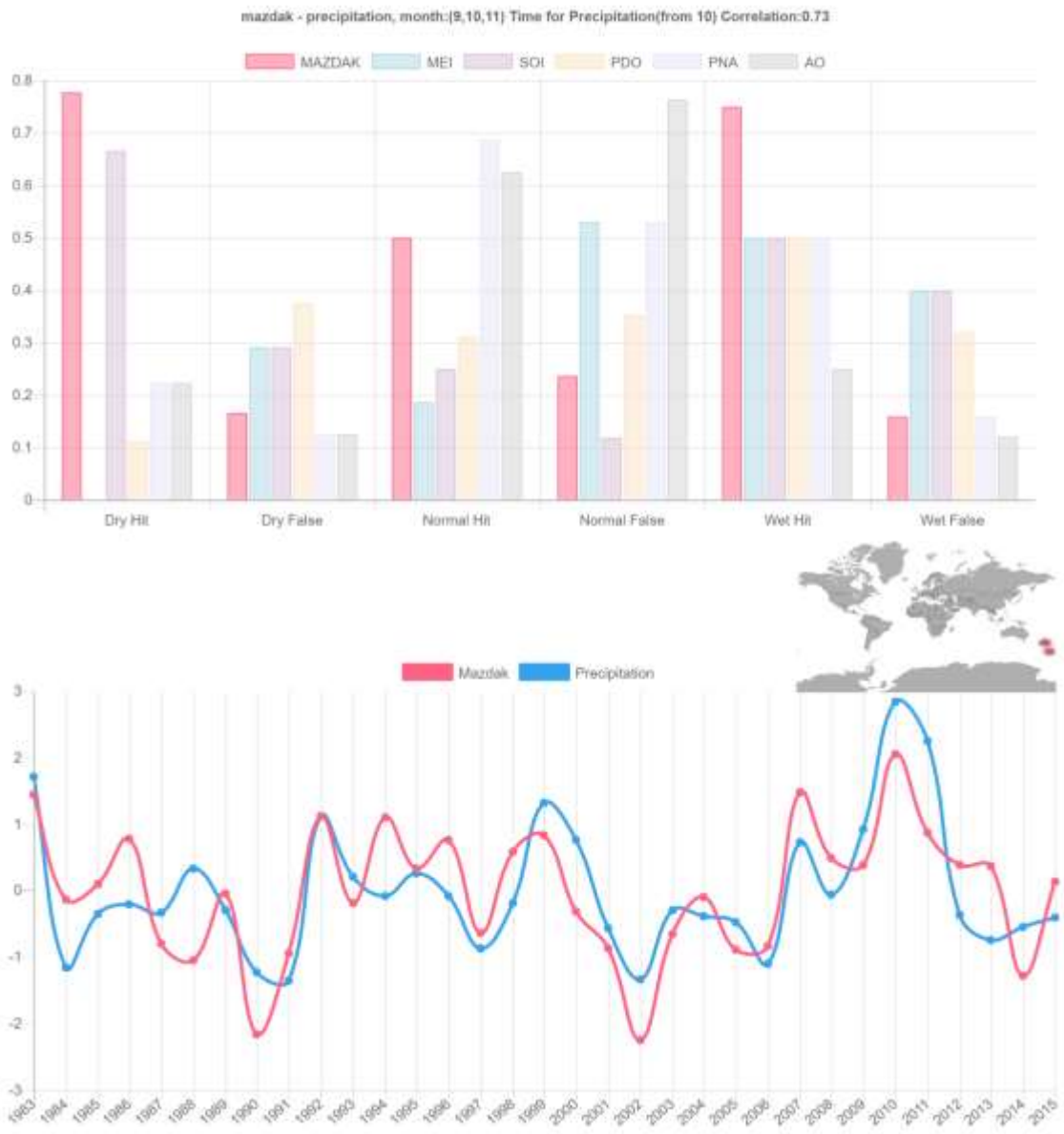



Figure S53| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

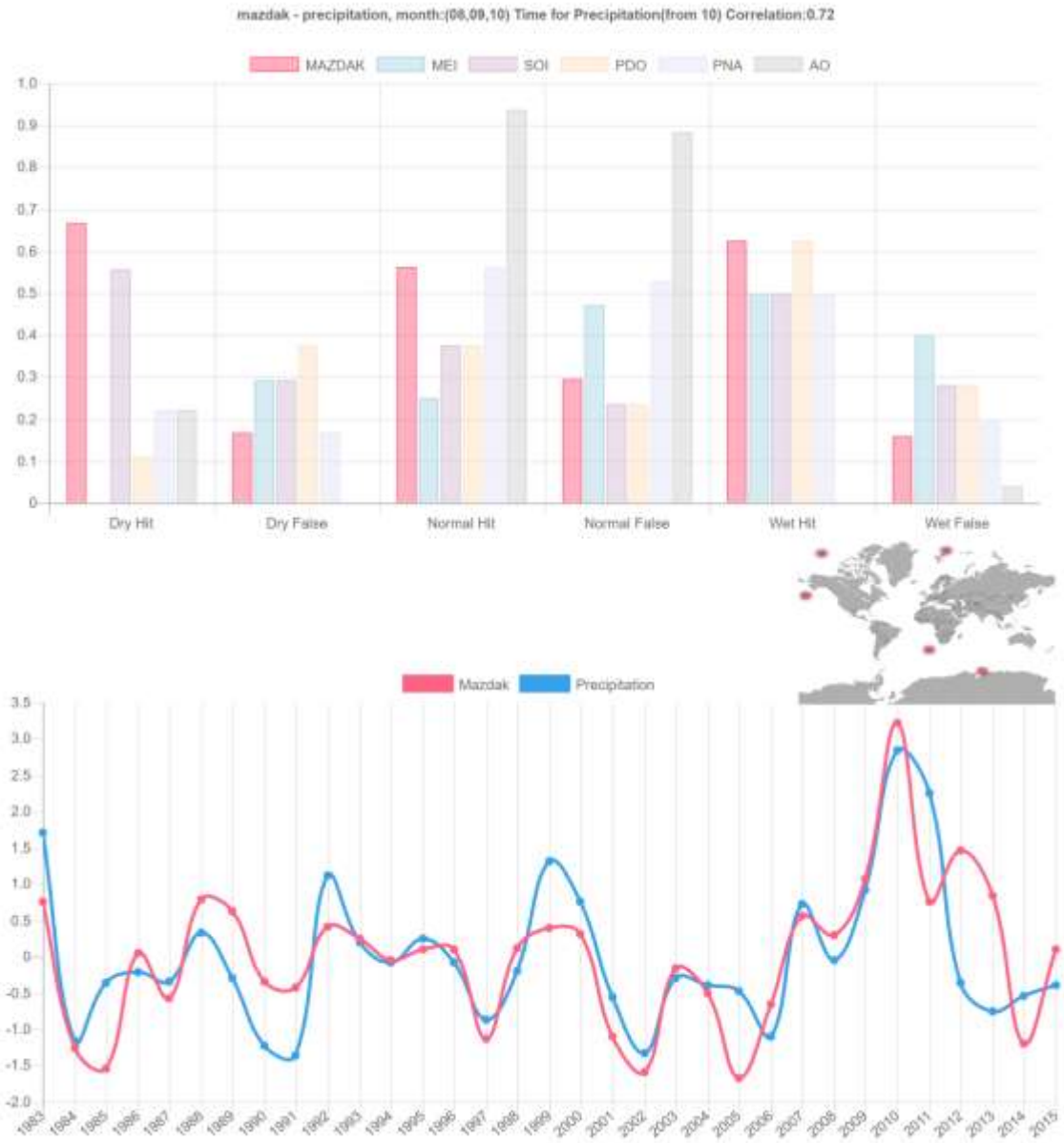



Figure S54| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

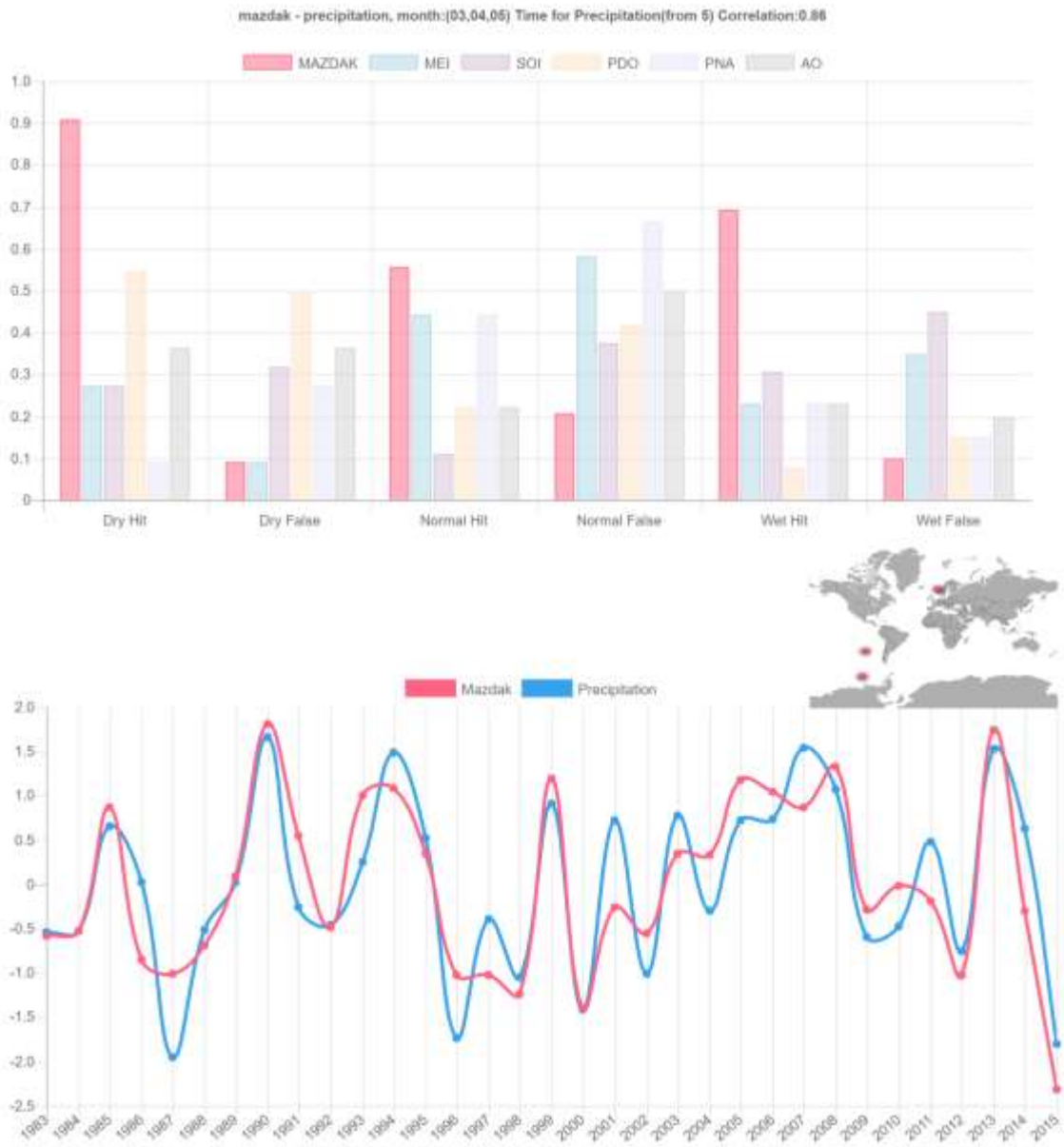


Figure S55| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S56| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

