CHARACTERIZATION OF NON-EL NIÑO INDUCED DRY CONDITIONS ACROSS THE U.S. AFFILIATED PACIFIC ISLANDS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ATMOSPHERIC SCIENCES

OCTOBER 2017

By

Alejandro Ludert

Dissertation Committee:

Bin Wang, Chairperson
Fei-Fei Jin
Yi-Leng Chen
Pao-Shin Chu
Eric Firing

Keywords: Pacific Islands, drought, ENSO, El Niño, La Niña

To my parents and Erin

ACKNOWLEDGMENTS

This project was funded by cooperative agreement NA17RJ1230 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein do not necessarily reflect the views of NOAA or any of its subdivisions.

I would like to thank Prof. Bin Wang for his support and advise during this project as well as Prof. Mark Merrifield for his support via JIMAR as well as his feedback and contributions to this work. I would also like to thank my committee members, Professors Fei-Fei Jin, Pao-Shin Chu and Yi-Leng Chen. Finally, I would like to extend my appreciation and gratitude to Prof. Eric Firing for agreeing to serve on my committee on short notice.

ABSTRACT

The U.S. Affiliated Pacific Islands (USAPIs), located in the Western Pacific, have limited water resources making them very susceptible to severe drought conditions. The annual cycle and ENSO response of rainfall differs between USAPIs north of 7°N and those to the south. Southern stations show a canonical negative correlation between dry season (December to May) rainfall and ENSO. Northern stations, on the other hand, show little correlation with ENSO if the three super El Niños are excluded. Instead, they exhibit two distinct rainfall regimes, the Canonical regime, and a Non-Canonical regime, in which the dry season rainfall is positively correlated with ENSO. Non-canonical years pose an important forecasting challenge. Cool Dry events are of particular interest because they have coincided with several emergency and disaster-level droughts across the Northern USAPIs. Composite analysis of the Canonical and Non-Canonical regimes show stark differences between dry season atmospheric and SST patterns. Compared to Canonical composites, the Non-Canonical composites show clear and previously undescribed anomaly patterns during the dry season. In Cool Dry events, circulation anomalies over the Western Pacific are anticyclonic, with a band of anomalous dry conditions extending from the central Pacific towards Micronesia that causes unexpected droughts in the Northern USAPIs. Canonical Cool Wet events, on the other hand, show cyclonic West Pacific circulation anomalies and a La Niña like horseshoe rainfall pattern over the Pacific Basin. Non-canonical Cool events also show SST anomalies narrowly constrained near the dateline, while Canonical Cool events show their largest anomaly magnitude east of the dateline. Both Non-Canonical and Canonical Cool events show negative rainfall and Western Pacific anticyclonic anomalies before the onset of the Dec-May dry season. In Non-Canonical events, these anomalies persist throughout the dry season, while for Canonical events they shift, rapidly becoming positive rainfall and cyclonic circulation anomalies during the dry season. SST anomalies also evolve differently, with Non-Canonical Cool events showing anomalies that extend eastward from the central Pacific rather than intensifying in place over the eastern Pacific. The features are similar and opposite for Canonical and Non-Canonical Warm events. Differences in the evolution of anomalies suggest that the physical mechanisms governing the Non-Canonical and Canonical ENSO regimes are distinct. These differences have been leveraged to develop a novel 2-tier forecasting methodology that combines logistic and linear regression to forecast the Dec-May dry season Standardized Precipitation Index in the Northern USAPIs. This 2-tier methodology achieves significant improvement in the forecast of Dec-May rainfall anomalies as compared

to a benchmark forecast. This type of improved forecasts will help provide local governments and decision makers with guidance for mitigation and relief during Non-Canonical events.

TABLE OF CONTENTS

Acknowledgments . . . . . **iv**

Abstract . . . . . **v**

List of Tables . . . . . **ix**

List of Figures . . . . . **xi**

1 Introduction **1**

2 Data and Methods **8**

2.1 Data 8

2.1.1 Station Rainfall Data 8

2.1.2 Gridded Data 8

2.1.3 PEAC Impact Reports 9

2.2 Methods 9

2.2.1 Determining Similarities in Station Rainfall Annual Cycle 9

2.2.2 Standardized Precipitation Index 12

2.2.3 Composite Analysis and Significance Testing 15

2.2.4 S-EOF Analysis 15

2.2.5 Physical-Empirical Forecasts and Predictor Selection 15

2.2.6 Linear, Stepwise and Logistic Regression and Cross Validation Techniques 16

3 Observed Features of Seasonal Rainfall Variability in the USAPIs **20**

3.1 Different USAPI Rainfall Regimes as Described Using K-Means 20

3.2 Impacts of ENSO North vs. South USAPI Stations 24

3.3 December to May Dry Season Composites 27

3.4 Contrasting Seasonal Evolution of SST and Atmospheric Anomalies Between the Canonical and Non-Canonical Responses 35

3.5 Discussion 48

4 Forecasting 52

4.1 Simple Linear Regression Model 53

4.2 2-tier Logistic Regression Forecasting Model 60

4.2.1 Logistic Regression Forecast for Canonical vs. Non-Canonical Years . 61

4.2.2 Linear Prediction Equations for the SPI During Canonical and Non-Canonical Years 72

4.2.3 Test Period Forecast 82

4.2.4 Forecast Discussion 85

5 Conclusions 88

A Complete Predictor Correlation Tables 91

B Supplemental Figures 94

Bibliography 95

LIST OF TABLES

3.1 Classification of all events selected for our composite analysis according to CPC operational methodology, Kug et al. (Kug et al. (2009)) and Murphy et al. (Murphy et al. (2014)). WP and Mix represent the terms Warm Pool and Mixed El Nino respectively. 26

4.1 Description of all predictors developed for the 1-tier Northern Station Dec-May SPI forecast model. Includes predictor name, variable used, region and a short description of the physical mechanism it represents. 57

4.2 Correlation table of predictors and target Northern Station Dec-May SPI for the 1-tier P-E model. Values in bold represent correlations significant at the 95% test level for a 58 year sample size. 60

4.3 Description of all predictors developed for the logistic regression forecast in the 2-tier forecasting model. 67

4.4 Correlation table for predictors developed for the logistic regression forecast in the 2-tier forecasting model. All predictors with the exception of the Mid Latitude SST Tendency index are correlated to the Canonical/Non-Canonical Classification time series with significance at the 90% confidence level using a Student t-test. 68

4.5 Contingency tables for the Train (1950-2007) and Test (2008-2017) periods for our Logistic Regression forecast of Canonical vs. Non-Canonical events. . 70

4.6 Description of all predictors developed for the Canonical linear regression model in the 2-tier forecasting model. 74

4.7 Correlation table for predictors developed for the Canonical linear regression forecast in the 2-tier forecasting model. All predictors are correlated to the Canonical SPI time series with significance at the 90% confidence level using a Student t-test. 76

4.8 Description of all predictors developed for the Non-Canonical linear regression model in the 2-tier forecasting model. 79

4.9 Correlation table for predictors developed for the Non-Canonical linear regression forecast in the 2-tier forecasting model. All predictors with the exception of the Western Pacific Stream Function Index are correlated to the Non-Canonical SPI time series with significance at the 90% confidence level using a Student t-test. 81

4.10 Contingency tables for the Test (2008-2017) periods for our Logistic Regression forecast of Canonical vs. Non-Canonical events. 84

A.1 Correlation Table for all predictors considered for the benchmark 1 tier model 93

LIST OF FIGURES

1.1 Map showing the US territories and independent nations that compose the U.S. Affiliated Pacific Islands. Marked in red are the stations for which the PEAC Center produces seasonal rainfall forecasts. 1

1.2 Time series of Hiedki Skill Score for each model in the PEAC Center operational forecasting suite (color lines), the consensus of the suite (dashed black line) and the PEAC Center final public forecast (solid black line). Color bars represent the Oceanic Niño Index colored in red for positive values above 0.5C and blue for negative values below -0.5C 5

1.3 Reproduction of a section of a USAID document from September 30th, 2013, detailing the drought relief efforts in the Northern USAPIs and the Republic of the Marshall Islands. 7

2.1 Changes in the probability density function and cumulative probability function of the gamma distribution when the shape k and scale θ parameters are changed. Example plots taken from Wikipedia (2017). 13

2.2 Schematic figure of the calculation of the Standardized Precipitation Index. . 14

2.3 Example of a linear (blue) and logistic (red) regression as fit to a classification dataset (black dots) with 1 predictor variable. Taken from scikit-learn developers (2017d). 18

3.1 Mean Davies-Bouldin index values for K-Means clustering of USAPI climatological annual cycle of monthly rainfall as a function of number of clusters. Clustering and calculation of the Davies-Bouldin Index was repeated 200 times per value of k. Error bars represent the standard deviation of all Davies-Boulding Index values obtained. 21

3.2 Map showing the USAPI stations and their assigned K-Means cluster (color) using k=2, Monsoon Index (contours), and silhouette score for most relevant stations (number next to station name). 22

3.3 Average annual cycle of monthly rainfall for the northern and southern cluster as calculated using K-Means with 2 clusters. 23

3.4 Map showing the USAPI stations and their assigned K-Means cluster (color) using k=3, Monsoon Index (contours), and silhouette score for most relevant stations (number next to station name). 23

3.5 Average annual cycle of monthly rainfall for the northern and southern cluster as calculated using K-Means with 3 clusters. 24

3.6 Scatterplots of the dry season SPI (Dec to May) averaged for the northern (left) and southern USAPI stations (right) vs. the Oceanic Niño Index for the corresponding December-January-February season as published by NOAA CPC. Color markers represent years chosen for composite analysis. Diamond markers represent years for which rainfall anomalies in the Northern USAPIs are of opposite sign to what would be expected from the ONI conditions. We will refer to these as Non-Canonical years. Circular markers, in contrast, represent years for witch the rainfall anomalies are as expected considering the corresponding ONI value. We will refer to these as Canonical years. Vertical dashed lines represent the 0.5°C and -0.5°C ONI values. 25

3.7 Rainfall (shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in rainfall anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a Wilcoxon rank-sum test (Wilks (2005)). 28

3.8 stream function of the anomalous wind(shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry, and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a Wilcoxon rank-sum test (Wilks (2005)). . . . 30

3.9 SST anomaly (shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in SST anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a 2 tailed Student t-test (Wilks (2005)). 32

3.10 December-May SST anomaly (shading top), Precipitation anomaly (shading mid) and stream function (shading bottom), along with 850 hPa geopotential height (contours) and wind composites wet and dry La Niña events. Hatching represents areas where the difference in anomalies between wet and dry composites are significant at the 90% test level. 33

3.11 Time series of the Oceanic Niño Index for each year selected for composite analysis. ONI index as published by the Climate Prediction Center. Bars represent the average of all events in each category. Bars are colored red where the average is greater than 0.5°C and blue where it is less than -0.5°C. The time series are labeled using the year 0. Red arrows depict the extend of the climatological Dec-May dry season. 36

3.12 Time series of the northern USAPI (Kwajalein and Guam) Standardized Precipitation Index (SPI) averaged for the members of each composite categories (black lines). The SPI in this figure is calculated for consecutive 6 month seasons. Bars represent the composite time series of the average stream function of anomalous winds over the Tropical Western Pacific. Red arrows depict the extent of the climatological Dec-May dry season. 37

3.13 Seasonal composites of SST (shading), 850 hPa geopotential height (contours) and wind anomalies for the three 3-month seasons leading up to the December-May dry season, as well as the December-May dry season composite. 38

3.14 Seasonal composites of stream function (shading), 850 hPa geopotential height (contours) and wind anomalies for the three 3-month seasons leading up to the December-May dry season, as well as the December-May dry season composite. 39

3.15 Hovmöller diagrams of composite monthly SST (shading) and GOADS 20 Degree Celsius depth (contours) anomalies averaged from 6°S to 6°N. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season. 41

3.16 Hovmöller diagrams of composite monthly SST (shading) and GOADS 20°Celsius depth (contours) anomalies averaged from 6°S to 6°N for La Niña events and Canonical Cool Wet and Non-Canonical Cool Dry events. Black horizontal lines represent December and May which are the start and end of the northern USAPI dry season. 42

3.17 Hovmöller diagrams of composite monthly rainfall anomalies averaged from 5°N to 15°N (shading) and the stream function of the anomalous winds averaged between the equator and 30°N (contours). Contours range between $+/-3^6$ in intervals of 3^5 m^2/s. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season. . . 43

3.18 Hovmöller diagrams of composite monthly rainfall anomalies averaged from 5°N to 15°N (shading) and the stream function of the anomalous winds averaged between the equator and 30°N (contours) for Wet and Dry La Niña events and Cool Wet and Cool Dry events (see also Figure 3.17). Contours range between $+/-3^6$ in intervals of 3^5 m^2/s. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season. 44

3.19 As in Figure 3.17, but for monthly SST anomalies averaged from 20° and 30°N (shading) and 850 hPa geopotential height anomalies from 30° to 50°N (contours). 45

3.20 Seasonal EOF of rainfall in the region from 5° to 20°N and 140°E to the dateline. Panel A shows the PC time series for the first two S-EOF modes. Triangular markers represent the years chosen for the Non-Canonical composites, with red representing Cool Dry years and blue Warm Wet years. Panels B, C, D and E show the regression of the PC1 to rainfall, geopotential height and wind anomalies during the JJA, SON, DJF and MAM seasons, respectively. Panels F, G, H and I show the regression of PC1 to SST anomalies for the JJA, SON, DJF and MAM seasons. Panels J, K, L and M show the same as B through E but for PC2, while N trough Q are analogous to F through I but also for PC2. Circular hatching represents areas for which correlations are significant at the 90% confidence level. 47

3.21 Schematic diagram of relevant December to May ocean surface temperature and atmospheric circulation anomalies that lead to difference in rainfall conditions over the USAPI region. Rainfall anomalies are represented as contours, with red for dry and green for wet anomalies, SST anomalies are represented as filled shapes and geopotential height anomalies as grey contours, solid for positive and dashed for negative. Winds are represented as black arrows, with equatorial wind anomalies represented in bold black arrows. 49

4.1 Correlation coefficient maps of Dec-May SPI and SST (top) and Stream Function (bottom) of the anomalous wind fields (PSI) for the June-July-August (JJA) and September-October-November (SON) seasons. Hatching represents significance at the 95% confidence level using a Student t-test with a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 53

4.2 Correlation coefficient maps of Dec-May SPI and 850 hPa zonal wind (top) and geopotential height (bottom) for the June-July-August (JJA) and September-October-November (SON) seasons. Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 55

4.3 Correlation coefficient maps of Dec-May SPI and SST, 850 hPa stream function of anomalous winds, zonal winds and geopotential height tendency fields. Tendency fields are defined as the October-November mean anomaly minus August-September mean anomaly. Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 58

4.4 Predicted vs observed Northern USAPI Dec-May SPI form our 1-tier P-E model. The left panel shows the predicted v observed SPI for the model training period 1950-2007 while the right panel shows the same for the independent test period 2008-2017. 61

4.5 JJA (-1) anomaly composites for Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet Stream Function (top) and SST (bottom). The top panels are the same as in figure 3.14 and the bottom are the same as figure 3.13. The boxes show the regions where area averages are taken to develop the "SST and Stream Function Index". Below each panel the expected sign of the area average for each type of events is shown. At the bottom of the figure, the expected sign of the complete multiplicative "SST and Stream Function Index" in shown (more details in text). 63

4.6 Spatial patterns of the first two seasonal EOF modes (same as Figure 3.20) of rainfall in the region from 5°N to 20°N and 140°E to the dateline. The blue and red boxes represent the regions where area average of SST anomalies are taken to develop the "Tropical to Mid Latitude SST gradient Index". 64

4.7 Correlation fields between the Canonical/Non-Canonical "classification time series" and SST anomalies (top) and absolute values of SST anomalies (bottom) for the JJA and SON seasons leading up to the Dec-May dry season. Boxes show the regions where the "Mid latitude SST index" (black box in top right panel) and the "Absolute Value East Pacific SST Index" (blue box in bottom right panel). Hatching shows regions where correlation is significant at the 90% confidence level using a Student t-test. 65

4.8 Dec-May SPI and DJF ONI for the 1950-2007 training period. Colours plot represent the Canonical (green) and Non-Canonical (purple) classification. On the left panel, correct classification based on the observed Dec-May SPI and DJF ONI. on the right panel, colors represent the logistic regression prediction of Canonical or Non-Canonical for each individual year. 71

4.9 Correlation coefficient maps of Dec-May SPI for years in the Canonical training set and SON SST (top left), SON 850 hPa Geopotential Height (top right) and, 850 hPa zonal winds (lower left) anomalies, as well as ON-AS 850 hPa zonal wind anomaly tendency (lower right). Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 41 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 73

4.10 Correlation coefficient maps of Dec-May SPI for years in the Non-Canonical training set and SON SST (top left), SON 850 hPa Geopotential Height (top right) and, 850 hPa zonal winds (lower left) anomalies, as well as ON-AS SST anomaly tendency (lower right). Hatching represents significance at the 90% confidence level using a Student t-test for a sample size of 17 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 77

4.11 Predicted vs. Observed SPI for the Canonical training set (left) and the Non-Canonical training set (right). The blue line represents the 1 to 1 line while the red line represents the linear regression of observed SPI onto predicted SPI values. The correlation values are presented above and are significant at the 95% confidence level using a Student t-test in both cases (41 year sample size for Canonical years and 17 for Non-Canonical years) 82

4.12 As in Figure 4.8. Dec-May SPI and DJF ONI for the 2008-2017 test period. Colours plot represent the Canonical (green) and Non-Canonical (purple) classification. On the left panel, correct classification based on the observed Dec-May SPI and DJF ONI. on the right panel, colors represent the logistic regression prediction of Canonical or Non-Canonical for each individual year. 83

4.13 Scatterplot f the predicted vs observed Standardized Precipitation Index for our 1-tier benchmark model (left), our 2-tier P-E model (center) and the same 2-tier model with perfect logistic prediction of Canonical and Non-Canonical events (right). The blue lines represent the 1:1 line while the red line represent the achieved regression line between prediction ans observations. Correlation values between predicted and observed SPI values presented at the top of each panel. 85

B.1 Correlation coefficient maps of DJF Oceanic Niño Index for years in the Canonical training set and SST (top), 850mb Geopotential Height (middle). The bottom panel represent the correlation between Dec-May SPI for years in the Canonical training set and 850mb Geopotential Height. Hashing represents significance at the 95% test level for a sample size of 17 years. Boxes represent regions chosen for predictor index average calculations (more details in text). 94

CHAPTER 1
INTRODUCTION

The U.S.-Affiliated Pacific Islands (USAPIs) are located in the Western Pacific basin and are composed of U.S territories and independent nations in free association with the United States. They are comprised of the U.S. territories of Guam, The Commonwealth of the Northern Mariana Islands, and American Samoa as well as the Independent Nations of The Republic of Palau, The Republic of the Marshall Islands and the Federated States of Micronesia (see 1.1).

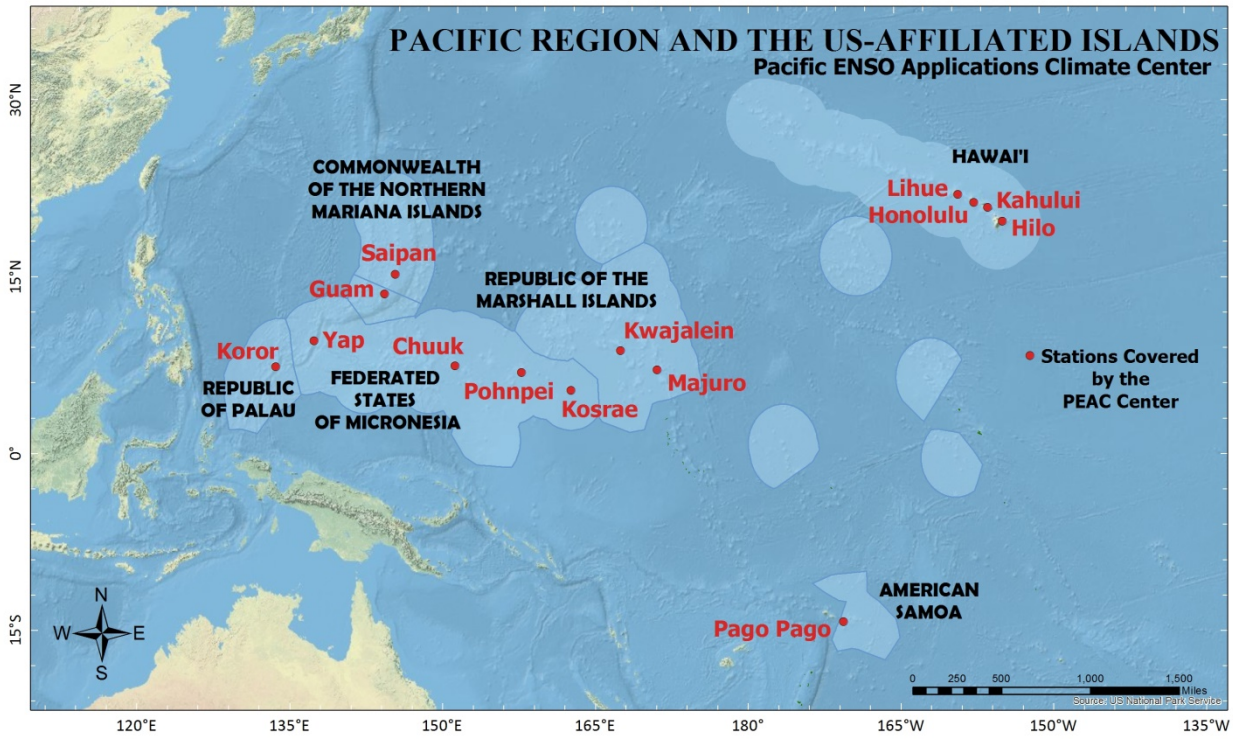


Figure 1.1: Map showing the US territories and independent nations that compose the U.S. Affiliated Pacific Islands. Marked in red are the stations for which the PEAC Center produces seasonal rainfall forecasts.

Islands in the region are a mix of low lying atolls and islands with higher elevation. In general, their small size, remote location and limited financial and natural resources make them very vulnerable to the impacts of climate variability (Shea and Coauthors (2001)). The Pacific ENSO Applications Climate Center (PEAC Center) was founded in 1994 as

multi-institutional partnership to provide governments and stakeholders throughout the USAPI communities with resources to support management and mitigation in climate sensitive sectors. PEAC Center activities include the production of tailored technical climate information and seasonal forecasts, the collection of impacts of climate and weather events in the region, research into Pacific climate variability, and outreach activities to governments and interested parties in the USAPI. A comprehensive account of PEAC Center activities and procedures can be found in Schroeder et al. (2012).

Most of the Northern USAPIs are encompassed within or lie very near the border of the monsoon region as defined by Ramage (1971, 2013). Their annual cycle of rainfall and temperature is in large part modulated by the northward and southward migration of the surface monsoon trough and the Intertropical Convergence Zone (Sadler (1978), BOM-CSIRO (2011)). Low latitude stations such as Koror, Yap, Chuuk, Pohnpei and Majuro tend to have their rainfall maxima during July/August, while Guam and Kwajalein show annual maxima during September/October (Yu et al. (1997a)).

The El Niño-Southern Oscillation (ENSO) is likely the largest source of climate variability in many regions across the globe (McPhaden et al. (2006), Collins et al. (2011)). In the USAPI, located in the Tropical Western Pacific, this is especially true (Bjerknes (1966), Bjerknes (1969), McPhaden et al. (2006)).

In the USAPI, ENSO is largely responsible for interannual sea level variability (Chowdhury et al. (2007)), with El Niño producing lower than average sea levels and La Niña higher than average anomalies. These sea level changes produce significant impacts. Low sea levels are associated with coral bleaching events, as well as damage to fisheries and the food security they provide. High sea levels produce coastal erosion, damage infrastructure and compromise water sources and food crops.

During the extreme phases of ENSO, El Niño and La Niña, climate variability in the USAPIs in the northern hemisphere is directly related to the shift in position of the Intertropical Convergence Zone (ITCZ) (Collins et al. (2011)). As El Niño conditions develop (spring, summer and fall of the developing year), ITCZ convection begins to shift eastward from the Pacific Warm Pool towards the dateline, producing wetter than normal conditions in the USAPIs during fall of the El Niño developing year. During the peak and decaying phase of El Niño events, enhanced convection is now located east of the dateline and is suppressed to the west. This leads to dry conditions in the USAPI region during the peak and decaying phase of El Niño, as were observed during early 1998 and 2016 (Ropelewski and Halpert (1987)). The intensity of El Niño-induced dry conditions across the USAPIs is proportional

the magnitude of the event, with the 1982/1983, 1997/1998 and 2015/2016 events producing intense and widespread droughts throughout the region (Yu et al. (1997b)).

The evolution of rainfall anomalies as El Niño events develop (mentioned above) is closely linked to the evolution of circulation anomalies over the Tropical Western Pacific. Circulation anomalies over the Tropical Western Pacific during El Niño events are observed to switch from cyclonic (enhanced rainfall), as the event develops, to anticyclonic (suppressed rainfall) during the peak and decaying phase of the event. This change in circulation anomalies has been proposed to be a result of coupled ocean-atmosphere dynamics over the Western Pacific (Wang and Zhang (2002)), as well as a result of the interaction between the ENSO cycle and the annual cycle resulting in the ENSO Combination Mode (Stuecker et al. (2013), Stuecker et al. (2015)). While the explanation of why the Western Pacific circulation changes sign from the developing to the decaying phase of El Niño events is still a subject for debate (Li et al. (2016), Stuecker et al. (2016)), the change in sign of circulation and rainfall anomalies is a robust feature.

Tropical cyclone behavior in the USAPIs is also strongly modulated by ENSO (Lander (1994)). Studies relating the total number of cyclones in the Western Pacific to ENSO have not yielded consistent results (Camargo and Sobel (2005)). Nonetheless, the influence of ENSO on the genesis location of TCs in the region is well documented (Chia and Ropelewski (2002), Wang and Chan (2002)). During El Niño years the genesis region of tropical cyclones in the Western Pacific shifts eastward and southwards away from the Maritime Continent. This shift means that the USAPIs tend to have more tropical cyclone activity, including landfalls and associated damages during El Niño years. During La Niña years, TC genesis tends to concentrate over the westernmost part of the basin, generally reducing tropical cyclone impact risk in the USAPIs.

Current understanding of ENSO has posed that El Niño events can display different "flavors", with distinct characteristics in SST anomaly patterns and local and global impacts (Kao and Yu (2009); Kug et al. (2009)). Eastern Pacific or Cold Tongue El Niño events are characterized by having their strongest SST anomalies over the Eastern Pacific, while Central Pacific or "Modoki" El Niño events show SST anomalies confined to the Central Pacific near the dateline. This difference in SST anomaly patterns leads to differences in atmospheric circulation anomalies, and hence results in differences in the regional and global impacts of these two types of El Niño events. While the existence of differences between types of ENSO events is well acknowledged, there is still significant debate as to whether these types represent distinct modes of variability or whether ENSO should be understood as a diverse

continuum (Capotondi et al. (2015)).

The oscillatory nature of ENSO is, in part due to the movement of warm water in the upper ocean layer (above the thermocline) from the equatorial region towards the mid latitude and back. This phenomenon is know and the discharge-recharge oscillator (Jin (1997)). The equatorial thermocline is shallower in the eastern Pacific than in the western part of the basin. For this reason changes in thermocline depth associated with the discharge-recharge process have a stronger effect on SST anomalies in the Eastern Pacific. Another important process in the formation of SST anomalies during ENSO events is the advection of SST gradients by the anomalous zonal currents (Jin and An (1999)). This zonal advective feedback is more effective at changing SST anomalies in the Western Pacific, where zonal SST gradients are strong in the region where the Pacific Cold Tongue and Warm Pool meet. Subsequently, the relative importance of these two dynamical processes during Eastern and Central Pacific El Niño changes, with thermocline anomalies and recharge-discharge dynamics being more prominent in Eastern Pacific El Niño events and zonal advective feedback more prominent in Central Pacific events (Kug et al. (2010), Capotondi (2013)). Observations also support the notion of zonal advection being of higher relative importance in the heat budget of SST anomalies in the Central Pacific but also suggest the importance of air-sea heat fluxes in this region (Kao and Yu (2009), Kug et al. (2009), Kug et al. (2010), Yu et al. (2010)).

Murphy et al. (2014) provide a detailed account of the impacts of different types of El Niño events on dry season rainfall across many Pacific Island stations, including most of the USAPI stations. They found that Cold Tongue El Niño events produced consistent dry conditions across all of the USAPI stations. Mixed El Niño events were found to produce dry conditions in the western USAPIs and Southern Marshall Islands only. Finally, Warm Pool El Niño events were found to produce dry conditions only in the westernmost USAPI station of Koror.

While the concept of different types of El Niño events is understood, the same is less true for La Niña events, for which differences in SST patterns do not allow for such a clear separation into types Kug et al. (2009))). This is because, in general, La Niña events tend to have peak SST anomalies farther west than El Niño events. This results in less inter event differences in SST patterns. As a result, the research into ENSO flavors has been concentrated on warm events exclusively. In fact, Murphy et al. (2014) study all la Niña events together and find that these produce wet conditions for all the USAPI stations.

On longer time scales, substantial climate variability is observed in the Pacfic Basin (Power and Smith (2007)). The dominant mode of decadal variability in the basin is the

ENSO-like natural mode called the Pacific Decadal Oscillation (PDO) (Mantua et al. (1997). Because PDO patterns are similar to ENSO, this decadal mode of variability has similar influence on the USAPIs as ENSO but at a lower frequency.

The PEAC Center issues seasonal rainfall forecasts for the USAPI stations on a monthly basis using a consensus methodology on a suite of 8 statistical and dynamical forecast models and has kept records of forecast performance since 2007 (more details in Schroeder et al. (2012)). Time series of the Heidke Skill Score for each model in the PEAC Suite, as well as the Final Consensus Forecast are shown in Figure 1.2. During the 2009/2010 and 2015/2016 El Niño events and their associated droughts, PEAC Center seasonal rainfall forecasts achieved skills well above that during ENSO neutral conditions. This is consistent with what is generally observed for dynamical models (Wang et al. (2009)) and statistical predictions (Chowdhury et al. (2007)).

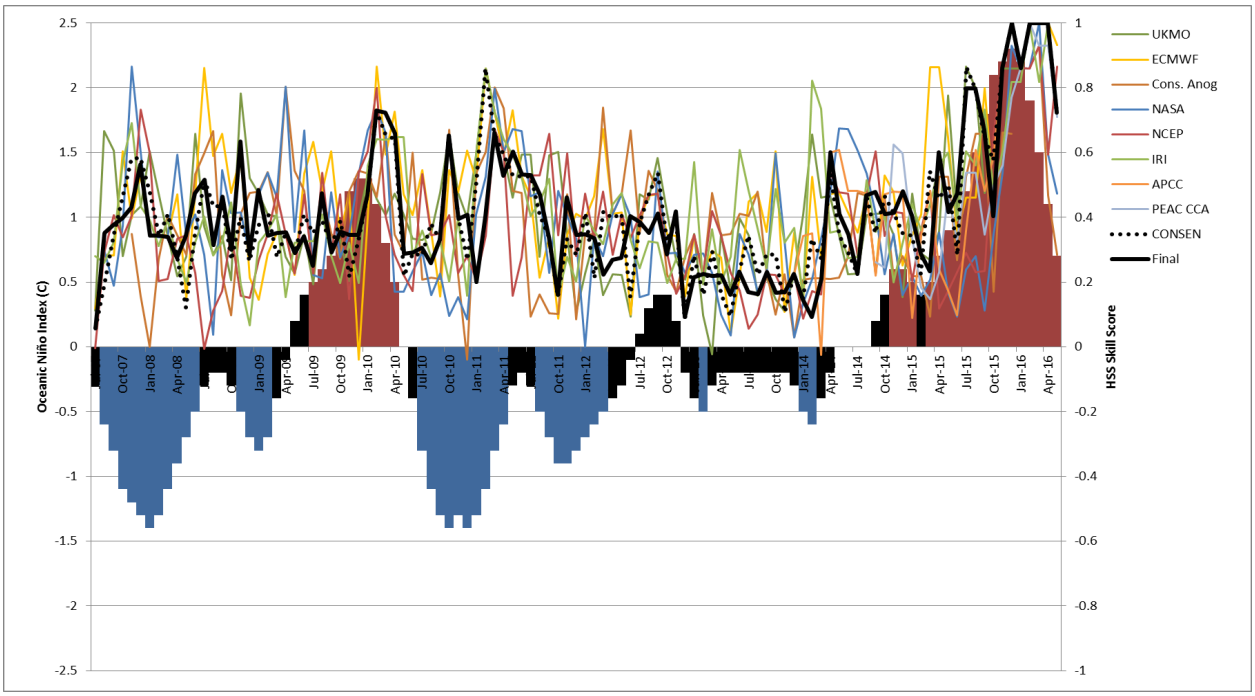
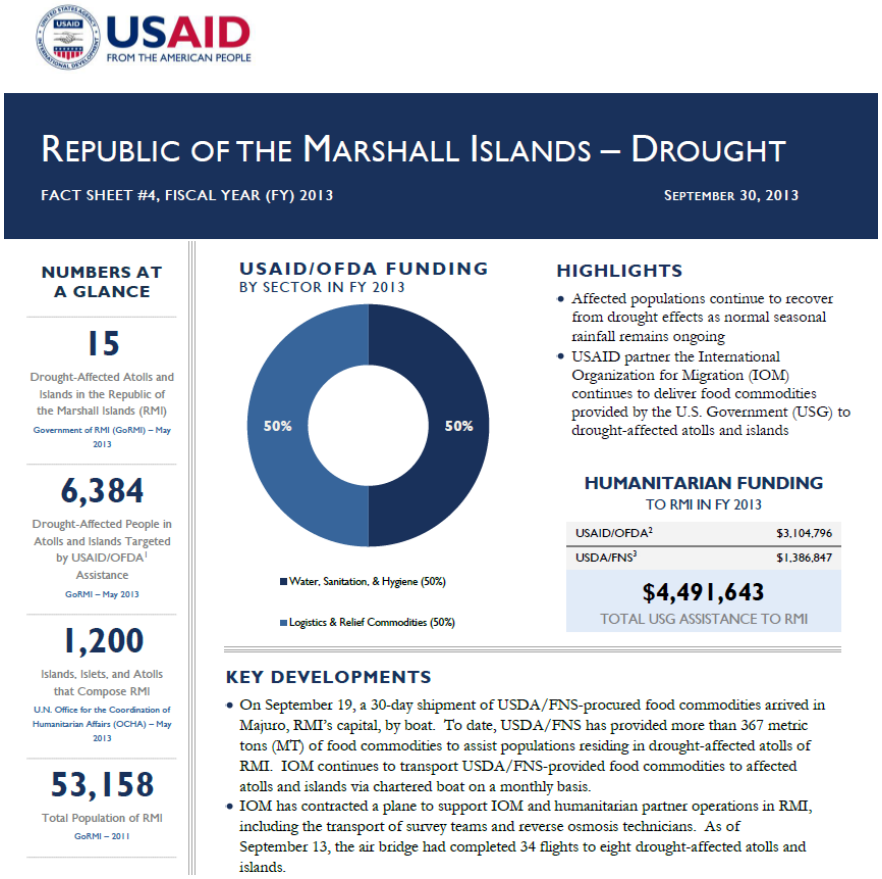


Figure 1.2: Time series of Hiedki Skill Score for each model in the PEAC Center operational forecasting suite (color lines), the consensus of the suite (dashed black line) and the PEAC Center final public forecast (solid black line). Color bars represent the Oceanic Niño Index colored in red for positive values above 0.5C and blue for negative values below -0.5C

Nonetheless, severe droughts can occur in the USAPIs during near neutral ENSO conditions. During the early part of 2013, dryer than normal conditions began affecting many of the USAPIs and were particularly persistent in the atolls and islands in the northern part of

the Republic of the Marshall Islands (RMI). In the Northern RMI, small island communities with little water storage capabilities are especially dependent on rainfall for drinking water and agriculture. Dry conditions over the northeastern USAPIs started during January 2013 and extended west. By April, most of the USAPI stations, with the exception of Pohnpei and Kosrae, were located on the dry side of a strong zonal rainfall gradient along 7°N, with dry conditions to the north and wet to the south. As the year progressed, wet conditions along the equator gave way to strong dry anomalies that extended northwestward and covered the entire Western North Pacific USAPIs by August 2013. The most severe drought impacts during early 2013 were when abnormally dry conditions coincided with the climatological dry season. From January to May 2013, Kwajalein, used here as a proxy for the Northern RMI, received less than 15 inches of rainfall per season. PEAC Center forecasters have determined that, as a rule of thumb, for many of the more water-precarious islands across the USAPIs, 4 inches of rain a month is the minimum needed to meet basic water needs, with 8 inches a month being enough to meet most water needs.

