Title
Empirical approaches for near-term climate predictions

Author
Faggiani Dias, Daniela

Publication Date
2020

UNIVERSITY OF CALIFORNIA SAN DIEGO

Empirical approaches for near-term climate predictions

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Daniela Faggiani Dias

Committee in charge:

Arthur J. Miller, Chair
Daniel R. Cayan, Co-Chair
Ian Eisenman
Amato Evan
Victor Ferreira
Sarah T. Gille

2020

The dissertation of Daniela Faggiani Dias is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2020

DEDICATION

To my parents, for their tireless effort to ensure my education, for their constant support and for their immeasurable love

EPIGRAPH

Our economic system and our planetary system are now at war. Or, more accurately, our economy is at war with many forms of life on earth, including human life. What the climate needs to avoid collapse is a contraction in humanity's use of resources; what our economic model demands to avoid collapse is unfettered expansion. Only one of these sets of rules can be changed, and it's not the laws of nature.

—Naomi Klein

A gente sabe que a gente está ativa,
está militando,
está resistindo
o tempo todo.

—Mariele Franco, presente

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

In times when individualism and selfishness seems to reign and when walls are built to isolate people in their mediocrity, I was very luck to have encountered an extreme helpful and collaborative community during these years of my PhD. I wish every graduate student had the same support, encouragement and opportunities that my advisors, mentors and collaborators provided me.

I thank my PhD advisor, Art Miller for his continued support as a mentor, giving me the independence to pursue my research interests, but helping me with careful insights and critiques. His curiosity-driven approach to the scientific questions and his genuine caring about the well-being of the others are values that really inspire me and I am going to carry them for my life. I am also so grateful to have had the opportunity to work with my PhD co-advisor, Dan Cayan. His patient and thoughtful insights and his way of thinking about problem-solving were fundamental to the development of this thesis and for my development as a scientist. I am also deeply grateful to my other mentors and close collaborators, Aneesh Subramanian and Sasha Gershunov. I thank Aneesh for his careful comments, for his patience in teaching me basic concepts and analysis since the beginning of my PhD and, foremost, for his kindness; and I am highly indebted to Sasha for his time in advising and teaching me and for his continued encouragement. They all made the road towards the end of my PhD much less bumpy.

I would also like to thank my other co-authors and committee members. I am thankful to Laure Zanna for her comments, ideas and reviews as well as for her patience to teach me about the nuances of the LIM; to Steve Yeager for all the help with the CESM-DPLE analysis and very careful and thoughtful insights; to Antonietta Capotondi for all the funny discussions that really shaped part of this work; and to those at NCAR that made my stay in Boulder extremely delightful (so much that I am coming back!). Thanks to Amato Evan, Sarah Gille, Ian Eisenman and Victor Ferreira. They all offered invaluable insights, while always challenging my ideas to make me think from a different perspective.

This work has been enabled by funding from many different sources: the science without borders program by the Brazilian National Council for Scientific and Technological Development (CNPq, 221222/2014-6), the National Science Foundation (OCE1419306), the National Oceanic and Atmospheric Administration (NOAA-MAPP; NA17OAR4310106), the NOAA's Regional Integrated Sciences and Assessments (RISA) California–Nevada Climate Applications Program (NA17OAR4310284), the California Energy Commission (PIR-15-005), and the Southwest Climate Adaptation Science Center (G18AC00320). I am also thankful to the Advanced Student Program from the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Thanks also to the SIO Graduate Department, the CASPO business office and the SIO Communications Office for all the support and opportunities offered along those years.

I thank the great community of friends that I have found here in San Diego. Thanks to my office mate, Nathali. Sharing an office with her (and many coffees) has made the process much more fun; thanks to Leticia, my roommate since the first day we arrived in the US for all the experiences shared, all the endless conversations and, more important, for all the food she cooked for me. Alfredo, Tiaguinho, Gabo, Gabriel, Tashiana, Liz, Jess, Emma, Rishi, Eriquita, Alyssa, Manuel: I am also extremely grateful to have met such wonderful people who became my dearest friends. I have learned so much with them as they have constantly restored my hope and my optimism for a better and fair world for everyone. Thanks also to Jenny Miller for making me feel at home since the first day we met. Thanks to my partner and friend, Lorenzo, for always giving me the best reasons to not work on weekends.

A special thanks to my family which did not really understand what I was doing and why I had to go so far away for that, but still gave me a lot of joy, support and love, unconditionally.

Finally, I respectfully acknowledge that UCSD occupies the territory of Kumeyaay Nation to whom I owe honor and gratitude.

Chapter 2, in full, is a reprint of the material as it appears in Dias, D.F., A. Subramanian, L. Zanna, and A. J. Miller (2018), Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study, Climate Dynamics. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is in revision for publication in *Journal of Climate*. Dias, D.F., A. Gershunov, D. Cayan and A.J. Miller (2020), The influence of sea surface temperature and soil moisture in seasonal predictions of air temperature over Western North America. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is in preparation for submission for publication in the *Proceedings of the National Academy of Sciences of the United States of America* as: Dias, D.F., A. Gershunov and D. Cayan (2020), A statistical benchmark for NMME: the case of seasonal predictions of minimum and maximum temperature over the Western North America. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, is in preparation for submission for publication as: Dias, Daniela Faggiani; Yeager, Stephen; Miller, Arthur J.; Capotondi, Antonietta; Cayan, Daniel R. Exploring sources of errors in decadal predictions of the Pacific surface and sub-surface ocean. The dissertation author was the primary investigator and author of this paper.

VITA

2010	B. S. in Oceanography, University of Monte Serrat, Brazil
2013	Master of Science in Remote Sensing, National Institute for Space Research, Brazil
2013-2015	Research Assistant, National Institute for Space Research, Brazil
2015-2020	Graduate Student Researcher, University of California, San Diego
2020	Doctor of Philosophy in Oceanography, University of California San Diego

PUBLICATIONS

Dias, D. F., D. R. Cayan, A. Gershunov, and A. J. Miller, 2020: The influence of sea surface temperature and soil moisture in seasonal predictions of air temperature over Western North America, *Journal of Climate*, **sub judice**

Cavole, L. M., S. Andrade-Vera, J. Marín-Jarrín, **D. F. Dias**, O. Aburto-Oropeza, and M. J. Barragán-Paladines, 2020: Using local ecological knowledge of fishers to infer the impact of climate variability in Galapagos small-scale fisheries, *Marine Policy*, **sub judice**

Capotondi, A., and Coauthors, 2019: Observational needs supporting marine ecosystems modeling and forecasting: From the global ocean to regional and coastal systems. *Frontiers in Marine Science*, 6, 623

Dias, D. F., A. Subramanian, L. Zanna, and A. J. Miller, 2018: Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study. *Climate Dynamics*, doi: 10.1007/s00382-018-4323-z

Dias, D. F., D. R. Cayan, and A. Gershunov, 2018: Statistical Prediction of Minimum and Maximum Air Temperature in California and Western North America. *California's Fourth Climate Change Assessment, California Energy Commission*, doi: CCCA4-CEC-2018-011

Gouveia, M. B., D. F. Gherardi, C. A. Lentini, **D. F. Dias**, and P. C. Campos, 2017: Do the Brazilian sardine commercial landings respond to local ocean circulation? *PLoS ONE*, 12 (5), 1–19, doi: 10.1371/journal.pone.0176808

Dias, D. F., L. P. Pezzi, D. F. M. Gherardi, and R. Camargo, 2014: Modeling the spawning strategies and larval survival of the Brazilian sardine (*Sardinella brasiliensis*). *Progress in Oceanography*, 123, 38–53, doi: 10.1016/j.pocean.2014.03.009

ABSTRACT OF THE DISSERTATION

Empirical approaches for near-term climate predictions

by

Daniela Faggiani Dias

Doctor of Philosophy in Oceanography

University of California San Diego, 2020

Arthur J. Miller, Chair
Daniel R. Cayan, Co-Chair

Climate variations on seasonal to decadal time scales can have enormous social, economical and environmental impacts. As such, the ability to make skilful and reliable climate predictions at these time scales offers many benefits for climate preparedness, adaptation and resilience. In the recent years, major progress has been made in the development of such predictions with the advent of simulations with global climate models that are initialized from the current climate state. However, many challenges remain including an understanding of the underlying physical mechanisms for skilful predictions and whether such predictions could be improved. The purpose of this thesis is to establish

new benchmarks for seasonal to decadal predictions in diverse components of the climate system and to provide some pieces of evidence that help to understand what are the drivers for these predictable patterns. Specifically, we use a suite of empirical models to perform predictions of oceanic and atmospheric variables together with initialized climate predictions to: 1. Understand the contribution of remote and local factors to the predictability of North and Tropical Pacific Oceans Sea Surface Temperature and Land Surface Temperature over Western North America; 2. Provide a higher baseline level skill for the state-of-art global prediction systems, from seasonal to decadal time scales; 3. Explore possible sources of errors in the global climate model simulations using statistical predictive models.

First, we isolate contributions to the forecast skill from different spatial and time scales in the Pacific Ocean using a Liner Inverse Modelling (LIM) approach, showing the importance of temporal scale interactions in improving the predictions on decadal time scales. Specifically, we show that the Extratropical North Pacific is a source of predictability for the tropics on seasonal to interannual time scales, while the tropics enhance the forecast skill for the decadal component. We then show that the skill for an empirically-built LIM is comparable to and sometimes better than that from two state-of-art global prediction systems, from seasonal to decadal timescales and for several regions around the globe. These results indicate that the evolution of the system in those areas may not be not fully driven by unpredictable dynamics and that there may be some room for improvement in the dynamical models predictions, given that a low-dimensional linear model is able to generate better skill than the fully-coupled nonlinear model. Bearing that in mind, we use the LIM linear feedback matrix to explore possible sources of errors in the dynamical model simulations and we find that some of the simulated atmospheric and oceanic local and remote feedbacks differ in several key regions from that obtained with observations. These results may indicate sources of error in the dynamical models and therefore in its prediction skill that merit focused attention.

We then investigate the role of remote and local predictors in seasonal predictors of minimum and maximum air temperatures over the Western North America, using a Canonical Correlation Analysis approach. We show that remote predictors, in the form of Pacific climate modes, provide the best predictive skill for temperature over land, particularly during wintertime. Lastly, considering that persistence is the widely-used measure when evaluating the predictive skill for dynamical models, we suggest the use of CCA as a much higher benchmark for seasonal predictions of land surface air temperatures.

Chapter 1

Introduction

The perfect harmony among the components of the Earth's climate provides just the right conditions to support life on this planet. Subtle changes in this delicate synchrony could have profound impacts on the diverse forms of life that it sustains, well beyond humankind. The indisputable evidence of anthropogenic-caused climate change shows that this balance in the climate system has been disturbed and the impacts are already felt. Therefore, understanding the complex interplay among the different components of this system to be able to anticipate its fluctuations has been, for centuries, a fundamental driver of inquiry across different societies.

All over the world, and since the beginning of civilization, people depend on or are influenced by weather and climate fluctuations to a varying extent. Around the Arctic circle, in the Inuit Nunangat (also known as Northern Canada), Inuit people have been living and coping with that rough environment for almost 4,000 years. For them, being able to predict the weather is a determining factor for their survival. Based on centuries of wisdom, careful observation techniques and a constant refinement in their knowledge, they have developed a deep understanding of the weather patterns and how the environment changes with them (Weatherhead et al., 2010). Far away from the Arctic civilizations, around 650 B.C., the Babylonians were also observing and recording weather patterns, particularly changes in the clouds, to predict short-term changes in the weather. In 340 B.C., the Greek philosopher Aristotle wrote *Meteorologica* which included his philosophical ideas about several aspects of the climate, such as formation of rain, clouds and winds.

It was not until about the 17th century, with the birth of meteorological instruments, that the records became quantifiable and theories for the observed patterns could be proposed, such as the cause of the Trade Winds by Edmond Halley in 1686. With the invention of the telegraph in 1835, the record of instantaneous collections of weather observations around different regions became possible. Robert Fitzroy, a true pioneer in the prediction science, made the use of the available technology to develop more accurate and advanced weather predictions than his contemporaries. The first attempt of a prediction through solving the equations of motion was made by Lewis Fry Richardson in 1922, but it was not until the birth of the electronic computer that such numerical weather predictions became practical. Since then, the interest in estimating future weather and climate conditions has increased. Lorenz (1963) demonstrated that chaotic dynamics of the climate system may set bounds on the predictability of weather and climate. As he pointed out: *"In view of the inevitable inaccuracy and incompleteness of weather observations, precise very-long-range forecasting would seem to be non-existent"*. As a result, if other components of the climate system such as the surface ocean or soil moisture are not considered, the limit of predictability for many aspects of the weather is about 2 weeks (Kirtman et al., 2013). So why are predictions beyond the weather predictability limit even possible? Although Lorenz (1963) demonstrated that there is this limit to precise predictions, he also adverted *"There remains the very important question as to how long is very-long-range. Conceivably it could be a few days or a few centuries"*. The scientific basis for climate predictions at longer lead times arises from the fact that variations in the slowly changing components of the climate system can affect atmospheric circulation and, accordingly, the climate at the Earth's surface. These slowly-changing components of the climate system are able to exert a sustained influence on climate anomalies that can extend over seasons or years, far beyond the limit of atmospheric predictability from initial conditions alone.

In recent years, near-term climate predictions (NTCP) has gained increased attention in the scientific community because of its considerable importance for many societal applications. Therefore, the World Climate Research Programme (WCRP) has recently recognized NTCP as one of the biggest challenges facing international climate research community (Boer et al., 2016). Near-term refers to predictions made from one year to several decades, in a period that the climate projections is generally similar (Kirtman et al., 2013). Different from weather forecasts, NTCP aim to provide information about the future evolution of the statistics of the climate system (Meehl et al., 2014); different from climate projections, which are run by prescribing external forcing, NTCP include information from the initial state, enhancing the predictive capacity on time scales longer than one year (Keenlyside et al., 2008; Pohlmann et al., 2009; Smith et al., 2007). Therefore, the primary goal of NTCP is to produce reliable prediction of the actual evolution of the climate system taking into account both internally-generated and externally-forced components. In this regard, the framework is based on running retrospective forecasts (also called "hindcasts") with fully coupled global climate models (GCM) using observationally based state information at prescribed intervals over the historical period. The skill of those hindcasts is estimated by comparing the predicted fields with observations.

Besides providing information on the historical skill, those forecasts can provide estimates of the predictability of the climate system, that arises from both internal generated and externally forced variability. Predictability studies can possibly indicate the upper bound and regions where it is possible to predict certain climate parameters on timescales from seasons to decades. Predictability can be estimated in different ways, for example examining the ensemble spread of GCM hindcasts initialized with observations (e.g., Smith et al., 2007; Yeager et al., 2012; Meehl and Teng, 2012); and using different statistic tools to identify where and at what timescales certain variables have potential to be predicted (e.g., Newman, 2007; DelSole and Tippett, 2009; Zanna, 2012; Zhang et al., 2017). These statistical approaches can be used as useful benchmarks for initialized decadal predictions.

Those predictable patterns in such timescales derive from the long-term memory in the ocean due to its thermal-inertia and its internal modes of variability at different timescales. In the North Atlantic, for example, the Atlantic Multidecadal Oscillation is a coherent mode of natural variability with a period of 60–80 years and it is estimated as the anomalies of sea surface temperature (SST) in the North Atlantic (AMO, Schlesinger and Ramankutty, 1994). In the Pacific, there are different well-known modes at different timescales. From seasonal to interannual timescales, the El Niño Southern Oscillation (ENSO) is the dominant source of variability for the Pacific Ocean and it has been identified as the highest source of predictability at these time scales for SST anomalies in the Tropical Pacific (Xue et al., 2013). The decadal variability in the Pacific is dominated by the Pacific Decadal Oscillation (PDO; Mantua et al., 1997), that represents the primary mode of SST variability across the midlatitude North Pacific. However, the decadal variability in the Pacific is not dominated by PDO to the same degree as the interannual variability is dominated by ENSO (Newman, 2013), and other modes do play a role. The North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al., 2008) represents the second mode of North Pacific SST and it characterizes the oceanic response to the atmospheric pressure pattern called North Pacific Oscillation (NPO; Linkin and Nigam, 2008; Ceballos et al., 2009).

Some studies have shown that those patterns may have some decadal predictive skill, although with different ranges. While the North Atlantic may be predictable in a decade or more in advance (Ding et al., 2016; Li et al., 2017), in the North Pacific demonstrable forecast skill on these time scales is still lacking (e.g., Smith et al., 2013). The current understand in the predictability of the decadal signals comes from the deterministic links among multi-time-scale modes. For example, the Central Pacific ENSO forces changes in the extra-tropical atmospheric circulation, that in turn drive to the decadal fluctuation of the North Pacific Gyre Oscillation. Additionally, an important fraction of the decadal variance is energized by internal stochastic variability of the coupled ocean/atmosphere system. However, it is unclear if what we currently refer to as stochastic variability is entirely unpredictable or if there are additional deterministic components of the variance that remain to be identified.

Sophisticated statistical models, including Linear Inverse Models (LIM's, Penland and Sardeshmukh, 1995), also yield important information on forecast skill for Pacific and Atlantic decadal climate variations. Previous work on the statistical predictability of these modes with LIM's suggest increased predictability in subsurface temperature over the traditionally-used SST, with potential predictability to over a decade (Branstator and Teng, 2010). In addition, it is apparent from other LIM studies that PDO both affects and is affected by ENSO (Alexander et al., 2008), necessitating studies of timescale coupling between different timescales. Additional work with LIM's and GCM's is needed to search for new modes of decadal variability and to diagnose physical processes and components at different time scales that are drivers of the low frequency modes in the global climate.

These low-frequency modes can influence weather patterns and climate statistics over land. For example, different phases of the AMO are associated with many regional effects over land, such as North Eastern Brazilian and African Sahel rainfall, Atlantic hurricanes and North American and European summer climate (e.g., Knight et al., 2006). In the Pacific, the PDO is known to influence synoptic weather patterns over the Western North America (Gershunov and Barnett, 1998; Hidalgo and Dracup, 2003). The downstream influence of Pacific decadal variability on vigorous weather events and their accumulated climatic effects is also evident in long-term changes in heat waves (Guirguis et al., 2011) and extreme precipitation statistics (DeFlorio et al., 2013). Such slow-varying processes can also give rise to predictability in the atmosphere from seasonal timescales and beyond, what could have many benefits for decision-making related to hydrology, agriculture, energy, and other sectors. Therefore, understanding the mechanisms that influence the predictability over land in seasonal timescales and beyond is of extreme importance to improve our predictive ability of variables that affect many sectors of our society.

Overview of the dissertation

The work in this dissertation uses sophisticated empirical models to establish new benchmarks for seasonal to decadal predictions in diverse components of the climate system and to provide some pieces of evidence that help to understand what are the drivers for these predictable patterns. The dissertation is organized s follows. In Chapter 2, we aim to identify how interactions between time scales affect predictability of Pacific Sea Surface Temperature anomalies (SST) and how Tropical and Extratropical Pacific affect each other in driving predictable components linked to ENSO or intrinsic mid-latitude interactions. In Chapter 3, we develop a statistical prediction model for seasonal minimum (Tmin) and maximum (Tmax) temperatures over Western North America (WNA) using remote and local predictors. The first predictor is sea surface temperature (SST) across the tropical and northern Pacific basin, representing the influence of large-scale climate variability patterns, which affect surface air temperature over land via association with atmospheric circulation patterns. The second predictor is soil moisture (SM), which is thought to represent the local effects on surface air temperature that affect the surface energy balance through partitioning of latent and sensible heating. We showed that both remote and local predictors influence the variability and, as consequence, the predictability of local Tmax and Tmin over WNA. The result of this work served as motivation for

Chapter 4, in which we propose a higher benchmark for seasonal predictions of Tmin and Tmax using the North American Multimodel Ensemble (NMME), a state-of-art global prediction system. Finally, in Chapter 5 we perform retrospective decadal predictions of global SST and sea surface height (SSH) anomalies using an empirically-built linear inverse models (LIM) and compare the results with that from another state-of-art decadal prediction system the Community Earth System Model Decadal prediciton Large Ensemble (CESM-DPLE). Guided by these comparisons, we investigate possible sources of errors in the CESM simulations, by exploring the LIM linear feedback matrix.