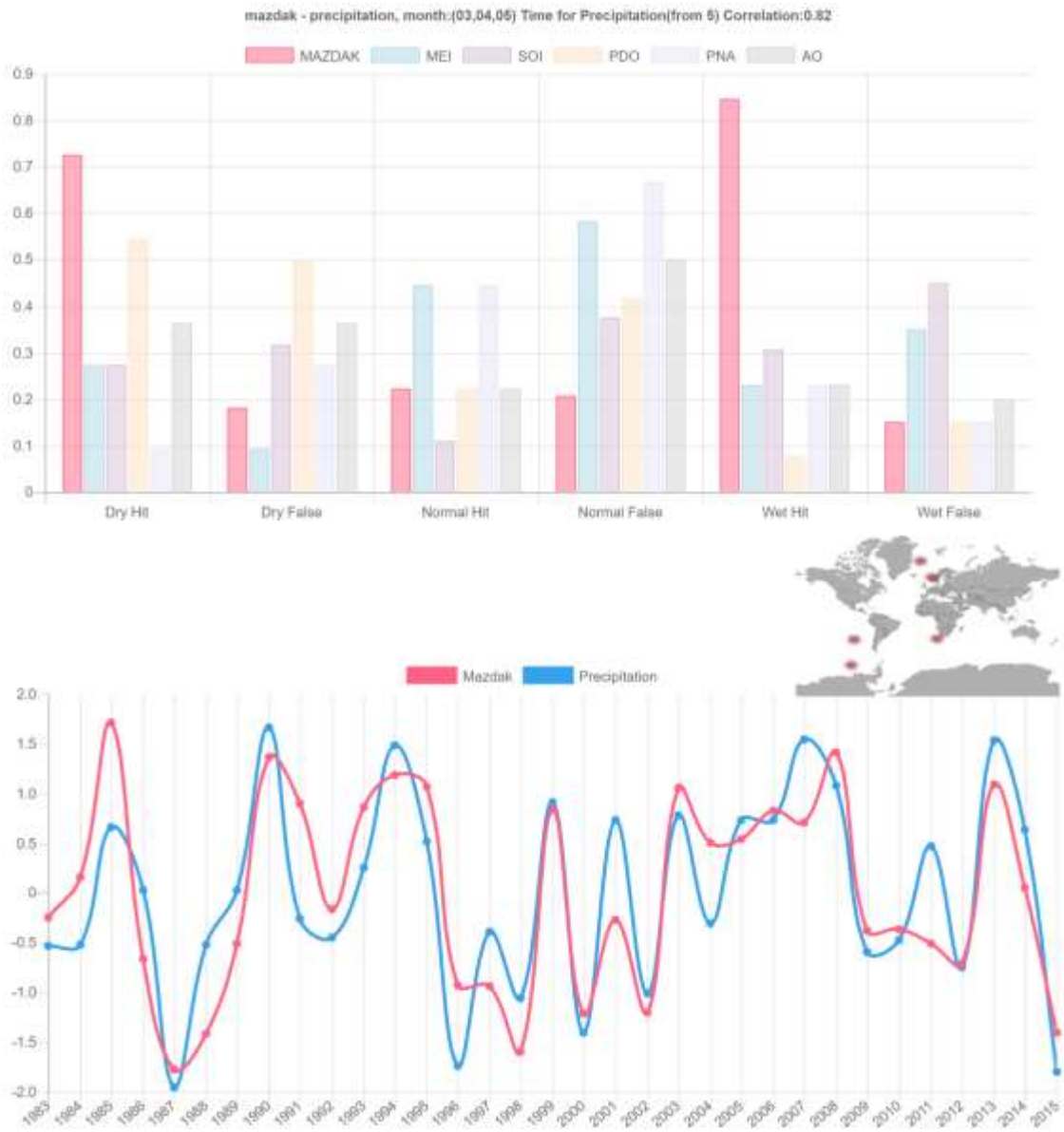



Figure S57| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

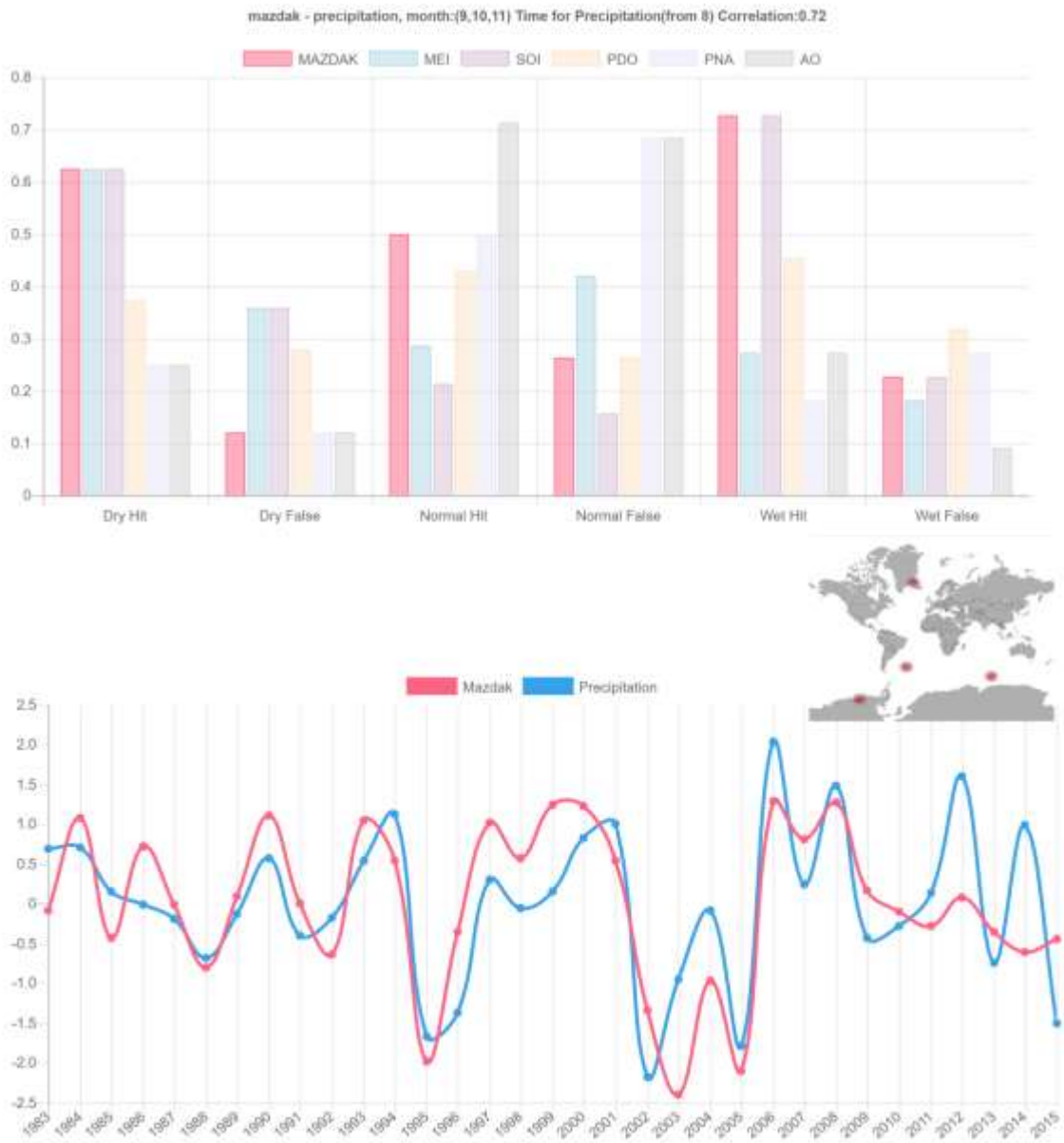



Figure S58| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

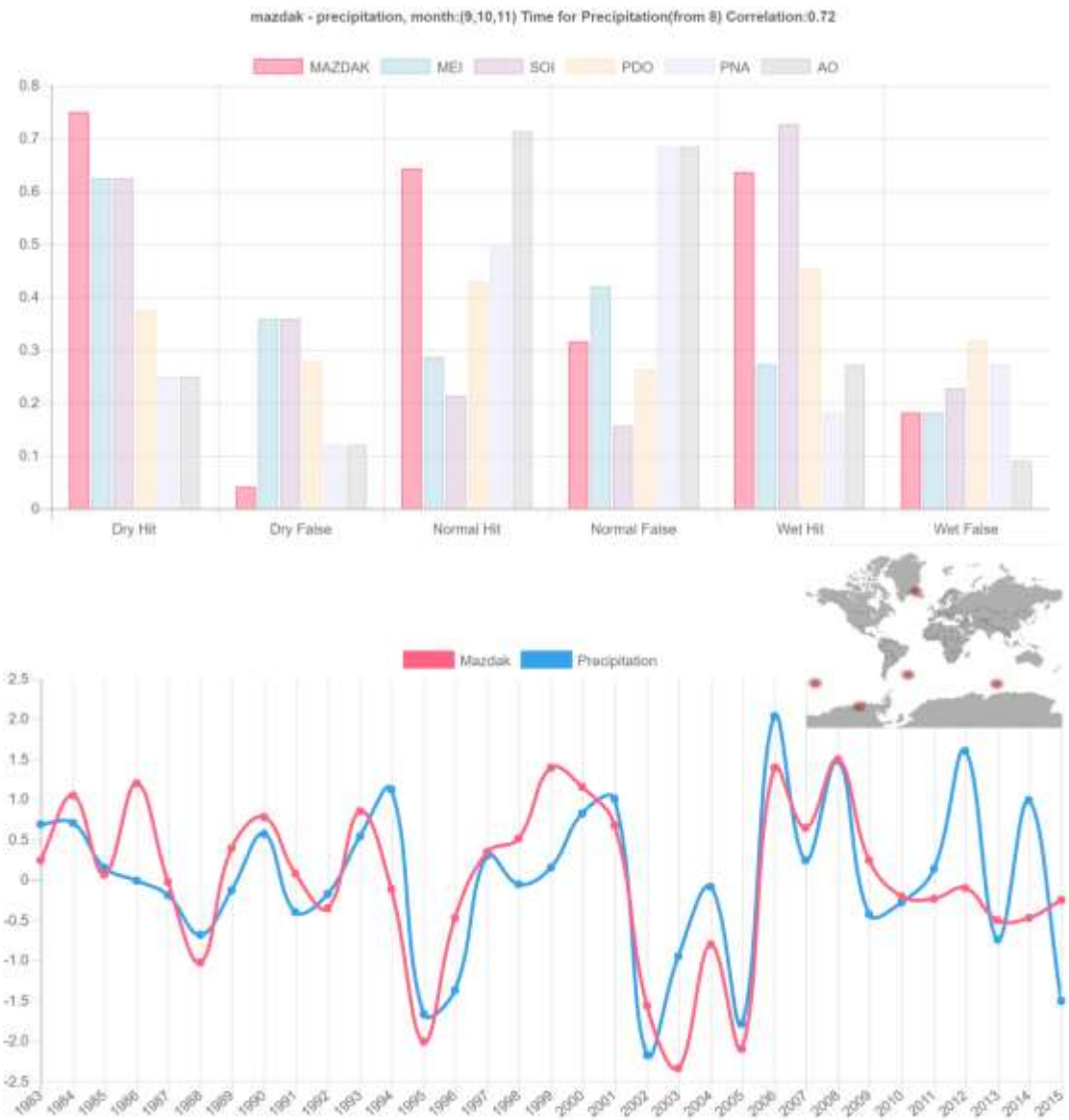



Figure S59| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

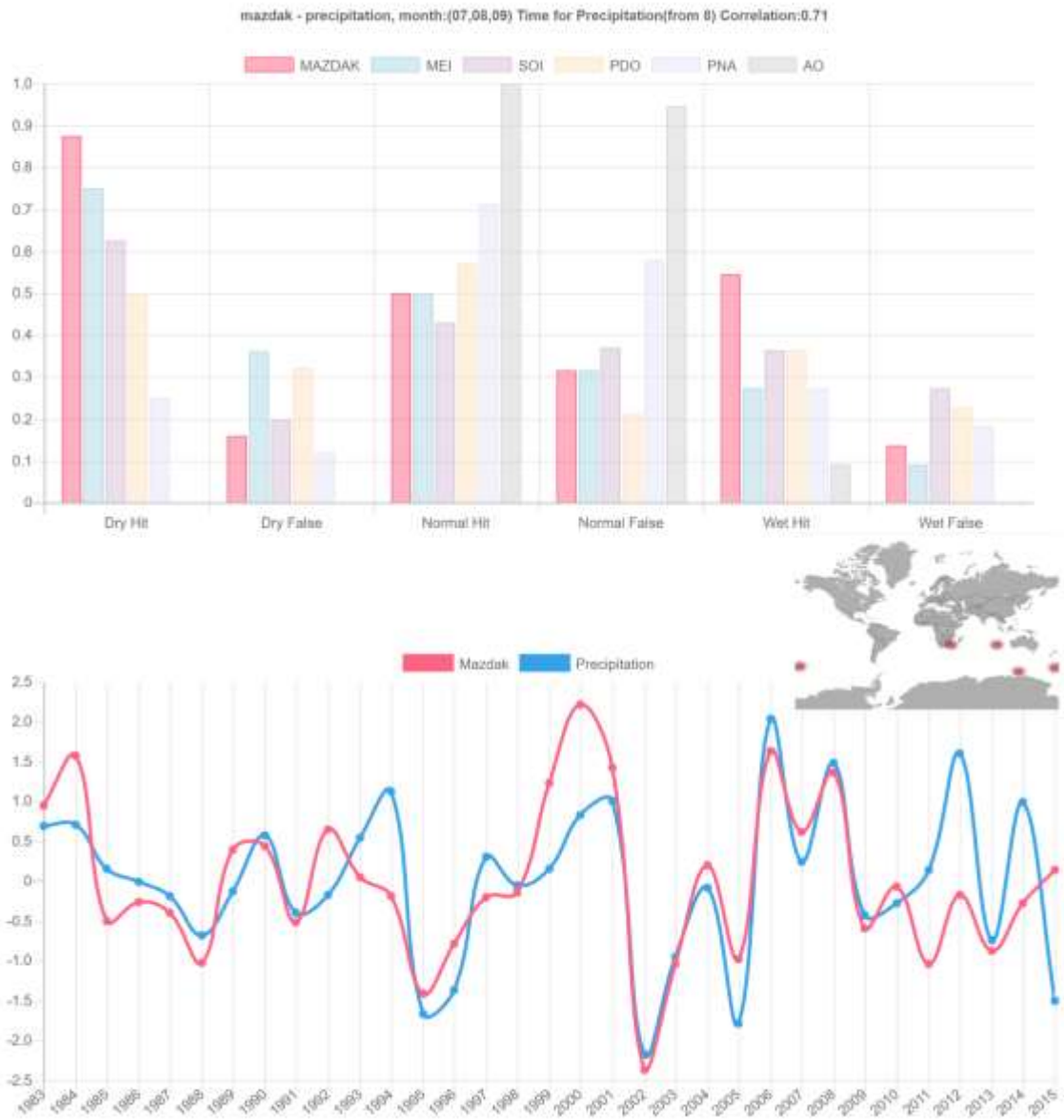



Figure S60| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

**Section 3:**
In this section the results are presented based on analysis of record over 1983-2015 and to find the locations that best estimate the precipitation and leave-one-out technique is used to evaluate the performance.


Figure S61| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S62| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S63| California-USA Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S64| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S65| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the Second highest correlation coefficient.

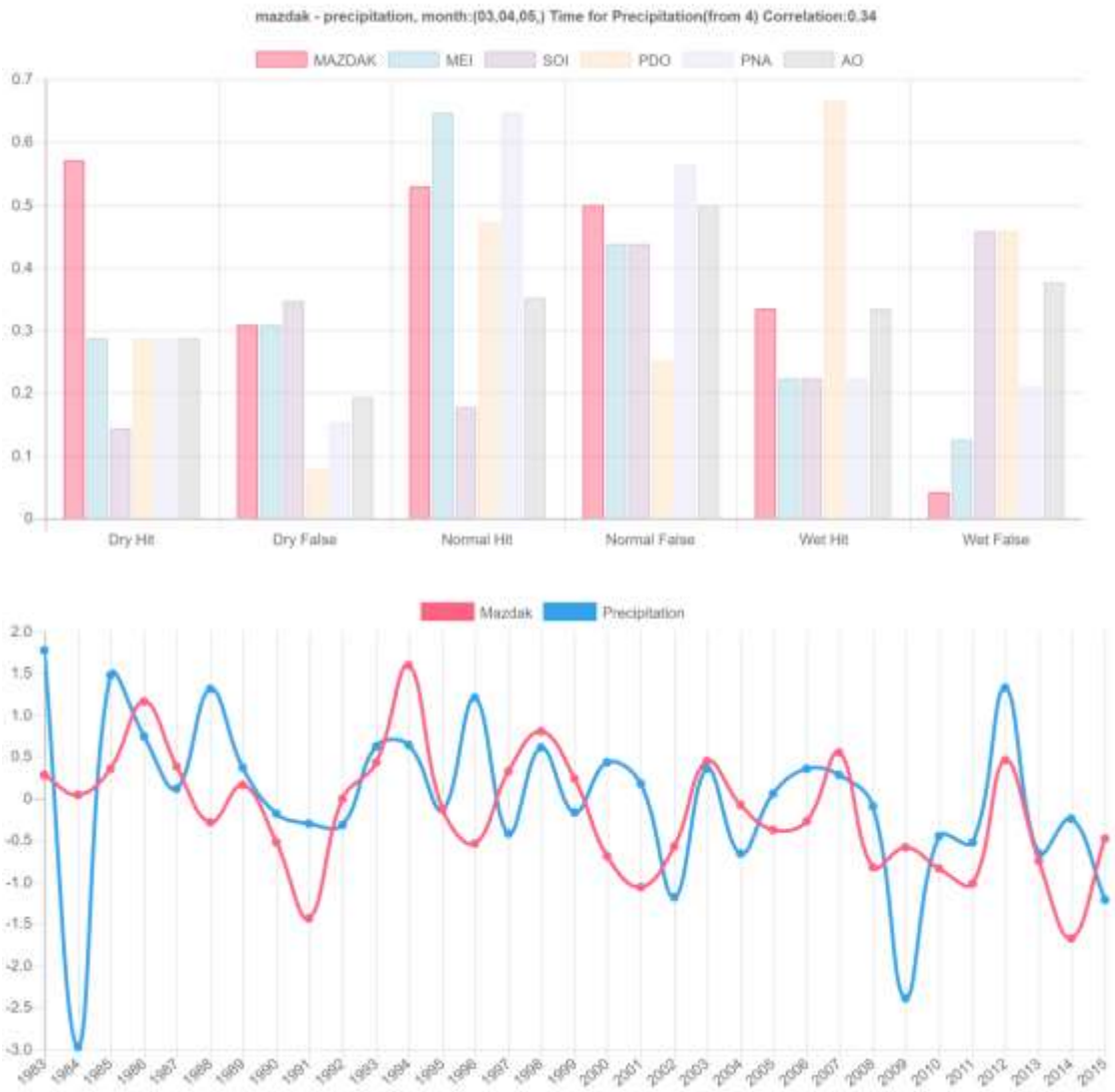



Figure S66| Ethiopia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

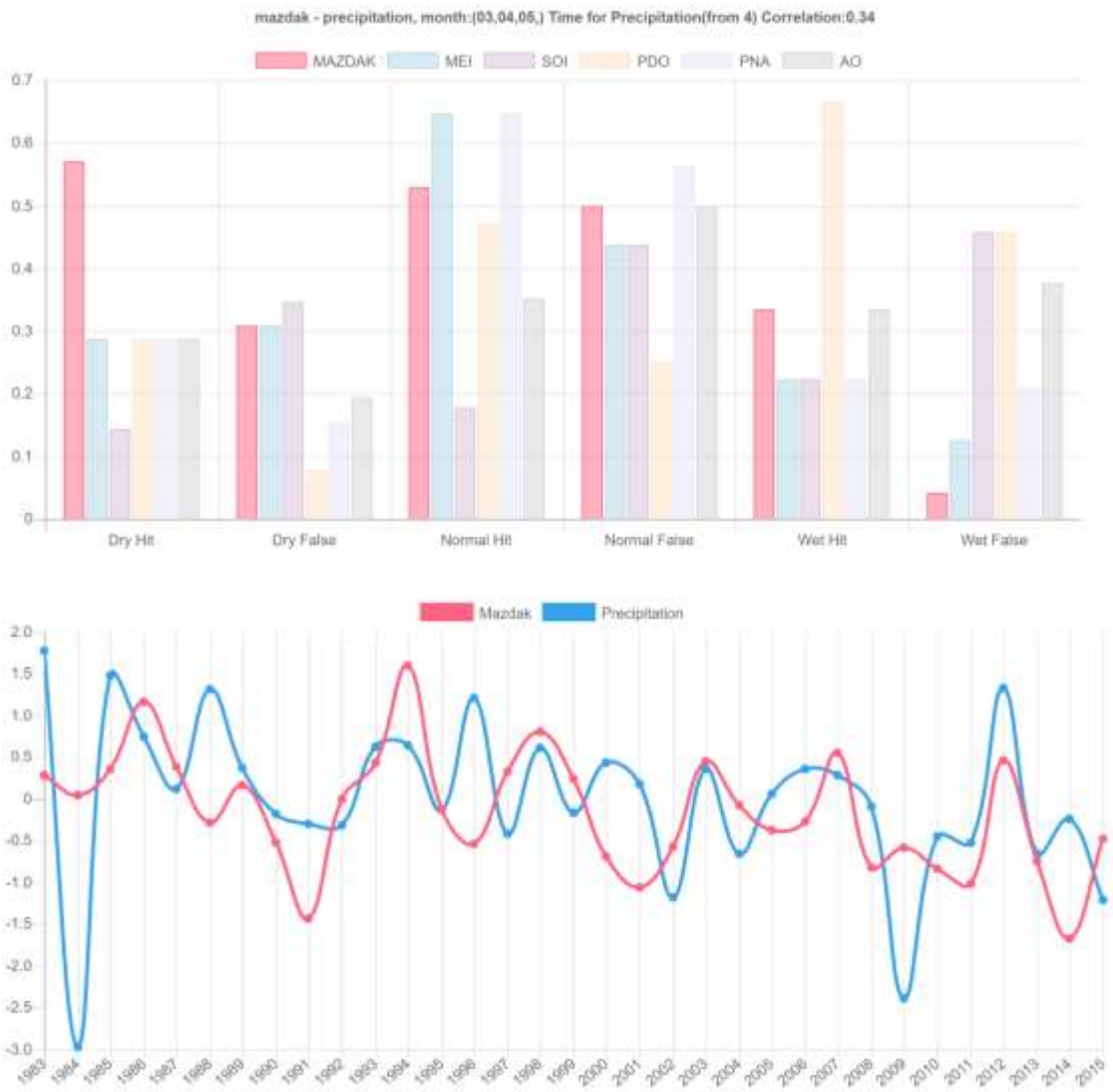



Figure S67| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S68| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S69| Goias-Brazil Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

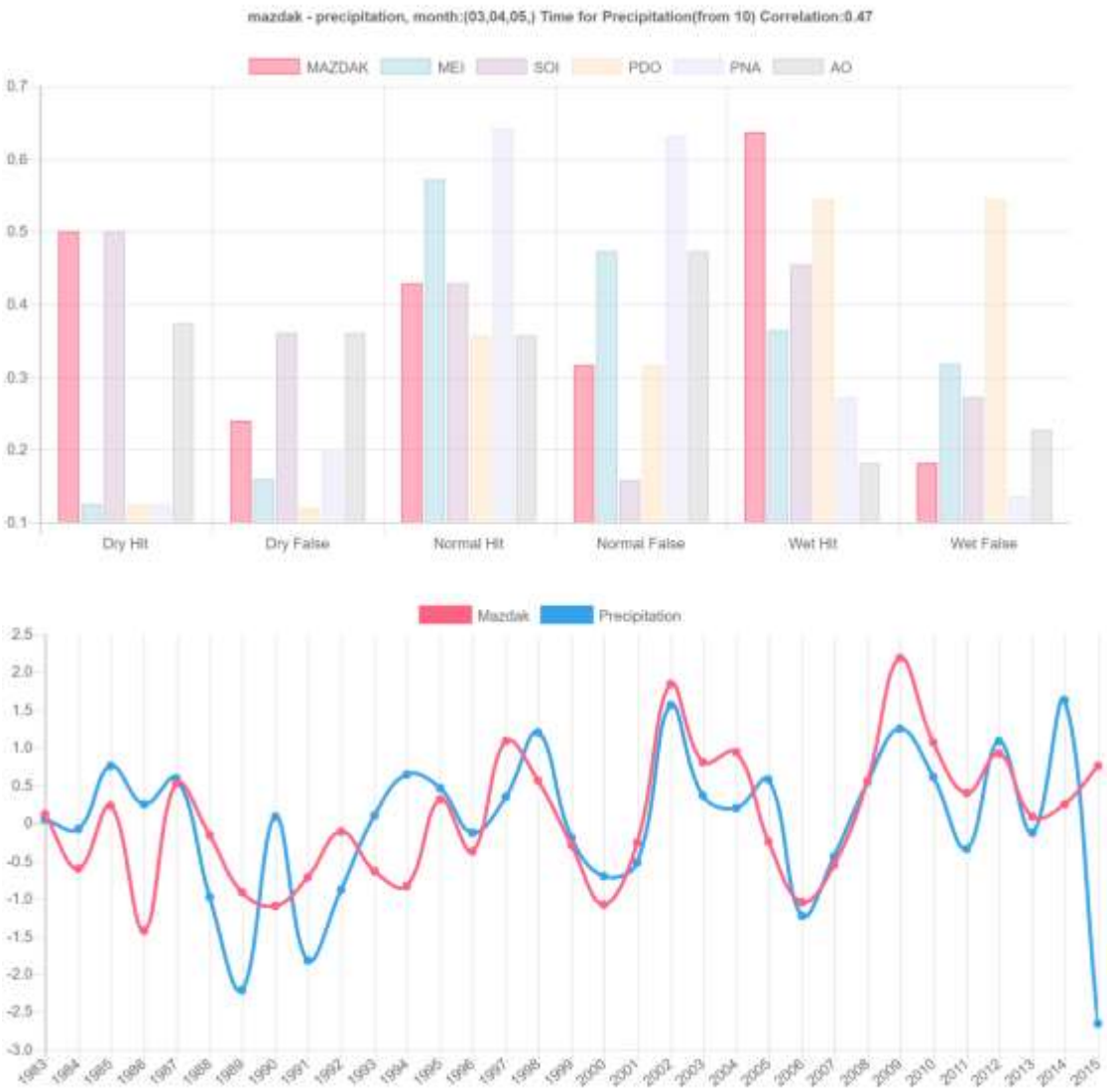



Figure S70| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.

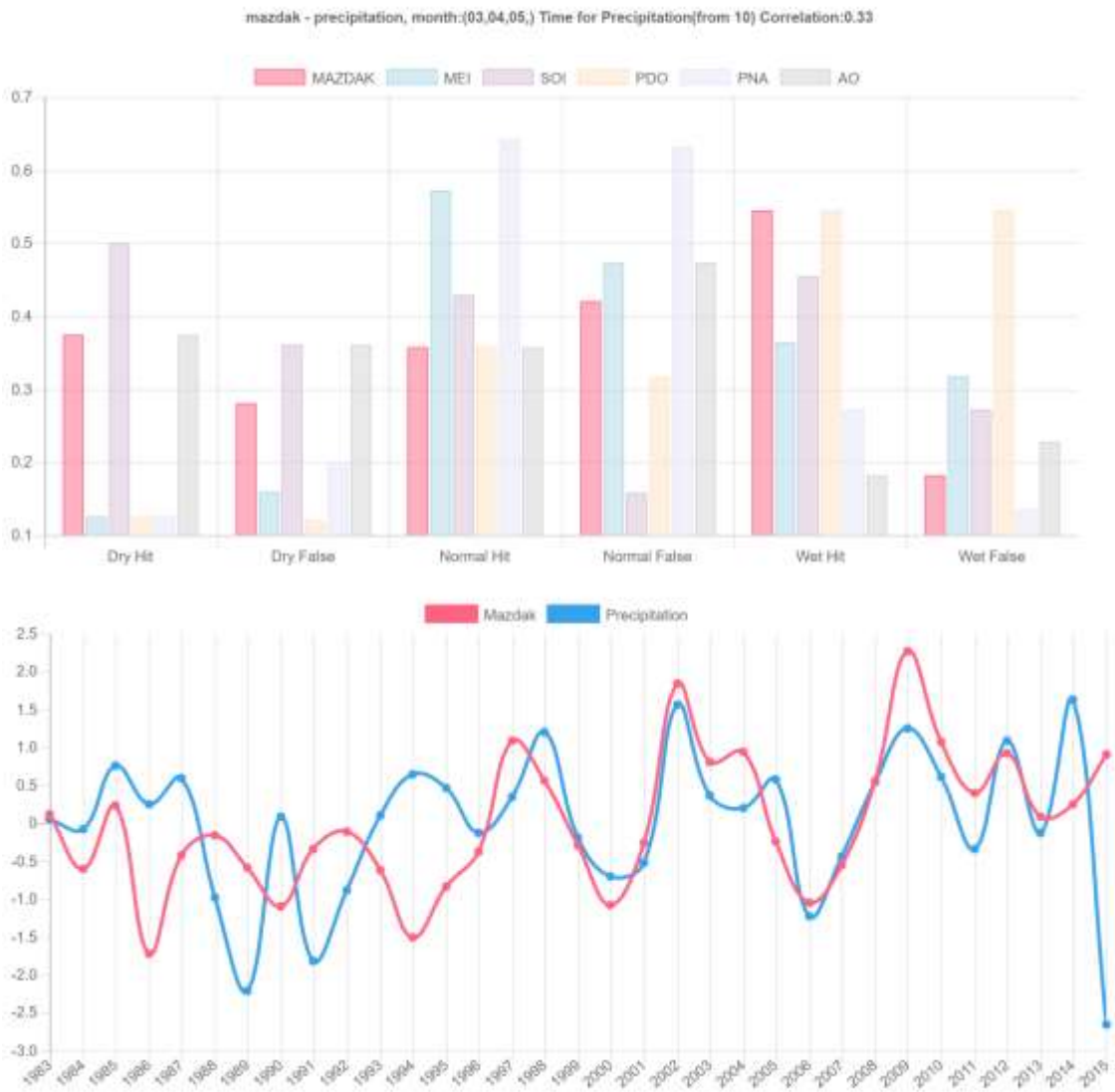



Figure S71| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S72| Greece Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S73| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S74| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

Figure S75| Hubei-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S76| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S77| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S78| Iran Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S79| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S80| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S81| Liaoning-China Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