The extended lack of sufficient rainfall observed during early 2013 produced severe water shortages across the northern parts of the RMI which required action from both local government and the international community. Climate impact reports collected by the PEAC Center during early March, 2013 show that low rainfall during the previous month had prompted the shipment of reverse osmosis units to the atolls of Mejit and Lae which were already experiencing water shortages. The RMI government began to hold meetings on drought assessment and mitigation during March, 2013. On April 19, 2013 the government of the RMI declared a state of emergency due to drought conditions. As dry conditions continued and drinking water availability deteriorated, the state of emergency declared by the RMI government was upgraded to a state of disaster by early May, 2013, due to the imminent threat to life, according to RMI authorities. Relief efforts on the part of the United States began on May 3. Initial aid funding was approved by the U.S. embassy in Majuro for relief efforts through USAID (see Figure 1.3) and The International Organization for Migration in order to distribute pre-positioned relief materials to drought-affected islands and atolls. Further relief funding by the U.S. came after U.S. President Barack Obama, at the request of RMI President Christopher Loeak, issued a declaration of disaster for the RMI under the Compact of Free Association (COFA) treaty between the two governments. By the end of this 2013 drought event, a total of about 4.5 million dollars in relief funds had been invested by the U.S. Government during a drought event that affected an estimated 6,400 people. PEAC Center rainfall forecasts showed low skill throughout this drought period, even below



USAID
FROM THE AMERICAN PEOPLE

REPUBLIC OF THE MARSHALL ISLANDS – DROUGHT

FACT SHEET #4, FISCAL YEAR (FY) 2013 — SEPTEMBER 30, 2013

NUMBERS AT A GLANCE

15
Drought-Affected Atolls and Islands in the Republic of the Marshall Islands (RMI)
Government of RMI (GoRMI) – May 2013

6,384
Drought-Affected People in Atolls and Islands Targeted by USAID/OFDA[1] Assistance
GoRMI – May 2013

1,200
Islands, Islets, and Atolls that Compose RMI
U.N. Office for the Coordination of Humanitarian Affairs (OCHA) – May 2013

53,158
Total Population of RMI
GoRMI – 2011

USAID/OFDA FUNDING
BY SECTOR IN FY 2013

HIGHLIGHTS

- Affected populations continue to recover from drought effects as normal seasonal rainfall remains ongoing
- USAID partner the International Organization for Migration (IOM) continues to deliver food commodities provided by the U.S. Government (USG) to drought-affected atolls and islands

HUMANITARIAN FUNDING
TO RMI IN FY 2013

USAID/OFDA[2]	$3,104,796
USDA/FNS[3]	$1,386,847

$4,491,643
TOTAL USG ASSISTANCE TO RMI

KEY DEVELOPMENTS

- On September 19, a 30-day shipment of USDA/FNS-procured food commodities arrived in Majuro, RMI's capital, by boat. To date, USDA/FNS has provided more than 367 metric tons (MT) of food commodities to assist populations residing in drought-affected atolls of RMI. IOM continues to transport USDA/FNS-provided food commodities to affected atolls and islands via chartered boat on a monthly basis.
- IOM has contracted a plane to support IOM and humanitarian partner operations in RMI, including the transport of survey teams and reverse osmosis technicians. As of September 13, the air bridge had completed 34 flights to eight drought-affected atolls and islands.

Figure 1.3: Reproduction of a section of a USAID document from September 30th, 2013, detailing the drought relief efforts in the Northern USAPIs and the Republic of the Marshall Islands.

the average skill shown by our forecast suite during other neutral periods in our forecast record.

In this dissertation, we will investigate if the 2013 drought in the USAPIs was a unique occurrence, and describe the features and evolution of this type of event. Chapter 2 describes the data and methodology used in selecting and examining the features associated with dry conditions in the USAPI during near neutral ENSO conditions. In Chapter 3, we present the results of our observational analysis of the dry season anomaly patterns and evolution of anomalies that produce dry conditions in the USAPIs during near neutral ENSO conditions. In Chapter 4, we develop and test a novel forecasting framework tailored to the USAPIs and designed to capture and forecast dry conditions in the USAPIs during near neutral ENSO conditions. In Chapter 5, we present final thoughts on our findings.

CHAPTER 2
DATA AND METHODS

2.1 Data

2.1.1 Station Rainfall Data

Precipitation data is recorded by the NOAA National Centers for Environmental Information (NCEI) through their Automated Surface Observing System (ASOS), as well as reported directly to the PEAC Center. PEAC Center rainfall data is consistent with what is archived for these stations by the NOAA National Centers for Environmental Information. Continuous monthly total rainfall data for Koror, Yap, Chuuk, Pohnpei, Guam, Kwajalein and Majuro is available from 1966 to present. For USAPI stations for which the PEAC Center does not archive rainfall data, monthly rainfall totals from the Cooperative Observer Network (COOP) stations were used. COOP station data are available for many USAPI stations, including many outer islands and atolls, and covers areas of interest such as the Northern RMI atolls of Enewetak and Utirik for which most of the severe impacts of the 2013 drought were observed. Unfortunately, data for many of these COOP stations in the USAPI suffers from large data gaps, including no data at many stations during 2013. A total of 35 COOP stations across the USAPI were used along with the 7 PEAC stations. COOP data was downloaded from the Western Regional Climate Center (http://www.wrcc.dri.edu/).

2.1.2 Gridded Data

Oceanic Niño Index values made public by the NOAA Climate Prediction Center (CPC) and based on the ERSST v4 (Boyin et al. (2015)) were used in this work as well as the Gridded ERSST v.4 data (Boyin et al. (2015)). NCEP/NCAR Reanalysis 1 (Kalnay et al. (1996)) was used for pressure level data of geopotential height and winds. GPCP V2.2 (Adler et al. (2003)) was used for large scale analysis of rainfall. ERSST, NCEP/NCAR reanalysis and GPCP data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

2.1.3 PEAC Impact Reports

The PEAC Center also collects detailed accounts of climate impacts from our contributors in the USAPIs. These recorded impacts are published in the PEAC Center Quarterly Pacific ENSO Update newsletter. PEAC Center newsletters extend back to 2005 in digital form (available to the public at www.weather.gov/peac/update), and back to 1996 in physical copies archived in our offices.

2.2 Methods

2.2.1 Determining Similarities in Station Rainfall Annual Cycle

Using the monthly data available for all COOP and PEAC stations in the USAPI, a climatological annual cycle of monthly rainfall was calculated for each station. The similarity of the annual cycles for different stations was estimated using the Monsoon Index (Wang and Ding (2008)) and K-Means clustering.

The Monsoon index was defined as the annual cycle range calculated as the climatological wet season rainfall (June to November), minus the climatological dry season rainfall (December to May), divided by the climatological total yearly rainfall and multiplied by 100.

The objective of any clustering algorithm is to separate a large number of observations into groups with similar features. The K-Means clustering algorithm separates data points into clusters by minimizing the within-cluster sum-of-square variance, also known as inertia. In our case, the annual cycle of monthly rainfall at each station represents one 12 dimensional observation. A total of $n = 42$ observations are available. These are partitioned into k clusters so that $k <= n$ by minimizing the cost function:

$$arg \min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{x \in \mathbf{S}_i} \|x - \mu_i\|^2, \tag{2.1}$$

where $\mathbf{S} = (S_1, S_2, \ldots, S_k)$ are the sets the data is being partitioned into, and μ_i is the position of the centroid of set S_i.

The calculation was done using the K-Means algorithm from the Python Scikit-Learn package. This K-Means implementation uses a method called Lloyd's algorithm (Lloyd (1982)) in which points in the dataset are first assigned to their closest existing cluster centroid. These cluster centroids are then updated as the mean position of all points assigned to the same cluster. This procedure is repeated until the assignment of points to clusters no

longer changes and thus the centroids no longer change position. This iterative procedure is equivalent to minimizing the cost function in equation 2.1. The number of clusters and their initial position are free parameters of the method that must be decided by the user and are discussed in the following paragraphs.

For an a priori chosen number of clusters to partition the data into, the K-Means algorithm requires an initial position for each one of these clusters. The final solution that the K-Means algorithm arrives at is sensitive to these initial centroid positions. However, this can be overcome by appropriate initial centroid selection, as well as by performing the K-Means partition multiple times with different initial conditions and choosing the partition with lowest inertia. In the most basic implementation of the algorithm, cluster centroids can be initialized as random points within the space that contains the data to be partitioned. A more advanced centroid initialization method used in this work is called K-Means++ (Arthur and Vassilvitskii (2007)). In this method, centroids are initialized to the positions of points in the data using the following steps:

- The first centroid is chosen from the data at random with uniform probability for all data points
- For each data point x, calculate the distance D_x to the nearest already chosen centroid
- Chose a new data point as a centroid using the weighted distribution so that the probability of a point being chosen is proportional to the distance D_x^2
- Repeat the last two steps until the desired number of k initial centroids have been chosen
- Carry out K-Means clustering

The K-Means++ method of centroid initialization performs well compared to more primitive methods. When combined with repetitive clustering, K-Means++ provides a good solution to avoid suboptimal clustering results.

The number of clusters k used in the K-Means clustering algorithm is also a free parameter. Usually, the number of clusters is chosen based on intuition of the expected number of clusters that the data should be able to be separated into. If no intuition is available, some methods exist for the evaluation of the final clustering which can lead to a selection of an appropriate number of clusters for the data. In our case, the Davies-Bouldin index (Davies and Bouldin (1979)) was used as a metric for clustering quality. The Davies-Bouldin Index

is defined as follows. Let C_i be a cluster of points in our data with centroid A_i and size T_i along with $X_1, \ldots, X_{T_i}$ all the data points assigned to that cluster. A measure of the scatter of the points assigned to that cluster can be defined as:

$$S_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i|^p \right)^{1/p}. \tag{2.2}$$

If p is chosen equal to 2, this scatter (Equation 2.2) is defined based on the Euclidean distance. Values higher than 2 can be used for cases where an Euclidean distance may not be appropriate. In our case however, the Euclidean distance is suitable as a measure, so p is chosen equal to 2. A measure of the distance between clusters C_i and C_j can be similarly defined as:

$$M_{i,j} = ||A_i - A_j||_p. \tag{2.3}$$

For two clusters C_i and C_j, a measure of how good the clustering is, $R_{i,j}$, can be defined as the sum of the scatters of each cluster divided by the distance between the cluster centers, such that:

$$R_{i,j} \equiv \frac{S_i + S_j}{M_{i,j}}. \tag{2.4}$$

The similarity of cluster C_i to its nearest neighbor, D_i, can be measured as:

$$D_i \equiv \underset{j \neq i}{max} R_{i,j}. \tag{2.5}$$

Finally, for a total number of clusters k, the Davies-Bouldin index is defined as:

$$DB \equiv \frac{1}{k} \sum_{i=1}^{k} D_i. \tag{2.6}$$

Defined in this way, the Davies-Bouldin index will be smaller for better clustering solutions. This index is defined only for clustering with $k \geq 2$. Other methods of choosing the optimal number of clusters exist, such as the elbow and silhouette methods (Ketchen and Shook (1996), Llet et al. (2004)) and information criteria methods such as the Akaike, Bayesian and deviance information criteria (Goutte et al. (2001)), among others. The Davies-Bouldin index method used in this work provides a simple, yet quantitative method of determining the optimal number of clusters in our data. The information criteria measure, in contrast, is

more complex, while the elbow and silhouette methods rely on visual interpretation of plots.

The optimal number of clusters selected by the user should be chosen considering the trade off between two extremes. The first is achieving the maximum possible rate of data compression, obtained by assigning all data points to a single cluster. This single cluster would retain all the variability of our original data without providing any additional information. On the other extreme every single data point could be assigned to its own cluster achieving maximum classification accuracy and no data compression (Pinto et al. (2015)). For our station data, clustering with k from 2 to 42 is performed and presented here. The final number of clusters k was chosen as being the smallest value that achieved a small the Davies-Bouldin Index.

2.2.2 Standardized Precipitation Index

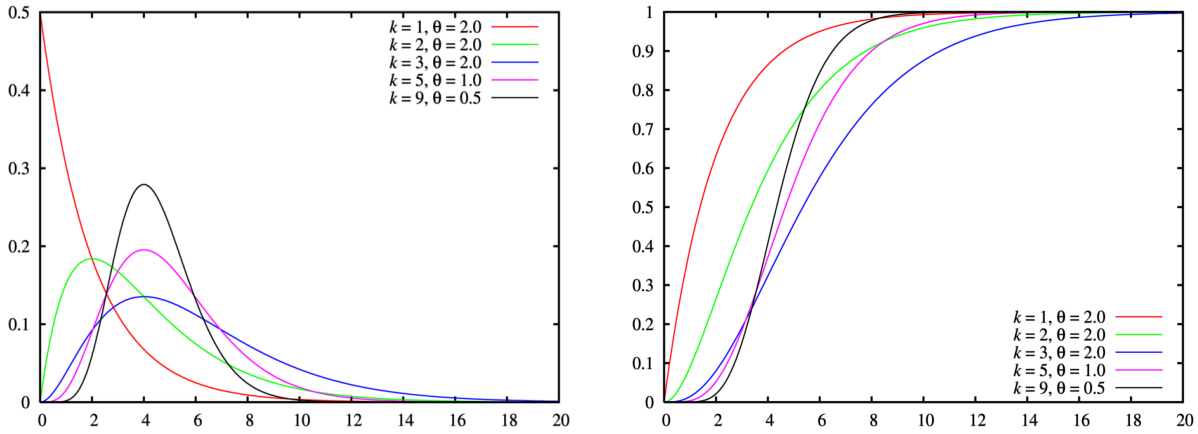
The Standardized Precipitation Index (SPI) (McKee et al. (1993)) is used as primary rainfall index. The SPI is calculated by fitting the observed station climatological rainfall data with a gamma distribution. Using the shape (k) and scale (θ) parametrization of the gamma distribution, the probability density function (pdf) and the cumulative probability functions (cdf) can be written as equations 2.7 and 2.8, respectively (Wilks (2005) uses the inverse scale parameter formulation $\beta = \frac{1}{\theta}$).

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \text{ for } x > 0 \text{ and } k, \theta > 0 \tag{2.7}$$

$$F(x; k, \theta) = \int_0^x f(u; k, \theta) du = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)} \tag{2.8}$$

In equations 2.7 and 2.8, $\Gamma(k)$ and $\gamma\left(k, \frac{x}{\theta}\right)$ represent the complete and incomplete forms of the gamma function. Figure 2.1 shows the changes in the probability density and cumulative probability functions as a result of different combinations of the shape (k) and scale (θ) parameters.

We use maximum likelihood estimation of the k and θ parameters. For any particular rainfall observation, the cumulative probability for that value is calculated based on the gamma distribution fitted to that station's climatological data. Using a Normal distribution for a random variable Z, with mean zero and standard deviation of 1, the value of the variable Z with equal cumulative probability as the rainfall observation is the corresponding SPI value for that rainfall observation. Figure 2.2 shows a schematic diagram of this process.



(a) Probability density function (b) Cumulative probability function

Figure 2.1: Changes in the probability density function and cumulative probability function of the gamma distribution when the shape k and scale θ parameters are changed. Example plots taken from Wikipedia (2017).

This type of transformation is equivalent to a "Z-score" calculation. Thus the SPI can be interpreted as the number of standard deviations away from the climatological mean that a particular observation falls (the SPI has no units).

The SPI is commonly used as an index to characterize drought conditions (Wu et al. (2001), Hayes et al. (1999)) and shows good performance as compared to other drought indices (Guttman (1998), Keyantash and Dracup (2002)) . The SPI can be calculated for cumulative rainfall deficits for time periods of 3 to 24 months. On short time scales it can be used as a measure of soil moisture, while on longer time scales it served as a good indicator of water resource availability (Svoboda et al. (2012)). The SPI is widely used by the operational drought monitoring community as a tool to quantify meteorological drought conditions (Keyantash and Staff (2016)). Values of the SPI are usually linked to descriptive categories of drought conditions. SPI values from -1 to -1.49 are considered to represent "Moderately Dry" conditions, SPI values between -1.5 and -1.99 are considered to represent "Severely Dry" conditions, and SPI values of -2 or less represent "Extremely Dry" conditions (NDMC (2017)).

The SPI calculation relies on precipitation data only, making it less complex to calculate than the Palmer Drought Index (Guttman (1998)) and a good tool for monitoring water resource availability. As a drawback the SPI, by not taking into account temperature and evapotranspiration, is less suited for long term monitoring of soil moisture in climate change applications. By virtue of being a Z-score transformation, the SPI is easy to compare across

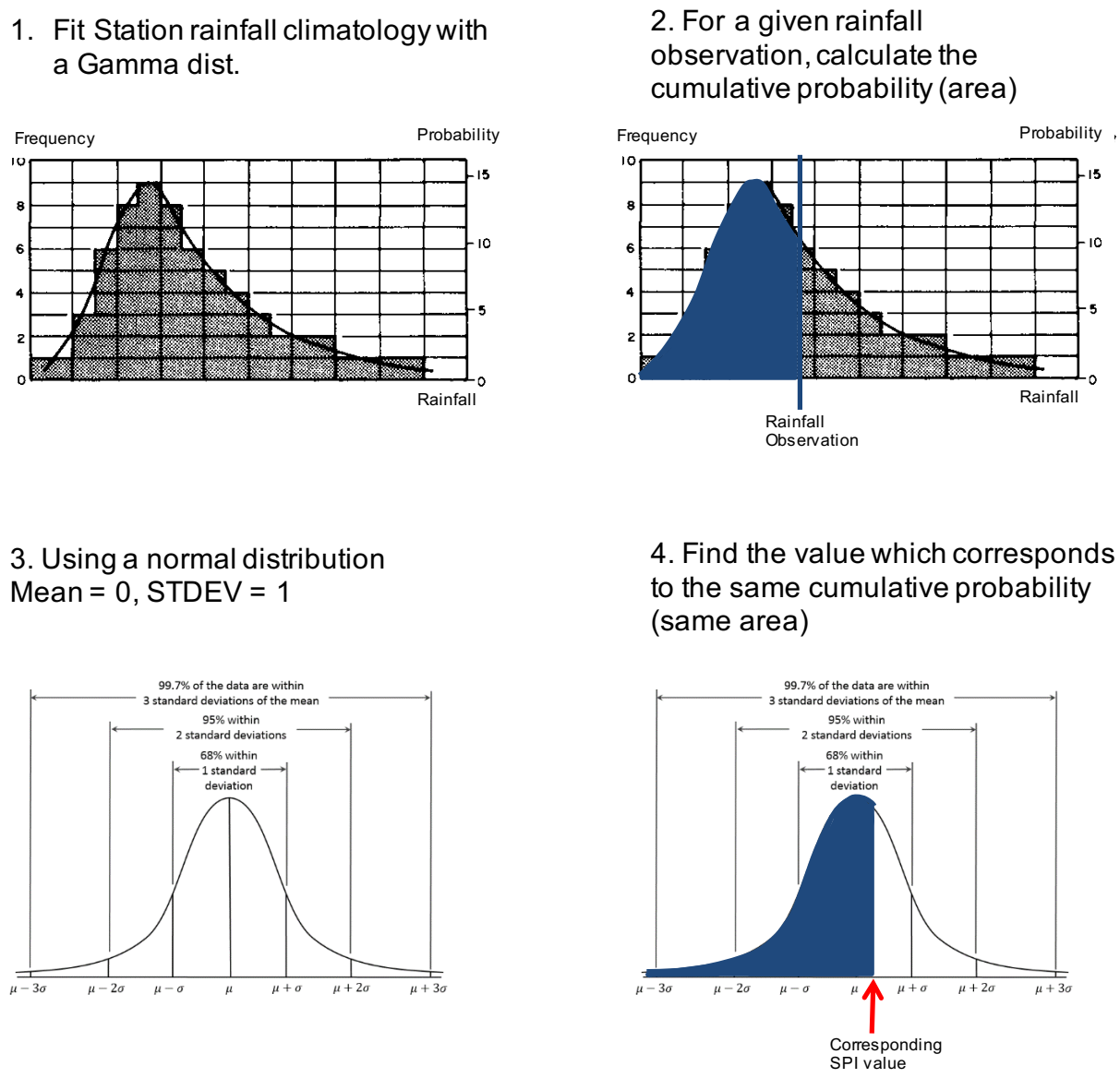


Figure 2.2: Schematic figure of the calculation of the Standardized Precipitation Index.

stations with different climatologies. This allows for the calculation of the SPI at each station for the common dry season and average across all coherent stations to form a single regional drought index.

The SPI can be susceptible to errors when data is scarce or unreliable, due to requiring a maximum likelihood fit of a distribution to the data. In our case, while data for the minor USAPIs can be scarce, good quality data is available for the major USAPI stations for which the SPI will be calculated.

2.2.3 Composite Analysis and Significance Testing

The Oceanic Niño Index and the Standardized Precipitation Index are used together to understand the relationship between ENSO and rainfall anomalies in the USAPIs. Once specific years are selected for study, composite maps of rainfall, SST, and stream function of anomalous winds are produced. This is done for the 6 month dry season for which anomalies are strongest, as well as the 3 month seasons leading up to the dry season.

The mean anomalies of contrasting composite fields were tested for statistically significant differences. In the case of SST anomalies, a Student t-test was used (Blair and Higgins (1980), Wilks (2005)). For rainfall and stream function anomalies, a Wilconxon rank-sum test is used (Wilcoxon (1945), Blair and Higgins (1980), Wilks (2005))

2.2.4 S-EOF Analysis

To understand the features of the different modes of variability as they evolve from season to season, Seasonal-Reliant EOF (S-EOF) analysis (Wang and An (2005)) is used in this work. In S-EOF, anomaly maps for consecutive seasons are treated as a single time step, and the resulting covariance matrix is diagonalized in the same way as in traditional EOF analysis. The resulting S-EOF modes will have a number of sequential patterns (4 in our case as will be discussed later on) showing the spatial features of the coherent modes of variability during each season and a single yearly principal component value for each year of observation.

2.2.5 Physical-Empirical Forecasts and Predictor Selection

A Physical-Empirical forecasting framework is one in which we aim to understand the physical or dynamic mechanisms of causality that link predictor to predictand variable (Yim et al. (2014a), Yim et al. (2015)). This in contrast to a purely empirical prediction framework in which predictors are chosen based on their high correlation to the target variable. This procedure tends to reduce overfitting, by eliminating predictors that represent the same physical mechanisms. Most importantly, it provides a context to interpret the predictions based on the contributions of different processes to the final forecast.

We will develop predictor candidates using two approaches, the first of which is to develop predictors based on the most prominent features of the observed anomaly fields in the seasons leading up to the target season of the forecast. Once these features are identified, a predictive index (the area average of anomalies in a certain region, for example) is formulated to represent them. The second is to use lead-lag correlation analysis with the

predictand and the relevant anomaly fields in the seasons leading up to it. In our study we use SST, and 850 hPa geopotential height, winds and stream function of the anomalous winds in the two 3 month seasons preceding our target forecast season. We also use the two month tendency fields of these anomalies (October-November minus August-September, for example) as possible predictor fields. All predictors are normalized by subtracting the mean and dividing by the standard deviation. This normalization allows for the comparison of the relative contributions of each predictors to the forecast anomaly by comparing the magnitude of the regression coefficients assigned to each predictor.

The best subset of these candidate predictors is selected using stepwise regression (described in the next section). Once the best set of predictors is chosen, a linear regression model between the target and predictor variables is formulated.

2.2.6 Linear, Stepwise and Logistic Regression and Cross Validation Techniques

In this work, we will extensively use linear regression analysis to formulate predictive models. We will use linear regression fitted by ordinary least squares as implemented in the Scikit-Learn Python library (scikit-learn developers (2017b), and underlying SciPy implementation scipy community (2017)). This algorithm calculates the best linear fit between predictor and predictand variables by minimizing the Euclidean 2-Norm S in equation 2.9:

$$S = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2, \tag{2.9}$$

where Y_i and $\hat{Y}_i$ represent the i_{th} observation and model prediction. We will use this basic linear regression model, as well as some variations and modifications to it explained in detail below.

Stepwise regression is a predictor selection methodology for linear regression models, used to select a subset of all possible predictor variables by choosing the most statistically significant among them (Draper and Smith (1981), Panofsky et al. (1953)). One predictor is added to the model at a time (this is called forward selection). After a predictor is added to the model, the significance of the improvement in the R^2 value of the resulting regression, based on a standard F test, is calculated. If the improvement in R^2 is significant at a given level, the predictor is kept in the regression model. At each step of the process, all predictors already in the model are removed one at a time and the R^2 is checked again to see if the

removal of that predictor reduces the R^2 value significantly. If the removal of a predictor already in the model does not reduce the R^2 significantly, the predictor is removed. This means that a predictor can be dropped from the model after it's added if another predictor takes its place. This differs, for example, from a "greedy forward" selection method, in which predictors are added to the model but not taken out. This is done because a predictor that may have added significant prediction skill at an earlier stage in the process, may later be made redundant by new variables in the model. This process stops once no more predictors are able to enter the model and no further predictors are removed.

Stepwise regression tends to choose predictors that maximize the correlation to the target variable and at the same time minimize the correlation between predictors. This is useful in our case because we wish to choose predictors that represent independent physical mechanisms and interpret the coefficients in the linear regression as relative contributions from each of these mechanisms. The significance levels for entry into and exit out of the model are set a priori by the user. In our case, we use 95% significance for both entry and exit. In this work, we use the implementation of the stepwise regression algorithm implemented in Matlab (Mathworks (2017)), which follows the algorithm presented By Draper and Smith (Draper and Smith (1981)).

As part of our forecasting methodology, we will implement a classification forecast. In this forecast the target variable is binary or logistic in type, predicting if a year will or will not be of a certain class and taking a value of 0 or 1 depending on the classification of the particular event. This is opposed to a continuous variable, as is commonly the case in linear regression. To fit this type of data and produce this model we will need to use a model called a logistic regression.

A logistic regression is a common machine learning algorithm which predicts the probability of a target variable Y being 0 or 1 given a set of predictor variables $X = x_{1...n}$. The predictive linear model h_Θ can be written as

$$h_\Theta(X) = \frac{1}{1 + e^{-\Theta^T X}}, \tag{2.10}$$

where the matrix Θ represents the model parameters. Unlike for a linear regression discussed before, there is no closed solution to find the values of the coefficients that maximize the likelihood function, so an iterative process such as gradient descent is implemented. This model can be thought of as a linear function of the variables given as an argument to a sigmoid or logistic function. Writing out the entire linear model, the resulting equation is:

$$h_\Theta(X) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n)}}. \tag{2.11}$$

Figure 2.3 shows an example of how a logistic regression would fit categorical data with 1 predictor as compared to a linear regression.

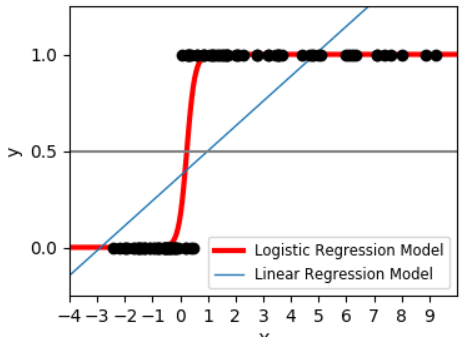


Figure 2.3: Example of a linear (blue) and logistic (red) regression as fit to a classification dataset (black dots) with 1 predictor variable. Taken from scikit-learn developers (2017d).

We will use the implementation of the logistic regression algorithm found in the Scikit-Learn Python library (scikit-learn developers (2017c)). For each data point, the logistic regression calculates the probability of that example belonging to one of the two categories based on the values of the predictors. The category for which the probability is highest is chosen as the prediction. In our work, we will use a threshold between prediction categories of 0.5.

To estimate the predictive power of the models developed here, we will use the cross validation techniques known as leave-one-out (LOO cross validation) and k-fold cross validation (Wilks (2005), Picard and Cook (1984)). In the leave-one-out cross validation method, as the name implies, the model is trained on all data of the training set with the exception of one year. The prediction error of the model is then calculated on the left out sample. This process is repeated for all years in the sample, and the error for each year is then averaged as an estimate of the model predictive power on independent data. This is similar to the jackknife method, but in that case statistics are computed on remaining samples and not left out samples (Efron and Stein (1981)). Similarly, with k-fold cross validation, the training set is divided into k equal parts. One of these subsets is set aside and the model trained on the k-1 remaining sets. The error is then calculated as the mean of the errors of the predictive model on the left out set, and then averaged out over all the sets (scikit-learn developers

(2017a)). Both of these methods are used because, on small training sets such as ours, the more commonly-used leave-one-out method tends to overestimate the error.

CHAPTER 3
OBSERVED FEATURES OF SEASONAL RAINFALL VARIABILITY IN THE USAPIS

3.1 Different USAPI Rainfall Regimes as Described Using K-Means

The areas in the Northern RMI most severely affected by the 2013 drought suffer from large rainfall data gaps, including no data at many stations during 2013. This includes the atolls of Enewetak and Utirik for which severe impacts during the 2013 drought were reported.

In an effort to overcome large data gaps for the stations most affected by possible non El Niño induced droughts, other USAPI stations with longer and complete rainfall records were chosen as proxies. To chose proxy stations, the first criterion applied was similarity in annual cycle. The Monsoon Index and K-Means clustering were used to determine the similarity of the climatological annual cycle of monthly rainfall for the USAPI stations.

To select the most appropriate number of clusters for our K-Means algorithm, the Davies-Bouldin index was calculated for K-Means clustering using from 2 to 42 clusters. Because K-Means clustering is sensitive to centroid initialization and may in some cases produce suboptimal solutions, the clustering was repeated 200 times for each value of k. The mean of all Davies-Bouldin index values for each number of clusters was used as a measure of the appropriateness of the number of clusters. Figure 3.1 shows the mean Davies-Bouldin index values for all k values in the range from 2 to 42, with the standard deviation of all values as error bars. Using 2 clusters provides the best clustering in our data, providing the lowest Davies-Bouldin Index value for a desirable low number of clusters.

Figure 3.2 is a spatial map of USAPI stations assigned into clusters according to their annual cycle using K-Means clustering and k=2 clusters. Using 2 clusters, the USAPI stations can be separated into two distinct categories, northern and southern stations. The mean annual cycle of monthly rainfall, which is also the centroid of each cluster, is presented in Figure 3.3.

Northern stations show moderate monsoonality to their annual cycle. They have a dry season from December to June and a wet season from July to November. The Monsoon Index for the northern stations is above 30% (the difference between wet and dry season rainfall is 30% of the total annual rainfall). Required values of the Monsoon Index for an

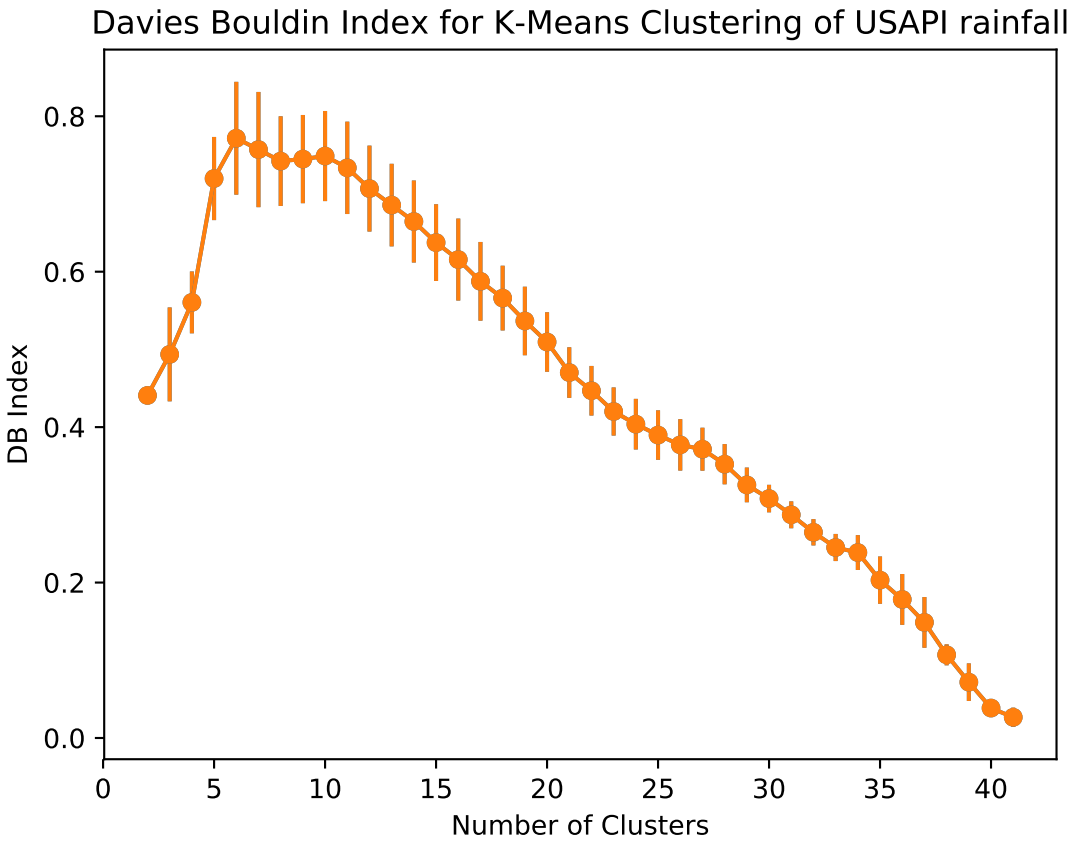


Figure 3.1: Mean Davies-Bouldin index values for K-Means clustering of USAPI climatological annual cycle of monthly rainfall as a function of number of clusters. Clustering and calculation of the Davies-Bouldin Index was repeated 200 times per value of k. Error bars represent the standard deviation of all Davies-Boulding Index values obtained.

area to be considered part of the Global Monsoon domain are of 50% and greater (Wang and Ding (2008)), placing these stations just outside of it. Stations most severely affected by the 2013 drought event, such as Enewetak, Utirik and Wotje, are all classified into this cluster. Anomalous dry conditions during early 2013 coincided with the climatological dry season and compounded to produce severe drought impacts in these stations.

Stations to the south show year-round rainfall, with slightly dryer than average conditions during February and October. This type of rainfall regime is consistent with a climate dominated the seasonal migration of the ITCZ. Previous analysis of the rainfall annual cycle on the USAPIs depicted northern and southern USAPI stations as having similar annual cycles (monsoon for both) but with a slight difference in the timing of the peak (Yu et al.

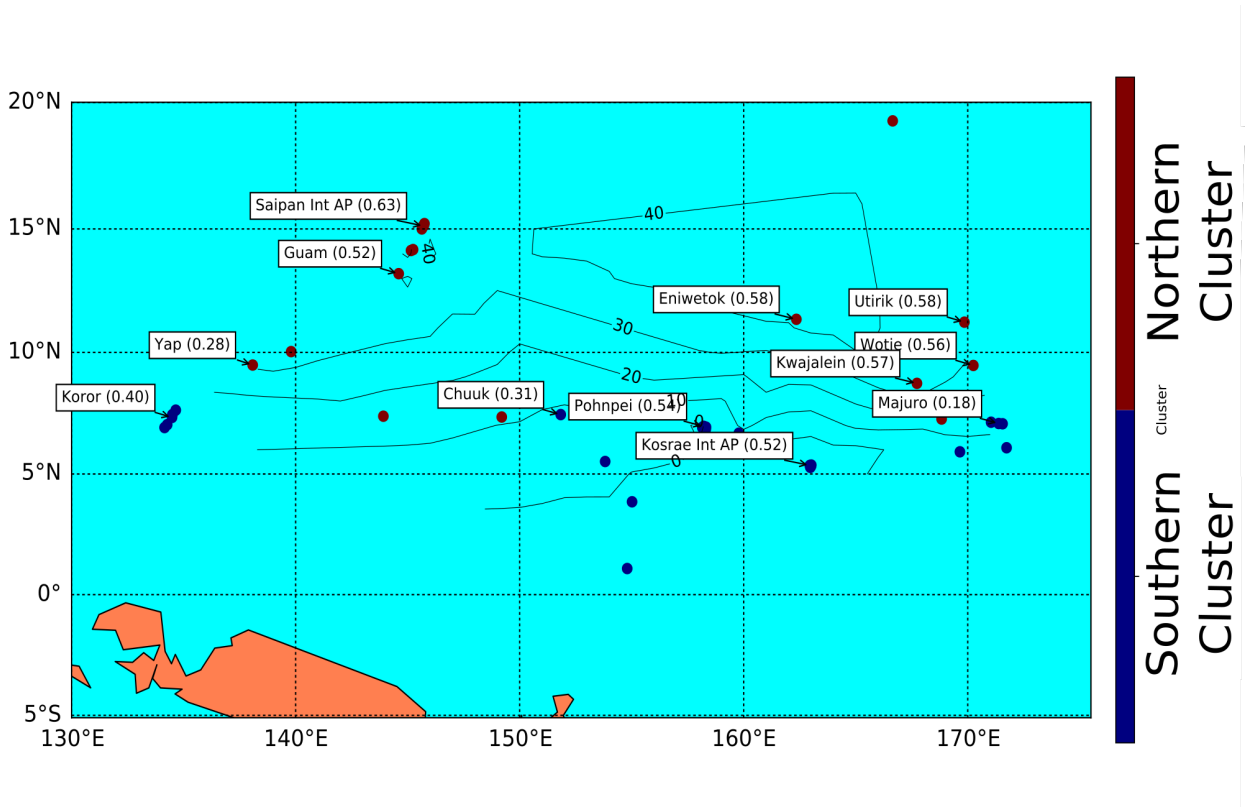


Figure 3.2: Map showing the USAPI stations and their assigned K-Means cluster (color) using k=2, Monsoon Index (contours), and silhouette score for most relevant stations (number next to station name).

(1997a)). We argue that the difference in annual cycle between northern and southern USAPIs is more significant than a shift in peak rainfall timing and justifies separate study of ENSO impacts on northern and southern stations.