Chapter 2

Remote and Local Influences in Forecasting Pacific SST: a Linear Inverse Model and a Multimodel Ensemble Study

Abstract

A suite of statistical linear inverse models (LIMs) are used to understand the remote and local SST variability that influences SST predictions over the North Pacific region. Observed monthly SST anomalies in the Pacific are used to construct different regional LIMs for seasonal to decadal predictions. The seasonal forecast skills of the LIMs are compared to that from three operational forecast systems in the North American Multi-Model Ensemble (NMME), revealing that the LIM has better skill in the Northeastern Pacific than NMME models. The LIM is also found to have comparable forecast skill for SST in the Tropical Pacific with NMME models. This skill, however, is highly dependent

on the initialization month, with forecasts initialized during the summer having better skill than those initialized during the winter. The data are also bandpass filtered into seasonal, interannual and decadal time scales to identify the relationships between time scales using the structure of the propagator matrix. Moreover, we investigate the influence of the tropics and extra-tropics in the predictability of the SST over the region. The Extratropical North Pacific seems to be a source of predictability for the tropics on seasonal to interannual time scales, while the tropics enhance the forecast skill for the decadal component. These results indicate the importance of temporal scale interactions in improving the predictions on decadal timescales. Hence, we show that LIMs are not only useful as benchmarks for estimates of statistical skill, but also to isolate contributions to the forecast skills from different timescales, spatial scales or even model components.

2.1 Introduction

The Pacific Ocean sea surface temperature (SST) exhibits variability on timescales from diurnal (Tanahashi, 2003; Clayson and Weitlich, 2007) to decadal and centennial (Mantua et al., 1997; Power et al., 1999; Yeh et al., 2011). These coherent large-scale SST anomalies observed in the Pacific Ocean also impact the weather and climate in regions around the Pacific and globally (e.g., Vimont et al., 2001; Grimm and Tedeschi, 2009; Allen et al., 2015; Capotondi et al., 2015; L'Heureux et al., 2015). SST fluctuations over the Pacific Ocean are caused by various mechanisms, both internal ocean variability as well as local and remote stochastic atmospheric heat and momentum flux forcings (Bjerknes, 1966; Battisti and Hirst, 1989; Newman et al., 2016). Different processes influence the SST anomaly evolution over this region, such as climate modes forced by stochastic atmospheric variability that manifest as oscillations in surface variables and in the coupled atmosphere-ocean system (Frankignoul and Hasselmann, 1977; Penland and Matrosova, 1994; Di Lorenzo et al., 2015).

For more than a century scientists have been exploring factors that impact the SST evolution over the North Pacific and especially over the California Coastal region. Mcewen (1914) explored impacts of temperature difference between the continent and the ocean on the large-scale atmospheric pressure and circulation over the Pacific, which then impacts the SST and leads to a feedback on the process on decadal timescales (Newman et al., 2016). Therefore, it is important to investigate not only which modes of variability in the Pacific dominate each timescale, but also the interactions among those scales of variability, in order to improve our understanding of the role of such interaction in the predictability.

For seasonal to interannual time scales, the El Niño Southern Oscillation (ENSO) is the dominant source of variability for the Pacific Ocean and it has been identified as the highest source of predictability in these time scales for SST anomalies in the Tropical Pacific (Xue et al., 2013). The predictability of ENSO variability has been tested using several statistical and physical methods, ranging from linear methods (e.g., Linear Inverse Models, Penland, 1989; Penland and Sardeshmukh, 1995, and canonical correlation analysis, Barnston and Ropelewski, 1992), nonlinear statistical methods (e.g., Eccles and Tziperman, 2004; MacMynowski and Tziperman, 2008; Chen and Majda, 2016a) to operational coupled general circulation models (e.g., North-American Multimodel Esemble - NMME, Kirtman et al., 2014). The latter is a global prediction system that was recently developed to exploit the idea of using multiple models to improve the skill of the forecasts. Indeed, Becker et al. (2014) showed that NMME mean forecasts have more skill in predicting ENSO-related variability than individual models.

The decadal variability in the Pacific is also well known, but the mechanism that control these long-term variations are still unclear. The Pacific Decadal Oscillation (PDO; Mantua et al., 1997) represents the primary mode of SST variability across the midlatitude North Pacific. Although PDO is correlated with a number of physical and biogeochemical variables, its dynamical mechanism is associated with the combination of a few phenomena (Miller and Schneider, 2000; Schneider and Cornuelle, 2005; Newman et al., 2016): teleconnections from the Tropical Pacific via “atmospheric bridges" (Alexander et al., 2002), large-scale stochastic atmospheric forcing via intrinsic variability of the Aleutian Low, ocean memory via reemergence mechanism (Alexander and Deser, 1995), and Rossby wave signatures along the Kuroshio-Oyashio Extension (KOE) region (Qiu, 2003; Taguchi et al., 2007). Moreover, the decadal variability in the Pacific is not dominated by PDO to the same degree as the interannual variability is dominated by ENSO (Newman, 2013), and other modes do play a role. The North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al., 2008) represents the second mode of North Pacific SST and it characterizes the oceanic response to the atmospheric pressure pattern called North Pacific Oscillation (NPO; Linkin and Nigam, 2008; Ceballos et al., 2009).

The ability to predict these fluctuations of Pacific SST anomalies has many obvious economical and societal benefits because it can help improve adaptation and mitigation to extreme weather and climate anomalies (Alexander et al., 2008). Yet, current generation of numerical models have many limitations in predicting the system accurately in this region (Meehl et al., 2014; Becker and van den Dool, 2016; Barnston et al., 2017; Newman and Sardeshmukh, 2017) and the long record of observations can be used to help inform the models better. Our goal in this study is to understand the remote and local SST variability that influences SST predictions over the North Pacific region and to investigate the interactions between dominant modes of variability on time scales from seasonal to decadal. With that, we expect to improve our understanding on how the long observed SST record can help better guide multi-model ensemble forecasts.

We use a linear inverse model (LIM, Penland and Magorian, 1993) to estimate the predictability of the SST anomalies in the North and Tropical Pacific at different time scales. We then use a suite of the seasonal forecasts from NMME models to investigate the limits on forecast skill of the system, by comparing a low-dimensional empirical linear model with a high-dimensional nonlinear coupled model. LIM constitutes the least complex form of a reduced stochastic-dynamic climate model (Majda et al., 2009) and it has been broadly used for diagnostic and prediction of Tropical and Extratropical weather and climate from diurnal to decadal time scales (Penland and Sardeshmukh, 1995; Alexander et al., 2008; Newman et al., 2009; Zanna, 2012; Cavanaugh et al., 2014; Capotondi and Sardeshmukh, 2015; Huddart et al., 2017; Newman and Sardeshmukh, 2017). Those studies have shown that LIMs have comparable predictive skill to high-dimensional global circulation models, despite their reduced number of degrees of freedom.

This chapter is organized as follows. In Section 2.2 we briefly introduce the LIM theory. Section 2.3 describes the data used and establishes the LIM model configuration and the experiments performed. Section 2.4 shows LIM forecasts skills for seasonal, interannual and decadal experiments as well as shows the comparison with NMME hindcasts skill. Finally, in Section 2.5 we summarize the main results and point some concluding remarks.

2.2 Linear Inverse Modeling

Linear Inverse Model (LIM) assumes that the evolution of some phenomena in a system can be represented as a linear process forced by stochastic noise. In other words, it can be separated into a linear deterministic part and a nonlinear part, represented by white noise fluctuation, which may be spatially correlated but temporally uncorrelated. Therefore, the governing dynamics of such system can be represented in the form:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{L}\boldsymbol{x} + \boldsymbol{\xi}, \tag{2.1}$$

where $\boldsymbol{x}$ is the system state vector, $\boldsymbol{L}$ is the time-independent linear operator matrix and $\boldsymbol{\xi}$ is the white stochastic forcing. A detailed description of LIM procedure have been broadly discussed in several papers (e.g., Newman et al., 2003; Newman, 2007; Penland and Sardeshmukh, 1995), so here we only provide some essential information necessary for discussion.

The linear operator $\boldsymbol{L}$ can be estimated from the observed statistics of the system under consideration if the relation in (2.1) can be used to describe this system. For that, we use the lag covariance matrix $\boldsymbol{C(\tau)}$ of the system with components $\boldsymbol{x}_i$, estimated at any fixed lag τ from the observations as $\boldsymbol{C}_{ij}(\tau) = \langle \boldsymbol{x}_i(t+\tau)\boldsymbol{x}_j(t)\rangle$, where the angle brackets denote a long-term average and the subscripts i and j represent the covarying observational time-series. Given that in linear inverse modeling one assumes that the relation in (2.1) is valid, the system satisfies the relation

$$\boldsymbol{C}(\tau) = \boldsymbol{B}(\tau)\boldsymbol{C}(0), \tag{2.2}$$

where $\boldsymbol{C}(0)$ is the lag 0 covariance matrix of the state vector $\boldsymbol{x}$ and $\boldsymbol{B}(\tau) = exp(\boldsymbol{L}\tau)$ is the propagator matrix that represents the evolution of the predictable signals at some lead time τ. Finally, for some chosen lag-time τ_0, the matrices $\boldsymbol{L}(\tau_0)$ and $\boldsymbol{B}(\tau)$ can be determined from the estimates of $\boldsymbol{C}(\tau_0)$ and $\boldsymbol{C}(0)$ as follows:

$$\boldsymbol{B}(\tau) = exp(\boldsymbol{L}\tau) = [\boldsymbol{C}(\tau_0)\boldsymbol{C}(0)^{-1}]^{\frac{\tau}{\tau_0}}. \tag{2.3}$$

The matrices $\boldsymbol{L}$ and $\boldsymbol{B}$ should be independent of the choice of τ_0 and how well this holds can be used as a measure of the efficacy of applying LIM to the system (Penland and Sardeshmukh, 1995). This is tested for the specific LIM's configurations used in this study by applying the so-called "τ test" for each of the time scales considered. This test can be performed in several different ways, and here we apply two methods. First, by comparing the observed lag-covariances at different lags with the ones estimated using our LIM configurations (e.g., Newman et al., 2011). Since the linear approximation in

(2.1) implies that $\boldsymbol{C}(\tau) = \boldsymbol{B}(\tau)\boldsymbol{C}(\mathbf{0})$ (Eq. 2.2), the model should be able to reproduce the observed lag-covariances at the different lags τ. Besides testing the independence of the matrix $\boldsymbol{B}$ from τ_0, this test also shows how well the relation in (2.1) describes the system. Our tests show that this is true for every LIM configuration used here, and for different lags. For example, the unfiltered LIM captures the local SST lag-covariance up to 12 months. This also is true for the filtered experiments, both interannual and decadal, in which the local SST lag-covariance is captured for lags ranging from 2 to 48 months (not shown). A second verification for the τ test was made by calculating the Euclidean norm of the sub-matrices in $\boldsymbol{L}$ as function of τ_0, following the methodology described in (Penland and Sardeshmukh, 1995). Our LIM also passes in this test, yielding similar results to (Penland and Sardeshmukh, 1995). Those results indicate that the linear approximation described in (2.1) is valid for the configurations used in this study.

Given the estimation of the deterministic part of the system using LIM, the forecast of $\boldsymbol{x}(t+\tau)$ can be made by analytically solving (2.1), which results in

$$\boldsymbol{x}(t+\tau) = \boldsymbol{B}(\tau)\boldsymbol{x}(t) + \boldsymbol{\varepsilon}, \tag{2.4}$$

where τ is the lead time, t is the initial condition and $\boldsymbol{B}(\tau)\boldsymbol{x}(t)$ will be the best forecast in the least square sense. $\boldsymbol{\varepsilon}$ is the error vector and represents the effect of the unpredictable stochastic forcing (term $\boldsymbol{\xi}$ in equation 2.1). The expected value of the global error covariances can be estimated by $\boldsymbol{E}(t,\tau) = \langle \boldsymbol{x}(t+\tau)\boldsymbol{x}^T(t+\tau)\rangle - \boldsymbol{B}(\tau)\langle \boldsymbol{x}(t)\boldsymbol{x}^T(t)\rangle \boldsymbol{B}^T(\tau)$, and can provide a measure for the expected forecast error due to unpredictable dynamics.

2.3 Model details and experiments

2.3.1 Data

The SST dataset used for the LIM analysis and forecasts are from the Hadley Center Sea Ice and Sea Surface Temperature (HadISST, Rayner et al., 2003), interpolated on a 1° by 1° grid for the North and Tropical Pacific (between 15°S to 60°N), and with monthly means from 1900 to 2016. The normalized anomalies (SSTa) are determined by first removing the climatological monthly mean for each grid cell at each month, then dividing by their standard deviation and finally weighting each grid cell by their surface area, in order to equally weight the Tropical and subtropical one degree cells. SSTa are filtered with a third order Butterworth filter for three different time scales: decadal (D), with periods greater than 10 years; interannual (I), with periods between 1 and 10 years; and seasonal (S), with periods less than 1 year.

Singular value decomposition (SVD) is applied for the unfiltered SSTa time series (U) and for each of those three filtered datasets separately (D, I and S). Therefore, the SSTa fields are decomposed into empirical orthogonal functions (EOFs) that describe the spatial pattern, and their associated principal components (PCs) that describe the time evolution. In addition to temporal fields, PCs and EOFs are also determined for different domains over the Pacific: tropics plus extratropics (PA, 15°S to 60°N), only tropics (TP, 15°N to 15°S), and only extratropics (ET, 16°N to 60°N).

2.3.2 LIM configuration

We choose to construct the SSTa state vector as a truncated time series of the leading PCs, and the form of $\boldsymbol{x}$ depends on which experiment is being performed, either unfiltered (for seasonal forecasts), or filtered SSTa (for interannual and decadal forecasts).

Unfiltered LIM

For the unfiltered experiments, the state vector $\boldsymbol{x}$ is built using the leading 20 PCs of SSTa ($\boldsymbol{x}_U$). Moreover, one state vector is built for each spatial field ($\boldsymbol{x}_{UPA}$, $\boldsymbol{x}_{UTP}$ and $\boldsymbol{x}_{UET}$). Therefore, Eq. (2.1) can be rewritten for the unfiltered experiments as:

$$\frac{d\boldsymbol{x}_{UPA}}{dt} = \boldsymbol{L}_{UPA}\boldsymbol{x}_{UPA} + \boldsymbol{\xi}_{UPA} \tag{2.5}$$

$$\frac{d\boldsymbol{x}_{UTP}}{dt} = \boldsymbol{L}_{UTP}\boldsymbol{x}_{UTP} + \boldsymbol{\xi}_{UTP} \tag{2.6}$$

$$\frac{d\boldsymbol{x}_{UET}}{dt} = \boldsymbol{L}_{UET}\boldsymbol{x}_{UET} + \boldsymbol{\xi}_{UET} \tag{2.7}$$

The amount of total variance explained by the 20 leading PCs for each of those experiments are very similar: the 20 PCs from the extratropics explain 87.4% (ET_U), from the tropics 91.2% (TP_U), and from the entire North Pacific (PA_U) they explain 86.2% (Table 2.1).

Table 2.1: Amount of variance explained by the 20 EOFs for each of LIM experiments.

	D	I	U
PA	99.4%	92.7%	86.2%
TP	99.7%	96.1%	91.2%
ET	99.5%	92.6%	87.4%

Filtered LIM

To construct the state vector $\boldsymbol{x}$ for the filtered experiments, we follow the methodology adopted in Huddart et al. (2017). We use the leading 20 PCs corresponding to each time scale (x_D, x_I and x_S), and combine them into a single 60-component vector. As in the unfiltered experiment, one state vector is built for each spatial field (x_{PA}, x_{TP} and

x_{ET}). So, the vectors $\boldsymbol{x}$ for the filtered experiments have the following form:

$$x_{PA} = \begin{bmatrix} x_{DPA} \\ x_{IPA} \\ x_{SPA} \end{bmatrix} \qquad x_{TP} = \begin{bmatrix} x_{DTP} \\ x_{ITP} \\ x_{STP} \end{bmatrix} \qquad x_{ET} = \begin{bmatrix} x_{DET} \\ x_{IET} \\ x_{SET} \end{bmatrix} \tag{2.8}$$

With this definition of $\boldsymbol{x}$, we can rewrite Eq. (2.1) in a general form as:

$$\frac{d}{dt}\begin{bmatrix} x_D \\ x_I \\ x_S \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{DD} & \boldsymbol{L}_{DI} & \boldsymbol{L}_{DS} \\ \boldsymbol{L}_{ID} & \boldsymbol{L}_{II} & \boldsymbol{L}_{IS} \\ \boldsymbol{L}_{SD} & \boldsymbol{L}_{SI} & \boldsymbol{L}_{SS} \end{bmatrix} \begin{bmatrix} x_D \\ x_I \\ x_S \end{bmatrix} + \begin{bmatrix} \xi_D \\ \xi_I \\ \xi_S \end{bmatrix} \tag{2.9}$$

We can observe in Eq. (2.9) that it is possible to diagnose linear relationships among the different time scales through the dynamical operator $\boldsymbol{L}$. Its diagonal components represent the interaction between each time scale with itself, while the off-diagonal components diagnose the interactions between two time scales. Filtering the data allows us to investigate how each time scale interacts with one another at different lead times, as well as the direction of this interaction (e.g $\boldsymbol{L}_{DI}$ represents the downscale interactions, hence how decadal modes drive interannual modes).

As in the unfiltered PCs, the 20 leading PCs for the filtered data explain a similar amount of variance for each region. The 20 leading PCs of the extratropics for the interannual time scale explain 92.6% ET_I, the ones from the tropics explain 96.1% (TP_I), and from them combined 92.7% (PA_I). For the decadal time scale, those three different domains (ET_D, TP_D, PA_D) explain at least 99% of the variance (Table 2.1).

2.3.3 LIM forecasts

Following the procedure described in Section 2.2, we use the propagator matrix $\boldsymbol{B}$, to do the forecasts of the PCs time series. First we do forecasts on seasonal time scale using the unfiltered SSTa time series for each spatial field, hence using the state vectors from Eqs. 2.5, 2.6 and 2.7.

The forecast made are 12-month long and they are initialized every month from January to December of the years 1995 to 2010 (experiments PA95, TP95 and ET95). The estimates of $\boldsymbol{B}$ and the forecast skill are made in a cross-validation sense by first excluding those 15 years of the SSTa time series (which we call here as verification period), and estimating $\boldsymbol{B}$ for the remainder years. The 12-months forecasts are then generated for every month of the independent 15 years, resulting in a total of 180 seasonal forecasts. In order to test the influence of the excluded data on the predictability, we used a different verification period to perform this LIM 12-months forecast (from 1950 to 1965, experiments PA50, TP50 and ET50).

The LIM′s seasonal forecast skill is then compared to that of the models of the North American Multimmodel Ensemble (NMME, Kirtman et al., 2014). The models included are all the GCMs from NMME phase 1 that have monthly SST hindcasts: the Community Earth System Model (CESM1) and the Canadian Center for Climate Modeling and Analysis (CanCM3 and CanCM4). The NMME dataset can be found in http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/. The NMME hindcasts are also 12-month long and are initialized every month from January to December of the years 1995 to 2010. We used the ensemble mean of those models and calculated the anomalies by subtracting the climatological mean of each month and each lead time of the hindcast.

To evaluate the forecast skill and the predictability, we use two measures: the anomaly correlation coefficient (ACC) and the root mean square error (RMSE). Those estimates are calculated by comparing LIM and NMME forecasts with the HadISST data. First we calculate the ACC for each grid cell at different lead times for the LIM forecasts (experiments PA50, TP50, ET50 and PA95, TP95, ET95). This test allows us to quantify the skill at a given lead time and at different parts of the domain. Therefore, by looking for regions in the domain where the predictable signals are high for longer lead times, we can estimate in which areas the SSTa evolution is approximated well by linear dynamics.