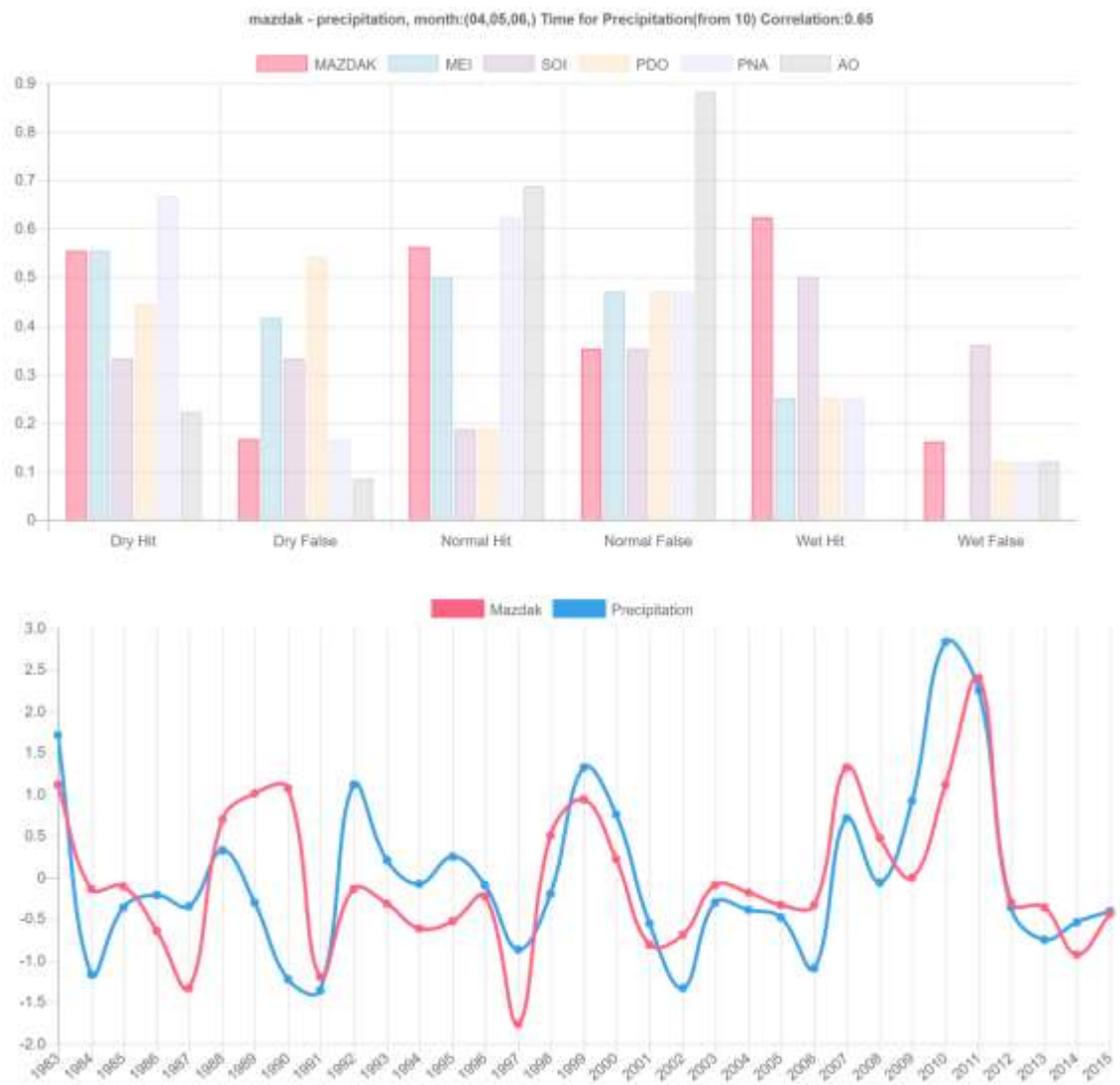



Figure S82| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S83| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.

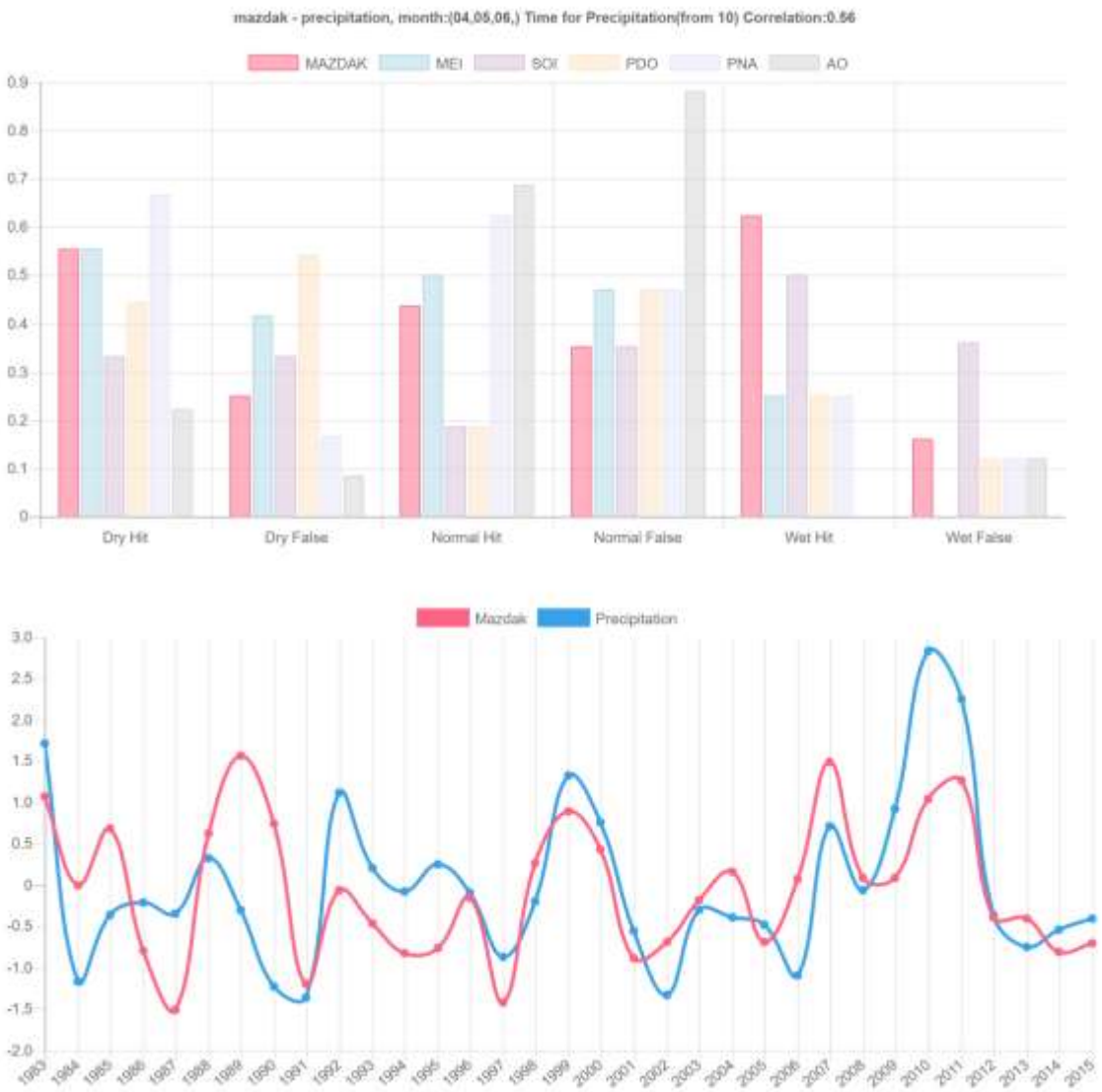



Figure S84| New South Wales-Australia Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S85| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S86| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S87| Odisha-India Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.


Figure S88| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the highest correlation coefficient.


Figure S89| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the second highest correlation coefficient.


Figure S90| Stockholm-Sweden Precipitation and MAZDAK bar graph, timeseries and location, for the location combination with the third highest correlation coefficient.

**Section 4: Correlation Maps**

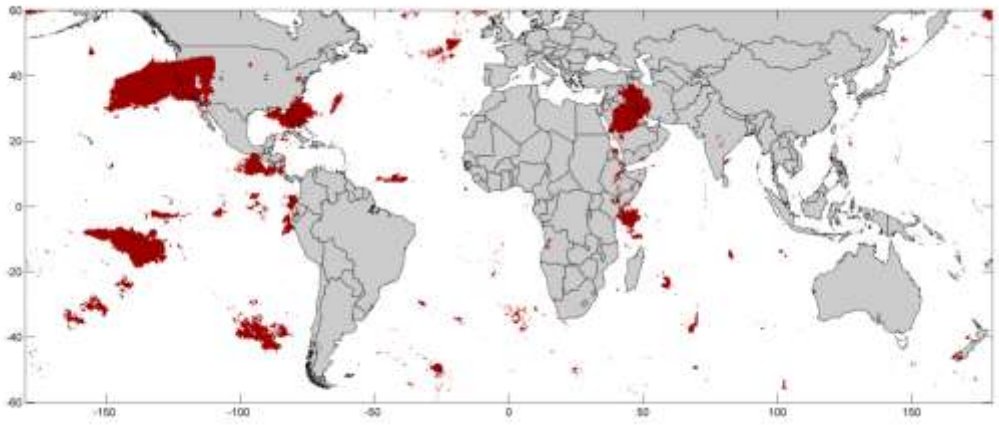


Figure S91| MAZDAK significant correlation map for California-USA.

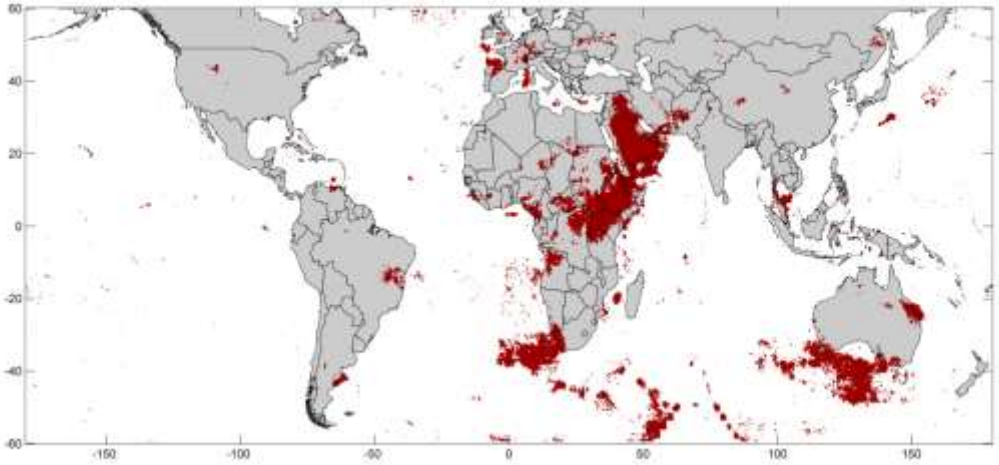


Figure S92| MAZDAK significant correlation map for Ethiopia.

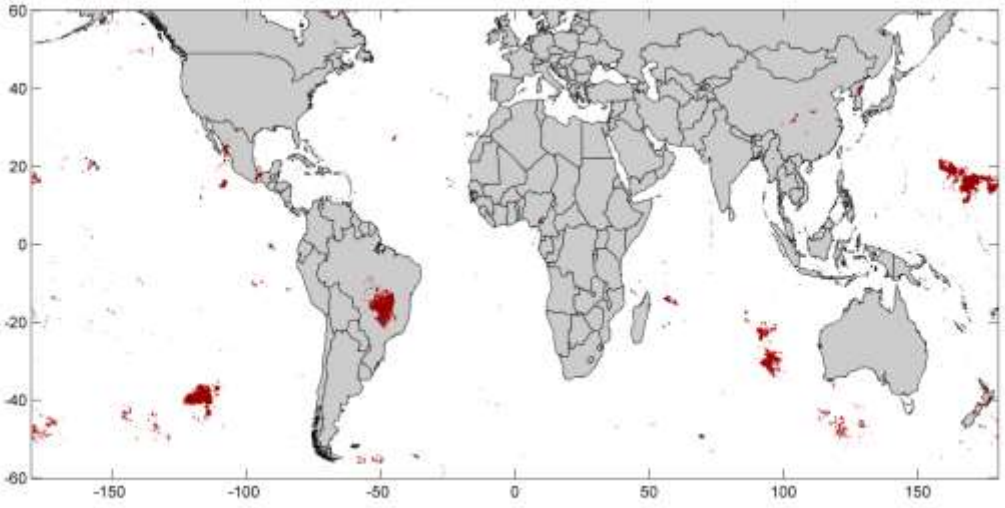



Figure S93| MAZDAK significant correlation map for Goias-Brazil.

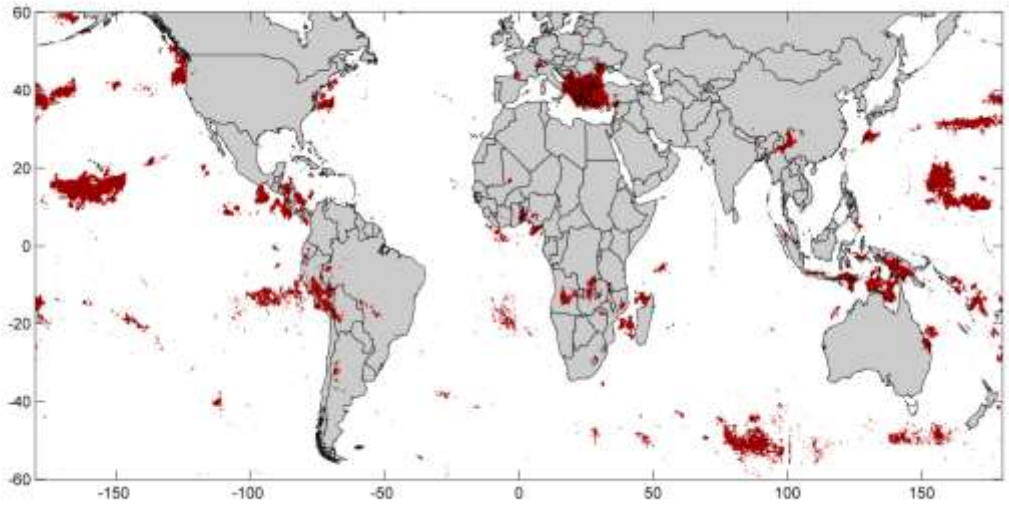



Figure S94| MAZDAK significant correlation map for Greece.

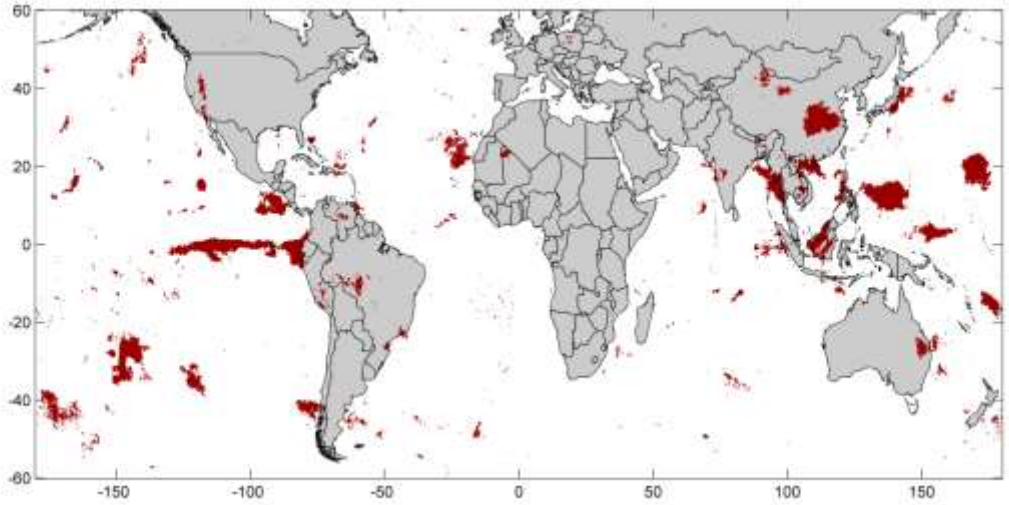



Figure S95| MAZDAK significant correlation map for Hubei-China.

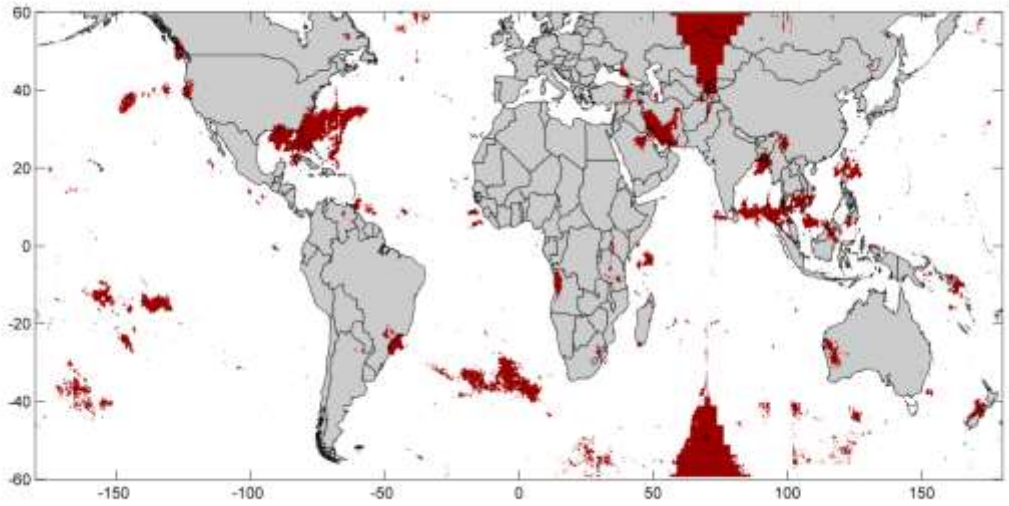



Figure S96| MAZDAK significant correlation map for Iran.

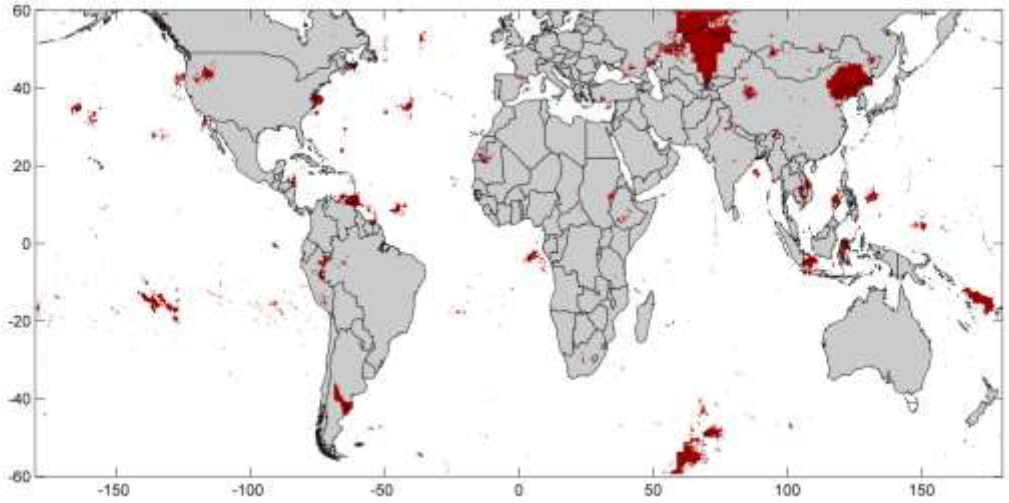



Figure S97| MAZDAK significant correlation map for Liaoning-China.

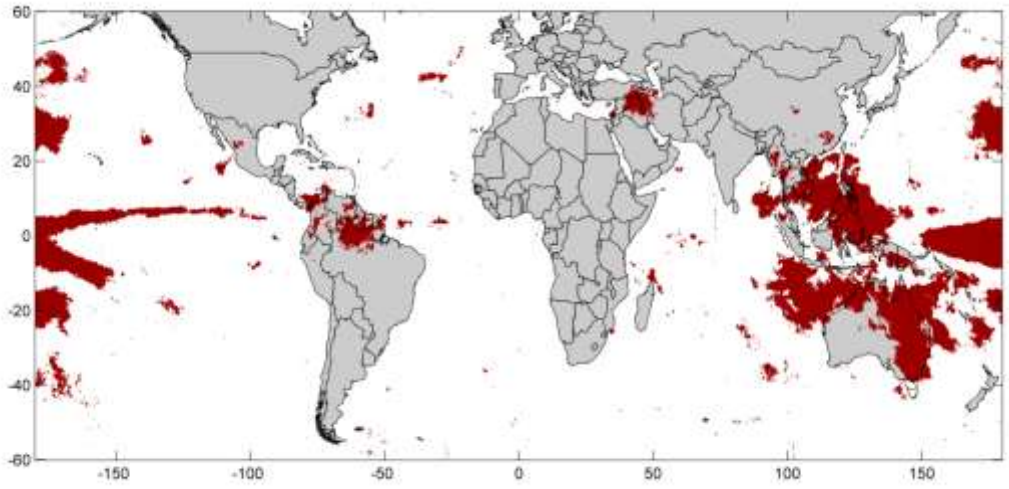



Figure S98| MAZDAK significant correlation map for New South Wales-Australia.