The separation of stations into northern ans southern stations is a robust feature, and is still present if 3 clusters are used. Figures 3.4 and 3.5 show the spatial map and average annual cycles of the USAPI stations as described by K-Means clustering using 3 clusters.

Rainfall records for Guam, Kwajalein and Yap are available since 1966 and are all classified as having the same type of annual cycle as those affected by the 2013 drought. The correlation between time series of December to May rainfall anomaly for these stations for the period of 1966 to present exceeds the 99% confidence level. The silhouette score, which measures the distance between a cluster point and the centroid of the nearest cluster, is 0.30 for Yap. This implies that the annual cycle in Yap is more similar to the mean annual cycle of southern stations than for other stations in its cluster with higher silhouette scores. Due

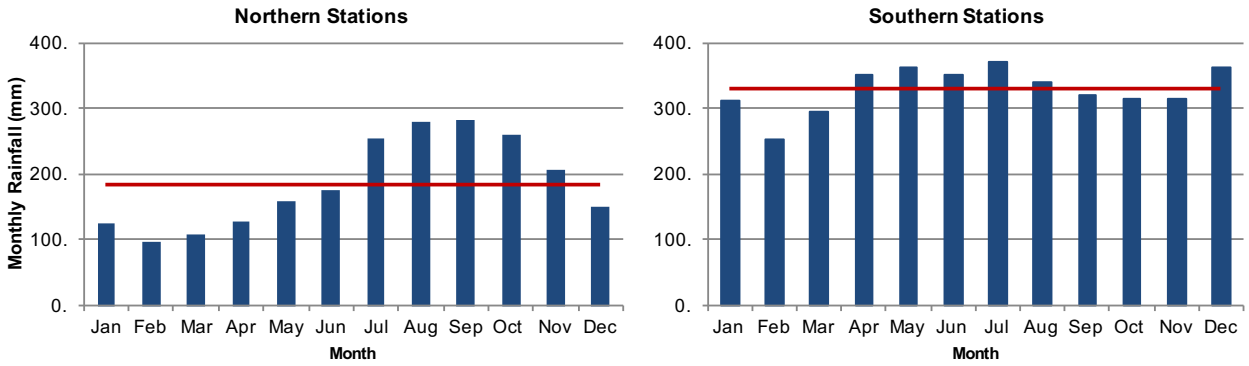


Figure 3.3: Average annual cycle of monthly rainfall for the northern and southern cluster as calculated using K-Means with 2 clusters.

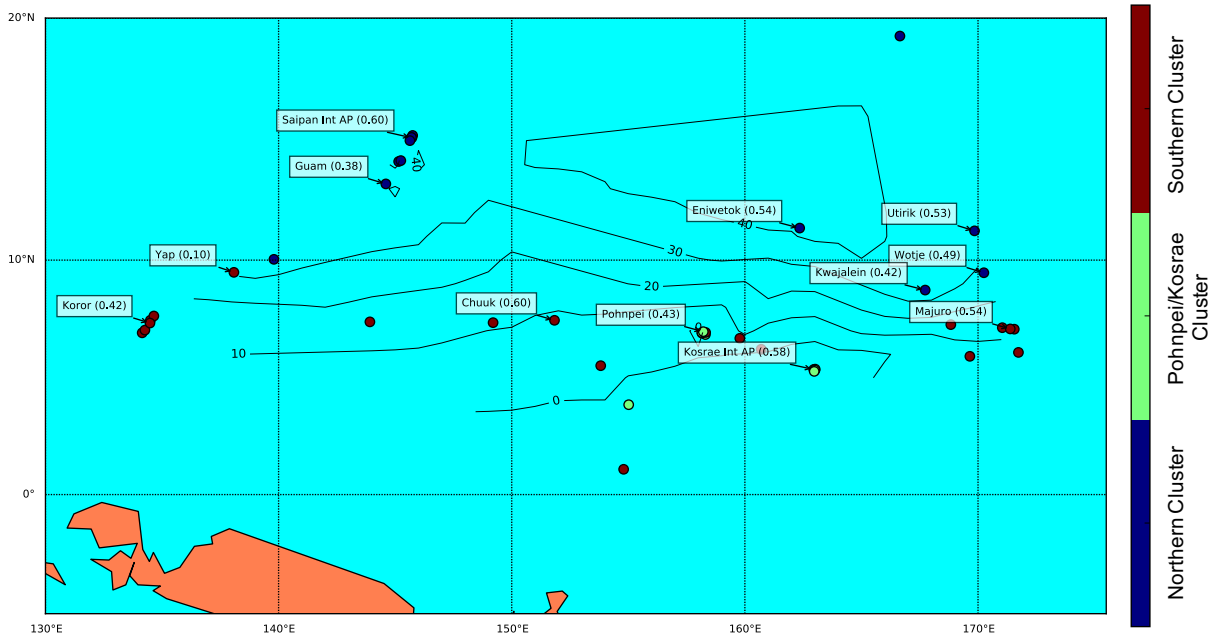


Figure 3.4: Map showing the USAPI stations and their assigned K-Means cluster (color) using k=3, Monsoon Index (contours), and silhouette score for most relevant stations (number next to station name).

to their very similar annual cycle and coherent interannual variability, Kwajalein and Guam are chosen as representative stations for rainfall variability in the northern USAPI region, whereas Koror, Chuuk, Pohnpei, and Majuro are used for the southern stations. The SPI for total rainfall from December to May was calculated individually for each station and

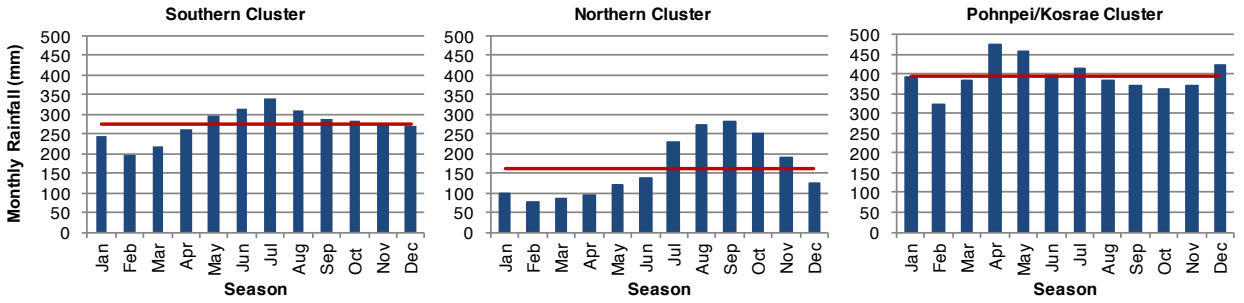


Figure 3.5: Average annual cycle of monthly rainfall for the northern and southern cluster as calculated using K-Means with 3 clusters.

averaged with the other proxy stations in that cluster.

3.2 Impacts of ENSO North vs. South USAPI Stations

The relationship between rainfall (SPI) and ONI differs between the northern and southern USAPI stations (Figure 3.6). For the strongest El Niño events (i.e., cold tongue events, ONI $> 2°$C) such as 1983, 1998 and 2016, severe drought conditions occur during the dry season (December of the previous year to May of the referred year) for both northern and southern USAPI stations. For ONI values less than 2°C, the southern USAPI stations exhibit negative correlation (correlation of -0.72, significant at the 99.9% confidence level) between SPI and ONI, with moderate El Niños producing drier than normal conditions and La Niñas producing wetter than normal conditions. On the other hand, rainfall at the northern USAPI stations does not correlate with the ONI, suggesting that other factors contribute to rainfall variability in this region beyond the amplitude and phase of ENSO.

The years selected for our composite analysis in Section 3.3 are shown in Figure 3.6 with color makers. We will focus our analysis on the ONI range between $-1°$ and 2°C, where rainfall variability in the northern USAPI region has highest amplitude. We have excluded from analysis years with SPI or ONI anomalies close to zero. We will consider years with ONI values of less than -1°C separately.

Years with anomalously warm ONI and dry SPI are referred to as Warm Dry years (red circles in Figure 3.6), and include 1988, 1992, 2005, 2007 and 2010. Similarly, years with anomalously cool ONI and wet SPI are referred to as Cool Wet years (blue circles in Figure 3.6), and include 1985, 1986, 1996, 1997 and 2014. In this work, we will consider Warm Dry and Cool Wet years in the northern USAPIs as Canonical years. This is because

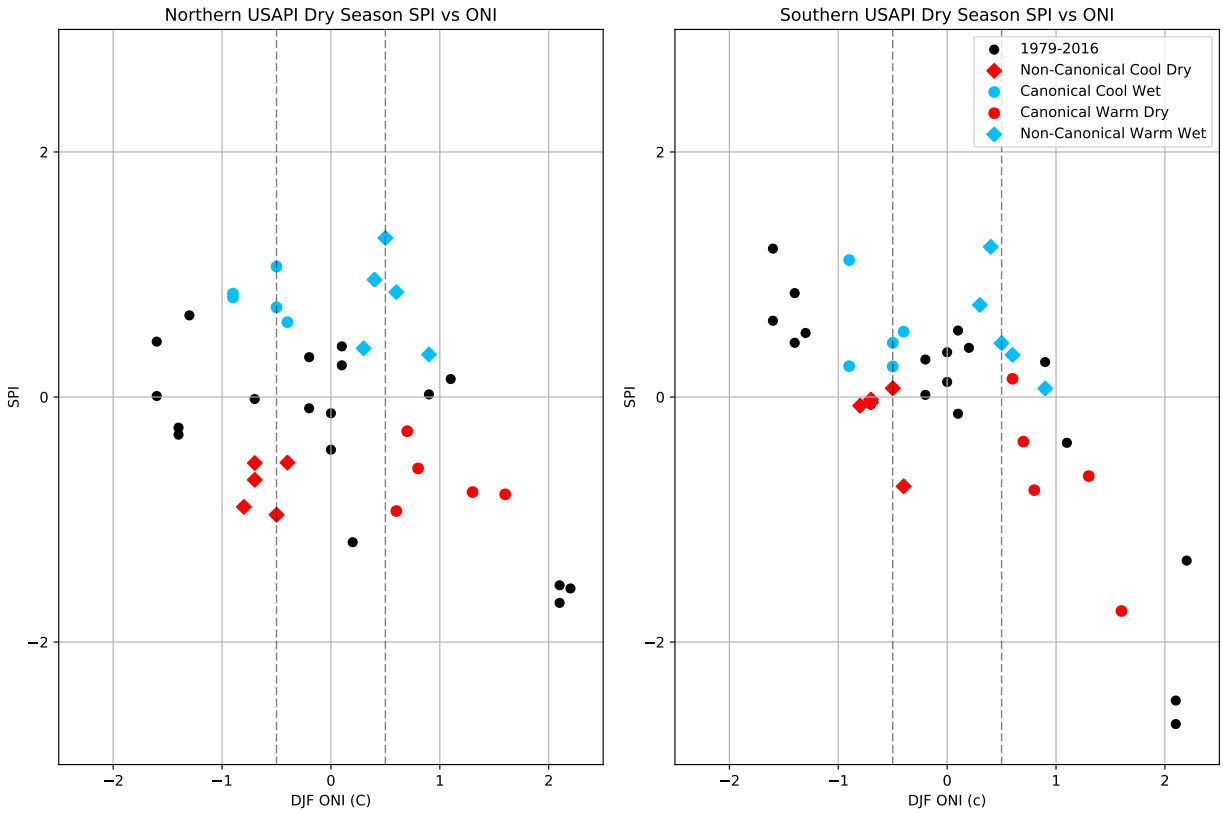


Figure 3.6: Scatterplots of the dry season SPI (Dec to May) averaged for the northern (left) and southern USAPI stations (right) vs. the Oceanic Niño Index for the corresponding December-January-February season as published by NOAA CPC. Color markers represent years chosen for composite analysis. Diamond markers represent years for which rainfall anomalies in the Northern USAPIs are of opposite sign to what would be expected from the ONI conditions. We will refer to these as Non-Canonical years. Circular markers, in contrast, represent years for witch the rainfall anomalies are as expected considering the corresponding ONI value. We will refer to these as Canonical years. Vertical dashed lines represent the 0.5°C and -0.5°C ONI values.

these years show the expected negative correlation between ENSO and rainfall in the USAPIs (Ropelewski and Halpert (1987), Murphy et al. (2014)). Conversely, Cool Dry (red diamonds in Figure 3.6) and Warm Wet (blue diamonds in Figure 3.6) represent years for which dry season rainfall and DJF ONI appear to be positively correlated. These years will be referred to as Non-Canonical years.

Table 3.1 shows the classification of each of the years chosen for composite analysis according to CPC operational methodology, as well as by Kug et al. (Kug et al. (2009)) and Murphy et al. (Murphy et al. (2014)). The methodology used to select composite member

years yields a very different year selection than previous works in which selection was based solely on sea surface temperature conditions. In most cases, years with dissimilar rainfall impacts over the northern USAPI are grouped together into composites, erasing the features that produce these diverse anomalies. Nonetheless, these differences in rainfall are significant, with the Warm Wet and Warm Dry events having a difference in SPI value significant at the 99.9% test level (with a p-value of 10^{-4}), and Cool Dry and Cool Wet events an even greater difference (p-value of 10^{-6}).

According to the PEAC Center's "Pacific ENSO Update" newsletter, dry conditions and drought impacts were also reported for years that coincided with cool Pacific SST conditions. These years are selected by our methodology to form part of the Cool Dry composite. The progression in time of these reported impacts is also similar to those observed in 2013. For example, dry conditions persisted from the second half of 2005 into 2006 in the Northern USAPIs. Guam reported a longer than normal wildfire season resulting in evacuations, and very low levels in the Fena reservoir, which provides water to about 20% of the population. In Majuro, water rationing was mandated by February 2006 and encouraged though the following months. Similarly, in 2009, persistent dry conditions in the Northern RMI required the delivery of drinking water to Utirik in late April and the issuing of special weather statements for the RMI by the Guam National Weather Service Forecast Office. Neither 2006 nor 2009 is classified as a La Niña events by current methodologies.

Canonical Cool Wet			
Year	CPC	Kug et al. 2009	Murphy et al. 2014
1985	Niña	Niña	–
1986	–	–	–
1996	Niña	–	–
1997	–	–	–
2014	–	**	–

Non-Canonical Warm Wet			
Year	CPC	Kug et al. 2009	Murphy et al. 2014
1980	Niño	–	–
1991	–	WP Niño	–
1995	Niño	WP Niño	WP Niño
2004	–	–	
2015	Niño	**	**

Non-Canonical Cool Dry			
Year	CPC	Kug et al. 2009	Murphy et al. 2014
1984	–	–	–
2001	Niña	Niña	–
2006	–	Niña	–
2009	–	**	–
2013	–	**	–

Canonical Warm Dry			
Year	CPC	Kug et al. 2009	Murphy et al. 2014
1988	Niño	Mix Niño	Mix Niño
1992	Niño	Mix Niño	Mix Niño
2005	Niño	WP Niño	WP Niño
2007	Niño	**	WP Niño
2016	Niño	**	**

Table 3.1: Classification of all events selected for our composite analysis according to CPC operational methodology, Kug et al. (Kug et al. (2009)) and Murphy et al. (Murphy et al. (2014)). WP and Mix represent the terms Warm Pool and Mixed El Nino respectively.

For ONI values of -1°C or less, both wet and dry rainfall conditions for the northern USAPIs occur. Rainfall anomalies are not as pronounced as during the Cool Dry years, which show weaker ONI anomalies. Years included in this category are 1989, 1999, 2000, 2008 and 2011. All of these years are considered La Niña years by CPC operational methodology, Murphy et al. (Murphy et al. (2014)) and Kug et al. (Kug et al. (2009)) this last reference leaving out 2008 and 2011 due to the time period considered.

It should also be noted that many of the years selected here which show the strong impacts over the northern USAPI region are not often studied because they are overlooked by most ENSO active phase selection criteria. In most cases, SST anomalies during these years do not meet magnitude or persistence criteria such as the CPC requirement for 5 consecutive seasons of ONI index being below -0.5°C. Nonetheless, their distinct impacts over the USAPI region and the severe drought conditions seen during years with negative rainfall anomalies warrant a closer look at these events.

The results in Figure 3.6 are robust and can be reproduced using GPCP rainfall anomalies averaged over the Northern USAPIs. The rainfall anomalies averaged in the region defined by 10°W to 15°N and 140°W to 170°W reproduce the results obtained for station data best, but results remain consistent to reasonable changes in the area used to define rainfall anomalies.

3.3 December to May Dry Season Composites

In this section, we contrast the composite climate anomalies between the Canonical and Non-Canonical regimes. In Figure 3.7, the top left and bottom right panels are the composites of rainfall anomalies for the Canonical years. Rainfall patterns during Canonical events are consistent with the familiar ENSO horseshoe shape. During Warm Dry events, wet anomalies occur over the central and eastern Pacific and dry anomalies over the western Pacific that extend northeast and southeast. During Cool Wet years, the opposite pattern occurs. The central Pacific rainfall anomalies during these events are confined to south of about 5°N west of the dateline and increase in latitudinal range to the east. The tropical northwest Pacific (0 to 30°N, 120°E to the dateline) anticyclonic low-level wind anomalies coincide with dry conditions in the region during Warm Dry years, while cyclonic anomalies coincide with wet conditions during the Cool Wet years.

In contrast, the lower left and upper right panels of Figure 3.7 show the same composites for the Non-Canonical Cool Dry and Warm Wet years. Because Non-Canonical Cool Dry

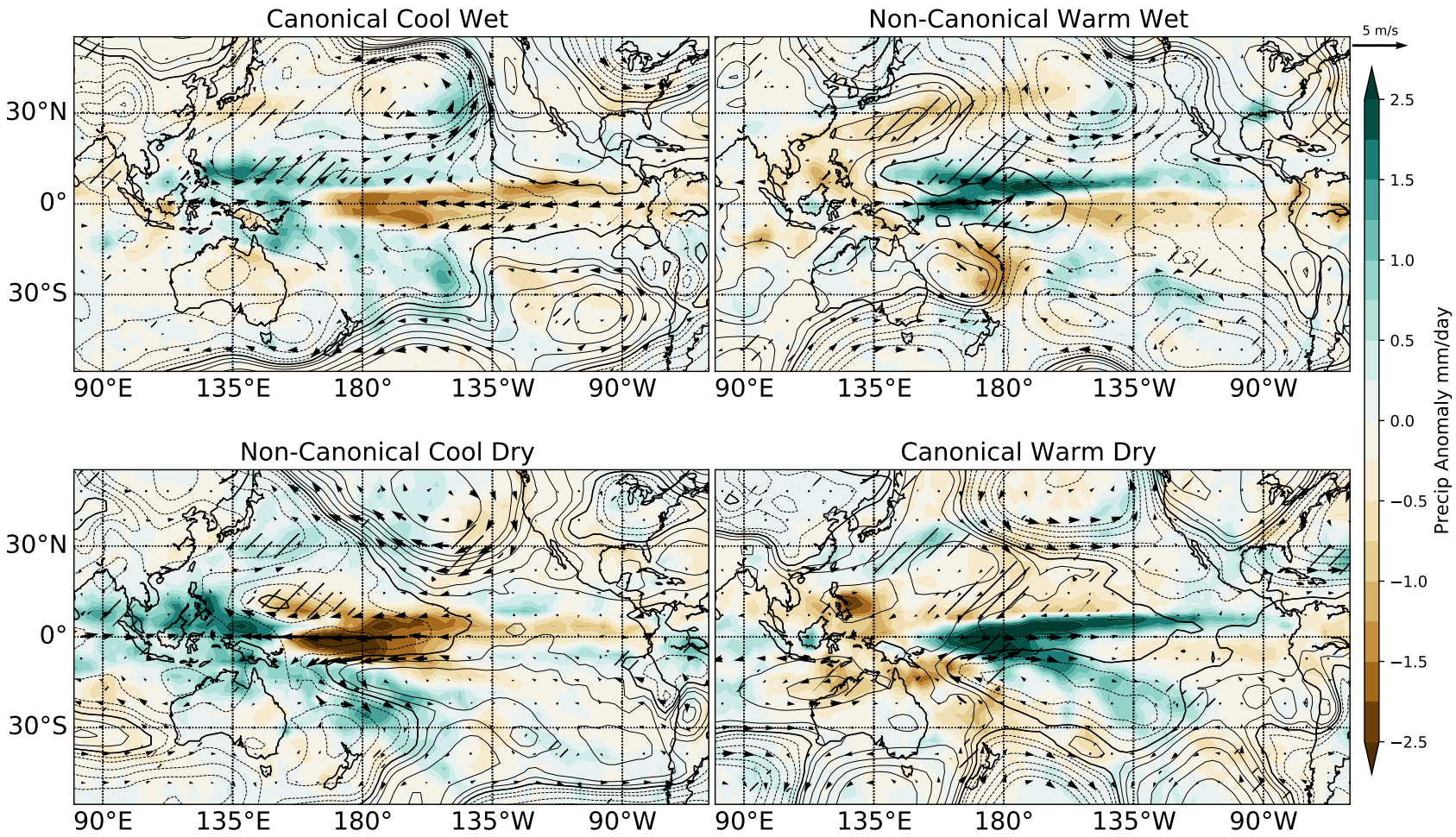


Figure 3.7: Rainfall (shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in rainfall anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a Wilcoxon rank-sum test (Wilks (2005)).

events are of particular interest as they bring unexpected droughts to the USAPIs, we will focus on comparing the features of these events with those of their Canonical Cool Wet counterparts. Comparisons for the Warm Dry and Warm Wet events will be assumed to be opposite (for the sake of brevity) unless otherwise pointed out.

In the Cool Dry composite, an anomalous rain belt extends the dry rainfall anomalies observed near the dateline toward Micronesia. This rain belt has it's strongest anomalies between about 7° and 15°N and is not present in Cool Wet events. Tropical northwest Pacific circulation during Cool Dry events is dominated by well-defined anticyclonic anomalies, consistent with the observed rainfall anomalies in the area, while Cool Wet events show the Western Pacific dominated by cyclonic anomalies.

Canonical Cool Wet and Non-Canonical Cool Dry events differ significantly in their circulation patterns over a large portion of the Pacific Basin, in addition to over the northwest tropical and subtropical Pacific (Figure 3.8). Strong circulation anomalies in the tropical western Pacific during the Dec-May dry season are cyclonic for Cool Wet events and anticyclonic for Cool Dry events. In both cases, Tropical Western Pacific circulation anomalies show support from mid-latitude circulation anomalies of the same sign as observed.

Over the eastern and southern Pacific basin, Cool Wet and Cool Dry events also exhibit significant circulation differences. In the tropical South Pacific, Non-Canonical Cool Dry events show very strong cyclonic circulation, while the Canonical Cool Dry composite shows weak anticyclonic anomalies over the entire region. In the tropical eastern Pacific (140°W), Cool Wet events show a strong anticyclonic circulation pattern similar to what is observed during the decaying phase of La Niña events, while during Cool Dry events the eastern tropical Pacific shows weak cyclonic anomalies. A significant difference between the Canonical Warm Dry composite and other composites is that the circulation anomalies over the northern central Pacific and the tropical western Pacific are of opposite sign.

The Canonical and Non-Canonical cool events, even having similar DJF ONI values (all between 0 and -1°C), are characterized by very distinct equatorial and off-equatorial spatial SST anomaly patterns (Figure 3.9). In the equatorial Pacific, Cool Wet years show strongest negative SST anomalies over the eastern Pacific from around 140°W extending towards the South American coast, with peak magnitude around 120°W. Cool Dry years show strongest negative anomalies to the east of 140°W and extend to the dateline. In the tropical western north pacific, Cool Wet years show weak warm SST anomalies from the equator to about 20°N, with cool anomalies to the north. During Cool Dry years, warm anomalies dominate the region.

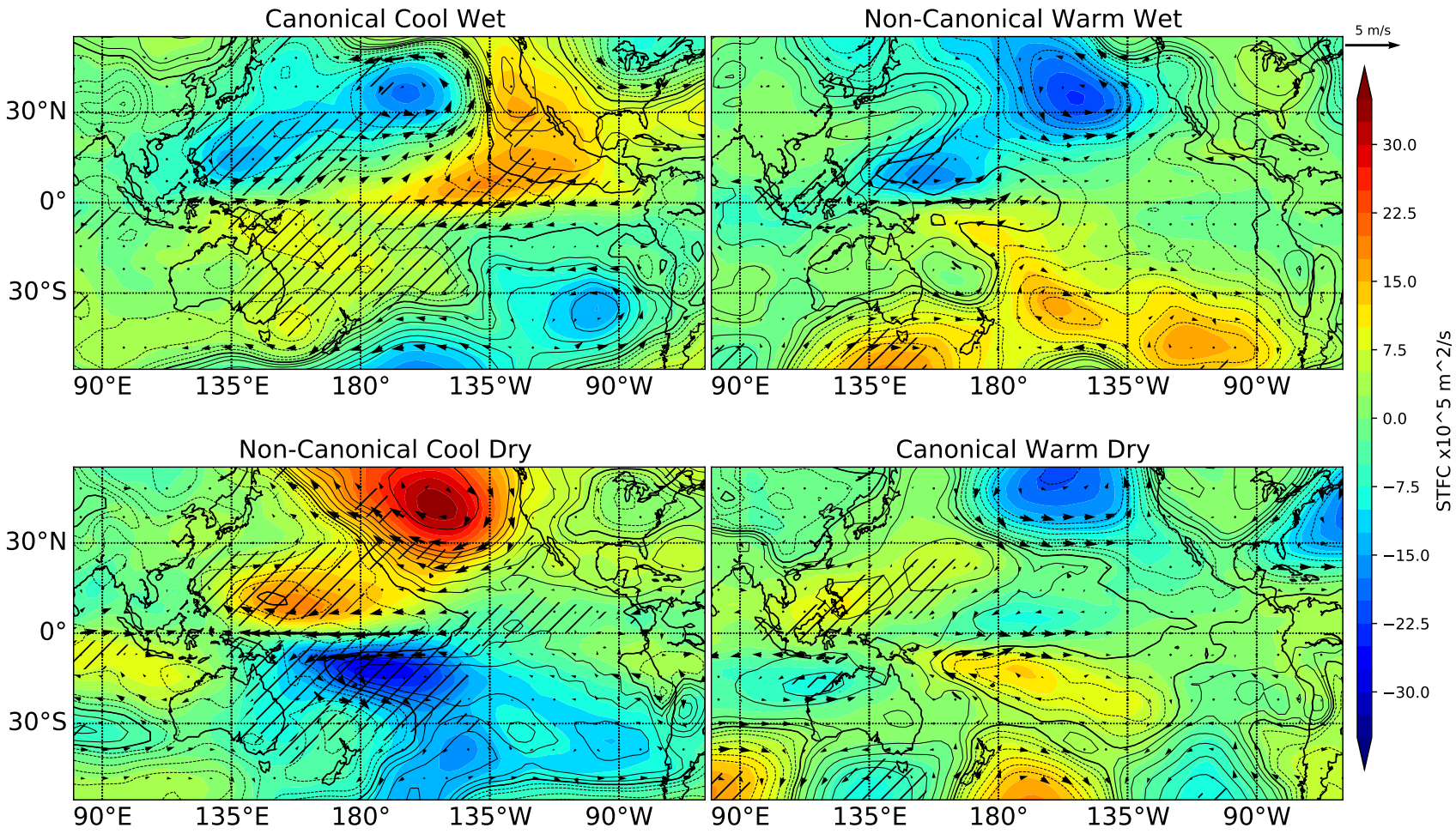


Figure 3.8: stream function of the anomalous wind(shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry, and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a Wilcoxon rank-sum test (Wilks (2005)).

Equatorial low-level easterly winds in the Cool Wet events are confined east of the dateline where the strongest SST gradients are observed. Westerly wind anomalies are present in the western part of the basin. These westerly winds form on the southern side of the regional cyclonic circulation anomalies observed in the Western Pacific during Cool Wet events. During Cool Dry events, easterly wind anomalies are only present west of 150°W, and extend all the way to the Maritime Continent. These easterly winds form the southern branch of the regional anticyclonic circulation that characterizes Cool Wet events. Weak wind anomalies are observed in the eastern Pacific during Cool Dry events.

In the mid latitudes, the Cool Wet composites show a strong low-pressure system and associated cyclonic circulation, as well as cool and warm SST anomalies (30°N and dateline and near the California coast respectively). Mid-latitude anomalies during Cool Dry events are opposite to those observed during Cool Wet events, with strong high pressure, anticyclonic circulation and cool SST anomalies to the east and south, and warm SST anomalies to the west.

Anomaly patterns for strong La Niña events are consistent with the Cool Dry and Cool Wet composites discussed above but with some modifications. Composites of La Niña years with wetter than normal (1989 and 2011) and drier than normal (1999 and 2008) conditions over the northern USAPIs are shown in Figure 3.10. SST anomalies for both Dry and Wet La Niña composites are considerably larger magnitude than those of Cool Dry and Cool Wet composites. The location of the highest SST anomalies is slightly more to the east in the wet La Niña composites than in the dry La Niña composites, but this shift is not as prominent and clear as in the Cool Dry and Cool Wet composites.

In spite of this more similar SST pattern between wet and dry La Niña events, local circulation and rainfall anomalies over Micronesia do show differences between the composites which are similar to the differences between the Cool Dry and Cool Wet composites. Dry La Niña events, like Cool Dry Non-Canonical events, show relative high pressure anomalies over the Micronesia region and large-scale anticyclonic circulation anomalies that produce a band of negative precipitation anomalies over the region. Wet La Niña composites show westerly winds over the Micronesia region north of about 5°N and accompanying positive rainfall anomalies. In the case of strong La Niña events, both dry and wet, tropical easterly anomalies are much stronger than those seen in the weak La Niña composites.

Both Dry and Wet La Niña events have stronger equatorial easterly wind anomalies along the equator than the Cool Wet and Cool Dry events. These stronger equatorial wind anomalies prevent local circulation anomalies over Micronesia from becoming as well defined

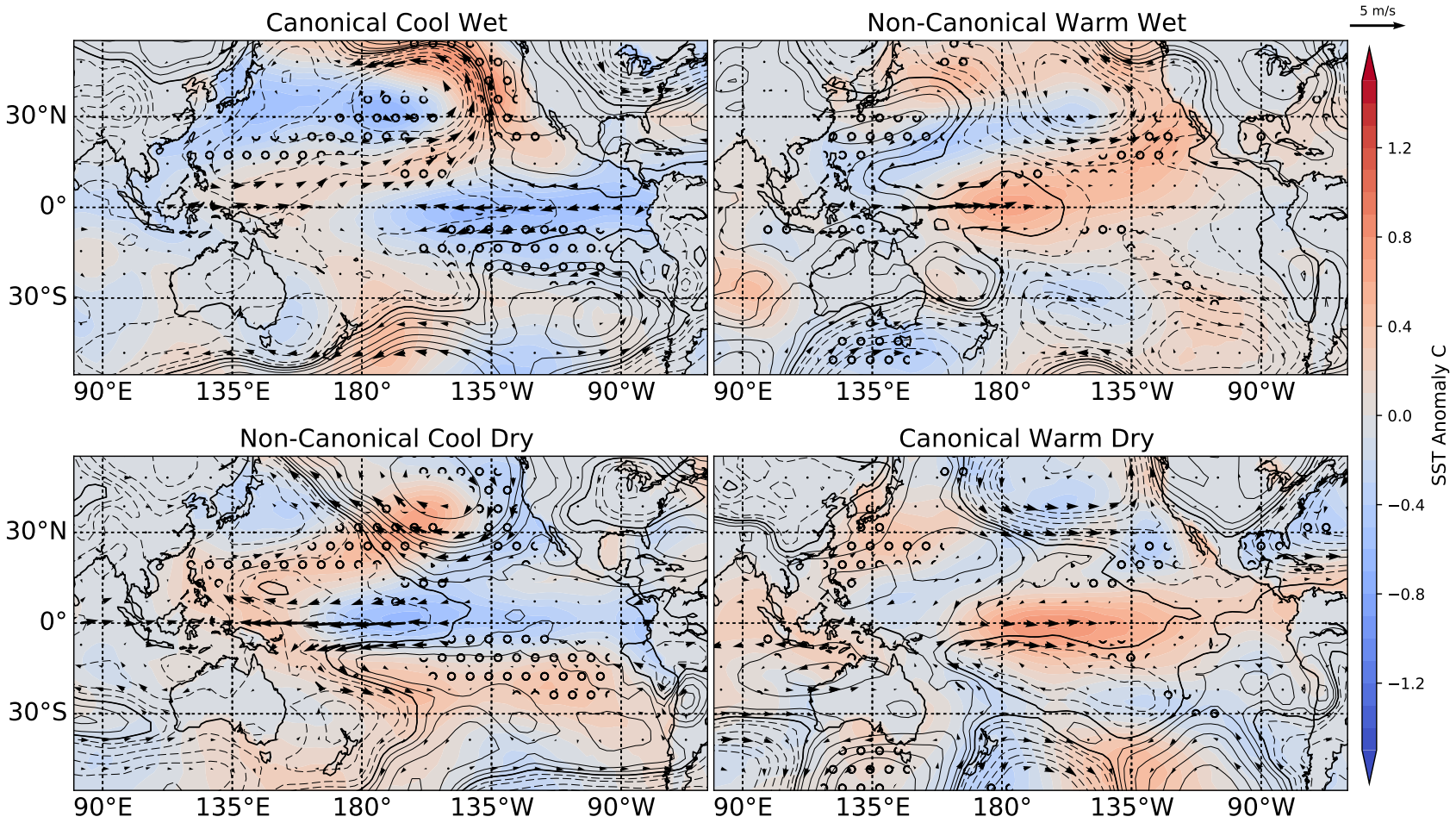


Figure 3.9: SST anomaly (shading), 850 hPa geopotential height (contours) and wind composites for the Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet years. Hatching represents areas where the difference in SST anomalies between composites of the same ONI character (Cool Wet vs. Cool Dry, for example) are significant at the 90% confidence level according to a 2 tailed Student t-test (Wilks (2005)).

during strong La Niña events as they are in the cool events (compare the bottom panels of Figure 3.10 with the left panels of Figure 3.8). This, in turn, is likely the reason why strong La Niña events produce similar but weaker rainfall anomalies over the northern USAPIs as the cool events.

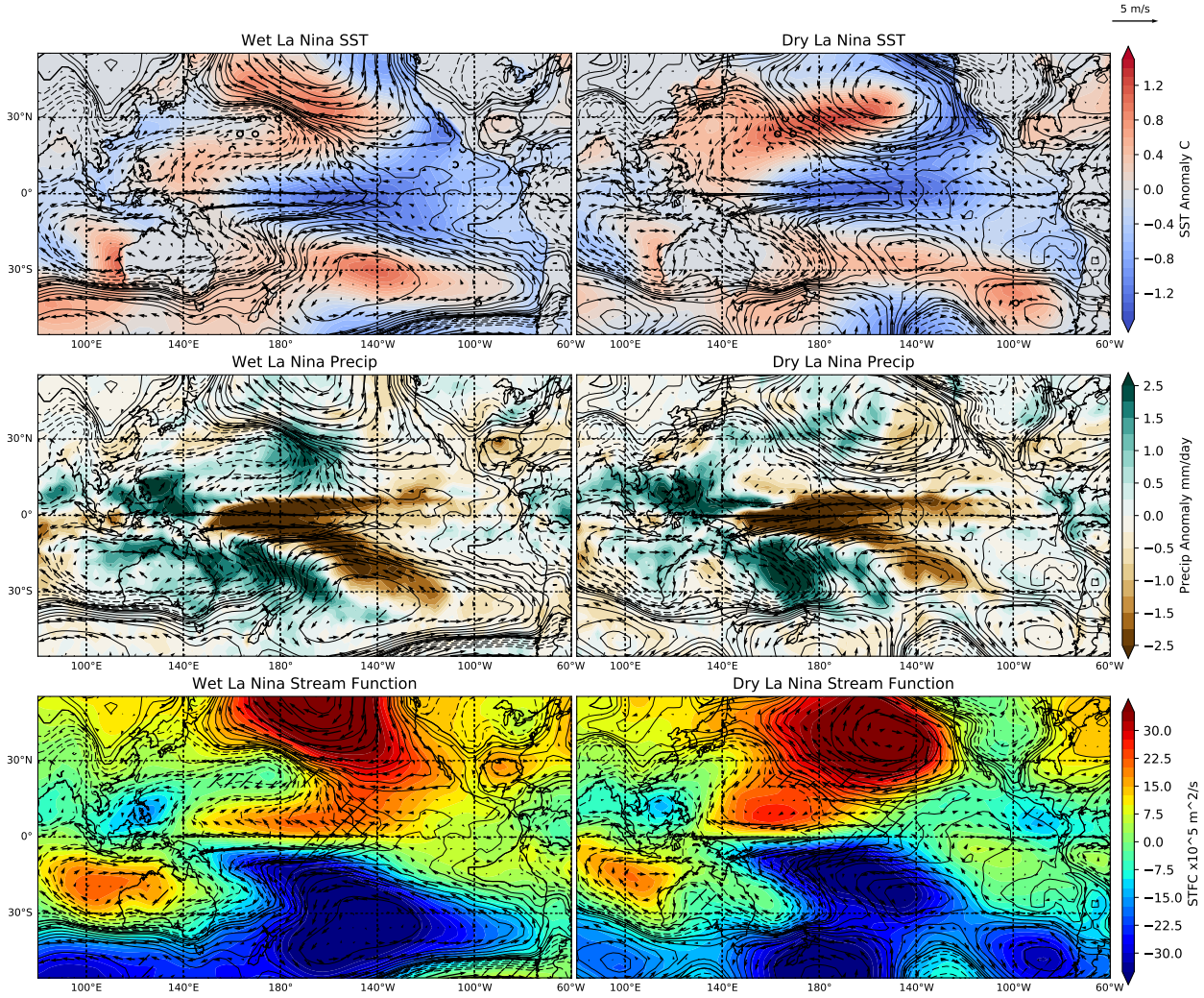


Figure 3.10: December-May SST anomaly (shading top), Precipitation anomaly (shading mid) and stream function (shading bottom), along with 850 hPa geopotential height (contours) and wind composites wet and dry La Niña events. Hatching represents areas where the difference in anomalies between wet and dry composites are significant at the 90% test level.