Three different domains are used to forecast the Pacific SSTa. First, the forecast is made for the whole North Pacific (PA) and then individually for the tropics (TP) and the extratropics (ET). It is important to address that the basis function in which each LIM spatial field is based is not exactly the same. That is, the 20 leading modes used here do not explain exactly the same amount of local and total variance. However, they explain at least 86% for the unfiltered experiment, 92% for the intearanual experiment, and 99% for the decadal, also the higher modes explain less than 1% of the total variance. Therefore, the results of the forecast skill for TP and ET can be compared to the results of the LIM forecasts trained with the whole domain, in order to investigate possible links between the Tropical and Extratropical SSTa evolution.

We then choose two regions in the Pacific to perform the LIM and NMME comparison: California Current System (CCS, 30° N to 40° N and 232° E to 244° E) and Niño 3.4 region (5° S to 5° N and 190° E to 240° E). We calculated the SSTa spatial mean within those regions for each of those datasets and the skill was evaluated by calculating the ACC between each forecast and the observations, for each lead time and each start month. Additionally, we computed the time evolution of RMSE between each forecast and observations, for each region.

Finally, we perform forecasts individually for the interannual and the decadal components, for each spatial field using the state vectors ($\boldsymbol{x}$) described in Eq. 2.8. Although the forecast is made individually for each time scale, the interactions among $\boldsymbol{x}_D$, $\boldsymbol{x}_I$ and $\boldsymbol{x}_S$ are explicitly included in $\boldsymbol{B}$. Therefore, the forecasts made for each individual time scale takes into account the linear relationship between each other.

The forecast skill calculations are also cross-validated and evaluated by using the maps of ACC. For this case, LIM-forecasts are 10-years long and they are also initialized every month from January to December of the years 1950-1960. We remove the 10 years of the SSTa time series, then we calculate the EOFs and PCs for the remainder years and we use those to estimate $\boldsymbol{B}$. The forecasts are then made for the independent 10 years removed and this procedure is repeated for each month along those 10 years, resulting in a total of 120 forecasts.

2.4 Results and discussion

2.4.1 Seasonal forecasts: unfiltered data

The predictability of North and Tropical Pacific SSTa was first evaluated using the ACC for the unfiltered data on a seasonal time scale. Figures 2.1 and 2.2 show the ACC for 3, 6, 9 and 12 forecast lead months for all the experiments described in Section 2.3.2 (50 and 95, respectively).

Overall, the spatial distribution of the forecast skill for all the experiments are very similar, with higher skill seen in three key regions, marked in the Figure 2.1a with a black circle and with the numbers 1 to 3: a circular-shaped region in the northwestern Pacific (1), a broad region along the coast of North America extending southwestward to the central Pacific (2), and a region east of the dateline in the border of the South tropics (3). Skill minima are found in the equatorial cold tongue region in the eastern Pacific, in regions surrounding the circular shaped region of maximum skill in the northern extra-tropics, and in a V-shaped region in the western Pacific. The decrease in skill with the increasing lead time is also similar for all the experiments. Not surprisingly, the regions of little skill are located along the regions of the higher amplitude of the first and second EOF modes of the Pacific. Those areas have the weakest forecast signal for the configuration of linear model used in this study. Different reasons can explain the lack of forecast skill: 1) the SST-only LIM may not be adequate for some regions, due to the lack of persistence. Indeed, it was shown that there is an improvement in the LIM's forecast skill for the Tropical Pacific when including subsurface Newman et al. (2011); 2) nonlinear processes dominate the variability, which are not fully captured by the LIM; 3) the predictability is low and the evolution of the system is dominated by stochastic processes. On the other hand, regions with the highest forecast skill are likely associated with ENSO or PDO related signals where SSTa evolution is largely driven by slower timescale processes which tend to be more linear.

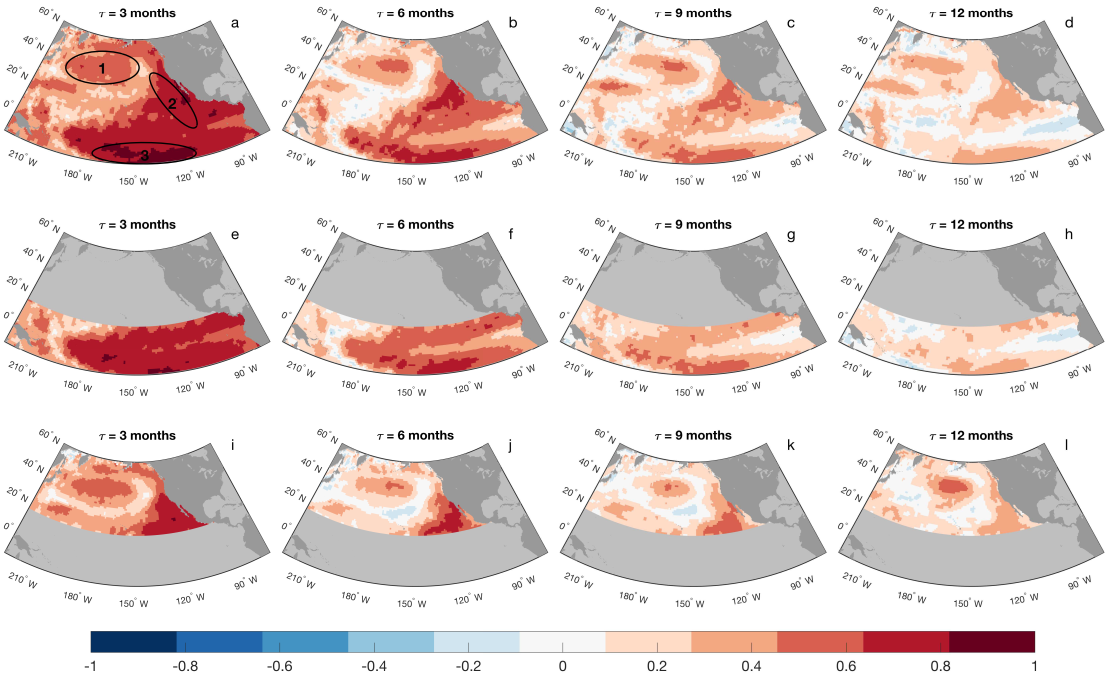


Figure 2.1: Maps of temporal correlation for the SST anomalies between the LIM seasonal forecast and observations, for different lead times. The verification period used was 1950 - 1965. **a - d**: LIM trained using data for the whole domain (experiment PA50); **e - h**: LIM trained only with data from the Tropical region (experiment TP50); **i - l**: LIM trained with data only from the extratropics (experiment ET50).

Excluding data from the tropics or the extratropics to make the forecast and comparing their skill with the forecast made with the whole domain allowed us to identify regions with possible links between the SSTa evolution in the tropics and in the extratropics. Including the extra-Tropical SST field in the forecast model helps improve the skill in predicting the Tropical SSTs (experiments PA50 and PA95 compared to TP50 and TP95).

By comparing Figures 2.1a-d and 2.2a-d to 2.1e-h and 2.2e-h, it is possible to identify a loss in skill for TP50 and TP95 experiments right in the first 3 months of forecast. For longer lead times this improvement in the skill is clearer in the PA95 experiment. This increase in the skill for the tropics when including data from the extratropics may be an indication that the SST-only LIM captures the influence of the extratropics in the Tropical variability. Some other studies that used a Tropical LIM extended far enough into the subtropics captured a SST pattern that leads to optimal ENSO excitation, described as the ENSO precursor mode originally by Penland and Sardeshmukh (1995). This pattern, also known as the Pacific Meridional Mode (PMM) acts as a conduit through which the Extratropical atmospheric variability is conducted to the tropics (Chiang and Vimont, 2004; Chang et al., 2007; Vimont et al., 2009). Additionally, several other studies performed over the past two decades have identified the mechanism behind this forcing. The atmospheric variability of the North Pacific has been identified as a forcing for the Tropical climate variability through the North Pacific Oscillation (NPO) and through the excitation of the PMM (e.g., Vimont et al., 2001, 2003, 2009; Chiang and Vimont, 2004; Chang et al., 2007; Di Lorenzo et al., 2015). During boreal winter, the southern branch of NPO modulates the strength of the northeasterly trade winds, leaving an anomalous imprint in the SST through latent heat flux changes. Those anomalies persist into boreal summer and induce changes in the Tropical atmospheric circulation. This is the so-called Seasonal Footprinting Mechanism (SFM) (Vimont et al., 2003). Even though our analysis included only SST data, LIM implicitly includes the impact of all other variables that are related to SST (Newman, 2007). Moreover, the persistence of high skill forecast up to 12 months in the region (2) in Figure 2.1a resembles the shape of PMM. Therefore, the linear model captured the effect of atmospheric Extratropical variability on the Tropical climate and the forecast skill improvement is an indication that the SST evolution as well as the interactions with the atmosphere are well approximated by linear dynamics. A recent study suggested that the influence of the midlatitudes SST anomalies associated with NPO and the propagation into the tropics through the PMM is essentially linear (Thomas and Vimont, 2016). Our results support this suggestion and add one more evidence for the linearity of this mechanism.

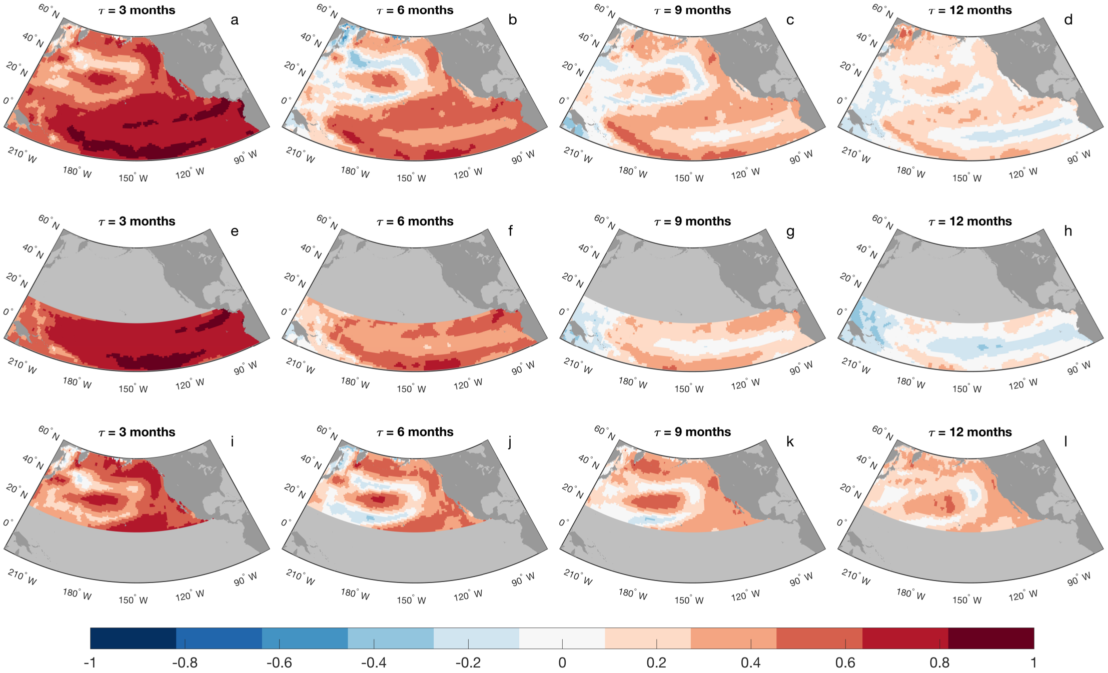


Figure 2.2: Same as Figure 2.1, but using data from 1995 to 2010 as the verification period (**a - d**: experiment PA95; **e - h**: experiment TP95; and **i - l**: experiment ET95).

Considerations on the spatial characteristics of ENSO can be made with our results. The SST anomalies are well approximated with the LIM in the central tropics, with ACC greater than 0.4 persisting until one year of forecast. On the other hand, the skill in the eastern tropics decreases much faster, with ACC close to zero after six months of forecast (Figure 2.1a-h and 2.2a-h). The geographical distribution of skill in the tropics resembles the two types of ENSO that are known: the Central-Pacific (CP), which produces warm SST anomalies in the tropics around the dateline, and the Eastern-Pacific (EP), where the warm SST anomalies are spread from the eastern tropics to the dateline. The precursors that influence the development of each type of ENSO have been recently investigated. Alexander et al. (2010) showed that the Extratropical atmospheric variability, through the PMM, is capable of triggering ENSO-like responses in the tropics and other studies have shown that this mechanism plays a key role in the development CP ENSO events (Yu et al., 2010; Yu and Kim, 2011; Vimont et al., 2014). However, there is no agreement in the literature that it is the only or the main cause. An additional explanation explores the influence of the initial thermocline state in the selection of a specific type of ENSO, in which a deeper thermocline in the eastern (central) tropics favors a EP (CP) ENSO events (Capotondi and Sardeshmukh, 2015), although the EP ENSO events rely more on thermocline-SST feedbacks whereas CP ENSO events may be influenced more by atmospheric forcing (Kao and Yu, 2009). Regarding the linearity of those mechanisms, Thomas and Vimont (2016) proposed that the thermocline parameters are strongly dependent on the nonlinear dynamics while PMM influence is mostly linear. In fact, the higher skills found in the central tropics suggest that the SST anomalies in this region are more influenced by linear processes than the SST anomalies in the eastern tropics. Therefore, our results are consistent with those findings and they highlight the importance of different contributors to the predictability in the Tropical Pacific. We reiterate, though, that other factors may play a role in the development of different types of ENSO, such as the nonlinear advection, that was shown to have a key role in the growth of central Tropical SST anomalies (Capotondi, 2013; Chen and Majda, 2016b).

Regarding the forecast for the extratropics, the inclusion of data from the tropics in the LIM configuration either do not improve the skill (experiment PA50, Figures 2.1a-d) or act to reduce it for most North Pacific region (experiment PA95, Figures 2.2a-d). For example, in PA95 the ACC decreases to 0.2 after 12 months of forecast, while it remains higher than 0.6 when the tropics are excluded (ET95). Curiously, it is known that the variability in the Tropical region, especially ENSO, influence the evolution of SST anomalies in the North Pacific, through atmospheric bridges (e.g., Alexander et al., 2002). Furthermore, those atmospheric bridges contribute to the SST variability on time scales longer than interannual (Zhang et al., 1997) and they are known to be one of the contributors for the development of PDO (Newman et al., 2016). For this reason, one would expect that including information from the tropics would increase the predictability for the North Pacific, which was not the case for our results. However, the Tropical forcing that contributes to the evolution of North Pacific SST is the low-frequency variability, and those results include the high-frequency data and were focused on seasonal forecast. Bearing that in mind, we explore the predictability with filtered data to evaluate the temporal interactions in improving the predictability for SST, and we present and discuss the results in Section 2.4.2.

Despite the general similarities of the skill for the experiments using two different verification periods, there are some outstanding differences between them, especially for the experiments TP50 and TP95. There is a clear loss in skill for longer lead times when forecasting the SST evolution between 1995 and 2010. A detailed explanation of the reasons behind this loss in skill is beyond the scope of this paper. However, these results are presented to show that similar models can have different results according to the data used to train the model. For example, in the period between 1995 and 2010 there was a major El Niño event (1997/1998), whose information probably would dominate the oscillatory pattern of the SSTa evolution, thereby adding predictability to the linear model. As this information was lost when cross validating, this could help to decrease the skill. Additionally, the background climate can act to decrease the predictability for certain periods (Zhao et al., 2016). The 1995 to 2010 period corresponds to a change in the background climate, represented by a transition between the positive (1977 - 1999) and the negative (1999 - 2010) phases of the Interdecadal Pacific Oscillation (IPO), while the years between 1950 and 1970 correspond entirely to the negative phase of the IPO (Henley et al., 2015).

In the next section we present the comparison between LIM and NMME models and present a more detailed discussion about the limits of predictability as well as remote influences in some key regions for the Pacific SST.

NMME and LIM comparison

In order to compare the forecast skill achieved by a simple LIM with a more complex and fully nonlinear GCM, the results of LIM experiments are contrasted with the NMME hindcasts. With that we aim to explore the influence of the nonlinearities represented in the NMME models to the LIM forecast skill, which may provide some hints of what may be the likely cause for the low forecast skill in LIM in certain regions. This is done first by analyzing the ACC spatial maps of the LIM95 (Figure 2.2) and the NMME forecasts (Figure 2.3). The latter is calculated as correlation between the observations and the mean of the hindcasts of the three NMME models used here over the period of 1995-2010. Subsequently, the LIM and NMME comparisons are also done for some specific regions and for different initialization times, by comparing their mean ACC and RMSE. Figures 2.4 and 2.6 show the correlations as a matrices with initialization month on the x-axis and lead time on the y-axis, for the regions Niño 3.4 and CCS, respectively. Figures 2.5 and 2.7 show the RMSE evaluated for each NMME model and the LIM experiment that were initialized for boreal winter (December, January and February) and summer (June, July and August). This approach allow us to study if the SST forecast skill in the region changes with different forecast periods, and also the linearity assumption for the SST dynamics during different periods.

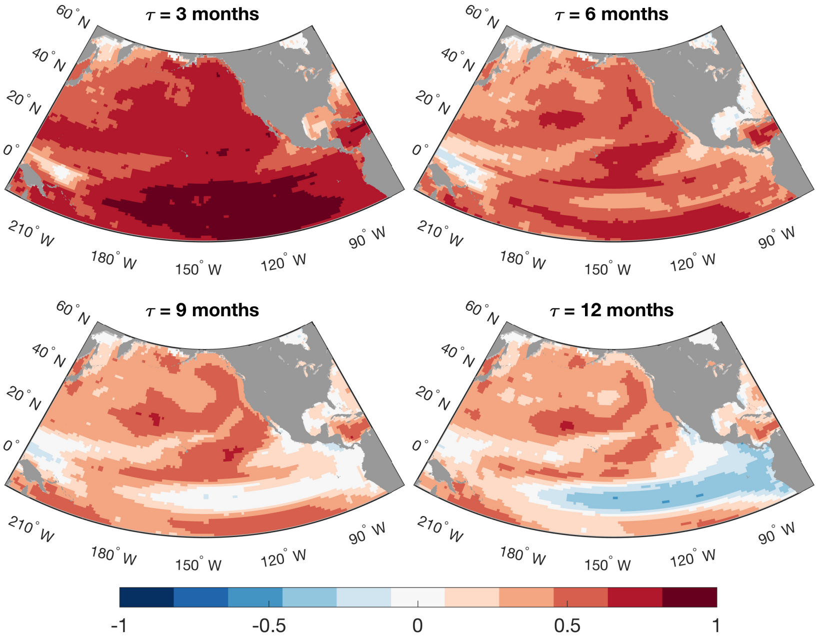


Figure 2.3: Maps of temporal correlation for the SST anomalies between the NMME hindcasts' ensemble mean and observations over the period of 1995-2010.