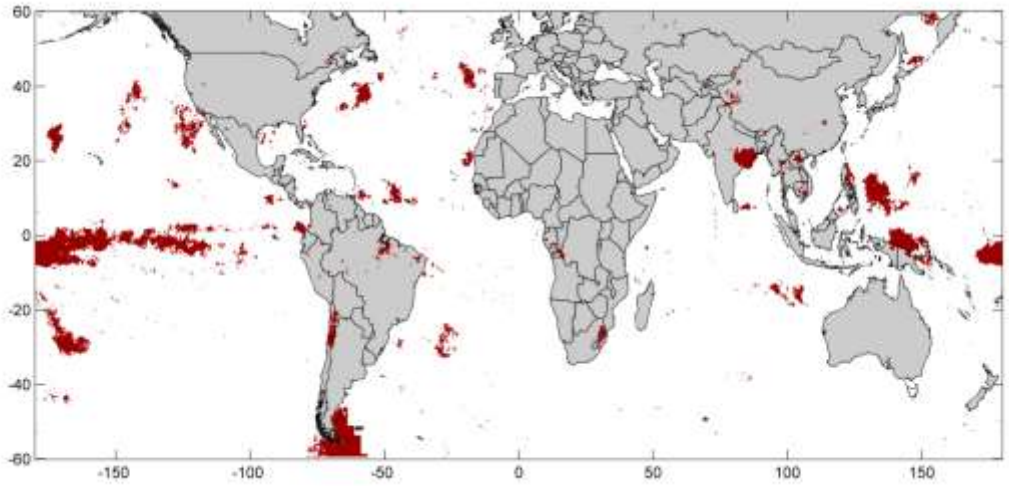



Figure S99| MAZDAK significant correlation map for Odisha-India.

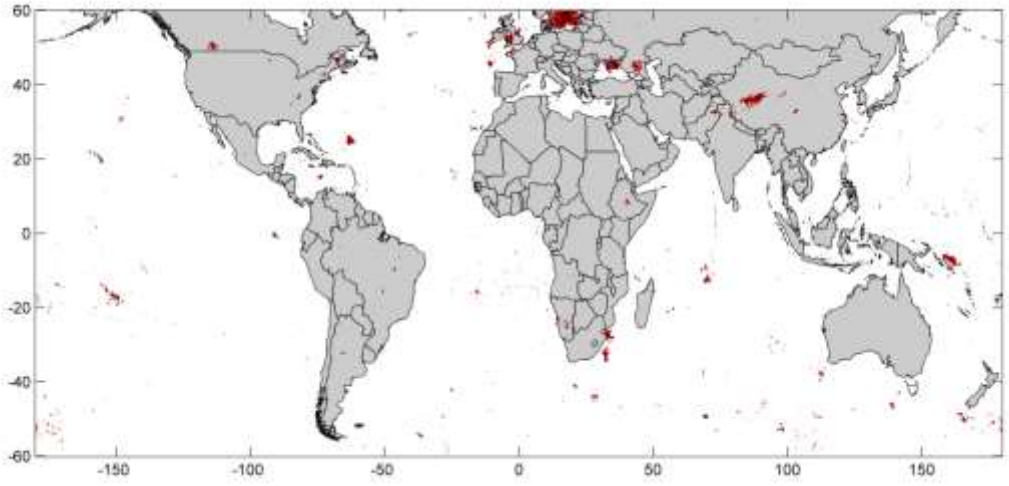


Figure S100| MAZDAK significant correlation map for Stockholm-Sweden.

**Section 5:**
In this section we show the locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with different lead-times.

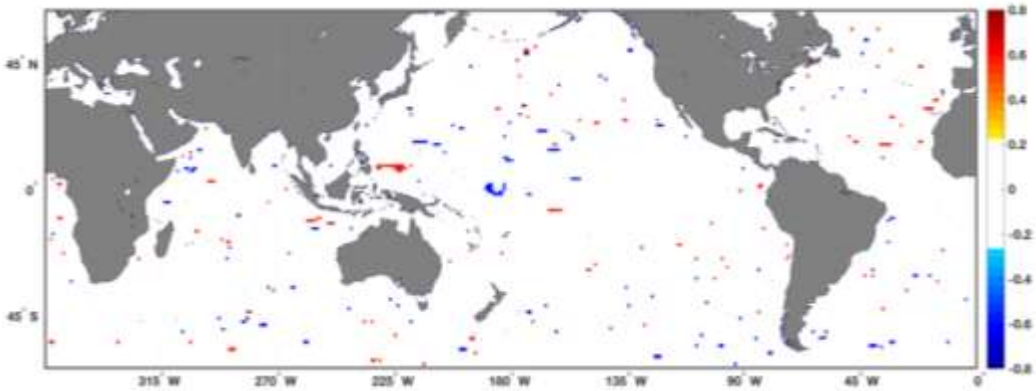


Figure S101| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 4 months lead-time.

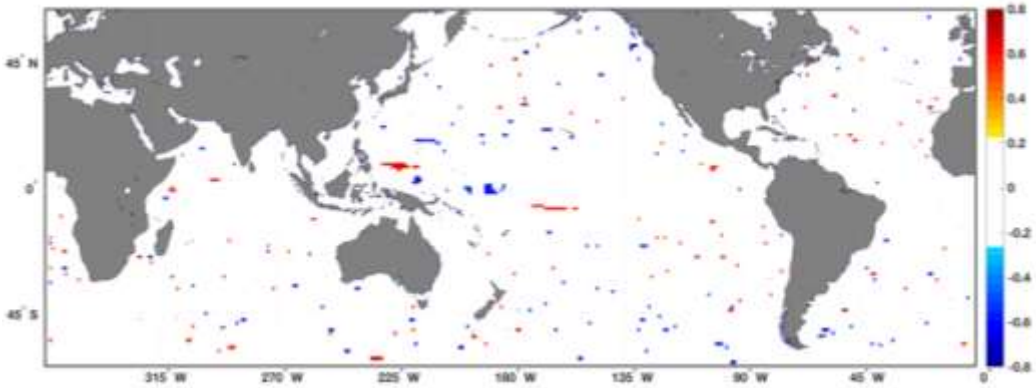


Figure S102| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 5 months lead-time.

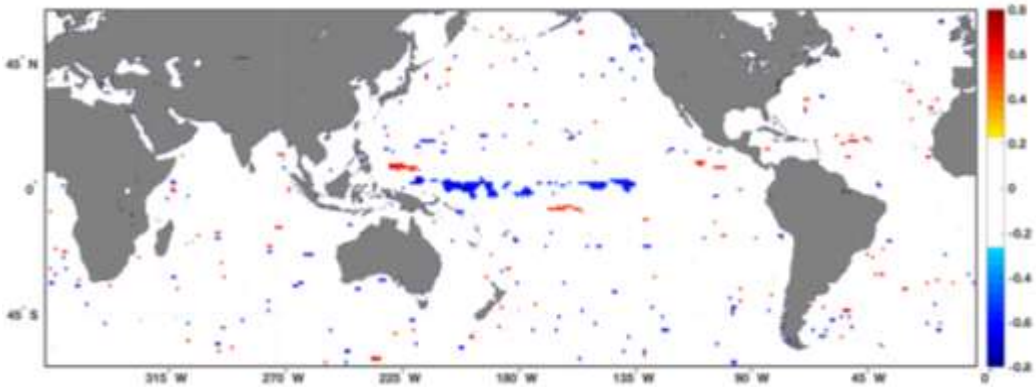


Figure S103| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 6 months lead-time.

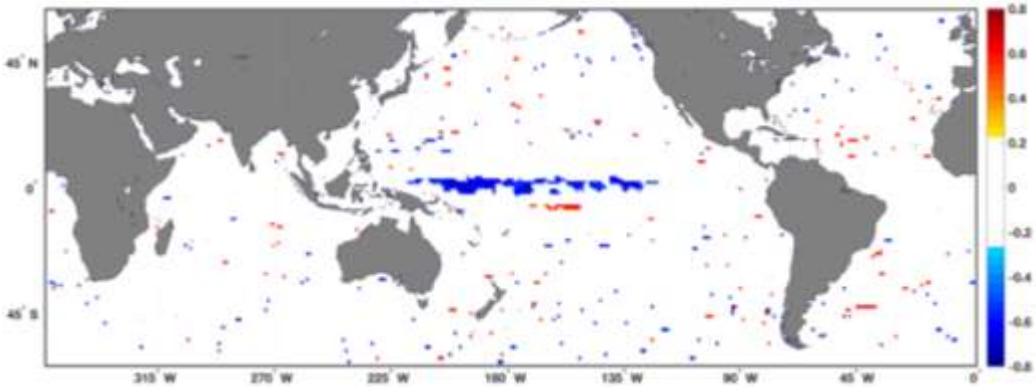


Figure S104| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 7 months lead-time.

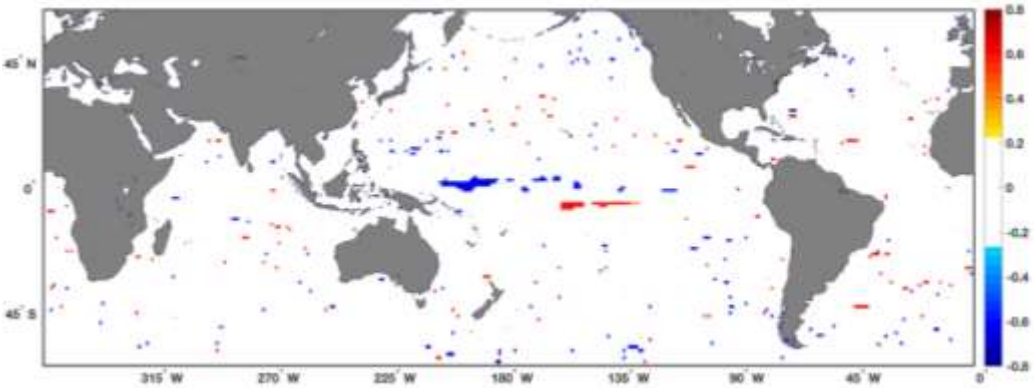


Figure S105| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 8 months lead-time.

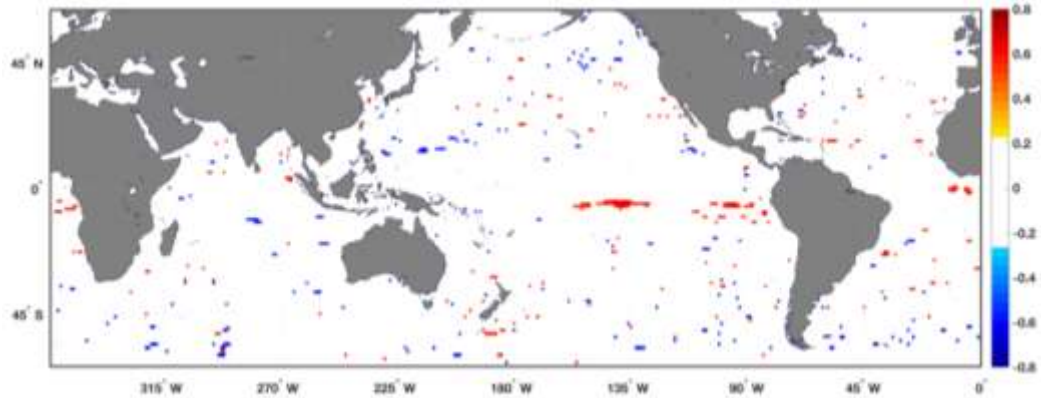


Figure S106| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 9 months lead-time.

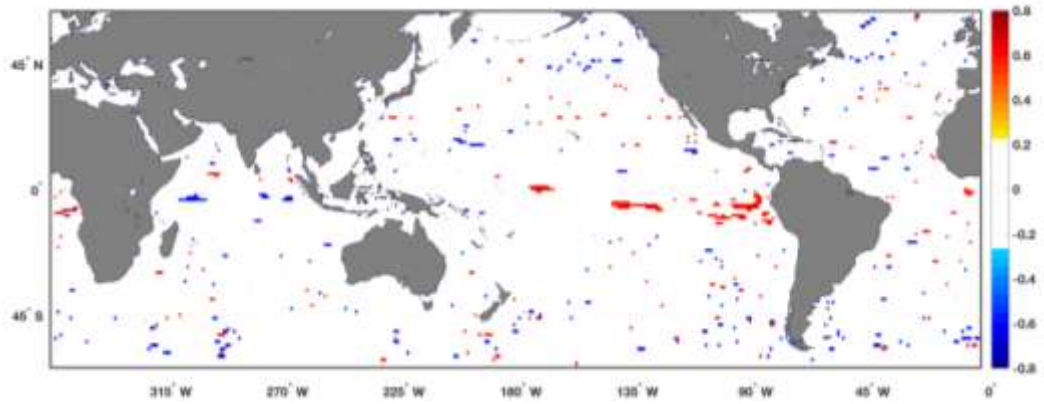


Figure S107| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 10 months lead-time.

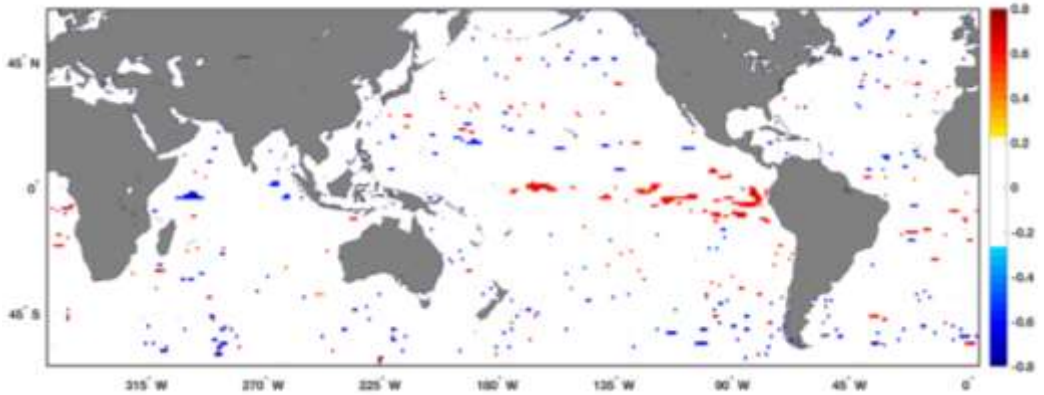


Figure S108| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 11 months lead-time.

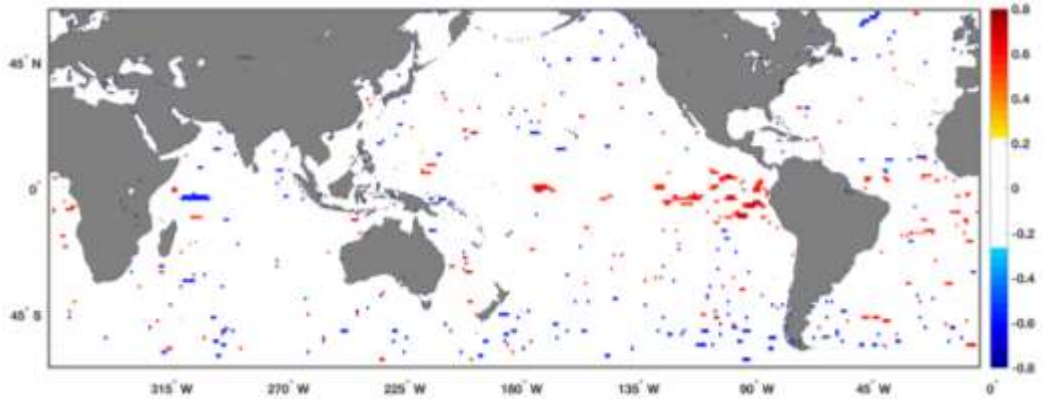


Figure S109| locations where ocean current velocity has a statistically significant correlation with California wet-season precipitation with 12 months lead-time.

**Section 6:**

In this section we show that tropopause level pressure at different locations show no statistically significant correlation with California wet-season precipitation with different lead-times.

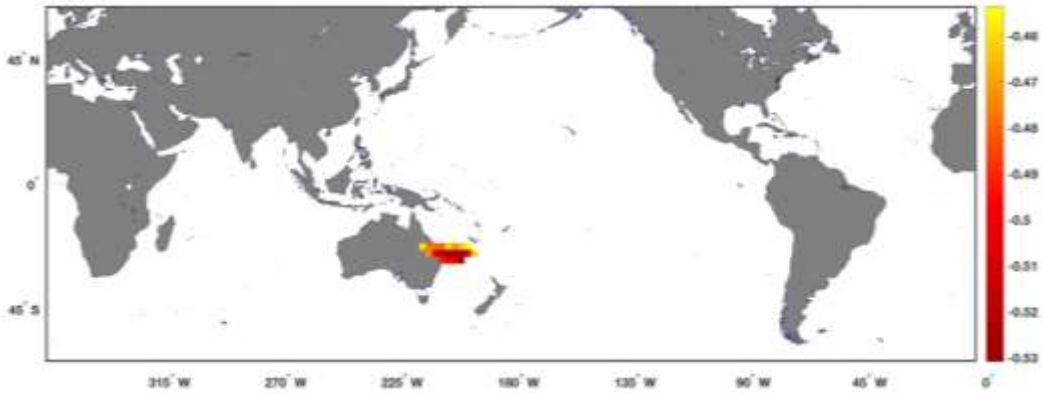


Figure S110| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 4 months lead-time.

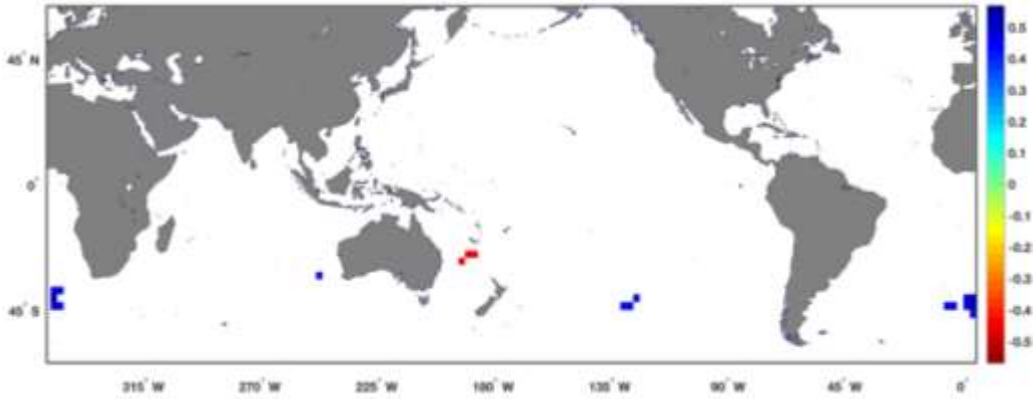


Figure S111| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 5 months lead-time.

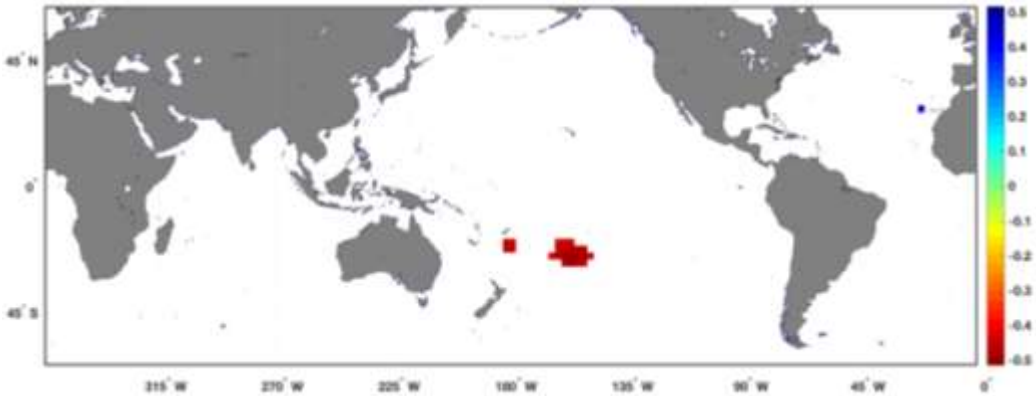


Figure S112| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 6 months lead-time.

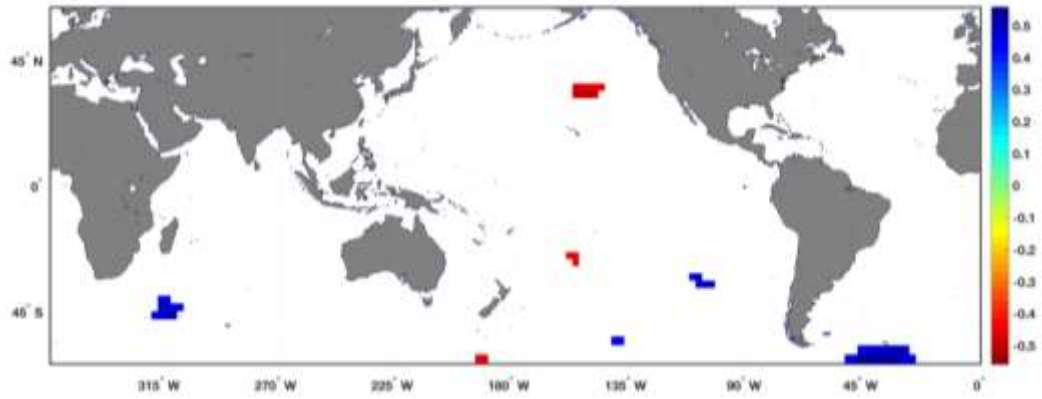



Figure S113| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 7 months lead-time.

