Chapman University

Chapman University Digital Commons

Computational and Data Sciences (Ph.D.)
Dissertations

Dissertations and Theses

1-2020

Long Term Ground Based Precipitation Data Analysis: Spatial and Temporal Variability

Luciano Rodriguez

Chapman University, rodri178@mail.chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/cads_dissertations

Part of the [Other Earth Sciences Commons](#)

Recommended Citation

L. Rodriguez, "Long term ground based precipitation data analysis: Spatial and temporal variability," Ph.D. dissertation, Chapman University, Orange, CA, 2019. <https://doi.org/10.36837/chapman.000121>

This Dissertation is brought to you for free and open access by the Dissertations and Theses at Chapman University Digital Commons. It has been accepted for inclusion in Computational and Data Sciences (Ph.D.) Dissertations by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

LONG TERM GROUND BASED PRECIPITATION DATA ANALYSIS: SPATIAL AND TEMPORAL VARIABILITY

A Dissertation by

Luciano Rodriguez

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational Sciences

January 2020

Committee in charge:

Hesham El-Askary, Ph.D., Chair

Cyril Rakovski, Ph.D.

Mohamed Allali, Ph.D.

The dissertation of Luciano Rodriguez is approved.

A handwritten signature in black ink, starting with a large 'H' and followed by a stylized, cursive name. The signature is written above a horizontal line.

Hesham El-Askary, Ph.D., Chair

A handwritten signature in black ink, consisting of a square symbol followed by the name 'Rakovski' in a stylized, blocky font. The signature is written above a horizontal line.

Cyril Rakovski, Ph.D.

A handwritten signature in black ink, featuring a large, stylized 'M' and 'A' followed by the name 'Allali' in a cursive script. The signature is written above a horizontal line.

Mohamed Allali, Ph.D.

December 2019

Long Term Ground Based Precipitation Data Analysis: Spatial and Temporal
Variability

Copyright © 2019

by Luciano Rodriguez

ACKNOWLEDGMENTS

I would like to thank Dr. El-Askary for his continued support and never giving up on me through the rollercoaster of graduate school and life. You pushed me through extremes that I didnt know I was capable of. Thank you for guiding me as your little brother.

To Dr. Rakovski, thank you for being you, always having a positive attitude, laughing about everything, and encouraging and being supportive when everything seemed wrong. “For sure!” :)

To Dr. Allali, thank you for always being helpful, even when it is the last minute. Thank you for not taking it personally when we played a friendly soccer match during the Math Picnic and becoming a great advisor.

Lastly, I would like to thank my family, Gloria Andrea Reynoso Mendez, and my friends for their support and understanding and for all the time that I was absent.

DEDICATION

To

My parents and family

Without their constant support, encouragement, and love, this dissertation would not have
been possible.

ABSTRACT

Long Term Ground Based Precipitation Data Analysis: Spatial and Temporal Variability

by Luciano Rodriguez

This dissertation evaluates response variables (classifiers) on various models applied to the detection of El Niño Southern Oscillation (ENSO) on California's seven climate divisions by using modeled and gauge (in-situ/ground) precipitation measurements and various climate indices. Three scientific studies were conducted as part of this research for evaluation of spatial and temporal ENSO events from modeled and gauge data using: 1) Wavelets 2) Autoregressive-moving-average (ARMA) model / Empirical Mode Decomposition (EMD) 3) Vector Generalized Linear Model (VGLM). This dissertation aims to propose and evaluate a methodology for developing a model to measure ENSO events accurately. The hypothesis is that precipitation data (either modeled or gauge) can be used to forecast ENSO events. One can create an assimilated index from various weighted indices as opposed to solely relying on popular climate indices as SOI, PDO, or Niño 3.4. Another objective is to identify how well modeled precipitation compares to gauge precipitation (ground truth) and if the composition of the indices are the same for both. A methodology for validating the generalization performance of the model is proposed and implemented (Controlled Parameter Cross-Validation). An analysis was performed using modeled data (regeneration of observable measurements and ground measurements) and gauge from seven climate divisions in California.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	IV
DEDICATION	V
ABSTRACT	VI
LIST OF TABLES	X
LIST OF FIGURES	XXII
1. Introduction	1
1.1 Goals of this Research	1
1.2 California’s Climate Background	2
1.3 Importance of Climate Change	3
1.4 Previous Studies of Extreme Precipitation	5
1.4.1 Gauge-Measured Data Versus Remotely-Sensed Data	6
2. First Study: In-Situ and Modeled Precipitation Measurements Analysis with SOI and PDO Using Wavelets	8
2.1 Introduction	8
2.2 Data Used	8
2.3 Methodology	12
2.3.1 Wavelet Function	14
2.3.2 The Continuous Wavelet Transform	15
2.3.3 Cross-Wavelet Transform	16
2.3.4 Wavelet Coherence	17
2.4 Results and Discussion	18
2.4.1 The Continuous Wavelet Transform	18
2.4.2 Cross-Wavelet Transform	26
2.4.2.1 Climate Division 6	26
2.4.2.2 Climate Divisions 1, 2, 3, 4, 5, and 7	34
2.4.3 Wavelet Coherence	37
2.4.3.1 Climate Division 6	37
2.4.3.2 Climate Divisions 1, 2, 3, 4, 5, and 7	43
2.4.4 Wavelet Spectra	45
2.5 Conclusion	49

3. Second Study: In-Situ and Modeled Precipitation Measurements Analysis With SOI, PDO, AMO, and Niño Indices Using Autoregressive-Moving-Average Model and Empirical Mode Decomposition	51
3.1 Introduction	51
3.2 Data Used	52
3.3 Methodology	54
3.3.1 ARMA	54
3.3.2 EMD	56
3.4 Results and Discussion	60
3.4.1 ARMA	60
3.4.1.1 Extreme Precipitation	61
3.4.1.2 Extreme Drought	68
3.4.2 Statistical Comparison for Modeled and Gauge Datasets	70
3.4.3 EMD	74
3.5 Conclusion	88
4. Third Study: Assimilation of a new Climate Index for Gauge and Modeled Data Using Autoregressive Moving Average and Advance Modeling Techniques	89
4.1 Introduction	89
4.2 Data Used	90
4.3 Methodology	95
4.3.1 Vector Generalized Linear Models	96
4.3.1.1 Main Effect	98
4.3.1.2 Two-Way Interaction	99
4.3.1.3 Three-Way Interaction	100
4.3.2 XGBoost	100
4.3.3 Random Forest Classifier	104
4.3.4 Extra Trees	105
4.3.5 Neural Networks	108
4.3.6 Lasso/Ridge Regression	110
4.4 Results and Discussion	113
4.4.1 Vector Generalized Linear Models	114
4.4.1.1 Main Effect	114
4.4.1.2 Two-Way Interaction	122
4.4.1.3 Three-Way Interaction	125
4.4.2 XGBoost	127
4.4.3 Random Forest Classifier	131
4.4.4 Extra Trees	136
4.4.5 Neural Networks	138

4.4.6	Ridge Regression	143
4.4.7	Lasso Regression	149
4.5	Conclusion	157
5.	Conclusion	159
	REFERENCES	162
	APPENDICES	189

LIST OF TABLES

	Page
2.1 Event years of El Niño (red) and La Niña (blue) categorized as Weak, Moderate, Strong, and Very Strong. The table was obtained from Golden Gate Weather Services [39]	11
2.2 Modeled and Gauge Residual Time-Series Wavelet Spectra	46
2.3 Gauge and SOI Residual Time-Series Wavelet Spectra	46
2.4 Gauge and PDO Residual Time-Series Wavelet Spectra	47
2.5 Modeled and SOI Residual Time-Series Wavelet Spectra	48
2.6 Modeled and PDO Residual Time-Series Wavelet Spectra	48
3.1 Excessive precipitation dates in climate division six	64
3.2 Drought dates in climate division six	69
3.3 Precipitation data from gauge time-series compared to SOI, PDO, and AMO. Each row block represents a climate division and underneath has a value of the number of stations in each division. Within each row block there are four rows that state the cross correlation, precipitation component, climate index component, and lag (in months). Each column block represents the climate index of interest.	78
3.4 Precipitation data from modeled time-series compared to SOI, PDO, and AMO. Each row block represents a climate division and underneath has a value of the number of stations in each division. Within each row block there are four rows that state the cross correlation, precipitation component, climate index component, and lag (in months). Each column block represents the climate index of interest.	80
3.5 Gauge and SOI EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	82

3.6	Modeled and SOI EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	83
3.7	Gauge and PDO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	84
3.8	Modeled and PDO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	85
3.9	Gauge and AMO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	86
3.10	Modeled and AMO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.	87
4.1	Extra-Trees splitting algorithm (for numerical attributes) from Geurts et al. [37]	106
4.2	F1 scores for training and testing for main effect on gauge time-series climate division six	115
4.3	F1 scores for training and testing for main effect on modeled time-series climate division six	116
4.4	Gauge predictors frequency for California training from 1948 to 2018	117
4.5	Model predictors frequency for California training from 1948 to 2018	120
4.6	Gauge and modeled predictors per climate division of the main effect in California training from 1948 to 2018	123
4.7	F1 scores for training and testing for two-way interaction on gauge time-series climate division six	124
4.8	F1 scores for training and testing for two-way interaction on modeled time-series climate division six	125
4.9	F1 scores for training and testing for three-way interaction on gauge time-series climate division six	126

4.10	F1 scores for training and testing for three-way interaction on modeled time-series climate division six	127
4.11	F1 scores for training and testing for XGBoost on gauge time-series climate division six	128
4.12	F1 scores for training and testing for XGBoost on modeled time-series climate division six	130
4.13	F1 scores for training and testing for Random Forest on gauge time-series climate division six	132
4.14	F1 scores for training and testing for Random Forest on modeled time-series climate division six	135
4.15	F1 scores for training and testing for Extra Trees on gauge time-series climate division six	142
4.16	F1 scores for training and testing for Extra Trees on modeled time-series climate division six	142
4.17	F1 scores for training and testing for Neural Network on gauge time-series climate division six	142
4.18	F1 scores for training and testing for Neural Network on modeled time-series climate division six	142
4.19	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division six	154
4.20	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division six	154
4.21	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division six	154
4.22	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division six	154
4.23	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division six	154
4.24	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division six	156
B.1	Excessive precipitation dates in climate division one	276
B.2	Excessive precipitation dates in climate division two	279
B.3	Excessive precipitation dates in climate division three	283

B.4	Excessive precipitation dates in climate division four	286
B.5	Excessive precipitation dates in climate division five	289
B.6	Excessive precipitation dates in climate division seven	293
B.7	Drought dates in climate division one	296
B.8	Drought dates in climate division two	298
B.9	Drought dates in climate division three	299
B.10	Drought dates in climate division four	300
B.11	Drought dates in climate division five	304
B.12	Drought dates in climate division seven	306
D.1	F1 scores for training and testing for main effect on gauge time-series climate division one	329
D.2	F1 scores for training and testing for main effect on gauge time-series climate division two	329
D.3	F1 scores for training and testing for main effect on gauge time-series climate division three	329
D.4	F1 scores for training and testing for main effect on gauge time-series climate division four	330
D.5	F1 scores for training and testing for main effect on gauge time-series climate division five	330
D.6	F1 scores for training and testing for main effect on gauge time-series climate division seven	330
D.7	F1 scores for training and testing for main effect on modeled time-series climate division one	330
D.8	F1 scores for training and testing for main effect on modeled time-series climate division two	331
D.9	F1 scores for training and testing for main effect on modeled time-series climate division three	331
D.10	F1 scores for training and testing for main effect on modeled time-series climate division four	331
D.11	F1 scores for training and testing for main effect on modeled time-series climate division five	331

D.12	F1 scores for training and testing for main effect on modeled time-series climate division seven	332
D.13	F1 scores for training and testing for two-way interaction on gauge time-series climate division one	339
D.14	F1 scores for training and testing for two-way interaction on gauge time-series climate division two	339
D.15	F1 scores for training and testing for two-way interaction on gauge time-series climate division three	339
D.16	F1 scores for training and testing for two-way interaction on gauge time-series climate division four	340
D.17	F1 scores for training and testing for two-way interaction on gauge time-series climate division five	340
D.18	F1 scores for training and testing for two-way interaction on gauge time-series climate division seven	340
D.19	F1 scores for training and testing for two-way interaction on modeled time-series climate division one	340
D.20	F1 scores for training and testing for two-way interaction on modeled time-series climate division two	341
D.21	F1 scores for training and testing for two-way interaction on modeled time-series climate division three	341
D.22	F1 scores for training and testing for two-way interaction on modeled time-series climate division four	341
D.23	F1 scores for training and testing for two-way interaction on modeled time-series climate division five	341
D.24	F1 scores for training and testing for two-way interaction on modeled time-series climate division seven	342
D.25	F1 scores for training and testing for three-way interaction on gauge time-series climate division one	349
D.26	F1 scores for training and testing for three-way interaction on gauge time-series climate division two	349
D.27	F1 scores for training and testing for three-way interaction on gauge time-series climate division three	349
D.28	F1 scores for training and testing for three-way interaction on gauge time-series climate division four	350

D.29	F1 scores for training and testing for three-way interaction on gauge time-series climate division five	350
D.30	F1 scores for training and testing for three-way interaction on gauge time-series climate division seven	350
D.31	F1 scores for training and testing for three-way interaction on modeled time-series climate division one	350
D.32	F1 scores for training and testing for three-way interaction on modeled time-series climate division two	351
D.33	F1 scores for training and testing for three-way interaction on modeled time-series climate division three	351
D.34	F1 scores for training and testing for three-way interaction on modeled time-series climate division four	351
D.35	F1 scores for training and testing for three-way interaction on modeled time-series climate division five	351
D.36	F1 scores for training and testing for three-way interaction on modeled time-series climate division seven	352
E.1	F1 scores for training and testing for XGBoost on gauge time-series climate division one	360
E.2	F1 scores for training and testing for XGBoost on gauge time-series climate division two	360
E.3	F1 scores for training and testing for XGBoost on gauge time-series climate division three	360
E.4	F1 scores for training and testing for XGBoost on gauge time-series climate division four	361
E.5	F1 scores for training and testing for XGBoost on gauge time-series climate division five	361
E.6	F1 scores for training and testing for XGBoost on gauge time-series climate division seven	361
E.7	F1 scores for training and testing for XGBoost on modeled time-series climate division one	361
E.8	F1 scores for training and testing for XGBoost on modeled time-series climate division two	362
E.9	F1 scores for training and testing for XGBoost on modeled time-series climate division three	362

E.10	F1 scores for training and testing for XGBoost on modeled time-series climate division four	362
E.11	F1 scores for training and testing for XGBoost on modeled time-series climate division five	362
E.12	F1 scores for training and testing for XGBoost on modeled time-series climate division seven	363
F.1	F1 scores for training and testing for Random Forest on gauge time-series climate division one	382
F.2	F1 scores for training and testing for Random Forest on gauge time-series climate division two	382
F.3	F1 scores for training and testing for Random Forest on gauge time-series climate division three	382
F.4	F1 scores for training and testing for Random Forest on gauge time-series climate division four	383
F.5	F1 scores for training and testing for Random Forest on gauge time-series climate division five	383
F.6	F1 scores for training and testing for Random Forest on gauge time-series climate division seven	383
F.7	F1 scores for training and testing for Random Forest on modeled time-series climate division one	383
F.8	F1 scores for training and testing for Random Forest on modeled time-series climate division two	384
F.9	F1 scores for training and testing for Random Forest on modeled time-series climate division three	384
F.10	F1 scores for training and testing for Random Forest on modeled time-series climate division four	384
F.11	F1 scores for training and testing for Random Forest on modeled time-series climate division five	384
F.12	F1 scores for training and testing for Random Forest on modeled time-series climate division seven	385
G.1	F1 scores for training and testing for Extra Trees on gauge time-series climate division one	404
G.2	F1 scores for training and testing for Extra Trees on gauge time-series climate division two	404

G.3	F1 scores for training and testing for Extra Trees on gauge time-series climate division three	404
G.4	F1 scores for training and testing for Extra Trees on gauge time-series climate division four	405
G.5	F1 scores for training and testing for Extra Trees on gauge time-series climate division five	405
G.6	F1 scores for training and testing for Extra Trees on gauge time-series climate division seven	405
G.7	F1 scores for training and testing for Extra Trees on modeled time-series climate division one	405
G.8	F1 scores for training and testing for Extra Trees on modeled time-series climate division two	406
G.9	F1 scores for training and testing for Extra Trees on modeled time-series climate division three	406
G.10	F1 scores for training and testing for Extra Trees on modeled time-series climate division four	406
G.11	F1 scores for training and testing for Extra Trees on modeled time-series climate division five	406
G.12	F1 scores for training and testing for Extra Trees on modeled time-series climate division seven	407
H.1	F1 scores for training and testing for Neural Network on gauge time-series climate division one	415
H.2	F1 scores for training and testing for Neural Network on gauge time-series climate division two	415
H.3	F1 scores for training and testing for Neural Network on gauge time-series climate division three	415
H.4	F1 scores for training and testing for Neural Network on gauge time-series climate division four	416
H.5	F1 scores for training and testing for Neural Network on gauge time-series climate division five	416
H.6	F1 scores for training and testing for Neural Network on gauge time-series climate division seven	416
H.7	F1 scores for training and testing for Neural Network on modeled time-series climate division one	416

H.8	F1 scores for training and testing for Neural Network on modeled time-series climate division two	417
H.9	F1 scores for training and testing for Neural Network on modeled time-series climate division three	417
H.10	F1 scores for training and testing for Neural Network on modeled time-series climate division four	417
H.11	F1 scores for training and testing for Neural Network on modeled time-series climate division five	417
H.12	F1 scores for training and testing for Neural Network on modeled time-series climate division seven	418
I.1	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division one	444
I.2	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division two	444
I.3	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division three	444
I.4	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division four	445
I.5	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division five	445
I.6	F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division seven	445
I.7	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division one	445
I.8	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division two	446
I.9	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division three	446
I.10	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division four	446
I.11	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division five	446
I.12	F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division seven	446

I.13	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division one	447
I.14	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division two	447
I.15	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division three	447
I.16	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division four	448
I.17	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division five	448
I.18	F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division seven	448
I.19	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division one	448
I.20	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division two	449
I.21	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division three	449
I.22	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division four	449
I.23	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division five	449
I.24	F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division seven	449
J.1	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division one	488
J.2	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division two	488
J.3	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division three	488
J.4	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division four	489
J.5	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division five	489

J.6	F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division seven	489
J.7	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division one	489
J.8	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division two	490
J.9	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division three	490
J.10	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division four	490
J.11	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division five	490
J.12	F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division seven	490
J.13	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division one	491
J.14	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division two	491
J.15	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division three	491
J.16	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division four	492
J.17	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division five	492
J.18	F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division seven	492
J.19	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division one	492
J.20	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division two	493
J.21	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division three	493
J.22	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division four	493

J.23	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division five	493
J.24	F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division seven	493

LIST OF FIGURES

	Page
1.1 7 Climate Divisions of California: 1 North Coast Drainage, 2 Sacramento Drainage, 3 Northeast Interior Basins, 4 Central Coast Drainage, 5 San Joaquin Drainage, 6 South Coast Drainage, 7 South Desert Basin	4
2.1 Locations of the 735 stations of interest within California’s Seven Climate Divisions	10
2.2 A) Model, Gauge, and SOI time-series with respective correlation for climate division one. B) Model, Gauge, and PDO time-series with respective correlation for climate division one	11
2.3 Gauge time-series decomposition for climate division six	13
2.4 Modeled time-series decomposition for climate division six	14
2.5 Continuous Wavelet power spectrum of gauge residuals in climate division six .	19
2.6 Continuous Wavelet power spectrum of modeled residuals in climate division six	20
2.7 Reproduced image from Golden Gate Weather Services [39] that shows the percentage of normal rainfall in California	23
2.8 Continuous wavelet power spectrum of SOI	25
2.9 Continuous wavelet power spectrum of PDO	26
2.10 Cross-Wavelet Transform of modeled and gauge residuals in climate division six	28
2.11 Cross-Wavelet Transform of gauge residuals and SOI in climate division six . .	30
2.12 Cross-Wavelet Transform of gauge residuals and PDO in climate division six .	31
2.13 Cross-Wavelet Transform of modeled residuals and SOI in climate division six .	33
2.14 Cross-Wavelet Transform of modeled residuals and PDO in climate division six	34
2.15 Wavelet Coherence of modeled and gauge residuals in climate division six . . .	38
2.16 Wavelet-Coherence of gauge residuals and SOI in climate division six	39
2.17 Wavelet-Coherence of gauge residuals and PDO in climate division six	41

2.18	Wavelet-Coherence of modeled residuals and SOI in climate division six	42
2.19	Wavelet-Coherence of modeled residuals and PDO in climate division six	43
3.1	EMD of gauge time-series for Climate Division 1. Decomposition of 10 IMF components and the final residual	59
3.2	Gauge monthly residual plot for climate division six	63
3.3	Modeled monthly residual plot for climate division six	64
3.4	Histogram of gauge minus modeled precipitation in climate division six	72
3.5	Climate division six EMD correlation between Modeled/Gauge and SOI/PDO/AMO	76
3.6	Climate division gauge-indices correlation	79
3.7	Climate division modeled-indices correlation	81
4.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division six	115
4.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division six	116
4.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division six	124
4.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division six	125
4.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division six	126
4.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division six	127
4.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division six	128
4.8	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division six	129
4.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division six	130
4.10	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division six	131
4.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division six	132

4.12	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division six	134
4.13	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division six	135
4.14	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division six	136
4.15	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division six	137
4.16	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division six	138
4.17	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division six	139
4.18	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division six	140
4.19	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division six	141
4.20	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division six	143
4.21	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division six	144
4.22	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division six	145
4.23	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division six	146
4.24	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division six	147
4.25	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division six	148
4.26	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division six	149
4.27	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division six	151
4.28	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division six	152

4.29	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division six	153
4.30	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division six	155
4.31	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division six	156
4.32	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division six	157
A.1	Continuous Wavelet power spectrum of gauge residuals in climate division one	191
A.2	Continuous Wavelet power spectrum of gauge residuals in climate division two	192
A.3	Continuous Wavelet power spectrum of gauge residuals in climate division three	193
A.4	Continuous Wavelet power spectrum of gauge residuals in climate division four	194
A.5	Continuous Wavelet power spectrum of gauge residuals in climate division five	195
A.6	Continuous Wavelet power spectrum of gauge residuals in climate division seven	196
A.7	Continuous Wavelet power spectrum of modeled residuals in climate division one	197
A.8	Continuous Wavelet power spectrum of modeled residuals in climate division two	198
A.9	Continuous Wavelet power spectrum of modeled residuals in climate division three	199
A.10	Continuous Wavelet power spectrum of modeled residuals in climate division four	200
A.11	Continuous Wavelet power spectrum of modeled residuals in climate division five	201
A.12	Continuous Wavelet power spectrum of modeled residuals in climate division seven	202
A.13	Cross-Wavelet Transform of modeled and gauge residuals in climate division one	203
A.14	Cross-Wavelet Transform of modeled and gauge residuals in climate division two	204
A.15	Cross-Wavelet Transform of modeled and gauge residuals in climate division three	205
A.16	Cross-Wavelet Transform of modeled and gauge residuals in climate division four	206
A.17	Cross-Wavelet Transform of modeled and gauge residuals in climate division five	207
A.18	Cross-Wavelet Transform of modeled and gauge residuals in climate division seven	208
A.19	Cross-Wavelet Transform of gauge residuals and SOI in climate division one . .	209
A.20	Cross-Wavelet Transform of gauge residuals and SOI in climate division two . .	210
A.21	Cross-Wavelet Transform of gauge residuals and SOI in climate division three .	211

A.22	Cross-Wavelet Transform of gauge residuals and SOI in climate division four . .	212
A.23	Cross-Wavelet Transform of gauge residuals and SOI in climate division five . .	213
A.24	Cross-Wavelet Transform of gauge residuals and SOI in climate division seven .	214
A.25	Cross-Wavelet Transform of gauge residuals and PDO in climate division one .	215
A.26	Cross-Wavelet Transform of gauge residuals and PDO in climate division two .	216
A.27	Cross-Wavelet Transform of gauge residuals and PDO in climate division three	217
A.28	Cross-Wavelet Transform of gauge residuals and SOI in climate division four .	218
A.29	Cross-Wavelet Transform of gauge residuals and PDO in climate division five .	219
A.30	Cross-Wavelet Transform of gauge residuals and PDO in climate division seven	220
A.31	Cross-Wavelet Transform of modeled residuals and SOI in climate division one	221
A.32	Cross-Wavelet Transform of modeled residuals and SOI in climate division two	222
A.33	Cross-Wavelet Transform of modeled residuals and SOI in climate division three	223
A.34	Cross-Wavelet Transform of modeled residuals and SOI in climate division four	224
A.35	Cross-Wavelet Transform of modeled residuals and SOI in climate division five	225
A.36	Cross-Wavelet Transform of modeled residuals and SOI in climate division seven	226
A.37	Cross-Wavelet Transform of modeled residuals and PDO in climate division one	227
A.38	Cross-Wavelet Transform of modeled residuals and PDO in climate division two	228
A.39	Cross-Wavelet Transform of modeled residuals and PDO in climate division three	229
A.40	Cross-Wavelet Transform of modeled residuals and PDO in climate division four	230
A.41	Cross-Wavelet Transform of modeled residuals and PDO in climate division five	231
A.42	Cross-Wavelet Transform of modeled residuals and PDO in climate division seven	232
A.43	Wavelet Coherence of modeled and gauge residuals in climate division one . . .	233
A.44	Wavelet Coherence of modeled and gauge residuals in climate division two . . .	234
A.45	Wavelet Coherence of modeled and gauge residuals in climate division three . .	235
A.46	Wavelet Coherence of modeled and gauge residuals in climate division four . .	236
A.47	Wavelet Coherence of modeled and gauge residuals in climate division five . . .	237
A.48	Wavelet Coherence of modeled and gauge residuals in climate division seven . .	238

A.49	Wavelet-Coherence of gauge residuals and SOI in climate division one	239
A.50	Wavelet-Coherence of gauge residuals and SOI in climate division two	240
A.51	Wavelet-Coherence of gauge residuals and SOI in climate division three	241
A.52	Wavelet-Coherence of gauge residuals and SOI in climate division four	242
A.53	Wavelet-Coherence of gauge residuals and SOI in climate division five	243
A.54	Wavelet-Coherence of gauge residuals and SOI in climate division seven	244
A.55	Wavelet-Coherence of gauge residuals and PDO in climate division one	245
A.56	Wavelet-Coherence of gauge residuals and PDO in climate division two	246
A.57	Wavelet-Coherence of gauge residuals and PDO in climate division three	247
A.58	Wavelet-Coherence of gauge residuals and PDO in climate division four	248
A.59	Wavelet-Coherence of gauge residuals and PDO in climate division five	249
A.60	Wavelet-Coherence of gauge residuals and PDO in climate division seven	250
A.61	Wavelet-Coherence of modeled residuals and SOI in climate division one	251
A.62	Wavelet-Coherence of modeled residuals and SOI in climate division two	252
A.63	Wavelet-Coherence of modeled residuals and SOI in climate division three	253
A.64	Wavelet-Coherence of modeled residuals and SOI in climate division four	254
A.65	Wavelet-Coherence of modeled residuals and SOI in climate division five	255
A.66	Wavelet-Coherence of modeled residuals and SOI in climate division seven	256
A.67	Wavelet-Coherence of modeled residuals and PDO in climate division one	257
A.68	Wavelet-Coherence of modeled residuals and PDO in climate division two	258
A.69	Wavelet-Coherence of modeled residuals and PDO in climate division three	259
A.70	Wavelet-Coherence of modeled residuals and PDO in climate division four	260
A.71	Wavelet-Coherence of modeled residuals and PDO in climate division five	261
A.72	Wavelet-Coherence of modeled residuals and PDO in climate division seven	262
B.1	Gauge monthly residual plot for climate division one	264
B.2	Gauge monthly residual plot for climate division two	265
B.3	Gauge monthly residual plot for climate division three	266

B.4	Gauge monthly residual plot for climate division four	267
B.5	Gauge monthly residual plot for climate division five	268
B.6	Gauge monthly residual plot for climate division seven	269
B.7	Modeled monthly residual plot for climate division one	270
B.8	Modeled monthly residual plot for climate division two	271
B.9	Modeled monthly residual plot for climate division three	272
B.10	Modeled monthly residual plot for climate division four	273
B.11	Modeled monthly residual plot for climate division five	274
B.12	Modeled monthly residual plot for climate division seven	275
B.13	Histogram of gauge minus modeled precipitation in climate division one	310
B.14	Histogram of gauge minus modeled precipitation in climate division two	311
B.15	Histogram of gauge minus modeled precipitation in climate division three . . .	312
B.16	Histogram of gauge minus modeled precipitation in climate division four	313
B.17	Histogram of gauge minus modeled precipitation in climate division five	314
B.18	Histogram of gauge minus modeled precipitation in climate division seven . . .	315
C.1	Climate division one EMD correlation between Modeled/Gauge and SOI/PDO/AMO316	
C.2	Climate division two EMD correlation between Modeled/Gauge and SOI/PDO/AMO317	
C.3	Climate division three EMD correlation between Modeled/Gauge and SOI/PDO/AMO318	
C.4	Climate division four EMD correlation between Modeled/Gauge and SOI/PDO/AMO319	
C.5	Climate division five EMD correlation between Modeled/Gauge and SOI/PDO/AMO320	
C.6	Climate division seven EMD correlation between Modeled/Gauge and SOI/PDO/AMO321	
D.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division one	322
D.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division two	323
D.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division three	323
D.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division four	324

D.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division five	324
D.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division seven	325
D.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division one	326
D.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division two	326
D.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division three	327
D.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division four	327
D.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division five	328
D.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division seven	328
D.13	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division one	332
D.14	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division two	333
D.15	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division three	333
D.16	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division four	334
D.17	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division five	334
D.18	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division seven	335
D.19	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division one	336
D.20	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division two	336
D.21	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division three	337

D.22	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division four	337
D.23	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division five	338
D.24	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division seven	338
D.25	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division one	342
D.26	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division two	343
D.27	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division three	343
D.28	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division four	344
D.29	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division five	344
D.30	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division seven	345
D.31	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division one	346
D.32	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division two	346
D.33	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division three	347
D.34	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division four	347
D.35	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division five	348
D.36	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division seven	348
E.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division one	353
E.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division two	354

E.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division three	354
E.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division four	355
E.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division five	355
E.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division seven	356
E.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division one	357
E.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division two	357
E.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division three	358
E.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division four	358
E.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division five	359
E.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division seven	359
E.13	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division one	363
E.14	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division two	364
E.15	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division three	365
E.16	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division four	366
E.17	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division five	367
E.18	XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	368
E.19	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division one	369

E.20	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division two	370
E.21	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division three	371
E.22	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division four	372
E.23	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division five	373
E.24	XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	374
F.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division one	375
F.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division two	376
F.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division three	376
F.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division four	377
F.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division five	377
F.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division seven	378
F.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division one	379
F.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division two	379
F.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division three	380
F.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division four	380
F.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division five	381
F.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division seven	381

F.13	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division one	385
F.14	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division two	386
F.15	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division three	387
F.16	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division four	388
F.17	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division five	389
F.18	Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	390
F.19	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division one	391
F.20	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division two	392
F.21	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division three	393
F.22	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division four	394
F.23	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division five	395
F.24	Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	396
G.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division one	397
G.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division two	398
G.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division three	398
G.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division four	399
G.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division five	399

G.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division seven	400
G.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division one	401
G.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division two	401
G.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division three	402
G.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division four	402
G.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division five	403
G.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division seven	403
H.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division one	408
H.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division two	409
H.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division three	409
H.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division four	410
H.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division five	410
H.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division seven	411
H.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division one	412
H.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division two	412
H.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division three	413
H.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division four	413

H.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division five	414
H.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division seven	414
H.13	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division one	418
H.14	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division two	419
H.15	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division three	420
H.16	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division four	421
H.17	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division five	422
H.18	Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	423
H.19	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division one	424
H.20	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division two	425
H.21	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division three	426
H.22	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division four	427
H.23	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division five	428
H.24	Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	429
I.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division one	430
I.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division two	431
I.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division three	431

I.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division four	432
I.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division five	432
I.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division seven	433
I.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division one	434
I.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division two	434
I.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division three	435
I.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division four	435
I.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division five	436
I.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division seven	436
I.13	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division one	437
I.14	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division two	438
I.15	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division three	438
I.16	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division four	439
I.17	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division five	439
I.18	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division seven	440
I.19	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division one	441
I.20	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division two	441

I.21	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division three	442
I.22	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division four	442
I.23	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division five	443
I.24	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division seven	443
I.25	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division one	450
I.26	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division two	451
I.27	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division three	452
I.28	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division four	453
I.29	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division five	454
I.30	Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	455
I.31	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division one	456
I.32	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division two	457
I.33	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division three	458
I.34	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division four	459
I.35	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division five	460
I.36	Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	461
I.37	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division one	462

I.38	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division two	463
I.39	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division three	464
I.40	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division four	465
I.41	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division five	466
I.42	Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	467
I.43	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division one	468
I.44	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division two	469
I.45	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division three	470
I.46	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division four	471
I.47	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division five	472
I.48	Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	473
J.1	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division one	474
J.2	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division two	475
J.3	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division three	475
J.4	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division four	476
J.5	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division five	476
J.6	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division seven	477

J.7	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division one	478
J.8	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division two	478
J.9	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division three	479
J.10	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division four	479
J.11	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division five	480
J.12	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division seven	480
J.13	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division one	481
J.14	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division two	482
J.15	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division three	482
J.16	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division four	483
J.17	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division five	483
J.18	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division seven	484
J.19	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division one	485
J.20	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division two	485
J.21	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division three	486
J.22	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division four	486
J.23	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division five	487

J.24	Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division seven	487
J.25	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division one	494
J.26	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division two	495
J.27	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division three	496
J.28	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division four	497
J.29	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division five	498
J.30	Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	499
J.31	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division one	500
J.32	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division two	501
J.33	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division three	502
J.34	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division four	503
J.35	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division five	504
J.36	Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	505
J.37	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division one	506
J.38	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division two	507
J.39	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division three	508
J.40	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division four	509

J.41	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division five	510
J.42	Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division seven	511
J.43	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division one	512
J.44	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division two	513
J.45	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division three	514
J.46	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division four	515
J.47	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division five	516
J.48	Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division seven	517

1. Introduction

1.1 Goals of this Research

The world's climate is a never-ending changing climate. Due to the various coupling of Earth's spheres, there exist many variations of local and regional climate change. Unfortunately, humans have not been living long enough to keep a record of changes in the weather and even a lesser chance of developing instruments to have accurate measurements to analyze the climate change properly. Ideally, to have a better understanding of the climate and changes within, one must know and understand the small processes that make up the significant climate changes over many years. Weather, for example, is one of the small changes that affect the climate. The higher the frequency of data taken from the weather will result in the capture of information in a higher temporal spectrum. This, in return, will result in an accurate representation of the changing climate. Patterns such as drought, precipitation, flash floods, thunderstorms, etcetera will be accurately detected and provide enough warning time for people to mitigate. This dissertation focuses on precipitation analysis to isolate extreme precipitation and drought using sophisticated univariate and multivariate models in seven regions of California [89] to map out patterns for the whole of California correctly.