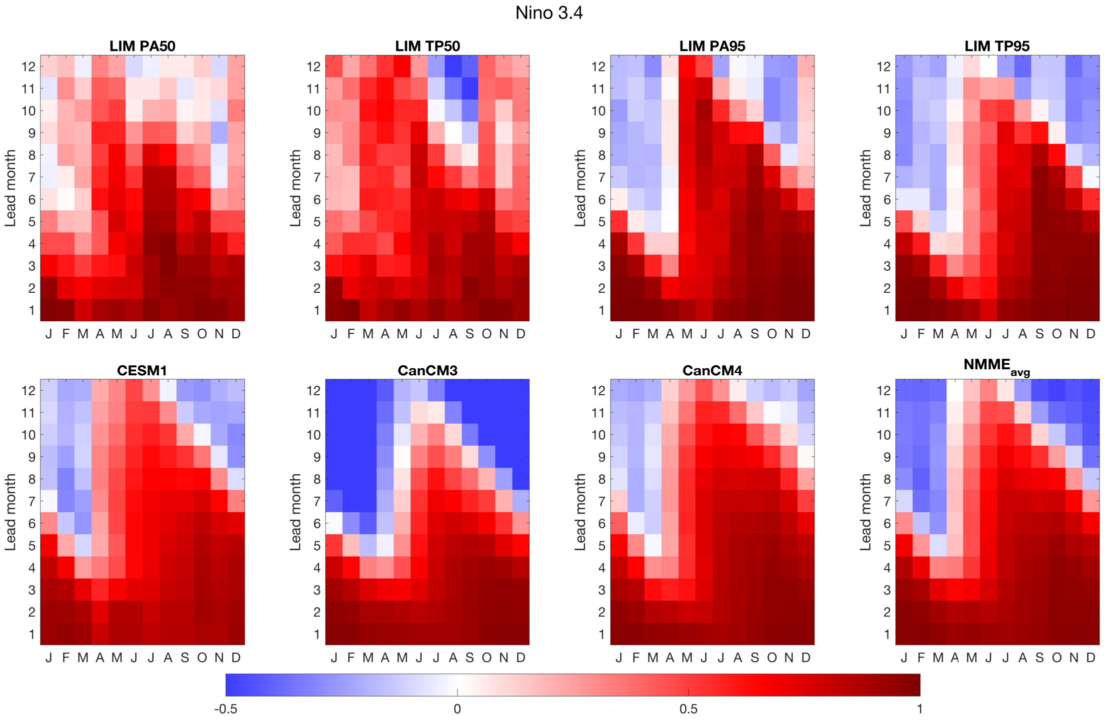


Figure 2.4: Correlations for each start month and lead time between LIMs forecasts and observations (upper panels), and between NMME models and observations (lower panels) for the Niño 3.4 region. NMME$_{\text{avg}}$ is the mean among the three NMME individual models.

In the tropics, the pattern of the NMME and LIM skills at different lead times are very similar: the skill maximum is located around the dateline and a minimum skill is found in the eastern Tropical Pacific and a westward extension, and in the far-western Tropical Pacific, particularly at longer lead times (9 and 12 months). On the other hand, in the extratropics there are some remarkable differences between those two forecasts systems. While the NMME models have skill greater than 0.4 for most of the extratropics and for longer lead times, the LIM positive skills are limited to some specific regions, like near the west coast of North America and a circular shaped region in the Northwest.

The forecasts are highly sensitive to initialization months for the Niño 3.4 region and this difference is more dramatic for the LIM forecast during late twentieth/early twenty-first centuries (PA95 and TP95) and for the NMME hindcasts (Figure 2.4). The forecasts initialized during boreal late spring and summer (from May to September) tend to have higher skill that persists for longer lead times. On the other hand, forecasts initialized before boreal spring (January to April) are much less successful, with high skill persisting for no longer than four months. Other studies have also seen this lack in forecast skill for ENSO forecasts initialized before spring of the El Niño year and this is known as the “Spring Predictability Barrier" (SPB) for ENSO (WEBSTER and YANG, 1992; McPhaden, 2003; Jin et al., 2008; Duan and Wei, 2013; Lopez and Kirtman, 2014). Another similar characteristic for those matrices is that they exhibit a higher skill along their diagonal. This occurs when the month that is being predicted has more skill independently of the lead time, which tend to arise when the skill is a result of impacts that depend on the season (Hervieux et al., 2017). It is noteworthy that the SST-only LIM overall captures the seasonal discrepancies of the NMME skill, for both the models′ average and for each individual model. This LIM ability in capturing the seasonal variations was also pointed by Newman and Sardeshmukh (2017), although those authors used a more comprehensive LIM, in which they included wind and sea surface height anomalies besides SSTa.

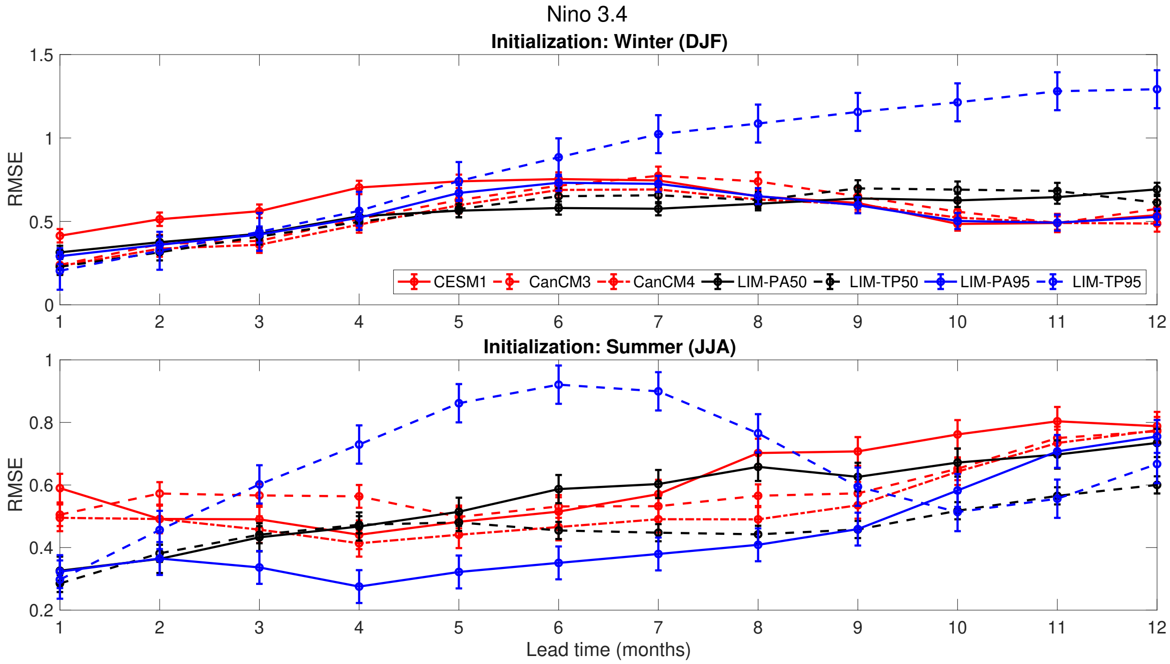


Figure 2.5: Seasonally averaged root mean square error of the SSTa forecast for Niño 3.4 made using four different LIM experiments (LIM-PA50, LIM-TP50, LIM-PA95 and LIM-TP95) and three NMME models (CESM1, CanCM3 and CanCM4). The vertical bars represent the 95% confidence interval standard errors.

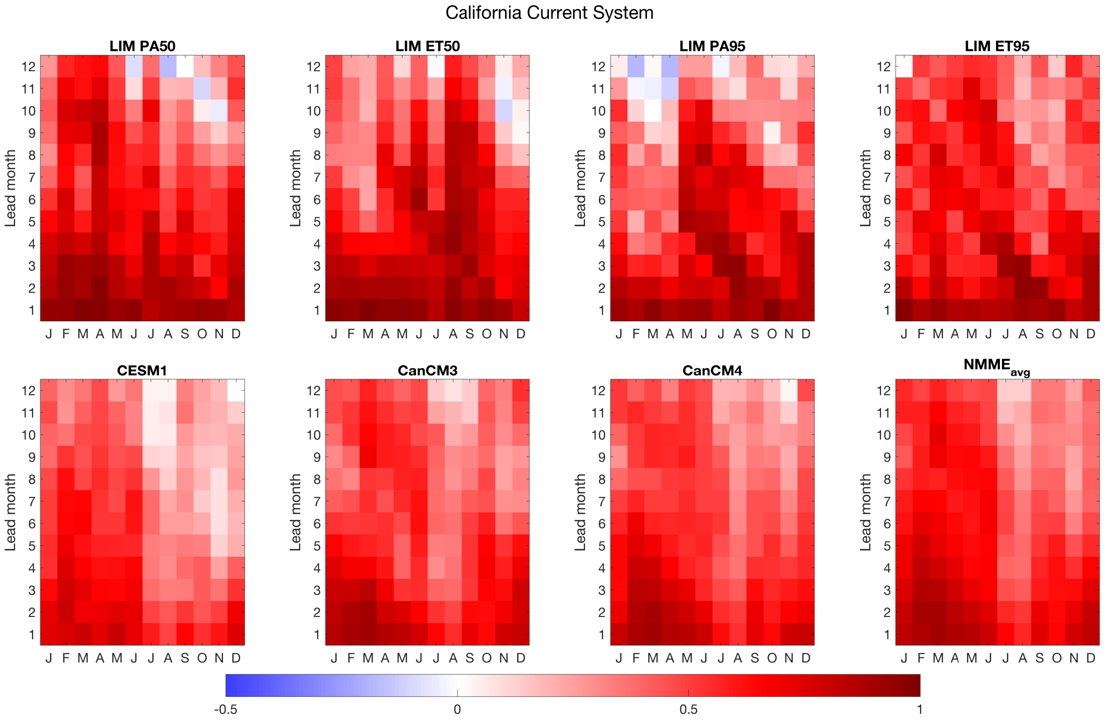


Figure 2.6: Same as Figure 2.4, but for the CCS region

By comparing two essentially different forecast systems, LIM, which is an empirical linear model, and the NMME models, which are high dimensional nonlinear coupled models, some considerations can be made about the predictability of the nonlinear signals. As stated before, LIM assumes that the nonlinear part of the signal is unpredictable, so the skill comes basically from the linear part. The similarity of NMME and LIM skills for the forecasts initialized during boreal late spring and summer can be an indication that the nonlinear signals for the following months are essentially unpredictable. When initialized during boreal winter, both LIM and NMME have low forecast skill at lead times longer than 4 months, disregarding some small differences. Therefore, it is possible that the predictability for the Niño3.4 region and for the months following the winter is intrinsically low. This is also true for the eastern and the far-western Tropical Pacific: with 6 months of lead time those regions have a considerable decrease in forecast skill and with 9 months the skill is completely lost, both for NMME mean and the LIM experiments 2.3. On the other hand, there are some differences between the LIM and NMME skill in certain regions of the extratropics. It is important to note that these discrepancies are not necessarily due to the nonlinearities in the predictable signals which the LIM will fail to capture. It may also be an indication that the SST-only LIM configuration is not the most adequate and there is room for improvement of the LIM skill by including other variables, such as thermocline depth and ocean heat content, which have a greater persistence. Moreover, the prediction skill can also be enhanced by adding additional stochastic noise to the system (e.g., Majda et al., 1999).

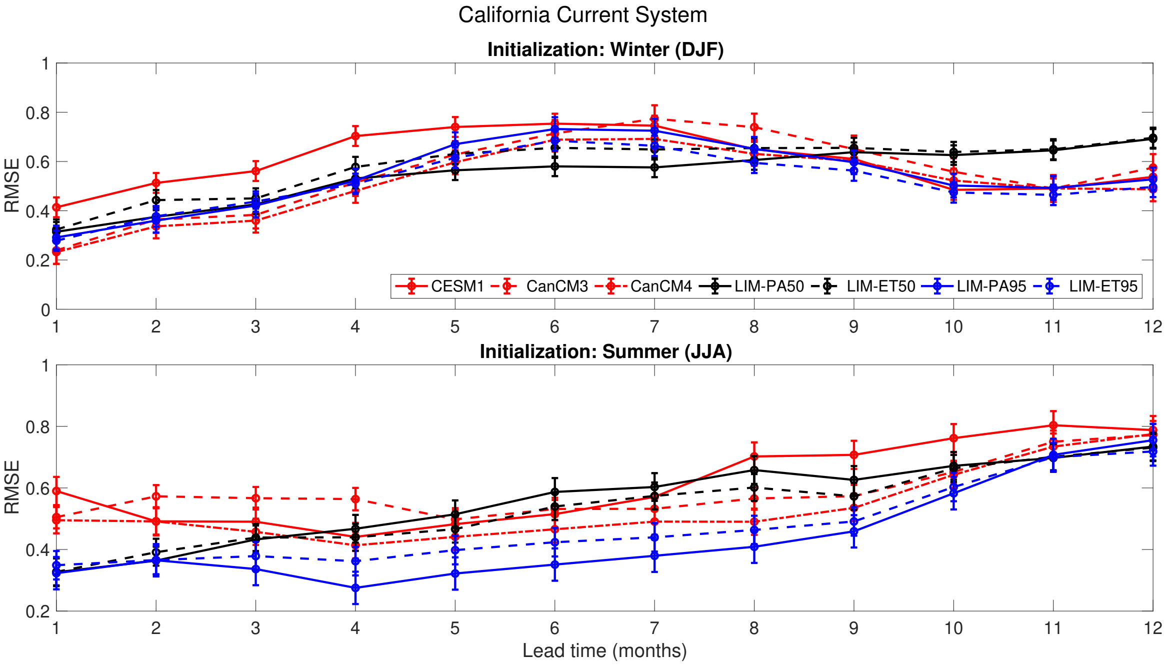


Figure 2.7: Same as Figure 2.5, but for the CCS region

Interestingly, there is a remarkable difference in the skill of the forecast for different periods. Although the forecast initialized during boreal late spring and summer have higher skill persisting up to 8 months, the SPB is much less clear for experiments PA50 and TP50. For example, for the TP50, there is positive skill up to 12 months for experiments initialized before June. Beyond the SPB, although, PA95 and TP95 perform better. A plausible explanation for those differences can be the occurrence of major El Niño events on the period that is being predicted, which can enhance the SPB. In the case of PA95 and T95, a major El Niño event has occurred (1997/98). On the contrary, during the period of PA50 and TP50 experiments, there was very weak El Niño interannual variability.

Those differences can be further explored using the evolution of RMSE with lead time for each NMME model and for each LIM experiment, with initialization during winter and summer months (Figure 2.5). For the initialization during the winter, the RMSE of the LIM and NMME models forecasts are not significantly different until 5 months of lead time. The exception is to the CESM1 model, that performs slightly worse than all the others. From six months of forecast onwards, the RMSE for the LIM experiment TP95 starts to grow fast, while the RMSE for the other models are still comparable. On the other hand, for experiments initialized during summer months, the differences between the RMSE of the models are much bigger: LIM experiment PA95 performs better than any other model up to 8 months, while PA50 and TP50 have comparable RMSE with NMME models. Curiously, when excluding the North Pacific for the verification period of 1995 - 2010, the error grows faster already in the second lead month, and in the third lead month it has larger errors than any other model. Those results reinforce our earlier discussion that in the North Pacific the inclusion of the Extratropical data in the LIM configuration increases the forecast skill for the tropics, especially when forecasting for the late twenty/early twenty-first centuries. This may be an indication that the SST-only LIM can capture some of the source of Extratropical predictability from the tropics.

When considering the CCS region, the LIM forecasts have, overall, better skill than the NMME models (Figure 2.6). LIM and NMME forecasts, although, demonstrate good skill across nearly all initialization months up to six months of lead time (the exception is for CESM1 experiments initialized during boreal late fall and winter). The matrices present the high skill diagonal pattern, as for Figure 2.4. For example, enhanced predictability is found when forecasting for late fall (October/November) and, less evident, for the early spring (February - April), which can be seen as two bands of high skill extending from the lower right to the upper left (exception again is for the CESM1 model). When the experiment is initialized during boreal winter months, LIM and NMME have similar skill and nearly no significant difference in the RMSE up to six months of lead time (Figure 2.7). The exception is for the CESM1 model, that presents the lowest skill and the biggest RMSE between all the models and experiments. Including the tropics (PA) or no (ET) in the analysis do not influence in the forecast, with the results being very similar when considering same verification period. However, for the forecasts initialized during the summer, the differences are more significant: in general, LIM experiments perform better than the NMME models, in particular for the ones using the verification period of 1995 - 2010. Moreover, including the tropics to train the LIM (PA95) makes that the RMSE is the lowest for almost all lead times. In opposition, the forecast made for 1950 - 1965 without the tropics (ET50) has a smaller RMSE than the one including the tropics (PA50). These results suggest that the Tropical variability can be a source of predictability for the CCS, depending on the period that is being considered. Specifically, the ENSO-related variability may add some seasonal predictability for the CCS, since in period with major ENSO variability (1995 - 2010) there is a decrease in the RMSE when including information from the tropics; on the other hand, in a period with less ENSO variability (1950 - 1965), local variability seems to play a more important role on the predictability, what can be seen by the decease in the RMSE when excluding the tropics. Indeed, in a recent study, Jacox et al. (2017) found that the skill above the persistence for the CCS derives primarily from predictable evolution of ENSO related variability, where the mechanism prevails during years with moderate to strong ENSO events.

2.4.2 Interannual and decadal forecasts

The predictability of the interannual and decadal components of SSTa in the North and Tropical Pacific Ocean is explored using a filtered field of SST to build the LIM, as explained in Section 2.3.1. The EOFs field for the decadal, interannual and intraanual components, which the correspondent PCs were used to build the propagator matrix, are presented in Figure 2.8. Not surprisingly, the well-known PDO pattern dominates the variability for the decadal component and ENSO dominates the variability for the interannual time scales.

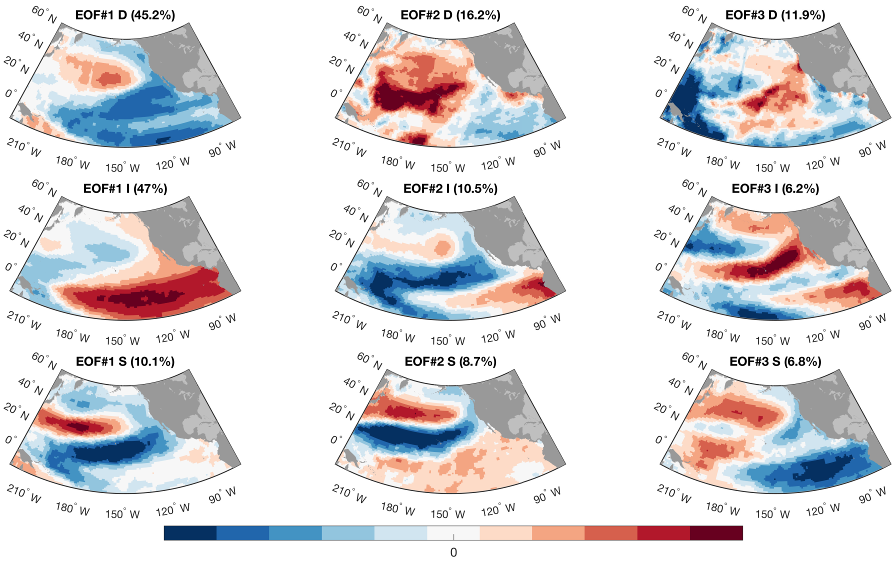


Figure 2.8: The three leading EOFs for the North Pacific and tropics for each time scale (decadal (D), interannual (I) and seasonal(S)), with the variance explained given in parenthesis.

The ACC maps for the forecast of the interannual component is presented in Figure 2.9. Although the highest skill regions up to six months are similar to those presented for the seasonal unfiltered forecast, the values of the ACC are smaller, with a maximum skill of 0.6 in those regions (against 0.8 in the unfiltered experiments). The Extratropical forcing to the tropics can also be identified for the interannual experiments by comparing Figures

2.9a with 2.9e: there is a clear loss in the skill for the tropics for the TP experiment, where the ACC has a typical value of 0.2. After one year of forecast, the skill is almost completely lost, with zero or negative values for mostly all tropics and extratropics (Figures 2.9b, 2.9f and 2.9i). Exception, though, can be found for the EP experiment (Figure 2.9i), where some low, yet positive, skill is presented on the coastal region of North America and in the Western Pacific. Although the skill values are smaller, the results are similar to those obtained for the unfiltered experiments: while the inclusion of the Extratropical data acts to increase the predictability in the tropics, the Tropical data degrades the skill for the North Pacific in seasonal to interannual time scales. Specifically, it is possible that the LIM configurations used here capture the SST signal in the PMM mechanism that adds some predictability to the tropics on seasonal time scales, and contributes to predictability on interannual timescales; moreover, the atmospheric bridge mechanism does not add forecast skill in the North Pacific for time scales shorter than decadal.