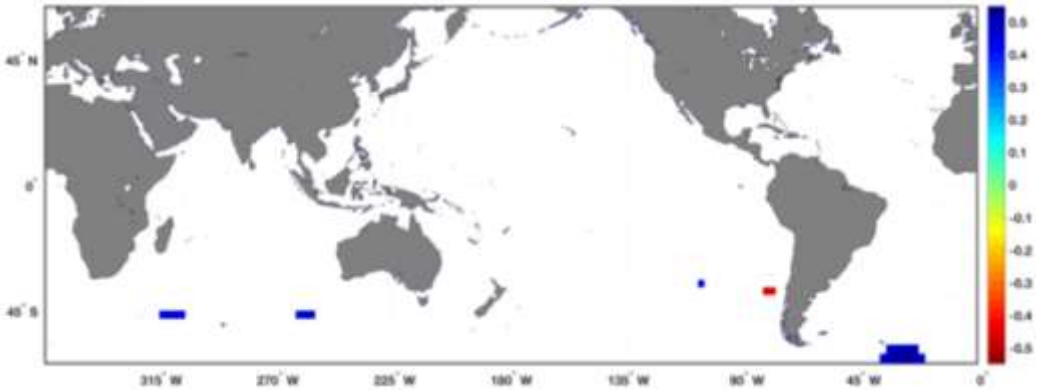



Figure S114| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 8 months lead-time.

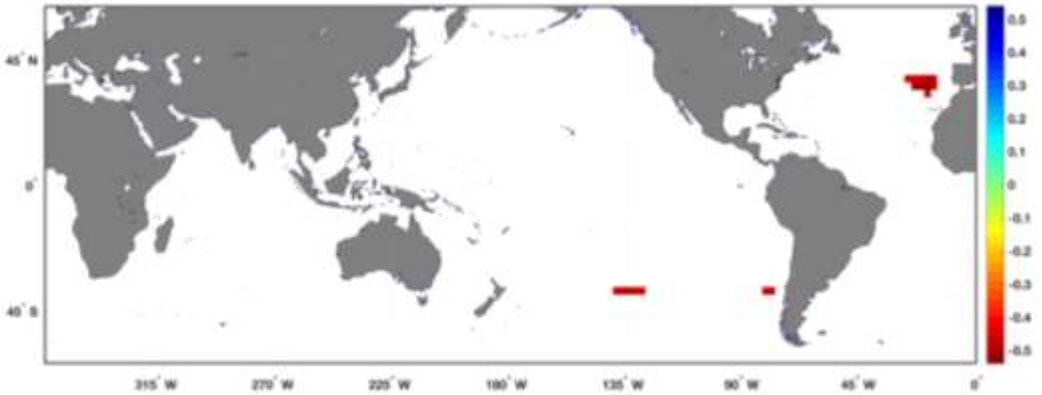



Figure S115| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 9 months lead-time.


Figure S116| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 10 months lead-time.

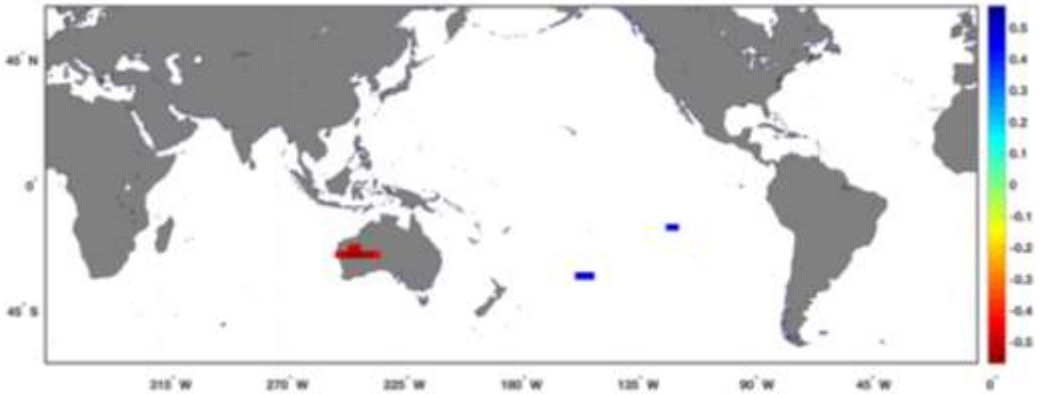



Figure S117| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 11 months lead-time.

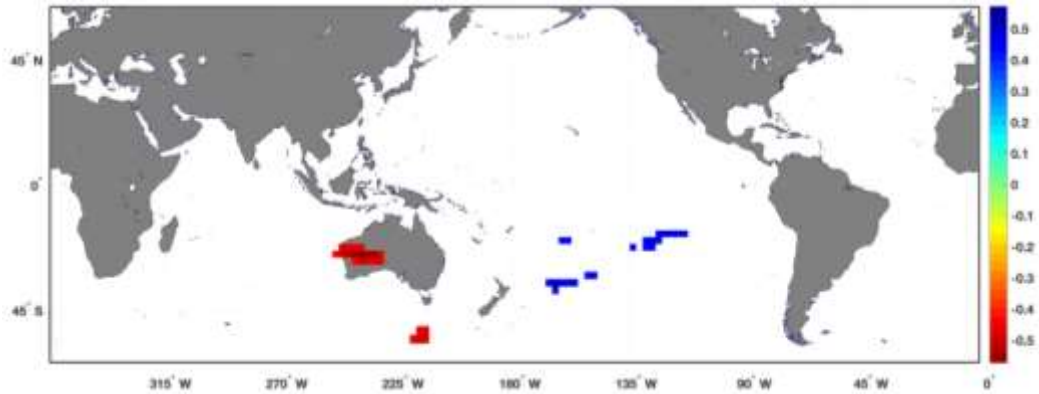



Figure S118| Correlation between tropopause level pressure at different locations and California wet-season precipitation with 12 months lead-time.

# Chapter 4

# Multivariate Indices for Climate Extremes: Heat Wave Intensity, Duration, Frequency Curve

**Atmospheric warming is projected to increase the frequency and severity of heat wave events. Heat waves can be quantified using multiple descriptors, including intensity, duration, frequency. However, most studies investigate one feature at a time using univariate analyses and consequentially lack a comprehensive measure to compare all heat wave characteristics simultaneously. I propose a multivariate approach to construct heat wave intensity, duration, frequency (HIDF) curves, which enables the concurrent analysis of all heat wave properties. As heat wave features (e.g., duration and intensity) are interdependent, ignoring the relationship between heat wave features can lead to substantial biases in frequency (risk) analyses. Here I show how HIDF curves can be used to describe the occurrence probability of heat waves with different duration and severity. I then show how HIDF curves can be used for attribution analysis of heat waves to anthropogenic warming through comparing historical climate model simulations (including anthropogenic emissions) with natural-only historical simulations. For example, in the city of Los Angeles, the HIDF shows that the likelihood of a four-day heat wave (temperature > 31 °C) has increased by 21% because of anthropogenic emissions. I also show that a six-day heat wave (temperature > 31 °C) has a 26% higher likelihood under the anthropogenic emission scenario relative to the natural-only scenario. The proposed approach can be applied in various locations to quantitatively describe the likelihood of heatwaves with different intensities and durations.**

## Introduction

Heat waves have significant negative implications on human health, urban air quality, ecological and environmental conditions, and agricultural and energy sectors [*Allen et al.*, 2010; *Jones et*

*al.*, 2018; *Santer et al.*, 2017; *Tebaldi and Lobell*, 2018; *Zampieri et al.*, 2017]. In addition, heatwaves have been connected to the increased risk of forest fires [*Koutsias et al.*, 2012]. Heatwaves are also considered to be one of the deadliest natural hazards, and cause high mortality rates in both developed and developing countries. For example, the 2003 European heat wave and 2010 Russian heat wave killed over 70,000 and 56,000 people, respectively [*Beniston*, 2004; *Grumm*, 2011; *Hauser et al.*, 2016; *Otto et al.*, 2012; *Rahmstorf and Coumou*, 2011]. The 2003 European heat wave also caused electricity demand to soar and energy efficiency to plummet [*Añel et al.*, 2017]. France, Europe's main electricity exporter, was forced to cut power exports by more than half during the heat wave, because power plants were operating at significantly reduced capacity [*Auffhammer et al.*, 2017; *De Bono et al.*, 2004]. Extreme temperatures and heatwave events have also caused problems in the transportation sector. Phoenix Sky Harbor Airport was forced to cancel nearly 50 flights due to extreme temperatures in the summer of 2017, when temperatures soared as high as 120 °F [*Magill,* 2017]. These cancellations produced a domino effect on the entire air transportation system, which demonstrate how impacts of regional heat waves can expand to a national or even global level.

Rising global temperatures are expected to increase the intensity, duration, and frequency of heat waves around the world [*Eltahir*, 2017; *Im et al.*, 2017; *Im et al.*, 2018; *G Meehl et al.*, 2018; *Perkins et al.*, 2012; *Shi et al.*, 2015], and are projected to further increase by the end of the century [*Alexander et al.*, 2006; *Perkins and Alexander*, 2013; *Sun et al.*, 2016]. Most studies investigate different features of heat waves independently, and ignore their relationships [*Kim et al.*, 2018; *Tebaldi and Wehner*, 2018; *Wehner et al.*, 2018a; *Wehner et al.*, 2018b]. Although

there is no universal definition, heat waves are typically described as a consecutive period of hot days with temperatures above a given threshold *[Mazdiyasni and AghaKouchak,* 2015*; Mazdiyasni et al.,* 2017*]*. The threshold is often based on a percentile of each month's daily temperatures or a fixed value [*Huynen et al.*, 2001; *G A Meehl and Tebaldi*, 2004]. Current metrics evaluate individual heat wave characteristics, such as the hottest day of each year or longest duration of consecutive hot days [*Mazdiyasni and AghaKouchak*, 2015; *Mazdiyasni et al.*, 2017; *G A Meehl and Tebaldi*, 2004; *Perkins and Alexander*, 2013; *Perkins et al.*, 2012; *Sun et al.*, 2017]. However, current univariate indicators often underestimate the impacts of heat waves because they fail to characterize the extreme event in a comprehensive manner [*Sun et al.*, 2016]. The risks from individual heat wave characteristics can be amplified when considered concurrently (e.g. high intensity and long duration heat wave events vs high intensity, short duration heat wave events). The significant impacts and increasing intensity and duration of extreme heat wave events highlight the need for a comprehensive metric, accounting for all heat wave characteristics simultaneously.

In this paper, I propose a heat wave intensity-duration-frequency (HIDF) model. I use multivariate copula functions to link heat wave durations and intensities. The use of these functions allows for the presentation of heat wave frequency information with different combinations of intensity and severity. Copulas have been used for linking different features of drought and precipitation extremes such as duration and severity [*Kao and Govindaraju*, 2010; *J-T Shiau and Modarres*, 2009; *J Shiau*, 2006; *Singh and Zhang*, 2007]. In this paper, I consider heat wave intensity as the average of mean daily temperature throughout the duration of heat wave. My results portray and compare HIDF curves for six cities in the United States, using daily

mean temperature data from 1979 – 2016. I also compare HIDF curves for historical model simulations (including anthropogenic emissions) against natural-only historical (i.e. pre-industrial) model simulations to investigate the impacts of anthropogenic emissions on extreme heat events, using daily mean temperature data from 1850 - 2005. Using this novel approach, I concurrently compare the differences in heat wave intensity, duration, and frequency for a more comprehensive analysis of anthropogenic climate change impacts on heat waves.

**Methods**

We create the HIDF curves by determining the non-exceedance probability of heat wave duration and intensity. I define heat wave intensity as the average daily temperature throughout the duration of the event. Similar to traditional precipitation IDF curves, I use the block maxima method to produce the HIDF curves. I determine the hottest heat wave events in each year with durations ranging from one to ten consecutive days. I then calculate the intensity of each event using,

$$\forall y \quad I_y = \max\left\{\frac{\sum_i^{i+D-1} t_i^y}{D}\right\}, \qquad (1)$$

$$\forall D \quad D = 1{:}\,10,$$

where *y* is the year, *i* is the first day in the moving calculating window that ranges between one and the number of days in the year, $t_i^y$ is the average daily temperature at day *i* of year *y*, and *D* is the length of the running window (i.e. heat wave duration).

To model the HIDF curves, I determine the average temperature of the one to ten hottest consecutive days in each year. I then use the multivariate copula functions to find the non-

exceedance joint probability cumulative distribution function of heat wave duration and heat wave intensity [*Gräler et al.*, 2013; *Madadgar et al.*, 2016; *Nelsen*, 2007; *Salvadori and De Michele*, 2007; *Salvadori et al.*, 2011; *Salvadori et al.*, 2013]. I determine the best fitting copula family out of the 26 families of the Multivariate Copula Analysis Toolbox [*Sadegh et al.*, 2017] based on the Bayesian Information Criterion. Refer to Table 1 of [*Sadegh et al.*, 2017] for more information regarding the copula families.

A copula function is defined as the multivariate distribution function

$$\mathrm{F_{X_1 \dots X_n}}(\mathrm{x_1}, \dots, \mathrm{x_i}, \dots, \mathrm{x_n}) = \mathrm{C}\left[\mathrm{F_{X_1}}(\mathrm{x_1}), \dots, \mathrm{F_{X_i}}(\mathrm{x_i}), \dots, \mathrm{F_{X_n}}(\mathrm{x_n})\right] = \mathrm{C}(\mathrm{U_1}, \dots, \mathrm{U_i}, \dots, \mathrm{U_n}) \quad (2)$$

where C is the cumulative distribution function (CDF) of the copula and $\mathrm{F_{X_i}}(\mathrm{x_i})$ is the non-exceedance probability of $\mathrm{x_i}$ (marginal distribution). For the study, I use the bivariate copula to estimate the joint probability distribution of heat wave duration ($X$), and heat wave intensity ($Y$),

$$\mathrm{F_{XY}}(\mathrm{x}, \mathrm{y}) = \mathrm{C}[\mathrm{F_X}(\mathrm{x}), \mathrm{F_Y}(\mathrm{y})] \quad (3)$$

We then calculate the joint return periods for different duration and frequency following Salvadori et al. (2013). I also determine the conditional probability of intensity exceeding a threshold (Y>y) at a certain duration (X=x), that is $\mathrm{f_{Y|X}}(\mathrm{Y} > \mathrm{y} \mid \mathrm{X})$ through the conditional marginal PDF [*Madadgar and Moradkhani*, 2013]:

$$\mathrm{f_{Y|X}}(\mathrm{y} \mid \mathrm{x}) = \mathrm{c}[\mathrm{F_X}(\mathrm{x}), \mathrm{F_Y}(\mathrm{y})] \,.\, \mathrm{f_Y}(\mathrm{y}) \quad (4)$$

where c is the copula probability density function (PDF) and $\mathrm{f_Y}(\mathrm{y})$ is the heat wave intensity marginal distribution. Once I choose a conditional PDF from Eq. 3, I calculate the probability of intensity (Y) exceeding a particular threshold (y) from the area under the curve, delineated by:

$f_{Y|X}(Y > y \mid x)$. I apply this technique to calculate $f_{Y|X}(y \mid x)$ for different values of x (e.g., duration=5 days).

**Data**

The proposed methodology is generalized and can be applied to different locations. Here, I used daily average temperatures for Atlanta (33.7490° N, 84.3880° W), Chicago (41.8781° N, 87.6298° W), Denver (39.7392° N, 104.9903° W), Houston (29.7604° N, 95.3698° W), Los Angeles (34.0522° N, 118.2437° W), and Phoenix (33.4484° N, 112.0740° W) from the Climate Prediction Center (CPC) global air temperature dataset provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/). This dataset includes near surface air temperature with a 0.5 degree spatial resolution and daily temporal resolution. I interpret observed temperature data from the grid encompassing each city to be representative of that particular city.

For attribution analysis, I use the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical and natural-only historical simulations from 1850-2005 to quantify the impact of anthropogenic climate change on daily average temperature values [*Taylor et al.*, 2012]. CMIP5 is an ensemble of climate model experiments intended to improve my understanding of my pre-industrial, historical, and projected climate [*Taylor et al.,* 2009/2011]; the spatial resolution of the models used in my study are listed in Table S1. The historical experiment imposes conditions – such as anthropogenic and natural trends and variability – that reflect what has been seen in the observations, including changes in atmosphere due to human and volcanic emissions, solar forcing, aerosols, and human land use [*Taylor et al.,* 2009/2011]. The natural-

only historical simulations capture natural trends and variability without anthropogenic forcing [*Taylor et al.*, 2012]. With the climate simulations, I could attribute differences between the two simulations to anthropogenic climate change. I show an example application for the city of Los Angeles, California (34.0522, -118.2437).

**Results**

Figure 1 shows the heat wave IDF curves for Atlanta, Chicago, Denver, Houston, Los Angeles, and Phoenix. Each subplot depicts the joint probabilities for different combinations of heat wave duration and heat wave intensity. In addition to providing heat wave risk information, HIDF curves can be used for comparing the risk of heat waves in different locations. For example, I observe that a six-day heat wave with an average temperature of 38 °C in Chicago corresponds to a 2-year event (probability = 0.5), while a similar six-day heat wave with a 2-year return period in Phoenix would have a much higher intensity of 47.5 °C. Similar comparisons can be made based on the durations or frequency of events in different regions. Figure 1 demonstrates the flexibility of the proposed HIDF curves for describing the probability of occurrence of different combinations of heat wave duration and intensity. The figure also shows different combinations of heat wave duration and intensity that lead to the same return period. For example, Figure 1 shows that an 7-day heat wave with an intensity of 47 °C is equally likely as a 10-day heat wave with an intensity of 45 °C (here, both are 2.5 year events), in Phoenix, AZ.

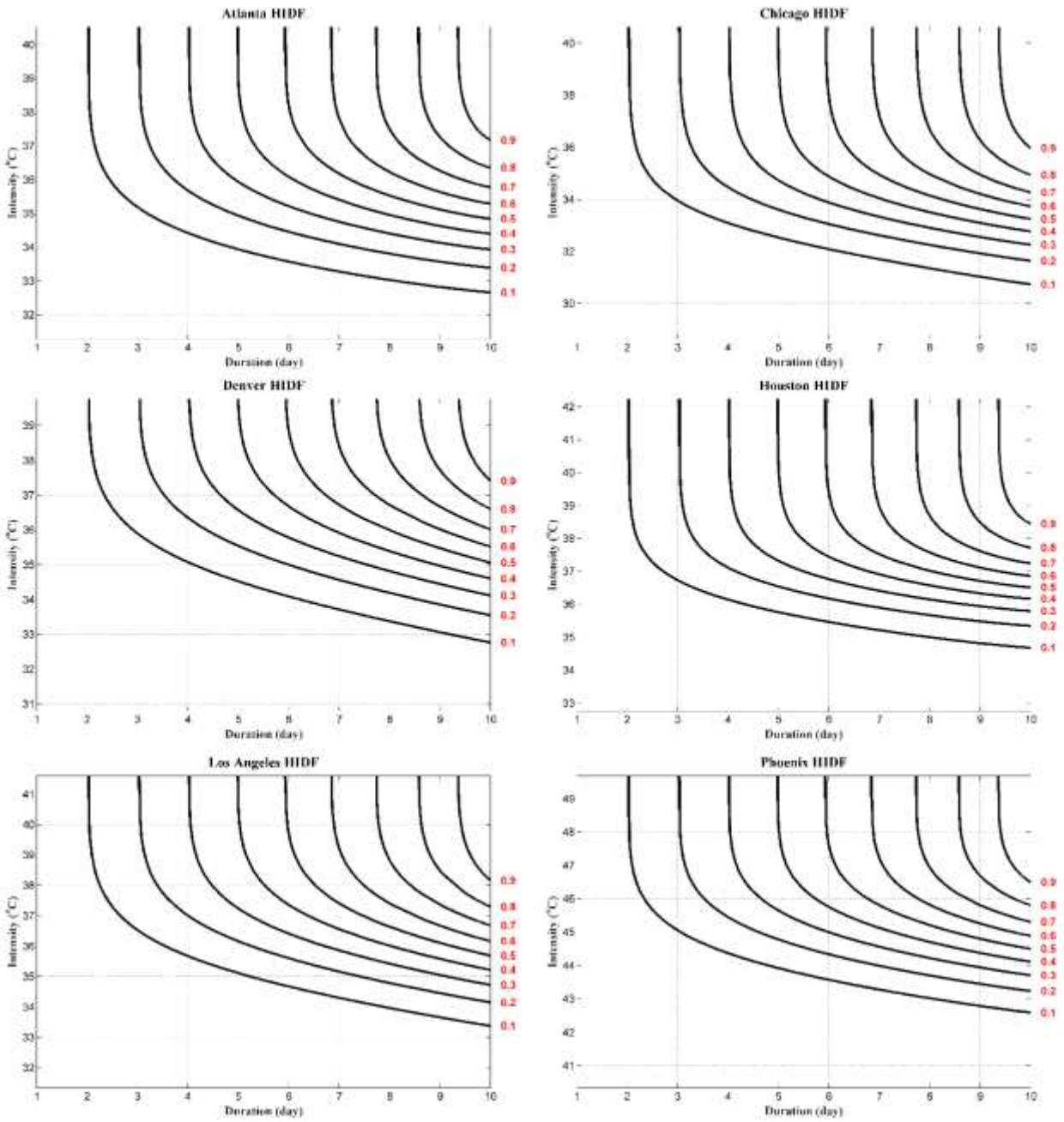



**Figure 1 | Heat wave intensity, duration, frequency (HIDF) curves** for six major cities across the United States (Atlanta, Chicago, Denver, Houston, Los Angeles, and Phoneix.