The results of Murphy et al. (2014) are consistent with the observations made here. A composite of rainfall anomalies for Central Pacific El Niño events (their Figure 5) captures weak positive rainfall anomalies that extend from the central Pacific into the northern USAPI

region. This composite is similar to our composite for Non-Canonical Warm Wet events (top right panel of Figure 3.7). In their case, mean anomalies in this rain belt are not significantly different from zero. This is because this composite is made by using a mix of Canonical Warm Dry years and Non-Canonical Warm Wet years. The Central Pacific El Niño events used for that composite are 1995 (considered in this study to be a Non-Canonical Warm Wet year) and 2005 and 2007 (considered in this study to be Canonical Warm Dry years).

Similarly, in their composite for La Niña events (also their Figure 5), a belt of anomalous negative rainfall extends from the central Pacific toward northern Micronesia (not statistically significant). This composite is similar to our composite of Non-Canonical Warm Dry events (bottom left panel of Figure 3.7). The years used in their La Niña composite are 1989, 2011 (considered as wet La Niña events in our case), 1999 and 2008 (considered as dry La Niña events) and 2000 which is left out of our analysis due to its near normal rainfall anomalies over the northern USAPI.

Murphy et al. (2014) comment on this belt of rainfall anomalies during Central Pacific El Niño events. In their SST composites for Central Pacific El Niño events, warm SST anomalies are observed in the Tropical Western Pacific over Micronesia.The authors argue that these warm SST anomalies serve to directly enhance convection over the region, resulting in the formation of the anomalous rainfall belt. Our SST composites for Warm Wet events during the Dec-May season show warm SST anomalies confined near the dateline, with cool anomalies to the west. This SST dipole suggests that coupled ocean-atmosphere dynamics play a role in forcing regional circulation anomalies and subsequent rainfall anomalies rather than direct thermodynamic forcing. Murphy et al. (Murphy et al. (2014)) do not highlight the other features described here.

It is clear that during the dry season of December to May, Canonical and Non-Canonical warm and cool events show a distinct set of atmospheric and oceanic anomalies that explain why their impacts on the northern USAPI region are so diverse. We will discuss in the following section how the evolution of these anomalies differs between Canonical and Non-Canonical events.

3.4 Contrasting Seasonal Evolution of SST and Atmospheric Anomalies Between the Canonical and Non-Canonical Responses

The evolution of the ONI index leading up to the December to May dry season for Canonical events resembles that of typical ENSO events. The onset of the events tends to occur in the late spring to early summer and peak by the end of the year (Figure 3.11).

In Non-Canonical Warm Wet events, the evolution of the ONI matches that seen during the Canonical events, with onset during late spring to early fall and intensification and peak during the winter. For Non-Canonical Cool Dry events, the seasonal evolution of the ONI shows less consistency. The onset of these events happens later in the year toward late summer, with ONI values during the preceding JJA season showing a wide range of values. For the 2013 event, ONI values are weak positive in 2012 as late as the November-December-January season, much later than any of the other years in this category.

Rainfall over the northern USAPI is strongly modulated by the anomalous circulation over the tropical western north Pacific. Figure 3.12 shows the composite time series of northern USAPI 6-month period SPI overlaid on the tropical western Pacific stream function of anomalous winds averaged from the equator to 30°N and 120°W to the dateline.

During Canonical events, we observe how rainfall anomalies over the northern USAPI shift sign from the developing to the decaying stage of the events. This shift in rainfall anomalies coincides with a shift in sign of circulation anomalies over the tropical western Pacific. For Canonical Warm Dry events, a wetter than normal June to November wet season transitions to a drier than normal December to May dry season. This transition in rainfall anomalies was described for El Niño events by Ropelewski and Halpert (1987). In compliment to this, for Canonical Cool Wet composites dry anomalies leading up to the Dec-May dry season sharply transition to wet anomalies.

Non-Canonical events, on the other hand, show persistence of rainfall anomalies from the June to November wet season through the December to May dry season that are linked to the persistent anomalous circulation anomalies over the tropical western Pacific. For Cool Dry events, a drier than normal June to November wet season is then followed by a dry December-May dry season. The effects of these long term dry conditions are compounded, making for severe drought impacts across the northern USAPI during Non-Canonical Cool Dry events.

Figure 3.13 shows a progression of seasonal SST, 850 hPa winds and geopotential height

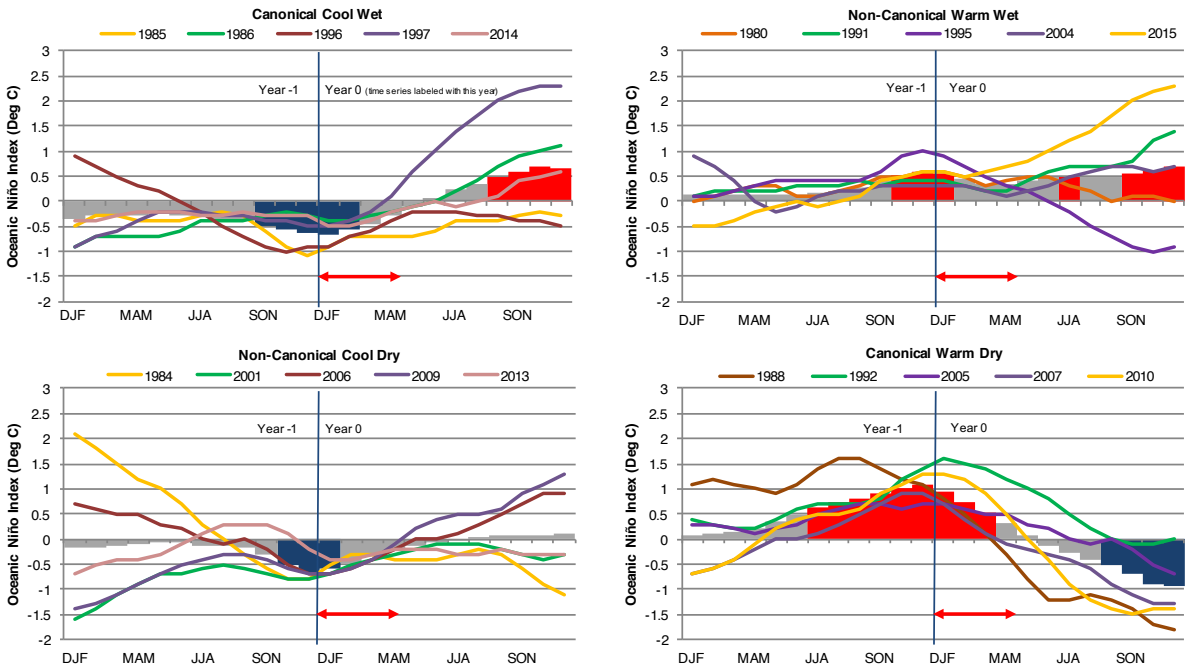


Figure 3.11: Time series of the Oceanic Niño Index for each year selected for composite analysis. ONI index as published by the Climate Prediction Center. Bars represent the average of all events in each category. Bars are colored red where the average is greater than 0.5°C and blue where it is less than -0.5°C. The time series are labeled using the year 0. Red arrows depict the extend of the climatological Dec-May dry season.

anomaly composites from the March-April-May season on to the December-May dry season for each of our composite categories. These composites show the differences in evolution of equatorial SST and tropical western north Pacific circulation between Canonical and Non-Canonical events that we have discussed before. These composites also show differences in the evolution of off-equatorial SST anomalies and their link to circulation anomalies.

For Cool Dry events, the tropical western Pacific anticyclonic anomalies observed during the dry season (and produce dry season droughts) are already established in the March-April-May season (MAM). These tropical western Pacific circulation anomalies are linked to anticyclonic anomalies over the subtropical central Pacific. These subtropical circulation anomalies provide support for the localized central Pacific easterly wind anomalies observed in the MAM, JJA and SON seasons leading up to the Dec-May dry season in the Cool Dry composites. SST anomalies in these leading seasons are cool to the east and south of the tropical and subtropical circulation anomalies, and warm to the north and west. This coinciding pattern of circulation and SST anomalies suggests ocean-atmosphere coupling

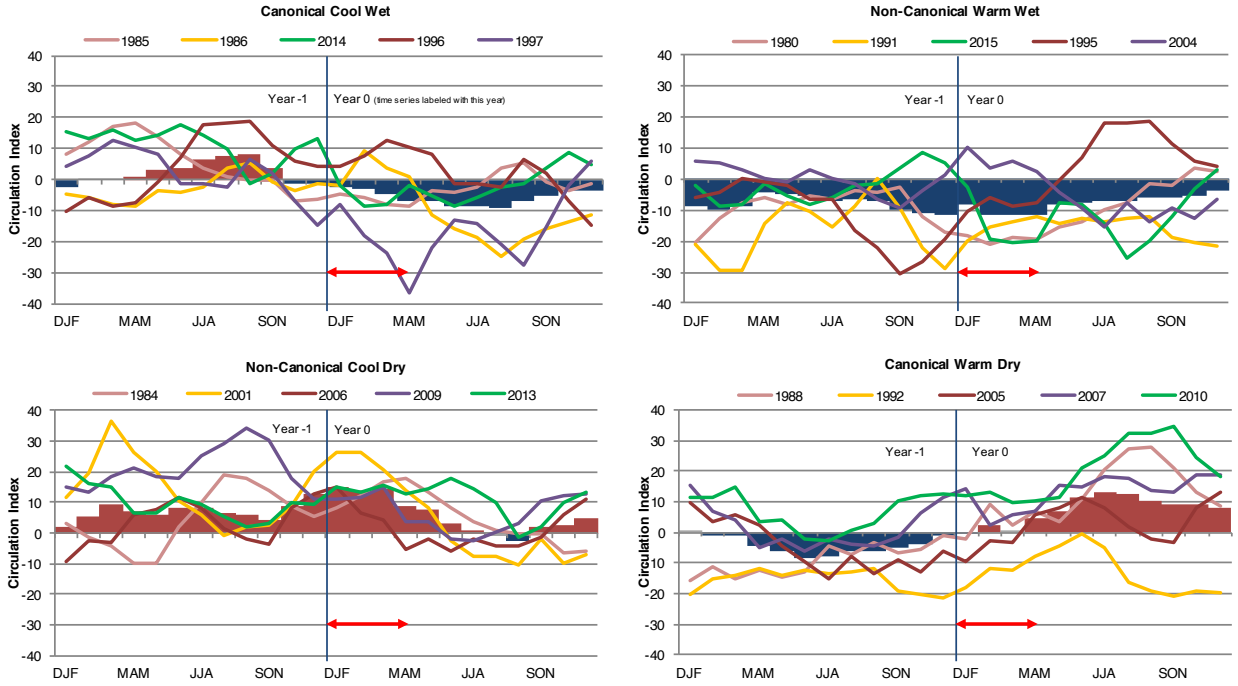


Figure 3.12: Time series of the northern USAPI (Kwajalein and Guam) Standardized Precipitation Index (SPI) averaged for the members of each composite categories (black lines). The SPI in this figure is calculated for consecutive 6 month seasons. Bars represent the composite time series of the average stream function of anomalous winds over the Tropical Western Pacific. Red arrows depict the extent of the climatological Dec-May dry season.

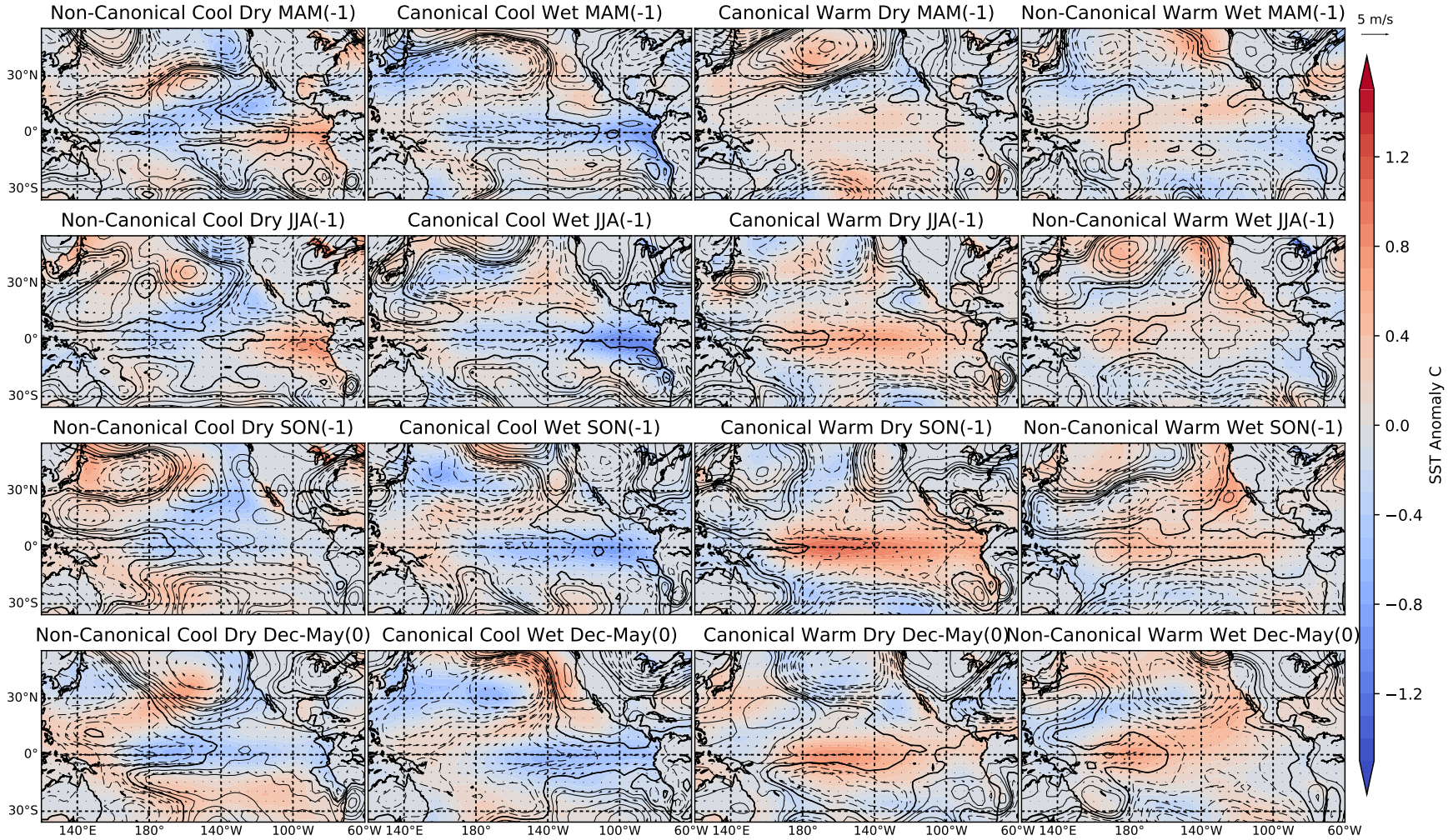


Figure 3.13: Seasonal composites of SST (shading), 850 hPa geopotential height (contours) and wind anomalies for the three 3-month seasons leading up to the December-May dry season, as well as the December-May dry season composite.

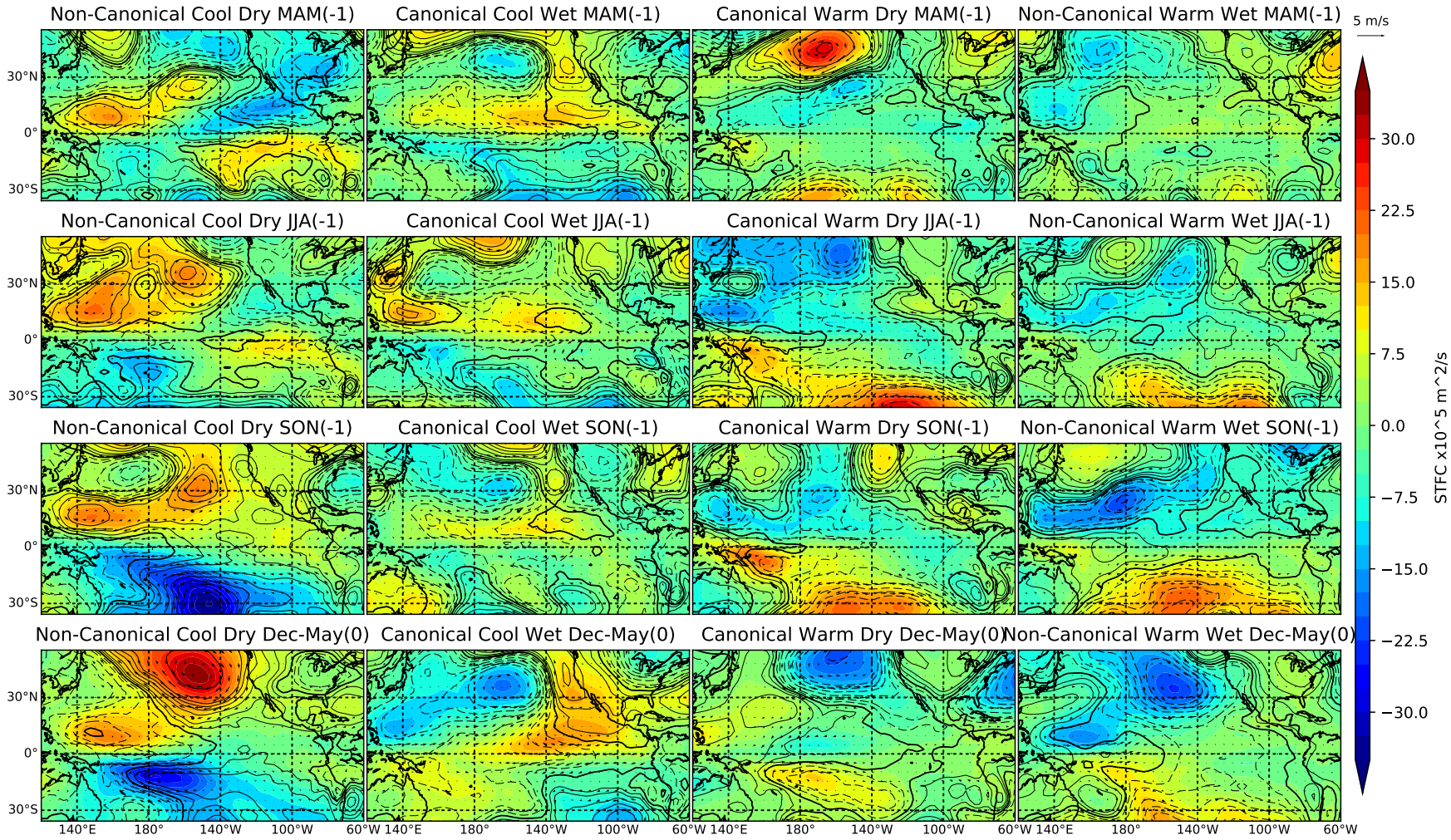


Figure 3.14: Seasonal composites of stream function (shading), 850 hPa geopotential height (contours) and wind anomalies for the three 3-month seasons leading up to the December-May dry season, as well as the December-May dry season composite.

could play an important role in the development and maintenance of these anomalies. While the link between subtropical and tropical circulation anomalies appears clear in the Cool Dry events, it is less so for their Warm Wet Non-Canonical counterparts, which show similar and opposite SST anomaly patterns but less clear circulation anomalies during the MAM and JJA seasons. For Canonical events, the off-equatorial SST and wind anomalies are weaker than for the Non-Canonical cases, with the exception of the JJA and Dec-May seasons of the Cool Wet event composites, which show an opposite SST and circulation pattern to the same seasons in the Cool Dry composite.

Anomalous SST along the equator for Canonical and Non-Canonical events also evolves differently from each other (Figure 3.15). During Canonical events, SST anomalies dominate the entire basin well ahead of the Dec-May dry season, extending from about 160°E to the South American coast. In the case of Cool Wet events, strong cool SST anomalies appear over the eastern Pacific off the South American coast during April and May of the developing year. As the event evolves, these precursor SST anomalies intensify and extend from east to west, peaking near 140°W (160°W for Warm Dry events).

Non-Canonical events are characterized by an evolution of equatorial SST that is quite different (Figure 3.15). In the late spring and early summer of the developing year, Non-Canonical years are characterized by an SST dipole pattern between the central and eastern Pacific. Cool Dry events show warm SST anomalies off the South American Coast and and cool SST anomalies over the central Pacific (bottom left panel of Figure 3.15). These cool central Pacific anomalies then extend eastwards, covering most of the region east of the dateline but with peak intensity strongly localized near the dateline. Thermocline anomalies mirror this difference in zonal propagation. Negative thermocline depth anomalies first appear off the South American Coast and propagate west during Canonical Cool Wet events, while during Non-Canonical Cool Dry events, they appear first over the Central Pacific and propagate east.

Wet and Dry La Niña events also exhibit this different evolution of equatorial SST anomalies (Figure 3.16). For wet La Nina events, as in Cool Wet events, strong negative SST and thermocline anomalies are observed off the South American coast as early as May leading up to the dry season. These negative SST and thermocline anomalies are then seen propagating westward. Dry La Niñas on the other hand, like Cool Dry events show warm SST anomalies off the South American coast. The propagation of cool SST anomalies from the dateline toward the east is very clear in Cool Dry events, but during Dry La Niña events this signal is much weaker, with negative SST anomalies appearing to first develop around 140°W and

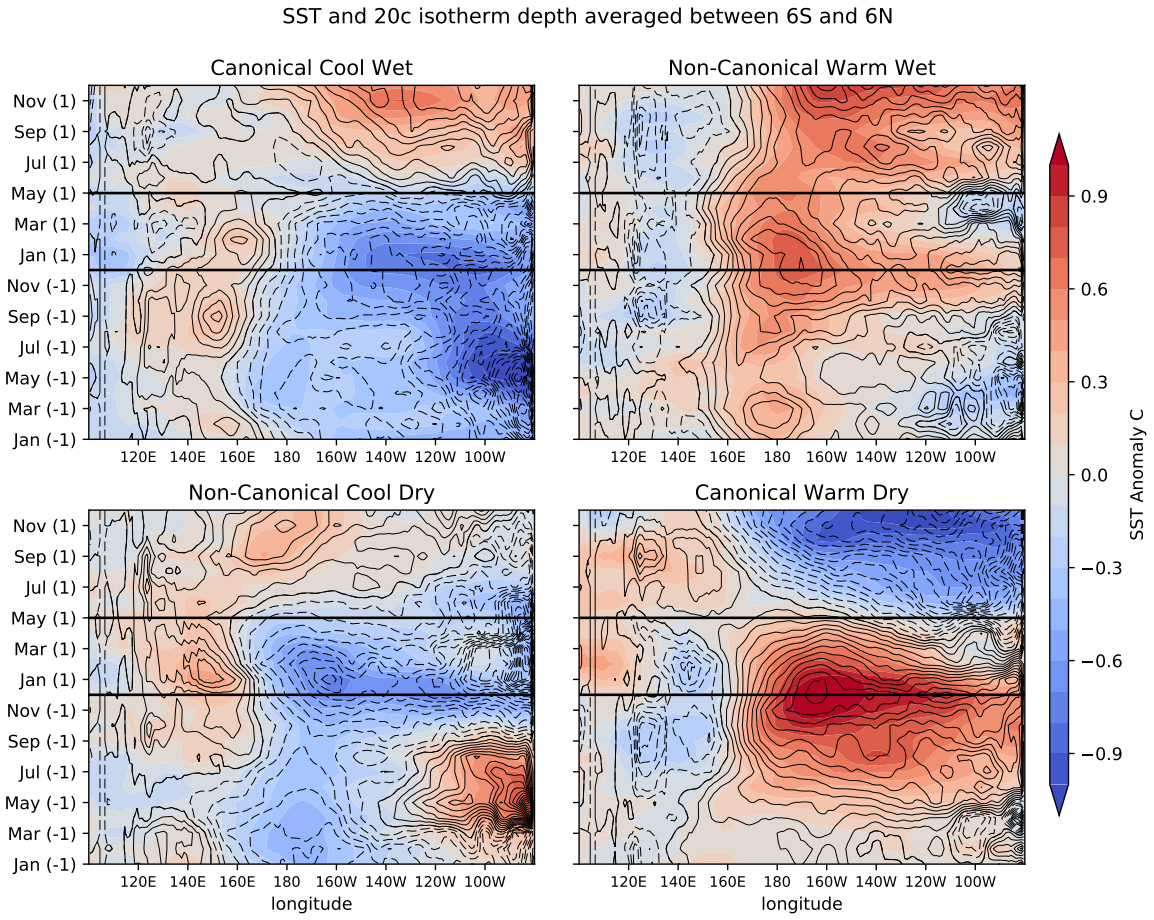


Figure 3.15: Hovmöller diagrams of composite monthly SST (shading) and GOADS 20 Degree Celsius depth (contours) anomalies averaged from 6°S to 6°N. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season.

propagate slightly westward.

Figure 3.17 shows Hovmöller diagrams of monthly rainfall anomaly composites averaged between 5°N and 15°N along with the stream function of the anomalous winds averaged from the equator to 30°N for each composite category. Over the western Pacific and the USAPI region (130°W to the dateline), monthly rainfall shows behavior consistent with what was described before using the time series of SPI. Canonical events show rainfall anomalies that change sign as the event progresses and Non-Canonical events show rainfall anomalies that persist.

Canonical events are also characterized by an eastward propagation of rainfall and circulation anomalies from the western Pacific towards the east (top left and bottom right

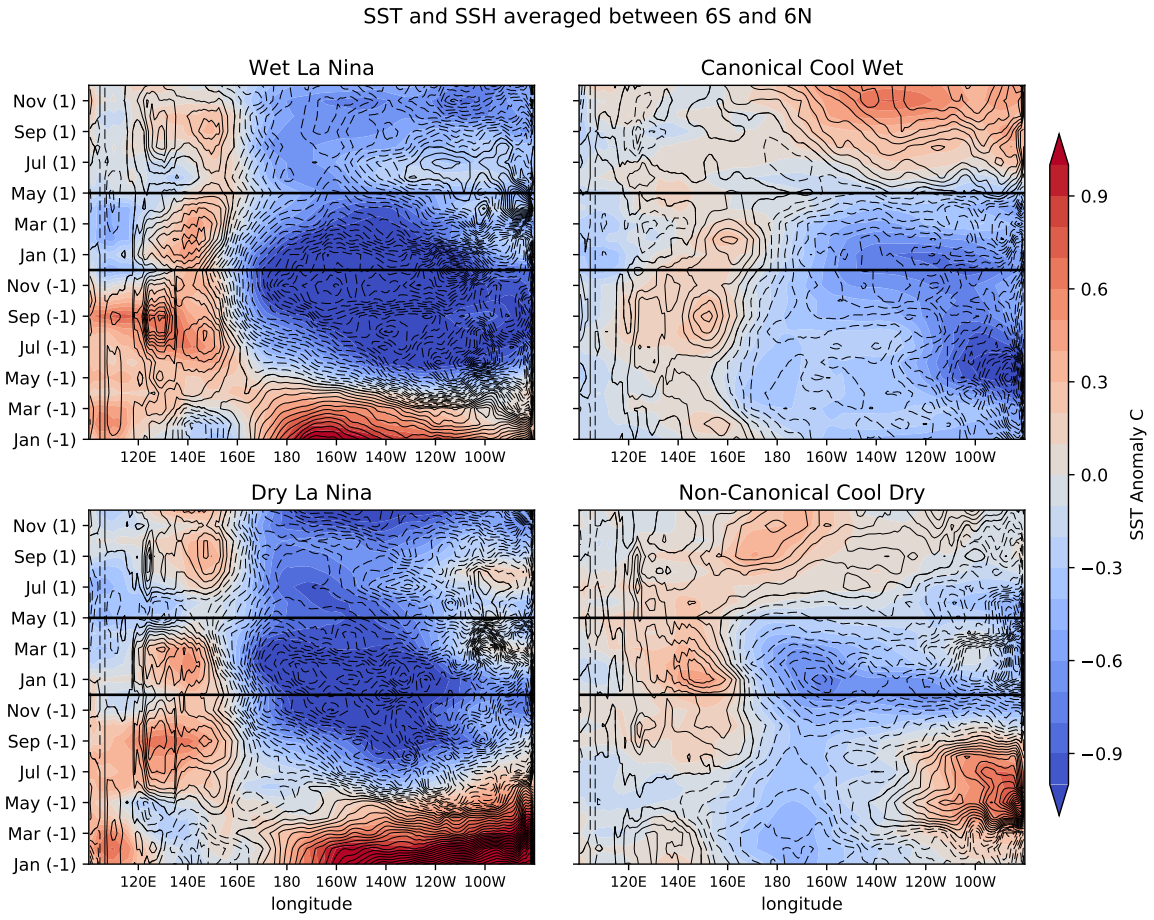


Figure 3.16: Hovmöller diagrams of composite monthly SST (shading) and GOADS 20°Celsius depth (contours) anomalies averaged from 6°S to 6°N for La Niña events and Canonical Cool Wet and Non-Canonical Cool Dry events. Black horizontal lines represent December and May which are the start and end of the northern USAPI dry season.

panels of Figure 3.17). During Cool Wet events, negative rainfall and anticyclonic circulation anomalies first appear over the western Pacific as early as March and propagate east, being confined east of 160°W by the start of the Dec-May dry season. This progression is consistent with previous observations (Wang and Zhang (2002)).

Eastward propagation of rainfall and circulation anomalies is not observed in the Non-Canonical composites (bottom left and top right panels of Figure 3.17). During Non-Canonical years, homogeneous rainfall and circulation anomalies occupy most of the Pacific basin and persist from the developing year of the events through the dry season of the target year. It is seen that Non-Canonical events are characterized by the persistence of anomalies,

not only over the tropical western Pacific, but also across a large part of the central Pacific, from about 140°E to 120°W.

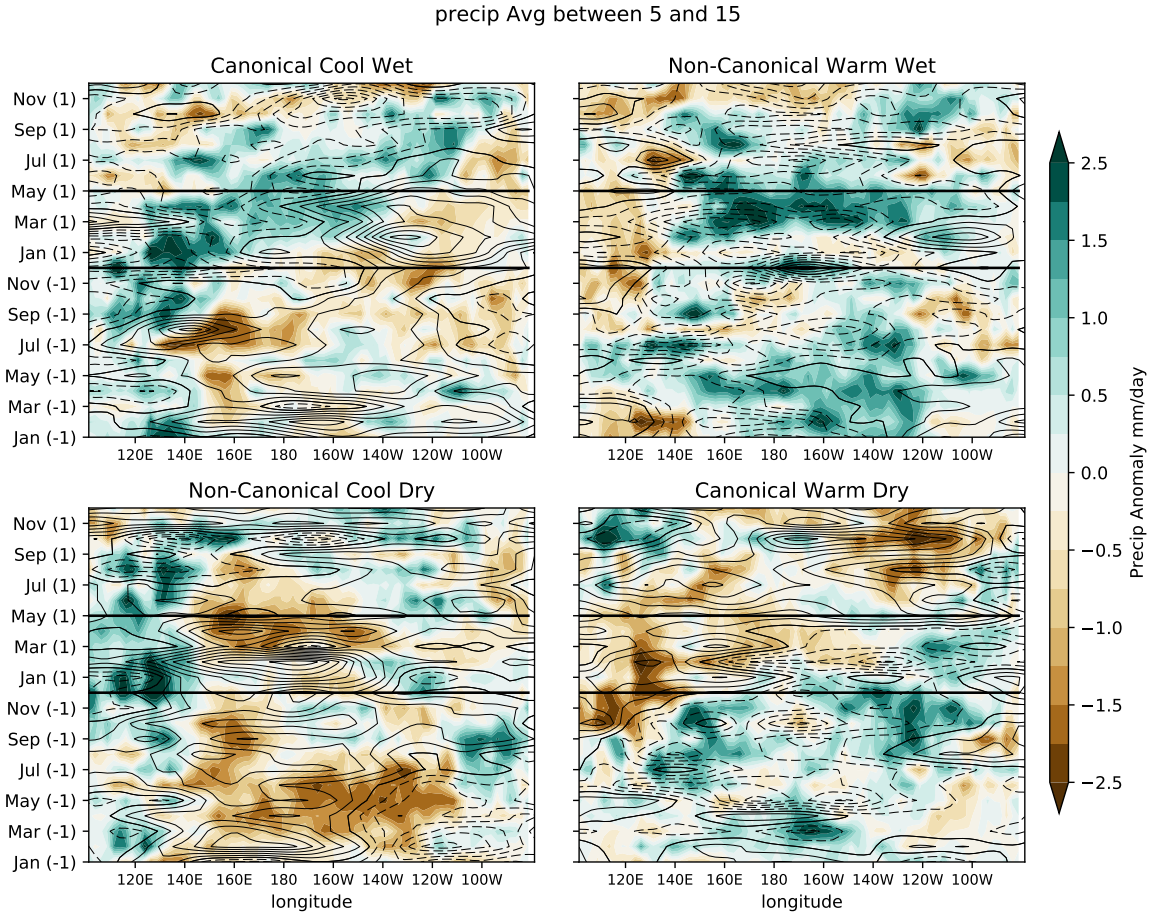


Figure 3.17: Hovmöller diagrams of composite monthly rainfall anomalies averaged from 5°N to 15°N (shading) and the stream function of the anomalous winds averaged between the equator and 30°N (contours). Contours range between $+/-3^6$ in intervals of 3^5 m^2/s. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season.

Wet and dry La Niña events are also characterized by propagation vs. persistence of rainfall anomalies over the western Pacific in a way similar to Canonical Cool Wet and Non-Canonical Cool Dry events (Figure 3.18). Wet La Niña events show eastward propagation of rainfall anomalies that is less pronounced than in the Cool Wet events. Similarly, Dry La Niña events show some propagation of rainfall anomalies toward the west, instead of no propagation as is seen in Cool Dry events.

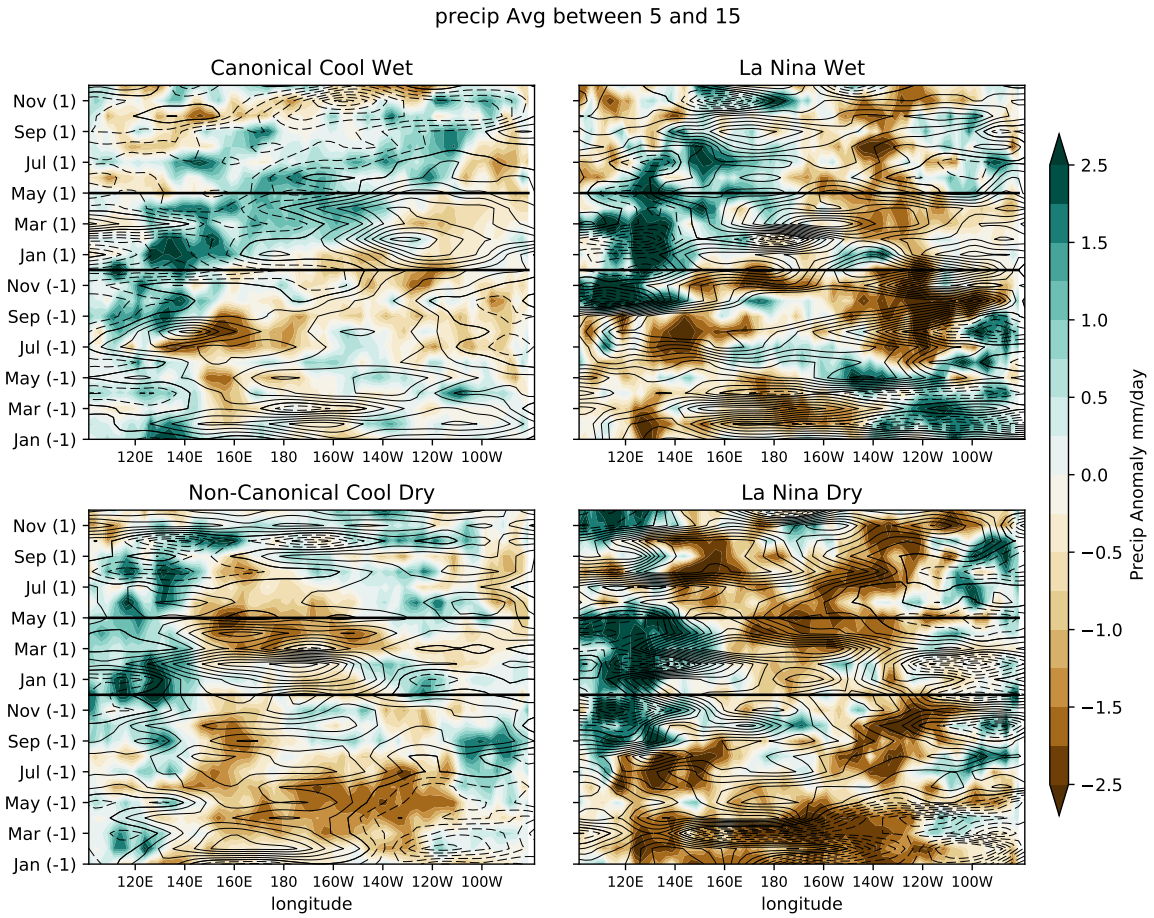


Figure 3.18: Hovmöller diagrams of composite monthly rainfall anomalies averaged from 5°N to 15°N (shading) and the stream function of the anomalous winds averaged between the equator and 30°N (contours) for Wet and Dry La Niña events and Cool Wet and Cool Dry events (see also Figure 3.17). Contours range between $+/-3^6$ in intervals of 3^5 m^2/s. Black horizontal lines represent December and May, which are the start and end of the northern USAPI dry season.