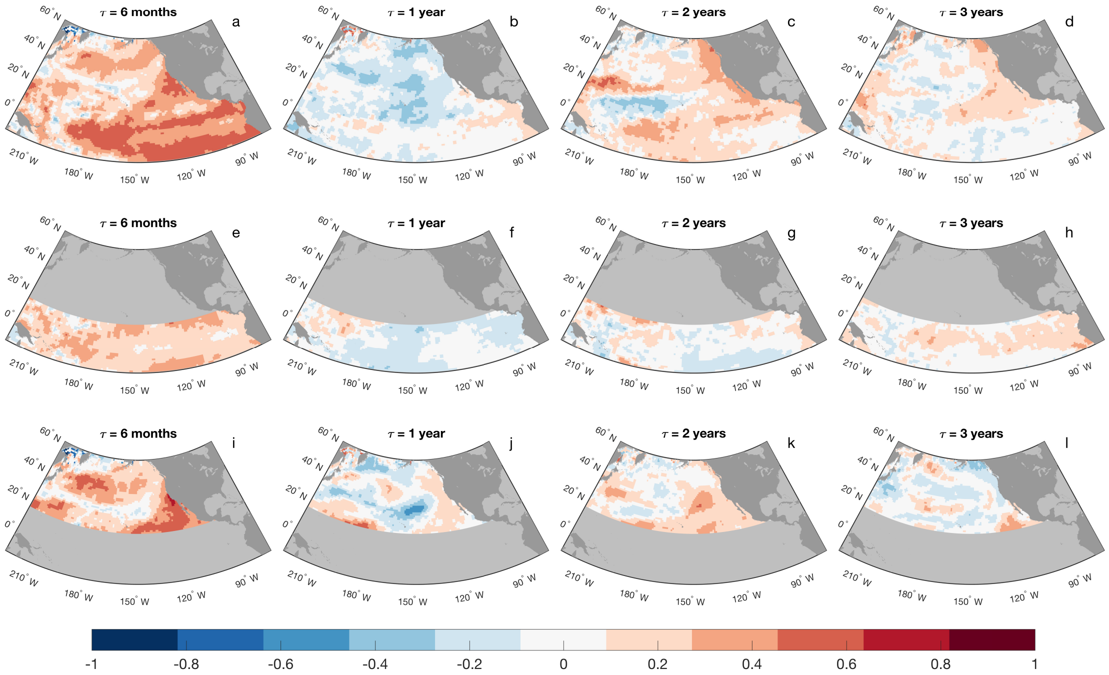


Figure 2.9: Maps of temporal correlation for the interannual component of the SST anomalies, for different lead times. LIM was constructed using both the decadal and the higher frequency components. **a - d**: LIM trained using data for the whole domain (experiment PA); **e - h**: LIM trained only with data from the Tropical region (experiment TP); **i - l**: LIM trained with data only from the extratropics (experiment ET).

One would think that after the skill is lost it can not be recovered. Curiously, this is not true for those forecasts: at two years, there is a significant increase in the skill for all the experiments, in particular for the experiment including tropics and extratropics. This gain in skill is seen in the Northeastern Pacific along the coast, in the Western Pacific north of the tropics, and in the tropics close to the dateline. The skill decreases for the forecast after three years, but it remains higher than the one at one year. Although LIM is able to identify decaying predictable signals at a certain lead time τ (represented by $\boldsymbol{B\tau}$ in Equation 2.3), a cyclic signal can be recovered by the forecast and expressed as an increase in the skill. Particularly, this might be expression of the SSTa "reemergence mechanism" (Alexander and Deser, 1995). According to this mechanism, the decorrelation time scale of midlatitude SSTa in successive winters is generally greater than one year. During winter time, temperature anomalies are formed throughout the deep mixed layer. In the spring the mixed layer suddenly shallows, and the winter temperature anomalies can remain below the thin mixed layer during spring and summer. Finally, during the fall the mixed layer deepens again, and the those deeper temperature anomalies are mixed back toward the surface. Although the reemergence is seasonally formed, it can affect the PDO variability on interannual to decadal time scales (Newman et al., 2003; Schneider and Cornuelle, 2005). Therefore, the recover in skill after one year of its damping might be an expression of the interannual influence of the reemergence mechanism over the North Pacific.

The interaction between a seasonal process driving an interannual expression can be better explored by analyzing the propagator matrix $\boldsymbol{B}$, as described in Sections 2.3.2 and 2.3.3. Figure 2.11 shows $\boldsymbol{B}$ at different lead times ($\boldsymbol{\tau}$), up to 10 years. The upscale interaction between seasonal and interannual (seasonal modes driving interannual modes) is shown in the submatrix $\boldsymbol{B}_{SI}$. This interaction decreases with lead time, but it persists up to two years, which can represent the reemergence mechanism. On the other hand, the downscale interaction (interannual modes driving seasonal modes) do not occur at any lead time (submatrix $\boldsymbol{B}_{IS}$).

The forecast for the decadal component of SSTa is presented in Figure 2.10. There is high prediction skill of the decadal part of the signal up to one year for the whole North Pacific and tropics, with skill values greater than 0.8. The exceptions are a circular-shaped region of low skill in the Northwest Pacific, being more clear in the ET experiment (Figure 2.10i). This circular region correspond to the nodal region of the first decadal EOF, as shown in Figure 2.8a. When calculating the EOFs excluding the tropics, this border is much more well defined (Figures not shown). Indeed, for the ET experiment the correlations in the circular-shaped region decrease to zero or some negative values in the second year and the skill is almost completely lost in the third year of forecast, with the exception of a coastal region between North Canada and Alaska.

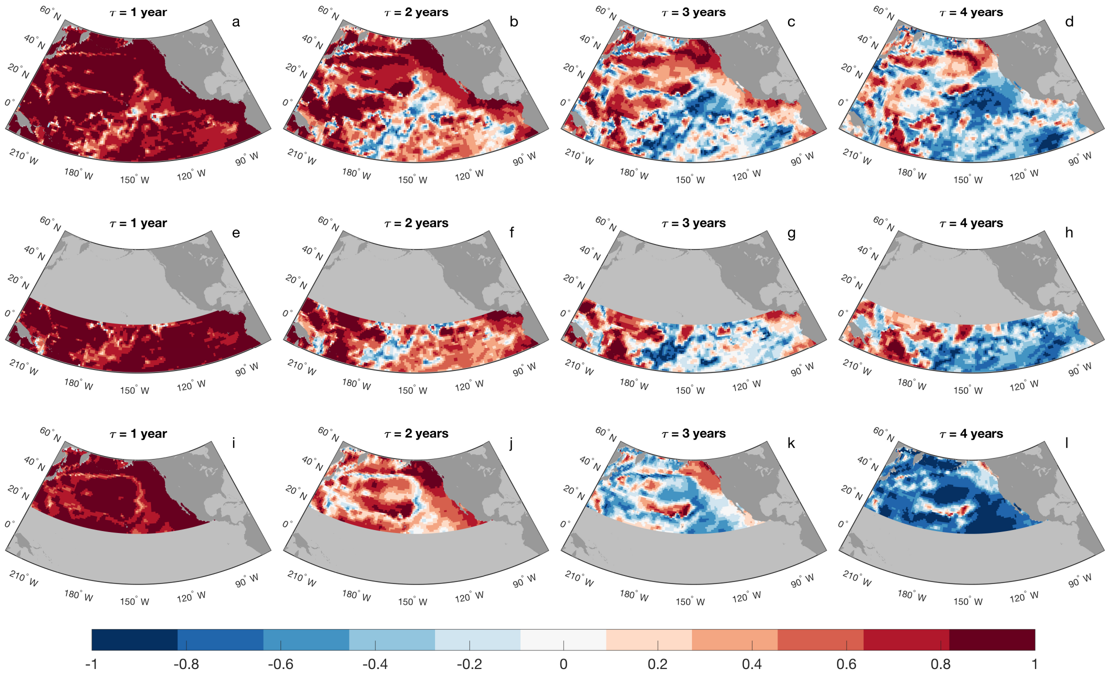


Figure 2.10: Same as Figure 2.9, but for the decadal component of SST anomalies.

The predictability of the decadal part of SSTa for the Tropical region does not seem to be very sensitive to the inclusion of the North Pacific to build the LIM, differently from the higher frequency variability, which has a significant gain in the forecast skill when including Extratropical information. There is a loss in skill in the tropics close to the dateline in the first year and beyond, when this low skill is spread eastward. This area with low skill is close to the nodal region of the EOF2 pattern for the decadal component (Figure 2.8b), which could explain the difficult of a linear model to forecast the SSTa in this region. On the other hand, the skill for the western of the dateline remains high until the fourth year of forecast (Figure 2.10a to 2.10h).

The decadal variability of SSTa in North Pacific, especially in the Western part, seems to be highly influenced by the Tropical variability. There is a clear loss in the predictability for the extratropics when excluding the Tropical region in the analysis. For the PA experiment, there is predictability up to the fourth year for some regions, while in the ET experiment the predictability is almost completely lost in the third year. Therefore, this shows a one-way interaction between North Pacific and tropics for the decadal part of the signal, where the tropics are a source of predictability for the North Pacific. On the other hand, and differently from the higher frequency signals, the North Pacific does not act to increase the predictability for the tropics. This forcing from the tropics to extratropics is consistent with several studies that have identified a communication of the Tropical decadal variability with the North Pacific, via the atmospheric bridges (Alexander et al., 2002, 2010; Zhang et al., 1997). Moreover, this process is identified as one of the contributors to the PDO related variability (Newman et al., 2016).

The propagator matrices (Figure 2.11) can be used to better explore how this forcing works. The modal interactions between decadal variability with itself (submatrix $\boldsymbol{B}_{DD}$) starts to play a role in the third month of lead time, and gets stronger with the increasing of lead time. This modal interaction can be represented by the communication of Tropical decadal variability to the extratropics, as stated before (e.g, Newman et al., 2016). Interestingly, though, the interannual variability appears to drive decadal variability at longer lead times. This effect is shown in the submatrices correpondent to the upscale interaction between interannual and decadal modes (submatrices $\boldsymbol{B}_{ID}$, interannual modes driving decadal modes). We observe that this modal interaction starts with about six months and increases up to 10 years. This suggests that not only the decadal Tropical variability adds some predictability to the extratropics, but also the interannual variability. Although with the propagator matrix it is not possible to identify the regions where this variability comes from, the leading EOFs allow us to identify the patterns with that dominate the variability at each time scale (Figure 2.8). We showed that a PDO-like patterns dominates the decadal variability, while an ENSO-like patterns dominates the interannual variability. Therefore, when excluding the tropics from the model, the predictable signals from the tropics, either the decadal and the interannual ENSO, are lost and the predictability for the PDO region is damped.

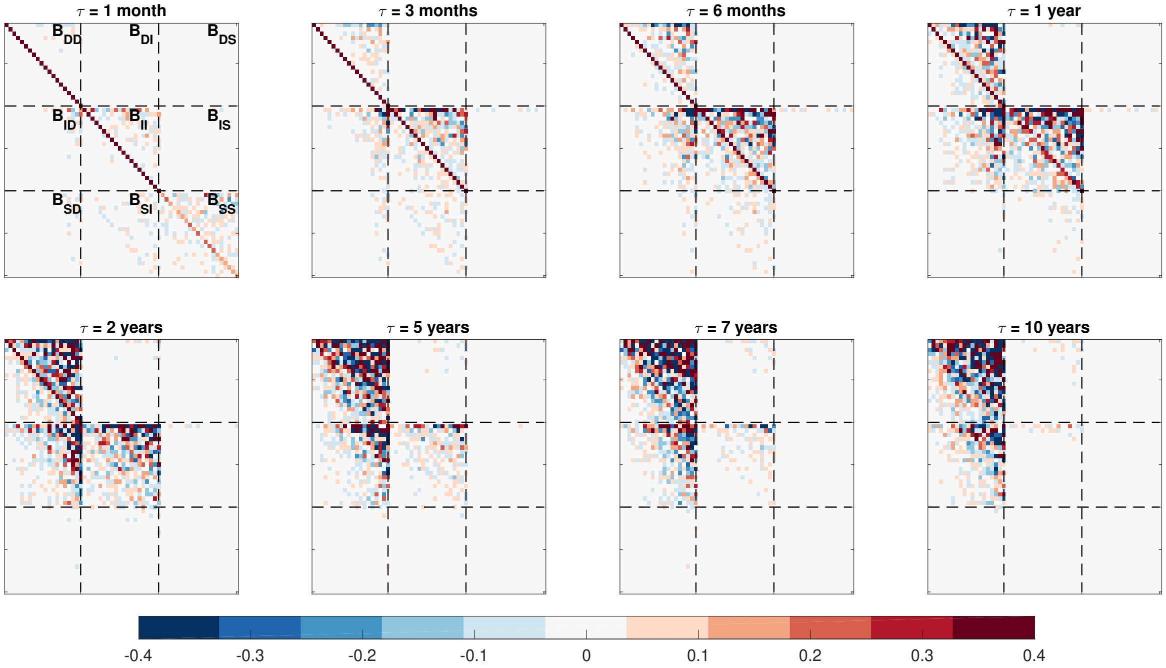


Figure 2.11: Propagator matrices constructed from the 20 lead EOFs of each time scale using data from the whole domain (PA), at various lead times. The dotted lines indicated are the boundaries for the submatrices of the interactions between each time scale. $\boldsymbol{B}_{DD}$, $\boldsymbol{B}_{II}$ and $\boldsymbol{B}_{II}$ represent the interaction between each time scale with itself; $\boldsymbol{B}_{DI}$ and $\boldsymbol{B}_{DS}$ represent the decadal downscale interactions (decadal modes driving interannual and seasonal modes); $\boldsymbol{B}_{IS}$ represents the interannual downscale interaction (interannual modes driving seasonal modes); $\boldsymbol{B}_{ID}$ represents the interannual upscale interactions (interannual modes driving decadal modes); and $\boldsymbol{B}_{SD}$ and $\boldsymbol{B}_{SI}$ represent the seasonal upscale interactions (seasonal modes driving interannual and decadal modes).

2.5 Summary and concluding remarks

The forecast skill of Pacific sea surface temperature anomalies (SSTa) on seasonal, interannual, and decadal time scales has been tested using a suite of linear inverse models (LIM). The forecast skill was first evaluated to the North Pacific plus the Tropical Pacific (15° S to 60° N) and then separately for each of those regions. By separating the variability into these time scales and regions, we aimed to identify how interactions between time scales affect predictability and how the two regions affect each other in driving predictable components linked to ENSO or intrinsic mid-latitude interactions.

For the seasonal time scale, our results revealed that LIM has a good forecast skill ($\boldsymbol{\rho} > 0.6$) in some areas up to 9 months. However, this skill varied for each region included in the model. Specifically, we showed that the inclusion of data from the extratropics enhanced the forecast skill of the tropics, suggesting that there is a communication between those two regions on seasonal time scales. Moreover, since LIM assumes that the evolution of a system is approximated by linear dynamics, the mechanism associated with this interaction might be essentially linear. Therefore, we suggested that this enhanced predictability seen in the Tropical regions for LIM-PA can be due to the Pacific Meridional Mode (PMM) propagating the mid-latitude SST anomalies associated with the North Pacific Oscillation (NPO).

Those results were also contrasted to the GCMs of the North-American Multi Model Ensemble (NMME) for two regions in the Pacific: Niño 3.4 and the California Current System (CCS). Both NMME and LIM forecats have comparable skill for those regions. Specifically, in Niño 3.4 region, there is a clear expression of the Spring Predictability Barrier (SPB), when there is a clear decrease in the skill to forecast the SSTa for the boreal spring. LIM assumes that the nonlinear part of the signal is unpredictable, so this skill comes exclusively from the linear part. Since the skill of an empirical linear model (LIM) is comparable to the skill of a high dimensional nonlinear coupled GCM (NMME) for certain months, we suggested that the nonlinear SSTa variability for this period is either essentially stochastic or not well represented in current GCMs.In the CCS, LIM overall also has a similar performance with the NMME models: both models have good skill across nearly all the initialization months up to six months of forecast. By comparing LIM experiments (including or excluding the tropics), Tropical and local variability seems to play a role in the CCS variability, depending on the period that is being considered. In a period with high (weak) ENSO variability, the Tropical signal increases (decreases) the skill for the

CCS forecast.

The forecast skill for a LIM formulated by using modes of variability from a filtered time series was performed to evaluate the long term predictions. For the interannual part of the signal, the extratropics act as a source of predictability to the tropics. Additionally, the regions with the highest skill in the interannual forecasts are very similar to the ones in the seasonal forecasts. We suggested that the PMM mechanism adds predictability to the tropics in seasonal to interannual time scales when forecasting up to one year. The LIM configured only with the decadal component has good forecast ($\boldsymbol{\rho} > 0.8$) up to three years for most part of the Pacific, especially for the extratropics and for the experiment including the tropics. Therefore, the interaction between tropics and extratropics for the decadal components acts the other way around when compared to the higher frequency variability: there is a significant increase in the forecast skill in the extratropics when including the tropics, although the Extratropical region does not act to improve the forecast skill for the tropics. We suggested that this forcing occurs due to the atmospheric bridge mechanism, that is one of the contributors to the PDO-related variability.

Finally, we explore the coupling among time scales using the propagator matrices (Eq. 2.3), that show the modal interaction between the leading modes of each filtered time scale, as presented in Eq. 2.9. We show that the upscale interaction between interannual and decadal modes persists up to 10 years, which is an indication that the predictability that is added from the tropics corresponds not only to Tropical decadal variability, but also the interannual. These results indicate the importance of temporal scale interactions in improving the forecast skill on decadal timescales.

Acknowledgements

We are grateful for the National Science Foundation (OCE1419306) and the National Oceanic and Atmospheric Administration (NOAA-MAPP; NA17OAR4310106) for funding that supported this research. DFD was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the grant 221222/2014-6. We also thank an anonymous reviewer for the comments and suggestions.

Chapter 2, in full, is a reprint of the material as it appears in Dias, D.F., A. Subramanian, L. Zanna, and A. J. Miller (2018), Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study, Climate Dynamics. The dissertation author was the primary investigator and author of this paper.

Chapter 3

The influence of sea surface temperature and soil moisture in seasonal predictions of air temperature over Western North America

Abstract

Seasonal predictions have the potential to improve the management of different sectors of the society by anticipating climate fluctuations and possible weather extremes. Such forecasts must contend with a high level of natural variability as well as challenges posed by climate change. However, they are constrained by limited understanding of local and regional atmospheric predictability. Here, a canonical correlation analysis (CCA) prediction model of minimum and maximum air temperature anomalies (Tmin and Tmax) over Western North America (WNA) is developed. Remote and local predictors are used: sea surface temperature (SST) across the Pacific and local soil moisture (SM). The evaluation of the skill of predicted air temperature using historical observations indicates that CCA can provide skillful predictions for seasonal anomalies of air temperature over the region. However, skill is found to vary over seasons, location and combination of predictor and predictand variables. SST yields the best predictive skill for Tmax and Tmin during wintertime, but for spring and early-summer its influence is mostly on Tmin. Remote large-scale patterns, in the form of climate indices, are captured by the CCA canonical modes and it is shown that they can be responsible for this predictive ability. On the other hand, the influence of SM is restricted to Tmax and only during the winter, when it is shown that SM has the highest autocorrelation for the region. The results demonstrate the importance of careful analyses that consider season, variable being predicted, and predictors in forming statistical forecast models to be used for decision making.

3.1 Introduction

Skillful predictions of climate fluctuations at seasonal time scales have many economical and societal applications. Anticipating the climate fluctuations, and possible associated weather extremes, at one or more month in advance would benefit decision-making relating to hydrology, agriculture, health, energy, and other sectors. Although the dynamics of the climate system display important nonlinear features, seasonal predictions can be made with linear statistical methods with skill comparable to that obtained from nonlinear statistical methods (Tang et al., 2000; Van den Dool, 2007). The linear approaches typically seek to estimate linearly related patterns of variability in the predictor and the predictand fields, such as by canonical correlation analysis (CCA, Barnston, 1994; Gershunov and Cayan, 2003; Wilks, 2008). The advantage of such statistical forecasts is that they can be formulated and implemented easily and they perform economically, and the forecast skill they produce can be evaluated in a straightforward manner. Moreover, such techniques are not “black boxes", in the sense that the sources of predictability can be easily identified and examined.