1.2 California's Climate Background

The climate of California is heterogeneous throughout the various climate divisions. It varies from arid to subarctic depending on the latitude, elevation, near effects from the coast line, and time-varying land use patterns [1]. Since the majority of the land is next to the ocean, there exists an extreme influence from oceanic properties. It creates a bipolar attribute to the seasons where the winter season is warmer and summer season is cooler. These properties induced by the ocean changes the temperature and climate of inland areas with subtropical traits where the summer has dry season and cool, foggy, and rainy seasons. In addition to temperature changes, westerly winds deriving from the ocean bring moisture. Northern California is subtle to more rain than Southern California. Some of the areas that experience the most rain are northwestern California and west-facing mountain slopes. The Californian mountains experience abundant rainfall due to the moisture-laden air from the west cools as it ascends through the mountains. Snow is also present in various mountains where over 10 feet of snow can be obtained within a season. Furthermore, during July through September to map out patterns for all of California properly, various desert regions are affected by the Mexican Monsoon (southwest monsoon) where moisture originates from the tropical Pacific, Gulf of California, and/or Gulf of Mexico which contributes to the formations of thunderstorms. California and Californians are fortunate to not have any formation of tropical cyclones due to the cold current from the North Pacific Ocean [184].

In addition to heavy precipitation from monsoon activities and pineapple express, wild-fires are extremely common during the summer. The season brings hot and dry weather,

mostly in Southern California, to ignite wildfires. A catalyst of wildfires originates from the infamous Santa Ana winds. They bring the heat and high winds from the eastern mountains towards coastal areas in Southern California during the spring and fall. They are generated from air pressure buildup between the Sierra Nevada and Rocky Mountains. Wildfires are caused by the Santa Ana winds due to their combination of high velocity wind speeds, heat, and dryness that increases friction on dry vegetation causing the spark of fire to fuel the area. These combinations of natural hazards can be determined based on the characteristics and analysis of precipitation.

California is partitioned into seven distinct climate divisions defined by the National Climatic Data Center (NCDC) where the topography of California plays a great part. The boundaries are reflected upon the alignment of the watersheds, agricultural administrative districts, geographic convenience, and distinct judgment of the committee [42]. Figure 1.1 shows the topography of California with the changes of elevation where green represents sea level and increasing color to white represents high elevation. Yosemite National Park which is part of the Sierra Nevada is one of the highest peaks in the topography where the boundaries between three divisions occur (2, 3, and 5).

1.3 Importance of Climate Change

The objective of studying the change of the climate not only for California but the world is to properly model the spatial and temporal variability of the climate in 4 distinct

Figure 1.1: 7 Climate Divisions of California: 1 North Coast Drainage, 2 Sacramento Drainage, 3 Northeast Interior Basins, 4 Central Coast Drainage, 5 San Joaquin Drainage, 6 South Coast Drainage, 7 South Desert Basin

areas described by Abatzoglou et al. [1]: 1) have a clear method for explaining climate change to the general public, 2) is logistically and practically computable, 3) has a strong correlation with local and regional climate, 4) outputs can be analyzed to understand the climate process through the temporal variation and spatial variability. Flooding is one of the catalysts of precipitation that has been causing a major concern for the United States west coast. Between 1980 and 2008, there have been four major flooding events along the west coast of the U.S. causing \$11 billion of damages [80]. The increasing quantity of events of precipitation has been documented by many governmental and nongovernmental organizations suggesting that it is due to anthropogenic global warming [128]. The increasing

threat suggests implementing new standards for public safety, management of rivers for fisheries, power generation, and flooding as various other applicable safety measures [80].

1.4 Previous Studies of Extreme Precipitation

Precipitation patterns at extreme events are closely related to the change in the moisture content of the air. Allen and Ingram [2] have suggested that extreme precipitation is proportional to water vapor content by $7.5\%^{\circ}\text{C}^{-1}$ of warming. Many studies of heavy precipitation trends in the U.S. found a heterogeneous pattern of temporal trends specifically in the western part of the United States. Some of the trends performed by researches show minimal trend changes if any in the west coast. However, Madsen and Figdor [76] studied the frequency of 1-day precipitation with a recurrence interval of one year from 1948-2006 using National Climatic Data Center (NCDC) Cooperative and Summary of the Day dataset and found a decrease in precipitation in northern California and an increase over central and southern California coastal zones. Similarly, Dulière et al. [26] observed and simulated extreme precipitation in the west coast of the U.S. from 1970-2007 and found increases over the coast of California. The authors suggested that the events of extreme precipitation of the west coast is depicted by natural variability as oppose to the influence of anthropogenic global warming [26]. Thus, the inconsistency from researches for the cause of extreme precipitation suggests a more detailed analysis of west coast variations should be taken. However, the method of studying precipitation patterns between rain gauges and remote sensing depends on the region of study.

1.4.1 Gauge-Measured Data Versus Remotely-Sensed Data

Remotely-sensed rainfall is one of the awaited measuring data that many researchers have been expecting. The Tropical Rainfall Measuring Mission (TRMM) precipitation data is currently one of the widely used remotely-sensed data available for free, which has a resolution of 5 kilometers, swath of 247 kilometers, detects rain and snow from the surface up to a height of 20 kilometers, and has an accuracy of measuring rainfall at 0.7 millimeters [4]. The successor of TRMM, Global Precipitation Measurement (GPM) is currently in orbit as of launch date in June 2014. It has a resolution of 250 meters, swath of 885 kilometers, and provides enough data to create a 4-Dimensional model [50]. Remotely-sensed data have an advantage over gauge-measured data since it is uniform and consistent spatially and temporally. Rain gauge, on the other hand, are becoming outdated, not uniformly distributed, and not kept with maintenance especially in open waters and unpopulated areas. These attributes may cause inconsistencies in coverage (spatially and temporally) [135]. Another disadvantage of gauge-measured data is at high elevations precipitation is in the form of snow which is usually under predicted since it can flow around the sensor (even more so if there is a strong wind present) [9]. A caveat of using remotely-sensed data is the resolution and how precipitation is detected. Precipitable water concentration is measured which means that the water does not necessarily reach the Earth's surface [135]. Nezhlin and Stein [135] found in their research a high correlation between remotely-sensed data and gauge-measured data. However, it was noted that satellite data measure precipitation much lower than gauge-measure. They concluded that remotely-sensed data is inappropriate for long-term regional hydrologic or climate modeling focused on trends and patterns of rainfall

in southern California [135] due to the amount of data that was collected over many years that is available through rain gauge instruments. A refined study of the gauge framework is needed to better analyze the spatial and temporal resolutions of the in-situ instruments or a higher resolution of satellite imagery (GPM in the near future). Therefore, this dissertation is an in-depth analysis of gauge-measured data to reflect ENSO occurrences in California's seven climate divisions.

2. First Study: In-Situ and Modeled Precipitation Measurements Analysis with SOI and PDO Using Wavelets

2.1 Introduction

Various works of literature have shown that the Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO) are strongly correlated to ENSO events [55, 56, 75, 133, 134, 165, 169, 171, 177, 189]. SOI is closely related to ENSO due to the ocean and atmosphere interaction in the central Pacific Ocean [55, 56, 165, 169, 171, 177, 189]. PDO measures the anomalous temperature in the Pacific Ocean, where El Niño is abnormally warm and La Niña is abnormally cold [75, 133, 134, 171]. The concept of this study is to understand and demonstrate whether each climate division in California has a strong signal with SOI and PDO indices in the power wavelet spectrum.

2.2 Data Used

We used hourly ground (in-situ) precipitation data from the National Climatic Data Center and NOAA's Forecast Systems Laboratory [131], where there are more than 2500 active stations and 7000 total stations. The dates for the stations in California range from

1948 through 2019. However, some stations outside of California begin as early as 1900. The hourly data extracted (735 stations from 1948 to 2019) concentrates on California’s seven climate divisions as shown in Figure 2.1. The data obtained has missing hours, days, and up to a few weeks. Therefore, the total precipitation of monthly data was calculated to solve the discontinuous problem by inputting missing values. We arithmetically averaged all stations in their respective climate divisions.

Monthly modeled precipitation data from NCDC gridded dataset [182] for California’s seven climate divisions were used to compare with the rain gauge (in-situ) measurements. The data begins from 1895 through 2019 for all climate divisions. This data set is an improvement of standard gauge data as it includes additional station networks and has quality assurance reviews. The extensive improvement is the computational approach for climatologically aided interpolation [182]. The data was derived from area-weighted averages of grid-point estimates interpolated from station data and any other observable means. A resolution of 5 km was used to ensure all divisions had sufficient spatial sampling and to address topographic and network variability [182]. To accurately compare both time-series, we cropped the monthly modeled data to the respective timestamp of the gauge data per climate division.

The Southern Oscillation Index (SOI) [66, 138, 168, 170] and the Pacific Decadal Oscillation (PDO) [79, 137] were used as references to the correlation between the signals of ENSO to compare the modeled and the rain gauge data set. Figure 2.2 shows the time-series of SOI and PDO with the gauge and modeled time-series.

Figure 2.1: Locations of the 735 stations of interest within California’s Seven Climate Divisions

We used Golden Gate Weather Services [39] for the years of El Niño and La Niña classification throughout this study as shown in Table 2.1. The determinations of each ENSO phase used the Oceanic Niño Index (ONI) as a three month mean of sea surface temperature anomaly of the Niño 3.4 region. ENSO events were identified by five consecutive three-month periods greater than or equal to $+0.5^\circ$ anomaly for El Niño events and less than or equal to -0.5° anomaly for La Niña [39]. The identification was enhanced to a different threshold

Figure 2.2: A) Model, Gauge, and SOI time-series with respective correlation for climate division one. B) Model, Gauge, and PDO time-series with respective correlation for climate division one

to classify each event as Weak, Moderate, Strong, and in some cases, Very Strong. ENSO events with three consecutive three-month periods with a 0.5 to 0.9 SST anomaly, 1.0 to 1.4, 1.5 to 1.9, and greater than 2.0 were classified as Weak, Moderate, Strong, and Very Strong respectively [39]. In this dissertation, we will focus on the cases of Very Strong for El Niño and Strong for La Niña.

2.3 Methodology

In order to model precipitation patterns to extract the anomalies, the gauge data must be consistent in time (hourly, daily, monthly, or yearly). As described before, from the inconsistency of the data, monthly sums were calculated from the hourly measurements.

Table 2.1: Event years of El Niño (red) and La Niña (blue) categorized as Weak, Moderate, Strong, and Very Strong. The table was obtained from Golden Gate Weather Services [39]

El Niño				La Niña		
Weak -11	Moderate -7	Strong -5	Very Strong -3	Weak -11	Moderate -4	Strong -7
1952-53	1951-52	1957-58	1982-83	1954-55	1955-56	1973-74
1953-54	1963-64	1965-66	1997-98	1964-65	1970-71	1975-76
1958-59	1968-69	1972-73	2015-16	1971-72	1995-96	1988-89
1969-70	1986-87	1987-88		1974-75	2011-12	1998-99
1976-77	1994-95	1991-92		1983-84		1999-00
1977-78	2002-03			1984-85		2007-08
1979-80	2009-10			2000-01		2010-11
2004-05				2005-06		
2006-07				2008-09		
2014-15				2016-17		
2018-19				2017-18		

Then we performed an average on the precipitation value over the same timestamp from all the gauges that belong in each climate division. Next, the data was converted from hundredths of an inch to inches to have the same units as the modeled data. The monthly precipitation was calculated over 71 years (1948 - 2019) and was implemented in a time-series decomposition to remove seasonal and trend components. This work used local linear regression (loess) [14] to remove such trends into a composition of a seasonal, trend, and residual components (STL decomposition) [28]. It is given by,

$$Y_t = S_t + T_t + R_t. \quad (2.1)$$

Equation 2.1 decomposes the time series (Y_t) into three distinct components. The seasonal and trend components are of no interest to this study, but the residual (R_t) is due to its fluctuation of anomalies. Moreover, an implementation of wavelets are used to analyze the stationary time-series (Y_t) and the residual (R_t). Figure 2.3 and Figure 2.4 shows the decomposition of the two time-series in climate division six (South Coast Drainage,

California). Figure 2.3 shows the gauge time-series decomposed into the trend, seasonal, and residual (remainder) components. Similarly, Figure 2.4 shows the decomposition but for the modeled time-series. This chapter will focus specifically on the residual components since they withhold the anomalous information.

Figure 2.3: Gauge time-series decomposition for climate division six

Figure 2.4: Modeled time-series decomposition for climate division six

2.3.1 Wavelet Function

Wavelet is a function that transforms any dataset (in this case, a time-series) into the frequency domain to reveal hidden features. Fourier, similar to wavelets, would have been an excellent choice to use; however, it does not report when the frequencies occur with respect to time. Wavelet transform, on the other hand, converts the time-series into time-frequency

space [41].

2.3.2 The Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) will be used since it is ideal for feature extraction purposes. The wavelet that is used in this work is the Morlet wavelet, as it is a great choice for feature extraction purposes [41, 70]. The Morlet wavelet is defined as:

$$\Psi_0(\eta) = \pi^{-\frac{1}{4}} e^{i\omega_0\eta} e^{-\frac{1}{2}\eta^2}, \quad (2.2)$$

where ω_0 is the dimensionless frequency and η is dimensionless time. The CWT applies the Morlet wavelet as a band-pass filter to the time-series. The CWT of a time-series $(x_n, n = 1, \dots, N)$ with uniform time steps δt , is defined as the convolution of x_n defined as,

$$W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n=1}^N x_n \Psi_0 \left[(n' - n) \frac{\delta t}{s} \right]. \quad (2.3)$$

The CWT has edge artifacts since the wavelet is not fully localized in time. Thus, it is best to introduce a Cone of Influence (COI) where the edges effects cannot be ignored [28]. Also, the statistical significance of wavelet power is retrieved from the null hypothesis that a stationary process generates the signal with a given background power spectrum (P_k). Grinsted et al. [41] suggests that much geophysical time-series can be modeled by a first-order autoregressive (AR1) process. The Fourier power spectrum of an AR1 process with

lag-1 autocorrelation α is given by,

$$P_k = \frac{1 - \alpha^2}{|1 - \alpha e^{-2i\pi k}|^2}, \quad (2.4)$$

where k is the Fourier frequency index. If the time-series is a stationary process with the power spectrum P_k , the variance at a given wavelet scale is the variance in the corresponding band P_k .

Grinsted et al. [41] adds that the probability of the wavelet power, which is a process with a given power spectrum (P_k), for which is higher than p is given as,

$$D \left(\frac{|W_n^X(s)|^2}{\sigma_X^2} < p \right) = \frac{1}{2} P_k \chi_v^2(p), \quad (2.5)$$

where v is equal to 1 for real and 2 for complex wavelets.

2.3.3 Cross-Wavelet Transform

The Cross-Wavelet Transform (XWT) of two time-series x_n and y_n is given as,

$$W^{XY} = W^X W^{Y*}, \quad (2.6)$$

where $*$ represents the complex conjugate. W^{XY} is interpreted as the relative local phase between x_n and y_n in time-frequency space. The distribution of the cross wavelet power of two time-series with background spectra P_k^X and P_k^Y is given by Grinsted et al. [41] as,

$$D \left(\frac{|W_n^X(s)W_n^{Y*}(s)|}{\sigma_X^2 \sigma_Y} < p \right) = \frac{Z_v(p)}{v} \sqrt{P_k^X P_k^Y}, \quad (2.7)$$

where $Z_v(p)$ is the confidence level associated with the probability of p , a pdf defined by the square root of the product of two χ^2 distributions.

2.3.4 Wavelet Coherence

Cross-wavelet power discloses areas with high correlation. The wavelet coherence measures how coherent the cross-wavelet transform is in the time-frequency space. Grinsted et al. [41] defines the wavelet of coherence of two time-series as,

$$R_n^2(s) = \frac{\left| S \left(s^{-1} W_n^{XY}(s) \right) \right|^2}{S \left(s^{-1} \left| W_n^X(s) \right|^2 \right) \cdot S \left(s^{-1} \left| W_n^Y(s) \right|^2 \right)}, \quad (2.8)$$

where S is a smoothing operator. Grinsted et al. [41] mentions that Equation 2.8 looks like the correlation coefficient, so it is best to think of the wavelet coherence as a localized correlation coefficient in the time-frequency space. The smoothing operator S is defined as,

$$S(W) = S_{scale} (S_{time} (W_n(s))), \quad (2.9)$$

where S_{scale} denotes smoothing along the wavelet scale axis and S_{time} smoothing in time. The Morlet wavelet has a smoothing operator replicated by Grinsted et al. [41] as,

$$S_{time}(W)|_s = \left(W_n(s) * c_1^{\frac{-t^2}{2s^2}} \right) \Big|_s, \quad (2.10)$$

$$S_{time}(W)|_s = (W_n(s) * c_2 \Pi(0.6s)) \Big|_n,$$

where c_1 and c_2 are normalization constants and Π is the rectangle function. The value of 0.6 is the empirically determined scale decorrelation length for the Morlet wavelet [41].

2.4 Results and Discussion

2.4.1 The Continuous Wavelet Transform

Referring to Figure 2.5 and Figure 2.6, they both show the CWT of gauge and modeled datasets for climate division six (climate divisions one, two, three, four, five, and seven and located in Appendix A). The black contour designates the 5% significance level against red noise. The cone of influence (COI) shown in white solid lines is where the edge effects might distort the picture shown outside the COI. The wavelet power spectrum images are essentially a three-dimensional image on a two-dimensional plane. It demonstrates the intensity of a signal over time and the reoccurrence rate. We will focus on the strong signals of precipitation during periods 2 to 8, which corresponds to the ENSO phase that occurs every 2 to 8 years.

Figure 2.5: Continuous Wavelet power spectrum of gauge residuals in climate division six

Figure 2.6: Continuous Wavelet power spectrum of modeled residuals in climate division six

We will analyze the isolated frequencies that are surrounded by the black contour. We will focus on the three Very Strong El Niños of 1982/1983, 1997/1998, and 2015/2016, as shown in Table 2.1. Almost all climate divisions have successfully shown a strong signal of precipitation within the period of 2 to 8 years for both modeled and gauge datasets. The remaining contours are all associated with the remaining categories of El Niño for Strong, Moderate, and Weak. The contours that are close to the zero-period represent cases of an infrequent high frequency. Such cases are related to extreme precipitation that is associated

with flooding.

As previously stated, all climate divisions for the modeled and majority of the gauge data sets well represent ENSO. Climate division six for the gauge dataset is the only case where an El Niño event was not represented. As the gauge dataset represents the accurate precipitation measurements, we can see in Figure 2.5 that for the ENSO event of 2015/2016 there is no strong frequency significance. During the 2015/2016 timeframe, climate scientists were addressing that it was a strong El Niño year as all the right conditions were met [54, 72, 82]. This was by referring to all Niño indices to show there was an extremely warm tongue in the Pacific Ocean and the Southern Oscillation Index (SOI). Although this may be true and all climate divisions for the modeled dataset and most of the gauge dataset did represent such El Niño phase, it was untrue for the gauge climate division six dataset. The Niño indices and SOI represent a global representation of ENSO but do not necessarily imply specific locations will experience extreme precipitation. Le et al. [71] demonstrated/forecasted in their paper using Palmer Z-indices (PZI) that climate division six will not experience heavy precipitation for 2015/2016 El Niño.

Golden Gate Weather Services [39] published on their website the percentage of normal rainfall in California. An image of average rainfall is reproduced in Figure 2.7. A value of 100% means California is within a typical region of precipitation. Anything above or below 100% represents excessive precipitation or below-average precipitation (drought possibly), respectively. Figure 2.7 shows that in the area of climate division six, the majority of the measured precipitation values are below 100%. They are almost as low as half the usual

amount of precipitation, given that it is considered a Very Strong El Niño year. An interesting observation is given the classification of El Niño, one would expect the amount of excessive rainfall to be much more significant than 100%. This is related to the complexity of the environmental fields. Even though there is a Very Strong El Niño event, there may be a similar strength event that diminishes the magnitude of the heavy precipitation [84]. Based on the results of the amount of precipitation received in 2015/2016 for California in various regions, we can concur with the results from the wavelet analysis and Le et al. [71] study.

Figure 2.7: Reproduced image from Golden Gate Weather Services [39] that shows the percentage of normal rainfall in California

The CWT of SOI and PDO are displayed in Figure 2.8 and Figure 2.9. The wavelet plot for SOI shows some signals are occurring in the periods 2 to 8. More specifically, there is a presence of a strong signal for the 1982/1983, 1997/1998, and 2015/2016 El Niños. Furthermore, the PDO plot shows a similar pattern as SOI for periods 2 to 8. However, no extreme incidents of high frequency occur after 1975. A more detailed analysis will be conducted in the later section discussing the correlation between the two time-series (gauge and modeled) with SOI and PDO.

Figure 2.8: Continuous wavelet power spectrum of SOI

Figure 2.9: Continuous wavelet power spectrum of PDO

2.4.2 Cross-Wavelet Transform

2.4.2.1 Climate Division 6

In this chapter, we will compare in full detail the results between the gauge and modeled residuals for climate division six. A comparison of the gauge and model original time-series was done but did not provide as much information as the residuals. The residuals are not

biased to the strong correlation of seasonal and trend components as the raw time-series. The cross-wavelet transform will show the correlation of anomalies over time. Figure 2.10 shows the cross-wavelet transform of the model and gauge in climate division six. To reiterate, anything that is in red has a high correlation; specifically, those in the 5% that are enclosed in a thick black contour. One of the expectations of the comparison between the two time-series is that most of the wavelet should be in red. It is evident that it is not and that there exist many inconsistencies. However, there are some similarities over the entire time-series, but they are sparse between the periods 2 to 8. Also, the arrows suggest the type of similarity they have [40]. Arrows that are pointing to the right signify that the two time-series are in phase. Arrows that point to the left mean that they are in anti-phase. Arrows that point up mean that the model leads the gauge time-series by $\frac{\pi}{2}$. Arrows pointing down mean that the gauge leads the model by $\frac{\pi}{2}$. The lead value $\frac{\pi}{2}$ means that there is a lag of $\frac{\pi}{2}$ in the frequency domain [41]. It is a similar concept when we calculate a cross-correlation of a time-series and one of the time-series lags by a certain amount of units in the time domain. Referring to Figure 2.10, one can see the arrows in the contours are only pointing to the right. This suggests that the time-series are in phase in the frequency domain. Furthermore, the El Niño years of 1982/1983, 1997/1998, and 2015/2016 are highly correlated between the modeled and gauge datasets, as they are shown as a strong signal in the periods 2 to 8.

Figure 2.10: Cross-Wavelet Transform of modeled and gauge residuals in climate division six

Figure 2.11 shows that there are similarities with both time-series within 2 to 8 periods. There is a mixture of the direction of the arrows, but it is consistently pointing the left, which signifies there are in anti-phase. Furthermore, the cross-wavelet transform shows the arrows pointing downwards at times which means that the gauge SOI leads the gauge precipitation. The contours show where there is the highest similarity. The time location of the contour matches the period of reported heavy precipitation in climate division six for El

Niño years 1982/1983 and 1997/1998. However, there is a small cell for 2015/2016 that is in the border of period eight, which is on the outskirts of ENSO. Compared to the cells for El Niño years 1982/1983 and 1997/1998, it is clear that El Niño is present, but not as strong in this division. Golden Gate Weather Services [39] showed that climate division six did not experience the expected rainfall as shown in Figure 2.7. After the events unfolded, many climate scientist studied the cases of the two strongest El Niños (1997/1998 and 2015/2016) to attempt to postulate why the 2015/2016 El Niño was not as strong [71, 160, 173, 175, 198]. They claimed it was due to climatic opposing forces and not all the right conditions were met, such as central Pacific El Niño dynamics were not aligned in the increasing stages of El Niño but at the decaying phase [160].

Figure 2.11: Cross-Wavelet Transform of gauge residuals and SOI in climate division six

The comparisons of gauge and PDO have distinctly different cross wavelet transform, as seen in Figure 2.12. First and foremost, there are more strong correlations signals between the gauge data set and PDO than compared to SOI. However, one that is strongly noticeable is the similarity within period 2 to 8 for the El Niño 1965/1966. Surprisingly, the Very Strong El Niños are not visibly strong as they were with SOI. In addition, the arrows within the plot show that there are out of phase by approximately 45° . Comparing the two plots of Figure 2.11 and Figure 2.12, it is clear that the gauge dataset for climate division

six is strongly correlated with SOI than PDO.

Figure 2.12: Cross-Wavelet Transform of gauge residuals and PDO in climate division six

The model residual time-series compared with SOI shows a similar result to the gauge residual time-series, as seen in Figure 2.13. There is a substantial similarity with SOI for periods 2 to 8, with some being very strong, as shown by the contours. The majority of the arrows are pointing leftwards, which again signifies there is anti-phase. The contours also depict the time locations of anomalies of high precipitation. Moreover, the modeled residual

time-series shows a consistent similarity to PDO as with SOI as seen in Figure 2.14. The directions of the arrows are the same as the SOI plot. In the first half of the timespan and within periods 2 to 8, the modeled dataset has a higher amount of correlation with PDO. In the second half of the timespan, the modeled dataset has a stronger correlation with SOI. Overall, it appears that the climate division six gauge residual dataset is synchronized with SOI and the modeled residual time-series in the present also with SOI. In literature, there is a mixture of results on which climate index is strongly correlated in the north California versus the south [75, 169, 186].

Figure 2.13: Cross-Wavelet Transform of modeled residuals and SOI in climate division six

Figure 2.14: Cross-Wavelet Transform of modeled residuals and PDO in climate division six

2.4.2.2 Climate Divisions 1, 2, 3, 4, 5, and 7

The correlation of anomalies in the remaining climate divisions (1, 2, 3, 4, 5, and 7) for the modeled and gauge datasets are located in Appendix A. They all demonstrate a strong signal for El Niño years of 1982/1983, 1997/1998, and 2015/2016 as they both concur. The remaining strong signals that lie within the period of 2 to 8 years also match with Weak, Moderate, and Strong El Niños from Table 2.1 [39].

The cross wavelet transform of Gauge and SOI shows the anomalies that intertwine with both time-series. The Very Strong El Niños match up perfectly in correlation within the periods of 2 to 8 years. Furthermore, there is a high frequency associated with 2015/2016 El Niño, which was not present in climate division six. Historically, the 2015/2016 El Niño was one of the most influential El Niño events recorded and it is clearly noted by the high frequency at a low period, which shows the rarity of the event) [11, 54]. The high-frequency contour is an indicator of flooding events that match with Figure 2.10. The relationship between Gauge and SOI is an anti-phase due to the arrows pointing majority to the left. Moreover, there is a definite anomaly correlation centered around 1951/1952 in climate division three that was not as visible in the other climate divisions. Reno Gazette Journal [167] reported the top ten snowiest and wettest years in Reno, which is neighboring to climate division three. The year 1951/1952 was number one on their list for the snowiest years (59.3 inches) and number eight for the wettest years (11.19 inches) that they have experienced. The large significant anomaly shows this in the CWT that spans from a 0 to 12 period. The frequency goes to zero due to the rare extreme of the precipitation event as it was noted in Reno Gazette Journal [167] as the snowiest year they have experience. PDO shows a strong anomalous correlation for 1951/1952, but it is not as strong as SOI. The climate index SOI has a sturdy overall correlation with the western part of the United States and PDO with the central United States [68].

The CWT of gauge and PDO is like gauge and SOI, where there is a high signal during Very Strong El Niños. However, there is no presence of a high frequency for the 2015/2016

El Niño. Instead, there is a high frequency located only for the year 1995/1996. During this timeframe, there is a moderate La Niña. The high-frequency occurrence is most likely related to PDO having a stronger relationship with droughts (La Niñas) [172].