Figure 2 compares HIDF curves in Los Angeles, CA using the mean of the historical (black) vs historical natural-only (red) CMIP5 simulations. Fig. 2 shows differences in the joint probability of heat wave duration and intensity between historical and natural-only historical simulations.

This figure demonstrates that heat wave events are generally shorter and less intense under natural-only forcings (without anthropogenic forcing) in relation to historical conditions. In other words, a heat wave with the same intensity and duration under natural-only historical conditions has a lower frequency (and probability of occurrence) than a heat wave occurring under historical conditions perturbed by anthropogenic emissions. For example, this figure shows that an extreme ten-year, ten-day heat wave event would be classified as an intensity of 31.8 °C under natural-only conditions, while the event would be classified as an intensity of 32.1 °C under historical conditions. This average difference of 0.3 °C over a period of ten days can have significant implications on human health, agriculture, the environment, and the electric grid [*Mazdiyasni et al.*, 2017]. Figure S1 portrays the HIDF curves of the individual models.

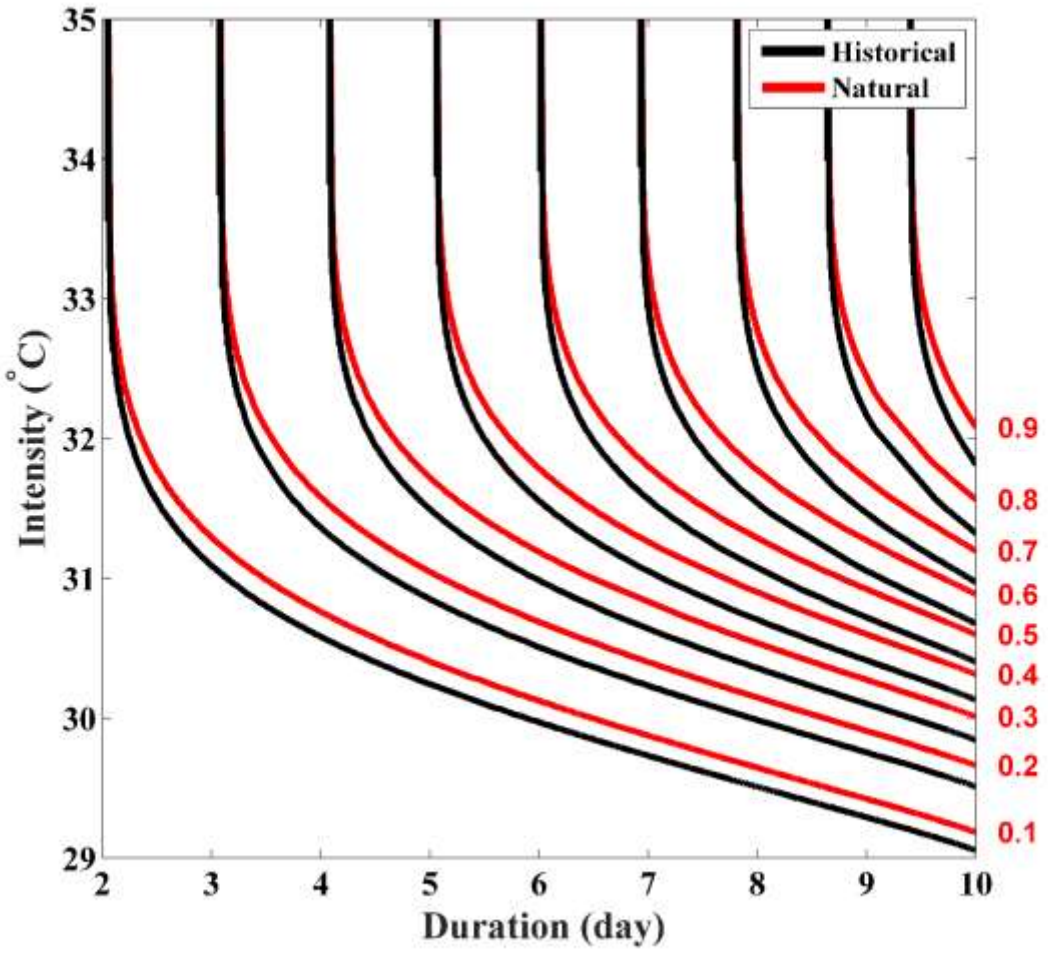



**Figure 2 | Mean heat wave intensity-duration-frequency (HIDF) curves** for historical (including anthropogenic forcings) vs natural-only historical simulations from selected CMIP5 models.

Figure 3 displays the difference in heat wave intensity given heat wave duration between historical vs natural-only historical conditions, using the mean of the four models. Figure S2 shows the conditional probability of each individual model. I show that the probability of heat wave intensity being greater than 30 °C given a duration of four days is six percent greater under historical conditions in comparison to historical natural-only conditions (81% vs 75%). I also show that the probability of intensity being greater than 31 °C given a heat wave duration

of four days is eight percent greater under historical conditions vs historical natural-only conditions (41% vs 33%, respectively). Figure 3 displays that an 8% increase in the likelihood of a four-day heat wave (temperature > 30 °C) and a 24% increase in a four-day, more intense heat wave (temperature > 31 °C) can be attributed to anthropogenic warming. Figure S3 shows similar results for six-day heat waves; anthropogenic warming has increased the likelihood of a six-day heat wave (temperature > 30 °C) by 10%, and a six-day heat wave (temperature > 31 °C) by 29%. From Figure 3 and S3, I also observed greater increases in the likelihood of longer and more intense heat waves in the historical simulations relative to the historical natural-only. Therefore, I can conclude that greater increases in the likelihood of extreme (in intensity and duration) heat wave events have been driven by anthropogenic warming.

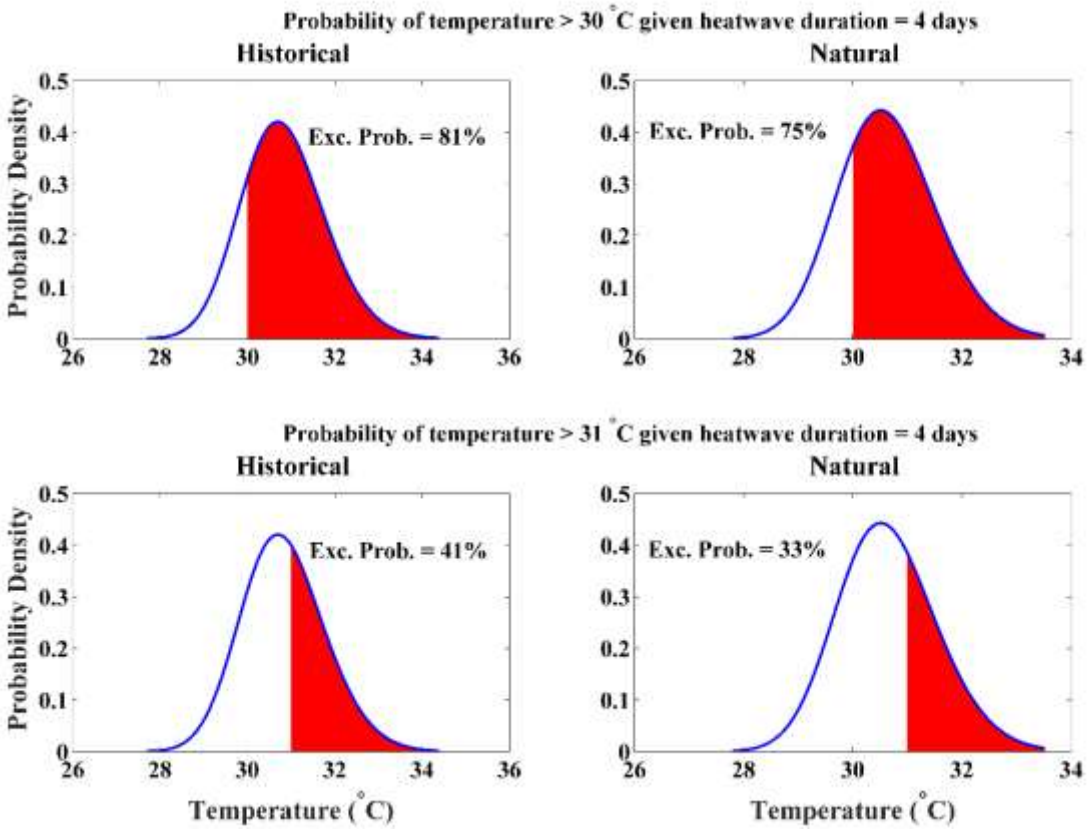



**Figure 3 | Comparison between historical (including anthropogenic forcings) vs natural-only historical parametric conditional probability density functions (PDFs)** using the mean of CMIP5 simulations for heat wave intensity given heat wave duration equal to four days.

## Discussion and Conclusion

Global warming is causing an increase in the frequency and severity of heat wave events, which increases the importance of using a robust model in understanding heat waves and quantifying heat wave properties. I propose a multivariate approach to construct heat wave intensity, duration, frequency (HIDF) curves, which allows for the frequency (risk) analysis of extreme heat events, while concurrently accounting for both the intensity and duration of heat waves. Since these heat wave features are interdependent, it is important to model the relationship in a manner that avoids biases in the frequency analyses. The main objective of this paper is to present the HIDF methodology and show different types of applications including describing

heat waves features and attribution analysis. An attribution analysis using the proposed HIDF curves shows that the anthropogenic emissions have increased the likelihood of a four-day heat wave (temperature > 31 °C) by 21%. I further show a six-day heat wave (temperature > 31 °C) has a 26% higher likelihood under the anthropogenic emission scenario relative to the natural-only scenario. The proposed method is general and can be used for similar analysis in different locations and different combinations for heat wave durations and intensities.

The HIDF curve can potentially be used by for design of infrastructure systems such as electric grids and power plants. An electric grid, for example, is designed for peak demand and it is important to consider both duration and intensity of heat wave events simultaneously; similar to the use of precipitation IDF curves for highway culvert design. Although analyzing heat waves with univariate indices have provided useful information in the past, my proposed metric can change how I view the extremeness of a heat wave event moving forward.

## Appendix 4

Table S1: List of climate models, their spatial resolutions, and Modeling Centers.

| Modeling Center | Institute ID | Model Name | Spatial resolution (lat, lon) |
|---|---|---|---|
| National Center for Atmospheric Research | NCAR | CCSM4 | 0.9424, 1.25 |
| Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence | CSIRO-QCCCE | CSIRO-Mk3.6.0 | 1.8653, 1.875 |
| Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies | MIROC | MIROC-ESM | 2.7906, 2.8125 |
| Norwegian Climate Centre | NCC | NorESM1-M | 1.8947, 2.5 |

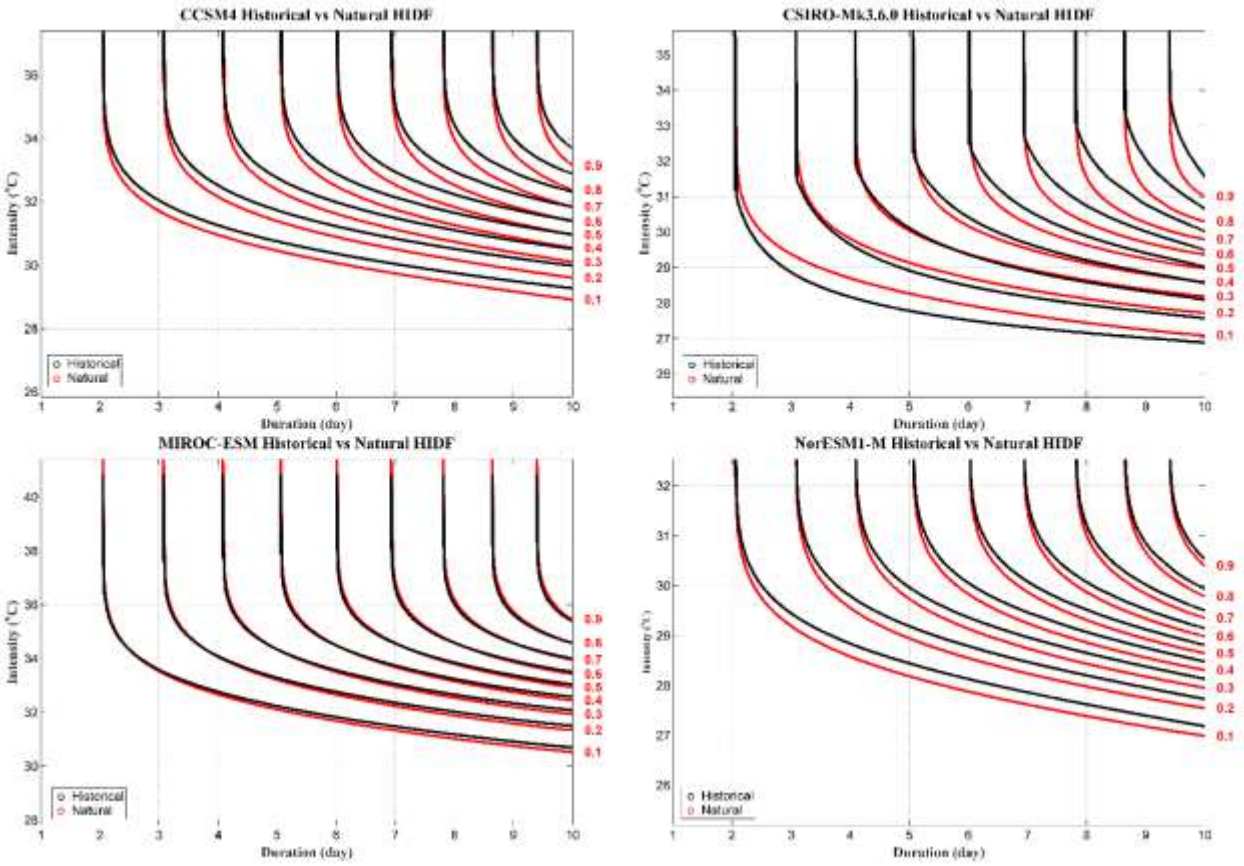



**Figure S1: Heat wave intensity-duration-frequency (HIDF) curves for historical (including anthropogenic forcings) vs natural-only historical simulations from selected CMIP5 models.**

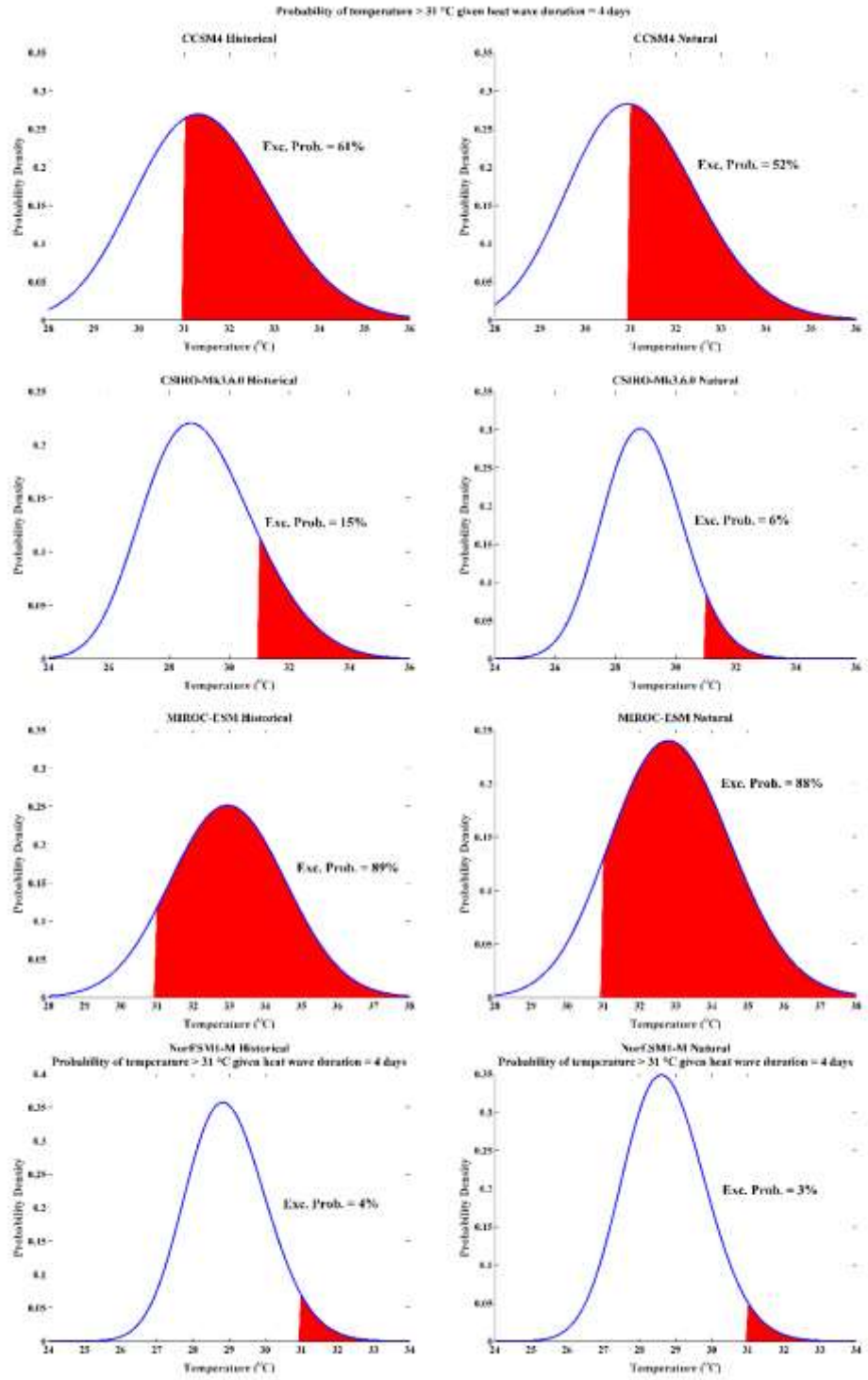



**Figure S2: Comparison between historical (including anthropogenic forcings) vs natural-only historical parametric conditional probability density functions (PDFs) for heat wave intensity given heat wave duration equal to four days.**

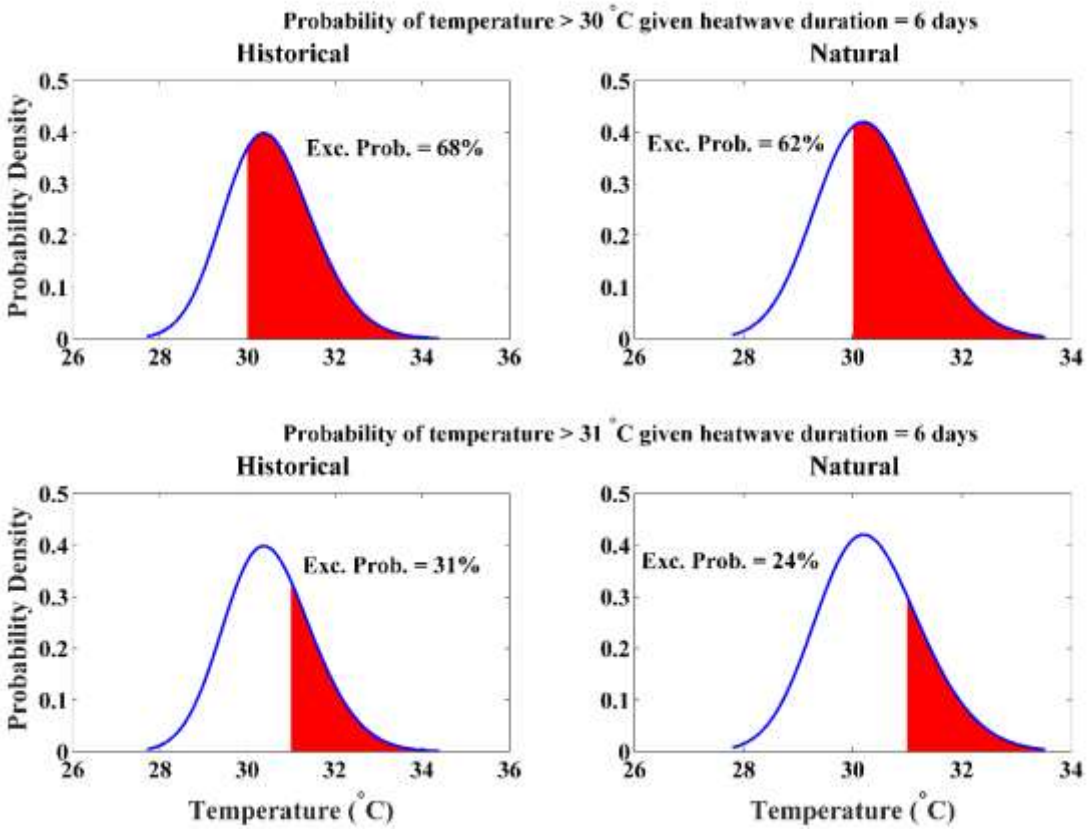



**Figure S3: Comparison between historical (including anthropogenic forcings) vs natural-only historical parametric conditional probability density functions (PDFs) using the mean of CMIP5 simulations for heat wave intensity given heat wave duration equal to six days.**

## Summary and Conclusion

Climate change and variability are likely to affect physical and hydrometeorological conditions and to interact with and possibly exacerbate ongoing environmental change. Numerous studies have demonstrated that climate and weather extremes are sensitive to climate change and thus, climate assessment activities are fundamental for understanding climate change and its impacts. Thus far, a myriad of indices have been developed mainly to investigate trends and changes to individual climate variables.