The scientific basis for seasonal atmospheric climate predictability arises from the fact that variations in slowly changing influences, such as oceanic and terrestrial conditions manifested at the Earth’s surface, in particular those varying on time scales of several days to several months, can affect atmospheric circulation and, accordingly, the climate at the surface. Ocean thermal anomalies often persist for or evolve systematically over several months, and many studies have shown that slowly evolving sea surface temperature (SST) anomaly patterns can be a source of predictability for seasonal weather and climate anomalies a months in advance (e.g, Barnett and Preisendorfer, 1987; Barnston, 1994; Gershunov and Barnett, 1998; Gershunov and Cayan, 2003; Xoplaki et al., 2003). This is particularly important for North America, where low-frequency variability in Pacific SST climate modes influences atmospheric circulation over the Pacific-North America sector (Hartmann, 2015; Horel and Wallace, 1981; Mantua et al., 1997).

Regional influences may also play a role in influencing the variability of air temperature over land. In particular, the variability of soil moisture has been shown to affect the surface energy balance (Koster et al., 2000) and thereby enhance predictability for seasonal time scales (Alfaro et al., 2006; Seneviratne et al., 2010). This may be the case when the impact of SST fields on the atmospheric variability is modest, which has been demonstrated during the summer season in mid-latitudes (Koster et al., 2000; Shukla, 1998; Trenberth et al., 1998). Moreover, the predictability of hot extremes over most areas of South and North America, Europe, Australia, and parts of China was enhanced when antecedent precipitation was introduced (Quesada et al., 2012). Therefore, soil moisture may increase the skill for seasonal predictions at some specific locations and seasons.

Alfaro et al. (2006) studied associations of Pacific SST and the Palmer Drought Severity Index (PDSI) with summer maximum and minimum temperatures over the central and western United States during 1950-2001. They found that large-scale and remote climate conditions exert their influence more strongly on minimum rather than maximum temperature. Interestingly, summer surface maximum air temperatures were more predictable than those in winter, despite summer temperatures having less spatially coherent and lower amplitude anomaly patterns than those in winter. This was attributed to the additive effect of soil moisture, presumably augmenting effects of large-scale patterns, represented by SST.

The goal of this paper is to use a linear statistical model to evaluate the predictive skill of seasonal average maximum and minimum land surface temperature over WNA using up-to-date finely resolved temperature and land surface moisture datasets. Additionally, the role of local and remote predictors in the predictive skill is investigated. The paper is organized as follows. Section 3.2 describes the predictive approaches used in this study, including the description of the statistical model, the data used and the skill evaluation and comparison. Section 3.3 shows the results of the optimized models that were used to make the seasonal predictions of Tmax and Tmin. Section 3.4 explores the climate patterns and their relationships that are responsible for the predictive skill and Section 3.5 shows the results for the statistical forecasts. Finally, Section 3.6 presents a summary of the main findings and some concluding remarks.

3.2 Data and Methods

The domain includes the southwestern Canada and the western United States (including the Rocky Mountains), and northwestern Mexico. Detail of the domain with the topography is found in Figure 3.1. Following results demonstrated in Alfaro et al. (2006), we adopt a linear statistical approach using canonical correlation analysis. The variables and the data used as predictors and predictands, as well as the details of the implemented model, are described in this section.

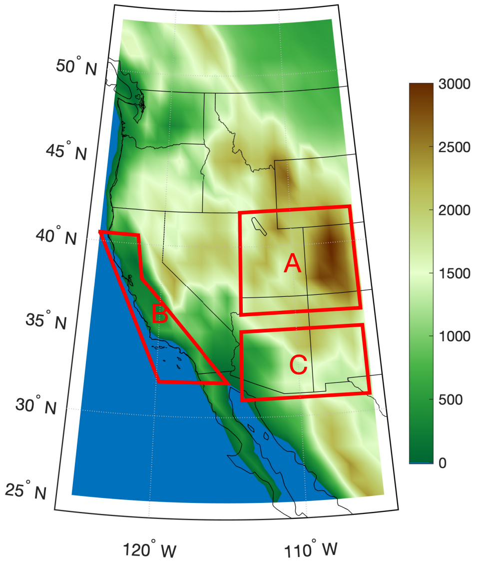


Figure 3.1: Region defined in this study as the WNA and the topography (in meters). Red boxes correspond to the three regional indexes used, named: A) Central Rockies; B) Coastal California; and C) Inland Southwest.

3.2.1 Data

Predictands

The variability of daytime temperatures, represented by time averages of the daily maximum temperature (Tmax), differ from that of nighttime temperatures, represented by time averages of daily minimum temperature (Tmin). From a statistical perspective, Tmax anomalies have higher amplitude (larger daily, monthly and seasonal fluctuations) than do Tmin, especially in summer months (Alfaro et al., 2006). Additionally, seasonal anomalies of Tmax and Tmin over WNA are at best (in winter months) correlated moderately, and at worst (late summer-fall) correlated poorly (Figure 3.2). This lack of strong correlation and high variability throughout the year evidently reflects the different mechanisms that affect Tmax and Tmin and, quite possibly, differences in the pattern and degree of their predictability in different seasons. Additionally, Tmax and Tmin anomalies may have different societal consequences. From an energy utility perspective, Tmax and Tmin anomalies may have different effects on the supply and on the demand for energy. For example, in summer, extreme hot Tmax increase demand for electricity because of the use of air conditioners, but usually occur under clear skies, which results in high solar energy production. Warmer summer Tmin would diminish beneficial diurnal cooling relief from hot days and thus might increase the magnitude and duration of air conditioning electrical load, which could increase system failures. In winter, other kinds of impacts may occur. In cooler parts of the domain, warm Tmin anomalies may lead to less space heating and thus lower electrical and natural gas demand, whereas cool Tmax anomalies may provoke higher indoor energy requirements from increased space heating and other indoor utility use. From a public health perspective, the expression of the impacts may differ between Tmin and Tmax. Over the Southwest of North America, heat waves are becoming not only stronger, longer and more extensive, but also more humid. High humidity, which exacerbates the nighttime expression of heat waves, can have greater and specific impacts on public health that may require different intervention strategies to manage (Gershunov et al., 2009).

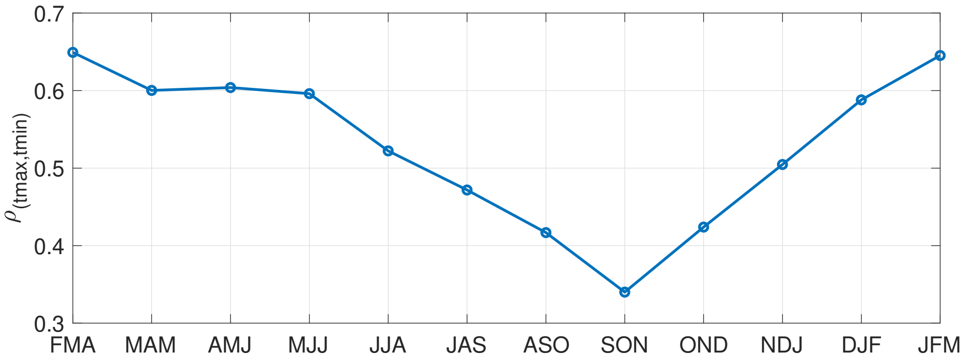


Figure 3.2: Correlation between seasonal averages of Tmax and Tmin anomalies averaged over the WNA. Each season corresponds to a 3-months average of Tmax and Tmin anomalies, represented as: **FMA**: February, March and April; **MAM**: March, April and May; **AMJ**: April, May and June; **MJJ**: May, June and July; **JJA**: June, July and August; **JAS**: July, August and September; **ASO**: August, September and October; **SON**: September, October and November; **OND**: October, November and December; **NDJ**: November, December and January; **DJF**: December, January and February; **JFM**: January, February and March.

In view of the above, in our study of air temperature predictability over the WNA (latitudes 25°N to 53°N and longitudes 105°W to 125°W), we investigate a separate set of forecast models for Tmax and Tmin. The 1950-2013 period is used in the model development and validation. The predictands variables considered are the seasonally averaged Tmin and Tmax, respectively from the Livneh CONUS near-surface gridded meteorological data (Livneh et al., 2015), provided by the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (ESRL/NOAA). These datasets are gridded at a spatial resolution of 1/16° by 1/16° (approximately 6km square) and are derived from daily temperature observations from approximately 20,000 NOAA Cooperative Observer (COOP) stations (Livneh et al., 2015). We use seasonal averages of Tmax and Tmin anomalies because they represent short period climate fluctuations over multi-day time scales. The predictand datasets are obtained by first calculating the monthly anomalies for each variable, and then calculating seasonal means of the Tmin and Tmax anomalies for 12 different seasons, each of those represented as a 3-months average of temperature anomalies.

Predictors

Two fields of predictor variables are used. The first predictor field is Sea Surface Temperatures (SST) anomalies for the Pacific Basin (15°S to 60°N and 135°E to 110°W), which are obtained from the Hadley Center Sea Ice and Sea Surface Temperature (HadISST, Rayner et al., 2003). The HadISST data has a resolution of 1° by 1° and it can be obtained at the following website: https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html.

The second predictor field is soil moisture (SM) anomalies, which is applied individually and also in combination with SST (SM+SST). In contrast to prior studies, that have used soil moisture indices such as the Palmer Drought Severity Index (PDSI), SM employed here is estimated from an existing hydrological model reanalysis produced by the Variable Infiltration Capacity (VIC) model, forced by observed daily precipitation and temperature. VIC is a macroscale hydrological water and energy balance accounting model (Liang et al., 1994; Cherkauer, 2003) that has been used in numerous studies of climate and hydrological variability and changes (e.g., Hamlet et al., 2007; Das et al., 2009). The forcing data is the Livneh et al. (2015) gridded daily precipitation and temperature historical dataset, which covers the conterminous United States and adjoining portions of Canada and Mexico. This version of VIC is forced at the same resolution as the Livneh predictand fields, resulting in SM data that is distributed over the identical 1/16° grid. The VIC SM domain is also identical to that of the Tmax and Tmin predictands, covering the land area from 25°N to 53°N and 105°W to 125°W.

SST and SM were selected as predictors because they contain some measure of climate memory and they are broadly representative of large, regional and local scales climate measures that vary over time scales that are relevant to seasonal air temperature fluctuations. While Tmax and Tmin are affected by several other variables such as winds, clouds, and topographic influences, we have not included those here since our purpose is to explore linear prediction using a plausible, readily available, and manageable predictor set.

Therefore, it is necessary to include a set of predictors that contain a measure of climate memory, so that it makes sense to use antecedent observations of these fields as predictors. Both predictor datasets consist of monthly means from 1950 to 2013, from which monthly anomalies are calculated by removing the climatological (1950–2013) monthly mean for each grid cell and each month.

3.2.2 Model details

The statistical model framework implemented to make the seasonal predictions of air temperature over the WNA uses Canonical Correlation Analysis (CCA), following an approach that has been used in several prior climate prediction and analysis studies (e.g., Barnett and Preisendorfer, 1987; Gershunov et al., 2000; Gershunov and Cayan, 2003; Alfaro et al., 2006). This method explores and identifies the linear combinations between two sets of variables that have the greatest correlation with each other, seeking to match patterns in the predictor fields (SST and SM) with patterns in the predictand fields (Tmin and Tmax) whose temporal evolutions are optimally correlated. For many climate-related phenomena, the CCA approach condenses much of the spatial and temporal co-variability into a few modes and it provides a simple and cheap way to predict one field of variables from another.

This methodology follows the hypotheses addressed by Alfaro et al. (2006): on one side, the anomalous Pacific SST causes changes in the large-scale atmospheric circulation, which in turn influences the land surface temperatures over a broad sector; on the other side, soil moisture modulates the local air temperature via the interaction between latent and sensible heating. With those hypotheses addressed, a CCA prediction model is constructed for each season, employing predictors formed by one month antecedent SST and SM. For example, to make the Tmin one-month lag winter prediction using SST as predictor, we select all Novembers of SST and all DJFs of Tmin.

The predictor and predictand fields are pre-filtered separately with the same number of p principal components (PCs), which are statistically orthogonal patterns of spatial and temporal variability, ordered by amount of variance explained. Those patterns are then related to each other using the q canonical correlates (CCs) extracted from the CCA analysis. The number of p PCs and q CCs are determined using a skill optimization scheme (Section 3.23.2.3). For the predictor using both SST and SM combined (SST+SM), we separately calculate their p PCs and then we join those PCs together ($PC_{SST+SM} = PC_{SST} + PC_{SM}$). After that, the procedure is similar to that for single-variable predictors with p and q for SST+SM taken to be the sum of p and q obtained for SST and SM separately.

Importantly, to avoid artificial skill from over-fitting, all results pertaining to model performance are cross-validated, wherein the year of a given prediction has been left out of the model development. Separate models are made for each year that is being predicted, in which the training period of each year's model is made using a period that excludes the season and adjacent months being predicted. This is implemented by excluding the year predicted from the training period (month for the predictor and lagged-season for the predictand). To calculate the confidence intervals for our CCA forecasts, we perform some randomized CCA experiments: first, the time series of the predictors are randomized and used to build the CCA model to forecast air temperature. This procedure is repeated 1,000 times and the forecast skill for each of those repetitions is calculated. Finally, using the skill obtained with this randomized experiments, for all these repetitions, the confidence intervals are calculated.

3.2.3 Skill optimization

Given the cross-validated forecasts, an evaluation is conducted of the yearly sequence of resulting forecasts compared to the observed seasonal temperatures to estimate forecast skill. To avoid artificial skill from over-fitting, the models are optimized, wherein skill over a range of predictor and predictand membership is evaluated under a cross-validation scheme. For each of the forecast model cases (season, predictor, and predictand), a certain number of p PCs and q CCs is determined by finding the optimum model, i.e., the model complexity (the combination of PCs and CCs) that maximize the forecast skill (Gershunov and Cayan, 2003). The model is calculated using all different combinations of PCs and CCs, from 2 to 18 and with PCs $\leq$ CCs. Therefore, $p \leq q \leq T$, where T is the number of years available for the training period.

For each option of model complexity, the cross-validated skill is estimated for each grid point and then summarized in one value by averaging the correlation of all the points. In practice, we use the field averaged correlation between the predicted and observed air temperature as a measure of forecast skill when determining the optimum model. Therefore, for each model combination (i.e., predictor-predictand-season), a different p and $q$are selected to maximize the forecast skill.

3.2.4 Skill evaluation

We first evaluate the skill of all the CCA forecast models by calculating the anomaly correlation coefficient (ACC) between each model's forecast result and the correspondent observation. We analyze the seasonal variations of the predictive skill for each combination of predictor, predictand and season by calculating the field-averaged ACC. The same is done for the randomized experiments to verify the significance of the skill values obtained with the CCA forecasts.

To explore the independence of SST and SM in producing predictive skill, i.e., how much skill can be added for each predictor separately, we perform some CCA forecast experiments for the residuals: first, the difference between Tmin and Tmax predicted and observed fields is calculated (residuals); secondly, those residuals fields are used as a set of new predictands, wherein the residuals obtained with the model using SST are predicted with the SM field (SST_{res}) and likewise the residuals obtained with SM are predicted with the SST field (SM_{res}); finally, the skill is calculated for each of those residuals experiment and it is contrasted with the original CCA forecast skill.

Once we verify how the skill varies seasonally in field-averaged sense, we also explore regional differences in the skill by calculating maps of ACC for the CCA models.

3.3 Identification of the optimum model

The optimum model complexity (i.e., numbers of PCs and CCs) used for each experiment is presented in Table A.1. Those choices were made based on the skill optimization matrices, in which the field-averaged skill is calculated for each combination PCs and CCs. To reduce the number of calculations required in this exercise, the number of PCs representing the predictor and predictand fields is fixed as the same. As an example, Figure 3.3 shows the skill optimization matrices for JFM using all the combinations of predictors and predictands. The skill values vary smoothly as the number of PCs (x-axis) and CCs (y-axis) changes, therefore, similar predictive skill is achieved by different model complexities (p PCs and q CCs). In proceeding with the model experiments, in the spirit of parsimony, i.e., “simpler is better”, we chose the simplest model complexity (the fewer PCs/CCs number) that yields near-optimal predictive skill. Surveying the model results, the optimization produces different combinations of predictors/predictands and canonical correlations produced for different seasons. The model complexity ranges from very simple models, where few PCs and CCs results in the maximum predictive skill, to more complex models, where a greater number of modes is required. For example, the FMA SST-Tmin model gives the greatest skill when using only four PCs and CCs modes, while the NDJ SST-Tmax is optimized when $p = 18$ and $q = 10$ (Table A.1). This is a hint that the sources of predictability for Tmin and Tmax vary according to the predictor and the season that is

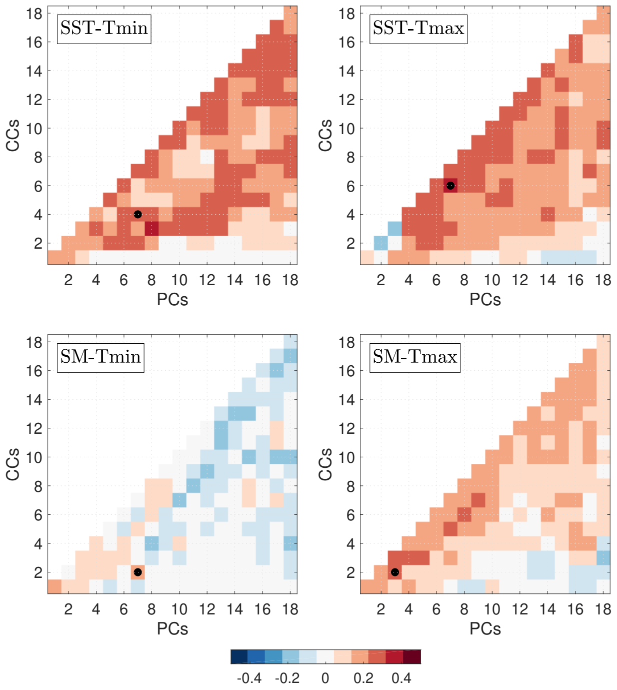


Figure 3.3: Skill optimization matrices for JFM season for CCA forecasts using different combinations of predictors (SST and SM), predictands (Tmin and Tmax), and seasons (FMA to JFM). Each matrix represents one specific predictor-predictand combination (named in the top left of each individual matrix). Color scale represents the cross-validated field-averaged skill, expressed as the correlation between the predicted and the observed temperatures, using different combination of the Principal Components (PCs) and the Canonical Correlates (CCs). The optimum is defined the number of PC and CC modes that yielded the highest cross validated skill (black dots in each matrix).

being analyzed. Therefore, it appears to be important to perform this skill optimization analysis across each case to maximize the potential air temperature forecast skill.