The modeled residual dataset versus SOI residual dataset demonstrates there are matches for the Very Strong El Niños as well as the Strong, Moderate, and Weak. Precisely like the gauge residual dataset versus SOI, we see there is a high-frequency for El Niño for 2015/2016 due to the intensity of the precipitation. Furthermore, the strength of the 1951/1952 has diminished when compared to the gauge. The result is surprising because of the construct of the modeled time-series, which is a similar construct that uses gauge datasets [181].

The cross-wavelet transformation of modeled and PDO dataset has a strong correlation for anomalies the 1982/1983 and 1997/1998 El Niños. However, the presence of the 2015/2016 El Niño is nonexistent. The highest significance is just outside the eighth period, which is outside the window for ENSO (2 to 8 year period). Just like the gauge and SOI, we see a high frequency again for La Niña in 1995/1996, which hints again about PDO's relationship with droughts [172]. the El Niño of 1951/1952 is not as strong, but it is still significant. However, both gauge and modeled residual time-series have a higher significant level for El Niño of 2015/2016 with an anti-phase. Thus, the results found by Kurtzman and Scanlon [68] and Shukla et al. [172] are not absolute. SOI and PDO in the wavelets we have seen thus far have produced relatively strong anomalous correlations.

2.4.3 Wavelet Coherence

2.4.3.1 Climate Division 6

The next analysis done is the wavelet coherence between all the time-series. It is the same concept of correlation but in the frequency domain. The comparison in this analysis is with the gauge and modeled residuals in climate division six. Figure 2.15 shows that there is a powerful correlation between the two time-series (given by the dark red). Anything on the light side coloring, such as blue, shows there is no correlation. Since the majority of the frequencies are correlated, we will address the ones where there is little correlation. One example of such a case is the period between 8 and 16 around the year 1952. It may be due to the deviation of the modeled data given a particular initial condition, while the gauge data is always the accurate precipitation data. The calculation of early precipitation values for the gauge time-series is low due to the number of in-situ measurements. However, as time progresses, the number of in-situ instruments increases and that is where the gauge matches the modeled time-series since the modeled time-series takes measurements from various instruments [182]. Figure 2.15 supports this by the arrows pointing to the right (where the residuals are in phase).

Figure 2.15: Wavelet Coherence of modeled and gauge residuals in climate division six

The same method was done on the gauge residual time-series. The wavelet coherence showed that there are a few times where there is a strong correlation between the two, as shown in Figure 2.16 and Figure 2.17. The gauge dataset and SOI are not entirely related to one another. It is noted by the direction in which the arrows point (upwards and downwards), implying a phase shift of 90 degrees. The years of the larger El Niños are aligned, which shows that at times SOI may and may not be the best candidate for a predictor of rainfall for California's climate division six. The points where there is a strong correlation are between

periods 4 and 8. It is believed to be the correlation due to the ENSO phase. In addition there are also strong correlations in later periods, 32 to 128, which may be due to the powerful storms and droughts that occur less frequently than the average wetter and dryer seasons.

Figure 2.16: Wavelet-Coherence of gauge residuals and SOI in climate division six

Similar results are found when comparing the gauge residual time-series to PDO. Everything applies as it was discussed for SOI, but there are some grand differences. As it was discussed before, SOI has a stronger correlation than PDO. In Figure 2.17, it shows there is

a mismatch during the strong El Niño years. In general, there is little correlation with PDO. Some notable correlations are in period 8 around 1956 and period 4-8 in 2002. The two years are a mixture of events between El Niño and La Niña, respectively. Thus, extreme events in found in PDO have an inconsistent pattern with ENSO events. The remainder of the strong correlations is out of the scope of the conic sections and is therefore unreliable. An interesting observation between the two comparisons is that they show the same strong correlation both in period and time. The findings demonstrate that SOI and PDO both are related to precipitation but SOI is more strongly linked.

Figure 2.17: Wavelet-Coherence of gauge residuals and PDO in climate division six

Lastly, the modeled residual time-series were correlated with SOI and PDO. Figure 2.18 and Figure 2.19 shows the resulting correlation. Referring to Figure 2.18, one can see a similar result as we did with the gauge. A noticeable difference is that the gauge dataset has a strong correlation in the same areas as the modeled dataset. It is a negative contribution to the modeled as it is not characterizing the attributes of the precipitation phenomena accurately. In retrospect, it shows that there is a decent correlation every 4 to 8 periods, which agrees with the ENSO cycle.

Figure 2.18: Wavelet-Coherence of modeled residuals and SOI in climate division six

Moreover, the PDO coherence with the model shows similar results to the coherence with SOI. This one again shows a weaker correlation to SOI within the middle of the conic section. Though towards 2015/2016 around period 2, there is a small strong cell that can signify a delayed connection to ENSO in 2015/2016. Figure 2.19 shows very little correlation to one would expect to have an ENSO like correlation since there are few instances of period 2 to 8. Overall, one can agree that the gauge does a better job matching the ENSO phase

when it is specifically compared to SOI.

Figure 2.19: Wavelet-Coherence of modeled residuals and PDO in climate division six

2.4.3.2 Climate Divisions 1, 2, 3, 4, 5, and 7

The analysis for the wavelet coherence for Climate Divisions 1, 2, 3, 4, and 7 are located in Appendix A. The comparison with the gauge and modeled residuals shows that there is a powerful correlation between the two time-series (given by the dark red) for al-

most all climate divisions. Anything on the light side coloring, such as blue, shows there is no correlation. Climate division three and seven have many timestamps that are not well correlated like the rest of the divisions. This could be the low representation of gauges (18 for climate division three and 60 for climate division seven) and the difference in terrain from the mountains and to the deserts. Some of these key differences occur during El Niño events, which unfortunately faults the forecasting if we cannot well represent the region.

The same method was done on the gauge residual time-series with SOI. The wavelet coherence showed that there is a correlation between all climate divisions during the strong ENSO events. However, the majority of the time, there is a positive time lag, which agrees with the direction of the arrows majority pointing to the left. Such direction implies that they are in anti-phase with gauge precipitation. Furthermore, the the arrows within COI point upwards, which means there is a lag in the frequency domain, which we can translate as a lag in the time scale as well.

Similar results are found when comparing the gauge residual time-series to PDO. For all climate divisions, there are signals present for El Niño but varied in minor strength. However, the El Niño of 2015/2016 has a consistently strong signal for the northern climate divisions. For climate divisions five and seven, the signal was present, but not as strong as the other climate divisions. This may be related to the association that PDO may have with droughts, as that is what was experienced in southern California [172].

The wavelets for the modeled data sets versus SOI shows consistency with the majority

of the climate divisions. For climate division one, there is no signal present for the strong El Niño of 1997/1998. Literature does not mention any localized precipitation, just general of the whole California. This finding demonstrates that given that the timespan overlaps a strong El Niño, it does not mean that they will be rainfall in every location.

The analysis of the modeled data sets versus PDO shows there is consistency with all climate divisions and ENSO events. The modeled time series does not vary as the gauge and this is due to the construct of the data set [182].

2.4.4 Wavelet Spectra

The last comparison was to use the built-in function `wdist` from the `BiWavelet` package [40] in R Statistical Program [164] to calculate the dissimilarity between two wavelet spectra. The first comparison was made between the model and gauge residuals for all seven climate divisions. A value close to zero is accepted to be very similar. The larger the value, the higher the dissimilarity is between the two wavelet spectra `BiWavelet`. Table 2.2 shows the spectra value between the modeled and gauge residual time-series. Climate division four has the lowest value of 392.4202, which signifies that in terms of the spectral frequency, the modeled and gauge residual values are the most similar. On the contrary, climate division six has the highest wavelet spectra value of 533.0569. Between the modeled and gauge residual values, the modeled dataset does not appropriately resemble the gauge dataset. Furthermore, none of the climate divisions are close to zero which illustrates the lack of precision on the modeled data set.

Table 2.2: Modeled and Gauge Residual Time-Series Wavelet Spectra

CD	Value
1	431.0109
2	399.5148
3	435.6474
4	392.4202
5	403.3535
6	533.0569
7	424.8643

The second comparison is with the gauge residual time-series with SOI and PDO. Table 2.3 lists the spectra value between the gauge and SOI residual time-series. The table shows that SOI has a stronger frequency correlation with the southern climate divisions of California. The relationship between SOI and southern climate divisions is documented in current research, such as from Fierro for SOI having a stronger relationship with SO-CAL [31, 74] and in Rodriguez et al. [169].

Table 2.3: Gauge and SOI Residual Time-Series Wavelet Spectra

CD	Value
1	668.5899
2	657.3492
3	561.8172
4	543.3671
5	555.1155
6	595.3408
7	432.8211

The comparison between gauge residual time-series and PDO, Table 2.4, shows similar values when compared to SOI with a couple of differences. The first difference is the

maximum spectra value. PDO has a max spectra value of 588.4987 while SOI has 668.5899. This implies there is a closer frequency relationship between the gauge residual time-series with PDO residual time-series than SOI. Furthermore, there is no consistent pattern as SOI, where the spectra values are decreasing from Northern California to Southern California. This could be due to the nature of how PDO is calculated and is based on a cycle of more than 30 years [69]. However, we do see that the lowest spectra value resides with climate division 7 for both SOI and PDO.

Table 2.4: Gauge and PDO Residual Time-Series Wavelet Spectra

CD	Value
1	561.9301
2	588.4987
3	467.7484
4	507.7063
5	476.8725
6	529.6231
7	355.0294

The last comparison is with the modeled residual time-series with SOI and PDO demonstrated in Table 2.5. First and foremost, we see a reduction in spectra values from North to South California as we did with gauge residual time-series and SOI. One apparent observation is that most of the spectra values are almost twice as large when compared to the gauges, while a few are lower. For example, from Table 2.5, we can assess that the modeled residual time-series (385.6459) has a stronger correlation with climate division seven than the gauge residual time-series (424.8643). This is most likely due to the construct of the modeled time-series as it incorporates SOI into its calculation. Furthermore, we have established

from previous literature that SOI has a strong influence on southern climate divisions [31, 74].

Table 2.5: Modeled and SOI Residual Time-Series Wavelet Spectra

CD	Value
1	657.8282
2	626.4349
3	501.0882
4	515.2916
5	499.3537
6	508.2890
7	385.6459

The comparison between modeled residual time-series and PDO, Table 2.6, has an overall lower wavelet spectra value for all climate divisions. It appears that the modeled time-series is far strictly related to PDO than SOI. In addition, there is a clear decreasing value from north to south that was not evident when the modeled time-series was compared with SOI. The modeled time-series in its acquisition appears to closely resemble the algorithm in which PDO was calculated and implemented in its precipitation measurements.

Table 2.6: Modeled and PDO Residual Time-Series Wavelet Spectra

CD	Value
1	613.7658
2	555.8268
3	429.1110
4	504.9540
5	475.9818
6	452.6316
7	325.8304

2.5 Conclusion

Forecasting atmospheric hazards using historical precipitation data is what scientists/researchers are attempting to do well enough to be able to avoid catastrophic events. However, the usage is different from the researcher and location. Therefore, this study focused on historical hourly precipitation data to model extreme precipitation (related to ENSO) to have a proper input for other models (climate, hydrological, crop, and etcetera) to ensure the public with food, water, and shelter. The data consisted of 71 years of hourly observation at various regions in California and was translated into monthly data to keep consistent with modeled data. The data provided insight into the hidden patterns using wavelet analysis and correlation with the gauge, modeled, SOI, and PDO time-series.

The results from the wavelets analysis have overall shown that the gauge and modeled data set for the majority are very similar but are different in their overall representation of precipitation patterns in California. The algorithm that was used to populate the data set for the modeled time-series incorporate area weights that are beyond the climate divisional boundaries, include data sets from other observable means than gauge, and the method used to populate missing data. The differences are numerically clear with the high values obtained from the spectra analysis. The key differences have shown how well they are able to represent ENSO.

SOI well represents the precipitation patterns in California compared to PDO for both gauge and modeled data sets. However, different climate divisions and climate indices re-

sponded better to the general concept of SOI as the key representation of California. Thus, the study found that it is naive to to state that SOI is the swole representation of precipitation and ENSO. This study can be extended to other climate indices and possibly find stronger indices that are related to each climate division.

3. Second Study: In-Situ and Modeled Precipitation Measurements Analysis With SOI, PDO, AMO, and Niño Indices Using Autoregressive-Moving-Average Model and Empirical Mode Decomposition

3.1 Introduction

A few years ago, from the date of this dissertation, it was reported by the National Oceanic and Atmospheric Administration (NOAA) that there is a very high chance not only for El Niño to continue through Northern Hemisphere winter 2015-16, but also a remarkable chance for El Niño to last into early spring 2016, due to the analysis of all the popular climate indices (SOI, PDO, etcetera) met the right conditions [83, 85, 86, 176]. This research aims at 1) investigating the impact of El Niño on precipitation on California's seven climate divisions 2) Analyzing the precipitation of California using the Empirical Mode Decomposition Method (EMD) 3) Comparing precipitation data with SOI, AMO, and PDO.

3.2 Data Used

We used hourly ground (in-situ) precipitation data from the National Climatic Data Center and NOAA’s Forecast Systems Laboratory [131], where there are more than 2500 active stations and 7000 total stations. The dates for the stations in California range from 1948 through 2019. However, some stations outside of California begin as early as 1900. The hourly data extracted (735 stations from 1948 to 2019) concentrates on California’s seven climate divisions as shown in Figure 1.1. The data obtained has missing hours, days, and up to a few weeks. Therefore, the total precipitation of monthly data was calculated to solve the discontinuous problem by inputting missing values. We arithmetically averaged all stations in their respective climate divisions.

Monthly modeled precipitation data from NCDC gridded dataset [182] for California’s seven climate divisions were used to compare with the rain gauge (in-situ) measurements. The data begins from 1895 through 2019 for all climate divisions. The data was derived from area-weighted averages of grid-point estimates interpolated from station data. A resolution of 5 km was used to ensure all divisions had sufficient spatial sampling. To accurately compare both time-series, we cropped the monthly modeled data to the respective timestamp of the gauge data per climate division.

The Southern Oscillation Index (SOI) [66, 138, 168, 170], the Pacific Decadal Oscillation (PDO) [79, 137] 1997), and the Atlantic Multidecadal Oscillation (AMO) [29, 136] indices were used as references to the correlation between the signals of ENSO to compare

the modeled and the rain gauge data set. Figure 2.2 shows the time-series of SOI and PDO with the gauge and modeled time-series.

The Southern Oscillation index is calculated using the difference in sea level pressure between Tahiti Island (largest Island of French Polynesia) and Darwin in Australia [66, 168, 170]. The data used in this study begins from 1866 to 2019.

The Pacific Decadal Oscillation index was developed by Hare and Francis [46] and Zhang [200] as the first-order principle component from an unrotated EOF analysis of monthly differences between observed anomalies and mean global average North Pacific sea surface temperature anomalies [77, 199]. A feature of PDO is that it remains relatively constant for multiple years and decades with little sign changes [47, 77, 78, 88, 199]. Furthermore, PDO has a substantial influence in various areas of North America [77]. The dates used in this study begins from 1900 to 2018.

The Atlantic Multidecadal Oscillation (AMO) was first coined by Kerr [57]. AMO signal is global and is positively correlated with oscillations in the North Pacific and strongly with the North Atlantic [29]. When AMO has a positive phase, the United States has below normal rainfall. During the winter, AMO has a convincing changed for the interannual rainfall variability associated with El Niño Southern Oscillation [29]. The dates used in this study begins from 1856 to 2019.

3.3 Methodology

The rain gauge and modeled precipitation time-series was decomposed to remove seasonal and trend components. This work used local linear regression (loess) [14] to remove such trends into a composition of seasonal, trend, and residual components (STL decomposition) using Equation 2.1 [28]. The seasonal and trend components are of no interest to this study, but the residual (R_t) due to its fluctuation of anomalies.

3.3.1 ARMA

An implementation for analysis of the stationary time-series autoregressive-moving average (ARMA) model was used to analyze the residual component. Autoregressive-moving average model is defined as,

$$R_t = \sum_{i=1}^p \alpha_i R_{t-i} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}, \quad (3.1)$$

where α_i , β_j , and ε_{t-j} are the parameters of the ARMA model. The error to this model is given by the following properties: $\varepsilon_{t-j} \sim N(0, \sigma^2)$, $cov(\varepsilon_t, \varepsilon_{t-j}) = 0$, and $cov(\varepsilon_t, y_{t-j}) = 0$ [28]. This model can be rewritten to establish a maximum likelihood error as defined in El-Askary et al. [28],

$$\phi_\alpha(L)R_t = \phi_\beta(L)\varepsilon_t, \quad (3.2)$$

where $\phi_\alpha(z) = 1 - \sum_{i=1}^p \alpha_i z^i$, $\phi_\beta(z) = \sum_{j=0}^p \beta_j z^j$, and $LR_t = R_{t-1}$, which is the lag operator. Using equation (3.2), one can obtain the maximum likelihood estimator for the residuals (R). Approaches to the fitting the ARMA models to the residuals are based on Akaike Information Criterion (AIC) minimization. The AIC model is given by,

$$AIC(\phi_\alpha, \phi_\beta, \sigma^2) = -2 \ln[L(\phi_\alpha, \phi_\beta, \sigma^2)] + p + q + 1. \quad (3.3)$$

The AIC model is optimal when the forecast mean square error (FMSE) measure is no more than a certain threshold [28].

Once the autocorrelation is calculated using equation (3.2), a derivation of 95% confidence band is used to detect extreme values,

$$\hat{R}_t \pm 1.96 \sqrt{MSE(\hat{R}_t)}, \quad (3.4)$$

where $\hat{R}_t$ corresponds to the residual output from the ARMA model and MSE is the mean square error. Equation (3.4) reflects the confidence intervals of anomalies beyond the common observations in the remainder decomposition of the data. Thus the extreme observations (E) are given by,

$$E = \{y_t | r_t > \hat{R}_t + 1.96 \sqrt{MSE(\hat{R}_t)}\} \cup \{y_t | r_t < \hat{R}_t - \sqrt{MSE(\hat{R}_t)}\}, \quad (3.5)$$

where y_t are time dependent observations and r_t are time dependent residual values from Equation (3.2). This methodology that was perform in El-Askary et al. [28] is used in this dissertation to model precipitation patterns and extract ENSO anomalies in California. We used the R statistical software [164] and the `arma` function in the package `forecast` [52, 53].

3.3.2 EMD

The empirical mode decomposition (EMD) is designed to analyze efficiently nonlinear and nonstationary properties of time-series data [23, 38, 51, 58]. The concept of EMD is to decompose a time-series into intrinsic modes of oscillations. The modes are referred to as intrinsic mode functions (IMFs). The algorithm of EMD decomposes time-series into intrinsic oscillations using local temporal and structural characteristics of the data. We use the empirical mode decomposition to decompose all gauges, modeled, and climate indices into local orthogonal modes.

There are various algorithms for an EMD [32] . However, we will explain the algorithm used in Khademul Islam et al. [58]. Given time-series data $x(t)$, it can be decomposed by EMD in the following steps:

1. Identify the extreme (maxima and minima) of data $x(t)$;
2. Generate the upper and lower envelopes $h(t)$ and $l(t)$, respectively, by connecting the maxima and minima points separately with cubic spline interpolation;

3. Determine the local mean $m_1(t) = \frac{(h(t) + l(t))}{2}$;
4. IMF should have a zero-local mean, subtract out m_1 from $x(t)$

$$g_1(t) = x(t) - m_1(t); \quad (3.6)$$

5. Test whether $g_1(t)$ is an IMF or not;
6. Repeat steps 1 to 5 and end up with an IMF $g_1(t)$.

When the first IMF is calculated, we can define $C_1(t) = g_1(t)$ as the shortest period component of the time-series $x(t)$. To calculate the remaining IMFs, we calculate the residual $r_1(t)$ from the time-series by subtracting $C_1(t)$ from it as,

$$r_1(t) = x(t) - C_1(t). \quad (3.7)$$

The residual $r_1(t)$ contains information about the components of more extended periods. The residual time-series $r_1(t)$ will be treated as a brand-new time-series and resifted to find any additional components. The sifting process (finding new components) will continue until the stopping criteria described in Flandrin et al. [32] is satisfied. The subsequent IMFs are $r_2(t) = r_1(t) - C_2(t)$, $r_3(t) = r_2(t) - C_3(t)$, $\dots$,

$$r_n(t) = r_{n-1}(t) - C_n(t), \quad (3.8)$$

where $r_n(t)$ is either a constant, a monotonic function, or a function with only maxima and one minima from which no more IMFs can be derived [51, 58, 59]. The original time-series $x(t)$ can be expressed as a sum of all the residuals/signals (IMFs):

$$x(t) = \sum_{i=1}^n C_i(t) + r_n(t), \quad (3.9)$$

where n is the number of IMFs and $r_n(t)$ is the final residual. Figure 3.1 shows the decomposition of the gauge climate division one time-series. The first row represents the original time-series. Rows two to eleven are the ten distinct decompositions of the original time-series. The last row is the residual; What remains after removing all ten decompositions from the gauge time-series.

Using empirical mode decomposition to decompose precipitation time-series is ideal because it can remove harmonic components [58]. Furthermore, EMD is very useful with climate data because the IMFs components are normally distributed and the energy density of the IMFs are χ^2 -distributions [58, 187]. We used the R statistical software [164] and the package **EMD** [60] for the function **emd** [61, 62, 63] to calculate the empirical mode decomposition of all the data used in this study.

An essential part of the assessment is to know what each component represents. According to Wu et al. [188], they interpret the meaning of each component for SOI only, however, a similar assumption can be made for other indices. The first component, c_1 , is related to physical phenomena or simply noise in the time-series. The second component is

Figure 3.1: EMD of gauge time-series for Climate Division 1. Decomposition of 10 IMF components and the final residual

known as a semiannual cycle given from the migration of the sun that crosses the equator twice a year. The third component is the annual cycle of Tahiti and Darwin pressures. The fourth component, known as a biennial cycle, is the nonlinear response of the tropical atmosphere to the seasonally varying force. In addition, it can also be the interaction between South Asian Monsoon and large-scale circulations in the tropics and extra tropics. The fifth and sixth components are considered as an interannual mode varying on average 3.3 years and 5.9 years, respectively. They represent the oscillations induced by atmosphere-ocean interaction. Components greater than 7 are decadal modes that are possibly related to the atmosphere-ocean interaction of longer timescales. In addition, they are responses of the

tropical Pacific due to decadal forcing. We discuss the results of such comparison in the later sections.

3.4 Results and Discussion

3.4.1 ARMA

The data obtained from NCDC and NOAA are given as an argument to a time-series function `stlplus` [43] in R statistical software [14, 164] to remove the seasonal, trend, and residual components from the data. Figure 2.3 and Figure 2.4 show the various components of the gauge and model data for climate division six. Recall the top row represents the raw data obtained from NCDC and NOAA. The second row is the seasonal component, which shows a consistent pattern throughout the years of the study. This is calculated by averaging each month for the 71 years and subtracting each monthly value from the mean. Following the seasonal component is the trend component. The trend component is calculated as a linear equation that best fits the raw data and each monthly value is again subtracted from the trend that was just calculated [14]. This plot shows an increase or decrease of rain over the years. Lastly, row four shows the main interest of the study, the residual component. As stated before, it shows and grants us the ability to detect the anomalies of precipitation over the years that may be due to ENSO processes. All the computational modeling was performed using the R statistical software [164].

Referencing Figure 2.3 and Figure 2.4, the raw data have similar traits where they both obtain their major peaks around the same time frame. However, the magnitude is of great difference by 5 inches. This high amount is crucial since it would be the determinant of flooding in a crowded city. The seasonal component for both are similar. The trends are not vastly different but show some distinctions between the model and gauge in terms of number of peaks around 1985. The gauge shows around 1985 that there were three inflation points as oppose to the model depicting two. A second observation to note is again the magnitude of the trend. It is much stronger for the gauge measurement. The residual (also known as remainder) is discussed in the following section. Similar results can be found in the remaining climate divisions.

3.4.1.1 Extreme Precipitation

Extreme events are those that lie beyond the red lines, as they rely outside the boundaries of the 95% confidence interval. Anything that is above the red line is considered as an event of extreme precipitation. In respect to the negative red line, anything below is considered as an abnormal drought occurrence. Comparing the patterns of Figure 3.2 and Figure 3.3, which corresponds to gauge and modeled respectively for climate division six, one can see that they have a distinct range of frequency. It should be noted that the gauge precipitation measurement are given in hundredths of an inch while the model is in inches. Focusing on the extreme precipitation events, there is a higher frequency in the model than the gauge data, within the timestamp of the gauge data set. From those that are similar

between both plots, the gauge data displays a higher frequency than the modeled. Table 3.1 shows the exact dates of high precipitation for both the gauge (left) and modeled (right) based on the ARMA model. One of the first corresponding dates from the two data sets is March 1952. The remainder value in comparison to the confidence interval band for the gauge is much higher than the modeled data (not show in the table). The values described is the difference between the frequency and the positive confidence interval. This alludes to the gauge shows a stronger event for this particular time. One of the strongest El Niño events of 1997/1998 was found in both data sets on February 1998. However, our results shows that the very strong El Niño of 2015/2016 appears in the modeled data sets on January 2017, but does not appear for the gauge data sets. This again concurs with the findings of Le et al. [71] for climate division six.

The climate divisions 1, 2, 3, 5, and 7 have different findings than climate division six. They all concur with the three very strong El Niños, as they are properly mapped out within their appropriate time frame. Climate division four shows the same results as climate division six. There is no evidence from the ARMA model that there was a strong El Niño in 2016/2017 for the gauge data set. However, the same cannot be said for the modeled time-series. This could be due to the effects of the ocean and land dynamics as climate divisions four and six are coastal lands [80, 198].

Figure 3.2: Gauge monthly residual plot for climate division six

Figure 3.3: Modeled monthly residual plot for climate division six

Table 3.1: Excessive precipitation dates in climate division six

Gauge	Modeled
Jan 1952	Jan 1895
Mar 1952	Nov 1900
Jan 1954	Feb 1905

Continued on next page

Table 3.1 – *Continued from previous page*

Gauge	Modeled
Mar 1954	Mar 1906
Jan 1957	Jan 1907
Feb 1962	Mar 1907
Apr 1965	Jan 1909
Nov 1965	Dec 1909
Dec 1965	Jan 1911
Dec 1966	Mar 1912
Apr 1967	Feb 1913
Jan 1969	Jan 1914
Feb 1969	Feb 1915
Nov 1970	Jan 1916
Dec 1971	Mar 1918
Jan 1974	Dec 1921
Feb 1976	Apr 1926
Sep 1976	Feb 1927
Feb 1978	Dec 1931
Mar 1978	Feb 1932
Feb 1980	Jan 1933
Mar 1983	Feb 1936

Continued on next page

Table 3.1 – *Continued from previous page*

Gauge	Modeled
Dec 1984	Dec 1936
Mar 1991	Feb 1937
Feb 1992	Mar 1938
Jan 1993	Dec 1938
Feb 1993	Mar 1941
Jan 1995	Dec 1941
Mar 1995	Jan 1943
Feb 1998	Dec 1943
Feb 2000	Feb 1944
Jan 2008	Nov 1944
Dec 2010	Dec 1945
	Nov 1946
	Jan 1952
	Mar 1952
	Nov 1960
	Feb 1962
	Apr 1965
	Nov 1965
	Dec 1965

Continued on next page

Table 3.1 – *Continued from previous page*

Gauge	Modeled
	Dec 1966
	Apr 1967
	Nov 1967
	Jan 1969
	Feb 1969
	Nov 1970
	Dec 1971
	Feb 1973
	Jan 1974
	Mar 1978
	Feb 1980
	Mar 1983
	Dec 1984
	Nov 1985
	Mar 1991
	Feb 1992
	Jan 1993
	Feb 1993
	Jan 1995

Continued on next page

Table 3.1 – *Continued from previous page*

Gauge	Modeled
	Mar 1995
	Feb 1998
	Jan 2005
	Feb 2005
	Jan 2008
	Jan 2010
	Dec 2010
	Dec 2014
	Jan 2017
	Feb 2019

3.4.1.2 Extreme Drought

Droughts as mentioned previously are those that are below the negative red line. For climate division six shown in Figure 3.2 and Figure 3.3 is the opposite from the extreme precipitation events. Here, there are more drought anomalies in the modeled than the gauge dataset. However, the magnitude of the similar events are larger in the gauge time-series. Referring to Table 3.2, the first agreeable drought event for both datasets is for February 1952. Unfortunately, the year of 1952 is the end of a moderate El Niño and has no connection

with a drought. The year 2011 drought was the only capture for the gauge data set. However, the modeled data set, like the gauge data set, was able to classify only a few instances of droughts. Though it was able to correctly classify in January 1975 and January 1976 of an occurrence of a weak La Niña.

The climate divisions 1, 2, 3, 4, 5, and 7 have similar findings to climate division six. They all very similar in sporadically correctly classifying La Niña events. Overall, compared to El Niño events, the drought events are not as well predicted.

Table 3.2: Drought dates in climate division six

Gauge	Modeled
Feb 1952	Feb 1896
Jan 1966	Jan 1912
Feb 1967	Feb 1912
Mar 1969	Jan 1918
Jan 2003	Feb 1924
Jan 2011	Jan 1927
	Jan 1936
	Nov 1940
	Feb 1952
	Feb 1967
	Dec 1968
	Feb 1974

Continued on next page

Table 3.2 – *Continued from previous page*

Gauge	Modeled
	Jan 1975
	Jan 1976
	Nov 1977
	Dec 1979
	Nov 1992
	Dec 1994
	Jan 2003

3.4.2 Statistical Comparison for Modeled and Gauge Datasets

Theoretically, the gauge dataset is the true measurement of precipitation while the model contains many inputs, such as tree ring data, ice core data, satellite data, etcetera, in order to properly estimate the amount of precipitation. To compare statistically the differences between the two time-series, we looked at their differences, $Y_t = X_t - Z_t$ where X_n is the modeled time-series and Z_t is the gauge time-series. Theoretically, if both the modeled and gauge time-series are the same it should equal zero or it would have a zero mean at worst case scenario. Plotting a histogram of the difference between gauge and modeled data shows it is extremely dominant in the positive side as seen in Figure 3.4 for climate division six, respectively. This is due to the gauge being almost always larger than

the modeled dataset. The average difference between all the climate divisions is 0.5 inches of rain. The results of the remaining climate divisions can be found in Appendix B. The other climate divisions show a mixture of difference between the quantity of precipitation. At times the gauge has a higher amount of precipitation and others the modeled time-series does. The difference in the amount of rainfall between the true measurements is not drastically different as they tend to be almost centered around zero. A few years ago, when the analysis was first done up to the year of 1995, it did not show such promising results. The extra 15 years of data and possibly the correction of the modeled time-series data bridged a closer gap between the two data sets.