Current extremes detection indices can be categorized into three types: (a) indices based on daily, monthly, and/or annual maxima or minima of a certain climate variable (i.e. temperature or precipitation amounts); (b) indices based on duration of an extreme condition (i.e. number of consecutive days in a year above a relative extreme threshold, such as the long-term $90^{th}$ percentile); and (c) indices based on duration and/or frequency of an absolute extreme exceedance threshold (i.e. annual number of frost days (minimum temperature below 0 °C) or number of consecutive frost days in a year). While all three types of indicators are important and informative, they do not provide any information on compound and inter-related (concurrent, simultaneous or joint) extremes such as droughts and heatwaves. As such, one cannot investigate historical and projected changes in compound extremes using current climate indicators. The main reason is that current indicators are univariate and do not account for interdependence between multiple climate variables (e.g., temperature and precipitation). The combination or sequence of climate extreme events may cause a significant impact on the ecosystem and society though individual events involved may not be severe themselves.

The notion of compound extremes (also referred to as simultaneous extremes, or concurrent extremes) as defined by the IPCC report is [*Field et al.,* 2012]: "(1) two or more extreme events occurring simultaneously or successively; (2) combinations of extreme events with underlying conditions that amplify the impact of the events; (3) combinations of events that are not themselves extremes but lead to an extreme event or impact when combined." Extreme events that contribute to compound extremes can be of the same type (clustered multiple events such as simultaneous flooding over a large spatial scale) or different types such as droughts and heatwaves. Understanding the risk these compound extremes present and estimating their occurrence in the future requires rigorous methods due to the small number of samples of extremes. This dissertation addresses these research gaps and outlines methods and indicators to investigate and model compound extremes.

One method to analyze changes in concurrent climatic events is the empirical counting method. The first chapter of this dissertation focuses on the changes in concurrences of heatwaves and meteorological droughts from 1960 – 2010. Despite an apparent hiatus in rising temperature and no significant trend in droughts, the results show a substantial increase in concurrent droughts and heatwaves across most parts of the United States, and a statistically significant shift in the distribution of concurrent extremes. While commonly used trend analysis methods do not show any trend in concurrent droughts and heatwaves, a unique statistical approach discussed in this study exhibits a statistically significant change in the distribution of the data.

The analysis in Chapter 2a investigates the change in the probability of an extreme event (i.e. mortality) under the condition of another extreme event (heat waves) using a parametric apporach. In this chapter, I show the changes in summer temperatures, the frequency, severity

and duration of heatwaves, and heat-related mortality in India between 1960 and 2009, using data from the India Meteorological Department. Mean temperatures across India have risen by more than 0.5 °C over this period, with statistically significant increases in heatwaves. Using a novel probabilistic model, I show that the increase in summer mean temperatures in India over this period corresponds to a 146% increase in the probability of heat-related mortality events of more than 100 people. In turn, my results suggest that future climate warming will lead to substantial increases in heat-related mortality, particularly in developing, low-latitude countries such as India where heatwaves will become more frequent and populations are especially vulnerable to these extreme temperatures. My findings indicate that even moderate increases in mean temperatures may cause great increases in heat-related mortality, and support efforts of governments and international organizations to build-up the resilience of these vulnerable regions to more and more severe heatwaves.

Chapter 2b of this dissertation portrays the probability of a climatic event (precipitation) given certain climatic condition (pressure) using an empirical approach. This chapter shows impacts of tropopause level pressure in the Northeastern Pacific Ocean on California precipitation using an empirical conditional probability methodology. Sea surface temperatures and teleconnection patterns such as El Nino/La Nina are considered the main culprits behind major California droughts. However, the underlying relationship between sea surface temperatures (SSTs) and precipitation anomalies is relatively weak. Here I show that tropopause level pressure in a region in the northeastern Pacific Ocean (dubbed the PARS-NEP region) plays a major role in whether California will experience a wet or dry year and often dominates the role of SST-based teleconnections. My results indicate that pressure in the PARS-NEP region

Granger-Causes precipitation in California during the wet season. I show that when pressure in the PARS-NEP region is in the lower (upper) tertile, 85% of wet seasons across California have a positive (negative) precipitation anomaly. The observed relationship between PARS-NEP and California precipitation is stronger than all the commonly used SST-based climatic indictors frequently used for understanding causes of droughts.

Chapter 3 demonstrates how multiple co-occurring drivers (i.e. sea surface temperatures) have a cascading impact on a climatic phenomenon (i.e. precipitation). I use an exhaustive, bottom-up, brute-force search algorithm for discovering predictive relationships between specific ocean-land regions. Based on 32 years of reanalysis data (1983-2015), I discover linkages that are distinct from known teleconnections, but which better predict precipitation with 4-12 months lead-time. I further evaluate my empirical approach by reversing the analysis, demonstrating that the predictive skill I observe is not mere coincidence; the results have strong spatial coherence. The teleconnections revealed here may have immediate practical value in anticipating seasonal weather events, and also help to focus future research on the physical mechanisms underlying the planet's strongest and most important teleconnections.

Finally, I use a multivariate approach to analyze the return period (frequency or risk) of a climatic event with multiple dimensions (i.e. heat waves), simultaneously. Heat waves can be quantified using multiple descriptors, including intensity, duration, frequency. However, most studies investigate one feature at a time using univariate analyses and consequentially lack a comprehensive measure to compare all heat wave characteristics simultaneously. I propose a multivariate approach to construct heat wave intensity, duration, frequency (HIDF) curves, which enables the concurrent analysis of all heat wave properties. As heat wave features (e.g.,

duration and intensity) are interdependent, ignoring the relationship between heat wave features can lead to substantial biases in frequency (risk) analyses. Here I show how HIDF curves can be used to describe the occurrence probability of heat waves with different duration and severity. I then show how HIDF curves can be used for attribution analysis of heat waves to anthropogenic warming through comparing historical climate model simulations (including anthropogenic emissions) with natural-only historical simulations. For example, in the city of Los Angeles, the HIDF shows that the likelihood of a four-day heat wave (temperature > 31 °C) has increased by 21% because of anthropogenic emissions. I also show that a six-day heat wave (temperature > 31 °C) has a 26% higher likelihood under the anthropogenic emission scenario relative to the natural-only scenario. The proposed approach can be applied in various locations to quantitatively describe the likelihood of heatwaves with different intensities and durations.

References

AghaKouchak, A., L. Cheng, O. Mazdiyasni, and A. Farahmand (2014), Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought, *Geophysical Research Letters*, *41*(24), 8847-8852.

Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. Klein Tank, M. Haylock, D. Collins, B. Trewin, and F. Rahimzadeh (2006), Global observed changes in daily climate extremes of temperature and precipitation, *Journal of Geophysical Research: Atmospheres*, *111*(D5).

Allen, C. D., A. K. Macalady, H. Chenchouni, D. Bachelet, N. McDowell, M. Vennetier, T. Kitzberger, A. Rigling, D. D. Breshears, and E. T. Hogg (2010), A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *Forest ecology and management*, *259*(4), 660-684.

Anderson, T. W. (1962), On the distribution of the two-sample Cramer-von Mises criterion, *The Annals of Mathematical Statistics*, 1148-1159.

Anderson, T. W., and D. A. Darling (1952), Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes, *The annals of mathematical statistics*, 193-212.

Añel, J. A., M. Fernández-González, X. Labandeira, X. López-Otero, and L. de la Torre (2017), Impact of Cold Waves and Heat Waves on the Energy Production Sector, *Atmosphere*, *8*(11), 209.

Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat (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.

Auffhammer, M., P. Baylis, and C. H. Hausman (2017), Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States, *Proceedings of the National Academy of Sciences*, 201613193.

Basu, R. (2009), High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008, *Environmental Health*, *8*(1), 1.

Beniston, M. (2004), The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations, *Geophysical Research Letters*, *31*(2).

Bouchama, A., and J. P. Knochel (2002), Heat stroke, *New England Journal of Medicine*, *346*(25), 1978-1988.

Bradley, R., H. Diaz, G. Kiladis, and J. Eischeid (1987), ENSO signal in continental temperature and precipitation records, *Nature*, *327*(6122), 497-501.

Bücher, A., and I. Kojadinovic (2016), A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing, *Bernoulli*, *22*(2), 927-968.

Bücher, A., I. Kojadinovic, T. Rohmer, and J. Segers (2014), Detecting changes in cross-sectional dependence in multivariate time series, *Journal of Multivariate Analysis*, *132*, 111-128.

Cayan, D. R., and J. O. Roads (1984), Local relationships between United States West Coast precipitation and monthly mean circulation parameters, *Monthly Weather Review*, *112*(6), 1276-1282.

Cayan, D. R., K. T. Redmond, and L. G. Riddle (1999), ENSO and hydrologic extremes in the western United States, *Journal of Climate*, *12*(9), 2881-2893.

Cayan, D. R., M. Tyree, M. D. Dettinger, H. León, G. Hugo, T. Das, E. P. Maurer, P. Bromirski, N. Graham, and R. Flick (2009), Climate change scenarios and sea level rise estimates for California 2008 Climate Change Scenarios Assessment.

Ciais, P., M. Reichstein, N. Viovy, A. Granier, J. Ogée, V. Allard, M. Aubinet, N. Buchmann, C. Bernhofer, and A. Carrara (2005), Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, *437*(7058), 529.

Clarke, R. T. (2010), On the (mis) use of statistical methods in hydro-climatological research, *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, *55*(2), 139-144.

Cramér, H. (1928), On the composition of elementary errors: First paper: Mathematical deductions, *Scandinavian Actuarial Journal*, *1928*(1), 13-74.

Damberg, L., and A. AghaKouchak (2014), Global trends and patterns of drought from space, *Theoretical and applied climatology*, *117*(3-4), 441-448.

De Bono, A., P. Peduzzi, S. Kluser, and G. Giuliani (2004), Impacts of summer 2003 heat wave in Europe.

DeFlorio, M. J., D. W. Pierce, D. R. Cayan, and A. J. Miller (2013), Western US extreme precipitation events and their relation to ENSO and PDO in CCSM4, *Journal of Climate*, *26*(12), 4231-4243.

Dholakia, H. H., V. Mishra, and A. Garg (2015), Predicted Increases in Heat related Mortality under Climate Change in Urban India*Rep.*, Indian Institute of Management Ahmedabad, Research and Publication Department.

Diffenbaugh, N. S., and M. Scherer (2011), Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries, *Climatic Change*, *107*(3-4), 615-624.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns (2000), Climate extremes: observations, modeling, and impacts, *science*, *289*(5487), 2068-2074.

Eltahir, E. (2017), Future Heat Waves In Asia, paper presented at AGU Fall Meeting Abstracts.

Gombay, E., and L. Horváth (1999), Change-points and bootstrap, *Environmetrics*, *10*(6), 725-736.

Gräler, B., M. van den Berg, S. Vandenberghe, A. Petroselli, S. Grimaldi, B. De Baets, and N. Verhoest (2013), Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation, *Hydrology and Earth System Sciences*, *17*(4), 1281-1296.

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

Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne (2014), Global assessment of trends in wetting and drying over land, *Nature geoscience*, *7*(10), 716.

Grimaldi, S., A. Petroselli, G. Salvadori, and C. De Michele (2016), Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses, *Advances in Water Resources*, *90*, 8.

Grumm, R. H. (2011), The central European and Russian heat event of July–August 2010, *Bulletin of the American Meteorological Society*, *92*(10), 1285-1296.

Hajat, S., R. S. Kovats, and K. Lachowycz (2007), Heat-related and cold-related deaths in England and Wales: who is at risk?, *Occupational and environmental medicine*, *64*(2), 93-100.

Hamoudi, A. A., and J. D. Sachs (1999), Economic consequences of health status: a review of the evidence*Rep.*, Center for International Development at Harvard University.

Hansen, J., M. Sato, and R. Ruedy (2012), Perception of climate change, *Proceedings of the National Academy of Sciences*, *109*(37), E2415-E2423.

Hao, Z., and A. AghaKouchak (2013), Multivariate standardized drought index: a parametric multi-index model, *Advances in Water Resources*, *57*, 12-18.

Hao, Z., A. AghaKouchak, and T. J. Phillips (2013), Changes in concurrent monthly precipitation and temperature extremes, *Environmental Research Letters*, *8*(3), 034014.

Hao, Z., A. AghaKouchak, N. Nakhjiri, and A. Farahmand (2014), Global integrated drought monitoring and prediction system, *Scientific Data*, *1*, 140001.

Harto, C., Y. Yan, Y. Demissie, D. Elcock, V. Tidwell, K. Hallett, J. Macknick, M. Wigmosta, and T. Tesfa (2012), Analysis of drought impacts on electricity production in the Western and Texas interconnections of the United States*Rep.*, Argonne National Laboratory (ANL).

Hauser, M., R. Orth, and S. I. Seneviratne (2016), Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia, *Geophysical Research Letters*, *43*(6), 2819-2826.

Hawkins, D. M., and Q. Deng (2010), A nonparametric change-point control chart, *Journal of Quality Technology*, *42*(2), 165-173.

Hegerl, G. C., H. Hanlon, and C. Beierkuhnlein (2011), Climate science: Elusive extremes, *Nature Geoscience*, *4*(3), 142.

Held, I. M. (2013), Climate science: The cause of the pause, *Nature*, *501*(7467), 318.

Hoerling, M., and A. Kumar (2003), The perfect ocean for drought, *Science*, *299*(5607), 691-694.

Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager (2014), Causes and predictability of the 2012 Great Plains drought, *Bulletin of the American Meteorological Society*, *95*(2), 269-282.

Hoerling, M., A. Kumar, R. Dole, J. W. Nielsen-Gammon, J. Eischeid, J. Perlwitz, X.-W. Quan, T. Zhang, P. Pegion, and M. Chen (2013), Anatomy of an extreme event, *Journal of Climate*, *26*(9), 2811-2832.

Hoerling, M. P., A. Kumar, and M. Zhong (1997), El Niño, La Niña, and the nonlinearity of their teleconnections, *Journal of Climate*, *10*(8), 1769-1786.

Holmes, M., I. Kojadinovic, and J.-F. Quessy (2013), Nonparametric tests for change-point detection à la Gombay and Horváth, *Journal of Multivariate Analysis*, *115*, 16-32.

Huynen, M.-M., P. Martens, D. Schram, M. P. Weijenberg, and A. E. Kunst (2001), The impact of heat waves and cold spells on mortality rates in the Dutch population, *Environmental health perspectives*, *109*(5), 463.

Im, E.-S., J. S. Pal, and E. A. Eltahir (2017), Deadly heat waves projected in the densely populated agricultural regions of South Asia, *Science advances*, *3*(8), e1603322.

Im, E.-S., S. Kang, and E. A. Eltahir (2018), Projections of rising heat stress over the western Maritime Continent from dynamically downscaled climate simulations, *Global and Planetary Change*.

IMD (2009), Disastrous Weather Event, *IMD Annual Publications*.

Jones, B., C. Tebaldi, B. C. O'Neill, K. Oleson, and J. Gao (2018), Avoiding population exposure to heat-related extremes: demographic change vs climate change, *Climatic Change*, *146*(3-4), 423-437.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, and J. Woollen (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American meteorological Society*, *77*(3), 437-471.

Kam, J., J. Sheffield, X. Yuan, and E. F. Wood (2014), Did a skillful prediction of sea surface temperatures help or hinder forecasting of the 2012 Midwestern US drought?, *Environmental Research Letters*, *9*(3), 034005.

Kao, S.-C., and R. S. Govindaraju (2010), A copula-based joint deficit index for droughts, *Journal of Hydrology*, *380*(1-2), 121-134.

Karl, T. R., A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, T. C. Peterson, R. S. Vose, and H.-M. Zhang (2015), Possible artifacts of data biases in the recent global surface warming hiatus, *Science*, aaa5632.

Kendall, M. (1975), Rank Correlation Measures; Charles Griffin, *London, 202p*.

Kendall, M. G. (1948), Rank correlation methods.

Khalil, A. F., H. H. Kwon, U. Lall, M. J. Miranda, and J. Skees (2007), El Niño–Southern Oscillation–based index insurance for floods: Statistical risk analyses and application to Peru, *Water Resources Research*, *43*(10).

Kim, Y.-H., S.-K. Min, D. A. Stone, H. Shiogama, and P. Wolski (2018), Multi-model event attribution of the summer 2013 heat wave in Korea, *Weather and Climate Extremes*.

Kintisch, E. (2016), How a 'Godzilla'El Niño shook up weather forecasts, *Science*, *352*(6293), 1501-1502.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter III, D. A. Paolino, Q. Zhang, H. Van Den Dool, S. Saha, M. P. Mendez, and E. Becker (2014), The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, *Bulletin of the American Meteorological Society*, *95*(4), 585-601.

Knowlton, K., S. P. Kulkarni, G. S. Azhar, D. Mavalankar, A. Jaiswal, M. Connolly, A. Nori-Sarma, A. Rajiva, P. Dutta, and B. Deol (2014), Development and implementation of South Asia's first heat-health action plan in Ahmedabad (Gujarat, India), *International journal of environmental research and public health*, *11*(4), 3473-3492.

Kojadinovic, I. (2013), A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing.

Korf, R. E. (1985), Depth-first iterative-deepening: An optimal admissible tree search, *Artificial intelligence*, *27*(1), 97-109.

Kothawale, D., A. Munot, and K. K. Kumar (2010), Surface air temperature variability over India during 1901–2007, and its association with ENSO, *Climate Research*, *42*(2), 89-104.

Koutsias, N., M. Arianoutsou, A. S. Kallimanis, G. Mallinis, J. M. Halley, and P. Dimopoulos (2012), Where did the fires burn in Peloponnisos, Greece the summer of 2007? Evidence for a synergy of fuel and weather, *Agricultural and Forest Meteorology*, *156*, 41-53.

Kumar, P., A. Wiltshire, C. Mathison, S. Asharaf, B. Ahrens, P. Lucas-Picher, J. H. Christensen, A. Gobiet, F. Saeed, and S. Hagemann (2013), Downscaled climate change projections with

uncertainty assessment over India using a high resolution multi-model approach, *Science of the Total Environment*, *468*, S18-S30.

Kurtzman, D., and B. R. Scanlon (2007), El Nino–Southern Oscillation and Pacific Decadal Oscillation impacts on precipitation in the southern and central United States: Evaluation of spatial distribution and predictions, *Water Resources Research*, *43*(10).

Leonard, M., S. Westra, A. Phatak, M. Lambert, B. van den Hurk, K. McInnes, J. Risbey, S. Schuster, D. Jakob, and M. Stafford-Smith (2014), A compound event framework for understanding extreme impacts, *Wiley Interdisciplinary Reviews: Climate Change*, *5*(1), 113-128.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier (2013), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions, *Journal of Climate*, *26*(23), 9384-9392.

Madadgar, S., and H. Moradkhani (2013), A Bayesian framework for probabilistic seasonal drought forecasting, *Journal of Hydrometeorology*, *14*(6), 1685-1705.

Madadgar, S., A. AghaKouchak, S. Shukla, A. W. Wood, L. Cheng, K. L. Hsu, and M. Svoboda (2016), A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States, *Water Resources Research*, *52*(7), 5095-5110.

Maurer, E. P., A. Wood, J. Adam, D. P. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *Journal of climate*, *15*(22), 3237-3251.

Mazdiyasni, O., and A. AghaKouchak (2015), Substantial increase in concurrent droughts and heatwaves in the United States, *Proceedings of the National Academy of Sciences*, *112*(37), 11484-11489.

Mazdiyasni, O., A. AghaKouchak, S. J. Davis, S. Madadgar, A. Mehran, E. Ragno, M. Sadegh, A. Sengupta, S. Ghosh, and C. Dhanya (2017), Increasing probability of mortality during Indian heat waves, *Science advances*, *3*(6), e1700066.

McCABE, G. J., and M. D. Dettinger (1999), Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States, *International Journal of Climatology*, *19*(13), 1399-1410.

McKee, T. B., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, paper presented at Proceedings of the 8th Conference on Applied Climatology, American Meteorological Society Boston, MA.

Meehl, G., C. Tebaldi, S. Tilmes, J.-F. Lamarque, S. Bates, A. G. Pendergrass, and D. Lombardozzi (2018), Future heat waves and surface ozone, *Environmental Research Letters*.

Meehl, G. A., and C. Tebaldi (2004), More intense, more frequent, and longer lasting heat waves in the 21st century, *Science*, *305*(5686), 994-997.

Mueller, B., and S. I. Seneviratne (2012), Hot days induced by precipitation deficits at the global scale, *Proceedings of the national academy of sciences*, *109*(31), 12398-12403.

Murari, K. K., S. Ghosh, A. Patwardhan, E. Daly, and K. Salvi (2015), Intensification of future severe heat waves in India and their effect on heat stress and mortality, *Regional Environmental Change*, *15*(4), 569-579.

Nelsen, R. B. (1999), An introduction to copulas, volume 139 of Lecture Notes in Statistics, edited, Springer-Verlag, New York.

Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science & Business Media.