3.4 Relationships between climate patterns

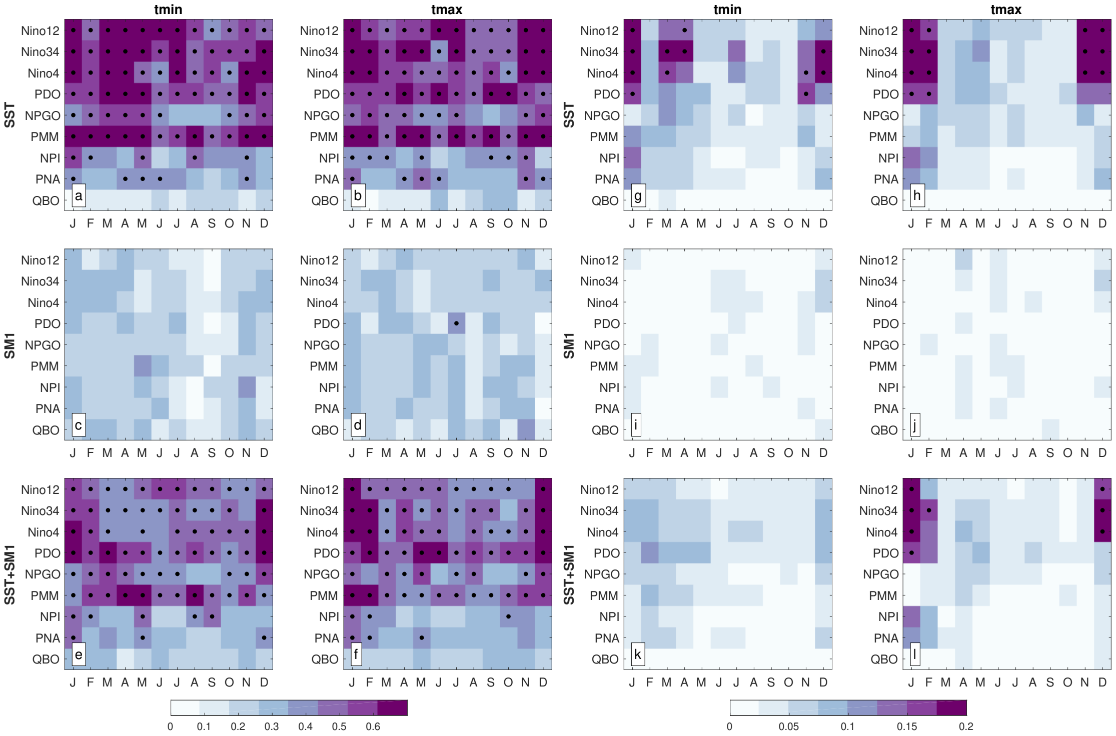
Before exploring the predictive skill, it is important to elucidate the spatial patterns and their time variations as described by the diagnostic application of CCA. We aim to provide some insight into the physical sources of predictability that arises from this analysis. As shown in Section 3.23.2.2, CCA condenses the spatial and temporal variability of the variables that are being used to build the model based on the salient patterns. As a result, we have a set of spatial patterns and their time variations, known as canonical correlates or canonical modes (CCs), which provide a set of modes ordered by the correlation between the CCs of the predictor (SST and SM) and the predictands (Tmax and Tmin). Those CC modes can be interpreted as climate-scale associations between predictors and predictand and they can be tested for possible associations with commonly known climate variation patterns at different timescales. To do so, we compare each CC mode from all the CCA models with several well-known oceanic and atmospheric climate indices in the Pacific (Table 3.1). This comparison is made by first calculating the correlation between each climate mode with the temporal variation of each CC mode for all the combination of the three predictors, two predictands and 12 seasons. After that, the absolute value is calculated (since both high positive or negative correlations are significant and may represent some physical association). Additionally, this correlation value is weighted by the amount of variance explained in each of those CC modes. With that, it is possible to get some insight about the physical patterns that may give rise to seasonal predictability skill for air temperature in the WNA.

Table 3.1: Description of each climate index used in comparison with canonical correlates in Figure 3.4

Index	**Name**	**Description**
Nino1.2	ENSO	Extreme Eastern Tropical Pacific SST: SST anomalies between 0 – 10°S, 90°W – 80°W
Nino3.4	ENSO	East Central Tropical Pacific SST: SST anomalies between 5°N – 5°S, 160°E – 150°W
Nino4	ENSO	Central Tropical Pacific SST: SST anomalies between 5°N – 5°S, 170°W – 120°W
PDO	Pacific Decadal Oscillation	The leading principal component of SST anomalies in the North Pacific (Mantua et al., 1997)
NPGO	North Pacific Gyre Oscillation	The second dominant mode of sea surface height variability in the Northeast Pacific (Di Lorenzo et al., 2008)
PMM	Pacfic Meridional Mode	Meridional variability in the tropical Pacific Ocean. It is defined as the first mode of a maximum covariance analysis (MCA) of SST and the zonal and meridional components of the 10m wind field, over the region defined between 21°S-32°N and 74°W-15°E (Chiang and Vimont, 2004)
NPI	North Pacific Index	The area-weighted sea level pressure over the region 30°N-65°N,160°E-140°W (Trenberth and Hurrell, 1994)
PNA	Pacific North American Pattern	Rotated Principal Component Analysis applied to monthly mean standardized 500-mb height anomalies over the region between 20°N-90°N (Barnston and Livezey, 1987)
QBO	Quasi-Biennale Oscillation	Oscillation of the equatorial wind between easterlies and westerlies in the tropical stratosphere (Baldwin et al., 2001)

The climate modes considered here have been identified in numerous previous studies, capturing regional to global scale variability at the ocean surface in the Pacific basin and in the atmosphere near Earth's surface and aloft in the mid-troposphere. Figure 3.4a-f shows the maximum absolute correlation value among all the canonical modes used for each CCA experiment (the number of canonical modes included is shown in Figures 3.3). High values indicate that the time series of the canonical mode, and as consequence its spatial pattern, resembles the climate mode that is being analyzed.

Figures 3.4a and 3.4b show that throughout the year several canonical modes are closely related to large-scale features in the Tropical and Extra-tropical Pacific ocean. The modes derived from SST exhibit significant and substantial correlation with oceanic climate indices over a range of time scales: seasonal, including the the Pacific Meridional Mode (PMM); interannual, such as several indices used to represent El Niño Southern Oscillation (Niño 1.2, Niño 3.4 and Niño 4); and decadal, such as the Pacific Decadal Oscillation (PDO). For both Tmin and Tmax, there are significant correlation for all the seasons (black dots). However, the canonical modes derived for Tmin have stronger correlation throughout the year, particularly for the winter (December and January) and for the spring (April and May). In both of these seasons, the correlation is higher than 0.6 for all the modes cited above (PMM, ENSO and PDO). The correlations with the atmospheric modes (NPI and PNA) are weaker ($\rho < 0.45$), but still significant for the late fall and winter. For Tmax, significant correlations are also found for those Tropical and Extratropical SST-related modes, but the strongest correlations are found for late fall and winter (November to February).



(a) (b)

Figure 3.4: Maximum absolute correlation between all the canonical modes (CCs) from each CCA experiment and selected climate modes **(a - f)**; maximum absolute correlation weighted by the variance explained by each canonical mode **(g - l)**. Each matrix is a combination of a predictor (SST, SM and SST+SM) and a predictand (tmin and tmax) wherein the columns represent the months that each CCA model was initialized (January to December) and the lines represent the climate modes (named along the lines of each matrix and described in the Table 3.1). Black dots indicate values that have a statistical significance greater than 99%.

The order of the canonical modes that represent those known-climate variations provides insight into the common variability between the predictor and the lagged-predictand and and thus provide a better understanding of the nature of the predictive skill. The modes in CCA are ordered by the correlation between the predictor and the lagged-predictand canonical correlates. If the climate patterns are represented only in lower-order canonical correlates, the variability shared between predictor and predictand is probably smaller and it may decrease the ability to predict the variations of air temperature for that season. To investigate that, we show in Figures 3.5a and 3.5b in which canonical modes of the

SST-Tmin and SST-Tmax experiments the majority of the climate indices are represented. This figure presents the number of times that the correlation between a canonical mode and a climate mode is larger than 0.6, for each month and canonical mode number. For example, in Figure 3.5a, in January the second canonical mode presents correlation higher than 0.6 with three climate modes (the three ENSO indices), whereas the third canonical mode has correlation higher than 0.6 with one climate climate mode (PDO). In June, only the 12th canonical mode has substantial correlations with two climate modes (Niño 1.2 and Niño 3.4). In general, for both Tmin and Tmax, the higher number of those strong correlations occur within the first five CC modes. More occurrences are concentrated in the second canonical mode during the late fall and winter (for Tmax) and during the winter and the spring (for Tmin). In the spring, association between those climate patterns and the Tmin occur between the second and the fifth canonical modes.

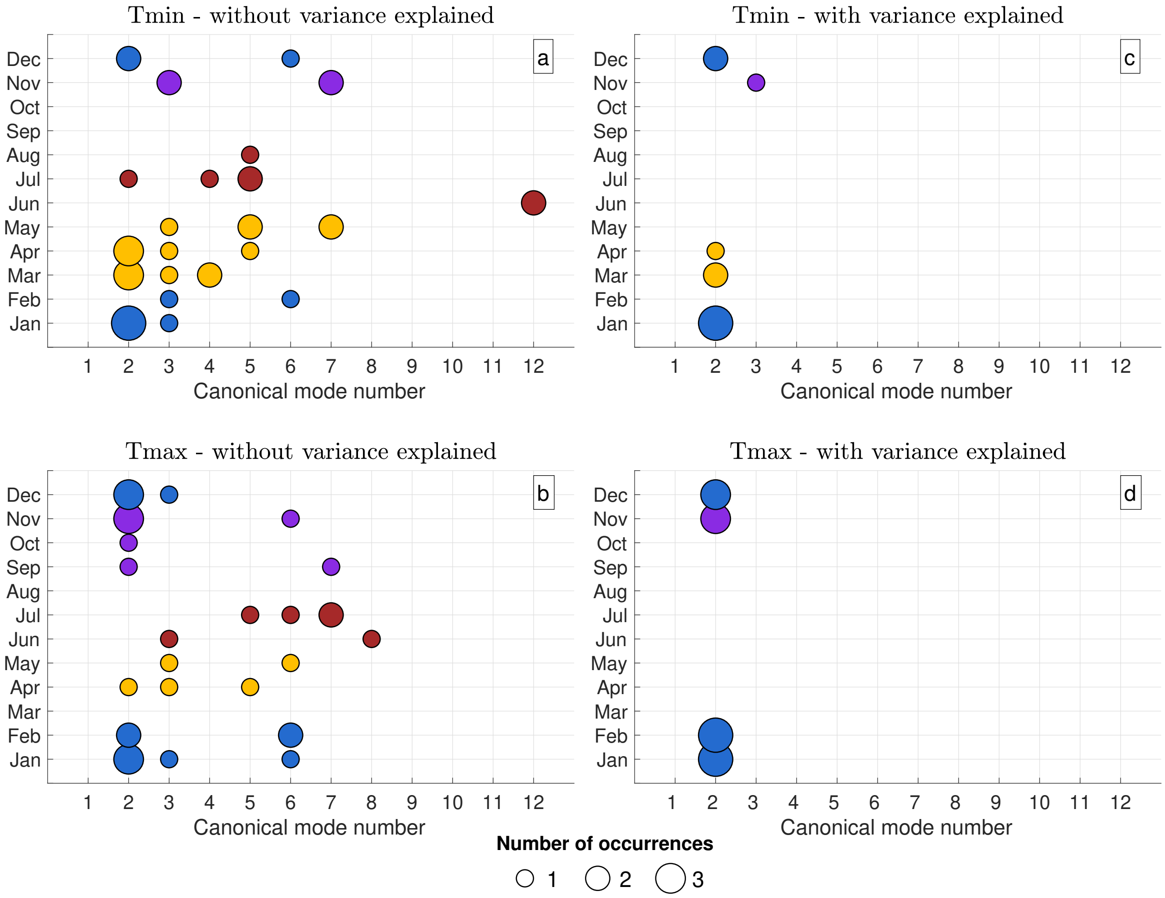


Figure 3.5: Number of occurrences that the correlation between the SST canonical modes and the climate mode is bigger than a threshold, for each initialization month (y-axis), and each canonical mode number (x-axis). **(a)** and **(b)** bring the results for Tmin and Tmax, respectively, not considering the variance explained and which the threshold value is 0.6; **(c)** and **(d)** bring the results for Tmin and Tmax, respectively, weighted by the amount of variance explained for each canonical mode and which threshold value is 0.2.

It is important to consider how much of the variance is explained by each of the canonical modes and how this is related to the climate patterns they represent. We show in Figures 3.4g-l the same correlations of Figures 3.4a-f weighted by the variance explained for each canonical mode. Despite the fact that the canonical modes present significant and substantial correlations with the climate modes throughout the year, when considering the variance explained the significant correlations only hold for the late fall and winter (for Tmax) and for winter and spring (for Tmin). This can also be seen in Figures 3.5c and 3.5d, showing that high values are restricted mostly to the leading CCA modes during the winter and fall. The only exception is for Tmin in April and May, when the second canonical mode displays significant correlation and variance explained (Figure 3.5c), represented exclusively by the Niño 3.4 index (Figure 3.4g). These results are consistent with many previous studies that have explored the effect of those climate patterns in the WNA. For the winter, the ENSO-PDO-related variability has been shown to affect circulation anomalies that, in turn, influence synoptic weather patterns over the WNA (e.g, Gershunov and Barnett, 1998; Hidalgo and Dracup, 2003; Thompson and Wallace, 2001; Guzman-Morales et al., 2016; Gershunov et al., 2017; Guirguis et al., 2019) and have implications for the predictability of its hydroclimate (Gershunov and Cayan, 2003). However, the influence of these large-scale features in the weather of Western U.S. is not restricted to winter. Alfaro et al. (2006) showed that the PDO is responsible in part for the coupled variability between the climate in May and summer air temperatures over Western U.S., as well as for its predictive capability. Our results for the models built with SST as predictor show that the leading CCA modes reflect the influence of the Tropical and Extra Tropical Pacific ocean-atmospheric climate throughout most of the year and over a range of time scales. Moreover, we show the greatest SST and air temperature related variability is present in the winter, given that this is when the CCA modes present the highest amount variance explained. Those CCA patterns have a physical meaning and are not simply statistical artifacts and this indicates that the forecast models developed here using SST as predictor rely on the effect of the large-scale climate variations over the regional variability.

In contrast to the SST canonical modes, those derived from soil moisture do not correlate strongly with any of the climate modes (Figures 3.4c and 3.4d). As consequence, when weighting by the variance explained by each mode, the values drop significantly (Figures 3.4i and 3.4j). However, SM canonical modes are more strongly correlated locally with the surface air temperature variability, particularly with Tmax for fall and winter, as it is shown in Figure A.1. This figure presents the pairs of SM and Tmin and SM and Tmax patterns that constitute the first two SM-Tmin and SM-Tmax canonical modes. SM and Tmax patterns for fall and winter are co-located, suggesting that the SM influence on subsequent surface temperature occurs locally. This influence is quantified in Figure A.2, showing the SM-Tmin and SM-Tmax patterns weighted by the variance explained by each of the SM canonical modes. We observe that the only cases when there is still significant values occur for Tmax during fall and winter (Figure A.2(b)). When both SST and SM are used to build the CCA model, the correlations between the canonical modes and the climate indices are similar to those obtained using only SST. However, the variance explained is significantly reduced. The inclusion of SM as a second predictor reduces the variance explained by each of those modes (Figures 3.6a to 3.6f). Implications for predictability and predictive skill of the CCA models are discussed in the next section.

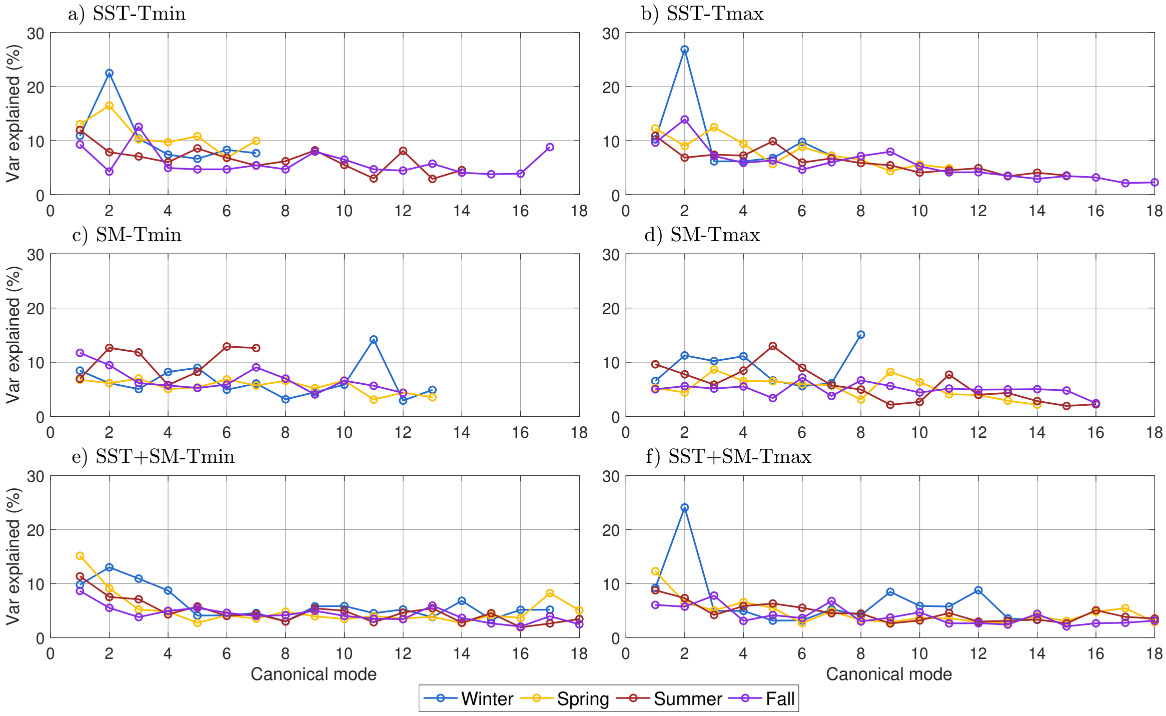


Figure 3.6: Variance explained by each canonical mode from the CCA models, averaged by seasons. The colors show the values referent for each season: winter includes those experiments initialized from December to February; spring, from March to May; Summer, from June to August; and fall, from September to November.

3.5 Predictive relationships and skill

In this section we present an overview of the annual cycle of the cross-validated seasonal forecast skill, calculated as the correlation between observed and predicted air temperature anomalies, obtained with the CCA models for the different combinations of predictors (SST, SM and SST+SM) and predictands (Tmin and Tmax).

Figure 3.7 shows annual cycle of the field-averaged skill of all CCA models and the field-averaged skill for the randomized experiments. Skill values that are outside of the shaded area in Figure 3.7 are considered significant with 95% confidence, representing a result of deterministic processes between the predictors and the predictands.

Remote SSTs yield more seasonal skill at one month lead time than does local soil moisture, for both Tmax and Tmin (Figures 3.7a and 3.7b). Additionally, there is a noteworthy difference in the annual cycle of the field-averaged skill between the two predictors: while for SM significant skill occurs only during late fall and winter, SST yields significant skill throughout the year for both set of predictands, with the the highest skill values occurring during late winter/early spring (FMA and MAM) and summer (JJA and JAS).

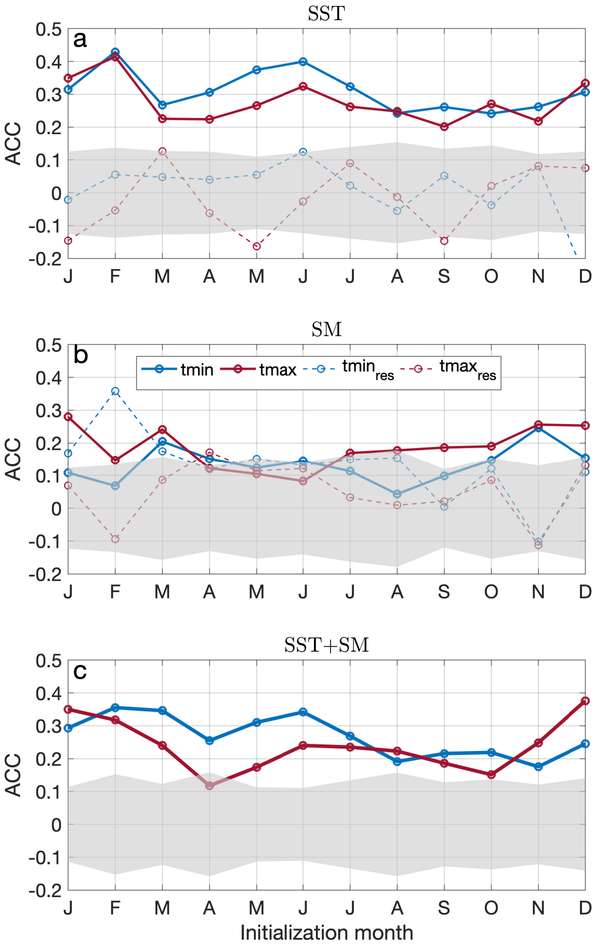


Figure 3.7: Annual cycle of field-averaged skill (average skill among all the grid points) obtained from the CCA model run at one month lead time using (a) SST as predictor, (b) SM as predictor, and (c) SST and SM combined. Solid blue line is the skill for Tmin and solid red line is the skill for Tmax. Shaded areas are the confidence interval obtained from the randomized experiments: values that fall outside this shaded area have statistical significance greater than 95%. Dashed lines are the annual cycle for the field-averaged skill of the residuals, where in (a) are the residuals from SST predicted with SM and in (b) are the residuals from SM predicted with SST (see Section 3.23.2.4 for details).