Figure 3.4: Histogram of gauge minus modeled precipitation in climate division six

The results from the histogram suggests that the difference time-series are not zeroed mean. This may be the case that the time series have a trend component. In order to test the hypothesis, we used Mann-Kendall Test which is given by:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(z_j - z_i), \quad (3.10)$$

where,

$$\text{sign}(z_j - z_i) = \begin{cases} 1, & > (z_j - z_i) \\ 0, & = (z_j - z_i) \\ -1, & < (z_j - z_i) \end{cases} \quad (3.11)$$

The trend returns a p -value signifying the presence of a trend based on hypothesis testing:

$$\begin{cases} H_O : & \text{No Trend.} \\ H_A : & \text{Presence of a Trend.} \end{cases} \quad (3.12)$$

For all climate divisions, except for climate divisions four and five, we obtained a high p -value, where we fail to reject the null hypothesis that there is no trend. In addition, we can check if the difference time series (Z_t) for climate divisions is stationary using the Augmented Dickey-Fuller Unit-Root Test:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t, \quad (3.13)$$

where,

$$\begin{cases} H_O : & \gamma = 0 \\ H_A : & \gamma < 0, \end{cases} \quad (3.14)$$

and,

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}. \quad (3.15)$$

The results of the Augmented Dickey-Fuller Unit-Root Test suggests that all of the difference time series are stationary. Since they are stationary we can test if they have a zero mean. The zero mean test is based on the z-scores and obtaining p-values in the manner of hypothesis testing:

$$\begin{cases} H_O : & \mu(t) = 0 \\ H_A : & \mu(t) \neq 0. \end{cases} \quad (3.16)$$

All the p-values are below 0.05 signifying that we reject the null hypothesis to be in favor of the alternative where the it is a non-zero time series. This result shows that the modeled is always underestimating the gauge.

3.4.3 EMD

Empirical mode decomposition (EMD) was performed on the rain gauges and modeled data sets and SOI, PDO, and AMO indices. The raw datasets have an insignificant relationship with each other, and thus EMD was performed. In the first row of Figure 3.5, one can see that the modeled/gauge has almost a zero correlation with SOI, PDO, and AMO as raw datasets. A low correlation is due to the high amount of noise within the datasets.

Not surprisingly, the modeled and gauge time-series have a high correlation of 0.96, which solidifies a great correspondence amongst the datasets. In the second and third row, one can see there is a much better correlation amongst the time-series. The noted correlations are for the strongest correlation among all components and lag times. A positive lag means that the modeled or gauge is ahead by n months. A negative lag means that SOI, PDO, or AMO is ahead by n months. For example, referring to the correlation between the gauge time-series and SOI, it has a high correlation of 0.39 with a negative lag of three months. We interpret that the gauge time-series is three months behind of the SOI time-series. The summary of the results of Figure 3.5 are discussed next in Table 3.3.

Figure 3.5: Climate division six EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Table 3.3 shows all seven climate divisions with its respective precipitation EMD components compared to the three indices for the gauge time-series. In the column of CD (Climate Division), it lists the total amount of gauge stations that were used. For each climate index, the table illustrates the cross-correlation with the gauge time-series. Furthermore, it list the component with which the highest possible correlation exist between a lag of -3 to 3.

The results show that for SOI, the majority of the components rely among the 8th and 9th component. This correlates with the longevity cycle of having a strong El Niño [188]. The lag value of -3 signifies that SOI precedes the precipitation as it should due to the ocean and atmosphere dynamics [188]. PDO and AMO also have similar results to SOI when we refer to the EMD components and the lag. The lag is mostly consistent shifting between -3 and 3.

To further explain the results from Table 3.3, our results are displayed visually in Figure 3.6. Our results demonstrate that there is a fluctuation with the dominate index from north California to the south. In northern California, we show that it is dominated by PDO and SOI. However, from climate division four to seven, it is completely dominated by AMO. This is a new regional and long-term finding that does not exist in literature. The concept of having a mixture of climate indices representing California and predominately AMO having the largest correlation around 80% is unorthodox. Furthermore, we show the difficulty in finding a reasonable correlation with climate division 7, which may also be due to the low number of gauge stations.

The next results is with the modeled time-series EMD components with the same three climate indices shown in Table 3.4. The results show that SOI's, PDO's, and AMO's EMD components rely mostly among the 7th and 8th component, which is similar to the the gauge time-series findings. This correlates continues to agree with the longevity cycle of having a strong El Niño [188].

Table 3.3: Precipitation data from gauge time-series compared to SOI, PDO, and AMO. Each row block represents a climate division and underneath has a value of the number of stations in each division. Within each row block there are four rows that state the cross correlation, precipitation component, climate index component, and lag (in months). Each column block represents the climate index of interest.

CD		SOI	PDO	AMO
1 (112 Stations)	CC	0.83	0.85	0.56
	PrecipComp	9	10	9
	IndexComp	8	8	7
	Lag	-3	3	3
2 (136 Stations)	CC	0.76	0.60	0.75
	PrecipComp	8	10	8
	IndexComp	8	8	7
	Lag	-3	3	3
3 (18 Stations)	CC	0.66	0.25	0.28
	PrecipComp	8	7	6
	IndexComp	8	5	7
	Lag	-3	3	-3
4 (97 Stations)	CC	0.31	0.32	0.83
	PrecipComp	7	5	8
	IndexComp	8	5	7
	Lag	-3	2	-3
5 (103 Stations)	CC	0.24	0.27	0.48
	PrecipComp	6	6	8
	IndexComp	5	5	7
	Lag	-3	3	-3
6 (209 Stations)	CC	0.39	0.36	0.80
	PrecipComp	7	6	8
	IndexComp	8	6	7
	Lag	-3	-3	-3
7 (60 Stations)	CC	0.10	0.21	0.39
	PrecipComp	7	7	7
	IndexComp	8	5	6
	Lag	3	-3	-3

To further explain the results from Table 3.4, our results are displayed visually in Figure 3.7. Our results shows a distinct pattern from the gauge time-series. There is a mixture of dominate climate indices with an extremely low occurrence of SOI as the highest correlated climate index. SOI has the highest correlation with climate division five. In addition,

Figure 3.6: Climate division gauge-indices correlation

PDO occurs a total of four times (including the cases when PDO and AMO both have the highest correlation in climate division six), while AMO has three occurrences (again double counting for climate division five). The findings lead to new understanding of not depending on SOI for ENSO but also referring to other climate indices as potential predictors in local regions as California and even into its climate divisions. However, this does not suggest that PDO and AMO should now be a worldwide predictor for ENSO as SOI may still be the best globally.

Expanding the results shown in Table 3.3 and Table 3.4, they are reproduced with added results in Tables 3.5, 3.6, 3.7, 3.8, 3.9, and 3.10. The tables not only demonstrate the highest lag correlation component for each climate division, but it also shows the rate of occurrence (cycle) for the component that they correlate with. We are familiar with the concept of El Niño repeating every 2 to 8 years [28, 71] so we expect that the cycle of the sine-wave should be around the same time period. Our results show that the gauge time

Table 3.4: Precipitation data from modeled time-series compared to SOI, PDO, and AMO. Each row block represents a climate division and underneath has a value of the number of stations in each division. Within each row block there are four rows that state the cross correlation, precipitation component, climate index component, and lag (in months). Each column block represents the climate index of interest.

CD		SOI	PDO	AMO
1	CC	0.55	0.27	0.69
	PrecipComp	8	5	8
	IndexComp	8	5	7
	Lag	-3	-1	3
2	CC	0.26	0.33	0.27
	PrecipComp	6	6	7
	IndexComp	5	5	7
	Lag	-3	-3	-3
3	CC	0.50	0.89	0.20
	PrecipComp	9	9	7
	IndexComp	9	8	7
	Lag	3	0	-3
4	CC	0.31	0.51	0.96
	PrecipComp	6	10	10
	IndexComp	6	8	9
	Lag	3	-3	-3
5	CC	0.84	0.43	0.24
	PrecipComp	8	9	6
	IndexComp	8	8	9
	Lag	3	3	-1
6	CC	0.19	0.33	0.33
	PrecipComp	8	7	8
	IndexComp	7	5	7
	Lag	-3	-3	-3
7	CC	0.92	0.71	0.33
	PrecipComp	8	9	7
	IndexComp	8	8	7
	Lag	-3	0	-3

series correlated with the highest possible component with SOI is around the 20 year mark.

If we refer to Table 2.1 and look closely to the Very Strong years, we notice that the distance between those are roughly 15 to 20 years apart. The modeled time-series with SOI, shown in Figure 3.6, shows almost the same pattern as the gauge time-series, with the exception of a few cycles that are at 9 and 6 years. Figure 3.7 shows a much closer relationship to

Figure 3.7: Climate division modeled-indices correlation

the general cycle of ENSO of 2 to 8 years. Climate division one and two have NA's because the EMD component is not cyclic. Figure 3.8 shows the comparison between the modeled time-series and PDO and has varying results compared to its counterpart. For the majority of the climate divisions, the cycle is between 2 to 8 years. However, the scale is much larger for climate division four and five. The reason those climate divisions are have a larger cyclic pattern may be due to the expanded range of PDO [133]. The last two cases is with the AMO for both gauge and modeled time-series, which are displayed in Figure 3.9 and Figure 3.10. They both show an extensive range in cycle of about 6 to 40 years. This range is within the boundaries of repetitions of all ENSO events and their particular relationship within a localized regions across California.

Table 3.5: Gauge and SOI EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

CD	Gauge Comp	SOI Comp	0 Lag Cor.	Lag Cor.	Lag	Gauge ROO Months	Gauge ROO Years	SOI ROO Months	SOI ROO Years
1	9	8	0.82	0.83	-3	317.00	26.42	434.00	36.17
2	8	8	0.74	0.76	-3	271.33	22.61	434.00	36.17
3	8	8	0.65	0.66	-3	299.00	24.92	245.67	20.47
4	7	8	0.30	0.31	-3	231.67	19.31	434.00	36.17
5	6	5	0.22	0.24	-3	106.14	8.85	50.20	4.18
6	7	8	0.39	0.39	-3	237.33	19.78	434.00	36.17
7	7	8	0.09	0.10	3	145.40	12.12	245.67	20.47

Table 3.6: Modeled and SOI EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

CD	Modeled Comp	SOI Comp	0 Lag Cor.	Lag Cor.	Lag	Modeled ROO Months	Modeled ROO Years	SOI ROO Months	SOI ROO Years
1	8	8	0.53	0.55	-3	256.50	21.38	434.00	36.17
2	6	5	0.23	0.26	-3	111.43	9.29	50.20	4.18
3	9	9	0.48	0.50	3	466	38.83	347.00	28.92
4	6	6	0.29	0.31	3	75.64	6.30	104.43	8.70
5	8	8	0.83	0.84	3	317.50	26.46	434.00	36.17
6	8	7	0.18	0.19	-3	277.00	23.08	203.67	16.97
7	8	8	0.90	0.92	-3	199.75	16.65	245.67	20.47

Table 3.7: Gauge and PDO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

C/D	Gauge Comp	PDO Comp	0 Lag Cor.	Lag Cor.	Lag	Gauge ROO Months	Gauge ROO Years	PDO ROO Months	PDO ROO Years
1	10	8	0.84	0.85	3	NA	NA	NA	NA
2	10	8	0.59	0.61	3	358.00	29.83	NA	NA
3	7	5	0.25	0.25	3	157.50	13.13	61.23	5.10
4	5	5	0.32	0.32	2	48.50	4.04	61.23	5.10
5	6	5	0.24	0.27	3	106.14	8.85	61.23	5.10
6	6	6	0.35	0.36	-3	111.86	9.32	126.00	10.50
7	7	5	0.21	0.21	-3	145.40	12.12	61.23	5.10

Table 3.8: Modeled and PDO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

CD	Modeled Comp	PDO Comp	0 Lag Cor.	Lag Cor.	Lag	Modeled ROO Months	Modeled ROO Years	PDO ROO Months	PDO ROO Years
1	5	5	0.27	0.27	-1	52.33	4.36	61.23	5.10
2	6	5	0.32	0.33	-3	111.43	9.29	61.23	5.10
3	9	8	0.89	0.89	0	NA	NA	NA	NA
4	10	8	0.49	0.51	-3	501.00	41.75	NA	NA
5	9	8	0.40	0.43	3	450.00	37.50	NA	NA
6	7	5	0.30	0.33	-3	148.00	12.33	61.23	5.10
7	9	8	0.71	0.71	0	262.00	21.83	NA	NA

Table 3.9: Gauge and AMO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

CD	Gauge Comp	AMO Comp	0 Lag Cor.	Lag Cor.	Lag	Gauge ROO Months	Gauge ROO Years	AMO ROO Months	AMO ROO Years
1	9	7	0.53	0.56	3	317.00	26.42	271.67	22.64
2	8	7	0.72	0.75	3	271.33	22.61	271.67	22.64
3	6	7	0.28	0.28	-3	105.86	8.82	244.00	20.33
4	8	7	0.82	0.83	-3	511.00	42.58	271.67	22.64
5	8	7	0.45	0.48	-3	438.00	36.50	271.67	22.64
6	8	7	0.79	0.80	-3	517.00	43.08	271.67	22.64
7	7	6	0.38	0.39	-3	145.40	12.12	106.57	8.88

Table 3.10: Modeled and AMO EMD correlation with rate of occurrence. “0 Lag Cor.” corresponds to the correlation of the two EMD components at zero lag time. “Lag. Cor.” corresponds to the correlation at “Lag” value. “ROO” means Rate Of Occurrence (cycle) of each component in terms of months or years.

CD	Modeled Comp	AMO Comp	0 Lag Cor.	Lag Cor.	Lag	Modeled ROO Months	Modeled ROO Years	AMO ROO Months	AMO ROO Years
1	8	7	0.68	0.69	3	256.50	21.36	271.67	22.64
2	7	7	0.26	0.27	-3	183.50	15.29	271.67	22.64
3	7	7	0.19	0.20	-3	193.50	16.13	244.00	20.33
4	10	9	0.95	0.96	-3	501.00	41.75	495.00	41.25
5	6	9	0.24	0.24	-1	71.90	5.99	495.00	41.25
6	8	7	0.32	0.33	-3	277.00	23.08	271.67	22.64
7	7	7	0.32	0.33	-3	116.14	9.68	244.00	20.33

3.5 Conclusion

The results of this study demonstrate that SOI is not necessarily the highest correlated climate index among the seven climate divisions of California. PDO and AMO are great representatives of precipitation patterns in California as it was shown in an ARMA and EMD analysis. The ARMA analysis showed the comparison between SOI, AMO, and PDO by removing the seasonal and trend components. However, EMD decomposed the signal further than STL and we compared each precipitation component with each climate index component at varying lags and find the highest correlation. The results were very promising and expose that SOI may be a naive way of associating ENSO as a local contributor to precipitation. From this study a question may rise, are there any other climate indices that are correlate with California's precipitation patterns? And if none exist, can one create an assimilated climate index? Chapter 4 will answer those questions.

4. Third Study: Assimilation of a new Climate Index for Gauge and Modeled Data Using Autoregressive Moving Average and Advance Modeling Techniques

4.1 Introduction

In Chapters Two and Three, we noticed that there were a few different climate indices that could reliably represent ENSO events in California. The most common climate index that is used globally is the Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), NIÑO 1+2, NIÑO 3, NIÑO 4, and NIÑO 3.4. SOI measures the difference in sea level pressure between the Eastern and Western hemisphere (Tahiti and Darwin). The concept measures the connection of atmosphere and ocean interaction. PDO measures the sea surface temperature anomalies of the North Pacific basin (Pacific Ocean). NIÑO 1+2, NIÑO 3, NIÑO 4, and NIÑO 3.4 measure anomalous sea surface temperature across the central Pacific Ocean. These climate indices are ideal because they measure the conditions of an ENSO phase. However, not all ENSO phases results in accurate measurement of abnormal precipitation or drought. Furthermore, climate scientists have religiously used SOI, PDO, and NIÑO indices as ENSO signs but are there other indices that can measure ENSO better

or possibly a weighted combination of them? Therefore, the focus of this chapter is to use true abnormal precipitation measurements and build a model using an assembled climate index to forecast such events in Californias seven climate divisions.

4.2 Data Used

We used hourly ground (In-situ) precipitation data from the National Climatic Data Center and NOAA’s Forecast Systems Laboratory [131], where there are more than 2500 active stations and 7000 total stations. The dates for the stations in California range from 1948 through 2019. However, some stations outside of California begin as early as 1900. The hourly data extracted (735 stations from 1948 to 2019) concentrates on California’s seven climate divisions as shown in Figure 1.1. The data obtained has missing hours, days, and up to a few weeks. Therefore, the total precipitation of monthly data was calculated to solve the discontinuous problem by inputting missing values. We arithmetically averaged all stations in their respective climate divisions.

Monthly modeled precipitation data from NCDC gridded dataset [182] for California’s seven climate divisions were used to compare with the rain gauge (in-situ) measurements. The data begins from 1895 through 2019 for all climate divisions. The data was derived from area-weighted averages of grid-point estimates interpolated from station data. A resolution of 5 km was used to ensure all divisions had sufficient spatial sampling. To accurately compare both time-series, we cropped the monthly modeled data to the respective timestamp of

the gauge data per climate division.

A total of 69 climate indices were used to determine whether a specific index can represent California or a combination of them. The dates vary from each index, where they are dependent on how well maintain they are. Below is a list of all the indices that were used in this study:

1. AAO (Antarctic Oscillation) [139]
2. AMM (Atlantic Meridional Mode) [140]
3. AMO Smoothed (Atlantic Multidecadal Oscillation Smoothed) [141]
4. AMO Unsmoothed (Atlantic Multidecadal Oscillation Unsmoothed) [142]
5. AO (Arctic Oscillation) [143]
6. Atlantic Tripole SST EOF [144]
7. CAR (Caribbean Index) [145]
8. Central Indian Precipitation [146]
9. Darwin Sea Level Pressure [111]
10. DMI (Dipole Mode Index) [147]
11. EA WR (Eastern Asia / Western Russia) [148]
12. ENSO Precipitation Index [149]

13. EP NP (East Pacific / North Pacific Oscillation) [150]
14. Equatorial Eastern Pacific SLP [112]
15. Equatorial SOI [114]
16. Equatorial SOI Indonesia SLP [113]
17. Gibraltar SLP [15]
18. Global Mean Land Ocean Temperature Index [151]
19. Globally Integrated Angular Momentum [152]
20. Global Tropics SST [115]
21. Hurricane Activity (Atlantic Hurricanes) [153]
22. Iceland SLP [16]
23. Indonesia SLP [116]
24. Madras SLP [17]
25. MEI (Multivariate ENSO Index) [154]
26. Nagasaki SLP [19]
27. NAO (North Atlantic Oscillation) [155]
28. NAO AZO (North Atlantic Oscillation Azores) [18]
29. NAO Jones (North Atlantic Oscillation Jones) [90]

30. NIÑO 1 + 2 (Extreme Eastern Tropical Pacific SST) [156]
31. NIÑO 3 (Eastern Tropical Pacific SST) [158]
32. NIÑO 3.4 (East Central Tropical Pacific SST) [157]
33. NIÑO 4 (Central Tropical Pacific SST) [159]
34. NOI (Northern Oscillation Index) [91]
35. North Atlantic SST [92]
36. Northeast Brazil Rainfall Anomaly [93]
37. NPI (North Pacific Index) [129]
38. NTA (North Tropical Atlantic Index) [94]
39. OLR Central Pacific (Outgoing Long Wave Radiation Central Pacific) [117]
40. OLWR Equator (Outgoing Long Wave Radiation Equator) [118]
41. ONI (Oceanic Nio Index) [95]
42. Pacific Warmpool [96]
43. PDO (Pacific Decadal Oscillation) [137]
44. PNA (Pacific North American Index) [97]
45. QBO U30 Index (Quasi-Biennial Oscillation 30mb zonal wind) [119]
46. QBO U50 Index (Quasi-Biennial Oscillation 50mb zonal wind) [120]

47. RAO (Reconstructed North Atlantic Oscillation) [130]
48. Sahel Rainfall [98]
49. SAM (Southern Annular Mode) [99]
50. SOI (Southern Oscillation Index) [138]
51. SOI NIO 3.4 [100]
52. Solar Flux [101]
53. South Atlantic SST [121]
54. Sunspot Index [102]
55. SW Monsoon Region Rainfall [103]
56. Tahiti Sea Level Pressure [122]
57. TNA (Tropical Northern Atlantic Index) [104]
58. TNI (Trans- Nio Index) [105]
59. TPI (Trans Polar Index) [20]
60. TPI IPO (Tripole Index for the Interdecadal Pacific Oscillation) [106]
61. Trade Winds Index 135E 180W 5N 5S West Pacific [123]
62. Trade Winds Index 135W 120W 5N 5S East Pacific [124]
63. Trade Winds Index 175W 140W 5N 5S Central Pacific [125]

- 64. Tropical Pacific SST EOF [107]
- 65. TSA (Tropical Southern Atlantic Index) [108]
- 66. WHWP (Western Hemisphere Warm Pool) [109]
- 67. WP (Western Pacific Index) [110]
- 68. Zonal Winds Equator 200mb [126]
- 69. Zonally Average 500mb Temperature Anomalies [127]

4.3 Methodology

The rain gauge and modeled precipitation time-series was decomposed to remove seasonal and trend components. This work used local linear regression (loess) [14] to remove such trends into a composition of seasonal, trend, and residual components (STL decomposition) using Equation 2.1 [28]. The seasonal and trend components are of no interest to this study, but the residual (R_t) due to its fluctuation of anomalies. The residual time-series were then fitted to an ARMA model as described in Equation 3.1. Using the model's values, we calculated the extreme events using Equation 3.5. We classified the positive extreme events as El Niño and negative extreme events as La Niña. Using the values of extreme events, we created a new time-series that consists of multiple classifications (0 - No ENSO event, 1 - El Niño event, and 2 - La Niña event).

In a similar matter, all 69 climate indices were decomposed and an ARMA model was created from it. We found extreme events and labeled them in a binary format (0 - No extreme event

and 1 - extreme event present). For each climate index, we created two extreme time-series for events that we called above and below, where above signifies the positive 95% confidence interval of extreme events and below signifies the negative 95% confidence interval of extreme events.

The objective of this study is to hindcast each time stamp as an ENSO event using vector generalized linear models (VGLMs), Random Forest Classifier, Extra Trees, Neural Network, Ridge, and Lasso. We compared the results to all models in the discussion section. The complete dataset that we used to hindcast ENSO events from the calculated multiple classification time-series per climate division is the original climate indices time-series, the 95% above time-series, and the 95% below time-series. This gives a total of 207 predictors. The definitions and algorithms of each model is in the next few sections.

4.3.1 Vector Generalized Linear Models

The vector generalized linear models (VGLMs) is an extension of the generalized linear models (GLM), which was proposed by Nelder and Wedderburn [132]. They showed that the Poisson regression, logistic regression, and other linear regressions can be represented as special cases of GLMs. Unfortunately, GLMs are restricted to one-parameter distribution. Therefore, VGLMs were introduced as an upgraded linear model for multiple parameters as there are in this study (0 - No ENSO event, 1 - El Niño, and 2 - La Niña). The general definition of VGLMs is adopted from Yee [193] [195] .

If we let y represent a Q -dimensional vector, then Y will be a conditional distribution that is based on the condition of x in the following form:

$$f(y|x; B) = h(y, \eta_1, \dots, \eta_M) \quad (4.1)$$

for some known function $f(\cdot)$, where $B = (\beta_1 \beta_2 \dots \beta_M)$ is a $p \times M$ matrix of regression coefficients, and the j th linear predictor is

$$\eta_j = \beta_j^T x = \sum_{k=1}^p \beta_{(j)k} x_k, \quad j = 1, \dots, M, \quad (4.2)$$

where $x = (x_1, \dots, x_p)^T$ with $x_1 = 1$ if there is an intercept. It should be noted that, in general, there is no relationship between Q and M . The relationship, if any, will depend on the model or the distribution to be fitted. More likely than not, M is the number of independent parameters.

VGLMs are estimated by iteratively reweighted least squares (IRLS) algorithm. For data (x_i, y_i) , $i = 1, \dots, n$, most models that can be fitted have a log-likelihood

$$l = \sum_{i=1}^n w_i l_i. \quad (4.3)$$

The w_i are known and fixed positive prior weights. If we let x_i denote the explanatory vector for the i th observation, then we can write

$$\eta_i = \begin{pmatrix} \eta_1(x_i) \\ \vdots \\ \eta_M(x_i) \end{pmatrix} = B^T x_i = \begin{pmatrix} \beta_1^T x_i \\ \vdots \\ \beta_M^T x_i \end{pmatrix}. \quad (4.4)$$

The adjusted dependent vector for an IRLS $z_i = \eta_i + W_i^{-1}d_i$ is regressed upon a large model matrix with $d_i = w_i\delta^2 l_i/(\delta\eta_i\delta\eta_i^T)$ or $W_i = -w_i E[\delta^2 l_i/(\delta\eta_i\delta\eta_i^T)]$, giving rise to the Newton-Raphson and Fisher-scoring algorithms respectively [193, 195]. Fisher-scoring is preferred for numerical stability because W_i are positive-definite over a larger region of parameter space. The disadvantage for a numerical stability is a slower convergence rate and less accurate standard errors [27, 193, 195].

The work used three distinct methods of VGLMs, described in the next few sections, to test the ability of varying the models. We also used R [164] to run the model and perform the analysis with the package **VGAM** and use the function **vglm** [191, 192, 193, 194, 195, 196, 197].

4.3.1.1 Main Effect

In the first part of the research, we studied the main effects of each climate index as a predictor to an ENSO event. We did a custom made forward selection and backwards elimination that adds and removes a climate index predictor. The first step in the building process is to find the climate index that has the strongest influence to the model by finding which one produces the smallest p -value. After that, we add one more climate index (the

68 that are left over) and find which one of those also produces the smallest p -value. This process continues until there are no more climate indices that we can add, where the smallest value is greater than 0.05. As we add each new predictor, we check if any of the predictor's p values became insignificant (greater than 0.05). If they have, we remove them from the model and process them in the following steps. If the p -value changed to be within the boundaries $0.05 < x < 0.25$, we kept the climate index as a potential to return to the model as a significant predictor. If the p -value was greater than 0.25, we removed it forever as it will never become significant. The results of the main effect are described in the next section.

4.3.1.2 Two-Way Interaction

In the second part of the research, we studied the two-way interaction of each climate index as a predictor to an ENSO event. Similar to the main effect, we constructed a forward selection and backwards elimination for the interaction of two climate index predictors. When we add the first predictor, it is exactly the same procedure as main effect in the selection process. However, when we add the next climate index predictor, we find the best index by finding the lowest p -value of the interaction between the two climate indices. If any of the interaction is greater than 0.05, we process it in a similar way as main effect. The only difference is that we check if all two-way interactions of the climate index that causes the increase in p -value is larger than we remove it. If there is at least one that is smaller than 0.05, then we do not remove it. Furthermore, if all two-way interactions of the climate index is greater than 0.25, then we remove it from being a potential predictor. The results

of the two-way interaction are described in the next section.

4.3.1.3 Three-Way Interaction

In the third part of the research, we studied the three-way interaction of each climate index as a predictor to an ENSO event. Like the two-way interaction, we constructed a forward selection and backwards elimination for the intersection of three climate index predictors. When we add the first predictor, we use the same process as the main effect in the selection process. When we add the second predictor, we use the same process as the two-way interaction selection process. However, when we add the third climate index, assuming all are significant, we look at the p -values of the three-way correlation. If all associations of the new climate index predictor is above 0.25, we remove it completely from ever being added to the model. If all associations of the new climate index is $0.05 < x < 0.25$, then we remove it from the model with the potential of becoming significant in the model in the future as predictors fluctuate. The results of the three-way interaction are described in the next section.

4.3.2 XGBoost

One method we used to compare our model to is the Extreme Gradient Boosting (XGBoost) model. It is a decision tree based ensemble machine learning algorithm that uses a gradient boosting framework. The XGBoost algorithm is always used for classification

and regression, which is ideal for our three parameter classification of no ENSO event, El Niño or La Niña event. XGBoost is a second order method that is improved from Friedman et al. [35]. The objective is if we were given a data set with n examples and m features $D = \{(x_i, y_i)\}$ ($|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}$). There are K additive functions to predict the output, given as:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in F, \quad (4.5)$$

where $F = \{f(x) = w_{q(x)}\} (q : \mathbb{R} \rightarrow T, w \in \mathbb{R}^T)$ is the space of regression trees, which is also known as CART [12]. The variable q represents the structure of each tree that maps an example to the corresponding leaf index. The variable T is the number of leaves in the tree structure. Each f_k corresponds to an independent tree structure q and its leaf weights are w . Each regression tree contains a cumulative score on each of the leaves, which is represented as w_i for the i th leaf. Decision rules in the trees, given by q , are used to classify a particular case into its appropriate leaf and calculate the final prediction by sum of the score of the corresponding leaves [12]. The learning set of functions used in the model is minimized by a regularized objective:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad (4.6)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$. The variable l is a differentiable convex loss function between the prediction $\hat{y}_i$ and the target y_i . The term Ω penalizes the complexity of the model. The

remaining terms in Equation 4.6 aid in smoothing the final learned weights to avoid overfitting [12].

Equation 4.6 is a tree ensembled model that includes functions as parameters and cannot be optimized in the Euclidean space [12]. In order to optimize it, the model is trained in an additive manner. If we let $\hat{y}_i^{(t)}$ be the prediction of the i th instance at the t iteration, then we need to add f_t to minimize the regularized learning objective, given as:

$$L^{(t)} \simeq \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t). \quad (4.7)$$

The term f_t improves the model through its included summation. With a second-order approximation, the regularized learning objective can be rewritten as:

$$L^{(t)} \simeq \sum_{i=1}^n \left[l\left(y_i, \hat{y}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \quad (4.8)$$

where $g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right)$ and $\partial_{\hat{y}^{(t-1)}}^2 l\left(y_i, \hat{y}^{(t-1)}\right)$ are first and second order gradient statistics on the loss function. At a particular step t , we can remove the constant terms to obtain the following model:

$$\tilde{L}^{(t)} \simeq \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t). \quad (4.9)$$

For an instance of leaf j , we can define $I_j = \{i | q(x_i) = j\}$. Rewriting Equation 4.9 by expanding Ω gives us:

$$\begin{aligned}
\tilde{L}^{(t)} &= \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\
&= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T.
\end{aligned} \tag{4.10}$$

If $q(x)$ represents a fixed structure, we can compute the optimal weight, w_j^* , of leaf j as:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \tag{4.11}$$

and calculate the corresponding optimal value as:

$$\tilde{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \tag{4.12}$$

Equation 4.12 is used as a scoring function that measures the quality of a tree structure q . In order to enumerate all tree structures q , a greedy algorithm is used where it starts from a single leaf and iteratively adds branches to the tree [12]. If we left I_L and I_R represent instances of left and right nodes after a split, then the loss reduction after split is:

$$L_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma, \tag{4.13}$$

where $I = I_L \cup I_R$.

The work used R [164] to run the model and perform the analysis with the package/function `XGBoost` [13].

4.3.3 Random Forest Classifier

The random forest classifier is a well-known tree classifier where each classifier is generated using a random vector sampled independently from an input vector [161]. Each tree votes on the most favored class to classify the input vector [5, 6, 161]. We used the random forest classifier since it consists of using randomly selected features or a combination of features for each node to grow a tree. It optimizes the performance by using bagging, which is a method of generating a training data set by randomly selecting with replacement K examples, where K is the size of the original training data set [5, 6, 48, 161]. The decision tree design is required for selection attribute measurement and pruning. There exists various approaches for the selection of attributes, but most assign a quality measure to each attribute [161]. The most popular approaches in decision tree induction are Information Gain Ratio [163] and Gini Index [7]. The Gini Index measures the impurity of an attribute with respect to each class [161] and is used in the Random Forest Classifier. Given a training data set, T , selecting one case at random and classifying it to class C_i , the Gini Index can be written as:

$$\sum \sum_{j \neq i} \left(\frac{f(C_i, T)}{|T|} \right) \left(\frac{f(C_j, T)}{|T|} \right), \quad (4.14)$$

where $\frac{f(C_i, T)}{|T|}$ is the probability that the selected case belongs to class C_i

Trees that are full-grown to the maximum depth on a new training data set that using a combination feature are not pruned [161]. The concept of not pruning is large advantage for random forest classifier compared to the proposition of Quinlan [163]. The performance of the tree-based classifiers are affected if there is a presence of pruning compared to attribute selection measurements [87, 161, 162]. Breiman [5] discusses that as the number of trees increases, the generalization error will always converge with or without pruning and overfitting will not occur due to the Strong Law of Large Numbers [30]. A random forest is generated based on the number of features used at each node and the user-defined number of trees, N . Each node searches for the selected features and is split based on the findings. The classification of a new data set is passed down to the N trees and the forest chooses a class that has the more votes out of N for every case [161].