To understand the evolution of mid-latitude anomalies and their possible role in the support of tropical anomalies during Non-Canonical events, Figure 3.19 represents the evolution of SST anomalies between 20° and 30°N and, in contours, the 850 hPa geopotential height anomalies averaged from 30° to 50°N. Over the northern Pacific, Non-Canonical events show a strong relationship between geopotential height anomalies north of 30°N and SST anomalies south of 30°N. Pressure anomalies over the northern Pacific and associated anomalous circulation coincide with strong SST anomalies. In the case of Cool Dry events, high pres-

sure anomalies to the north are followed by cool SST anomalies east of about 160°E and warm SST anomalies to the west. For Canonical events, Cool Wet events show a similar but opposite behavior over the northern pacific as Cool Dry events, but Warm Dry events show weak mid-latitude anomalies both in geopotential height and SST, suggesting that Warm Dry events have less of a mid latitude influence than their Cool Wet Canonical counterparts and Non-Canonical events.

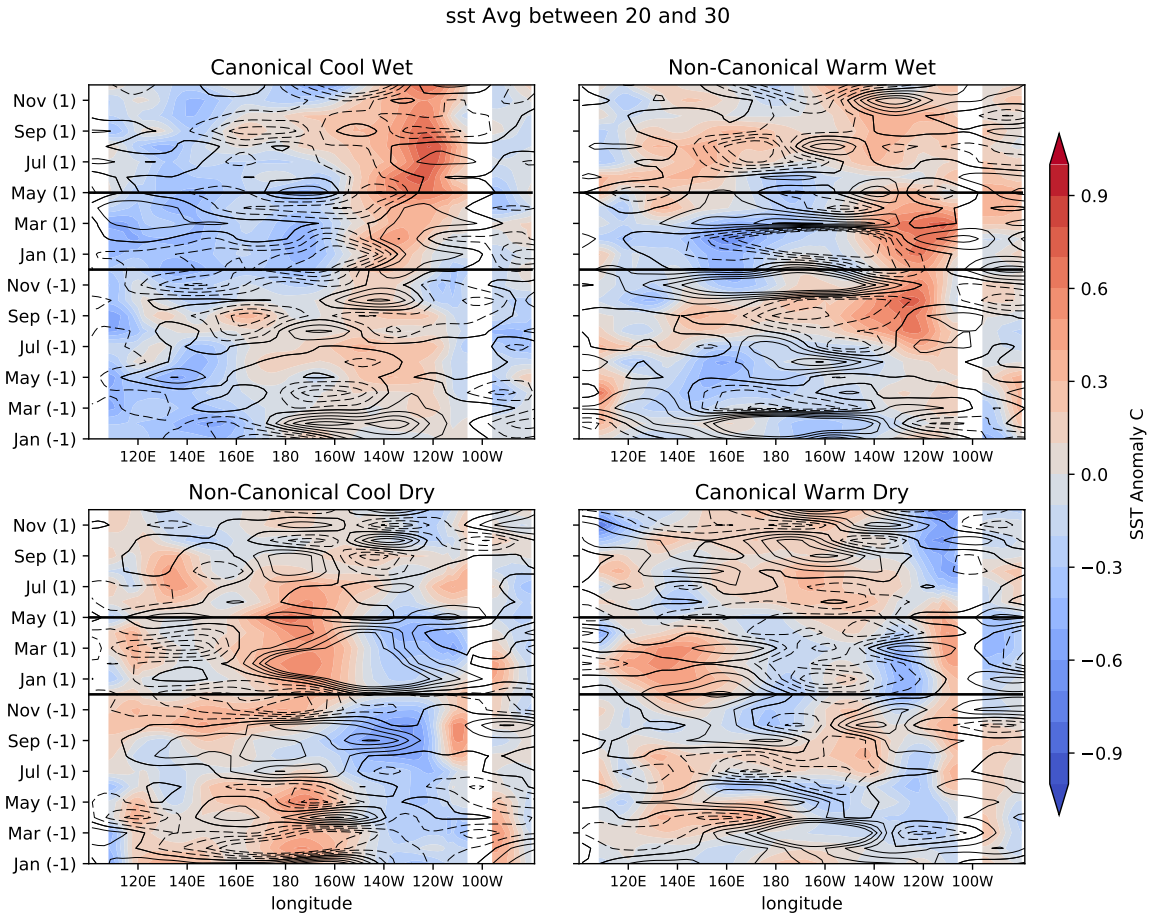


Figure 3.19: As in Figure 3.17, but for monthly SST anomalies averaged from 20° and 30°N (shading) and 850 hPa geopotential height anomalies from 30° to 50°N (contours).

The differences between peak dry season SST maxima coupled with their differences in evolution for Canonical and Non-Canonical events described here are consistent with some previous results. Modeling studies have shown that Eastern Pacific and Central Pacific El Niño events have peak SST anomaly patterns that evolve from distinct initial condition patterns (Newman et al. (2011a), Newman et al. (2011b)). Precursor conditions in those

studies were derived using observed SST, thermocline and surface wind stress anomalies. The precursor conditions derived in those studies for Eastern Pacific El Niño events are similar to the precursor season composite maps observed here for Canonical Warm Dry events, with SST anomalies first appearing off the South American coast and extending west as the event matures (Newman et al. (2011b)). Likewise, the precursor conditions derived for Central Pacific El Niño events are remisniscent of the composites produced here for Non-Canonical Warm Dry events. In these cases, an SST dipole is observed between the central and eastern Pacific (Newman et al. (2011b)).

Observations of the seasonal evolution of anomalies of Eastern and Central Pacific El Niño events also show that their seasonal evolution of SST anomalies are different (Kao and Yu (2009), Yeh et al. (2014)). Eastern Pacific El Niño events show SST anomalies that appear off the coast of South America during the spring and extend west as the event matures, just as is observed for Canonical Warm Dry events. On the other hand, similar to our Non-Canonical events, Central Pacific El Niño events show SST anomalies that extend from the eastern subtropical Pacific towards the central Pacific during spring and summer.

S-EOF analysis is performed on GPCP gridded rainfall anomalies for the JJA, SON, DJF and MAM seasons over the northern USAPI region defined between 5° to 20°N and 140°E and the dateline (Figure 3.20). This type of seasonal analysis is capable of separating modes of variability that evolve differently from season to season, even if during one particular season these modes show similar patterns. In our case, years in the Canonical Warm Dry and Non-Canonical Cool Dry composites have similar rainfall anomalies over the USAPI region. However, the seasonal evolution of these rainfall anomalies is different (as revealed by previous analysis). S-EOF analysis is particularly suited for this scenario and allows for the separation of Canonical and Non-Canonical variability into orthogonal S-EOF modes. Regressions of each PC time series to rainfall, SST, winds and geopotential height are used to capture the large scale spatial features of these S-EOF modes across the Pacific basin.

The first two S-EOF modes of precipitation anomalies over the northern USAPI capture the Non-Canonical and Canonical modes (respectively) of variability previously described using composite analysis. These modes account for 17.5% and 12.8% of the total precipitation variance in the region. The PC time series of these first two S-EOF modes are given in Figure 3.20.

The first S-EOF mode captures, as in the previous composite analysis, the persistence of circulation anomalies over the western tropical Pacific from the JJA to the MAM season, causing the continuous rainfall anomalies in the northern USAPI that characterize the Non-

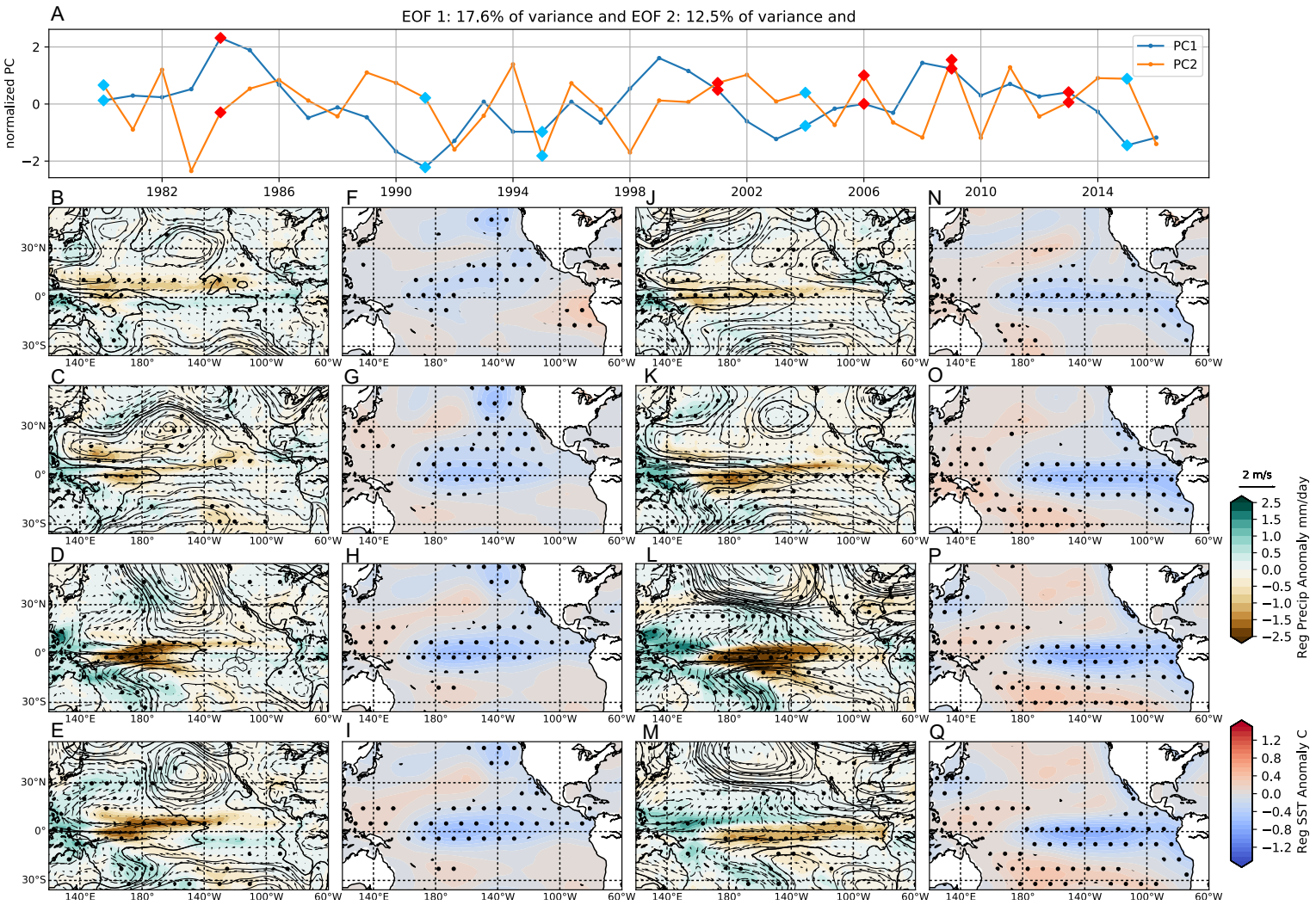


Figure 3.20: Seasonal EOF of rainfall in the region from 5° to 20°N and 140°E to the dateline. Panel A shows the PC time series for the first two S-EOF modes. Triangular markers represent the years chosen for the Non-Canonical composites, with red representing Cool Dry years and blue Warm Wet years. Panels B, C, D and E show the regression of the PC1 to rainfall, geopotential height and wind anomalies during the JJA, SON, DJF and MAM seasons, respectively. Panels F, G, H and I show the regression of PC1 to SST anomalies for the JJA, SON, DJF and MAM seasons. Panels J, K, L and M show the same as B through E but for PC2, while N trough Q are analogous to F through I but also for PC2. Circular hatching represents areas for which correlations are significant at the 90% confidence level.

Canonical events. Linked to these circulation anomalies, the first S-EOF mode also captures the persistent rainfall anomalies that extend eastward along 10°N across a large portion of the Pacific basin observed in Non-Canonical event composites. This suggests that this mode captures the meridional modulation of the ITCZ activity by atmospheric and ocean anomalies. The PC time series of the S-EOF 1 mode has positive values for 4 out of 5 of the years chosen by our previous methodology to be in our Non-Canonical Cool Dry composite, and negative values for 4 out of 5 of the years in our Warm Wet composite.

The second S-EOF mode encompasses many of the hallmark features of the development, peak and decay of ENSO events. Circulation anomalies in this mode shift from cyclonic to anticyclonic over the tropical western north Pacific as the events develop, and SST anomalies develop close to the equator and intensify as the events progress. The PC time series of S-EOF 2 has strong positive values for the Eastern Pacific El Niño events of 1983, 1998 and 2016, as well as positive values for all of the years chosen to be members of the Canonical Warm Dry composite, and 4 out of 5 member years of the Cool Wet composite.

3.5 Discussion

A schematic diagram of anomaly differences for Canonical and Non-Canonical events during the peak dry season from December to May illustrate the principal differences between the two categories (Figure 3.21). We contend that Canonical events represent current understanding of the ENSO events. Rainfall anomalies in Canonical events are consistent with what is generally described as the ENSO horseshoe rainfall pattern over the Pacific, with opposite anomalies in the central and eastern Pacific and the westernmost part of the basin. Canonical events are characterized by having strongest SST anomalies east of the dateline and opposite wind anomalies on either side of the dateline.

Non-Canonical events, on the other hand, show a modified horseshoe rainfall anomaly pattern characterized by a belt of rainfall anomalies that extends from the central Pacific toward Micronesia. This rainfall belt is driven by western Pacific circulation anomalies of opposite sign than those observed for Canonical events of similar ONI sign and intensity. Non-Canonical events have the strongest SST anomalies over the central Pacific near the dateline, with weak SST anomalies to the east and opposite signed anomalies to the west, as well as having the strongest equatorial wind anomalies in the western Pacific.

The difference in evolution of anomalies leading up to the December to May dry season suggests different physical mechanisms behind the Canonical and Non-Canonical modes of

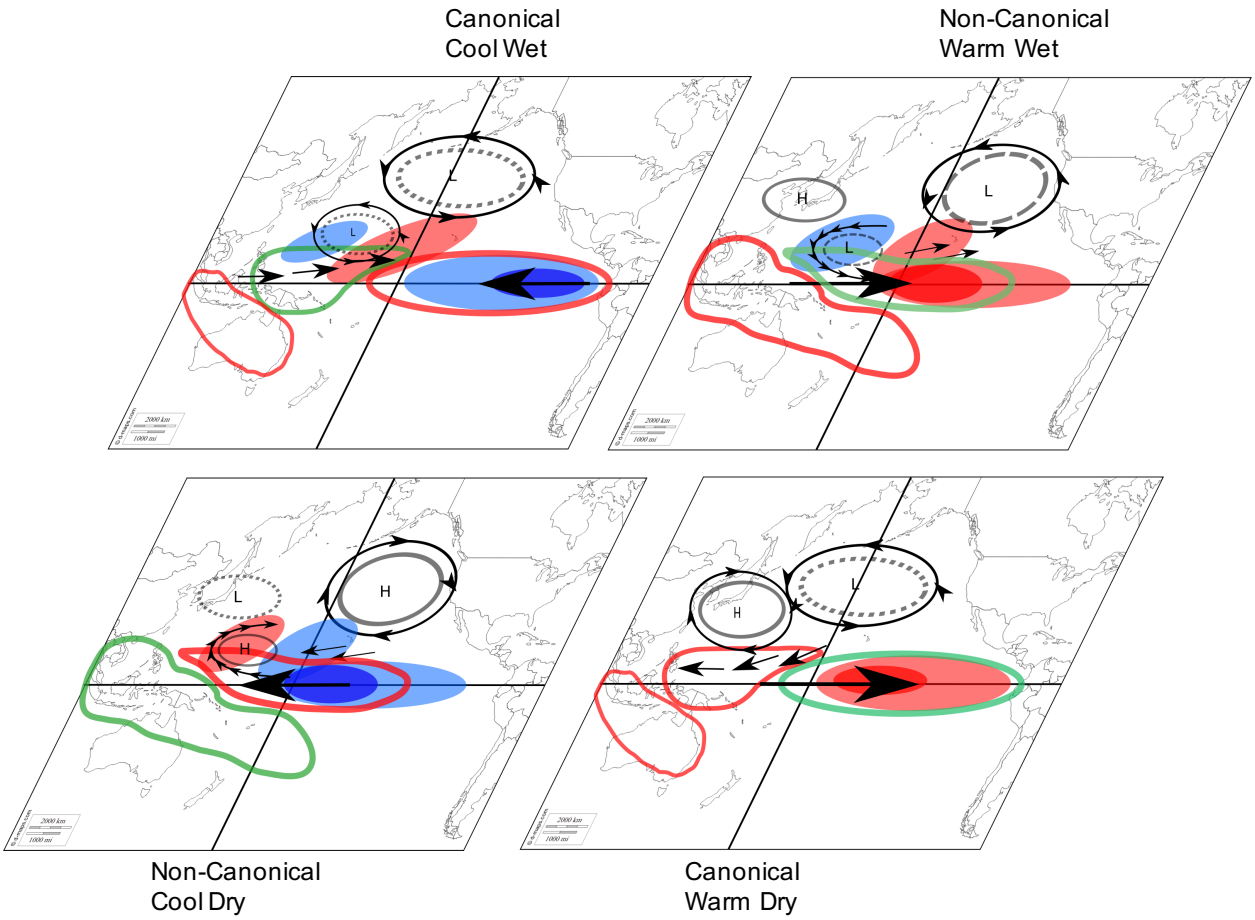


Figure 3.21: Schematic diagram of relevant December to May ocean surface temperature and atmospheric circulation anomalies that lead to difference in rainfall conditions over the USAPI region. Rainfall anomalies are represented as contours, with red for dry and green for wet anomalies, SST anomalies are represented as filled shapes and geopotential height anomalies as grey contours, solid for positive and dashed for negative. Winds are represented as black arrows, with equatorial wind anomalies represented in bold black arrows.

variability. During Canonical Cool Wet events for example, anticyclonic circulation anomalies over the western Pacific are observed in the 6 months previous to the Dec-May dry season. The southern branch of this cyclonic pattern are the anomalous easterly winds near the equator and west of the dateline that trigger the development via Kelvin Waves of cool SST anomalies over the eastern Pacific. Evidence for this is the observed thermocline deepening anomalies that appear off the South American Coast preceding the SST anomalies in the same region. As eastern Pacific SST anomalies intensify, the anticyclonic circulation over the western Pacific migrates eastward, and allows for cyclonic circulation anomalies to de-

velop over the Western Pacific. This eastward migration of anticyclonic circulation anomalies produces the shift in rainfall anomalies, from drier than normal during the months leading up to the Dec-May dry season, to wetter than normal during the peak and post dry season. Canonical Warm Wet events show a similar progression of events with opposite sign.

Non-Canonical Cool Dry events also show anticyclonic circulation anomalies over the tropical western Pacific in the months leading up to the Dec-May dry season, but these fail to trigger the development of more intense SST anomalies over the eastern Pacific. Anticyclonic circulation anomalies hence do not migrate eastward during these events and remain anchored over the western Pacific, with regional air-sea interaction likely playing a significant role in their maintenance (as suggested by the complementary circulation and SST anomaly patterns observed). The persistent anticyclonic circulation over Micronesia and resulting persistent dry rainfall anomalies in the northern USAPI region then lead to the severe droughts observed during these types of events. Because Non-Canonical Warm Wet events show similar but opposite dry season anomalies as well as anomaly evolution, it seems appropriate to think of these events as the two sides of a single mode of variability.

Strong La Niña events can also produce dry or wet conditions over the northern USAPI region. Rainfall anomalies (both wet and dry) during strong La Niña events are of lesser magnitude than those observed for the Cool Dry and Cool Wet years. Stronger equatorial easterlies during strong La Niña events prevent local circulation anomalies over Micronesia from developing. During dry La Niña events, equatorial easterlies push farther west past Micronesia toward the maritime continent, and stronger high pressure anomalies over the Northern Pacific prevent the formation of a strong anomalous anticyclonic circulation over Micronesia, producing weaker dry conditions than those observed during Cool Dry events. During wet La Niña events, strong easterlies push anomalous westerlies over Micronesia north, extending equatorial dry conditions northward and weakening wet anomalies over Micronesia. The difference in the location of maximum SST anomalies between wet and dry La Niña is not as large as observed between Cool Dry and Cool Wet events. Dry La Niña events show SST anomalies slightly shifted toward the dateline compared to wet La Niña, however, they are not located as far west as in Cool Dry events. For this reason, we observe some migration of rainfall anomalies from east to west during dry La Niña events. It appears that La Niña events follow similar dynamics as Cool Wet and Cool Dry events, but stronger SST anomalies with less difference in their location result in events which show less distinct characteristics.

The stark separation between Canonical and Non-Canonical events used in this study

is valuable to the PEAC Center because these events show such distinct impacts on the Center's area of responsibility, although it may not be adequate for the ENSO community at large. As pointed out in the previous sections, many of the most distinct features of Non-Canonical events have not been analyzed in detail. Nonetheless, some of the results presented here are consistent with previous findings from ENSO diversity studies. Canonical Warm Dry events show peak dry season anomalies and seasonal evolution of those anomalies consistent with some previous observation and modeling results found for Eastern Pacific El Niño events. Similarly, our results for Non-Canonical events show some overlap with previous results obtained for Central Pacific El Niño events. (See Capotondi et al. (2015), Newman et al. (2011a), Newman et al. (2011b), Kao and Yu (2009), Yeh et al. (2014), Wang and Zhang (2002), Ropelewski and Halpert (1987), and Stuecker et al. (2015)). It may be that Non-Canonical events can be incorporated into the larger landscape as being thought to represent the most extreme of Central Pacific ENSO events. In these extreme Central Pacific events, SST anomalies are so strongly constrained to the central Pacific that the evolution of western Pacific circulation, and subsequently, rainfall, is dramatically different from that of events in which SST anomalies peak farther east. This is not unlike the fact that the most extreme Eastern Pacific El Niño events, such as 1983-83, 1997-98 and 2015-16, share their own set of distinguishing features that separate them from events with peak SST anomalies farther to the west.

The event selection methodology used in this work allows for a framework to examine the causes of dry season rainfall anomalies in the northern USAPI region outside traditional ENSO events. It has the virtue of highlighting years that do not show the expected impacts of ENSO over the region, such as 1995 which produced wetter than normal conditions over the northern USAPI, or would be overlooked by selection criteria based only on SST, such as 2013 whose drought impacts were described in the introduction.

The analysis is limited by the small sample size of 5 members per composite category out of 36 years. This makes detailed analysis of the finer differences in our composites contentious, but the most important features highlighted in our work are considered statistically significant. While the analysis shown here is suggestive of differences in physical mechanisms acting during Canonical and Non-Canonical events, future work should be aimed at verifying the physical mechanisms proposed here.

CHAPTER 4
FORECASTING

In this chapter we will develop forecasting methodology better suited to handle the diverse impacts of Canonical and Non-Canonical events on rainfall anomalies in the Northern USAPIs. Our forecasting target will be the Standardized Precipitation Index (SPI) in the Northern USAPIs during the December-May dry season and our methods will be based on the Physical-Empirical forecast methodology (Yim et al. (2014b), Yun et al. (2015)). In section 4.1, we will establish a simple benchmark model to predict Dec-May SPI which we will refer to as a 1-tier model. In Section 4.2, we will develop a 2-tier prediction methodology that will rely first on predicting if an event will be Canonical or Non-Canonical, and then predict their different impacts on rainfall in the Northern USAPIs.

Data used in this section is extended to the periods of 1950 to present. We will separate our data into a training period from 1950 to 2007 and an independent test period, on which we do not train the model, spanning from 2008 to 2017. This selection of test period is made because it coincides with the span of the PEAC Center operational forecast database with which we wish to compare the performance of our forecasting scheme.

As described in the methodology section, a Physical-Empirical forecasting methodology requires us to choose predictors that not only show statistical correlation with the target variable, but can also be understood in a physically meaningful context. The first step is to identify possible predictors from the lower level boundary conditions that are most likely to affect the evolution of relevant climate fields. In our case we will use Sea Surface Temperature (SST) and 850 hPa geopotential height as possible predictor fields. We will also use 850 hPa winds and stream function of the anomalous winds. Although wind fields are not lower boundary conditions and therefore tend to make less powerful predictors, we include them in the predictor search because their evolution differs markedly between Canonical and Non-Canonical events. We will construct a 0 lead forecasting scheme using the anomaly fields in the June-July-August and September-October-November seasons prior to the target December-May dry season. We will also use anomaly tendency fields, defined as the October-November anomaly minus the August-September anomaly. All predictive indexes will be normalized by subtracting the mean and dividing by the standard deviation of the index.

4.1 Simple Linear Regression Model

As a benchmark, we will establish a forecasting model that forecasts Dec-May SPI under all conditions (Canonical or Non-Canonical) using the observed anomaly fields in the preceding seasons. To establish predictor candidates, the correlation maps of Dec-May SPI with preceding season anomaly fields were calculated. We begin with a large set of predictors that may be physically meaningful and then use stepwise regression to select a small subset.

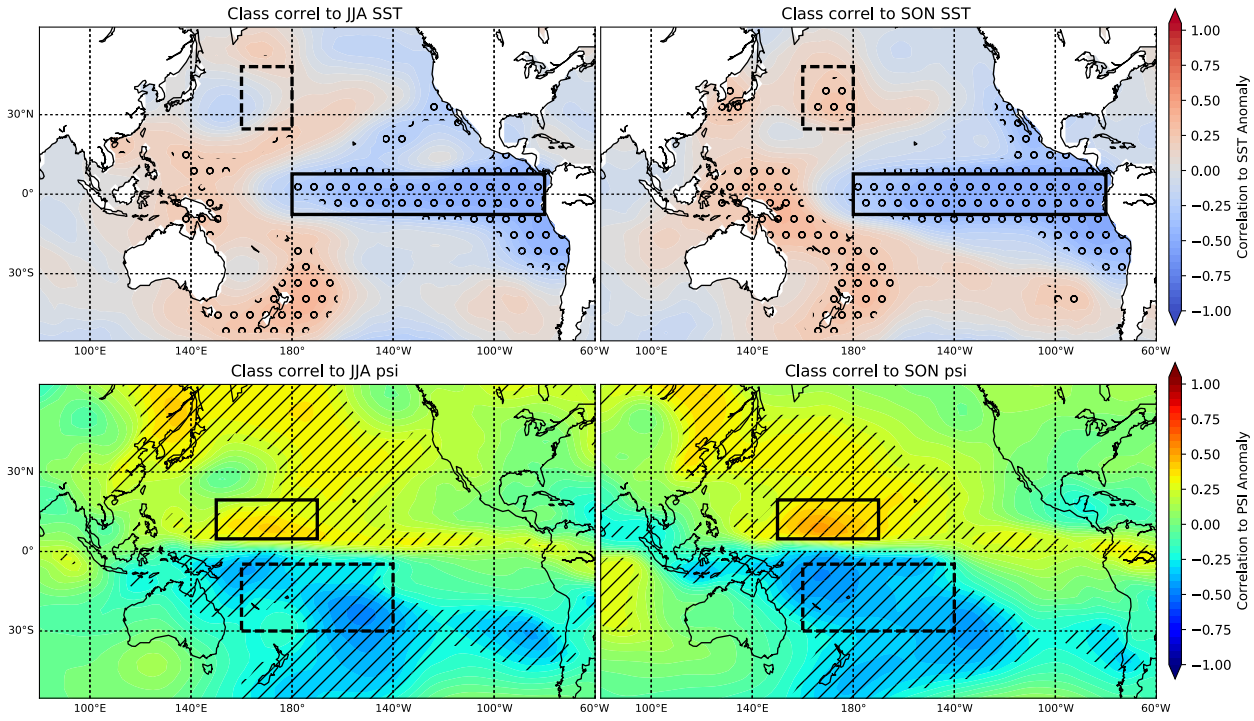


Figure 4.1: Correlation coefficient maps of Dec-May SPI and SST (top) and Stream Function (bottom) of the anomalous wind fields (PSI) for the June-July-August (JJA) and September-October-November (SON) seasons. Hatching represents significance at the 95% confidence level using a Student t-test with a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

The correlation maps of Dec-May SPI to preceding SST and zonal wind component anomalies for the preceding JJA and SON seasons are given in Figure 4.1. In the tropical Pacific, the correlation to SST maps show a clear ENSO 'horseshoe' pattern with negative correlation in the cold tongue region and the central Pacific east of the dateline, and positive correlation to the west. This pattern captures the Canonical relationship between ENSO and Northern USAPI Dec-May SPI, with El Niño producing negative SPI anomalies and La

Niña positive SPI anomalies. As a predictor for the impact of ENSO on Northern USAPI rainfall, the area averaged SST anomaly in the region from the dateline to 80°W is defined as the Central Pacific SST Index (4.1, solid black rectangle in top panels).

A similar procedure will be done for all other fields, constructing predictive indexes as area averages of the anomaly maps where high correlation is observed. Table 4.1 shows the name, region and a short description of all predictor candidates developed for this benchmark model.

In the subtropical northwestern Pacific, an area of high correlation between SST anomaly and Dec-May SPI is also observed. This feature may form part of the ENSO SST anomaly pattern, but it may also capture local ocean-atmosphere interaction between the regional circulation anomaly in the western Pacific and the upper ocean. The development of this western Pacific circulation also plays an important role in the modulation of rainfall anomalies in the Northern USAPIs. The dashed box in the top panels of Figure 4.1 represents the area used for the anomaly index representative of this feature.

The bottom panels of Figure 4.1 represent the correlation maps of Dec-May SPI to JJA and SON stream function of the anomalous wind field. In both the JJA and SON seasons, the western Pacific shows areas with high correlation between Dec-May SPI and stream function north and south of the Equator. This correlation pattern is dominated by the Canonical correlation between Northern USAPI SPI and ENSO. Circulation anomalies in this region are cyclonic during the onset of El Niño and anticyclonic during the onset of La Niña. North of the equator, positive stream function is collocated with anticyclonic circulation, so during the onset of a Canonical Cool Wet event, the western Pacific is dominated by anticyclonic circulation (positive stream function) that then shifts to cyclonic during the dry season, producing above average rainfall and positive SPI values. This relationship results in correlations between Dec-May SPI and stream function in the western Pacific in the previous season (the correlation is negative in the south Pacific because the relationship between stream function and cyclonic/anticyclonic circulation reverses). From this correlation map we define two predictor candidates, the West Pacific Stream Function Index (solid box in bottom panels of Figure 4.1, 5°N to 15°N, 150°E to 170°W) and the South Pacific Stream Function Index (dashed box in bottom panels of Figure 4.1, 5°S to 30°S, 160°E to 140°W).

Figure 4.2 shows the correlation patterns of Dec-May SPI to 850 hPa zonal wind and geopotential height. The top panels in Figure 4.2 show strong correlation between Dec-May SPI and zonal winds along the equatorial belt in the central Pacific. These equatorial wind anomalies are a crucial part of the onset of ENSO events and thus should be considered as a

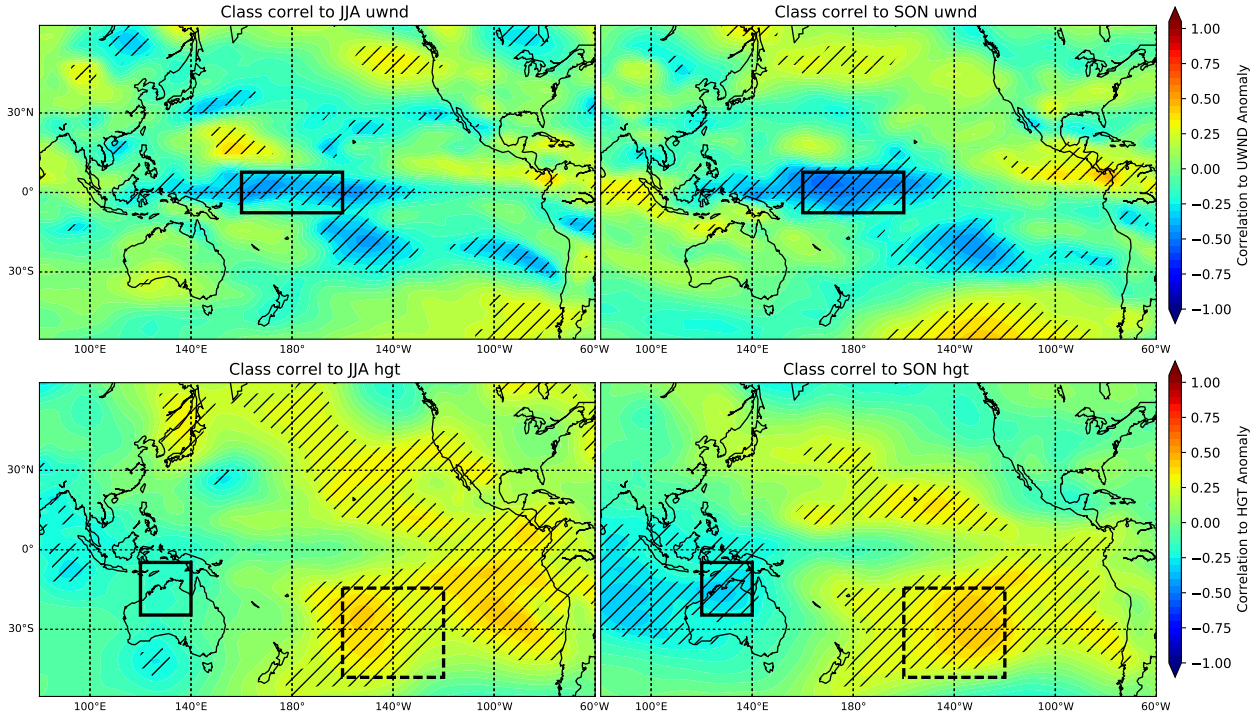


Figure 4.2: Correlation coefficient maps of Dec-May SPI and 850 hPa zonal wind (top) and geopotential height (bottom) for the June-July-August (JJA) and September-October-November (SON) seasons. Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

1-tier Model Predictors			
Predictor Name	Variable	Region	Description
Central Pacific SST Index	SST Fig 4.1	Dateline to 80°W, 6°S to 6°N	The SST anomalies in this region capture the intensity of ENSO (Bamston et al. (1997)).
Subtropical northwestern Pacific SST Index	SST Fig. 4.1	160°E to the Dateline, 25°N to 45°N	SST anomalies in this region likely to reflect the Ocean-Atmosphere interaction between circulation in the western Pacific and underlying ocean (Wang and Zhang (2002)).
West Pacific Stream Function Index	Stream Function Fig 4.1	5°N to 15°N, 150°E to 170°W	Captures the anomalous circulation in the western Pacific linked to the evolution of ENSO (Wang and Zhang (2002), Li et al. (2016))
South Pacific Stream Function Index	Stream Function Fig 4.1	5°S to 30°S, 160°E to 140°W	Captures the anomalous circulation in the south Pacific linked to the evolution of ENSO (Trenberth (1997))
Equatorial Wind Index	850 hPa Zonal Wind Fig 4.2	5°S to 5°N, 160°E to 160°W	Captures the anomalous winds in the western Pacific linked to the onset of ENSO events (Yu et al. (2003))
South Pacific Pressure Gradient Index	850 hPa Geopotential Height Fig 4.2	5°S to 25°S, 120°E to 140°E minus 15°S to 45°S, 120°W to 160°W	Captures the Southern Oscillation, atmospheric component of ENSO (Trenberth (1997)).
Central Pacific SST Tendency Index	SST Tendency Fig 4.3	5°S to 5°N, from the Dateline to 120°W	The SST anomalies in this region capture the development of ENSO (Bamston et al. (1997)).