O'Neill, M. S., A. Zanobetti, and J. Schwartz (2005), Disparities by race in heat-related mortality in four US cities: the role of air conditioning prevalence, *Journal of Urban Health*, *82*(2), 191-197.

Otto, F. E., N. Massey, G. Oldenborgh, R. Jones, and M. Allen (2012), Reconciling two approaches to attribution of the 2010 Russian heat wave, *Geophysical Research Letters*, *39*(4).

Perkins, S., and L. Alexander (2013), On the measurement of heat waves, *Journal of Climate*, *26*(13), 4500-4517.

Perkins, S., L. Alexander, and J. Nairn (2012), Increasing frequency, intensity and duration of observed global heatwaves and warm spells, *Geophysical Research Letters*, *39*(20).

Politis, D. N., and H. White (2004), Automatic block-length selection for the dependent bootstrap, *Econometric Reviews*, *23*(1), 53-70.

Poumadere, M., C. Mays, S. Le Mer, and R. Blong (2005), The 2003 heat wave in France: dangerous climate change here and now, *Risk analysis*, *25*(6), 1483-1494.

Rahmstorf, S., and D. Coumou (2011), Increase of extreme events in a warming world, *Proceedings of the National Academy of Sciences*, *108*(44), 17905-17909.

Rasmusson, E. M., and J. M. Wallace (1983), Meteorological aspects of the El Nino/southern oscillation, *Science*, *222*(4629), 1195-1202.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *Journal of climate*, *15*(13), 1609-1625.

Ropelewski, C. F., and M. S. Halpert (1986), North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO), *Monthly Weather Review*, *114*(12), 2352-2362.

Ross, G. J., and N. M. Adams (2012), Two nonparametric control charts for detecting arbitrary distribution changes, *Journal of Quality Technology*, *44*(2), 102-116.

Sadegh, M., E. Ragno, and A. AghaKouchak (2017), Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resources Research*.

Saini, H. S., and M. E. Westgate (1999), Reproductive development in grain crops during drought, in *Advances in agronomy*, edited, pp. 59-96, Elsevier.

Salvadori, G., and C. De Michele (2007), On the use of copulas in hydrology: theory and practice, *Journal of Hydrologic Engineering*, *12*(4), 369-380.

Salvadori, G., F. Durante, and C. De Michele (2011), On the return period and design in a multivariate framework.

Salvadori, G., F. Durante, and C. Michele (2013), Multivariate return period calculation via survival functions, *Water Resources Research*, *49*(4), 2308-2311.

Santer, B. D., J. C. Fyfe, G. Pallotta, G. M. Flato, G. A. Meehl, M. H. England, E. Hawkins, M. E. Mann, J. F. Painter, and C. Bonfils (2017), Causes of differences in model and satellite tropospheric warming rates, *Nature Geoscience*, *10*(7), 478.

Schubert, S., Y. Chang, H. Wang, R. Koster, and M. Suarez (2016), A Modeling Study of the Causes and Predictability of the Spring 2011 Extreme US Weather Activity, *Journal of Climate*, *29*(21), 7869-7887.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2008), Potential predictability of long-term drought and pluvial conditions in the US Great Plains, *Journal of Climate*, *21*(4), 802-816.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär (2006), Land–atmosphere coupling and climate change in Europe, *Nature*, *443*(7108), 205.

Seneviratne, S. I., M. G. Donat, B. Mueller, and L. V. Alexander (2014), No pause in the increase of hot temperature extremes, *Nature Climate Change*, *4*(3), 161.

Seneviratne, S. I., N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, and M. Rahimi (2012), Changes in climate extremes and their impacts on the natural physical environment.

Sharkey, P., and R. Killick (2014), Nonparametric methods for online changepoint detection, *STOR601 Research Topic II*.

Sheffield, J., E. F. Wood, and M. L. Roderick (2012), Little change in global drought over the past 60 years, *Nature*, *491*(7424), 435.

Shepard, D. S. (1984), Computer mapping: The SYMAP interpolation algorithm, in *Spatial statistics and models*, edited, pp. 133-145, Springer.

Shi, L., I. Kloog, A. Zanobetti, P. Liu, and J. D. Schwartz (2015), Impacts of temperature and its variability on mortality in New England, *Nature climate change*, *5*(11), 988-991.

Shiau, J.-T., and R. Modarres (2009), Copula-based drought severity-duration-frequency analysis in Iran, *Meteorological Applications*, *16*(4), 481-489.

Shiau, J. (2006), Fitting drought duration and severity with two-dimensional copulas, *Water resources management*, *20*(5), 795-815.

Shukla, S., M. Safeeq, A. AghaKouchak, K. Guan, and C. Funk (2015), Temperature impacts on the water year 2014 drought in California, *Geophysical Research Letters*, *42*(11), 4384-4393.

Singh, V. P., and L. Zhang (2007), IDF curves using the Frank Archimedean copula, *Journal of Hydrologic Engineering*, *12*(6), 651-662.

Srivastava, A., M. Rajeevan, and S. Kshirsagar (2009), Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region, *Atmospheric Science Letters*, *10*(4), 249-254.

Steinschneider, S., and U. Lall (2016), El Niño and the US precipitation and floods: What was expected for the January–March 2016 winter hydroclimate that is now unfolding?, *Water Resources Research*, *52*(2), 1498-1501.

Stocker, T., D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, B. Bex, and B. Midgley (2013), IPCC, 2013: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.

Sun, Q., C. Miao, A. AghaKouchak, and Q. Duan (2016), Century-scale causal relationships between global dry/wet conditions and the state of the Pacific and Atlantic Oceans, *Geophysical Research Letters*, *43*(12), 6528-6537.

Sun, Q., C. Miao, A. AghaKouchak, and Q. Duan (2017), Unraveling anthropogenic influence on the changing risk of heat waves in China, *Geophysical Research Letters*.

Swain, D. L., D. E. Horton, D. Singh, and N. S. Diffenbaugh (2016), Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California, *Science advances*, *2*(4), e1501344.

Swain, D. L., M. Tsiang, M. Haugen, D. Singh, A. Charland, B. Rajaratnam, and N. S. Diffenbaugh (2014), The extraordinary California drought of 2013/2014: Character, context, and the role of climate change, *Bulletin of the American Meteorological Society*, *95*(9), S3.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, *93*(4), 485-498.

Tebaldi, C., and R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 2053-2075.

Tebaldi, C., and M. F. Wehner (2018), Benefits of mitigation for future heat extremes under RCP4. 5 compared to RCP8. 5, *Climatic Change*, *146*(3-4), 349-361.

Tebaldi, C., and D. Lobell (2018), Estimated impacts of emission reductions on wheat and maize crops, *Climatic Change*, *146*(3-4), 533-545.

Teng, H., and G. Branstator (2016), Causes of extreme ridges that induce California droughts, *Journal of Climate*(2016).

Trenberth, K. E. (1984), Signal versus noise in the Southern Oscillation, *Monthly Weather Review*, *112*(2), 326-332.

Trenberth, K. E. (1997), The definition of el nino, *Bulletin of the American Meteorological Society*, *78*(12), 2771.

Trenberth, K. E., A. Dai, G. Van Der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield (2014), Global warming and changes in drought, *Nature Climate Change*, *4*(1), 17.

Vautard, R., C. Honore, M. Beekmann, and L. Rouil (2005), Simulation of ozone during the August 2003 heat wave and emission control scenarios, *Atmospheric Environment*, *39*(16), 2957-2967.

Wehner, M., D. R. Easterling, J. H. Lawrimore, R. R. Heim Jr, R. S. Vose, and B. D. Santer (2011), Projections of future drought in the continental United States and Mexico, *Journal of Hydrometeorology*, *12*(6), 1359-1377.

Wehner, M., D. Stone, H. Shiogama, P. Wolski, A. Ciavarella, N. Christidis, and H. Krishnan (2018a), Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ Detection and Attribution multi-model ensemble, *Weather and Climate Extremes*.

Wehner, M., D. Stone, D. Mitchell, H. Shiogama, E. Fischer, L. S. Graff, V. V. Kharin, L. Lierhammer, B. Sanderson, and H. Krishnan (2018b), Changes in extremely hot days under stabilized 1.5 and 2.0° C global warming scenarios as simulated by the HAPPI multi-model ensemble, *Earth System Dynamics*, *9*(1), 299.

Westerling, A. L., and T. W. Swetnam (2003), Interannual to decadal drought and wildfire in the western United States, *EOS, Transactions American Geophysical Union*, *84*(49), 545-555.

White, W. B., Y. M. Tourre, M. Barlow, and M. Dettinger (2003), A delayed action oscillator shared by biennial, interannual, and decadal signals in the Pacific Basin, *Journal of Geophysical Research: Oceans*, *108*(C3).

Widmann, M., and C. S. Bretherton (2000), Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States, *Journal of Climate*, *13*(11), 1936-1950.

Wilhite, D. A. (2005), *Drought and water crises: science, technology, and management issues*, CRC Press.

Wilker, E. H., G. Yeh, G. A. Wellenius, R. B. Davis, R. S. Phillips, and M. A. Mittleman (2012), Ambient temperature and biomarkers of heart failure: a repeated measures analysis, *Environmental health perspectives*, *120*(8), 1083.

Wolter, K., and M. S. Timlin (1998), Measuring the strength of ENSO events: How does 1997/98 rank?, *Weather*, *53*(9), 315-324.

Wolter, K., and M. S. Timlin (2011), El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI. ext), *International Journal of Climatology*, *31*(7), 1074-1087.

Wright, S. (1921), Correlation and causation, *Journal of agricultural research*, *20*(7), 557-585.

Zaitchik, B. F., A. K. Macalady, L. R. Bonneau, and R. B. Smith (2006), Europe's 2003 heat wave: A satellite view of impacts and land–atmosphere feedbacks, *International Journal of Climatology*, *26*(6), 743-769.

Zampieri, M., A. Ceglar, F. Dentener, and A. Toreti (2017), Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales, *Environmental Research Letters*, *12*(6), 064008.

Zamuda, C., B. Mignone, D. Bilello, K. Hallett, C. Lee, J. Macknick, R. Newmark, and D. Steinberg (2013), US energy sector vulnerabilities to climate change and extreme weather*Rep.*, Department of Energy Washington DC.

AghaKouchak, A., A. Bárdossy and E. Habib (2010a). Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. *Advances in Water Resources* 33(6): 624-634.

Bacro, J.-N. and C. Gaetan (2012). A review on spatial extreme modelling. Advances and Challenges in Space-time Modelling of Natural Events, Springer**:** 103-124.

Ballani, F. and M. Schlather (2011). A construction principle for multivariate extreme value distributions. *Biometrika* 98(3): 633-645.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water resources research* 42(11): W11416.

Bárdossy, A. and S. Pegram (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences* 13(12): 2299-2314.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*: 318-355.

Beirlant, J., Y. Goegebeur, J. Segers, et al. (2006). Statistics of extremes: theory and applications, Wiley. com.

Beniston, M. (2009). Trends in joint quantiles of temperature and precipitation in Europe since 1901 and projected for 2100. *Geophysical Research Letters* 36(7): L07707.

Bortot, P. and C. Gaetan (2013). Multivariate extremes. *Encyclopedia of Environmetrics*.

Brooks, H. E. (2009). Proximity soundings for severe convection for Europe and the United States from reanalysis data. *Atmospheric Research* 93(1): 546-553.

Brooks, H. E., J. W. Lee and J. P. Craven (2003). The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmospheric Research* 67: 73-94.

Chiang, F., O. Mazdiyasni, and A. AghaKouchak. Amplified Warming of Droughts in Observations and Model Simulations. *In review.*

Coles, S. (2001). An introduction to statistical modeling of extreme values, Springer.

Coles, S., J. Heffernan and J. Tawn (1999). Dependence measures for extreme value analyses. *Extremes* 2(4): 339-365.

Coles, S. and F. Pauli (2002). Models and inference for uncertainty in extremal dependence. *Biometrika* 89(1): 183-196.

Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*: 377-392.

Cooley, D., J. Cisewski, R. J. Erhardt, et al. (2012). A survey of spatial extremes: measuring spatial dependence and modeling spatial effects. *REVSTAT–Statistical Journal* 10(1): 135-165.

Cooley, D., R. A. Davis and P. Naveau (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis* 101(9): 2103-2117.

Das, B. and S. I. Resnick (2011). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli* 17(1): 226-252.

Davis, R. A., C. Klüppelberg and C. Steinkohl (2013). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Davis, R. A., C. Klüppelberg and C. Steinkohl (2013 (2011)). Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society*.

Davis, R. A. and T. Mikosch (2008). Extreme value theory for space–time processes with heavy-tailed distributions. *Stochastic Processes and their Applications* 118(4): 560-584.

Davison, A. C., S. Padoan and M. Ribatet (2012). Statistical modeling of spatial extremes. *Statistical Science* 27(2): 161-186.

Draisma, G., H. Drees, A. Ferreira, et al. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* 10(2): 251-280.

Eastering, D. R., G. A. Meehl, C. Parmesan, et al. (2000). Climate Extremes: Observations, Modeling, and Impacts. *Science* 289(5487): 2068-2074.

Field, C. B., V. Barros, T. F. Stocker, et al. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, UK, and New York, NY, USA*.

Fischer, E. and R. Knutti (2012). Robust projections of combined humidity and temperature extremes. *Nature Climate Change* 3: 126–130.

Galambos, J. (1987). The asymptotic theory of extreme order statistics, Krieger Malabar.

Gallant, A. J. and D. J. Karoly (2010). A combined climate extremes index for the Australian region. *Journal of climate* 23(23): 6153-6165.

Gaume, J., N. Eckert, G. Chambon, et al. (2013). Mapping extreme snowfalls in the French Alps using max-stable processes. *Water resources research* 49(2): 1079-1098.

Gilleland, E., B. G. Brown and C. M. Ammann (2013). Spatial Extreme Value Analysis to Project Extremes of Large-Scale Indicators for Severe Weather. *Environmetrics*.

Gilleland, E., M. Ribatet and A. G. Stephenson (2013). A software review for extreme value analysis. *Extremes* 16(1): 103-119.

Gleason, K. L., J. H. Lawrimore, D. H. Levinson, et al. (2008). A revised US climate extremes index. *Journal of climate* 21(10): 2124-2137.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. Copula Theory and Its Applications, Springer**:** 127-145.

Guillotte, S. and F. Perron (2008). A bayesian estimator for the dependence function of a bivariate extreme-value distribution. *Canadian Journal of Statistics* 36(3): 383-396.

Guillotte, S., F. Perron and J. Segers (2011). Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(3): 377-406.

Hao, Z. and A. AghaKouchak (2013). Multivariate Standardized Drought Index: A Multi-Index Approach for Drought Analysis. *Advances in Water Resources*.

Hao, Z., A. AghaKouchak and T. J. Phillips (2013). Changes in concurrent monthly precipitation and temperature extremes. *Environmental Research Letters* 8(3): 034014.

Heffernan, J. and S. Resnick (2005). Hidden regular variation and the rank transform. *Advances in applied probability* 37(2): 393-414.

Heffernan, J. E. and S. I. Resnick (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*: 537-571.

Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(3): 497-546.

Huser, R. and A. C. Davison (2012). Space-time modelling of extreme events. *arXiv preprint arXiv:1201.3245*.

Jeon, S. and R. L. Smith (2012). Dependence structure of spatial extremes using threshold approach. *arXiv preprint arXiv:1209.6344*.

Joe, H. (1997). Multivariate Models and Dependence Concepts. London, Chapman & Hall.

Joe, H., R. L. Smith and I. Weissman (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*: 171-183.

Jonathan, P., K. Ewans and J. Flynn (2012). Joint modelling of vertical profiles of large ocean currents. *Ocean Engineering* 42: 195-204.

Jonathan, P., K. Ewans and D. Randell (2012b). Joint modelling of environmental parameters for extreme sea states incorporating covariate effects. *preparation for Ocean Engineering.* .

Jonathan, P., K. Ewans and D. Randell (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering* 79: 22-31.

Jonathan, P., J. Flynn and K. Ewans (2010). Joint modelling of wave spectral parameters for extreme sea states. *Ocean Engineering* 37(11): 1070-1080.

Kabluchko, Z., M. Schlather and L. de Haan (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*: 2042-2065.

Kao, S. C. and R. S. Govindaraju (2008). Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resour. Res.* 44(2): W02415, doi:02410.01029/02007WR006261.

Karl, T. R., R. W. Knight, D. R. Easterling, et al. (1996). Indices of climate change for the United States. *Bulletin of the American Meteorological Society* 77(2): 279-292.

Katz, R. W., M. B. Parlange and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25(8-12): 1287-1304.

Katz, R. W., M. B. Parlange and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25(8): 1287-1304.

Keef, C., I. Papastathopoulos and J. A. Tawn (2012). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*.

Keef, C., C. Svensson and J. A. Tawn (2009). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology* 378(3): 240-252.

Kotz, S., N. Balakrishnan and N. L. Johnson (2000). Continuous multivariate distributions, models and applications, Wiley-Interscience.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83(1): 169-187.

Ledford, A. W. and J. A. Tawn (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59(2): 475-499.

Leonard, M., S. Westra, A. Phatak, et al. (2013). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*.

Maulik, K. and S. Resnick (2004). Characterizations and examples of hidden regular variation. *Extremes* 7(1): 31-67.

Mazdiyasni, Omid, and Amir AghaKouchak. "Substantial increase in concurrent droughts and heatwaves in the United States." *Proceedings of the National Academy of Sciences* 112.37 (2015): 11484-11489.

Mazdiyasni, Omid, et al. "Increasing probability of mortality during Indian heat waves." *Science Advances* 3.6 (2017): e1700066.

Nadarajah, S. (1999). A polynomial model for bivariate extreme value distributions. *Statistics & probability letters* 42(1): 15-25.

Perkins, S. and L. Alexander (2012). On the measurement of heatwaves. *Journal of climate*(2012).

Pickands, J. (1981). Multivariate extreme value distributions. Proceedings 43rd Session International Statistical Institute.

Ramos, A. and A. Ledford (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(1): 219-241.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* 5(4): 303-336.

Rocklöv, J. and B. Forsberg. (2009). Comparing approaches for studying the effects of climate extremes - a case study of hospital admissions in Sweden during an extremely warm summer. *Global Health Action* 2(1). doi: 10.3402/gha.v2i0.2034.

Sabourin, A. and P. Naveau (2013). Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis*.

Sabourin, A., P. Naveau and A.-L. Fougères (2013). Bayesian model averaging for multivariate extremes. *Extremes*: 1-26.

Sadegh, M., E. Ragno, and A. AghaKouchak (2017), Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resour. Res.*, 53, 5166–5183, doi:10.1002/2016WR020242.

Salvadori, G. and C. De Michele (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water resources research* 46: W10501, doi:10510.11029/12009WR009040.

Salvadori, G., C. de Michele, N. Kottegoda, et al. (2007). Extremes in Nature: An Approach Using Copulas. New York, Springer.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8: 229–231.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript, Univer*.

Steadman, R. G. (1979). The assessment of sultriness. Part I: A temperature-humidity index based on human physiology and clothing science. *Journal of Applied Meteorology* 18(7): 861-873.

Steadman, R. G. (1984). A universal scale of apparent temperature. *Journal of Climate and Applied Meteorology* 23(12): 1674-1687.

Stephenson, A. G. (2009). HIGH-DIMENSIONAL PARAMETRIC MODELLING OF MULTIVARIATE EXTREME EVENTS. *Australian & New Zealand Journal of Statistics* 51(1): 77-88.

Svensson, C. and D. A. Jones (2002). Dependence between extreme sea surge, river flow and precipitation in eastern Britain. *International Journal of Climatology* 22(10): 1149-1168.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* 75(3): 397-415.

Vandenberghe, S., N. E. C. Verhoest, C. Onof, et al. (2011). A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-

Lewis modeled rainfall. *Water resources research* 47: W07529, doi:07510.01029/02009WR008388.

Weller, G. B., D. S. Cooley and S. R. Sain (2012). An investigation of the pineapple express phenomenon via bivariate extreme value theory. *Environmetrics* 23(5): 420-439.

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, *44*(2), 182-198.

Aas, K., & Berg, D. (2009). Models for construction of multivariate dependence–a comparison study. *The European Journal of Finance*, *15*(7-8), 639-659. http://dx.doi.org/10.1080/13518470802588767

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, *32*(1-4), 245-268. https://doi.org/10.1023/A:1016725902970

Brechmann, E. C., & Schepsmeier, U. (2013). Modeling dependence with C-and D-vine copulas: The R-package CDVine. *Journal of Statistical Software*, *52*(3), 1-27. https://doi.org/10.18637/jss.v052.i03

Bedford, T., & Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 1031-1068. https://doi.org/10.1214/aos/1031689016

Czado, C., Schepsmeier, U., & Min, A. (2012). Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling*, *12*(3), 229-255. https://doi.org/10.1177/1471082X1101200302

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, 120-141.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Kurowicka, D., & Cooke, R. M. (2007). Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics & Data Analysis*, *51*(6), 2889-2906. https://doi.org/10.1016/j.csda.2006.11.043

Liu, Z., Cheng, L., Hao, Z., Li, J., Thorstensen, A., Gao, H., 2017, A Framework for Exploring Joint Effects of Conditional Factors on Compound Floods, *Water Resources Research*, doi: 10.1002/2017WR021662

Schirmacher, D., & Schirmacher, E. (2008). *Multivariate dependence modeling using pair-copulas*. Paper presented at The 2008 Enterprise Risk Management Symposium, Chicago.