The work used R [164] to run the model and perform the analysis with the package/function `randomForest` [73].

4.3.4 Extra Trees

The Extra-Trees algorithm is a batch-mode supervised learning with an emphasis of working with large numerical data sets. It builds an ensemble of unpruned decision or regression trees [37]. There are two differences between Extra-Trees and other tree-based

algorithms: 1) Splits nodes by selecting cut-points completely at random and 2) Uses the entire learning data set to grow trees as oppose to using a bootstrap replica [37]. The predictions of the trees are aggregated to yield the predictions by majority vote for a classification problems and arithmetic average for regression problems [37]. The Extra-Trees algorithm, as shown in Table 4.1, optimizes in the use of randomize cut-points and ensemble averaging to reduce the variance. The three major parameters are K , $n_{\min}$, and M . The term K determines the strength of the attribute selection, $n_{\min}$ is the strength averaging output noise, and M is the strength of the variance reduction of the ensembled model aggregation [37]. The growth of the trees, assuming they are well-balanced, has a Big-O of $N \log N$, where N represents the sample size [37].

Table 4.1: Extra-Trees splitting algorithm (for numerical attributes) from Geurts et al. [37]

Split_a_node(S)

Input: the local learning subset S corresponding to the node we want to split

Output: a split $[a < a_c]$ or nothing

-If **Stop_split**(S) is TRUE then return nothing.

-Otherwise select K attributes $\{a_1, \dots, a_k\}$ among all non constant (in S) candidate attributes;

-Draw K splits $\{s_1, \dots, s_k\}$, where $s_i = \mathbf{Pick_a_random_split}(S, a_i), \forall i = 1, \dots, K$;

-Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1, \dots, K} \text{Score}(s_i, S)$.

Continued on next page

Table 4.1 – *Continued from previous page*

Pick_a_random_split(S, a)

Inputs: a subset S and an attribute a

(Output): a split

-Let $a_{\max}^S$ and $a_{\min}$ denote the maximal and minimal value of a in S ;

-Draw a random cut-point a_c uniformly in $[a_{\min}^S, a_{\max}^S]$;

-Return the split $[a < a_c]$.

Stop_split(S)

Inputs: a subset S

(Output): a boolean

-If $|S| < n_{\min}$, then return TRUE; -If all attributes are constant in S , then return TRUE;

-If the output is constant in S , then return TRUE;

-Otherwise, return FALSE.

The work used R [164] to run the model and perform the analysis with the package/function `extraTrees` [174].

4.3.5 Neural Networks

A flexible generalized linear regression functions are feed-foward neural networks. They are non-linear regression models that are very flexible due to the great amount of parameters they have which makes it easy to approximate any smooth function. In a one hidden layer, the input values are distributed to the inputs of the hidden units in the second layer. The hidden units aggregate their inputs, add a constant bias, and take a fixed function ϕ_h of the result [180]. The output units of the hidden layer are the same as the input units but with an output function ϕ_o . The output is given as:

$$y_k = \phi_o \left(\alpha_k + \sum_h w_{hk} \phi_h \left(\alpha_h + \sum_i w_{ih} x_i \right) \right) \quad (4.15)$$

The ‘activation function’ ϕ_h of the hidden layer units is almost always a logistic function:

$$l(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (4.16)$$

and the output units are linear logistic, or threshold units [180]. Equation 4.15 can be rewritten to allow more than one hidden layer and allow ‘skip-layer’ connects from input to output as:

$$y_k = \phi_o \left(\alpha_k + \sum_{i \rightarrow k} w_{ik} x_i + \sum_{j \rightarrow k} w_{jk} \phi_h \left(\alpha_j + \sum_{i \rightarrow j} w_{ij} x_i \right) \right). \quad (4.17)$$

Equation 4.17 allows non-linear units to perturb functional form [180]. Thus the regression function can be parameterized by weights, w_{ij} , for every link in the network (zeros for non-existent links).

Feed-forward neural networks can parametrize a general non-linear function [180]. Cybenko [24], Funahashi [36], Hornik et al. [49] have shown that neural networks with linear output units can approximate any continuous function uniformly on compact sets by decreasing the size of the hidden layer [180]. The weights of the model are chosen so that it minimizes the fitting criterion in a form like least squares:

$$E = \sum_p \|t^p - y^p\|^2, \quad (4.18)$$

where t^p is the target and y^p is the output. Other minimizing methods have been established, such as maximum likelihood given as:

$$E = \sum_p \sum_k \left[t_k^p \log \frac{t_k^p}{y_k^p} + (1 - t_k^p) \log \frac{1 - t_k^p}{1 - y_k^p} \right]. \quad (4.19)$$

For a multinomial log-linear model with K classes uses a neural network with K outputs and has a negative conditional log-likelihood given as:

$$E = \sum_p \sum_k -t_k^p \log p_k^p, \quad (4.20)$$

where $p_k^p = \frac{e^{y_k^p}}{\sum_{c=1}^K e^{y_k^c}}$. Equation 4.20 has one of the targets, t_k^p as one for the class that occurred while the rest will zeros, which is known as 'softmax' [180]. This particular equation is the one used in the study where $K = 3$ for our three distinct classes (0 - No ENSO event, 1 - El Niño, and 2 - La Niña).

The work used R [164] to run the model and perform the analysis with the package/function `nnet` [180].

4.3.6 Lasso/Ridge Regression

A popular regression method are lasso and ridge, which was introduced by Tibshirani [178], that uses an l_1 and l_2 penalty to obtain a sparse solution [34]. If we have a response variable $Y \in \mathbb{R}$ and a predictor vector $X \in \mathbb{R}^p$, then we can approximate the regression function as a linear model:

$$E(Y|X = x) = \beta_0 + x^T \beta. \quad (4.21)$$

The linear regression works with N observation pairs (x_i, y_i) . In general, for an unstandardized case, Friedman et al. [34] proposes an elastic net give as:

$$\min_{\beta_0, \beta \in \mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right], \quad (4.22)$$

where

$$P_\alpha(\beta) = (1 - \alpha)\frac{1}{2}\|\beta\|_{l_2}^2 + \alpha\|\beta\|_{l_1} \quad (4.23)$$

$$= \sum_{j=1}^p \left[\frac{1}{2}(1 - \alpha)\beta_j^2 + \alpha|\beta_j| \right] \quad (4.24)$$

is the *elastic-net penalty* [34, 201]. The term P_α dictates the type of penalty if it is the ridge-regression ($\alpha = 0$), the lasso penalty ($\alpha = 1$), or a combination of the two ($0 < \alpha < 1$). The penalty is ideal in cases when p is much greater than N or when there are many correlated predictor variables [34].

The Ridge regression decreases the coefficients of correlated predictors towards each other. If there are k identical predictors, they each get coefficients of $\frac{1}{k}$ th size. The use of a ridge regression is ideal when there are any predictors and all have non-zero coefficients [34].

The Lasso works quite a bit different from ridge. Instead of providing equal weight to all the predictors, it will tend to pick one and ignore the rest [34]. The Lasso penalty is related to Laplace prior, which expects many coefficients to be close to zero and a small subset to be larger and nonzero [34, 35].

The elastic net is when there is a combination of the Ridge regression and Lasso. If $\alpha = 1 - \epsilon$ for some small $\epsilon > 0$, then it performs as if it were a Lasso. However, this removes

any degeneracies and abnormal behavior caused by extreme correlations [34]. As α increases from 0 to 1, the sparsity of Equation 4.22 increases monotonically from 0 to the sparsity of the lasso solution.

Using Equation 4.22 and if we have estimates $\tilde{\beta}_0$ and $\tilde{\beta}_l$ where $l \neq j$, then we can partially optimize with respect to β_j . To compute the gradient at $\beta_j = \tilde{\beta}_j$, it only exists if $\tilde{\beta}_j \neq 0$. For the case when $\tilde{\beta}_j > 0$, then we can compute it as:

$$\frac{\partial R}{\partial \beta_j} \Big|_{\beta=\tilde{\beta}} = -\frac{1}{N} \sum_{i=1}^N x_{ij} \left(y_i - \tilde{\beta}_0 - x_i^T \tilde{\beta} \right) + \lambda(1 - \alpha) + \lambda\alpha. \quad (4.25)$$

A similar expression exists for the cases when $\tilde{\beta}_j < 0$ and $\tilde{\beta}_j = 0$. Performing some calculus manipulation shown by Donoho and Johnstone [25] reveals that the coordinate-wise update as:

$$\tilde{\beta}_j \leftarrow \frac{S \left(\frac{1}{N} \sum_{i=1}^N x_{ij} \left(y_i - \tilde{y}_i^{(j)} \right), \lambda\alpha \right)}{1 + \lambda(1 - \alpha)}, \quad (4.26)$$

where

- $\tilde{y}_i^{(j)} = \tilde{\beta}_0 + \sum_{l \neq j} x_{il} \tilde{\beta}_l$ is the fitted value excluding the contribution from x_{ij} , which shows that $y_i - \tilde{y}_i^{(j)}$ is the *partial residual* for fitting β_j . This is due to the standardization given as, $\frac{1}{N} \sum_{i=1}^N x_{ij} \left(y_i - \tilde{y}_i^{(j)} \right)$, is the simple least-squares coefficient when fitting the partial residual to x_{ij} .

- $S(z, \gamma)$ is the soft-thresholding operator with the value give as:

$$\text{sign}(z)(|z| - \gamma)_+ = \begin{cases} z - \gamma & \text{if } z > 0 \text{ and } \gamma < |z| \\ z + \gamma & \text{if } z < 0 \text{ and } \gamma < |z| \\ 0 & \text{if } \gamma \geq |z|. \end{cases} \quad (4.27)$$

The details of the derivation of the equations are noted in Friedman et al. [33].

Kooij et al. [67] suggested that the algorithm is optimal since it computes the least-squares on the partial residual, applies soft-thresholding to take care of the lasso contribution to the penalty, and applies a proportional shrinkage for the ridge penalty [34].

The work used R [164] to run the model and perform the analysis with the package/function `glmnet` [34].

4.4 Results and Discussion

All models were trained on monthly data sets from 1948 until 1994 (a few years before El Niño of 1997/1998 to hyndcast whether or not the model can realistically isolate an ENSO event. The notation of class 0 signifies the event of no ENSO event, class 1 is an El Niño event and class 2 is La Niña event. We use such class definition unless explicitly listed in

each section. We provide detailed results from each of the competing models for climate division six. These results include confusion matrices and F1 values. Climate divisions 1, 2, 3, 4, 5, and 7 are located in Appendix D through J.

4.4.1 Vector Generalized Linear Models

First, we present the results from the three VGL models.

4.4.1.1 Main Effect

The main effects models attained a very poor classification accuracy for both gauge and modeled data as evidenced by the two zeroes on the diagonals of the confusion matrices corresponding to El Niño and La Niña (see Figure 4.1, Figure 4.2, Table 4.2 and Table 4.3). On the testing data El Niño and La Niña were completely misclassified. Even on the training data, only zero out of 29 and three out of 24 months El Niño months were correctly classified and only zero out of four and two out of two months La Niña months were correctly classified.

Figure 4.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division six

Table 4.2: F1 scores for training and testing for main effect on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NaN	0.50

Figure 4.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division six

Table 4.3: F1 scores for training and testing for main effect on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.54	Class 0:	0.99	0.50
Class 1:	0.20	0.56	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	NA

In Chapter Two and Three, we identified that SOI was not the best candidate to associate ENSO events for the majority of California’s seven climate divisions. Table 4.4 lists in frequency order the positive predictors for the gauge time-series for all climate divisions. Even though the model did not perform well due to the limited amount of ENSO events, it provided a listing of positive predictors to represent California. The first two listings show an obvious zero to demonstrate that the two popular climate indices, SOI and PDO, are not meaningful predictors to ENSO. Moreover, our results show that PDSI, a drought index, is

highly subtle to precipitation variation across the state. Such claim is clearly highlighted in Le et al. [71] article where they used a similar drought index, PZI, to measure precipitation in climate division six and seven. Furthermore, the table also shows climate indices like Iceland SLP and Negasaki SLP that are uncommon to use as precipitation predictors but have a strong correlation. The fourth important predictor is what we called PDO_BIN_BELOW which are PDO negative extremes at 95% confidence interval from an ARMA model. Thus, the representation of a decomposed time-series of PDO is what the main effect VGL model finds meaningful. Furthermore, the model finds Darwin SLP and Tahiti SLP as positive predictors as oppose to SOI, even though Darwin SLP and Tahiti SLP are the composites of SOI [168].

Table 4.4: Gauge predictors frequency for California training from 1948 to 2018

Freq.	Index
0	SOI
0	PDO
5	PDSI
4	ICELAND_SLP
3	NEGASAKI_SLP
3	PDO_BIN_BELOW
3	SW_MONSOON_REGION_RAINFALL

Continued on next page

Table 4.4 – *Continued from previous page*

Freq.	Index
2	WP
2	OBO_U30_INDEX
2	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX_BIN_ABOVE
2	DARWIN_SEA_LEVEL_PRESSURE
2	PNA
2	PDO_BIN_BELOW
2	TAHITI_SEA_LEVEL_PRESSURE
1	TSA
1	AO
1	NTA
1	NORTH ATLANTIC_SST
1	TSA_BIN_BELOW
1	ATLANTIC_TRIPOLE_SST_EOF_BIN_BELOW
1	TSA_BIN_ABOVE
1	NORTH EAST_BRAZIL_RAINFALL_ANOMALY
1	GILBRALTAR_SLP_BIN_ABOVE
1	MADROS_SLP
1	TRADE_WINDS_INDEX_175W_5N_5S_CENTRAL_PACIFIC
1	ENSO_PRECIP_INDEX

Continued on next page

Table 4.4 – *Continued from previous page*

Freq.	Index
1	SAM
1	TNI
1	EA_WR
1	EP_NP
1	SOLAR_FLUX_BIN_ABOVE
1	ZONAL_WINDS_EQUATOR_200MB
1	NPI_BIN_BELOW
1	PACIFIC_WARMPOOL
1	NINO1_2
1	NAO
1	TRADE_WINDS_INDEX_135W_120W_5N_5S_EAST PACIFIC
1	OLWR_EQUATOR_BIN_BELOW
1	WHWP

Hyndcasting the modeled time-series also generated a listing of positive predictors. Table 4.5 contains such listing. Again, we notice that SOI and PDO have zero occurrences of all of California. Not surprising, PDSI again results in the most occurrence but only in three climate divisions as oppose to five. The predictor Nagasaki SLP, Darwin SLP, and Tahiti SLP make their appearances as well. The results closely reflect what was found in the

gauge time-series, even though we discussed and found key differences between both data sets. Table 4.6 shows all the positive predictors for each climate division for the gauge and modeled data sets. The majority of the predictors are the unprocessed climate indices with a few mixture of above (positive) and below (negative) extremes.

Table 4.5: Model predictors frequency for California training from 1948 to 2018

Freq.	Index
0	SOI
0	PDO
3	PDSI
3	AMM_BIN_ABOVE
3	NAGASAKI_SLP
3	PDSI_BIN_BELOW
2	WP
2	TSA
2	DARWIN_SEA_LEVEL_PRESSURE
2	TAHITI_SEA_LEVEL_PRESSURE
2	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX_BIN_ABOVE
2	NINO1_2
2	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX_BIN_BELOW
2	QBO_U30_INDEX

Continued on next page

Table 4.5 – *Continued from previous page*

Freq.	Index
2	ER_WR
2	SW_MONSOON_REGION_RAINFALL
1	AMO_SMOOTHED_BIN_ABOVE
1	TNA_BIN_BELOW
1	ZONALLY_AVERAGE_500MB_TEMP_ANOMALIES_BIN_BELOW
1	TRADE_WINDS_INDEX_175W_140W_5N_5S_CENTRAL_PACIFIC
1	PNA
1	NAO_AZO_BIN_BELOW
1	TROPICAL_PACIFIC_SST_EOF
1	PDO_BIN_BELOW
1	ATLANTIC_TRIPOLE_SST_EOF_BIN_BELOW
1	TRADE_WINDS_INDEX_135W_120W_5N_5S_EAST_PACIFIC
1	GILBRALTAR_SLP
1	SUNSPOT_INDEX_BIN_BELOW
1	TNI_BIN_BELOW
1	NPI
1	ATLANTIC_TRIPOLE_SST_EOF_BIN_ABOVE
1	ICELAND_SLP
1	AO_BIN_BELOW

Continued on next page

Table 4.5 – *Continued from previous page*

Freq.	Index
1	ONI_BIN_ABOVE
1	NINO4_BIN_ABOVE
1	TSA_BIN_BELOW
1	AMO_SMOOTHED
1	CENTRAL_INDIAN_PRECIP

4.4.1.2 Two-Way Interaction

The two-way interaction models again attained a very poor classification accuracy for both gauge and modeled data as evidenced by the two zeroes on the diagonals of the confusion matrices corresponding to El Niño and La Niña (see Figure 4.3, Figure 4.4, Table 4.7 and Table 4.8). On the testing data El Niño and La Niña were completely misclassified. Even on the training data, only zero out of 29 and zero out of 27 months El Niño months were correctly classified and only zero out of two La Niña months were correctly classified.

Table 4.6: Gauge and modeled predictors per climate division of the main effect in California training from 1948 to 2018

CD	Gauge	Modeled
1	PDSI	PDSI
	NAGASAKI.SLP	AMM.BIN.ABOVE
	TSA	NAGASAKI.SLP
	AO	AMO.SMOOTHED.BIN.ABOVE
	TNA	WP
	WP	TSA
	QBO.U30.INDEX	TNA.BIN.BELOW
2	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.ABOVE	PDSI.BIN.BELOW
	NORTH_ATLANTIC_SST	AMM.BIN.ABOVE
	DARWIN_SEA_LEVEL_PRESSURE	DARWIN_SEA_LEVEL_PRESSURE
	TSA.BIN.BELOW	PDSI
	ATLANTIC_TRIPOLE_SST_EOF.BIN.BELOW	ZONALLY_AVERAGE_500MB_TEMP_ANOMALIES.BIN.BELOW
	TSA.BIN.ABOVE	TRADE_WINDS_INDEX_175W_175W_140W_5N_5S_CENTRAL_PACIFIC
	NORTH_EAST_BRAZIL_RAINFALL_ANOMALY	WP
3		PNA
	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX	PDSI.BIN.BELOW
	PDO.BIN.BELOW	NAO_AZO.BIN.BELOW
	DARWIN_SEA_LEVEL_PRESSURE	DARWIN_SEA_LEVEL_PRESSURE
	GILBRALTAR_SLP_BIN.ABOVE	TAHITI_SEA_LEVEL_PRESSURE
	QBO.U30.INDEX	TROPICAL_PACIFIC_SST_EOF
	PDSI	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.ABOVE
4		TSA
	MADRAS.SLP	PDSI.BIN.BELOW
	TRADE_WINDS_INDEX_175W_140W_5N_5S_CENTRAL_PACIFIC	NAGASAKI.SLP
	ENSO_PRECIP_INDEX	AMM.BIN.ABOVE
	ICELAND.SLP	TAHITI_SEA_LEVEL_PRESSURE
	SAM	NINO1.2
	PDSI	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.BELOW
	PNA	PDO.BIN.BELOW
	WP	ATLANTIC_TRIPOLE_SST_EOF.BIN.BELOW
	PDO	QBO.U30.INDEX
	TNI	EA_WR
	EA_WR	
	EP_NP	
	SW_MONSOON_REGION_RAINFALL	
5	OLWR_EQUATOR	
	SOLAR_FLUX_BIN.ABOVE	
	ICELAND.SLP	PDSI
	PDSI	NAGASAKI.SLP
	NAGASAKI.SLP	QBO.U30.INDEX
	PDO	GILBRALTAR.SLP
	PNA	TRADE_WINDS_INDEX_135W_120W_5N_5S_EAST_PACIFIC
6	ZONAL_WINDS_EQUATOR_200MB	EA_WR
	NPI.BIN.BELOW	SUNSPOT_INDEX.BIN.BELOW
	ICELAND.SLP	TNI.BIN.BELOW
	SW_MONSOON_REGION_RAINFALL	NPI
	PACIFIC_WARMPOOL	ATLANTIC_TRIPOLE_SST_EOF.BIN.ABOVE
	PDO.BIN.BELOW	SW_MONSOON_REGION_RAINFALL
7	TAHITI_SEA_LEVEL_PRESSURE	NINO1.2
	ZONAL_WINDS_EQUATOR_200MB	
	NINO1.2	ICELAND.SLP
	SW_MONSOON_REGION_RAINFALL	AO.BIN.BELOW
	ICELAND.SLP	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.BELOW
	PDO.BIN.BELOW	SW_MONSOON_REGION_RAINFALL
	TAHITI_SEA_LEVEL_PRESSURE	ONI.BIN.ABOVE
	TRADE_WINDS_INDEX_135W_120W_5N_5S_EAST_PACIFIC	NINO4.BIN.ABOVE
	OLWR_EQUATOR.BIN.BELOW	TSA.BIN.BELOW
	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.ABOVE	AMO.SMOOTHED
	WHWP	GLOBAL_MEAN_LAND_OCEAN_TEMPERATURE_INDEX.BIN.ABOVE
	NAO	CENTRAL_INDIAN_PRECIP
	NAGASAKI.SLP

Figure 4.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division six

Table 4.7: F1 scores for training and testing for two-way interaction on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.53	Class 0:	0.97	0.48
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.67	0.75	Class 2:	NaN	0.48

Figure 4.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division six

Table 4.8: F1 scores for training and testing for two-way interaction on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	NA

4.4.1.3 Three-Way Interaction

The three-way Interaction models again attained a very poor classification accuracy for both gauge and modeled data as evidenced by the two zeroes on the diagonals of the confusion matrices corresponding to El Niño and La Niña (see Figure 4.5, Figure 4.6, Table 4.9 and Table 4.10. On the testing data El Niño and La Niña were completely misclassified. Even on the training data, only zero out of 29 and zero out of 27 months El Niño months were

correctly classified and only zero out of four and zero out of two La Niña months were correctly classified.

Figure 4.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division six

Table 4.9: F1 scores for training and testing for three-way interaction on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NaN	0.50

Figure 4.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division six

Table 4.10: F1 scores for training and testing for three-way interaction on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	NA

4.4.2 XGBoost

The XGBoost classifier overfits the training data and attains perfect classification but still completely misclassifies the testing data (see Figure 4.7, Figure 4.9, Table 4.11 and Table 4.12). The testing data only zero out of 4 and zero out of 9 months El Niño months were correctly classified and only zero out of two and zero out of one month La Niña months were correctly classified.

Figure 4.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division six

Table 4.11: F1 scores for training and testing for XGBoost on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.50
Class 1:	1.00	1.00	Class 1:	NaN	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Figure 4.8 and Figure 4.10 show the top 20 positive predictors of the XGBoost model. The prediction power is measured but what percentage of each climate index is factored into the model. For example, referring to Figure 4.8, the SW Monsoon Region Rainfall corresponds to 12.63% to successfully predict an ENSO event. For both the modeled and gauge time-series, we clearly see that PDSI, Tahiti SLP, and Darwin SLP still remain as the top climate index predictors as we noted with the main effect of the VGL models. Furthermore, PDO is now a contender by contributing roughly 1.7% to the predictive model for both data

sets. However, we cannot say the same for SOI as the dominate climate index is again its index composition.

Figure 4.8: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division six

Figure 4.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division six

Table 4.12: F1 scores for training and testing for XGBoost on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NaN	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Figure 4.10: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division six

4.4.3 Random Forest Classifier

Similar to the XGBoost classifier, the Random Forest classifier overfits the training data and attains perfect classification but still completely misclassifies the testing data (see Figure 4.11, Figure 4.13, Table 4.13 and Table 4.14). The testing data results show only

zero out of one and zero out of one months El Niño months were correctly classified and only zero out of zero and zero out of zero months La Niña months were correctly classified. The quantity of the values are different from other models due to the Random Forest is incapable of working with missing values for the climate indices. Ideally one will find a suitable method to impute the values, but we decided not to as that will skew the findings.

Figure 4.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division six

Table 4.13: F1 scores for training and testing for Random Forest on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	NA	NA	Class 2:	NA	NA

The Random Forest positive predictors are displayed in Figure 4.12 and Figure 4.14. The results differ from the what was noted for the VGLMs and XGBoost. The most signif-

icant positive predictor is SW Monsoon Region Rainfall BIN_ABOVE for both gauge and modeled time-series at 2.57% and 3.54%. A new positive predictor that the model found as significant is Solar Flux. These new indices that are significant are not commonly associated with ENSO, but are clearly important in some form and further research should be considered. Another important result is again we do not have SOI and PDO as a positive predictor, but Tahiti SLP is.

Figure 4.12: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division six

Figure 4.13: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division six

Table 4.14: F1 scores for training and testing for Random Forest on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	NA

Figure 4.14: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division six

4.4.4 Extra Trees

Similar to the previous study tree-models, the Extra Trees model also displays some overfitting with the gauge training data. However, a poor performance of the model is seen with the tests data. In addition, we notice that the model was not able to classify observa-

tions of types class 1 (El Niño) and class 2 (La Niña) (see Figure 4.15, Figure 4.16, Table 4.15 and Table 4.16).

With the modeled time-series data, one can again see an overfitting of the trained model. On the corresponding test data, a good performance of the modeled while classifying observations of type class 0 is observed against a noticeable inability to well classify observations of types class 1 and class 2. This can be interpreted as the modeled predicts well when no event happens opposed to not being able to capture minimal events (low ENSO occurrence) that are taking place.

Figure 4.15: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division six

Figure 4.16: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division six

4.4.5 Neural Networks

In the case of the neural networks with the gauge time-series, one can observe that there is a poor performance of the models on the training and testing sets. In actuality, the neural networks are classifying very well when no event is taking place. On contrary, the neural networks displays a crucial inability to predict class 1 and class 2 events. Figure 4.17, Figure 4.19, Table 4.17, and Table 4.18 show the results. Similar results are seen with the neural networks in the case of modeled time series. Precisely stated, the model does a great a job on classifying when there is not a single event that is occurring, whereas it affords some inabilities to capture events of type class 1 and class 2. The testing data results show only zero out of one and zero out of one months El Niño months were correctly classified and only zero out of zero and zero out of zero months La Niña months were correctly classified.

Figure 4.17: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division six

The positive predictors of the neural networks are all insignificant. The percentage of contribution has a very low percentage where the highest value is 0.74%. All of the positive predictors are essentially the entirety of the data set as it lists all the cases that are above (positive extreme events). Unfortunately, little to no information can be obtain from the neural networks shown in Figure 4.18 and Figure 4.20.

Figure 4.18: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division six

(a) Training

(b) Testing

Figure 4.19: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division six

Table 4.15: F1 scores for training and testing for Extra Trees on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table 4.16: F1 scores for training and testing for Extra Trees on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table 4.17: F1 scores for training and testing for Neural Network on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.99	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	NA	Class 2:	NA	NA

Table 4.18: F1 scores for training and testing for Neural Network on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.97	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	NA

Figure 4.20: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division six

4.4.6 Ridge Regression

The ridge regression will deviate how we were conducting the models. Instead of having a three distinct cases of No ENSO event, El Niño event, and La Niña event, the ridge regression will be performed on a binary case. In other words, one model will focus in deciphering

between a no ENSO event and an El Niño event and the second model for a no ENSO event and La Niña event. The ridge regression model will still be performed for the gauge and modeled time series. As a note, the Ridge regression works with a certain amount positive events. If there are a few events, then the model cannot perform due to the inadequate representation of the event of interest. Due to this reason, climate division six for La Niña is not shown.

Referring to Figure 4.21, Figure 4.23, Figure 4.25, we note again that model is training well, specifically when we test it on what the model has already learned from. Not surprisingly, the testing on the last 14 years the model is not able to properly classify either an El Niño or La Niña events. The results from the confusion matrix are verified through the F1 scores in Table 4.19, Table 4.20, and Table 4.21

Figure 4.21: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division six

Figure 4.22: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division six

Similar to the neural networks, the positive predictors are given almost equal weight as it was noted by Friedman et al. [34]. There is a variation of predictors from the various models that we conducted, but the model still favors the first climate indices that it reads. These are all cases of what we classify as ABOVE, which signifies the extreme events of each climate index that is a positive extreme at the 95% confidence interval.

Figure 4.23: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division six

Figure 4.24: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division six

Figure 4.25: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division six

Figure 4.26: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division six

4.4.7 Lasso Regression

The Lasso regression is very similar to the Ridge Regression. Instead of having a three distinct cases of No ENSO event, El Niño event, and La Niña event, the lasso regression will be performed on a binary case. In other words, one model will focus in deciphering between

a no ENSO event and an El Niño event and the second model for a no ENSO event and La Niña event. The Lasso regression model will still be performed for the gauge and modeled time series. As a note, the Lasso regression also works with a certain amount positive events. If there are a few events, then the model cannot perform due to the inadequate representation of the event of interest. Due to this reason, climate division six for La Niña is not shown.

Referring to Figure 4.27, Figure 4.29, Figure 4.31, we note that the Lasso regression is very different than the Ridge. For all the training we have a perfect fit, and unfortunately, we overfit the model. The testing of the data set is not great either. For almost all the cases, the model classifies decently the cases when no ENSO event occurs. The case of more concern, the model misdiagnosis the cases of El Niño and La Niña. The only case where the model classifies well is for the testing of El Niño on the gauge time-series as seen in Figure 4.27b.

Figure 4.27: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division six

The top 20 positive predictors are very poor. The model classifies on average three climate indices, while the rest are essentially zero where they play little to no role in the model. Just as the ridge regression, the ABOVE predictors are classified as positive predictors, more or less to the order that the model processed the potential predictors.

Figure 4.28: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division six

(a) Training

(b) Testing

Figure 4.29: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division six

Table 4.19: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table 4.20: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table 4.21: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.50	Class 0:	1.00	NaN
Class 1:	0.00	0.50	Class 1:	NaN	NaN

Table 4.22: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.85	0.87
Class 1:	1.00	1.00	Class 1:	0.14	0.87

Table 4.23: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.47
Class 1:	1.00	1.00	Class 1:	0.00	0.47

Figure 4.30: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division six

Figure 4.31: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division six

Table 4.24: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division six

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	NaN
Class 1:	1.00	1.00	Class 1:	0.00	NaN

Figure 4.32: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division six

4.5 Conclusion

The objective of the study is to attempt to create an assimilated product of a new climate index to represent California's ENSO events. Overall, the models that we used did not perform to our expectations. This is due to the low sample size of the events of interest

within the timespan of the study. If the data were to be extended, then it is likely that the models will perform as intended and particular to our custom made model of VGL. Based on the various model simulations and learning patterns, results were shown as to which combination of climate indices represent El Niño. SOI and PDO are not a well representation of ENSO as we believed even though Chapters Two and Three showed they have strong correlations. Instead, SOI's composition of Darwin and Tahiti sea level pressure were a strong contender as a positive predictor. Thus the results allow climate scientist to reevaluate their methods when they attempt to forecast the next ENSO phase.