West Pacific Stream Function Tendency Index	Stream Function Tendency Fig 4.3	5°N to 30°N, 150°E to 140°W	Captures the intensification of anomalous circulation in the western Pacific linked to the evolution of ENSO (Wang and Zhang (2002))
South Pacific Stream Function Tendency Index	Stream Function Tendency Fig 4.3	The Equator to 30°S, 160°E to 160°W	Captures the intensification of anomalous circulation in the western Pacific linked to the evolution of ENSO (Trenberth (1997))
Equatorial Wind Tendency Index	850 hPa Zonal Wind Fig 4.2	5°S to 5°N, 160°E to 120°W	Captures the intensification of anomalous winds in the western Pacific linked to the onset of ENSO events (Yu et al. (2003))
Eastern Pacific Geopotential Height Tendency Index	850 hPa Geopotential Height Fig 4.2	15°N to 25°N and 110°W to 160°W	Captures the migration of circulation anomalies from the western Pacific to the east as ENSO events develop

Table 4.1: Description of all predictors developed for the 1-tier Northern Station Dec-May SPI forecast model. Includes predictor name, variable used, region and a short description of the physical mechanism it represents.

predictive feature. An Equatorial Wind Index is defined as the zonal wind anomalies from 6°S to 6°N and from 160°E to 160°W, where correlations are highest. For 850 hPa geopotential height, areas of high correlation are prominent in the southern Pacific. A see-saw pattern between the western Indian Ocean and the south Pacific near 140°W clearly captures the Southern Oscillation usually represented as the surface pressure gradient between Tahiti and Darwin in northern Australia. Similar to the Southern Oscillation Index, we will define a South Pacific Pressure Gradient Index as the difference between the 850 hPa geopotential height anomalies in the region over northern Australia (solid box in Figure 4.2 5°S to 25°S, 120°E to 140°E) and the southern Pacific (dashed box in Fig 4.2, from 15°S to 45°S, 120°W to 160°W).

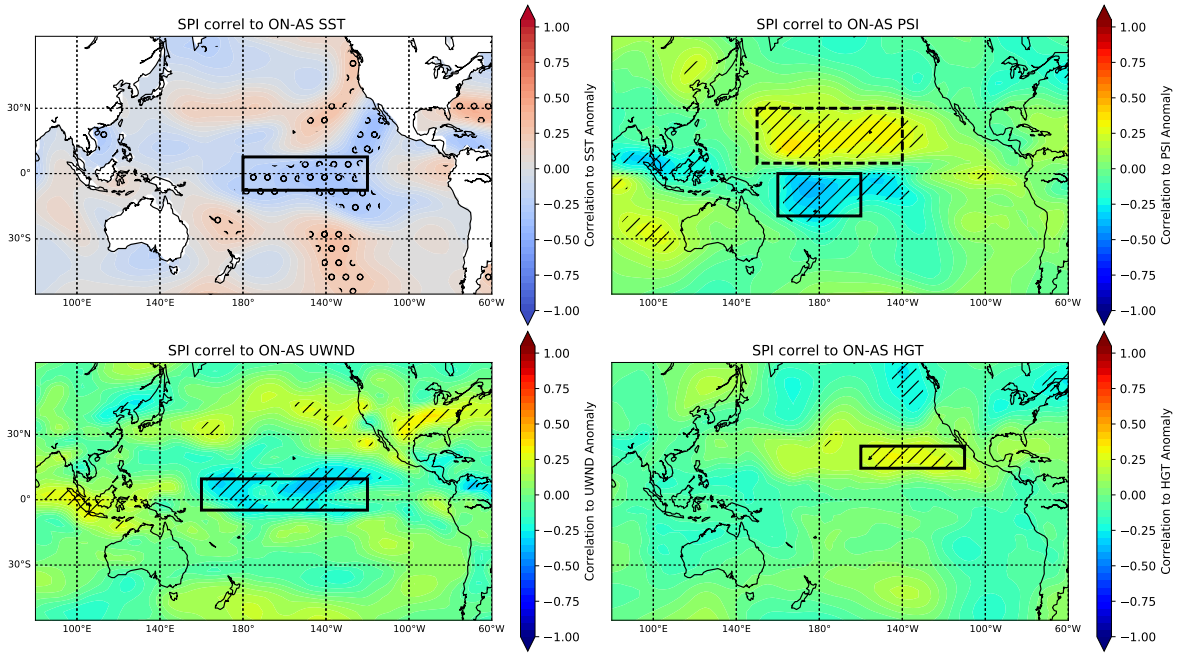


Figure 4.3: Correlation coefficient maps of Dec-May SPI and SST, 850 hPa stream function of anomalous winds, zonal winds and geopotential height tendency fields. Tendency fields are defined as the October-November mean anomaly minus August-September mean anomaly. Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 58 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

The last predictors developed for our 1-tier model are made using the anomaly tendency fields of SST and 850 hPa stream function of the anomalous wind field, zonal wind, and geopotential height. Tendency fields are defined as the October-November mean anomaly field minus the August-September mean anomaly field. Correlation maps between Dec-

May SPI and anomaly tendencies are shown in Figure 4.3. Strong negative correlation between Dec-May SPI and SST anomaly tendency is observed in Central and Eastern Pacific east of the dateline. Intensification of SST anomalies in this region is characteristic of the development of ENSO events with warming (positive tendency) related to El Niño and cooling (negative tendency) to La Niña resulting in negative correlation to SPI. An Equatorial SST Tendency Index is formulated as the area average of SST tendency from 5°S to 5°N, from the Dateline to 120°W (solid box in the top left panel of Figure 4.3).

For stream function we define predictors for the intensification of circulation anomalies in both the western north and south Pacific (dashed and solid box in the top right panel of Figure 4.3). Similarly, the intensification of zonal wind anomalies in the equatorial region are strongly correlated to Dec-May SPI. An Equatorial Zonal Wind Tendency Index is defined as the average zonal wind tendency from 5°S to 5°N and 160°E to 120°W (solid box in the bottom left panel of Figure 4.3). Finally, the geopotential height tendency over the tropical eastern Pacific shows significant correlation to Dec-May SPI. As discussed in the previous section, the circulation and geopotential height anomalies over the western Pacific during the onset of ENSO events migrate eastward as the events mature. The tendency of geopotential height anomalies in the eastern Pacific is likely related to this migration of circulation anomalies. To capture this, an Eastern Pacific Geopotential Height Tendency Index is formed as the average tendency in the region from 15°N to 25°N and 110°W to 160°W.

We use stepwise regression to reduce the number of predictors to those which hold the highest predictive power and are independent from each other. The stepwise regression methodology is set up with a 95% confidence level for both entry into and exit from the model. The selected predictors are the Equatorial Wind Index during the JJA(-1) season and the West Pacific Geopotential Height Tendency. Table 4.2 shows the correlation table of the chosen predictors for the 1-tier linear model with the target Dec-May SPI as well as with each other (a complete correlation table of all formulated predictors can be found in the appendix A.1).

Using the selected predictors, The Equatorial Wind Index in JJA(-a) and the Eastern Pacific Geopotential Height Tendency Index ON-AS(-1), a linear regression model is constructed to predict Dec-May SPI in the Northern USAPIs. Equation 4.1 is the prediction equation for our 1-tier P-E model.

Stepwise Regression chosen 1-tier Model Predictors correlation table			
	Dec-May SPI	Equatorial Wind Index JJA(-1)	Eastern Pacific Geopotential Height Tendency Index ON-AS(-1)
Dec-May	1	**-0.496**	**0.318**
Equatorial Wind JJA(-1)		1	-0.178
West Pac Hgt ON-AS(-1)			1

Table 4.2: Correlation table of predictors and target Northern Station Dec-May SPI for the 1-tier P-E model. Values in bold represent correlations significant at the 95% test level for a 58 year sample size.

$$SPI = -0.34 \times \begin{pmatrix} \text{Equatorial Wind} \\ \text{Index JJA(-1)} \end{pmatrix} + 0.22 \times \begin{pmatrix} \text{Eastern Pacific} \\ \text{Geopotential Height} \\ \text{Tendency Index} \\ \text{ON-AS(-1)} \end{pmatrix} + 0.07 \qquad (4.1)$$

Figure 4.4 shows the predicted vs. observed Northern USAPI Dec-May SPI for both the 1950-2007 train period as well as the 2008-2017 independent test period. The 1-tier predictive model achieves a correlation skill for the training set of 0.50 and 0.33 for the independent test set.

We compare the performance of this 1-tier model over the independent 2008-2017 test period with that of a more complex or 2-tier model that takes into account the existence of Canonical and Non-Canonical events explicitly.

4.2 2-tier Logistic Regression Forecasting Model

Our 2-tier model approach consists of two steps. We will first predict if the coming Dec-May season will show a Canonical or Non-Canonical relationship between Northern USAPI Dec-May rainfall and ENSO conditions. This will be done using a "classification" or "logistic" type of forecast. Depending on the result of this first forecast we will use one of two linear prediction equations (one for Canonical years and one for Non-Canonical years) for the Standardized Precipitation Index (SPI) for the Dec-May season. Details on the formulation

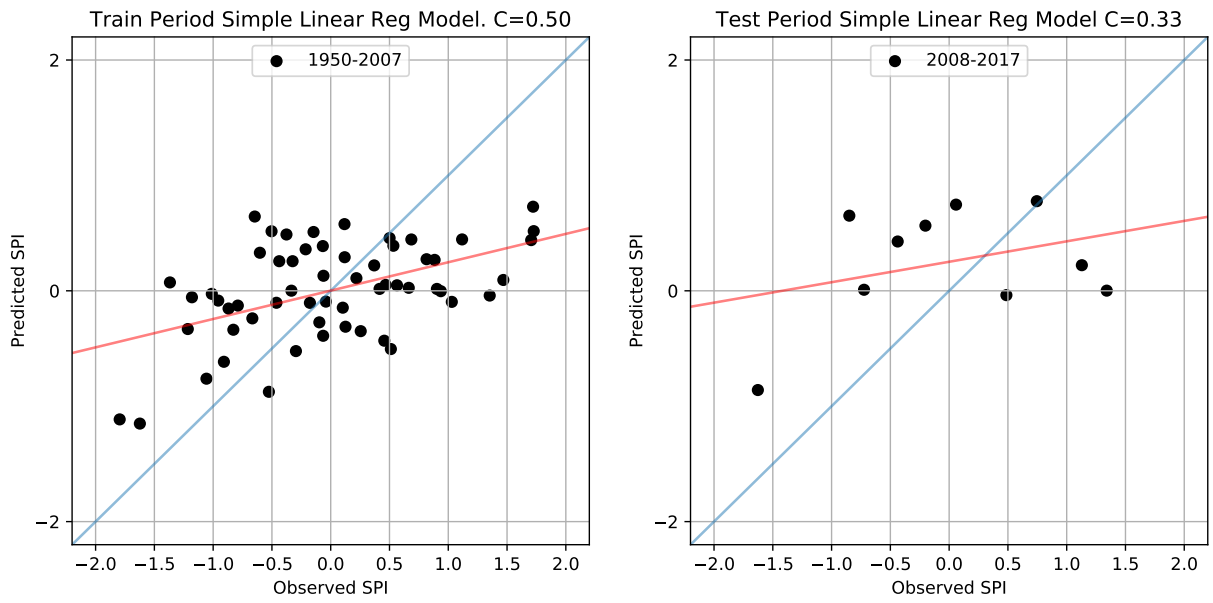


Figure 4.4: Predicted vs observed Northern USAPI Dec-May SPI form our 1-tier P-E model. The left panel shows the predicted v observed SPI for the model training period 1950-2007 while the right panel shows the same for the independent test period 2008-2017.

of the logistic forecast and the linear regressions are given in subsections 4.2.1 and 4.2.2 respectively.

4.2.1 Logistic Regression Forecast for Canonical vs. Non-Canonical Years

Our understanding of the difference in evolution of the anomaly patterns between Canonical and Non-Canonical years gleamed in Chapter 3 has been leveraged here to develop physically meaningful predictors that can differentiate between the two types of events. Once meaningful predictors are developed, a Logistic Regression will be trained on data from 1950-2007 to develop a predictive model capable of determining whether a particular year will evolve to have a Canonical or Non-Canonical relationship between rainfall and Central Pacific SST anomalies.

The first predictor developed to differentiate between Canonical and Non-Canonical events will be referred to as the "SST and Stream Function Index". As we discussed in Chapter 3 and showed in Figures 3.13 and 3.14 Canonical and Non-Canonical events of the same type (cold or warm) show stream function anomalies of the same sign over the western

Pacific and SST anomalies of opposite sign over the Eastern Pacific during the JJA season preceding the onset of the Dec-May dry season. Figure 4.5 reproduces the JJA(-1) panels from Figures 3.13 and 3.14 and with boxes for the regions used in the development of this Index.

For Cool cases for example, stream function anomalies in the western Pacific are anticyclonic, corresponding to positive stream function values, for JJA(-1) in both Canonical Cool Wet and Non-Canonical Cool Dry cases. If the area average of Stream Function anomalies over the western Pacific from the Equator to 30°N, 140°E to the dateline (boxes in top panels of Figure 4.5) is taken, positive values are expected for both Canonical Cool Wet and Non-Canonical Cool Dry years. Over the Eastern Pacific on the other hand, Canonical Cool Wet and Non-Canonical Cool Dry events show SST anomalies of opposite sign. Taking the area average of SST anomalies in the region from 15°S to 5°N, 80°W to 100°W (boxes in bottom panels of Figure 4.5) positive values are expected for Canonical Cool Wet years while negative values are expected for Non-Canonical Cool Dry years. These differences are equal but of opposite sign for the Warm counterparts of these events. By taking the product of the western Pacific area average stream function anomaly times the Eastern Pacific area average SST anomaly we obtain a predictor index that is expected to have negative sign for Canonical events and positive sign for Non-Canonical events, regardless of if the event is cold or warm, as shown in the bottom of figure 4.5. We are then able to separate all Canonical from Non-Canonical events with a single index. The area average stream function and SST anomalies are normalized by subtracting the mean and dividing the standard deviation before being multiplied to form the "SST and Stream Function Index".

Figure 4.6 shows the spatial patterns of the first two Seasonal EOF modes of rainfall in the region from 5°N to 20°N and 140°E to the dateline. As discussed in the previous section, Non-Canonical events show, in contrast to Canonical events, strong mid latitude support for the equatorial SST anomalies. To measure the relative strength of the mid latitude SST anomalies relative to the equatorial SST anomalies a "Tropical to Mid Latitude SST gradient Index" is developed. The absolute value of area average SST anomalies in the region from 120°W to 140°W, 10°N to 30°N (red box in Figure 4.6) are subtracted from the area average anomalies in the region from 120°W to 140°W, 30°N to 50°N (blue box in Figure 4.6). For Non-Canonical events, for which the mid latitude SST anomalies are of comparable magnitude as the tropical SST anomalies, the "Tropical to Mid Latitude SST gradient Index" is near 0 while for Canonical events, for which the equatorial SST anomalies are much greater than the mid latitude anomalies, this index should be strongly negative.

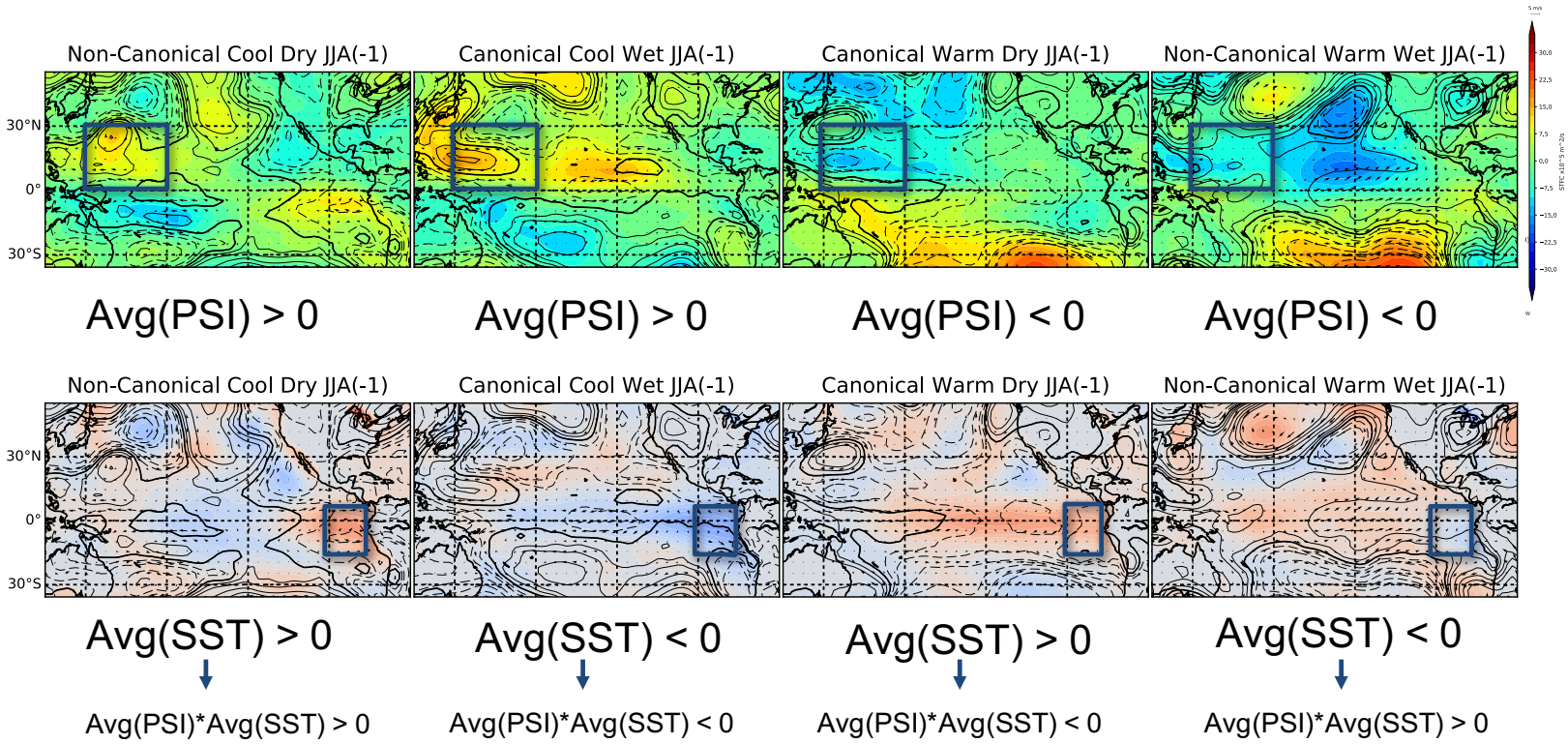


Figure 4.5: JJA (-1) anomaly composites for Canonical Cool Wet and Warm Dry and Non-Canonical Cool Dry and Warm Wet Stream Function (top) and SST (bottom). The top panels are the same as in figure 3.14 and the bottom are the same as figure 3.13. The boxes show the regions where area averages are taken to develop the "SST and Stream Function Index". Below each panel the expected sign of the area average for each type of events is shown. At the bottom of the figure, the expected sign of the complete multiplicative "SST and Stream Function Index" in shown (more details in text).

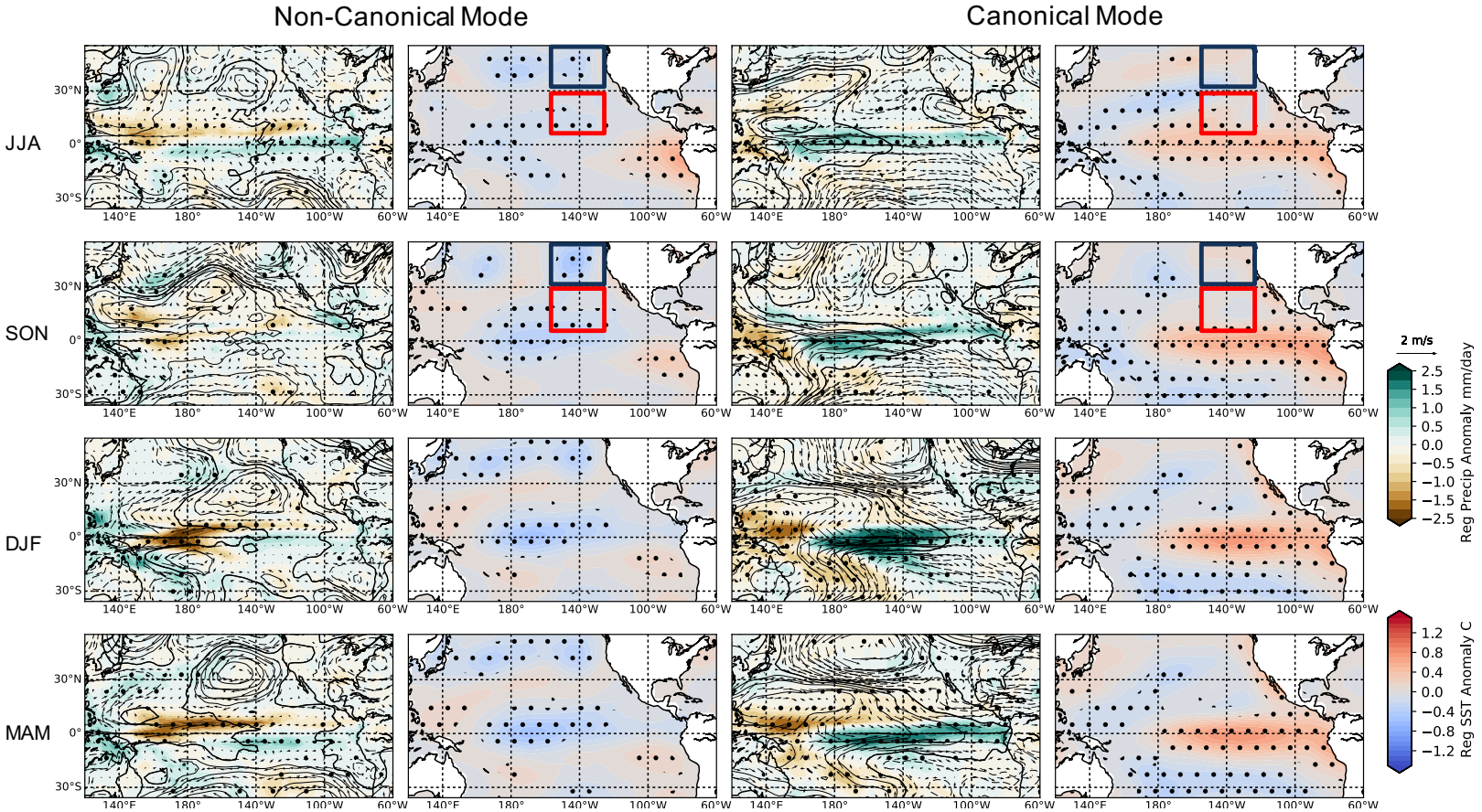


Figure 4.6: Spatial patterns of the first two seasonal EOF modes (same as Figure 3.20) of rainfall in the region from 5°N to 20°N and 140°E to the dateline. The blue and red boxes represent the regions where area average of SST anomalies are taken to develop the "Tropical to Mid Latitude SST gradient Index".

In an effort to supplement the two predictive features described before, a general analysis of the correlation between the incidence of Canonical and Non-Canonical events with SST and 850 hPa geopotential height, circulation, and zonal winds was performed. This was done by calculating the correlation between a "classification time series", or time series with 0 values for observed Canonical events both warm and cool and 1 values for Non-Canonical events, and anomaly fields of the variables mentioned. Correlation fields with both the anomaly and the absolute value of anomalies were explored.

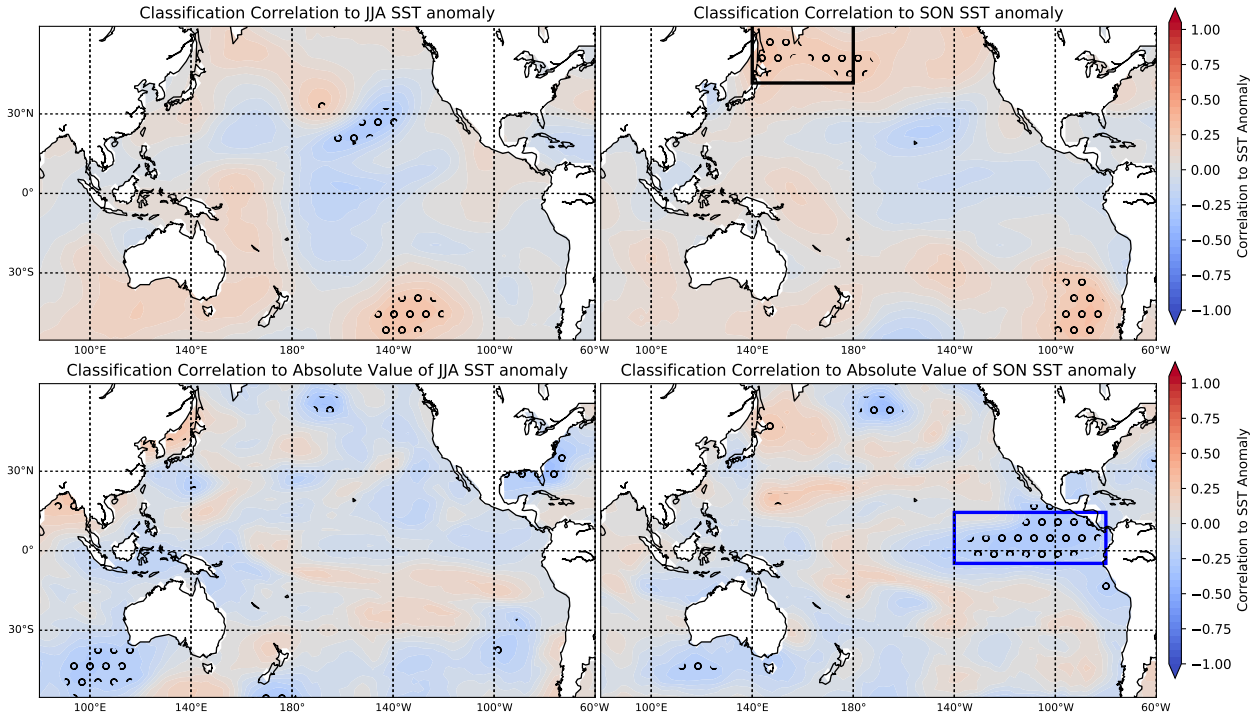


Figure 4.7: Correlation fields between the Canonical/Non-Canonical "classification time series" and SST anomalies (top) and absolute values of SST anomalies (bottom) for the JJA and SON seasons leading up to the Dec-May dry season. Boxes show the regions where the "Mid latitude SST index" (black box in top right panel) and the "Absolute Value East Pacific SST Index" (blue box in bottom right panel). Hatching shows regions where correlation is significant at the 90% confidence level using a Student t-test.

The top panels of Figure 4.7 show the correlation of the Canonical/Non-Canonical classification time series to SST anomalies in the JJA and SON seasons leading up to the Dec-May dry season. During SON, the region in the north Pacific from 40°N to 60°N and 140°E to the Dateline shows significant correlation between SST anomalies and the incidence of Non-Canonical events. The positive correlation in this region with our classification time series

implies that Non-Canonical events, both warm and cool show are linked to armer than normal SST anomalies in this region while Canonical events show cool SSTs in the same region. This is consistent with the SST anomaly composites for the SON season for Canonical and Non-Canonical events shown in Figure 3.13 and could be exploited as a predictive feature. The reason as to why Non-Canonical events both cool and warm show warm SST anomalies in this region while their Canonical counterparts show cool SST anomalies is not clear yet and will be investigated in the future. Nonetheless, a "Mid latitude SST index" index is formulated as the area average of SST anomalies in the region from 40°N to 60°N and 140°E to the dateline, which is shown as a black box in Figure 4.7. Comparing the top panels of Figure 4.7 we observe that the SST anomalies in the north western Pacific captured by the "Mid latitude SST" index develop remarkably fast from JJA to SON. To capture this rapid development, simple tendency predictor is also calculated as the difference between the absolute values of the SON(-1) and JJA(-1) "Mid latitude SST" index values.

The bottom panels of Figure 4.7 show the correlation of the Canonical/Non-Canonical classification time series and the absolute value of SST anomalies in the JJA and SON seasons leading up to the Dec-May dry season. Over the Eastern Pacific the absolute value of SST anomalies show a strong negative correlation to the classification time series. This implies that Canonical events, which are represented as 0 in our classification time series coincide with strong SST anomalies in the Eastern Pacific. This is consistent with our observations of the difference in SST patterns associated with Canonical and Non-Canonical events shown in Figure 3.9 for the dry season and Figure 3.13 for the SON season before the Dec-May dry season. Canonical events show SST anomalies with their largest magnitudes over the Cold Tongue region east of the dateline while Non-Canonical events show largest SST anomalies confined to the central Pacific. An "Eastern Pacific SST anomaly magnitude" index is developed as the absolute value of area average SST anomalies in the region from 5°S to 15°N, 80°W to 140°W to represent this feature. All predictors with the exception of the Mid Latitude SST Tendency index show correlations to the Canonical/Non-Canonical Classification time series with significance at the 90% confidence level (Table 4.4).

To chose the best subset of predictive indexes for our model, we use forward feature selection for classification to select the indexes that are predictive of the Canonical/Non-Canonical classification at the 90% confidence level using a Student t-test. The selected predictors for the final logistic regression model are the SST and Stream Function index and the Tropical to mid latitude SST gradient index. The prediction equation for our Logistic Regression model is shown in 4.2.

2-tier Model Logistic Regression Predictors			
Predictor Name	Variable	Region	Description
"SST and Stream Function index"	SST and Stream Function Fig 4.5	Stream Function anomalies. Equator-30°N, 140°E-Dateline. SST anomalies 15°S-5°N, 80°-100°W	Measures the relationship between western Pacific circulation and eastern Pacific SST anomalies. Index positive for Non-Canonical events, Negative for Canonical Events
"Tropical to mid latitude SST gradient index"	SST anomalies Fig 4.6	SST anomalies 120°W-140°W, 30°N-50°N minus SST anomalies 120°W-140°W, 10°N-20°N	Measures the relative strength of tropical vs. mid latitude SST anomalies. Near zero for Non-Canonical events Negative for Canonical events.
"Mid latitude SST anomaly index"	SST anomalies Fig 4.7	SST anomalies 40°N-60°N and 140°E to the Dateline	Measures the strength of SST anomalies over the north western Pacific. Warm
"Mid latitude SST anomaly tendency index"	SST anomalies Fig 4.7	Absolute value of "Mid latitude SST anomaly index" SON(-1) minus absolute value of "Mid latitude SST anomaly index" JJA(-1)	Measures the intensification of SST anomalies over the north western Pacific.
"Eastern Pacific SST anomaly magnitude index"	Absolute Value of SST anomalies Fig 4.7	Absolute Value of SST anomalies 5°S-15°N and 80°W-140°W	Reflects Eastern Pacific SST anomaly max during Canonical Events vs. Central Pacific max during Non-Canonical events.

Table 4.3: Description of all predictors developed for the logistic regression forecast in the 2-tier forecasting model.

2-tier Model Logistic Regression Predictor correlations						
	Canonical Classification	SST and Stream Function index JJA(-1)	Mid latitude SST anomaly index SON(-1)	Mid latitude SST anomaly tendency index	Tropical to mid latitude SST gradient index SON(-1)	Eastern Pacific SST anomaly magnitude index SON(-1)
Canonical Classification	1	0.361	0.183	0.233	0.301	-0.235
SST and Stream Function index JJA(-1)		1	-0.08	0.13	0.038	-0.712
Mid latitude SST anomaly index SON(-1)			1	-0.084	0.451	0.006
Mid latitude SST anomaly tendency index				1	0.151	-0.221
Tropical to mid latitude SST gradient index SON(-1)					1	0.003
Eastern Pacific SST anomaly magnitude index SON(-1)						1

Table 4.4: Correlation table for predictors developed for the logistic regression forecast in the 2-tier forecasting model. All predictors with the exception of the Mid Latitude SST Tendency index are correlated to the Canonical/Non-Canonical Classification time series with significance at the 90% confidence level using a Student t-test.

Because our model has a low number of predictors compared to the total number of observations, it is less likely that our linear regression will be over fit to the training data. Nonetheless, to further guard against over fitting a regularization parameter was added to our logistic regression cost function. Because the regularization parameter is an independent parameter, it most be optimized by minimizing the error of the linear model on an independent dataset. Because we want our independent test period, 2008-2017, to remain truly independent from all model development, we further divide the training period into a training set, 1950-1999 and a cross validation set 2000-2007 over which we will choose the regularization parameter.

Regularization parameters in the range of 10 to 10^{10} were used to fit the model on the 1950-1999 period and the error of each model was estimated on the 2000-2007 period. In our case, because the number of predictors is small and the predictors chosen are independent, the model does not greatly over fit the training data so the error over the 2000-2007 cross validation period remains constant at 0.75 for all values of the regularization parameter. For this reason we can choose a small regularization parameter of $c = 10$ that will not greatly penalize our model.

$$Canonical/Non-Nanonical = \frac{1}{1 + e^{-\left(1.2\times\begin{array}{c}\text{SST and}\\ \text{Stream Function}\\ \text{Index JJA(-1)}\end{array} +0.89\times\begin{array}{c}\text{Tropical to}\\ \text{mid latitude}\\ \text{SST gradient}\\ \text{Index SON(-1)}\end{array} -0.72\right)}} \tag{4.2}$$

This logistic regression model achieves a classification accuracy over the training set of 1950-2007 of 0.74 and a Hiedke Skill score of 0.43. Our logistic regression has a good performance for predicting Canonical events correctly. As table 4.5 shows, 31 out of 36 Canonical events during the training period are classified correctly. For Non-Canonical events, about half (12 out of 22 4.5) of events are correctly classified as correctly as Non-Canonical. This makes for a False Alarm Ratio of 0.24 for Canonical Events and 0.3 for Non-Canonical events.

Figure 4.8 shows a scatter plot of Dec-May SPI and DJF ONI for the 1950-2007 training period with colors representing the Canonical (green) and Non-Canonical (purple) classification. On the left, classification is shown correctly based on the observed Dec-May SPI and DJF ONI while on the right, colors represent the logistic regression prediction for each year.

Training Period

		Observations	
		Canonical	Non-Canonical
Forecast	Canonical	31	10
	Non-Canonical	5	12

Test Period

		Observations	
		Canonical	Non-Canonical
Forecast	Canonical	5	2
	Non-Canonical	1	2

Table 4.5: Contingency tables for the Train (1950-2007) and Test (2008-2017) periods for our Logistic Regression forecast of Canonical vs. Non-Canonical events.

A first test of our models predictive ability is to correctly classify extreme ENSO events. As we see in the right panel of Figure 4.8, our model is able to correctly classify strong Canonical El Niño and La Niña events, with years in the extreme upper left and lower right of this plot being classified correctly.

During the training period, our model performs well in classifying the Non-Canonical Cool dry events (lower left quadrant, right panel, Figure 4.8) missing only one strong event. We also observe that our model does not greatly misclassify Canonical Warm Dry events (lower right quadrant, right panel, Figure 4.8). With droughts being of grave concern in the USAPI, it is encouraging to observe our model has good performance during both Canonical Warm Dry and Non-Canonical Cool Dry events.

Furthermore, our model performs well in correctly classifying most Canonical Cool Wet events (upper left quadrant, right panel, Figure 4.8). This means that, at least during the training period, we observe a low false alarm rate for Non-Canonical Cool Dry events. The worst performance of our model is for Non-Canonical Warm Wet years (upper left quadrant, right panel, Figure 4.8).

The performance of this logistic regression model, while far from perfect with a prediction accuracy of 70%, is greatly complimentary to the performance of the PEAC Center operational forecasting suite. Drought conditions in the Northern USAPIs pose the high risk to the health, food security and livelihood of the local populations. PEAC Center models show good skill in forecasting Canonical Warm Dry or El Niño induced droughts but fail to capture Non-Canonical Cool Dry conditions for which our model performs well. A lingering issue with both PEAC Center forecasts and our logistic regression model is false alarms

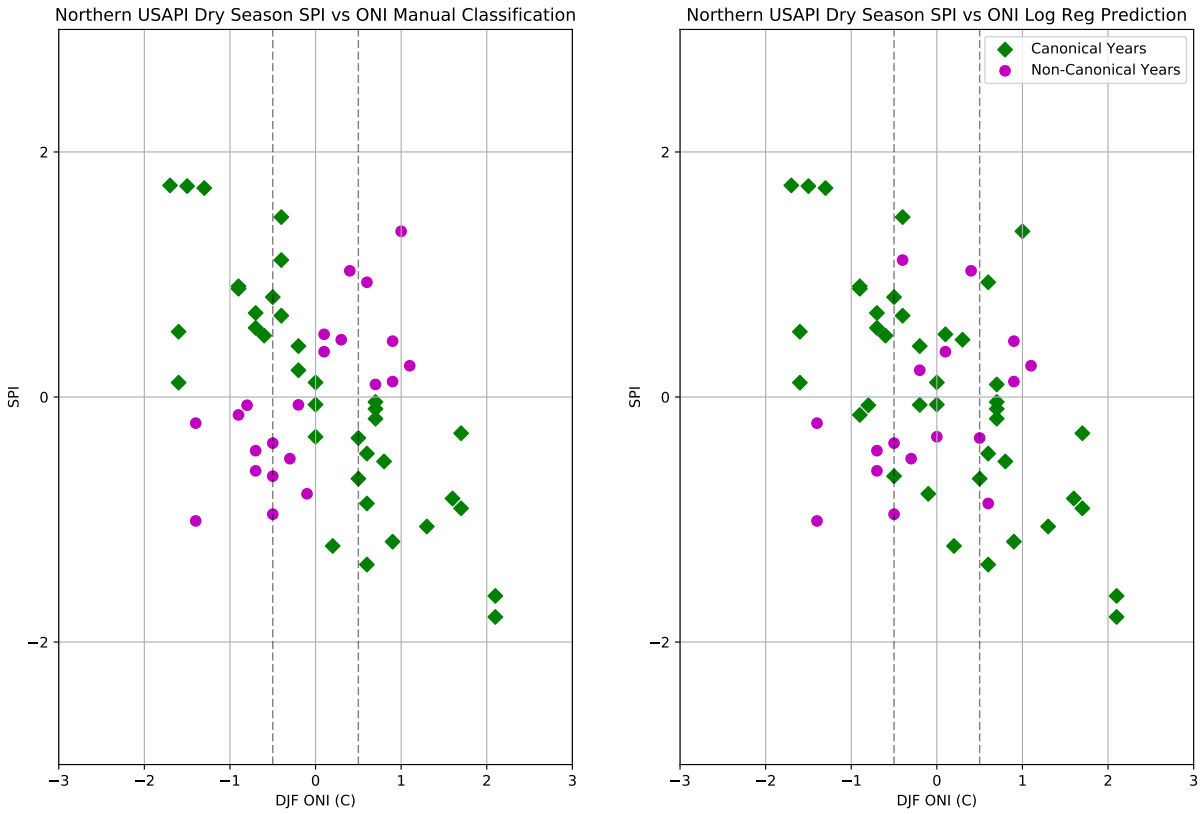


Figure 4.8: Dec-May SPI and DJF ONI for the 1950-2007 training period. Colours plot represent the Canonical (green) and Non-Canonical (purple) classification. On the left panel, correct classification based on the observed Dec-May SPI and DJF ONI. on the right panel, colors represent the logistic regression prediction of Canonical or Non-Canonical for each individual year.

for drought conditions during Non-Canonical wet events, which are not well captured by both PEAC Center forecasts and our logistic regression model. While these false alarms for drought during warm conditions pose a challenge in that they can erode user confidence in future forecasts, from an impact mitigation perspective they are less detrimental than failing to forecast droughts.

Using this ahead of time prediction of whether a particular year will show a Canonical or Non-Canonical relationship between rainfall and ENSO we will use one of the two linear regression prediction models developed in the following section to predict the Dec-May Standardized Precipitation Index for the Northern USAPIs.

4.2.2 Linear Prediction Equations for the SPI During Canonical and Non-Canonical Years

We will develop two separate linear prediction equations to be used depending on the result of the logistic regression forecast. The training data, 1950-2007, will be separated into two independent training sets. We will separate the training set using the prediction of our logistic regression model. Years in the training set predicted to be Canonical will be separated into a Canonical training set while years predicted to be Non-Canonical will be used to create the Non-Canonical training set. In the 1950-2007 training set, there are 41 years predicted to be Canonical (36 observed) and 18 years predicted to be Non-Canonical (22 observed).

We will search for predictors for each linear model independently. We will produce correlation maps of the SPI to JJA and SON anomaly fields of relevant variables as well as the tendency fields of anomalies calculated as the October-November anomaly minus August-September anomaly. From these correlation maps, the most relevant and physically meaningful predictors will be isolated and used in a stepwise regression model to select the most significant and produce a predictive model.

Figure 4.9 shows the correlation maps of SPI during Canonical years to Septermber-October-November (SON) SST anomaly, SON 850 hPa geopotential height anomalies, June-July-August (JJA) 850 hPa zonal wind anomalies and 850 hPa zonal wind tendency from August-September to October-November. We will describe the formulation of each predictor here in detail, but a short description of every predictor, the anomaly field and region used to define it as well as a short description of the mechanisms captured is found in Table 4.6. Table 4.7 shows a correlation table for all the Canonical predictors as well as the Dec-May SPI for Canonical events.

SPI in the Canonical training set show correlation to SST anomalies (Figure 4.9 upper left panel) consistent with the ENSO SST horseshoe pattern with negative correlations in the eastern Pacific east of the dateline and positive correlation to the west. To represent the relationship between canonical SPI and SST anomalies the area average of SST anomalies in the Nino3.4 region (120°W-170°W and 5°S- 5°N) is used as a predictive index. This predictor is as expected highly correlated with the Dec-May SPI during Canonical years with a correlation value to Dec-May SPI of -0.59 exceeding significance at the 99% confidence level for 41 years in the Canonical training set.

850 hPa geopotential height anomalies (Figure 4.9 upper right panel) show strong correlation to SPI in the Canonical training set over the south Pacific in a dipole pattern consistent

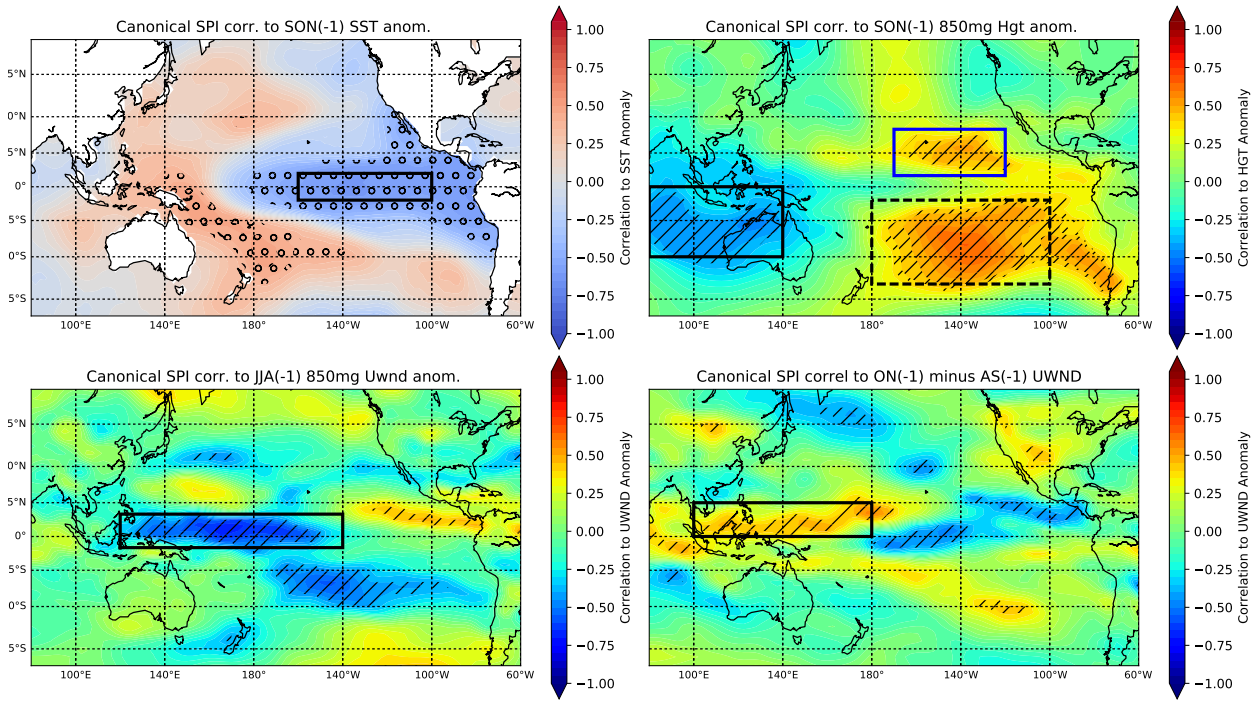


Figure 4.9: Correlation coefficient maps of Dec-May SPI for years in the Canonical training set and SON SST (top left), SON 850 hPa Geopotential Height (top right) and, 850 hPa zonal winds (lower left) anomalies, as well as ON-AS 850 hPa zonal wind anomaly tendency (lower right). Hatching represents significance at the 95% confidence level using a Student t-test for a sample size of 41 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

with the Southern Oscillation, with high pressure over the eastern south Pacific accompanied by low over north Australia and the Indian Ocean during negative phases on ENSO. This results in positive correlations to Dec-May SPI over the southeastern Pacific and Negative over the West. To capture these anomalies, a substractive index is defined as the area average 850 hPa geopotential height anomalies from 100°W to the Dateline 40°S to 5°S minus the area average anomalies in the region from 80°E to 140°E 30°S to the Equator (dashed and solid boxes in the top left panel of Figure 4.9). We will refer to this index as the "South Pacific Geopotential Height Gradient Index". From Table 4.7, this index is highly correlated to both the Dec-May SPI and the SON Nino3.4 index.

In the western tropical Pacific, the 850 hPa geopotential height anomalies show positive correlation to Dec-May SPI in the region from about 120°W to 160°W and 5°N to 25°N. The geopotential height anomalies are part of the ENSO surface pressure anomaly pattern during

the developing SON season. Figure B.1 in the appendix shows the correlation maps between DJF Oceanic Niño Index and preceding JJA and SON SST and 850 hPa geopotential height anomalies. Nonetheless, the area average of geopotential height anomalies in that region (blue box top right in Figure 4.9), referred to as the "Eastern Pacific Geopotential Height Index", shows low correlation to contemporary Nino3.4 values as seen from Table 4.7. So the "Eastern Pacific Geopotential Height Index" forms a good complementary index to the "South Pacific Geopotential Height Gradient Index" that measures the geopotential height and pressure anomalies in the northern Pacific and has high correlation with ENSO intensity in the following DJF season.

2-tier Model Canonical Linear Regression Predictors			
Predictor Name	Variable	Region	Description
"Nino3.4"	SST anomaly Fig4.9	SST anomalies 120°W-170°W and 5°S- 5°N.	Measures ENSO phase and intensity (Bamston et al. (1997))
"South Pacific Geopotential Height Gradient Index"	850 hPa Height anomaly Fig4.9	100°W to Dateline, 40°S to 5°S, minus 80°E to 140°E, 30°S to the Equator	Measures ENSO phase and intensity. Similar to the "Southern Oscillation Index" (Trenberth (1997)).
"Eastern Pacific Geopotential Height Index"	850 hPa Height anomaly Fig4.9	120°W to 160°W and 5°N to 25°N	Northern Hemisphere ENSO pressure anomalies.
"Equatorial Wind Index"	850 hPa Zonal wind anomaly Fig4.9	120°E to 140°W, 5°S to 5°N	Captures the western Pacific zonal wind anomalies related to ENSO onset (Yu et al. (2003)).
"Western Pacific Wind Tendency Index"	850 hPa Zonal wind anomaly tendency Fig4.9	Equator to 5°N, 140°E to the Dateline	Measures intensification of western Pacific zonal wind anomalies (Yu et al. (2003)).

Table 4.6: Description of all predictors developed for the Canonical linear regression model in the 2-tier forecasting model.

Zonal wind anomalies in the JJA(-1) ENSO developing season in the western Pacific (bottom left panel of Figure 4.9) as well as their tendency (bottom right in Figure 4.9) are also strongly correlated to SPI during Canonical years. An "Equatorial Wind Index" (120°E to 140°W, 5°S to 5°N, bottom left panel of figure 4.9) and "Western Pacific Wind Tendency Index" (Equator to 5°N, 140°E to the Dateline, solid box in the bottom right panel of Figure 4.9) are defined to capture these relationships. These indexes capture the wind anomalies observed in the onset of ENSO events. Both of these predictors show significant (99% confidence level) to the target Dec-May SPI as seen in Table 4.7.

Using a stepwise regression predictor selection methodology on these 5 predictors we find that the best predictors are the "South Pacific Geopotential Height Gradient Index" and the "Eastern Pacific Geopotential Height Index". Using these two predictors in a linear regression equation we find the predictive equation for the standardized Precipitation Index for the Dec-May season during Canonical years to be given by Equation 4.3.

$$CanonicalSPI = -0.51\times \begin{pmatrix}\text{South Pacific}\\ \text{Geopotential}\\ \text{Height}\\ \text{Gradient}\\ \text{Index}\\ \text{SON(-1)}\end{pmatrix} +0.24\times \begin{pmatrix}\text{Eastern}\\ \text{Pacific}\\ \text{Geopotential}\\ \text{Height Index}\\ \text{SON(-1)}\end{pmatrix} +0.09 \quad (4.3)$$

This prediction equation achieves a correlation skill score between observed and predicted SPI in the Canonical training set of 0.68 as shown in figure 4.11. The average leave one out cross validation skill score of the Canonical linear regression is 0.51 while the k-fold cross validation error with k=10 is 0.63 (consistent with the leave one out cross validation overestimating the error). As the linear relationship between predicted and observed SPI seen in Figure 4.11 as well as the cross validation scores suggest, our linear regression model for SPI during Canonical years suffers of high bias. This could be overcome by allowing for the inclusion of more predictors in the model, the creating of new predictors or using a higher order model to predict SPI. Nevertheless, all of these solutions are likely to result in over fitting and thus we run unto the trouble of the bias/variance trade off in the predictive model. For this work we wish to keep the predictive models simple so that the physical relationships between predictors and target variable can be understood, so this linear model to predict the SPI during Canonical years is appropriate.

2-tier Model Canonical Linear Regression Predictor correlations						
	Canonical SPI Dec-May	Nino3.4 SON(-1)	South Pacific Geopotential Height Gradient Index SON(-1)	Eastern Pacific Geopotential Height Index SON(-1)	Equatorial Wind Index JJA(-1)	Western Pacific Wind Tendency Index
Canonical SPI Dec-May	1	-0.592	-0.627	0.369	-0.588	0.54
Nino3.4 SON(-1)		1	0.857	-0.12	0.929	-0.785
South Pacific Geopotential Height Gradient Index SON(-1)			1	-0.182	0.859	-0.75
Eastern Pacific Geopotential Height Index SON(-1)				1	-0.076	0.073
Equatorial Wind Index JJA(-1)					1	-0.836
Western Pacific Wind Tendency Index						1

Table 4.7: Correlation table for predictors developed for the Canonical linear regression forecast in the 2-tier forecasting model. All predictors are correlated to the Canonical SPI time series with significance at the 90% confidence level using a Student t-test.

Predictors for SPI during Non-Canonical years were searched for in a similar fashion as for Canonical years. Figure 4.10 shows the correlation maps of SPI during Non-Canonical years to SON SST anomaly, JJA 850 hPa stream function of the anomalous winds, JJA 850 hPa geopotential height anomalies and SST anomaly tendency from August-September to October-November.

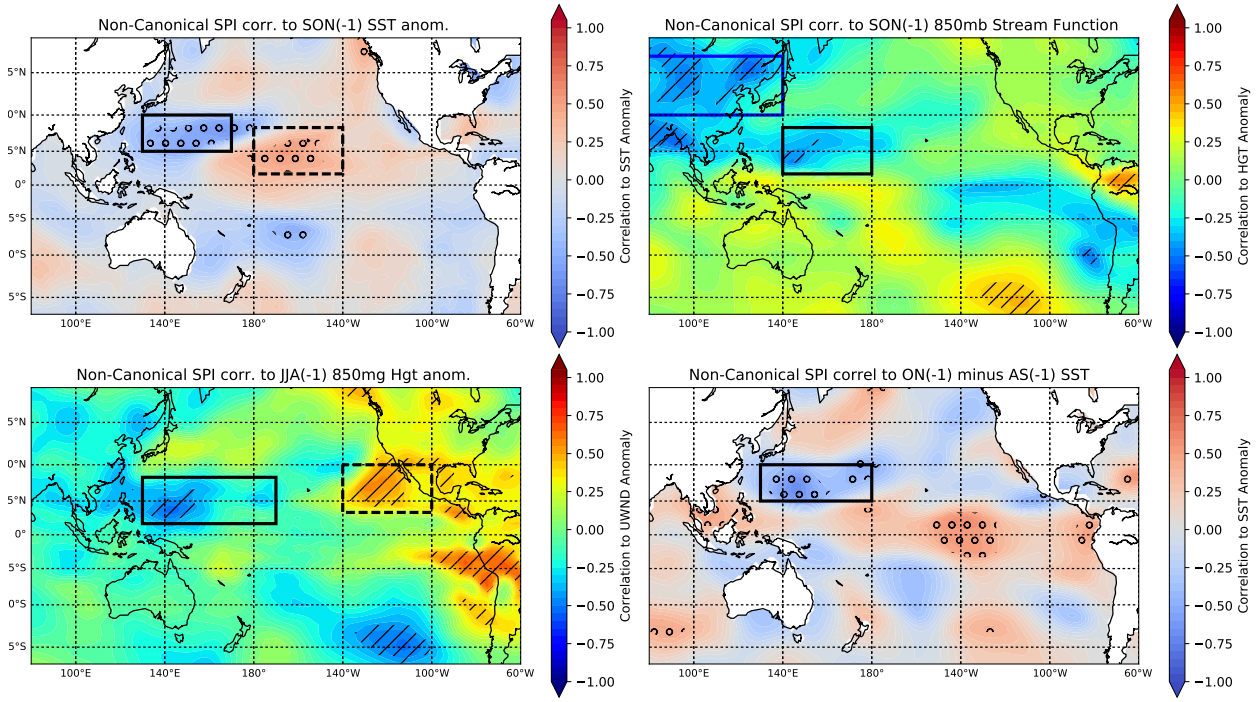


Figure 4.10: Correlation coefficient maps of Dec-May SPI for years in the Non-Canonical training set and SON SST (top left), SON 850 hPa Geopotential Height (top right) and, 850 hPa zonal winds (lower left) anomalies, as well as ON-AS SST anomaly tendency (lower right). Hatching represents significance at the 90% confidence level using a Student t-test for a sample size of 17 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

SPI in the Non-Canonical training set shows a correlation pattern to previous SON SST anomalies, shown in the top left panel of Figure 4.10, that is consistent with SST composites of Non-Canonical events shown in the previous section (Figure 3.9). SST anomalies in Non-Canonical events are confined to the central and western Pacific and form a dipole pattern with anomalies of opposite sign over the central Pacific and to the northwest. This SST pattern is representative of the Ocean/Atmosphere interaction between regional circulation anomalies in the western Pacific and underlying SST anomalies that characterize

the Non-Canonical events. The Non-Canonical SPI correlation to SST pattern differs with the composites in that, while Non-Canonical SST composites showed their strongest SST anomalies along the equator, the strongest Non-Canonical SPI and SST correlations are seen slightly further north.

To capture the effect of SST anomalies during the SON(-1) season on the Dec-May SPI during Non-Canonical years an SST gradient index is formulated between the area averaged SST anomalies in the region from 15°N to 30°N, 150°E to 170°E minus the area average anomalies in the region from 5°N to 25°N, 140°W to the Dateline. This predictor will be referred to as the "Tropical Western Pacific SST Gradient Index". Ithas a correlation to the Non-Canonical Dec-May SPI of -0.6 wich is significant at the 99% confidence level for 17 samples (see Table 4.9).

850 hPa stream function of the anomalous winds (top right panel of Figure 4.10) shows areas of strong correlation to Non-Canonical SPI located over the tropical western Pacific over the USAPIs and over Mainland Asia. As was discussed in the previous section based on composite analysis, circulation anomalies over the western Pacific during the JJA and SON seasons persist into the Dec-May dry season during Non-Canonical events (as opposed to migrating east as they do in Canonical events, see Figure 3.14). The stream function of anomalous winds in the western Pacific captures these regional circulation anomalies in the SON season by showing significant correlation to SPi during Non-Canonical events. To represent these regional circulation anomalies in our model, a index of area average stream function of anomalous 850 hPa winds is used, withing the region from 5°N to 25°N, 140°E to the Dateline, called the "Western Pacific Stream Function Index". This index shows a small correlation to Non-Canonical SPI of -0.29 which is not significant at the 90% confidence level using a Student t-test (see Table 4.9) but is highly correlated to the "Tropical Western Pacific SST Gradient Index" which represents underlying SST anomalies in the same region.

Over the Asian mainland, the stream function of anomalous 850 hPa winds shows strong correlation to Non-Canonical SPI. Circulation anomalies in this region, while of opposite sign, in Dec-May Non-Canonical composites (see Figure 3.8) are not strong. It is not clear why these anomalies show strong correlation to Non-Canonical SPI over a fairly large area. This correlation could be representative of the influence of anomalous heating/cooling over the Asian Mainland inducing circulation anomalies over the continent, but a corresponding signature in the 850 hPa geopotential height anomalies would be expected and is not observed (bottom left panel of Figure 4.10). Nonetheless, a predictive index with the are average values of stream function of the anomalous winds from 80°E to 140°E, 30°N to 50°N is considered

2-tier Model Non-Canonical Linear Regression Predictors			
Predictor Name	Variable	Region	Description
"Tropical Western Pacific SST Gradient Index"	SON SST anomalies 4.10	15°N-30°N, 130°E-170°E minus 5°N-25°N, 140°W-Dateline	Represents response of SST anomalies to Ocean Atmosphere interaction in the western Pacific (Wang and Zhang (2002))
"Western Pacific Stream Function Index"	SON Stream Function of anomalous 850 hPa winds 4.10	5°N-25°N, 140°E-Dateline	Represents the circulation response to Ocean Atmosphere interaction in the western Pacific
"Asian Continent Stream Function Index"	SON Stream Function of anomalous 850 hPa winds 4.10	30°N-50°N, 80°E-140°E	Measures the relationship between
"Tropical Pressure Gradient Index"	JJA 850 hPa Geopotential height anomalies 4.10	5°N-25°N, 130°E-170°W minus 10°N-30°N, 100°W-140°W	Measures the tropical pressure gradient related to Walker Circulation anomalies
"Tropical Western Pacific SST Tendency Index"	Aug-Sep to Oct-Nov SST tendency, 4.10	15°N-30°N, 130°W-Dateline	Measures the intensification of the SST anomaly dipole characteristic of Ocean/Atmosphere interaction during non-Canonical events.

Table 4.8: Description of all predictors developed for the Non-Canonical linear regression model in the 2-tier forecasting model.

in our model referred to as the "Asian Continent Stream Function Index". This index has a surprisingly high correlation value to non-Canonical SPI of -0.46, significant at the 95% confidence level using a Student t-test, but is also highly correlated to the "Tropical Western Pacific SST Gradient Index" (Table 4.9).

In the Northern Hemisphere, the 850 hPa geopotential height anomalies show a dipole pattern of anomalies strongly correlated to Non-Canonical SPI values (bottom left panel of Figure 4.10). This dipole of geopotential height is representative of the Walker Circulation anomalies linked to rainfall anomalies over the western Pacific. To capture the effect of these geopotential height anomalies, a gradient index is defined as the 850 hPa geopotential height anomalies area averaged in the region from 5°N to 25°N, 130°E to 170°W, minus the area average anomalies from 10°N to 30°N, 100°W to 140°W. This "Tropical Pressure Gradient Index" shows a correlation to Non-Canonical SPI of -0.54 (significant at the 95% confidence level using a Student t-test) and is also highly correlated to the "Tropical Western Pacific SST Gradient Index" and "Asian Continent Stream Function Index" (see Table 4.9).

Finally, as seen in the bottom right panel of Figure 4.10, the tendency of SST anomalies from August-September to October-November shows over the Central and western Pacific a pattern of correlated anomaly tendency consistent with the intensification of the Ocean/Atmosphere coupling between regional circulation and underlying SST anomalies that characterizes Non-Canonical events. To capture the intensification of this Ocean Atmosphere interaction several indexes were developed, including a gradient index between the areas of negative and positive correlation and independent area average indexes over each area. All of the resulting indexes were highly correlated, thus the most effective index, as measured by it's correlation to Non-Canonical SPI target variable was chosen. The resulting index was taken as the area average of SST anomaly tendency from 15°N to 30°N, 130°W to the Dateline and will be referred to as the "Tropical Western Pacific SST Tendency Index" which has a correlation of -0.59 to SPI values during Non-Cannical years (Table 4.9).

We use stepwise regression on the 5 predictors for the Non-Canonical SPI and find that the tropical western Pacific SST Gradient Index during SON is the only predictor selected for the linear model. Using this predictor in a linear reggression equation we find the predictive equation for the standardized Precipitation Index for the Dec-May season during Non-Canonical years to be given by Equation 4.4.

2-tier Model Non-Canonical Linear Regression Predictor correlations						
	Non Canonical SPI Dec-May	Tropical Western Pacific SST Grad Index SON(-1)	Western Pacific Stream Function Index SON(-1)	Asian Continent Stream Function Index SON(-1)	Tropical Pressure Gradient Index JJA(-1)	Tropical Western Pacific SST Tendency Index
Non Canonical SPI Dec-May	1	-0.6	-0.29	-0.46	-0.54	-0.59
Tropical Western Pacific SST Grad Index SON(-1)		1	0.78	0.41	0.48	0.45
Western Pacific Stream Function Index SON(-1)			1	0.33	0.23	0.17
Asian Continent Stream Function Index SON(-1)				1	0.54	0.4
Tropical Pressure Gradient Index JJA(-1)					1	0.67
Tropical Western Pacific SST Tendency Index						1

Table 4.9: Correlation table for predictors developed for the Non-Canonical linear regression forecast in the 2-tier forecasting model. All predictors with the exception of the Western Pacific Stream Function Index are correlated to the Non-Canonical SPI time series with significance at the 90% confidence level using a Student t-test.

$$Non-CanonicalSPI = -0.35 \times \begin{pmatrix} \text{Tropical} \\ \text{Western} \\ \text{Pacific} \\ \text{SST Grad} \\ \text{Index} \\ \text{SON(-1)} \end{pmatrix} - 0.026 \tag{4.4}$$

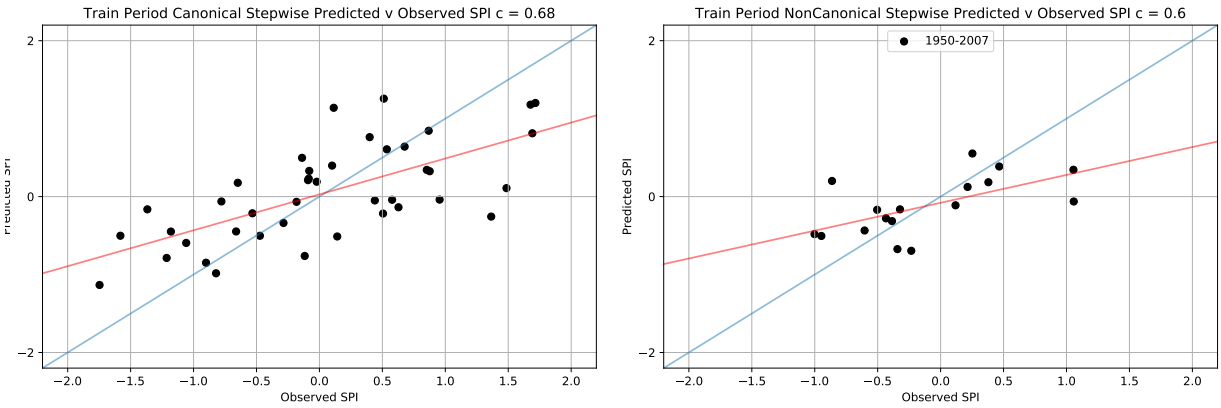


Figure 4.11: Predicted vs. Observed SPI for the Canonical training set (left) and the Non-Canonical training set (right). The blue line represents the 1 to 1 line while the red line represents the linear regression of observed SPI onto predicted SPI values. The correlation values are presented above and are significant at the 95% confidence level using a Student t-test in both cases (41 year sample size for Canonical years and 17 for Non-Canonical years)

This prediction equation achieves a correlation skill score between observed and predicted SPI in the Non-Canonical training set of 0.6 as shown in Figure 4.11 which is comparable to the 0.68 obtained by the Canonical prediction equation on the Canonical Training set. The average leave one out cross validation skill score of the Non-Canonical linear regression is 0.30 while the k-fold cross validation error with k=10 is 0.46. As in the Canonical case, the linear relationship between predicted and observed SPI seen in Figure 4.11 as well as the cross validation scores suggest, the Non-Canonical linear regression model suffers of high bias. This problem is more severe in the Non-Canonical case given that the training set size is about half of the Canonical set.

4.2.3 Test Period Forecast

In this section we will bring together the two tiers of our prediction methodology and produce a forecast for the 2008-2017 period which has been kept independent from any training

process of the models and should provide a measure of our models predictive ability.

Figure 4.12 shows the observed Dec-May SPI plotted vs the DJF ONI were the colors of the markers represent the correct Canonical/Non-Canonical classification, in the left panel, and the logistic regression forecast of the Canonical/Non-Canonical classification, in the right panel.

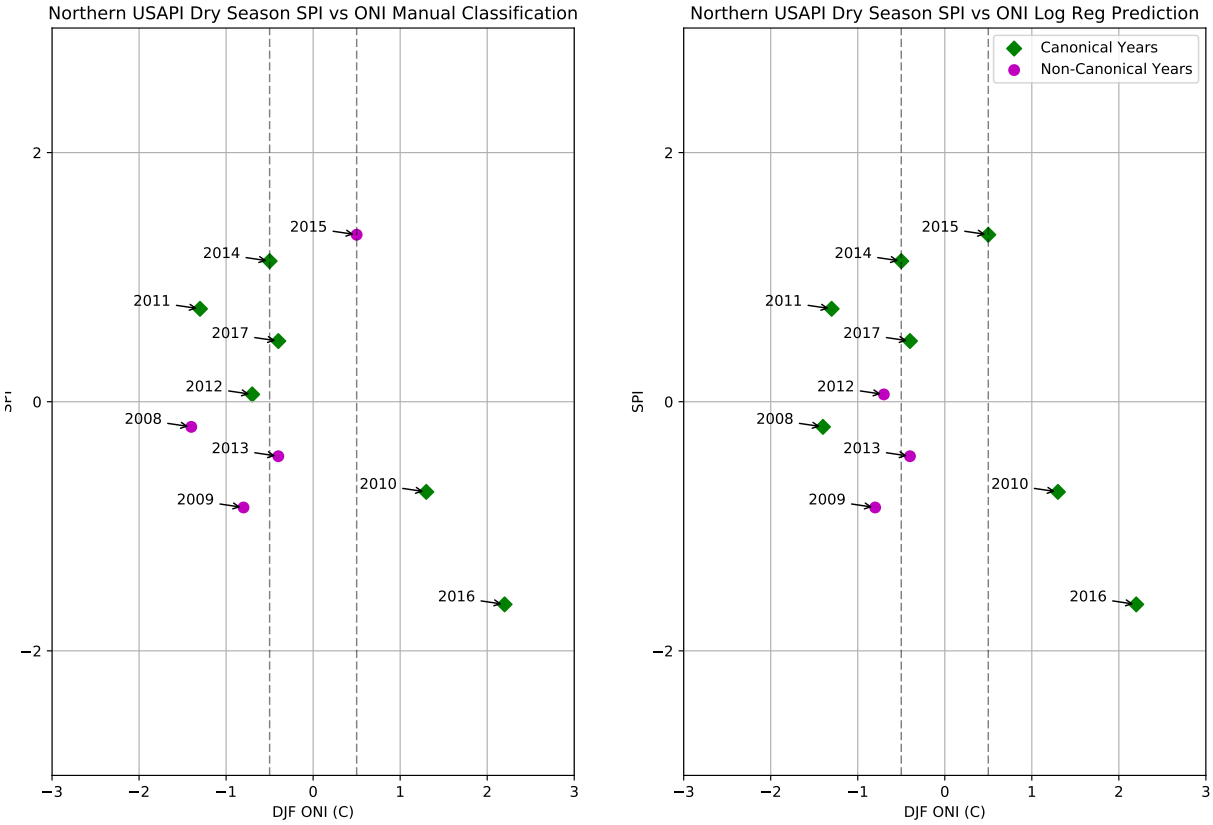


Figure 4.12: As in Figure 4.8. Dec-May SPI and DJF ONI for the 2008-2017 test period. Colours plot represent the Canonical (green) and Non-Canonical (purple) classification. On the left panel, correct classification based on the observed Dec-May SPI and DJF ONI. on the right panel, colors represent the logistic regression prediction of Canonical or Non-Canonical for each individual year.

During the independent test period our logistic regression achieves an accuracy of 0.7 and a Hiedke skill score of 0.35; 5 out of 6 Canonical events are classified correctly while half of Non-Canonical events are classified correctly (see Table 4.10). False Alarm Rates for the independent training period are 0.28 for Canonical events and 0.5 for Non-Canonical events. The strong El Niño of 2009/2010, the very strong El Niño of 2015/2016, and the strong La Niña event of 2010/2011 are correctly classified by our logistic regression method. The

Test Period

		Observations	
		Canonical	Non-Canonical
Forecast	Canonical	5	2
	Non-Canonical	1	2

Table 4.10: Contingency tables for the Test (2008-2017) periods for our Logistic Regression forecast of Canonical vs. Non-Canonical events.

strong El Niño of 2009/2010 as well as the very strong El Niño of 2015/2016 are correctly classified by our logistic regression forecast as well as the strong La Niña event of 2010/2011. Three years are misclassified by our logistic regression forecast: two La Niña events, 2008 and 2012, which produce weak SPI anomalies over the Northern USAPIs (2008 slightly dry and 2012 slightly wet) and the Non-Canonical Warm wet event of 2014/2015 which produced strong wet conditions during El Niño like SST conditions.

The behavior of our logistic regression forecast is very consistent with what was expected from our diagnostics of the model during the training set, with comparable accuracy and score and showing difficulties in classifying the same type of non-Canonical Warm Wet events like 2015.

Using our complete two tier model, logistic regression plus Canonical and Non-Canonical linear regressions, we forecast the SPI for the Dec-May dry season for the independent test set of 2008-2017. Results are shown in the center panel of Figure 4.13 along with the correlation values of predicted vs observed SPI. The left panel of Figure 4.13 shows the forecast for the test period of the benchmark 1-tier P-E model developed in Section 4.1 for comparison. The right panel of Figure 4.13 shows the result of applying the second stage of the 2-tier forecast after using the original manual classification of Canonical versus Non-Canonical events–that is, a perfect logistic forecast. This gives an upper-bound estimate of potential forecast skill.

The 1-tier Model shows a correlation of predicted vs. observed SPI during the training set of 0.33 while the 2-tier Model shows a correlation of 0.72 and under perfect logistic forecast a correlation of 0.83. Non-Canonical years in the test period are represented in figure 4.13 with diamond shaped markers, red for Non-Canonical Cool Dry events and blue for Non-Canonical Warm Wet events.

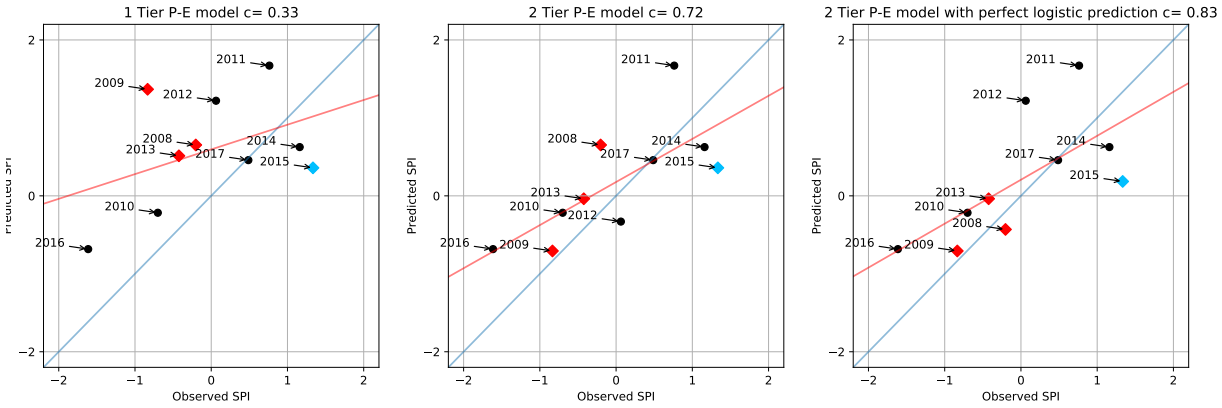


Figure 4.13: Scatterplot f the predicted vs observed Standardized Precipitation Index for our 1-tier benchmark model (left), our 2-tier P-E model (center) and the same 2-tier model with perfect logistic prediction of Canonical and Non-Canonical events (right). The blue lines represent the 1:1 line while the red line represent the achieved regression line between prediction ans observations. Correlation values between predicted and observed SPI values presented at the top of each panel.

4.2.4 Forecast Discussion

The observational analysis of the modes of rainfall variability in the Northern USAPIs carried out in Chapter 3 led us to conclude that there are two distinct modes of interannual variability that govern the tropical western Pacific. The first of these modes is what we call here the Canonical mode, which is consistent with previously described El Niño/Southern Oscillation SST, circulation and rainfall anomalies. The second of these modes, referred to here as the Non-Canonical mode, has not been described in detail prior to this work. We have shown that it differs markedly from the Canonical mode in the evolution of its anomaly patterns.

In this chapter we leverage the recognition of these distinct Canonical and Non-Canonical modes of variability to formulate a forecasting framework that is able to more accurately predict the Standardized Precipitation Index for the Dec-May dry season in the Northern USAPIs. This is achieved by formulating a 2-tier approach, in which we first predict if a particular year will be Canonical or Non-Canonical using a logistic regression approach and then, using separate linear regressions for each type of year, predict the SPI for the Dec-May dry season.

Predictors for the logistic forecast of Canonical or Non-Canonical conditions are developed in two ways. In the first method, predictors are crafted from the observed differences in the evolution of anomaly fields for each type of event. Predictor indexes are then formulated to represent these differences. In the second method, correlations between a Canonical/Non-

Canonical classification time series and relevant anomaly fields was carried out and predictor indexes are derived to represent regions with high correlations. Applying a stepwise regression predictor selection methodology we find that the most significant predictors those developed from the first method. When the prediction task is, as in this case, a complex one, where there is strong nonlinear behavior, formulation of predictors via correlation analysis can prove difficult. Careful analysis of the evolution of these events is important in the development of appropriate predictors.