5. Conclusion

Forecasting ENSO events is the unsolved problem of climate scientist. There are various methods researches have attempted and had some success on isolated events, but most have been unsuccessful. The El Niño year of 2015/2016 was a great example of placing the latest technology and understanding of the ENSO problem. Unfortunately, it was a huge miss and people, like scientist, were left puzzled why all the right conditions were met yet there was no precipitation. Thus, this dissertation focused on using historical precipitation data to attempt to understand and model precipitation patterns and find a connection with ENSO. The first study focused on historical hourly precipitation data and a modeled dataset to model extreme precipitation that are related to the three strongest El Niño years (1982/1983, 1997/1998, and 2015/2016). The data consisted of 71 years of hourly observation at various regions in California that was grouped into seven divisions that categorizes California's climate divisions. We used wavelets to unveil the hidden patterns and correlation between the gauge and modeled datasets with two climate divisions, Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO). The results showed that the gauge and modeled data set for the majority are very similar but are different in their overall representation of precipitation patterns in California. The algorithm that was used to populate the data set for the modeled time-series incorporate area weights that are beyond the climate divisional boundaries, include data sets from other observable means than gauge, and the method used to populate missing data. The differences are numerically clear with the high

values obtained from the spectra analysis. The key differences have shown how well they are able to represent ENSO, specifically PDO. Thus, the study found that it is naive to state that SOI is the swole representation of precipitation and ENSO.

The second study had two parts. The first focused on creating a model using an ARMA model where we used it to analyze the residual time-series of the gauge and modeled data sets. It focused on how well the true precipitation measurements correlated with defined ENSO years. The results showed that the patterns of excessive precipitation matched to all El Niño years for both gauge and modeled time-series. However, the excessive drought years had few matches, which resulted in a lower correlation to ENSO. This demonstrates that ENSO affects precipitation in California regionally then how the general audience interpret it as worldwide effect. The second part focused on an analysis of principle components from EMD of the gauge and modeled time-series with a comparison to three climate indices: SOI (Southern Oscillation Index), PDO (Pacific Decadal Oscillation), and AMO (Atlantic Multidecadal Oscillation). The results of this study demonstrate that SOI is not necessarily the highest correlated climate index among the seven climate divisions of California. PDO and AMO are great representatives of precipitation patterns in California as it was shown in an ARMA and EMD analysis. The analysis exposed SOI may be a naive way of associating ENSO as a local contributor to precipitation.

The third study focused on an assimilated product of a new climate index to represent California's ENSO events. Overall, the models that we used did not perform to our expectations. This is due to the low sample size of the ENSO events. Based on the various model

simulations and learning patterns, results were shown as to which combination of climate indices well-represent El Niño. SOI and PDO are not a well representation of ENSO as the previous studies showed. Instead, SOI's composition of Darwin and Tahiti sea level pressure were a strong contender as a positive predictor. For California, the best climate index representation is PDSI, which is a drought index. Thus the results allow climate scientist to reevaluate their methods when they attempt to forecast the next ENSO phase.

REFERENCES

- [1] J. T. Abatzoglou, K. T. Redmond, and L. M. Edwards, “Classification of Regional Climate Variability in the State of California,” *Journal of Applied Meteorology and Climatology*, vol. 48, no. 8, pp. 1527–1541, 2009.
- [2] M. R. Allen and W. J. Ingram, “Constraints on future changes in climate and the hydrologic cycle,” *Nature*, vol. 419, no. 6903, p. 228, 2002.
- [3] Z. Asong, M. Khaliq, and H. Wheeler, “Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models,” *Climate Dynamics*, vol. 47, no. 9/10, pp. 2901–2921, 2016. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=aph&AN=118732174&site=eds-live>
- [4] S. Braun and H. F. Pierce, “NASA Facts: TRMM Instruments,” http://trmm.gsfc.nasa.gov/overview_dir/pr.html, Jun. 2011. [Online]. Available: http://trmm.gsfc.nasa.gov/overview_dir/pr.html
- [5] L. Breiman, “Random forests,” *UC Berkeley TR567*, 1999.
- [6] —, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees.*, ser. Wadsworth Statistics/Probability Series.

- Wadsworth Advanced Books and Software, Belmont, CA, 1984. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=msn&AN=MR726392&site=eds-live>
- [8] R. Buizza, “Horizontal resolution impact on short- and long-range forecast error,” *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 649, pp. 1020–1035, 2010.
- [9] P. Caldwell, “California wintertime precipitation bias in regional and global climate models,” *Journal of Applied Meteorology and Climatology*, vol. 49, no. 10, pp. 2147–2158, 2010.
- [10] M. A. Cane, “The evolution of El Niño, past and future,” *Earth and Planetary Science Letters*, vol. 230, no. 3-4, pp. 227–240, 2005.
- [11] A. Capotondi, T. Wittenberg Andrew, M. Newmaman, E. Di Lorenzo, J.-Y. Yu, P. Braconnot, J. Cole, B. Dewitte, B. Giese, E. Guilyardi, F.-F. Jin, K. Karneuskas, B. Kirtman, T. Lee, N. Schneider, Y. Xue, and S.-W. Yeh, “UNDERSTANDING ENSO DIVERSITY,” *Bulletin of the American Meteorological Society*, vol. 96, no. 6, p. 921, 2015. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26219618&site=eds-live>
- [12] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794.

- [13] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, and Y. Li, *xgboost: Extreme Gradient Boosting*, 2019. [Online]. Available: <https://CRAN.R-project.org/package=xgboost>
- [14] R. W. Cleveland, “STL: a seasonal-trend decomposition procedure based on loess,” *Journal of official statistics : JOS ; an international quarterly*, vol. 6, no. 1, pp. 3–33, 1990. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edszbw&AN=EDSZBW259706647&site=eds-live>
- [15] Climatic Research Unit and University of East Anglia, “Gilbraltar SLP Index,” <https://crudata.uea.ac.uk/cru/data/nao/nao.dat>, Aug. 2019.
- [16] —, “Iceland SLP Index,” <https://crudata.uea.ac.uk/cru/data/nao/nao.dat>, Aug. 2019.
- [17] —, “Madras SLP Index,” <https://crudata.uea.ac.uk/cru/data/madrasmslp/madrasmslp.dat>, Aug. 2019.
- [18] —, “North Atlantic Oscillation Azores Index,” https://crudata.uea.ac.uk/cru/data/nao/nao_azo.dat, Aug. 2019.
- [19] —, “Nagasaki SLP Index,” <https://crudata.uea.ac.uk/cru/data/nagasaki/nagasakiPRES.dat>, Aug. 2019.
- [20] —, “Trans Polar Index,” <https://crudata.uea.ac.uk/cru/data/tpi/tpi.dat>, Aug. 2019.

- [21] M. Collins and B. Sinha, “Predictability of decadal variations in the thermohaline circulation and climate,” *Geophysical Research Letters*, vol. 30, no. 6, 2003.
- [22] M. Costa-Cabral, J. S. Rath, W. B. Mills, S. B. Roy, P. D. Bromirski, and C. Milesi, “Projecting and Forecasting Winter Precipitation Extremes and Meteorological Drought in California Using the North Pacific High Sea Level Pressure Anomaly,” *Journal of Climate*, vol. 29, no. 13, pp. 5009–5026, 2016. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=gft&AN=116817523&site=eds-live>
- [23] K. Coughlin and K.-K. Tung, “Elevenyear solar cycle signal throughout the lower atmosphere,” *Journal of Geophysical Research: Atmospheres*, vol. 109, no. D21, 2004.
- [24] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [25] D. L. Donoho and J. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [26] V. Dulière, Y. Zhang, and E. P. Salathé Jr, “Extreme precipitation and temperature over the US Pacific Northwest: A comparison between observations, reanalysis data, and regional models,” *Journal of Climate*, vol. 24, no. 7, pp. 1950–1964, 2011.
- [27] B. Efron and D. V. Hinkley, “Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information,” *Biometrika*, vol. 65, no. 3, pp. 457–483, 1978.

- [28] H. El-Askary, M. Allali, C. Rakovski, A. Prasad, M. Kafatos, and D. Struppa, “Computational methods for climate data,” *WIREs: Computational Statistics*, vol. 4, no. 4, pp. 359–374, 2012. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=aph&AN=76918190&site=eds-live>
- [29] D. B. Enfield, A. M. MestasNuñez, and P. J. Trimble, “The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental us,” *Geophysical Research Letters*, vol. 28, no. 10, pp. 2077–2080, 2001.
- [30] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 2008, vol. 2.
- [31] A. O. Fierro, “Relationships between California rainfall variability and large-scale climate drivers,” *International Journal of Climatology*, no. 13, p. 3626, 2014. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsgao&AN=edsgcl.397824941&site=eds-live>
- [32] P. Flandrin, G. Rilling, and P. Goncalves, “Empirical mode decomposition as a filter bank,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.
- [33] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani *et al.*, “Pathwise coordinate optimization,” *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.
- [34] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization Paths for Generalized

- Linear Models via Coordinate Descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010. [Online]. Available: <http://www.jstatsoft.org/v33/i01/>
- [35] J. H. Friedman, B. E. Popescu *et al.*, “Importance sampled learning ensembles,” *Journal of Machine Learning Research*, vol. 94305, pp. 1–32, 2003.
- [36] K.-I. Funahashi, “On the approximate realization of continuous mappings by neural networks,” *Neural Networks*, vol. 2, no. 3, pp. 183–192, 1989.
- [37] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [38] P. Gloersen and N. Huang, “Comparison of interannual intrinsic modes in hemispheric sea ice covers and other geophysical parameters,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 5, pp. 1062–1074, 2003.
- [39] Golden Gate Weather Services, “El Niño and La Niña Years and Intensities,” <https://ggweather.com/enso/oni.htm>, Aug. 2019.
- [40] T. C. Gouhier, A. Grinsted, and V. Simko, *R package biwavelet: Conduct Univariate and Bivariate Wavelet Analyses*, 2019. [Online]. Available: <https://github.com/tgouhier/biwavelet>
- [41] A. Grinsted, J. C. Moore, and S. Jevrejeva, “Application of the cross wavelet transform and wavelet coherence to geophysical time series,” *Nonlinear Processes in Geophysics*, no. 5/6, p. 561, 2004. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://>

//search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsdoj&AN=edsdoj.2bab1d2a7d8348848f7539933f63d184&site=eds-live

- [42] N. B. Guttman and R. G. Quayle, “A historical perspective of US climate divisions,” *Bulletin of the American Meteorological Society*, vol. 77, no. 2, pp. 293–304, 1996.
- [43] R. Hafen, *stlplus: Enhanced Seasonal Decomposition of Time Series by Loess*, 2016. [Online]. Available: <https://CRAN.R-project.org/package=stlplus>
- [44] Z. Hao and A. AghaKouchak, “Multivariate Standardized Drought Index: A parametric multi-index model,” *Advances in Water Resources*, vol. 57, pp. 12–18, 2013. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0309170813000493&site=eds-live>
- [45] Z. Hao, F. Hao, V. P. Singh, Y. Xia, W. Ouyang, and X. Shen, “A theoretical drought classification method for the multivariate drought index based on distribution properties of standardized drought indices,” *Advances in Water Resources*, 2016. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsgao&AN=edsgcl.517835865&site=eds-live>
- [46] S. Hare and R. Francis, “Climate change and salmon production in the Northeast Pacific Ocean,” *Canadian Special Publication of Fisheries and Aquatic Sciences*, pp. 357–372, 1995.

- [47] S. R. Hare, “Low frequency climate variability and salmon production,” Ph.D. dissertation, University of Washington, 1996.
- [48] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [49] K. Hornik, M. Stinchcombe, H. White *et al.*, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [50] A. Y. Hou, G. Skofronick-Jackson, and E. F. Stocker, “The global precipitation measurement (gpm) mission: Overview and us science status,” in *EGU General Assembly Conference Abstracts*, vol. 15, 2013, p. 13604.
- [51] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [52] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeeen, *forecast: Forecasting functions for time series and linear models*, 2019. [Online]. Available: <http://pkg.robjhyndman.com/forecast>
- [53] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: the forecast

- package for R,” *Journal of Statistical Software*, vol. 26, no. 3, pp. 1–22, 2008. [Online]. Available: <http://www.jstatsoft.org/article/view/v027i03>
- [54] M. G. Jacox, E. L. Hazen, K. D. Zaba, D. L. Rudnick, C. A. Edwards, A. M. Moore, and S. J. Bograd, “Impacts of the 2015–2016 El Niño on the California Current System: Early assessment and comparison to past events,” *Geophysical Research Letters*, vol. 43, no. 13, pp. 7072–7080, 2016.
- [55] A. Kawamura, A. I. McKerchar, R. H. Spigel, and K. Jinno, “Chaotic characteristics of the Southern Oscillation Index time series,” *Journal of Hydrology*, vol. 204, no. 1-4, pp. 168–181, 1998.
- [56] C. L. Keppenne and M. Ghil, “Adaptive filtering and prediction of the Southern Oscillation index,” *Journal of Geophysical Research: Atmospheres*, vol. 97, no. D18, pp. 20 449–20 454, 1992.
- [57] R. A. Kerr, “A North Atlantic Climate Pacemaker for the Centuries,” *Science*, vol. 288, no. 5473, pp. 1984–1986, 2000. [Online]. Available: www.jstor.org/stable/3075527
- [58] M. Khademul Islam, R. M. Sayedur, S. Akimasa, and B. Pabitra, “EMPIRICAL MODE DECOMPOSITION ANALYSIS OF CLIMATE CHANGES WITH SPECIAL REFERENCE TO RAINFALL DATA,” *Discrete Dynamics in Nature and Society*, p. 318, 2006. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarl&AN=edsarl.354458F2&site=eds-live>
- [59] D. Khedhaouria, A. Mailhot, and A.-C. Favre, “Daily Precipitation Fields Modeling

- across the Great Lakes Region (Canada) by Using the CFSR Reanalysis,” *Journal of Applied Meteorology and Climatology*, vol. 57, no. 10, pp. 2419–2438, 2018.
- [60] D. Kim and H.-S. Oh, *EMD: Empirical Mode Decomposition and Hilbert Spectral Analysis*, 2018.
- [61] —, “Emd: A package for empirical mode decomposition and hilbert spectrum,” *The R Journal*, vol. 1, no. 1, pp. 40–46, 2009.
- [62] D. Kim, K. O. Kim, and H.-S. Oh, “Extending the Scope of Empirical Mode Decomposition using Smoothing,” *EURASIP Journal on Advances in Signal Processing*, vol. 2012:168, pp. 1–17, 2012.
- [63] D. Kim, M. Park, and H.-S. Oh, “Bidimensional Statistical Empirical Mode Decomposition,” *IEEE Signal Processing Letters*, vol. 19, pp. 191–194, 2012.
- [64] K. M. Kim and K. M. Lau, “Dynamics of monsoon-induced biennial variability in enso,” *Geophysical Research Letters*, vol. 28, no. 2, pp. 315–318, 2001.
- [65] J. R. Knight, “A signature of persistent natural thermohaline circulation cycles in observed climate,” *Geophysical Research Letters*, vol. 32, no. 20, 2005.
- [66] G. P. Können, P. D. Jones, M. H. Kaltofen, and R. J. Allan, “Pre-1866 Extensions of the Southern Oscillation Index Using Early Indonesian and Tahitian Meteorological Readings,” *Journal of Climate*, vol. 11, no. 9, p. 2325, 1998. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26243055&site=eds-live>

- [67] A. J. Kooij *et al.*, *Prediction accuracy and stability of regression with optimal scaling transformations*. Child & Family Studies and Data Theory (AGP-D), Department of Education and , 2007.
- [68] D. Kurtzman and B. R. Scanlon, “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*, vol. 43, no. 10, 2007.
- [69] A. J. M. Kuss and J. J. Gurdak, “Groundwater level response in us principal aquifers to ENSO, NAO, PDO, and AMO,” *Journal of Hydrology*, vol. 519, pp. 1939–1952, 2014.
- [70] K.-M. Lau and H. Weng, “Climate signal detection using wavelet transform: How to make a time series sing,” *Bulletin of the American Meteorological Society*, vol. 76, no. 12, pp. 2391–2402, 1995.
- [71] J. A. Le, H. M. El-Askary, M. Allali, and D. C. Struppa, “Application of recurrent neural networks for drought projections in California,” *Atmospheric Research*, vol. 188, pp. 100–106, 2017.
- [72] A. F. Levine and M. J. McPhaden, “How the July 2014 easterly wind burst gave the 2015–2016 El Niño a head start,” *Geophysical Research Letters*, vol. 43, no. 12, pp. 6503–6510, 2016.
- [73] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” *R News*,

- vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: <https://CRAN.R-project.org/doc/Rnews/>
- [74] Y.-C. Liu, P. Di, S.-H. Chen, and J. DaMassa, “Relationships of Rainy Season Precipitation and Temperature to Climate Indices in California: Long-Term Variability and Extreme Events,” *Journal of Climate*, vol. 31, no. 5, pp. 1921–1942, 2018. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=gft&AN=128527451&site=eds-live>
- [75] G. M. MacDonald and R. A. Case, “Variations in the Pacific Decadal Oscillation over the past millennium,” *Geophysical Research Letters*, vol. 32, no. 8, 2005.
- [76] T. Madsen and E. Figdor, *When it rains, it pours: global warming and the rising frequency of extreme precipitation in the United States*. Environment Texas Research & Policy Center, 2007.
- [77] N. J. Mantua and S. R. Hare, “The Pacific decadal oscillation,” *Journal of Oceanography*, vol. 58, no. 1, pp. 35–44, 2002.
- [78] N. J. Mantua, S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, “A Pacific interdecadal climate oscillation with impacts on salmon production,” *Bulletin of the American Meteorological Society*, vol. 78, no. 6, pp. 1069–1080, 1997.
- [79] J. Mantua Nathan, R. Hare Steven, Y. Zhang, M. Wallace John, and C. Francis Robert, “A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production,” *Bulletin of the American Meteorological Society*, vol. 78, no. 6, p. 1069, 1997. [Online]. Avail-

able: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26207496&site=eds-live>

- [80] C. Mass, A. Skalenakis, and M. Warner, “Extreme Precipitation over the West Coast of North America: Is There a Trend?” *Journal of Hydrometeorology*, vol. 12, no. 2, pp. 310–318, 2011.
- [81] J. W. Mays, D. Niyogi, and K. Scheeringa, “An Introduction to ENSO and its impact on Indiana Weather and Climate,” <https://climate.agry.purdue.edu/climate/publications/enso7-easy.pdf>, Apr. 2009. [Online]. Available: <https://climate.agry.purdue.edu/climate/publications/enso7-easy.pdf>
- [82] S. McClatchie, “State of the California Current 2015–16: comparisons with the 1997–98 El Niño,” *California Cooperative Oceanic Fisheries Investigations. Data Report*, vol. 57, 2016.
- [83] M. J. McPhaden, “Playing hide and seek with El Niño,” *Nature Climate Change*, vol. 5, no. 9, p. 791, 2015.
- [84] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz, “ENSO as an integrating concept in earth science,” *Science*, vol. 314, no. 5806, pp. 1740–1745, 2006.
- [85] C. E. Menkes, M. Lengaigne, J. Vialard, M. Puy, P. Marchesiello, S. Cravatte, and G. Cambon, “About the role of Westerly Wind Events in the possible development of an El Niño in 2014,” *Geophysical Research Letters*, vol. 41, no. 18, pp. 6476–6483, 2014.

- [86] Q. Min, J. Su, R. Zhang, and X. Rong, “What hindered the El Niño pattern in 2014?” *Geophysical Research Letters*, vol. 42, no. 16, pp. 6762–6770, 2015.
- [87] J. Mingers, “An empirical comparison of pruning methods for decision tree induction,” *Machine Learning*, vol. 4, no. 2, pp. 227–243, 1989.
- [88] S. Minobe, “A 50–70 year climatic oscillation over the North Pacific and North America,” *Geophysical Research Letters*, vol. 24, no. 6, pp. 683–686, 1997.
- [89] B. G. Nathaniel and G. Q. Robert, “A Historical Perspective of U.S. Climate Divisions,” *Bulletin of the American Meteorological Society*, vol. 77, no. 2, p. 293, 1996. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26232706&site=eds-live>
- [90] National Oceanic and Atmospheric Administration, ESRL Physical Science Division, and CIRES, “North Atlantic Oscillation Jones Index,” <https://www.esrl.noaa.gov/psd/data/correlation/jonesnao.data>, Aug. 2019.
- [91] —, “Northern Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/noi.data>, Aug. 2019.
- [92] —, “North Atlantic SST Index,” https://www.esrl.noaa.gov/psd/data/correlation/NTA_ersst.data, Aug. 2019.
- [93] —, “Northeast Brazil Rainfall Anomaly Index,” <https://www.esrl.noaa.gov/psd/data/correlation/brazilrain.data>, Aug. 2019.

- [94] —, “North Tropical Atlantic Index,” https://www.esrl.noaa.gov/psd/data/correlation/NTA_ersst.data, Aug. 2019.
- [95] —, “Oceanic Nino Index,” <https://www.esrl.noaa.gov/psd/data/correlation/oni.data>, Aug. 2019.
- [96] —, “Pacific Warmpool Index,” <https://www.esrl.noaa.gov/psd/data/correlation/pacwarm.data>, Aug. 2019.
- [97] —, “Pacific North American Index,” <https://www.esrl.noaa.gov/psd/data/correlation/pna.data>, Aug. 2019.
- [98] —, “Sahel Rainfall Index,” <https://www.esrl.noaa.gov/psd/data/correlation/sahelrain.data>, Aug. 2019.
- [99] —, “Southern Annular Mode Index,” https://www.esrl.noaa.gov/psd/data/20thC_Rean/timeseries/monthly/SAM/sam.20crv2c.long.data, Aug. 2019.
- [100] —, “SOI NINO 3.4 Index,” <https://www.esrl.noaa.gov/psd/data/correlation/censo.data>, Aug. 2019.
- [101] —, “Solar Flux Index,” <https://www.esrl.noaa.gov/psd/data/correlation/solar.data>, Aug. 2019.
- [102] —, “Sunspot Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/sunspot.long.data, Aug. 2019.
- [103] —, “SW Monsoon Region Rainfall Index,” <https://www.esrl.noaa.gov/psd/data/correlation/swmonsoon.data>, Aug. 2019.

- [104] —, “Tropical Northern Atlantic Index,” <https://www.esrl.noaa.gov/psd/data/correlation/tna.data>, Aug. 2019.
- [105] —, “Trans-Nino Index,” <https://www.esrl.noaa.gov/psd/data/correlation/tni.data>, Aug. 2019.
- [106] —, “Tripole Index for the Interdecadal Pacific Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/timeseries/IPOTPI/ipotpi.hadisst2.data>, Aug. 2019.
- [107] —, “Tropical Pacific SST EOF Index,” <https://www.esrl.noaa.gov/psd/data/correlation/eofpac.data>, Aug. 2019.
- [108] —, “Tropical Southern Atlantic Index,” <https://www.esrl.noaa.gov/psd/data/correlation/tsa.data>, Aug. 2019.
- [109] —, “Western Hemisphere Warm Pool Index,” <https://www.esrl.noaa.gov/psd/data/correlation/whwp.data>, Aug. 2019.
- [110] —, “Western Pacific Index,” <https://www.esrl.noaa.gov/psd/data/correlation/wp.data>, Aug. 2019.
- [111] National Weather Service and Climate Prediction Center, “Darwin Sea Level Pressure Index,” <https://www.cpc.ncep.noaa.gov/data/indices/darwin>, Aug. 2019.
- [112] —, “Equatorial Eastern Pacific SLP Index,” https://www.cpc.ncep.noaa.gov/data/indices/repac_slpa.for, Aug. 2019.
- [113] —, “Equatorial SOI Indonesia SLP Index,” https://www.cpc.ncep.noaa.gov/data/indices/rindo_slpa.for, Aug. 2019.

- [114] —, “Equatorial SOI Index,” <https://www.cpc.ncep.noaa.gov/data/indices/reqsoi.for>, Aug. 2019.
- [115] —, “Global Tropics SST Index,” <https://www.cpc.ncep.noaa.gov/data/indices/sstoi.atl.indices>, Aug. 2019.
- [116] —, “Indonesia SLP Index,” https://www.cpc.ncep.noaa.gov/data/indices/rindo_slpa.for, Aug. 2019.
- [117] —, “Outgoing Long Wave Radiation Central Pacific Index,” <https://www.cpc.ncep.noaa.gov/data/indices/cpolr.mth.81-10.ascii>, Aug. 2019.
- [118] —, “Outgoing Long Wave Radiation Equator Index,” <https://www.cpc.ncep.noaa.gov/data/indices/olr>, Aug. 2019.
- [119] —, “Quasi-Biennial Oscillation 30mb zonal wind Index,” <https://www.cpc.ncep.noaa.gov/data/indices/qbo.u30.index>, Aug. 2019.
- [120] —, “Quasi-Biennial Oscillation 50mb zonal wind Index,” <https://www.cpc.ncep.noaa.gov/data/indices/qbo.u50.index>, Aug. 2019.
- [121] —, “South Atlantic SST Index,” <https://www.cpc.ncep.noaa.gov/data/indices/sstoi.atl.indices>, Aug. 2019.
- [122] —, “Tahiti Sea Level Pressure Index,” <https://www.cpc.ncep.noaa.gov/data/indices/tahiti>, Aug. 2019.
- [123] —, “Trade Winds Index 135E 180W 5N 5S West Pacific Index,” <https://www.cpc.ncep.noaa.gov/data/indices/wpac850>, Aug. 2019.

- [124] —, “Trade Winds Index 135W 120W 5N 5S East Pacific Index,” <https://www.cpc.ncep.noaa.gov/data/indices/epac850>, Aug. 2019.
- [125] —, “Trade Winds Index 175W 140W 5N 5S Central Pacific Index,” <https://www.cpc.ncep.noaa.gov/data/indices/cpac850>, Aug. 2019.
- [126] —, “Zonal Winds Equator 200mb Index,” <https://www.cpc.ncep.noaa.gov/data/indices/zwnd200>, Aug. 2019.
- [127] —, “Zonally Average 500mb Temperature Anomalies Index,” <https://www.cpc.ncep.noaa.gov/data/indices/z500t>, Aug. 2019.
- [128] National Wildlife Federation, “Global Warming is Affecting Weather,” <http://www.nwf.org/Wildlife/Threats-to-Wildlife/Global-Warming/Global-Warming-is-Causing-Extreme-Weather.aspx>, Aug. 2012. [Online]. Available: <http://www.nwf.org/Wildlife/Threats-to-Wildlife/Global-Warming/Global-Warming-is-Causing-Extreme-Weather.aspx>
- [129] NCAR and UCAR, “North Pacific Index,” <https://climatedataguide.ucar.edu/climate-data/north-pacific-np-index-trenberth-and-hurrell-monthly-and-winter>, Aug. 2019.
- [130] NCDC and NOAA, “Reconstructed North Atlantic Oscillation Index,” ftp://ftp.ncdc.noaa.gov/pub/data/paleo/historical/north_atlantic/nao_mon.txt, Aug. 2019.
- [131] NCEI GIS Agile Team, “Observational data map of hourly weather station networks from around the world,” <https://www.nci.noaa.gov/data/coop-hourly-precipitation/v2/archive/>, Aug. 2019.

- [132] J. A. Nelder and R. W. Wedderburn, “Generalized linear models,” *Journal of the Royal Statistical Society: Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972.
- [133] M. Newman, G. P. Compo, and M. A. Alexander, “ENSO-forced variability of the Pacific decadal oscillation,” *Journal of Climate*, vol. 16, no. 23, pp. 3853–3857, 2003.
- [134] M. Newman, M. A. Alexander, T. R. Ault, K. M. Cobb, C. Deser, E. Di Lorenzo, N. J. Mantua, A. J. Miller, S. Minobe, H. Nakamura *et al.*, “The Pacific decadal oscillation, revisited,” *Journal of Climate*, vol. 29, no. 12, pp. 4399–4427, 2016.
- [135] N. P. Nezlin and E. D. Stein, “Spatial and temporal patterns of remotely-sensed and field-measured rainfall in southern California,” *Remote Sensing of Environment*, vol. 96, no. 2, pp. 228–245, 2005.
- [136] NOAA and ESRL Physical Science Division, “Atlantic Multidecadal Oscillation (AMO),” <https://www.esrl.noaa.gov/psd/data/correlation/amon.us.long.data>, Aug. 2019.
- [137] —, “Pacific Decadal Oscillation (PDO),” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/pdo.long.data, Aug. 2019.
- [138] —, “Southern Oscillation Index (SOI),” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/soi.long.data, Aug. 2019.
- [139] NOAA, ESRL Physical Science Division, and CIRES, “Antarctic Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/aao.data>, Aug. 2019.

- [140] —, “Atlantic Meridional Mode Index,” <https://www.esrl.noaa.gov/psd/data/timeseries/monthly/AMM/ammsst.data>, Aug. 2019.
- [141] —, “Atlantic Multidecadal Oscillation Smoothed Index,” <https://www.esrl.noaa.gov/psd/data/correlation/amon.sm.data>, Aug. 2019.
- [142] —, “Atlantic Multidecadal Oscillation Unsmoothed Index,” <https://www.esrl.noaa.gov/psd/data/correlation/amon.us.data>, Aug. 2019.
- [143] —, “Arctic Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/ao.data>, Aug. 2019.
- [144] —, “Atlantic Tripole SST EOF Index,” <https://www.esrl.noaa.gov/psd/data/correlation/atlttri.data>, Aug. 2019.
- [145] —, “Caribbean Index,” https://www.esrl.noaa.gov/psd/data/correlation/CAR_ersst.data, Aug. 2019.
- [146] —, “Central Indian Precipitation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/indiamon.data>, Aug. 2019.
- [147] —, “Dipole Mod Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/DMI/, Aug. 2019.
- [148] —, “Eastern Asia / Western Russia Index,” <https://www.esrl.noaa.gov/psd/data/correlation/ea.data>, Aug. 2019.
- [149] —, “ENSO Precipitation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/espi.data>, Aug. 2019.

- [150] —, “East Pacific / North Pacific Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/epo.data>, Aug. 2019.
- [151] —, “Global Mean Land Ocean Temperature Index,” <https://www.esrl.noaa.gov/psd/data/correlation/gmsst.data>, Aug. 2019.
- [152] —, “Globally Integrated Angular Momentum Index,” <https://www.esrl.noaa.gov/psd/data/correlation/glaam.data.scaled>, Aug. 2019.
- [153] —, “Atlantic Hurricanes Index,” <https://www.esrl.noaa.gov/psd/data/correlation/hurr.data>, Aug. 2019.
- [154] —, “Multivariate ENSO Index,” <https://www.esrl.noaa.gov/psd/data/correlation/meiv2.data>, Aug. 2019.
- [155] —, “North Atlantic Oscillation Index,” <https://www.esrl.noaa.gov/psd/data/correlation/nao.data>, Aug. 2019.
- [156] —, “Niño 1+2 SST Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino12.long.data, Aug. 2019.
- [157] —, “Niño 3.4 SST Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.data, Aug. 2019.
- [158] —, “Niño 3 SST Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino3.long.data, Aug. 2019.
- [159] —, “Niño 4 SST Index,” https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino4.long.data, Aug. 2019.

- [160] H. Paek, J.-Y. Yu, and C. Qian, “Why were the 2015/2016 and 1997/1998 extreme El Niños different?” *Geophysical Research Letters*, vol. 44, no. 4, pp. 1848–1856, 2017.
- [161] M. Pal, “Random forest classifier for remote sensing classification,” *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.
- [162] M. Pal and P. M. Mather, “An assessment of the effectiveness of decision tree methods for land cover classification,” *Remote Sensing of Environment*, vol. 86, no. 4, pp. 554–565, 2003.
- [163] V. Quinlan, John R., *C4.5 : programs for machine learning.*, ser. The Morgan Kaufmann series in machine learning, 1993. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsbvb&AN=edsbvb.BV007222053&site=eds-live>
- [164] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: <https://www.R-project.org/>
- [165] R. Ranasinghe, R. McLoughlin, A. Short, and G. Symonds, “The Southern Oscillation Index, wave climate, and beach rotation,” *Marine Geology*, vol. 204, no. 3-4, pp. 273–287, 2004.
- [166] N. Rayner, D. E. Parker, E. Horton, C. K. Folland, L. V. Alexander, D. Rowell, E. Kent, and A. Kaplan, “Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century,” *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D14, 2003.

- [167] Reno Gazette Journal, “A look back: 8 worst of the storms in Northern Nevada since the 90s,” <https://www.rgj.com/story/news/2017/12/11/snow-flood-dust-storm-8-worst-storms-northern-nevada/903536001/>, Dec. 2017.
- [168] J. A. Robert, N. Neville, D. J. Phil, and J. B. Ian, “A Further Extension of the Tahiti–Darwin SOI, Early ENSO Events and Darwin Pressure,” *Journal of Climate*, vol. 4, no. 7, p. 743, 1991. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26196423&site=eds-live>
- [169] L. Rodriguez, C. Rakovski, M. Allali, and H. El-Askary, “Long-Term Variability of Gauged Precipitation Over California and Its Links to Circulation Patterns,” in *Patterns and Mechanisms of Climate, Paleoclimate and Paleoenvironmental Changes from Low-Latitude Regions*. Springer, 2019, pp. 95–98.
- [170] C. F. Ropelewski and P. D. Jones, “An extension of the Tahiti–Darwin Southern Oscillation index,” *Monthly Weather Review*, vol. 115, pp. 2161–2165, 1987. [Online]. Available: <http://libproxy.chapman.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=ofm&AN=501908453&site=eds-live>
- [171] N. Schneider and B. D. Cornuelle, “The forcing of the Pacific decadal oscillation,” *Journal of Climate*, vol. 18, no. 21, pp. 4355–4373, 2005.
- [172] S. Shukla, A. Steinemann, S. F. Iacobellis, and D. R. Cayan, “Annual Drought in California: Association with Monthly Precipitation and Climate Phases,” *Journal of*

- Applied Meteorology and Climatology*, vol. 54, no. 11, pp. 2273–2281, 2015. [Online]. Available: <https://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-15-0167.1>
- [173] N. Siler, Y. Kosaka, S.-P. Xie, and X. Li, “Tropical ocean contributions to Californias surprisingly dry El Niño of 2015/16,” *Journal of Climate*, vol. 30, no. 24, pp. 10 067–10 079, 2017.
- [174] J. Simm, I. M. de Abril, and M. Sugiyama, *Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression*, 2014. [Online]. Available: <http://CRAN.R-project.org/package=extraTrees>
- [175] D. Singh, M. Ting, A. A. Scaife, and N. Martin, “California winter precipitation predictability: Insights from the anomalous 2015–2016 and 2016–2017 seasons,” *Geophysical Research Letters*, vol. 45, no. 18, pp. 9972–9980, 2018.
- [176] Z. Song, Q. Shu, Y. Bao, X. Yin, and F. Qiao, “The prediction on the 2015/16 El Niño event from the perspective of FIO-ESM,” *Acta Oceanologica Sinica*, vol. 34, no. 12, pp. 67–71, 2015.
- [177] R. C. Stone, G. L. Hammer, and T. Marcussen, “Prediction of global rainfall probabilities using phases of the Southern Oscillation Index,” *Nature*, vol. 384, no. 6606, pp. 252–255, 1996.
- [178] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [179] K. E. Trenberth and D. J. Shea, “Atlantic hurricanes and natural variability in 2005,” *Geophysical Research Letters*, vol. 33, no. 12, 2006.

- [180] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002. [Online]. Available: <http://www.stats.ox.ac.uk/pub/MASS4>
- [181] R. S. Vose, S. Applequist, M. Squires, I. Durre, M. J. Menne, C. N. W. Jr., C. Fenimore, K. Gleason, and D. Arndt, “Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions,” *Journal of Applied Meteorology and Climatology*, vol. 53, no. 5, pp. 1232–1251, 2014. [Online]. Available: <https://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-13-0248.1>
- [182] R. S. Vose, S. Applequist, M. Squires, I. Durre, M. J. Menne, C. N. Williams, C. Fenimore, K. Gleason, and D. Arndt, “NOAA’s gridded climate divisional dataset (CLIMDIV). NOAA National Climatic Data Center,” <ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/>, Aug. 2019.
- [183] P. J. Ward, M. Kummu, and U. Lall, “Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis,” *Journal of Hydrology*, vol. 539, pp. 358–378, 2016.
- [184] Western Regional Climate Center, “Climate of California,” <https://wrcc.dri.edu/narratives/CALIFORNIA.htm>, Aug. 2019. [Online]. Available: <https://wrcc.dri.edu/narratives/CALIFORNIA.htm>
- [185] K. Wolter and M. S. Timlin, “El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (mei.ext),” *International Journal of Climatology*, vol. 31, no. 7, pp. 1074–1087, 2011.
- [186] D. Woolhiser, T. Keefer, and K. Redmond, “Southern Oscillation effects on daily

- precipitation in the southwestern United States,” *Water Resources Research*, vol. 29, no. 4, pp. 1287–1295, 1993.
- [187] Z. Wu and N. E. Huang, “A study of the characteristics of white noise using the empirical mode decomposition method,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.
- [188] Z. Wu, E. K. Schnieder, Z.-Z. Hu, and L. Cao, *The impact of global warming on ENSO variability in climate records*. Center for Ocean-Land-Atmosphere Studies Calverton, 2001, vol. 110.
- [189] H. Yan, L. Sun, Y. Wang, W. Huang, S. Qiu, and C. Yang, “A record of the Southern Oscillation Index for the past 2,000 years from precipitation proxies,” *Nature Geoscience*, vol. 4, no. 9, pp. 611–614, 2011.
- [190] T. W. Yee, *VGAM: Vector Generalized Linear and Additive Models*, 2019. [Online]. Available: <https://CRAN.R-project.org/package=VGAM>
- [191] —, “The VGAM Package for Categorical Data Analysis,” *Journal of Statistical Software*, vol. 32, no. 10, pp. 1–34, 2010. [Online]. Available: <http://www.jstatsoft.org/v32/i10/>
- [192] —, “Two-parameter reduced-rank vector generalized linear models,” *Computational Statistics and Data Analysis*, 2013. [Online]. Available: <http://ees.elsevier.com/csda>
- [193] —, *Vector generalized linear and additive models: with an implementation in R*. Springer, 2015.

- [194] T. W. Yee and A. F. Hadi, “Row-column interaction models, with an R implementation,” *Computational Statistics*, vol. 29, no. 6, pp. 1427–1445, 2014.
- [195] T. W. Yee and A. G. Stephenson, “Vector generalized linear and additive extreme value models,” *Extremes*, vol. 10, no. 1-2, pp. 1–19, 2007.
- [196] T. W. Yee and C. J. Wild, “Vector Generalized Additive Models,” *Journal of Royal Statistical Society, Series B*, vol. 58, no. 3, pp. 481–493, 1996.
- [197] T. W. Yee, J. Stoklosa, and R. M. Huggins, “The VGAM Package for Capture-Recapture Data Using the Conditional Likelihood,” *Journal of Statistical Software*, vol. 65, no. 5, pp. 1–33, 2015. [Online]. Available: <http://www.jstatsoft.org/v65/i05/>
- [198] A. P. Young, R. E. Flick, T. W. Gallien, S. N. Giddings, R. T. Guza, M. Harvey, L. Lenain, B. C. Ludka, W. K. Melville, and W. O’Reilly, “Southern California coastal response to the 2015–2016 El Niño,” *Journal of Geophysical Research: Earth Surface*, vol. 123, no. 11, pp. 3069–3083, 2018.
- [199] Y. Zang, J. Wallace, and D. Battisti, “Enso-like interdecadal variability: 1900–1993,” *Journal of Climate*, vol. 10, pp. 1004–1020, 1997.
- [200] Y. Zhang, “An observational study of atmosphere-ocean interactions in the northern oceans on interannual and interdecadal time-scales,” Thesis, 1996.
- [201] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

APPENDICES

APPENDIX A

Wavelets

A.1 The Continuous Wavelet Spectrum

A.1.1 Gauge

Figure A.1: Continuous Wavelet power spectrum of gauge residuals in climate division one

Figure A.2: Continuous Wavelet power spectrum of gauge residuals in climate division two

Figure A.3: Continuous Wavelet power spectrum of gauge residuals in climate division three

Figure A.4: Continuous Wavelet power spectrum of gauge residuals in climate division four

Figure A.5: Continuous Wavelet power spectrum of gauge residuals in climate division five

Figure A.6: Continuous Wavelet power spectrum of gauge residuals in climate division seven

A.1.2 Modeled

Figure A.7: Continuous Wavelet power spectrum of modeled residuals in climate division one

Figure A.8: Continuous Wavelet power spectrum of modeled residuals in climate division two

Figure A.9: Continuous Wavelet power spectrum of modeled residuals in climate division three

Figure A.10: Continuous Wavelet power spectrum of modeled residuals in climate division four

Figure A.11: Continuous Wavelet power spectrum of modeled residuals in climate division five

Figure A.12: Continuous Wavelet power spectrum of modeled residuals in climate division seven

A.2 Cross-Wavelet Transform

A.2.1 Modeled VS Gauge

Figure A.13: Cross-Wavelet Transform of modeled and gauge residuals in climate division one

Figure A.14: Cross-Wavelet Transform of modeled and gauge residuals in climate division two

Figure A.15: Cross-Wavelet Transform of modeled and gauge residuals in climate division three

Figure A.16: Cross-Wavelet Transform of modeled and gauge residuals in climate division four

Figure A.17: Cross-Wavelet Transform of modeled and gauge residuals in climate division five

Figure A.18: Cross-Wavelet Transform of modeled and gauge residuals in climate division seven

A.2.2 Gauge VS SOI

Figure A.19: Cross-Wavelet Transform of gauge residuals and SOI in climate division one

Figure A.20: Cross-Wavelet Transform of gauge residuals and SOI in climate division two

Figure A.21: Cross-Wavelet Transform of gauge residuals and SOI in climate division three

Figure A.22: Cross-Wavelet Transform of gauge residuals and SOI in climate division four

Figure A.23: Cross-Wavelet Transform of gauge residuals and SOI in climate division five

Figure A.24: Cross-Wavelet Transform of gauge residuals and SOI in climate division seven

A.2.3 Gauge VS PDO

Figure A.25: Cross-Wavelet Transform of gauge residuals and PDO in climate division one

Figure A.26: Cross-Wavelet Transform of gauge residuals and PDO in climate division two

Figure A.27: Cross-Wavelet Transform of gauge residuals and PDO in climate division three

Figure A.28: Cross-Wavelet Transform of gauge residuals and SOI in climate division four

Figure A.29: Cross-Wavelet Transform of gauge residuals and PDO in climate division five

Figure A.30: Cross-Wavelet Transform of gauge residuals and PDO in climate division seven

A.2.4 Modeled VS SOI

Figure A.31: Cross-Wavelet Transform of modeled residuals and SOI in climate division one

Figure A.32: Cross-Wavelet Transform of modeled residuals and SOI in climate division two

Figure A.33: Cross-Wavelet Transform of modeled residuals and SOI in climate division three

Figure A.34: Cross-Wavelet Transform of modeled residuals and SOI in climate division four

Figure A.35: Cross-Wavelet Transform of modeled residuals and SOI in climate division five

Figure A.36: Cross-Wavelet Transform of modeled residuals and SOI in climate division seven

A.2.5 Modeled VS PDO

Figure A.37: Cross-Wavelet Transform of modeled residuals and PDO in climate division one

Figure A.38: Cross-Wavelet Transform of modeled residuals and PDO in climate division two

Figure A.39: Cross-Wavelet Transform of modeled residuals and PDO in climate division three

Figure A.40: Cross-Wavelet Transform of modeled residuals and PDO in climate division four

Figure A.41: Cross-Wavelet Transform of modeled residuals and PDO in climate division five

Figure A.42: Cross-Wavelet Transform of modeled residuals and PDO in climate division seven

A.3 Wavelet Coherence

A.3.1 Modeled VS Gauge

Figure A.43: Wavelet Coherence of modeled and gauge residuals in climate division one

Figure A.44: Wavelet Coherence of modeled and gauge residuals in climate division two

Figure A.45: Wavelet Coherence of modeled and gauge residuals in climate division three

Figure A.46: Wavelet Coherence of modeled and gauge residuals in climate division four

Figure A.47: Wavelet Coherence of modeled and gauge residuals in climate division five

Figure A.48: Wavelet Coherence of modeled and gauge residuals in climate division seven

A.3.2 Gauge VS SOI

Figure A.49: Wavelet-Coherence of gauge residuals and SOI in climate division one

Figure A.50: Wavelet-Coherence of gauge residuals and SOI in climate division two

Figure A.51: Wavelet-Coherence of gauge residuals and SOI in climate division three

Figure A.52: Wavelet-Coherence of gauge residuals and SOI in climate division four

Figure A.53: Wavelet-Coherence of gauge residuals and SOI in climate division five

Figure A.54: Wavelet-Coherence of gauge residuals and SOI in climate division seven

A.3.3 Gauge VS PDO

Figure A.55: Wavelet-Coherence of gauge residuals and PDO in climate division one

Figure A.56: Wavelet-Coherence of gauge residuals and PDO in climate division two

Figure A.57: Wavelet-Coherence of gauge residuals and PDO in climate division three

Figure A.58: Wavelet-Coherence of gauge residuals and PDO in climate division four

Figure A.59: Wavelet-Coherence of gauge residuals and PDO in climate division five

Figure A.60: Wavelet-Coherence of gauge residuals and PDO in climate division seven

A.3.4 Modeled VS SOI

Figure A.61: Wavelet-Coherence of modeled residuals and SOI in climate division one

Figure A.62: Wavelet-Coherence of modeled residuals and SOI in climate division two

Figure A.63: Wavelet-Coherence of modeled residuals and SOI in climate division three

Figure A.64: Wavelet-Coherence of modeled residuals and SOI in climate division four

Figure A.65: Wavelet-Coherence of modeled residuals and SOI in climate division five

Figure A.66: Wavelet-Coherence of modeled residuals and SOI in climate division seven

A.3.5 Modeled VS PDO

Figure A.67: Wavelet-Coherence of modeled residuals and PDO in climate division one

Figure A.68: Wavelet-Coherence of modeled residuals and PDO in climate division two

Figure A.69: Wavelet-Coherence of modeled residuals and PDO in climate division three

Figure A.70: Wavelet-Coherence of modeled residuals and PDO in climate division four

Figure A.71: Wavelet-Coherence of modeled residuals and PDO in climate division five

Figure A.72: Wavelet-Coherence of modeled residuals and PDO in climate division seven

APPENDIX B

ARMA

B.1 Residual Plots

B.1.1 Gauge

Figure B.1: Gauge monthly residual plot for climate division one

Figure B.2: Gauge monthly residual plot for climate division two

Figure B.3: Gauge monthly residual plot for climate division three

Figure B.4: Gauge monthly residual plot for climate division four

Figure B.5: Gauge monthly residual plot for climate division five

Figure B.6: Gauge monthly residual plot for climate division seven

B.1.2 Modeled

Figure B.7: Modeled monthly residual plot for climate division one

Figure B.8: Modeled monthly residual plot for climate division two

Figure B.9: Modeled monthly residual plot for climate division three

Figure B.10: Modeled monthly residual plot for climate division four

Figure B.11: Modeled monthly residual plot for climate division five

Figure B.12: Modeled monthly residual plot for climate division seven

B.2 Extreme Precipitation Events

Table B.1: Excessive precipitation dates in climate division one

Gauge	Modeled
Dec 1952	Jan 1895
Dec 1955	Jan 1896
Feb 1958	Nov 1899
Oct 1962	Feb 1902
Dec 1964	Nov 1903
Jan 1967	Feb 1904
Jan 1969	Mar 1904
Jan 1970	Mar 1907
Nov 1970	Jan 1909
Nov 1973	Nov 1909
Mar 1975	Jan 1911
Feb 1983	Jan 1914
Mar 1983	Feb 1915
Nov 1983	Jan 1916
Dec 1983	Nov 1920
Nov 1984	Dec 1922
Feb 1986	Oct 1924

Continued on next page

Table B.1 – *Continued from previous page*

Gauge	Modeled
Dec 1987	Feb 1925
Nov 1988	Nov 1926
Mar 1989	Feb 1927
May 1990	Dec 1929
Mar 1991	Dec 1931
Jan 1995	Feb 1938
Mar 1995	Mar 1938
Dec 1995	Feb 1940
Dec 1996	Dec 1941
Feb 1998	Dec 1945
Nov 1998	Oct 1947
Feb 1999	Oct 1950
Dec 2002	Dec 1952
Dec 2005	Jan 1954
Jan 2008	Dec 1955
Mar 2011	Jan 1956
Mar 2012	Feb 1958
Dec 2012	Oct 1962
Oct 2016	Dec 1964

Continued on next page

Table B.1 – *Continued from previous page*

Gauge	Modeled
Feb 2019	Jan 1969
	Jan 1970
	Nov 1970
	Nov 1973
	Mar 1975
	Mar 1983
	Nov 1983
	Dec 1983
	Nov 1984
	Feb 1986
	Dec 1987
	Nov 1988
	Mar 1989
	Jan 1995
	Mar 1995
	Dec 1996
	Feb 1998
	Nov 1998
	Feb 1999

Continued on next page

Table B.1 – *Continued from previous page*

Gauge	Modeled
	Dec 2002
	Apr 2003
	Dec 2005
	Jan 2008
	Mar 2011
	Dec 2012
	Dec 2014
	Mar 2016
	Feb 2017
	Feb 2019

Table B.2: Excessive precipitation dates in climate division two

Gauge	Modeled
Dec 1952	Jan 1895
Dec 1955	Jan 1896
Feb 1958	Apr 1896
Feb 1962	Feb 1902

Continued on next page

Table B.2 – *Continued from previous page*

Gauge	Modeled
Oct 1962	Nov 1903
Dec 1964	Feb 1904
Jan 1967	Mar 1904
Jan 1969	Mar 1906
Jan 1970	Mar 1907
Nov 1970	Jan 1909
Dec 1970	Jan 1911
Jan 1973	Jan 1914
Nov 1973	Feb 1915
Feb 1975	Jan 1916
Jan 1978	Feb 1919
Nov 1981	Dec 1922
Nov 1983	Nov 1926
Dec 1983	Feb 1927
Nov 1984	Dec 1931
Feb 1986	Dec 1933
Nov 1988	Apr 1935
Mar 1989	Feb 1938
Mar 1991	Mar 1938

Continued on next page

Table B.2 – *Continued from previous page*

Gauge	Modeled
Jan 1993	Feb 1940
Jan 1995	Dec 1940
Mar 1995	Dec 1941
Dec 1996	Dec 1945
Feb 1998	Oct 1947
Feb 2000	Dec 1952
Dec 2002	Dec 1955
Dec 2003	Jan 1956
Dec 2005	Feb 1958
Mar 2012	Feb 1962
Dec 2012	Oct 1962
Dec 2014	Nov 1963
Mar 2016	Dec 1964
Jan 2017	Nov 1965
Feb 2017	Jan 1967
Mar 2018	Jan 1969
Feb 2019	Jan 1970
	Nov 1970
	Nov 1973

Continued on next page

Table B.2 – *Continued from previous page*

Gauge	Modeled
	Mar 1974
	Jan 1978
	Nov 1981
	Mar 1983
	Nov 1983
	Dec 1983
	Nov 1984
	Feb 1986
	Nov 1988
	Mar 1989
	Mar 1991
	Jan 1995
	Mar 1995
	Dec 1996
	Jan 1997
	Feb 1998
	Feb 2000
	Dec 2001
	Dec 2002

Continued on next page

Table B.2 – *Continued from previous page*

Gauge	Modeled
	Feb 2004
	Dec 2005
	Mar 2011
	Dec 2012
	Dec 2014
	Mar 2016
	Jan 2017
	Feb 2017
	Feb 2019

Table B.3: Excessive precipitation dates in climate division three

Gauge	Modeled
Dec 1948	Jan 1895
Jan 1950	Jan 1896
Apr 1950	Feb 1897
Nov 1950	Oct 1899
Dec 1955	Feb 1904

Continued on next page

Table B.3 – *Continued from previous page*

Gauge	Modeled
Jan 1957	Jan 1906
Feb 1959	Mar 1907
Feb 1962	Jan 1909
Oct 1962	Jan 1911
Dec 1964	Mar 1911
Nov 1965	Jan 1914
Jan 1969	Jan 1916
Jan 1970	Feb 1919
Dec 1971	Feb 1922
Feb 1975	Nov 1926
Nov 1981	Dec 1931
Nov 1983	Apr 1935
Nov 1984	Feb 1936
Feb 1986	Feb 1938
Mar 1991	Mar 1938
Jan 1995	Jan 1940
Mar 1995	Feb 1940
Dec 1996	Dec 1940
Jan 1997	Dec 1941

Continued on next page

Table B.3 – *Continued from previous page*

Gauge	Modeled
Dec 2005	Jan 1943
Jan 2008	Nov 1946
Jan 2017	Nov 1950
Feb 2017	Dec 1955
Mar 2018	Nov 1960
Feb 2019	Feb 1962
	Oct 1962
	Dec 1964
	Nov 1965
	Jan 1969
	Jan 1970
	Nov 1970
	Nov 1973
	Sep 1976
	Nov 1981
	Nov 1983
	Dec 1983
	Nov 1984
	Feb 1986

Continued on next page

Table B.3 – *Continued from previous page*

Gauge	Modeled
	Nov 1988
	Mar 1991
	Jan 1995
	Mar 1995
	Dec 1996
	Jan 1997
	Dec 2005
	Dec 2010
	Dec 2012
	Jan 2017
	Feb 2017
	Apr 2017
	Feb 2019

Table B.4: Excessive precipitation dates in climate division four

Gauge	Modeled
Mar 1949	Jan 1895

Continued on next page

Table B.4 – *Continued from previous page*

Gauge	Modeled
Nov 1950	Jan 1896
Jan 1952	Nov 1900
Dec 1952	Mar 1907
Dec 1955	Jan 1909
Feb 1959	Dec 1909
Sep 1959	Jan 1911
Nov 1960	Mar 1911
Feb 1962	Jan 1914
Nov 1963	Feb 1915
Dec 1964	Jan 1916
Nov 1965	Dec 1921
Jan 1969	Dec 1922
Feb 1969	Nov 1926
Jan 1970	Dec 1931
Nov 1970	Jan 1933
Dec 1970	Feb 1936
Dec 1971	Dec 1941
Oct 1975	Feb 1944
Nov 1983	Dec 1945

Continued on next page

Table B.4 – *Continued from previous page*

Gauge	Modeled
Dec 1983	Nov 1946
Nov 1984	Oct 1947
Feb 1986	Dec 1952
Mar 1991	Dec 1955
Jan 1993	Feb 1959
Jan 1995	Nov 1960
Mar 1995	Feb 1962
Dec 1996	Nov 1963
Jan 1997	Nov 1965
Feb 1998	Jan 1969
Feb 2000	Feb 1969
Dec 2002	Nov 1970
Dec 2003	Dec 1970
Jan 2008	Dec 1971
Dec 2012	Feb 1973
Dec 2014	Oct 1975
	Nov 1983
	Dec 1983
	Nov 1984

Continued on next page

Table B.4 – *Continued from previous page*

Gauge	Modeled
	Mar 1991
	Jan 1995
	Mar 1995
	Jan 1997
	Feb 1998
	Feb 2000
	Dec 2001
	Dec 2002
	Feb 2004
	Jan 2008
	Dec 2012
	Dec 2014
	Jan 2017

Table B.5: Excessive precipitation dates in climate division five

Gauge	Modeled
Nov 1950	Jan 1895

Continued on next page

Table B.5 – *Continued from previous page*

Gauge	Modeled
Dec 1951	Jan 1896
Dec 1952	Mar 1906
Dec 1955	Mar 1907
Mar 1958	Jan 1909
Apr 1958	Dec 1909
Feb 1959	Jan 1911
Feb 1962	Jan 1914
Dec 1964	Jan 1916
Nov 1965	Nov 1926
Apr 1967	Dec 1931
Jan 1969	Jan 1933
Feb 1969	Feb 1936
Jan 1970	Mar 1938
Nov 1970	Dec 1941
Dec 1971	Nov 1944
Oct 1975	Dec 1945
Feb 1980	Nov 1946
Mar 1982	Nov 1950
Nov 1983	Dec 1952

Continued on next page

Table B.5 – *Continued from previous page*

Gauge	Modeled
Nov 1984	Dec 1955
Feb 1986	Feb 1959
Mar 1991	Nov 1960
Jan 1995	Feb 1962
Mar 1995	Nov 1963
Feb 1998	Dec 1964
Apr 2006	Nov 1965
Jan 2008	Apr 1967
Dec 2010	Jan 1969
Mar 2011	Feb 1969
Jan 2017	Jan 1970
Feb 2017	Nov 1970
Mar 2018	Dec 1970
Feb 2019	Dec 1971
	Oct 1975
	Mar 1983
	Nov 1983
	Dec 1983
	Nov 1984

Continued on next page

Table B.5 – *Continued from previous page*

Gauge	Modeled
	Feb 1986
	Mar 1991
	Jan 1995
	Mar 1995
	Dec 1996
	Jan 1997
	Feb 1998
	Feb 2000
	Dec 2001
	Dec 2002
	Jan 2008
	Dec 2010
	Dec 2012
	Dec 2014
	Jan 2017
	Mar 2018

Table B.6: Excessive precipitation dates in climate division seven

Gauge	Modeled
Dec 1951	Jan 1895
Nov 1965	Feb 1901
Dec 1966	Mar 1907
Nov 1967	Oct 1907
Jan 1969	Dec 1909
Nov 1970	Mar 1912
Dec 1971	Jan 1914
Jan 1974	Jan 1916
Feb 1976	Mar 1918
Sep 1976	Dec 1921
May 1977	Feb 1927
Feb 1978	Dec 1927
Mar 1978	Feb 1932
Jan 1980	Jan 1933
Feb 1980	Feb 1936
Mar 1982	Mar 1938
Mar 1983	Dec 1938
Dec 1984	Sep 1939

Continued on next page

Table B.6 – *Continued from previous page*

Gauge	Modeled
Mar 1991	Dec 1941
Feb 1992	Jan 1943
Mar 1992	Dec 1943
Jan 1993	Feb 1944
Jan 1995	Nov 1944
Feb 1996	Dec 1945
Feb 1998	Nov 1946
Feb 2003	Jan 1949
Oct 2004	Dec 1952
Jan 2005	Nov 1960
May 2010	Nov 1965
Dec 2010	Dec 1965
Jan 2016	Dec 1966
Dec 2016	Jan 1969
Jan 2017	Feb 1969
Mar 2018	Nov 1970
	Dec 1971
	Jan 1974
	Feb 1976

Continued on next page

Table B.6 – *Continued from previous page*

Gauge	Modeled
	Sep 1976
	Feb 1980
	Mar 1983
	Aug 1983
	Dec 1983
	Dec 1984
	Mar 1991
	Mar 1992
	Jan 1993
	Feb 1993
	Jan 1995
	Feb 1998
	Feb 2004
	Feb 2005
	Oct 2005
	Jan 2008
	Jan 2010
	Dec 2010
	Jan 2017

B.3 Extreme Drought Events

Table B.7: Drought dates in climate division one

Gauge	Modeled
Mar 1956	Feb 1896
Dec 1956	Jan 1902
Feb 1967	Dec 1903
Feb 1971	Jan 1904
Dec 1978	Dec 1908
Jan 1984	Jan 1920
Jan 1985	Mar 1926
Dec 1989	Nov 1939
Feb 1995	Nov 1940
Nov 1995	Jan 1949
Dec 1997	Dec 1953
Dec 1999	Feb 1967
Jan 2009	Nov 1969
Jan 2011	Dec 1978
Dec 2011	Jan 1984

Continued on next page

Table B.7 – *Continued from previous page*

Gauge	Modeled
Dec 2013	Jan 1991
Jan 2014	Dec 1994
Dec 2017	Feb 1995
	Nov 1995
	Dec 1997
	Dec 1999
	Jan 2007
	Jan 2009
	Jan 2011
	Dec 2011
	Feb 2012
	Jan 2014
	Dec 2017
	Feb 2018

Table B.8: Drought dates in climate division two

Gauge	Modeled
Feb 1953	Feb 1896
Mar 1956	Feb 1899
Feb 1965	Jan 1902
Feb 1967	Jan 1904
Mar 1969	Dec 1905
Feb 1971	Nov 1906
Jan 1984	Dec 1908
Dec 1989	Feb 1912
Jan 1991	Jan 1918
Nov 1992	Jan 1920
Feb 1995	Mar 1926
Feb 1997	Nov 1939
Dec 1999	Feb 1948
Jan 2011	Jan 1949
Dec 2011	Feb 1967
Jan 2015	Jan 1975
Dec 2017	Jan 1991
	Nov 1992

Continued on next page

Table B.8 – *Continued from previous page*

Gauge	Modeled
	Feb 1995
	Nov 1995
	Dec 1997
	Dec 1999
	Jan 2007
	Jan 2011
	Dec 2011
	Feb 2012
	Jan 2014
	Feb 2018

Table B.9: Drought dates in climate division three

Gauge	Modeled
Dec 1961	Feb 1896
Feb 1965	Dec 1903
Feb 1967	Jan 1904
Mar 1969	Dec 1905

Continued on next page

Table B.9 – *Continued from previous page*

Gauge	Modeled
Jan 1984	Dec 1908
Feb 1995	Jan 1918
Feb 1997	Nov 1939
Jan 2011	Jan 1945
	Feb 1967
	Jan 1991
	Feb 1991
	Dec 1994
	Feb 1995
	Nov 1995
	Jan 2011
	Nov 2016
	Feb 2018

Table B.10: Drought dates in climate division four

Gauge	Modeled
Oct 1955	Feb 1896

Continued on next page

Table B.10 – *Continued from previous page*

Gauge	Modeled
Nov 1955	Feb 1899
Nov 1957	Jan 1902
Jan 1958	Jan 1904
Jan 1962	Dec 1905
Feb 1967	Oct 1906
Oct 1968	Nov 1906
Nov 1968	Feb 1907
Dec 1968	Dec 1908
Dec 1972	Nov 1910
Feb 1974	Dec 1910
Nov 1977	Feb 1911
Feb 1982	Feb 1912
Dec 1982	Oct 1913
Dec 1985	Dec 1913
Jan 1986	Nov 1914
Jan 1991	Nov 1915
Feb 1991	Jan 1918
Jan 1992	Nov 1921
Nov 1992	Mar 1926