Our logistic regression model achieves a classification accuracy of 74% over the 1950-2007 training set and 70% over the independent test set. This classification forecast shows good performance during Canonical events both Warm and Cool (El Niño and La Niña events) and Non-Canonical Cool events which represent the poorly forecasted droughts in the Northern USAPIs. The worst performance of this classification model is during Non-Canonical Warm Wet events, i.e. El Niño like conditions that do not produce droughts. While a drought forecast that does not verify is less of a problem from the operational forecast point of view than an unforecast drought, it still poses a challenge that should be addressed. This might be achieved by looking carefully at the anomaly fields of strong misclassified Non-Canonical Warm events (upper right quadrant, right panel Figure 4.8) to understand why these events are misclassified.

Using this logistic regression model, the training set was separated into two distinct training sets, Non-Canonical and Canonical, for which independent predictors and linear regressions between SPi and predictors were developed. In spite of the shortcomings of our logistic regression model, when used to separate the training set into Canonical and Non-Canonical years to train the two separate linear regressions, we find that all the predictors formed for each training set are strongly correlated to each other.

For the Canonical training set, the predictors selected using stepwise regression represent the intensity of ENSO as measured by the Southern Hemisphere pressure gradient and the Northern Hemisphere tropical pressure gradient. These effects are represented by the South Pacific Geopotential Height Gradient Index SON(-1), similar to the Southern Oscillation Index and the Eastern Pacific Geopotential Height Index SON(-1). Our linear regression model achieves a correlation score on the Canonical training set of 0.68. For the Non-Canonical training set, the predictor selected via stepwise regression represents the Ocean/Atmosphere interaction in the western Pacific that characterizes Non-Canonical events, with the linear regression model achieving a correlation score of 0.6 on the Non-Canonical training set.

The 2-tier forecasting methodology developed in this work, shows promise as a value

adding tool to the PEAC Center operational forecasting methodology. Our methodology performs well during the most dangerous of Non-Canonical events, the Cool Dry conditions that bring unexpected droughts to the Northern USAPIs; The rest of the PEAC Center forecast suite shows poor skill during these periods (as discussed in the introduction and seen in Figure 1.2). Both our methodology and the PEAC Center suite perform well during Canonical events. These features would most certainly make the this model a good addition go the PEAC Center operational forecasting suite.

CHAPTER 5
CONCLUSIONS

A thorough study of droughts in the U.S. Affiliated Pacific Islands (USAPIs) has yielded a more complete understanding of the mechanisms behind interannual rainfall variability in the region. Southern USAPI region tends to exhibit rainfall variability that is tied to ENSO. In the Southern USAPIs, El Niño is linked to dry conditions and La Niña to wet conditions in a way consistent with previously published works. The northern USAPIs on the other hand do not show such consistent impacts during different ENSO phases. In the Northern USAPIs, while many years conformed to the expected negative relationship between rainfall and ENSO (positive ONI value producing dry conditions and vice versa), a similar number of years appeared to have a positive relationship between rainfall and ENSO. Years with negative relationship between rainfall and ENSO in the Northern USAPIs are grouped into two categories, Cool Wet and Warm-dry, and together these categories are described here as having the Canonical response or simply as being Canonical years. Conversely, years with positive relationship between rainfall and ENSO are grouped into Cool Dry and Warm Wet years and these categories together are described as Non-Canonical.

We find that the 2013 disaster level drought that motivated this study and occurred during cool ENSO conditions was not an isolated event. Another 4 years in the 1979-2016 record were found for which dry conditions in the Northern USAPIs coincide with cool Pacific SST conditions. Along with 2013, these other 4 years compose the Non-Canonical Cool Dry category of events. According to reports collected in the PEAC Pacific ENSO Update newsletter, drought conditions similar to those observed in 2013 were also observed for other Non-Canonical Cool Dry years. In 2006, persistent dry conditions in Guam led to a longer than normal wildfire season resulting in evacuations, and very low levels in the Fena reservoir, which provides water to about 20% of the population. In Majuro, water rationing was mandated by February 2006 and encouraged though the following months. Similarly, in 2009 persistent dry conditions in the Northern RMI required the delivery of drinking water to Utirik in late April and the issuing of special weather statements for the RMI by the Guam National Weather Service Forecast Office. Both 2006 and 2009 are not classified as La Niña events but are singled out for analysis in our case.

Composites of SST, 850 hPa winds and geopotential height and precipitation were produced for the Canonical Warm Dry and Cool Wet as well as the Non-Canonical Warm Wet and Cool Dry years. Composite characteristics during the peak dry season as well as

the evolution of relevant anomaly fields show clear differences between Canonical and Non-Canonical events suggesting that the mechanisms that produce events in each category may be different.

During the Dec-May dry season, Canonical and Non-Canonical events show opposite circulation anomalies over the Tropical Western Pacific. Canonical Cool Wet events show cyclonic circulation anomalies over the western Pacific and rainfall anomaly patterns consistent with the commonly described as a La Nina horseshoe pattern, with wet anomalies to the west and dry anomalies over the central and eastern Pacific. Non-Canonical Cool Dry events on the other hand have anticyclonic anomalies over Tropical Western Pacific. This produces a dry anomaly belt that modifies the horseshoe pattern and extends dry conditions from the central and eastern Pacific into the Northern USAPIs. SST anomalies are also different, with Non-Canonical events characterized by strongest anomalies constrained to a narrow region of the Pacific near the dateline and Canonical showing anomalies further east. These differences are also true when comparing Canonical Warm Dry and Non-Canonical Warm Wet events, with circulation and rainfall anomalies being opposite.

Equally important as their differences during the Dry season, is the distinct way in which Canonical and Non-Canonical events evolve leading up to it. Canonical events are characterized by a transition of circulation anomalies over the western Pacific from the onset of the events in the spring to the Dec-May Dry season while Non-Canonical events are characterized by the persistence of these anomalies. For Canonical Cool Wet events, we observe that anticyclonic circulation anomalies are present in the western part of the basin in early spring and summer and transition to cyclonic anomalies as Dec-May approaches. Conversely, in Non-Canonical Cool Dry events western pacific anticyclonic anomalies are present as early as the preceding March-April-May season and linger through Dec-May. This makes long term persistent droughts brought on by Non-Canonical Cool Dry events a concern for the USAPIs. This difference in evolution is also reflected in SST anomalies, with east to west propagation observed in Canonical events and in situ development over the dateline observed for Non-Canonical events.

It is proposed here that the division of events into Canonical and Non-Canonical types can be integrated into the larger ENSO paradigms. It is thought that Non-Canonical events are the most extreme representatives of the Central Pacific type of ENSO events, with peak SST anomalies strongly restricted to the Central Pacific. Events with peak SST anomalies farther east compose the Canonical category. Because of this difference in SST anomaly location, Canonical and Non-Canonical events show a distinct seasonal evolution and set of

impacts.

Using the observed differences in the evolution of Canonical and Non-Canonical events, a forecasting methodology was developed to predict the Standardized Precipitation Index during both Canonical and Non-Canonical years. This methodology was based on a 2-tier approach. First, a logistic regression model was used to predict if a Dec-May season will be Canonical or Non-Canonical. Depending on the outcome of this classification forecast, one of two independent linear regression models, one for Canonical years and one for Non-Canonical years, is chosen to predict Dec-May SPI values. These linear regression models were trained on independent subsamples of the training set and with specific predictors crafter to capture the relevant features of the anomaly patterns of Canonical and Non-Canonical years.

Model performance was estimated over the training set (1950-2007) and an independent test set (2008-2017) which was excluded from all model training procedures. This 2-tier model approach provides a significant improvement of the SPI forecasts compared to a simple 1-tier model which predicts rainfall for all years with a single linear regression equation. Our 2-tier model also shows complementary skill to the PEAC Center operational forecasting suite of models witch performs poorly during Non-Canonical events but are well forecast in our 2-tier model. The success of this 2-tier approach highlights the importance of recognizing the existence of both Canonical and Non-Canonical modes of rainfall variability in the Western Pacific.

Previous published work links the rainfall variability and prediction in the USAPI exclusively to ENSO with dry conditions throughout the region occurring during El Niño and wet conditions occurring during La Niña. This understanding leaves little guidance for forecasting during neutral or weak ENSO conditions and does not provide context for years during which rainfall impacts are not as expected based on ENSO conditions. Our work proposes that differences in SST and circulation anomaly patterns, even during years that do not meet general criteria to be classified as ENSO events, can produce distinct and significant changes in local precipitation patterns that can produce severe drought across the USAPI. It is our hope that these new insights may provide PEAC Center forecasters with better tools in the forecasting of these types of dry event to improve our ability to deliver timely and accurate climate forecasts to the communities we serve.

APPENDIX A
COMPLETE PREDICTOR CORRELATION TABLES

	Dec-May	DJF	Central Pacific SST JJA(-1)	Central Pacific SST SON(-1)	MidLat Pacific SST JJA(-1)	MidLat Pacific SST SON(-1)	West Pac PSI JJA(-1)	West Pac PSI SON(-1)	South Pac PSI JJA(-1)	South Pac PSI SON(-1)	Equatorial Wind JJA(-1)	Equatorial Wind SON(-1)	South Pac Hgt Grad(-1)	South Pac Hgt Grad(-1)	Central Pacific SST ON-AS(-1)	West Pac PSI ON-AS(-1)	South Pac PSI ON-AS(-1)	Equatorial Wind ON-AS(-1)	East Pac Hgt ON-AS(-1)
Dec-May	1.000	-0.479	-0.430	-0.435	0.151	0.212	0.160	-0.103	-0.127	0.006	-0.457	-0.438	-0.390	-0.390	-0.422	-0.021	0.002	-0.438	0.298
DJF	-0.479	1.000	0.831	0.919	-0.254	-0.485	-0.262	0.026	0.477	0.087	0.825	0.870	0.845	0.845	0.905	-0.104	0.236	0.870	-0.399
Central Pacific SST JJA(-1)	-0.430	0.831	1.000	0.923	-0.278	-0.377	-0.236	0.028	0.275	0.041	0.835	0.832	0.842	0.842	0.930	-0.046	0.104	0.832	-0.166
Central Pacific SST SON(-1)	-0.435	0.919	0.923	1.000	-0.227	-0.404	-0.321	0.024	0.442	0.016	0.846	0.891	0.874	0.874	0.997	-0.077	0.115	0.891	-0.282
MidLat Pacific SST JJA(-1)	0.151	-0.254	-0.278	-0.227	1.000	0.672	-0.234	0.042	0.193	-0.239	-0.434	-0.333	-0.360	-0.360	-0.226	0.093	-0.092	-0.333	0.035
MidLat Pacific SST SON(-1)	0.212	-0.485	-0.377	-0.404	0.672	1.000	-0.072	0.036	-0.032	-0.143	-0.572	-0.521	-0.505	-0.505	-0.395	0.109	-0.117	-0.521	0.408
West Pac PSI JJA(-1)	0.160	-0.262	-0.236	-0.321	-0.234	-0.072	1.000	-0.177	-0.648	0.155	-0.206	-0.276	-0.281	-0.281	-0.332	-0.167	-0.045	-0.276	0.156
West Pac PSI SON(-1)	-0.103	0.026	0.028	0.024	0.042	0.036	-0.177	1.000	0.042	-0.469	0.081	-0.182	-0.044	-0.044	0.012	0.888	-0.291	-0.182	-0.004
South Pac PSI JJA(-1)	-0.127	0.477	0.275	0.442	0.193	-0.032	-0.648	0.042	1.000	-0.321	0.310	0.427	0.362	0.362	0.432	0.026	0.004	0.427	-0.286
South Pac PSI SON(-1)	0.006	0.087	0.041	0.016	-0.239	-0.143	0.155	-0.469	-0.321	1.000	0.003	0.202	0.085	0.085	0.022	-0.557	0.754	0.202	-0.108
Equatorial Wind JJA(-1)	-0.457	0.825	0.835	0.846	-0.434	-0.572	-0.206	0.081	0.310	0.003	1.000	0.893	0.869	0.869	0.841	-0.021	-0.013	0.893	-0.349
Equatorial Wind SON(-1)	-0.438	0.870	0.832	0.891	-0.333	-0.521	-0.276	-0.182	0.427	0.202	0.893	1.000	0.870	0.870	0.885	-0.272	0.251	1.000	-0.377
South Pac Hgt Grad(-1)	-0.390	0.845	0.842	0.874	-0.360	-0.505	-0.281	-0.044	0.362	0.085	0.869	0.870	1.000	1.000	0.877	-0.138	0.104	0.870	-0.223

South Pac Hgt Grad(-1)	-0.390	0.845	0.842	0.874	-0.360	-0.505	-0.281	-0.044	0.362	0.085	0.869	0.870	1.000	1.000	0.877	-0.138	0.104	0.870	-0.223
Central Pacific SST ON-AS(-1)	-0.422	0.905	0.930	0.997	-0.226	-0.395	-0.332	0.012	0.432	0.022	0.841	0.885	0.877	0.877	1.000	-0.077	0.106	0.885	-0.253
West Pac PSI ON-AS(-1)	-0.021	-0.104	-0.046	-0.077	0.093	0.109	-0.167	0.888	0.026	-0.557	-0.021	-0.272	-0.138	-0.138	-0.077	1.000	-0.406	-0.272	0.221
South Pac PSI ON-AS(-1)	0.002	0.236	0.104	0.115	-0.092	-0.117	-0.045	-0.291	0.004	0.754	-0.013	0.251	0.104	0.104	0.106	-0.406	1.000	0.251	-0.172
Equatorial Wind ON-AS(-1)	-0.438	0.870	0.832	0.891	-0.333	-0.521	-0.276	-0.182	0.427	0.202	0.893	1.000	0.870	0.870	0.885	-0.272	0.251	1.000	-0.377
East Pac Hgt ON-AS(-1)	0.298	-0.399	-0.166	-0.282	0.035	0.408	0.156	-0.004	-0.286	-0.108	-0.349	-0.377	-0.223	-0.223	-0.253	0.221	-0.172	-0.377	1.000

Table A.1: Correlation Table for all predictors considered for the benchmark 1 tier model

APPENDIX B
SUPPLEMENTAL FIGURES

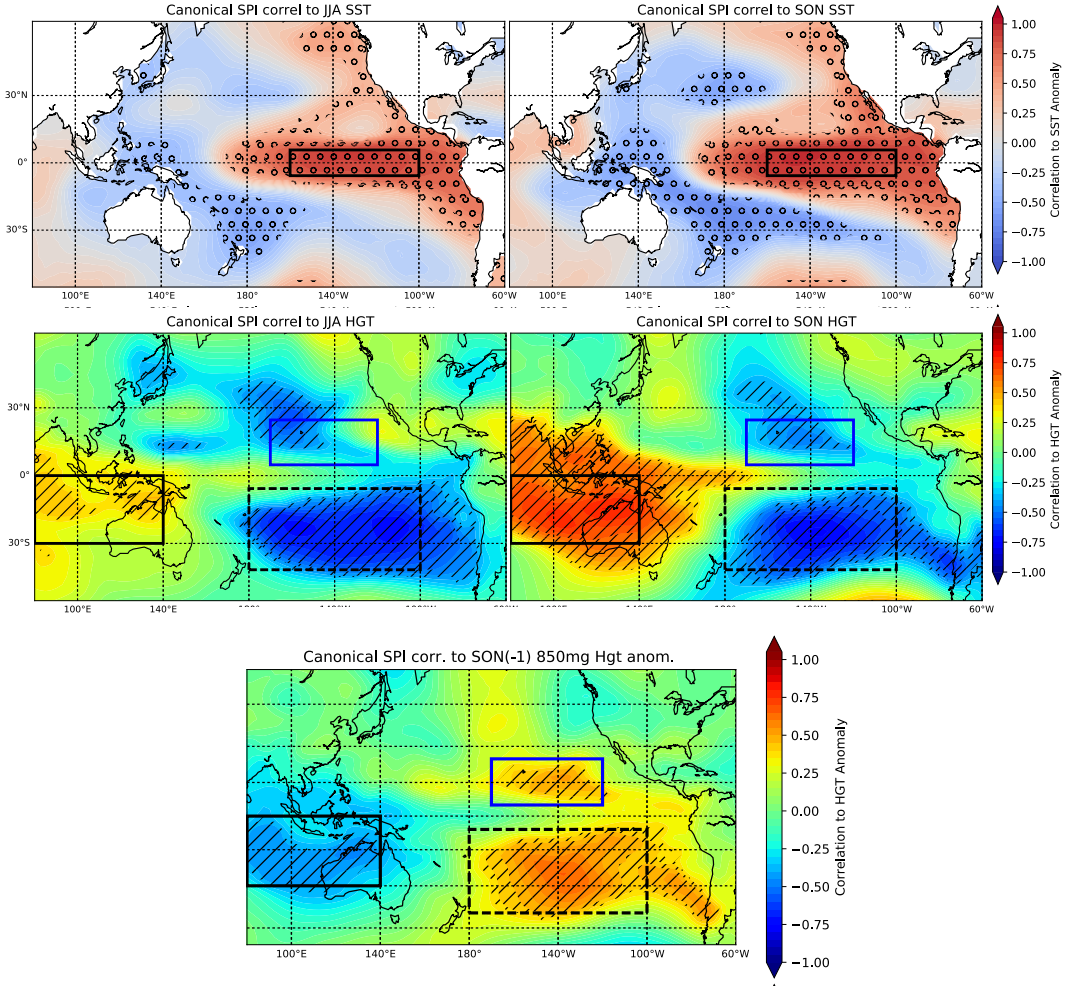


Figure B.1: Correlation coefficient maps of DJF Oceanic Niño Index for years in the Canonical training set and SST (top), 850mb Geopotential Height (middle). The bottom panel represent the correlation between Dec-May SPI for years in the Canonical training set and 850mb Geopotential Height. Hashing represents significance at the 95% test level for a sample size of 17 years. Boxes represent regions chosen for predictor index average calculations (more details in text).

Subsequent chapters are labeled with letters of the alphabet.

BIBLIOGRAPHY

Adler, R. F., and Coauthors. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979 to present). *Journal of Hydrometeorology*, 4(6):1147–1167, 2003. doi: 10.1175/1525-7541(2003)004⟨1147:TVGPCP⟩2.0.CO;2. URL https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Arthur, D., and S. Vassilvitskii. k-means++: the advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007.

Bamston, A. G., M. Chelliah, and S. B. Goldenberg. Documentation of a highly ensorelated sst region in the equatorial pacific: Research note. *Atmosphere-Ocean*, 35(3):367–383, 1997. doi: 10.1080/07055900.1997.9649597. URL https://doi.org/10.1080/07055900.1997.9649597.

Bjerknes, J. A possible response of the atmospheric hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18:820829, 1966.

Bjerknes, J. Atmospheric teleconnections from the equatorial pacific. *Monthly Weather Review*, 97:163172, 1969.

Blair, C., and J. J. Higgins. A comparison of the power of wilcoxon's rank-sum statistic to that of student's t statistic under various nonnormal distributions. *Journal of Educational Statistics*, 5(4):309–335, 1980. ISSN 03629791. URL http://www.jstor.org/stable/1164905.

BOM-CSIRO. Climate change in the pacific: Scientific assessment and new research. volume 1: Regional overview. volume 2: Country reports. Technical report, Australian Bureau of Meteorology and CSIRO, 2011. URL https://www.pacificclimatechangescience.org/publications/reports/report-climate-change-in-the-pacific-scientific-assessment-and-new-research/.

Boyin, H., and Coauthors. Extended reconstructed sea surface temperature version 4 (ERSST.v4). part i: Upgrades and intercomparisons. *Journal of Climate*, 28(3):911–930, 2015. doi: 10.1175/JCLI-D-14-00006.1. URL https://doi.org/10.1175/JCLI-D-14-00006.1.

Camargo, S. J., and A. H. Sobel. Western north pacific tropical cyclone intensity and enso. *Journal of Climate*, 18(15):2996–3006, 2005. doi: 10.1175/JCLI3457.1. URL https://doi.org/10.1175/JCLI3457.1.

Capotondi, A. ENSO diversity in the NCAR CCSM4 climate model. *Journal of Geophysical Research: Oceans*, 118(10):4755–4770, 2013. ISSN 2169-9291. doi: 10.1002/jgrc.20335. URL http://dx.doi.org/10.1002/jgrc.20335.

Capotondi, A., and Coauthors. Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, 96(6):921–938, 2015. doi: 10.1175/BAMS-D-13-00117.1. URL https://doi.org/10.1175/BAMS-D-13-00117.1.

Chia, H. H., and C. F. Ropelewski. The interannual variability in the genesis location of tropical cyclones in the northwest pacific. *Journal of Climate*, 15(20):2934–2944, 2002. doi: 10.1175/1520-0442(2002)015⟨2934:TIVITG⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0442(2002)015<2934:TIVITG>2.0.CO;2.

Chowdhury, M., P.-S. Chu, and T. Schroeder. ENSO and seasonal sea level variabilitya diagnostic discussion for the u.s.-affiliated pacific islands. *Theoretical and Applied Climatology*, 88:213224, 2007.

Collins, D., A. S. Gupta, and S. Power. *Regional Overview, Vol. 1, Climate Change in the Pacific: Scientific Assessment and New Research.*, volume 1. CSIRO, 2011.

Davies, D., and D. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligences*, page 1224227, 1979.

Draper, N., and H. Smith. *Applied regression analysis*. Number v. 1 in Wiley series in probability and statistics: Texts and references section. Wiley, 1981. ISBN 9780471170822. URL https://books.google.com/books?id=8n8pAQAAMAAJ.

Efron, B., and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 9(3):586–596, 05 1981. doi: 10.1214/aos/1176345462. URL https://doi.org/10.1214/aos/1176345462.

Goutte, C., L. K. Hansen, M. G. Liptrot, and E. Rostrup. Feature-space clustering for fmri meta-analysis. *Human Brain Mapping*, 13(3):165–183, 2001. ISSN 1097-0193. doi: 10.1002/hbm.1031. URL http://dx.doi.org/10.1002/hbm.1031.

Guttman, N. B. Comparing the palmer drought index and the standardized precipitation index1. *JAWRA Journal of the American Water Resources Association*, 34(1):113–121, 1998. ISSN 1752-1688. doi: 10.1111/j.1752-1688.1998.tb05964.x. URL http://dx.doi.org/10.1111/j.1752-1688.1998.tb05964.x.

Hayes, M. J., M. D. Svoboda, D. A. Wilhite, and O. V. Vanyarkho. Monitoring the 1996 drought using the standardized precipitation index. *Bulletin of the American Meteorological Society*, 80(3):429–438, 1999. doi: 10.1175/1520-0477(1999)080⟨0429:MTDUTS⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0477(1999)080<0429:MTDUTS>2.0.CO;2.

Jin, F.-F. An equatorial ocean recharge paradigm for ENSO. part i: Conceptual model. *Journal of the Atmospheric Sciences*, 54(7):811–829, 1997. doi: 10.1175/1520-0469(1997)054⟨0811:AEORPF⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.

Jin, F.-F., and S.-I. An. Thermocline and zonal advective feedbacks within the equatorial ocean recharge oscillator model for ENSO. *Geophysical Research Letters*, 26(19):2989–2992, 1999. ISSN 1944-8007. doi: 10.1029/1999GL002297. URL http://dx.doi.org/10.1029/1999GL002297.

Kalnay, E., and Coauthors. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471, 1996. doi: 10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kao, H.-Y., and J.-Y. Yu. Contrasting eastern-pacific and central-pacific types of el nio. *Journal of Climate*, 22:615632, 2009.

Keener, V., K. Hamilton, S. Izuka, and coathors. Part 8. climate of pacific islands. Technical report, NOAA NESDIS, 2013. URL https://www.nesdis.noaa.gov/sites/default/files/asset/document/NOAA_NESDIS_Tech_Report_142-8-Climate_of_the_Pacific_Islands.pdf.

Ketchen, D., and C. Shook. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6):441–458, 1996. ISSN 1097-0266. doi: 10.1002/(SICI)1097-0266(199606)17:6⟨441::AID-SMJ819⟩3.0.CO;2-G. URL http://dx.doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

Keyantash, J., and J. A. Dracup. The quantification of drought: An evaluation of drought indices. *Bulletin of the American Meteorological Society*, 83(8):1167–1180, 2002. doi: 10.1175/1520-0477(2002)083⟨1191:TQODAE⟩2.3.CO;2. URL https://doi.org/10.1175/1520-0477(2002)083<1191:TQODAE>2.3.CO;2.

Keyantash, J., and N. Staff. The climate data guide: Standardized precipitation index (spi). https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi, 2016.

Kug, J.-S., F. F. Jin, and S.-I. An. Two types of el niño events: Cold tongue el niño and warm pool el niño. *Journal of Climate*, 22:14991515, 2009.

Kug, J.-S., J. Choi, S.-I. An, F.-F. Jin, and A. T. Wittenberg. Warm pool and cold tongue el nio events as simulated by the GFDL 2.1 coupled GCM. *Journal of Climate*, 23 (5):1226–1239, 2010. doi: 10.1175/2009JCLI3293.1. URL https://doi.org/10.1175/2009JCLI3293.1.

Lander, M. A. An exploratory analysis of the relationship between tropical storm formation in the western north pacific and ENSO. *Monthly Weather Review*, 122:636651, 1994.

Li, T., B. Wang, and L. Wang. Comments on combination mode dynamics of the anomalous northwest pacific anticyclone. *Journal of Climate*, 29(12):4685–4693, 2016. doi: 10.1175/JCLI-D-15-0385.1. URL https://doi.org/10.1175/JCLI-D-15-0385.1.

Llet, R., M. Ortiz, L. Sarabia, and M. Snchez. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*, 515 (1):87 – 100, 2004. ISSN 0003-2670. doi: https://doi.org/10.1016/j.aca.2003.12.020. URL http://www.sciencedirect.com/science/article/pii/S0003267003016246. Papers presented at the 5th COLLOQUIUM CHEMIOMETRICUM MEDITERRANEUM.

Lloyd, S. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

Mantua, N. J., 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*, 78(6):1069–1079, 1997. doi: 10.1175/1520-0477(1997)078⟨1069:APICOW⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Mathworks. Matlab stepwise regression documentation. https://www.mathworks.com/help/stats/stepwisefit.html, 2017. Accessed: 2017-09-25.

McKee, T., N. Doesken, and J. Kliest. The relationship of drought frequency and duration to time scales. *In Proceedings of the 8th Conference of Applied Climatology*, pages 17–22, January 1993.

McPhaden, M. J., S. E. Zebiak, and M. H. Glantz. ENSO as an integrating concept in earth science. *Science*, 314:17401745, 2006.

Murphy, B. F., S. B. Power, and S. McGree. The varied impacts of el niño-southern oscillation on pacific island climates. *Journal of Climate*, 27(11):4015–4036, 2014. doi: 10.1175/JCLI-D-13-00130.1. URL https://doi.org/10.1175/JCLI-D-13-00130.1.

NDMC, T. N. D. M. C. Interpretation of 6-month standardized precipitation index map, 2017. URL \url{http://drought.unl.edu/MonitoringTools/ClimateDivisionSPI/Interpretation/6month.aspx}. [Online; accessed 13-Nov-2017].

Newman, M., M. A. Alexander, and J. D. Scott. An empirical model of tropical ocean dynamics. *Climate Dynamics*, 37(9):1823, Mar 2011a. ISSN 1432-0894. doi: 10.1007/s00382-011-1034-0. URL https://doi.org/10.1007/s00382-011-1034-0.

Newman, M., S.-I. Shin, and M. A. Alexander. Natural variation in ENSO flavors. *Geophysical Research Letters*, 38(14):n/a–n/a, 2011b. ISSN 1944-8007. doi: 10.1029/2011GL047658. URL http://dx.doi.org/10.1029/2011GL047658. L14705.

Panofsky, H., G. Brier, and P. S. C. S. of Mineral Industries. Mineral Industries Extension Services. *Some applications of statistics to meteorology*. Penn. State University, College of Earth and Mineral Sciences, 1953.

Picard, R. R., and R. D. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984. doi: 10.1080/01621459.1984.10478083. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10478083.

Pinto, T., G. Santos, L. Marques, T. M. Sousa, I. Praa, Z. Vale, and S. L. Abreu. *Distributed Computing and Artificial Intelligence, 12th International Conference*, chapter Solar Intensity Characterization Using Data-Mining to Support Solar Forecasting. Advances in Intelligent Systems and Computing. Springer International Publishing, 2015. ISBN 9783319196381. URL https://books.google.com/books?id=m7G4CQAAQBAJ.

Power, S. B., and I. N. Smith. Weakening of the walker circulation and apparent dominance of el nio both reach record levels, but has ENSO really changed? *Geophysical Research Letters*, 34(18):n/a–n/a, 2007. ISSN 1944-8007. doi: 10.1029/2007GL030854. URL http://dx.doi.org/10.1029/2007GL030854. L18702.

Ramage, C. *Monsoon Meteorology*. International geophysics series. Academic Press, 1971. URL https://books.google.com/books?id=i5kRAQAAIAAJ.

Ropelewski, C. F., and M. S. Halpert. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, 115(8):1606–1626, 1987. doi: 10.1175/1520-0493(1987)115⟨1606:GARSPP⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

Sadler, J. C. Mid-season typhoon development and intensity changes and the tropical upper tropospheric trough. *Monthly Weather Review*, 106(8):1137–1152, 1978. doi: 10.1175/1520-0493(1978)106⟨1137:MSTDAI⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0493(1978)106<1137:MSTDAI>2.0.CO;2.

Schroeder, T. A., M. R. Chowdhury, M. A. Lander, C. C. Guard, C. Felkley, and D. Gifford. The role of the pacific ENSO applications climate center in reducing vulnerability to climate hazards: Experience from the u.s.-affiliated pacific islands. *Bulletin of the American Meteorological Society*, 93(7):1003–1015, 2012. doi: 10.1175/BAMS-D-11-00109.1. URL https://doi.org/10.1175/BAMS-D-11-00109.1.

scikit-learn developers. Cross-validation: evaluating estimator performance. linear regression documentation. http://scikit-learn.org/stable/modules/cross_validation.html#cross-validation, 2017a. Accessed: 2017-11-18.

scikit-learn developers. Scikit-learn linear regression documentation. http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html, 2017b. Accessed: 2017-09-25.

scikit-learn developers. Logistic regression documentation. http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html, 2017c. Accessed: 2017-09-25.

scikit-learn developers. Logistic function example. http://scikit-learn.org/stable/auto_examples/linear_model/plot_logistic.html#

sphx-glr-auto-examples-linear-model-plot-logistic-py, 2017d. Accessed: 2017-09-25.

scipy community. Scipy linear regression (numpy.linalg.lstsq) documentation. https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.linalg.lstsq.html, 2017. Accessed: 2017-09-25.

Shea, E., and Coauthors. Preparing for a changing climate: The potential consequences of climate variability and change. Technical report, University of Hawaii at Manoa, 2001. URL http://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/3404/pacific-climate-assessment.pdf.

Stuecker, M. F., A. Timmermann, F.-F. Jin, S. McGregor, and H.-L. Ren. A combination mode of the annual cycle and the elniño/southern oscillation. *Nat Geos.*, 6:540, 2013. doi: 10.1038/ngeo1826. URL http://dx.doi.org/10.1038/ngeo1826.

Stuecker, M. F., F.-F. Jin, A. Timmermann, and S. McGregor. Combination mode dynamics of the anomalous northwest pacific anticyclone. *Journal of Climate*, 28(3):1093–1111, 2015. doi: 10.1175/JCLI-D-14-00225.1. URL https://doi.org/10.1175/JCLI-D-14-00225.1.

Stuecker, M. F., F.-F. Jin, A. Timmermann, and S. McGregor. Reply to comments on combination mode dynamics of the anomalous northwest pacific anticyclone. *Journal of Climate*, 29(12):4695–4706, 2016. doi: 10.1175/JCLI-D-15-0558.1. URL https://doi.org/10.1175/JCLI-D-15-0558.1.

Svoboda, M., M. Hayes, and D. Wood. Standardized precipitation index user guide. Technical Report WMO-No. 1090, World Meteorological Organization, Geneva, 2012.

Trenberth, K. E. The definition of el nio. *Bulletin of the American Meteorological Society*, 78 (12):2771–2777, 1997. doi: 10.1175/1520-0477(1997)078⟨2771:TDOENO⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

Wang, B., J. Lee, and I. Kang. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (19802004). *Climate Dynamics*, 33(93):40154036, 2009.

Wang, B., and S.-I. An. A method for detecting season-dependent modes of climate variability: S-EOF analysis. *Geophysical Research Letters*, 32(15):n/a–n/a, 2005. ISSN 1944-8007. doi: 10.1029/2005GL022709. URL http://dx.doi.org/10.1029/2005GL022709. L15710.

Wang, B., and J. C. L. Chan. How strong enso events affect tropical storm activity over the western north pacific. *Journal of Climate*, 15(13):1643–1658, 2002. doi: 10.1175/1520-0442(2002)015⟨1643:HSEEAT⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0442(2002)015<1643:HSEEAT>2.0.CO;2.

Wang, B., and Q. Ding. Global monsoon: Dominant mode of annual variation in the tropics. *Dynamics of Atmospheres and Oceans*, 44(3):165 – 183, 2008. ISSN 0377-0265. doi: https://doi.org/10.1016/j.dynatmoce.2007.05.002. URL http://www.sciencedirect.com/science/article/pii/S0377026508000055. Current Contributions to Understanding the General Circulation of the Atmosphere Part 2.

Wang, B., and Q. Zhang. Pacific-east asian teleconnection. part ii: How the Philippine Sea anomalous anticyclone is established during el niño development. *Journal of Climate*, 15 (22):3252–3265, 2002. doi: 10.1175/1520-0442(2002)015⟨3252:PEATPI⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0442(2002)015<3252:PEATPI>2.0.CO;2.

Wikipedia. Gamma distribution, 2017. URL \url{https://en.wikipedia.org/wiki/Gamma_distribution}. [Online; accessed 13-Nov-2017].

Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. ISSN 00994987. URL http://www.jstor.org/stable/3001968.

Wilks, D. *Statistical Methods in the Atmospheric Sciences*. International Geophysics. Elsevier Science, 2005. ISBN 9780080456225. URL https://books.google.com/books?id=_b4R9j2Iy7EC.

Wu, H., M. Hayes, A. Weiss, and Q. Hu. An evaluation of the standardized precipitation index, the china-z index and the statistical z-score. *International Journal of Climatology*, 21(6):745–758, 2001. ISSN 1097-0088. doi: 10.1002/joc.658. URL http://dx.doi.org/10.1002/joc.658.

Yeh, S.-W., J.-S. Kug, and S.-I. An. Recent progress on two types of el niño: Observations, dynamics, and future changes. *Asia-Pacific Journal of Atmospheric Sciences*, 50(1):69–81,

Jan 2014. ISSN 1976-7951. doi: 10.1007/s13143-014-0028-3. URL https://doi.org/10.1007/s13143-014-0028-3.

Yim, S.-Y., B. Wang, and W. Xing. Prediction of early summer rainfall over south china by a physical-empirical model. *Climate Dynamics*, 43(7):1883–1891, Oct 2014a. ISSN 1432-0894. doi: 10.1007/s00382-013-2014-3. URL https://doi.org/10.1007/s00382-013-2014-3.

Yim, S.-Y., B. Wang, W. Xing, and M.-M. Lu. Prediction of meiyu rainfall in taiwan by multi-lead physical–empirical models. *Climate Dynamics*, 44(11):3033–3042, Jun 2015. ISSN 1432-0894. doi: 10.1007/s00382-014-2340-0. URL https://doi.org/10.1007/s00382-014-2340-0.

Yim, S., B. Wang, and Q. Weng. Prediction of early summer rainfall over south china by a physical-empirical model. *Climate Dynamics*, 43(7):1883–1891, 2014b.

Yu, J.-Y., H.-Y. Kao, and T. Lee. Subtropics-related interannual sea surface temperature variability in the central equatorial pacific. *Journal of Climate*, 23(11):2869–2884, 2010. doi: 10.1175/2010JCLI3171.1. URL https://doi.org/10.1175/2010JCLI3171.1.

Yu, L., R. A. Weller, and W. T. Liu. Case analysis of a role of enso in regulating the generation of westerly wind bursts in the western equatorial pacific. *Journal of Geophysical Research: Oceans*, 108(C4):n/a–n/a, 2003. ISSN 2156-2202. doi: 10.1029/2002JC001498. URL http://dx.doi.org/10.1029/2002JC001498. 3128.

Yu, Z.-P., P.-S. Chu, and T. Schroeder. Predictive skills of seasonal to annual rainfall variations in the u.s.-affiliated pacific islands: Canonical correlation analysis and multivariate principal component regression approaches. *Journal of Climate*, 10:25862599, 1997a.

Yu, Z.-P., P.-S. Chu, and T. Schroeder. Predictive skills of seasonal to annual rainfall variations in the u.s. affiliated pacific islands: Canonical correlation analysis and multivariate principal component regression approaches. *Journal of Climate*, 10(10):2586–2599, 1997b. doi: 10.1175/1520-0442(1997)010⟨2586:PSOSTA⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0442(1997)010<2586:PSOSTA>2.0.CO;2.

Yun, K.-S., K.-J. Ha, S.-W. Yeh, B. Wang, and B. Xiang. Prediction of meiyu rainfall in taiwan by multi-lead physical-empirical models. *Climate Dynamics*, 44(7):1979–1992, 2015.