Continued on next page

Table B.10 – *Continued from previous page*

Gauge	Modeled
Dec 1994	Feb 1935
Feb 1995	Jan 1936
Dec 1997	Nov 1936
Dec 1999	Nov 1937
Nov 2004	Jan 1938
Nov 2005	Dec 1939
Feb 2006	Nov 1940
Jan 2011	Jan 1941
Feb 2016	Oct 1955
Nov 2016	Nov 1955
Dec 2016	Nov 1957
Dec 2018	Jan 1958
	Jan 1962
	Feb 1967
	Nov 1968
	Dec 1968
	Dec 1972
	Jan 1975
	Nov 1977

Continued on next page

Table B.10 – *Continued from previous page*

Gauge	Modeled
	Feb 1982
	Dec 1982
	Jan 1986
	Jan 1991
	Feb 1991
	Jan 1992
	Nov 1992
	Dec 1994
	Feb 1995
	Oct 1997
	Dec 1997
	Jan 1998
	Dec 1999
	Nov 2004
	Feb 2006
	Nov 2016
	Dec 2016
	Feb 2018

Table B.11: Drought dates in climate division five

Gauge	Modeled
Mar 1956	Feb 1896
Feb 1965	Feb 1899
Feb 1967	Dec 1900
Mar 1969	Mar 1901
Jan 1984	Jan 1902
Nov 1992	Dec 1905
Feb 1995	Feb 1906
Dec 1999	Nov 1906
Jan 2003	Dec 1908
Jan 2011	Dec 1910
Feb 2016	Feb 1912
	Oct 1913
	Nov 1915
	Jan 1918
	Nov 1921
	Mar 1923
	Jan 1925
	Jan 1936

Continued on next page

Table B.11 – *Continued from previous page*

Gauge	Modeled
	Nov 1936
	Nov 1937
	Jan 1938
	Dec 1939
	Nov 1940
	Jan 1945
	Oct 1955
	Nov 1955
	Mar 1956
	Jan 1958
	Jan 1962
	Feb 1967
	Nov 1968
	Dec 1968
	Jan 1975
	Oct 1977
	Nov 1977
	Dec 1979
	Feb 1982

Continued on next page

Table B.11 – *Continued from previous page*

Gauge	Modeled
	Jan 1986
	Jan 1991
	Feb 1991
	Nov 1992
	Dec 1994
	Feb 1995
	Dec 1997
	Dec 1999
	Feb 2006
	Jan 2011
	Nov 2016
	Dec 2016
	Feb 2018

Table B.12: Drought dates in climate division seven

Gauge	Modeled
Feb 1967	Feb 1896

Continued on next page

Table B.12 – *Continued from previous page*

Gauge	Modeled
Jan 1972	Dec 1900
Feb 1974	Dec 1904
Nov 1977	Oct 1906
Dec 1979	Feb 1912
Jan 1984	Dec 1913
Jan 1985	Dec 1915
Nov 1992	Nov 1921
Jan 2003	Jan 1927
	Mar 1931
	Jan 1932
	Jan 1938
	Nov 1940
	Jan 1941
	Dec 1942
	Jan 1944
	Feb 1952
	Jan 1958
	Oct 1965
	Dec 1968

Continued on next page

Table B.12 – *Continued from previous page*

Gauge	Modeled
	Jan 1976
	Oct 1977
	Nov 1977
	Dec 1977
	Nov 1979
	Dec 1979
	Oct 1982
	Nov 1991
	Jan 1992
	Sep 1992
	Nov 1992
	Nov 1994
	Dec 1994
	Feb 1995
	Jan 1998
	Dec 2000
	Jan 2003
	Jan 2004
	Sep 2004

Continued on next page

Table B.12 – *Continued from previous page*

Gauge	Modeled
	Nov 2004
	Nov 2009
	Nov 2016

B.4 Statistical Comparison for Modeled and Gauge Datasets

Figure B.13: Histogram of gauge minus modeled precipitation in climate division one

Figure B.14: Histogram of gauge minus modeled precipitation in climate division two

Figure B.15: Histogram of gauge minus modeled precipitation in climate division three

Figure B.16: Histogram of gauge minus modeled precipitation in climate division four

Figure B.17: Histogram of gauge minus modeled precipitation in climate division five

Figure B.18: Histogram of gauge minus modeled precipitation in climate division seven

APPENDIX C

EMD

Figure C.1: Climate division one EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Figure C.2: Climate division two EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Figure C.3: Climate division three EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Figure C.4: Climate division four EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Figure C.5: Climate division five EMD correlation between Modeled/Gauge and SOI/PDO/AMO

Figure C.6: Climate division seven EMD correlation between Modeled/Gauge and SOI/PDO/AMO

APPENDIX D

VGLM

D.1 Main Effect

D.1.1 Confusion Matrix of Main Effect

D.1.1.1 Gauge

Figure D.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division one

(a) Training

(b) Testing

Figure D.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division two

(a) Training

(b) Testing

Figure D.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division three

(a) Training

(b) Testing

Figure D.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division four

(a) Training

(b) Testing

Figure D.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division five

Figure D.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on gauge time-series climate division seven

D.1.1.2 Modeled

Figure D.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division one

Figure D.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division two

(a) Training

(b) Testing

Figure D.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division three

(a) Training

(b) Testing

Figure D.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division four

(a) Training

(b) Testing

Figure D.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division five

(a) Training

(b) Testing

Figure D.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for main effect on modeled time-series climate division seven

D.1.2 F1 Scores of Main Effect

D.1.2.1 Gauge

Table D.1: F1 scores for training and testing for main effect on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.58	Class 0:	0.97	0.58
Class 1:	0.27	0.58	Class 1:	0.27	0.58
Class 2:	0.27	0.60	Class 2:	0.27	0.60

Table D.2: F1 scores for training and testing for main effect on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.62	Class 0:	0.95	0.72
Class 1:	0.23	0.57	Class 1:	0.44	0.73
Class 2:	0.59	0.73	Class 2:	0.50	0.74

Table D.3: F1 scores for training and testing for main effect on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.57	Class 0:	0.95	0.63
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.80	0.83	Class 2:	0.25	0.96

Table D.4: F1 scores for training and testing for main effect on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.61	Class 0:	0.95	0.65
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.94	0.94	Class 2:	0.33	0.72

Table D.5: F1 scores for training and testing for main effect on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.65	Class 0:	0.64	0.57
Class 1:	NA	0.50	Class 1:	NaN	0.48
Class 2:	0.67	0.78	Class 2:	0.13	0.76

Table D.6: F1 scores for training and testing for main effect on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.66	Class 0:	0.93	0.46
Class 1:	0.15	0.54	Class 1:	NaN	0.49
Class 2:	1.00	1.00	Class 2:	NA	NA

D.1.2.2 Modeled

Table D.7: F1 scores for training and testing for main effect on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.54	Class 0:	0.95	0.50
Class 1:	0.21	0.56	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.8: F1 scores for training and testing for main effect on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.54	Class 0:	0.96	0.54
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	0.40	0.66

Table D.9: F1 scores for training and testing for main effect on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.52	Class 0:	0.97	0.58
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.29	0.58	Class 2:	NaN	0.48

Table D.10: F1 scores for training and testing for main effect on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.94	0.55	Class 0:	0.94	0.50
Class 1:	NaN	0.50	Class 1:	NA	0.50
Class 2:	0.24	0.57	Class 2:	NA	0.50

Table D.11: F1 scores for training and testing for main effect on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.62	Class 0:	0.95	0.61
Class 1:	0.31	0.60	Class 1:	0.50	0.67
Class 2:	0.34	0.63	Class 2:	NaN	0.49

Table D.12: F1 scores for training and testing for main effect on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.89	Class 0:	0.96	0.63
Class 1:	0.84	0.90	Class 1:	0.67	0.99
Class 2:	0.80	0.89	Class 2:	NaN	0.47

D.2 Two-Way Interaction

D.2.1 Confusion Matrix of Two-Way Interaction

D.2.1.1 Gauge

Figure D.13: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division one

Figure D.14: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division two

Figure D.15: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division three

(a) Training

(b) Testing

Figure D.16: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division four

(a) Training

(b) Testing

Figure D.17: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division five

Figure D.18: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on gauge time-series climate division seven

D.2.1.2 Modeled

Figure D.19: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division one

Figure D.20: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division two

(a) Training

(b) Testing

Figure D.21: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division three

(a) Training

(b) Testing

Figure D.22: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division four

(a) Training

(b) Testing

Figure D.23: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division five

(a) Training

(b) Testing

Figure D.24: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for two-way interaction on modeled time-series climate division seven

D.2.2 F1 Scores of Two-Way Interaction

D.2.2.1 Gauge

Table D.13: F1 scores for training and testing for two-way interaction on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.64	Class 0:	0.89	0.57
Class 1:	0.32	0.60	Class 1:	0.14	0.54
Class 2:	0.67	0.75	Class 2:	NaN	0.48

Table D.14: F1 scores for training and testing for two-way interaction on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.96	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.15: F1 scores for training and testing for two-way interaction on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.98	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.16: F1 scores for training and testing for two-way interaction on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.51	Class 0:	0.96	0.49
Class 1:	NaN	0.50	Class 1:	NaN	0.50
Class 2:	0.20	0.56	Class 2:	NaN	0.50

Table D.17: F1 scores for training and testing for two-way interaction on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.92	Class 0:	0.36	0.49
Class 1:	0.86	0.89	Class 1:	NaN	0.42
Class 2:	0.97	0.97	Class 2:	0.07	0.70

Table D.18: F1 scores for training and testing for two-way interaction on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.57	Class 0:	0.83	0.55
Class 1:	0.33	0.60	Class 1:	0.10	0.54
Class 2:	NaN	0.50	Class 2:	NaN	0.42

D.2.2.2 Modeled

Table D.19: F1 scores for training and testing for two-way interaction on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.96	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.20: F1 scores for training and testing for two-way interaction on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.62	Class 0:	0.83	0.59
Class 1:	0.13	0.54	Class 1:	0.13	0.53
Class 2:	1.00	1.00	Class 2:	NaN	0.44

Table D.21: F1 scores for training and testing for two-way interaction on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.59	Class 0:	0.47	0.56
Class 1:	0.09	0.52	Class 1:	NaN	0.42
Class 2:	0.73	0.83	Class 2:	0.03	0.74

Table D.22: F1 scores for training and testing for two-way interaction on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	0.99	Class 0:	0.65	0.68
Class 1:	0.97	0.98	Class 1:	NaN	0.43
Class 2:	1.00	1.00	Class 2:	0.17	0.67

Table D.23: F1 scores for training and testing for two-way interaction on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.87	Class 0:	0.84	0.58
Class 1:	0.67	0.78	Class 1:	0.15	0.54
Class 2:	0.95	0.97	Class 2:	0.13	0.68

Table D.24: F1 scores for training and testing for two-way interaction on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.77	Class 0:	0.94	0.78
Class 1:	0.33	0.60	Class 1:	NaN	0.48
Class 2:	0.90	0.95	Class 2:	NaN	0.45

D.3 Three-Way Interaction

D.3.1 Confusion Matrix of Three-Way Interaction

D.3.1.1 Gauge

Figure D.25: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division one

(a) Training

(b) Testing

Figure D.26: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division two

(a) Training

(b) Testing

Figure D.27: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division three

(a) Training

(b) Testing

Figure D.28: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division four

(a) Training

(b) Testing

Figure D.29: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division five

Figure D.30: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on gauge time-series climate division seven

D.3.1.2 Modeled

Figure D.31: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division one

Figure D.32: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division two

(a) Training

(b) Testing

Figure D.33: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division three

(a) Training

(b) Testing

Figure D.34: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division four

(a) Training

(b) Testing

Figure D.35: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division five

(a) Training

(b) Testing

Figure D.36: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for three-way interaction on modeled time-series climate division seven

D.3.2 F1 Scores of Three-Way Interaction

D.3.2.1 Gauge

Table D.25: F1 scores for training and testing for three-way interaction on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.51	Class 0:	0.94	0.53
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.18	0.55	Class 2:	0.08	0.54

Table D.26: F1 scores for training and testing for three-way interaction on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.96	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.27: F1 scores for training and testing for three-way interaction on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.98	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.28: F1 scores for training and testing for three-way interaction on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.51	Class 0:	0.97	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.20	0.56	Class 2:	NA	0.50

Table D.29: F1 scores for training and testing for three-way interaction on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.50	Class 0:	0.97	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NaN	0.50	Class 2:	NaN	0.50

Table D.30: F1 scores for training and testing for three-way interaction on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.51	Class 0:	0.98	0.49
Class 1:	0.08	0.52	Class 1:	NaN	0.49
Class 2:	NaN	0.50	Class 2:	NA	NA

D.3.2.2 Modeled

Table D.31: F1 scores for training and testing for three-way interaction on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.66	Class 0:	0.88	0.49
Class 1:	0.33	0.60	Class 1:	NaN	0.47
Class 2:	0.67	0.80	Class 2:	NaN	0.45

Table D.32: F1 scores for training and testing for three-way interaction on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.50	Class 0:	0.93	0.54
Class 1:	NaN	0.50	Class 1:	0.18	0.56
Class 2:	NA	0.50	Class 2:	NA	0.50

Table D.33: F1 scores for training and testing for three-way interaction on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.52	Class 0:	0.64	0.54
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.25	0.58	Class 2:	0.01	0.49

Table D.34: F1 scores for training and testing for three-way interaction on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.90	Class 0:	0.80	0.58
Class 1:	0.83	0.87	Class 1:	0.13	0.54
Class 2:	0.92	0.93	Class 2:	0.30	0.79

Table D.35: F1 scores for training and testing for three-way interaction on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.78	Class 0:	0.78	0.58
Class 1:	0.42	0.64	Class 1:	NaN	0.46
Class 2:	0.97	0.97	Class 2:	0.06	0.62

Table D.36: F1 scores for training and testing for three-way interaction on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.86	Class 0:	0.78	0.39
Class 1:	0.71	0.77	Class 1:	NaN	0.41
Class 2:	0.91	0.94	Class 2:	NaN	0.43

APPENDIX E

XGBoost

E.1 Confusion Matrix of XGBoost

E.1.1 Gauge

Figure E.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division one

(a) Training

(b) Testing

Figure E.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division two

(a) Training

(b) Testing

Figure E.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division three

(a) Training

(b) Testing

Figure E.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division four

(a) Training

(b) Testing

Figure E.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division five

Figure E.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on gauge time-series climate division seven

E.1.2 Modeled

Figure E.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division one

Figure E.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division two

(a) Training

(b) Testing

Figure E.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division three

(a) Training

(b) Testing

Figure E.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division four

(a) Training

(b) Testing

Figure E.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division five

(a) Training

(b) Testing

Figure E.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for XGBoost on modeled time-series climate division seven

E.2 F1 Scores of XGBoost

E.2.1 Gauge

Table E.1: F1 scores for training and testing for XGBoost on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.55
Class 1:	1.00	1.00	Class 1:	0.27	0.58
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table E.2: F1 scores for training and testing for XGBoost on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.52
Class 1:	1.00	1.00	Class 1:	NaN	0.49
Class 2:	1.00	1.00	Class 2:	0.29	0.58

Table E.3: F1 scores for training and testing for XGBoost on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.54
Class 1:	1.00	1.00	Class 1:	0.14	0.55
Class 2:	1.00	1.00	Class 2:	NaN	0.50

Table E.4: F1 scores for training and testing for XGBoost on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.53
Class 1:	1.00	1.00	Class 1:	0.13	0.54
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table E.5: F1 scores for training and testing for XGBoost on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.53
Class 1:	1.00	1.00	Class 1:	0.18	0.55
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table E.6: F1 scores for training and testing for XGBoost on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NaN	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

E.2.2 Modeled

Table E.7: F1 scores for training and testing for XGBoost on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.52
Class 1:	1.00	1.00	Class 1:	0.13	0.53
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table E.8: F1 scores for training and testing for XGBoost on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.54
Class 1:	1.00	1.00	Class 1:	0.24	0.57
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table E.9: F1 scores for training and testing for XGBoost on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.92	0.56
Class 1:	1.00	1.00	Class 1:	0.32	0.66
Class 2:	1.00	1.00	Class 2:	NaN	0.45

Table E.10: F1 scores for training and testing for XGBoost on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.49
Class 1:	1.00	1.00	Class 1:	NaN	0.50
Class 2:	1.00	1.00	Class 2:	NaN	0.49

Table E.11: F1 scores for training and testing for XGBoost on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.55
Class 1:	1.00	1.00	Class 1:	0.29	0.58
Class 2:	1.00	1.00	Class 2:	NaN	0.50

Table E.12: F1 scores for training and testing for XGBoost on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.56
Class 1:	1.00	1.00	Class 1:	0.40	0.63
Class 2:	1.00	1.00	Class 2:	NA	0.50

E.3 XGBoost Predictors

E.3.1 Gauge

Figure E.13: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate

Figure E.14: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure E.15: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure E.16: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure E.17: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure E.18: XGBoost gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

E.3.2 Modeled

Figure E.19: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure E.20: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure E.21: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure E.22: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure E.23: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure E.24: XGBoost modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

APPENDIX F

Random Forest Classifier

F.1 Confusion Matrix of Random Forest Classifier

F.1.1 Gauge

Figure F.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division one

(a) Training

(b) Testing

Figure F.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division two

(a) Training

(b) Testing

Figure F.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division three

(a) Training

(b) Testing

Figure F.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division four

(a) Training

(b) Testing

Figure F.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division five

Figure F.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on gauge time-series climate division seven

F.1.2 Modeled

Figure F.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division one

Figure F.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division two

(a) Training

(b) Testing

Figure F.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division three

(a) Training

(b) Testing

Figure F.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division four

(a) Training

(b) Testing

Figure F.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division five

(a) Training

(b) Testing

Figure F.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Random Forest on modeled time-series climate division seven

F.2 F1 Scores of Random Forest Classifier

F.2.1 Gauge

Table F.1: F1 scores for training and testing for Random Forest on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.93	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.2: F1 scores for training and testing for Random Forest on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.3: F1 scores for training and testing for Random Forest on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.4: F1 scores for training and testing for Random Forest on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.5: F1 scores for training and testing for Random Forest on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.6: F1 scores for training and testing for Random Forest on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	NA

F.2.2 Modeled

Table F.7: F1 scores for training and testing for Random Forest on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	0.97	Class 0:	0.93	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	0.89	0.90	Class 2:	NA	0.50

Table F.8: F1 scores for training and testing for Random Forest on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.9: F1 scores for training and testing for Random Forest on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	NA

Table F.10: F1 scores for training and testing for Random Forest on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.93	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.11: F1 scores for training and testing for Random Forest on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table F.12: F1 scores for training and testing for Random Forest on modeled time-series climate division seven

	(a) Training			(b) Testing	
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

F.3 Random Forest Classifier Predictors

F.3.1 Gauge

Figure F.13: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate

Figure F.14: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure F.15: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure F.16: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure F.17: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure F.18: Random Forest gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

F.3.2 Modeled

Figure F.19: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure F.20: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure F.21: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure F.22: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure F.23: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure F.24: Random Forest modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

APPENDIX G

Extra Trees

G.1 Confusion Matrix of Extra Trees

G.1.1 Gauge

Figure G.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division one

(a) Training

(b) Testing

Figure G.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division two

(a) Training

(b) Testing

Figure G.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division three

(a) Training

(b) Testing

Figure G.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division four

(a) Training

(b) Testing

Figure G.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division five

Figure G.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on gauge time-series climate division seven

G.1.2 Modeled

Figure G.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division one

Figure G.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division two

(a) Training

(b) Testing

Figure G.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division three

(a) Training

(b) Testing

Figure G.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division four

Figure G.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division five

Figure G.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Extra Trees on modeled time-series climate division seven

G.2 F1 Scores of Extra Trees

G.2.1 Gauge

Table G.1: F1 scores for training and testing for Extra Trees on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.2: F1 scores for training and testing for Extra Trees on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.3: F1 scores for training and testing for Extra Trees on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.4: F1 scores for training and testing for Extra Trees on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NaN	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.5: F1 scores for training and testing for Extra Trees on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.6: F1 scores for training and testing for Extra Trees on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

G.2.2 Modeled

Table G.7: F1 scores for training and testing for Extra Trees on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.8: F1 scores for training and testing for Extra Trees on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.9: F1 scores for training and testing for Extra Trees on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.10: F1 scores for training and testing for Extra Trees on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.11: F1 scores for training and testing for Extra Trees on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

Table G.12: F1 scores for training and testing for Extra Trees on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	NA	0.50
Class 2:	1.00	1.00	Class 2:	NA	0.50

APPENDIX H

Neural Networks

H.1 Confusion Matrix of Neural Networks

H.1.1 Gauge

Figure H.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division one

Figure H.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division two

Figure H.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division three

(a) Training

(b) Testing

Figure H.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division four

(a) Training

(b) Testing

Figure H.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division five

Figure H.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on gauge time-series climate division seven

H.1.2 Modeled

Figure H.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division one

Figure H.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division two

(a) Training

(b) Testing

Figure H.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division three

(a) Training

(b) Testing

Figure H.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division four

(a) Training

(b) Testing

Figure H.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division five

(a) Training

(b) Testing

Figure H.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Neural Network on modeled time-series climate division seven

H.2 F1 Scores of Neural Networks

H.2.1 Gauge

Table H.1: F1 scores for training and testing for Neural Network on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	NA	0.50	Class 0:	NA	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.06	0.50	Class 2:	0.08	0.50

Table H.2: F1 scores for training and testing for Neural Network on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.82	0.38	Class 0:	0.81	0.38
Class 1:	NaN	0.50	Class 1:	NA	0.50
Class 2:	NaN	0.38	Class 2:	NaN	0.38

Table H.3: F1 scores for training and testing for Neural Network on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.4: F1 scores for training and testing for Neural Network on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.95	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.5: F1 scores for training and testing for Neural Network on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.50	Class 0:	0.97	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.6: F1 scores for training and testing for Neural Network on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.98	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	NA

H.2.2 Modeled

Table H.7: F1 scores for training and testing for Neural Network on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.95	0.50	Class 0:	0.93	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.8: F1 scores for training and testing for Neural Network on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.50	Class 0:	0.95	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.9: F1 scores for training and testing for Neural Network on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	NA	0.50	Class 0:	NA	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	0.06	0.50	Class 2:	NA	NA

Table H.10: F1 scores for training and testing for Neural Network on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.95	0.50	Class 0:	0.93	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.11: F1 scores for training and testing for Neural Network on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.95	0.50	Class 0:	0.95	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

Table H.12: F1 scores for training and testing for Neural Network on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.95	0.50	Class 0:	0.98	0.50
Class 1:	NA	0.50	Class 1:	NA	0.50
Class 2:	NA	0.50	Class 2:	NA	0.50

H.3 Neural Networks Predictors

H.3.1 Gauge

Figure H.13: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on

Figure H.14: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure H.15: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure H.16: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure H.17: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure H.18: Neural Network gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

H.3.2 Modeled

Figure H.19: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure H.20: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure H.21: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure H.22: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure H.23: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure H.24: Neural Network modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

APPENDIX I

Ridge Regression

I.1 Confusion Matrix of Ridge Regression

I.1.1 El Niño

I.1.1.1 Gauge

Figure I.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division one

(a) Training

(b) Testing

Figure I.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division two

(a) Training

(b) Testing

Figure I.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division three

Figure I.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division four

Figure I.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division five

Figure I.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on gauge time-series climate division seven

I.1.1.2 Modeled

Figure I.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division one

Figure I.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division two

Figure I.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division three

Figure I.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division four

(a) Training

(b) Testing

Figure I.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division five

(a) Training

(b) Testing

Figure I.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on El Niño on modeled time-series climate division seven

I.1.2 La Niña

I.1.2.1 Gauge

Figure I.13: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division one

(a) Training

(b) Testing

Figure I.14: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division two

(a) Training

(b) Testing

Figure I.15: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division three

Figure I.16: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division four

Figure I.17: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division five

Figure I.18: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on gauge time-series climate division seven

I.1.2.2 Modeled

Figure I.19: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division one

Figure I.20: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division two

(a) Training

(b) Testing

Figure I.21: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division three

(a) Training

(b) Testing

Figure I.22: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division four

(a) Training

(b) Testing

Figure I.23: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division five

(a) Training

(b) Testing

Figure I.24: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Ridge on La Niña on modeled time-series climate division seven

I.2 F1 Scores of Ridge Regression

I.2.1 El Niño

I.2.1.1 Gauge

Table I.1: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.96	0.54	Class 0:	0.96	0.50
Class 1:	0.13	0.54	Class 1:	0.00	0.50

Table I.2: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.3: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.4: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.5: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.6: F1 scores for training and testing for Ridge on El Niño on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

I.2.1.2 Modeled

Table I.7: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.96	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.8: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.97	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.9: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.10: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.11: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.97	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.12: F1 scores for training and testing for Ridge on El Niño on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

I.2.2 La Niña

I.2.2.1 Gauge

Table I.13: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.14: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.94	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.15: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.16: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.17: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.18: F1 scores for training and testing for Ridge on La Niña on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.50	Class 0:	1.00	NaN
Class 1:	0.00	0.50	Class 1:	NaN	NaN

I.2.2.2 Modeled

Table I.19: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.20: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.99	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.21: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	1.00	NaN
Class 1:	0.00	0.50	Class 1:	NaN	NaN

Table I.22: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.97	0.50	Class 0:	0.96	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.23: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.98	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

Table I.24: F1 scores for training and testing for Ridge on La Niña on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	0.98	0.50	Class 0:	0.99	0.50
Class 1:	0.00	0.50	Class 1:	0.00	0.50

I.3 Ridge Regression Predictors

I.3.1 El Niño

I.3.1.1 Gauge

Figure I.25: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division one

Figure I.26: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure I.27: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure I.28: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure I.29: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure I.30: Ridge on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

I.3.1.2 Modeled

Figure I.31: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure I.32: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure I.33: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure I.34: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure I.35: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure I.36: Ridge on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

I.3.2 La Niña

I.3.2.1 Gauge

Figure I.37: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division one

Figure I.38: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure I.39: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure I.40: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure I.41: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure I.42: Ridge on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

I.3.2.2 Modeled

Figure I.43: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure I.44: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure I.45: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure I.46: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure I.47: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure I.48: Ridge on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

APPENDIX J

Lasso Regression

J.1 Confusion Matrix of Lasso Regression

J.1.1 El Niño

J.1.1.1 Gauge

Figure J.1: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division one

(a) Training

(b) Testing

Figure J.2: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division two

(a) Training

(b) Testing

Figure J.3: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division three

Figure J.4: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division four

Figure J.5: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division five

(a) Training

(b) Testing

Figure J.6: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on gauge time-series climate division seven

J.1.1.2 Modeled

Figure J.7: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division one

Figure J.8: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division two

Figure J.9: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division three

Figure J.10: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division four

(a) Training

(b) Testing

Figure J.11: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division five

(a) Training

(b) Testing

Figure J.12: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on El Niño on modeled time-series climate division seven

J.1.2 La Niña

J.1.2.1 Gauge

Figure J.13: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division one

(a) Training

(b) Testing

Figure J.14: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division two

(a) Training

(b) Testing

Figure J.15: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division three

Figure J.16: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division four

Figure J.17: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division five

(a) Training

(b) Testing

Figure J.18: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on gauge time-series climate division seven

J.1.2.2 Modeled

Figure J.19: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division one

Figure J.20: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division two

Figure J.21: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division three

Figure J.22: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division four

(a) Training

(b) Testing

Figure J.23: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division five

(a) Training

(b) Testing

Figure J.24: Training (1948-1994) and Testing (1995-2019) Confusion Matrix for Lasso on La Niña on modeled time-series climate division seven

J.2 F1 Scores of Lasso Regression

J.2.1 El Niño

J.2.1.1 Gauge

Table J.1: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.93	0.70
Class 1:	1.00	1.00	Class 1:	0.40	0.70

Table J.2: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.73
Class 1:	1.00	1.00	Class 1:	0.40	0.73

Table J.3: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.96
Class 1:	1.00	1.00	Class 1:	0.50	0.96

Table J.4: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.86	0.88
Class 1:	1.00	1.00	Class 1:	0.35	0.88

Table J.5: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.99
Class 1:	1.00	1.00	Class 1:	0.67	0.99

Table J.6: F1 scores for training and testing for Lasso on El Niño on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.47
Class 1:	1.00	1.00	Class 1:	0.00	0.47

J.2.1.2 Modeled

Table J.7: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.73
Class 1:	1.00	1.00	Class 1:	0.50	0.73

Table J.8: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.82
Class 1:	1.00	1.00	Class 1:	0.67	0.82

Table J.9: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.99	0.99
Class 1:	1.00	1.00	Class 1:	0.80	0.99

Table J.10: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.94	0.71
Class 1:	1.00	1.00	Class 1:	0.29	0.71

Table J.11: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.81
Class 1:	1.00	1.00	Class 1:	0.57	0.81

Table J.12: F1 scores for training and testing for Lasso on El Niño on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.46
Class 1:	1.00	1.00	Class 1:	0.00	0.46

J.2.2 La Niña

J.2.2.1 Gauge

Table J.13: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.47
Class 1:	1.00	1.00	Class 1:	0.00	0.47

Table J.14: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.74
Class 1:	1.00	1.00	Class 1:	0.50	0.74

Table J.15: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	1.00	1.00
Class 1:	1.00	1.00	Class 1:	1.00	1.00

Table J.16: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.97	0.50
Class 1:	1.00	1.00	Class 1:	0.00	0.50

Table J.17: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.92	0.45
Class 1:	1.00	1.00	Class 1:	0.00	0.45

Table J.18: F1 scores for training and testing for Lasso on La Niña on gauge time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	NaN
Class 1:	1.00	1.00	Class 1:	0.00	NaN

J.2.2.2 Modeled

Table J.19: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division one

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.95	0.47
Class 1:	1.00	1.00	Class 1:	0.00	0.47

Table J.20: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division two

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.96	0.72
Class 1:	1.00	1.00	Class 1:	0.33	0.72

Table J.21: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division three

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	NaN
Class 1:	1.00	1.00	Class 1:	0.00	NaN

Table J.22: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division four

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.91	0.45
Class 1:	1.00	1.00	Class 1:	0.00	0.45

Table J.23: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division five

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.98	0.50
Class 1:	1.00	1.00	Class 1:	0.00	0.50

Table J.24: F1 scores for training and testing for Lasso on La Niña on modeled time-series climate division seven

(a) Training			(b) Testing		
	F1	Balanced Accuracy		F1	Balanced Accuracy
Class 0:	1.00	1.00	Class 0:	0.86	0.38
Class 1:	1.00	1.00	Class 1:	0.00	0.38

J.3 Lasso Regression Predictors

J.3.1 El Niño

J.3.1.1 Gauge

Figure J.25: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division one

Figure J.26: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure J.27: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure J.28: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure J.29: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure J.30: Lasso on El Niño gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

J.3.1.2 Modeled

Figure J.31: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure J.32: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure J.33: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure J.34: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure J.35: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure J.36: Lasso on El Niño modeled time-series top 20 predictors from 1948 to 2018 on climate division seven

J.3.2 La Niña

J.3.2.1 Gauge

Figure J.37: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division one

Figure J.38: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division two

Figure J.39: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division three

Figure J.40: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division four

Figure J.41: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division five

Figure J.42: Lasso on La Niña gauge time-series top 20 predictors from 1948 to 2018 on climate division seven

J.3.2.2 Modeled

Figure J.43: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division one

Figure J.44: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division two

Figure J.45: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division three

Figure J.46: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division four

Figure J.47: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division five

Figure J.48: Lasso on La Niña modeled time-series top 20 predictors from 1948 to 2018 on climate division seven