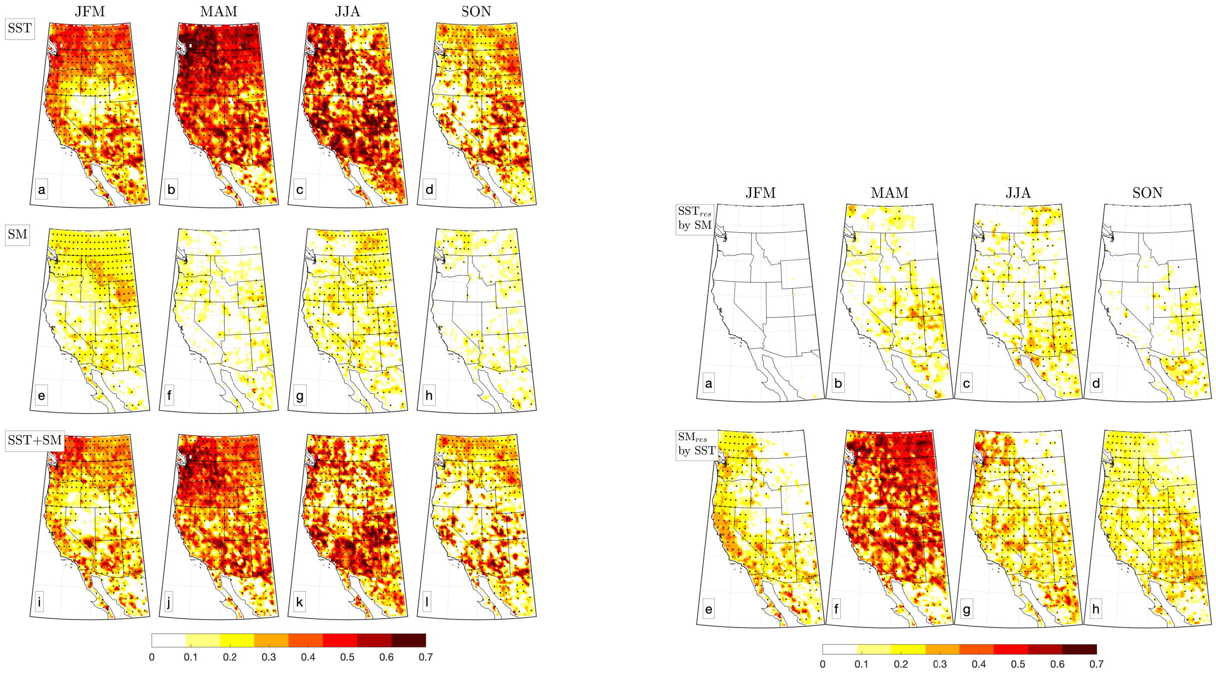
In comparison to predictions from SST alone, the combination of SM and SST predictors add only little to seasonal forecast skill (Figure 3.7c). However, some interesting features are found. In location and seasons where the skill for the SM-CCA model is the lowest, the inclusion of this predictor with SST tends to deteriorate the predictive skill. For example, during spring and summer (MAM to JAS), the skill to predict Tmax and Tmin with SM is at its lowest; those seasons correspond to a decrease in the skill for the SST+SM model compared to the SST-alone model (Figure 3.7c). It is surprising that the inclusion of an additional predictor deteriorates the skill. However, when there is not much influence, complicating the predictor field provides more opportunity for CCA to find optimally correlated modes that do not explain much variance, e.g., spurious or non-physical modes. This low amount of variance explained by the SST+SM model combined with the low skill achieved by the SM-alone model during spring and summer impacts the ability to predict land surface temperature for those seasons. On the other hand, during the winter (DJF, JFM and FMA), there is a slight improvement in the Tmax skill for the SST+SM model. This season corresponds to the best skill for the SM model and it is also when the leading CCA modes for SST+SM explain the greatest amount of variance (Figures 3.6e and 3.6f). This is not seen in the Tmin results, suggesting that the wintertime influence of soil moisture in Tmin variability is very low compared to that from Tmax.

The improvement in the SST+SM skill is very modest and it is much smaller than the sum of the skill for SST and SM exclusive models. This raises the question whether the skill achieved individually by each of those predictors is actually independent. To investigate that, we run the experiments with the residuals, when SM is used to predict the residuals of the SST CCA model and vice versa (see Section 3.23.2.4). The results for the skill for the prediction of the residuals are presented in Figures 3.7a and 3.7b, as the dashed lines. We observe that the skill for predicting the SST residuals with SM (for both Tmax and Tmin) is not statistically significant, explaining less than 1% of the variance ($\rho^2 < 0.01$) and its annual cycle is very noisy (dashed lines in Figure 3.7a). On the other hand, the soil moisture residuals predicted with SST present some noteworthy patterns. Although its skill is, in general, below the significance levels, the annual cycle is less noisy than that from the SST residuals predicted with SM. Moreover, the SM residuals for Tmin during the late winter and early spring can be skillfully predicted by SST (dashed blue line in Figure 3.7c). Those results show that SM cannot predict the residuals from the SST model, but SST can sometimes predict the residuals from the SM models, particularly when the original skill is at its lowest. This indicates that the skill obtained from SM is not completely independent of SST, contributing to the fact that combining SST and SM into a single predictor does not considerably improve the forecast skill.

The field-averaged skill analysis shows patterns in the annual variability of the predictive skill for the WNA. However, there are some regional variations in the skill pattern throughout the domain. These variations are explored in the maps of the cross-validated skill obtained with the three combinations of predictor fields for selected seasons (JFM, MAM, JJA and SON), and for Tmin (Figure 3.8(a)) and Tmax Figure 3.9(a)).

In general, these maps reflect what is seen in field-averaged skill analysis: SST is the best predictor for both Tmin and Tmax for all seasons, the inclusion of SM degrades the skill for Tmin throughout the year and for Tmax during Spring and summer, but it improves the predictive skill for Tmax during the winter. Specifically, SM yields the best skill in predicting Tmax in the central-east and south region during JFM (Figure 3.9(a)). The spatial pattern for the residuals of Tmax emphasize this result: for JFM, SM can significantly predict the Tmax residuals obtained with the SST model, particularly for the central-east portion of the domain (Figure 3.9(b), frame a). This region correspond to that where SST fails in predicting Tmax, leading to an improvement in the skill when SST and SM are combined (Figure 3.9(a), frame i).

For the remainder of the year, SM does not explain a lot of neither Tmin nor Tmax variability. Although there are statistically significant values, the skill is smaller than that of JFM ($\rho < 0.25$) and, differently from JFM, the inclusion of SM does not improve the skill. Moreover, the spatial pattern of the skill of SST residuals predicted with SM is not coherent and the values are small and in general not significant (Figures 3.9(b) and 3.8(b), frames a-d). On the other hand, for Tmin, the SM residuals can be skillfully predicted with SST, particularly for the spring (MAM) (Figure 3.8(b), frame f). In fact, MAM is the season with the highest predictive skill and this skill seems to be almost entirely derived from relationships with SST patterns. Among those patterns, the leading canonical mode corresponds to a warming trend (Figure A.4) and it accounts for almost 30% of the variance, the highest among all the 12 seasons (Figure A.3). Indeed, MAM is the season with the largest warming trend observed over WNA. Therefore, the warming trend accounts for an important part of the predictive skill for the spring, particularly for Tmin.

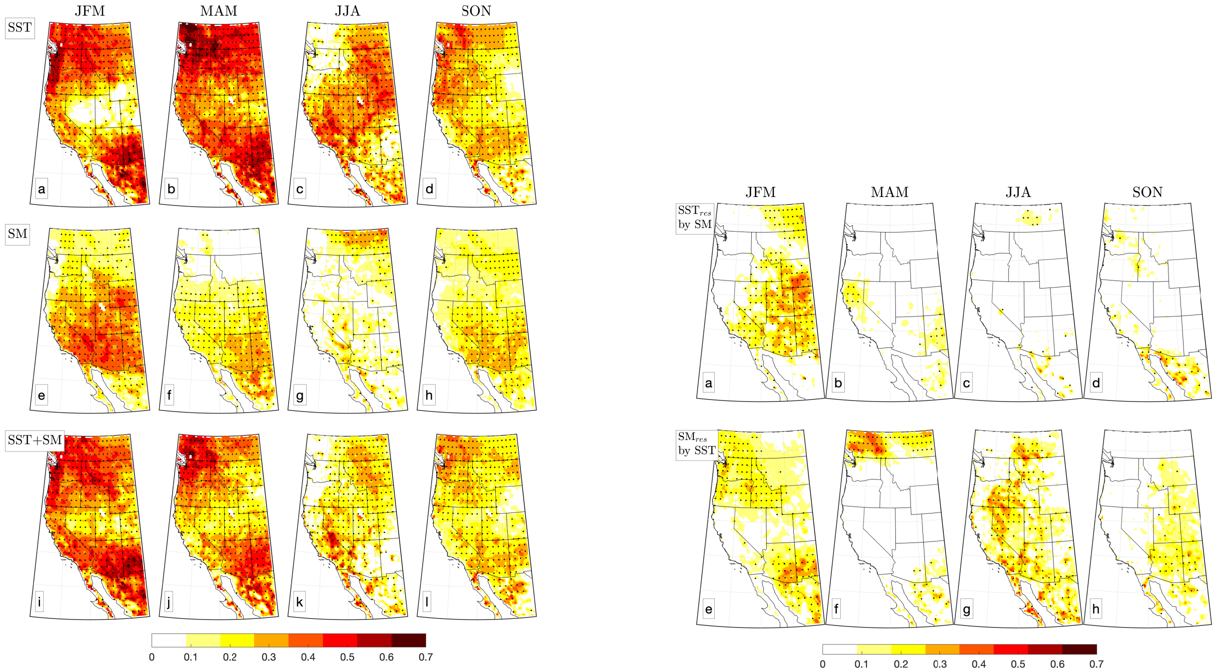


(a) Skill

(b) Residuals

Figure 3.8: Spatial pattern of Tmin forecast skill for selected seasons (shown above the maps in the upper panels). **(a)** Tmin seasonal forecast skill obtained with the model using SST (top panels), SM (middle panels) and SST+SM (lower panels) as predictors; **(b)** SST and SM Tmin residuals forecast skill. SST-Tmin residuals forecast using SM as predictor (upper panels) and SM-Tmin residuals forecast using SST as predictor (lower panels). Skill is expressed as correlation between the cross-validated CCA forecast and observation at each grid point in the WNA. Black dots represent values that have statistical significance greater than 95%.

Compared to models using SST predictors, those using SST + SM for spring, summer and fall show that there is a decrease in the skill in predicting Tmin and Tmax, especially in regions where SM skill is the lowest. This result contrasts with previous results from Alfaro et al. (2006), which also used Pacific SST and an index of soil moisture as predictors of seasonal air temperatures. The authors found that summertime air temperatures are more predictable than those in the winter. The main reason for this discrepancy is that, during the summer, both SST, in the form of the large-scale patterns, and local soil moisture provide a source for temperature variability, especially for Tmax. For the winter, Alfaro et al. (2006) found that the effects of soil moisture are less important and SST plays a more important role in the temperature variability, especially for Tmin. Here we found that SM does not improve the predictions for Tmin and only has an impact in the Tmax predictions during the winter. This disagreement with the results from Alfaro et al. (2006) is likely a result of the larger Pacific coast to the Mississippi basin domain of the their study, whose Great Plains region has been shown to harbor relatively strong soil moisture feedback to the overlying atmosphere during the summer period, while the West Coast United States does not (Huang et al., 1996; Zhang et al., 2008; Gershunov and Douville, 2008; Wing et al., 2016).



(a) Skill

(b) Residuals

Figure 3.9: Same as Figure 3.8, but for Tmax as predictand

The significant skill of winter Tmax forecasts from SM, to our best knowledge, has not been reported before. To understand the process that yielded those positive skills, we explore the annual variability of the SM autocorrelation at different lags over different regional domains compared to the annual cycle of the regional-averaged skill of Tmin and Tmax. Those regional domains are shown as the red boxes in Figure 3.1 and they were chosen based on regions where SM presented strong predictive skill for Tmax during wintertime: Central Rockies (region A), Coastal California (region B) and Inland Southwest (region C). The annual cycle of the regional-averaged skill for each of those regions is presented in Figures 3.10a-c for Tmin and in Figures 3.10d-f for Tmax. For Tmin the SST-alone model has a better predictive skill than SM and the inclusion of SM in general does not improve the skill. Moreover, the SM skill does not vary much throughout the year and the SST+SM skill variability is closely related to that from SST-alone. This is consistent with findings shown above that SM does not influence Tmin variability and therefore the predictability in any season. On the other hand, SM did exhibit association with Tmax variability. During wintertime, SM yields a significant predictive skill of Tmax, very close to the skill achieved by SST-alone. As a consequence, the combination of two predictors accounts for more skill than the individual predictors, particularly for Coastal California and the Inland Southwest during JFM. This annual variability in the SM predictive skill for those three regions is very similar to the annual cycle of the SM autocorrelation in each of these regions (Figure 3.10g-i). These figures show the correlation between the monthly and the seasonally-averaged SM at different lags. For example, January soil moisture is correlated with FMA soil moisture and it corresponds to lag 1; January with MAM corresponds to lag 2 and so forth. The figure shows that this autocorrelation is large and more persistent during the winter months (November to January). This shows that for these regions over WNA, soil moisture has a longer memory during the winter than during the summer. This is reflected in the seasonal predictions, since those are the months that the Tmax predictions from SM yielded significant skill.

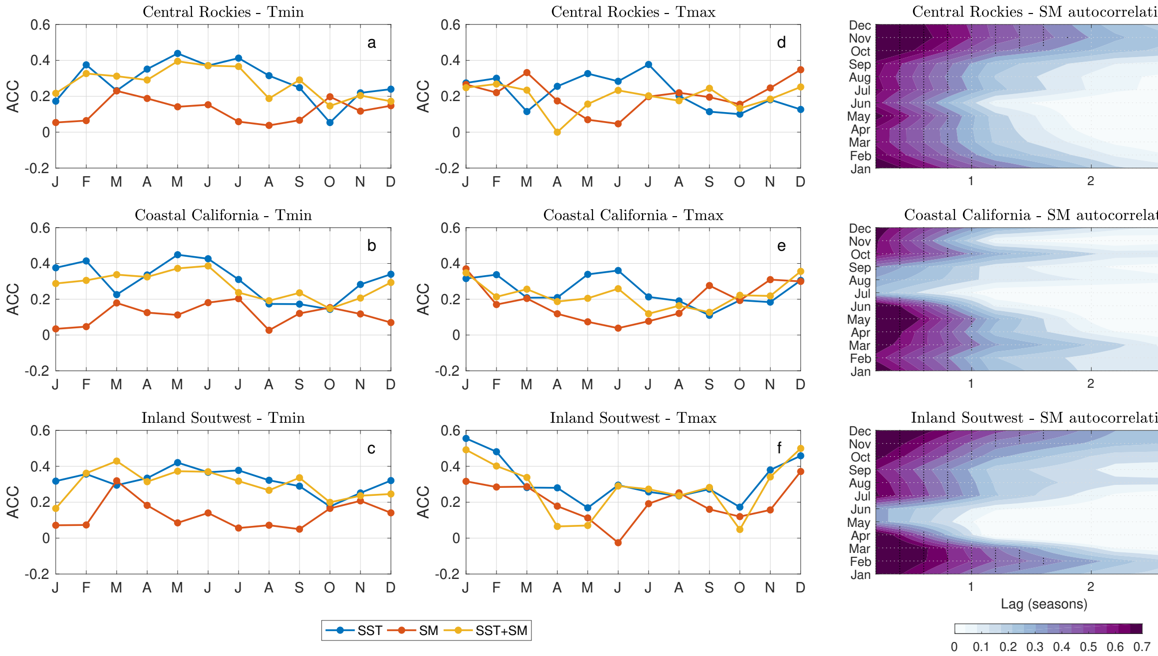


Figure 3.10: Annual cycle of seasonal predictive skill for (a-c) Tmin and (d-f) Tmax, and (g-i) annual cycle of soil moisture memory for the three regional domains shown in Figure 3.1.

3.6 Summary and Conclusions

We developed a statistical prediction model for seasonal Tmax and Tmin anomalies over WNA using remote and local predictors. The first predictor is sea surface temperature (SST) across the tropical and northern Pacific basin, representing the influence of large-scale climate variability patterns, which affect surface air temperature over land via association with atmospheric circulation patterns. The second predictor is soil moisture (SM), which is thought to represent the local effects on surface air temperature that affect the surface energy balance through partitioning of latent and sensible heating. We showed that both remote and local predictors influence the variability and, as consequence, the predictability of local Tmax and Tmin over WNA. Skills shown are cross-validated and are also significant when compared to a randomized experiment, indicating that the skill is real.

We explored the annual cycle of Tmax and Tmin predictive skill from SST, SM and from those two predictors combined. The general pattern in the annual cycle of the predictive skill obtained with SST is similar for both Tmin and Tmax in which the highest skill occurs during the late winter and in spring, in particular for the MAM season, that is shown to be the season with the highest predictive skill, in part because of the enhanced warming trend during that season. The spatial variation in the skill, shown as anomaly correlation coefficient maps, is also very similar for Tmin and Tmax, in which the highest skill occurs along the coast and in the northernmost and southernmost parts of the domain. For both predictands, the Central region, corresponding to Nevada, Utah and Colorado, presents very low predictive skill during the wintertime. During the summer season, SST yields higher skill for Tmin than Tmax.

The leading SST-WNA temperature canonical modes reflect distinct forms of remote climate variability operating via teleconnections, explored as correlations between the canonical modes and several known climate indices in the Pacific ocean and atmosphere. Such patterns are strongly correlated with the canonical modes throughout the year, but there is more variance explained by each individual mode during the wintertime, which accounts for more predictability for both Tmin and Tmax during winter. However, ENSO indices are highly correlated with the Tmin canonical patterns during late spring and early summer, leading to a higher predictive skill for Tmin compared to Tmax during summertime. Therefore, those large-scale climate conditions influence both Tmax and Tmin during the winter, but their influence in the summer mainly affects Tmin.

When considering soil moisture (SM), there are some noteworthy differences between the predictive ability for Tmin and Tmax. Soil moisture seems to not influence Tmin in any season, but it improved the predictability skill for Tmax during the wintertime, particularly for JFM. This result agreed with previous findings showing that soil moisture exerts higher influence on temperature variability mainly through Tmax. However, this result is not consistent with Huang et al. (1996) and Alfaro et al. (2006) results showing that the strongest influence and greatest predictability of Tmax from SM occurs in the summer. Those studies, however, were made over a larger domain that included the Great Plains region in the United States interior, where soil moisture has been shown to have stronger influence on surface air temperature. Additionally, we found that, for the central and southern parts of WNA, soil moisture displays a strong autocorrelation up to one season ahead during the winter months, therefore having more potential to influence the overlying atmosphere one season ahead, which likely explains the larger influence exerted on the predictive skill by soil moisture in late winter.

These results provide more insight into seasonal air temperature variability over WNA and advances the development of a linear statistical technique for seasonal forecasting, which could be easily operationalized. Evaluation of statistical methods relative to state-of-art operational dynamical approaches was not the goal of this paper, but the statistical model skills are comparable and in some cases superior to those from dynamical models. Exploring these comparative benefits will be the goal of a future study. Finally, prediction of Tmax and Tmin separately can provide useful information for stakeholders, given the distinct consequences that Tmin and Tmax have for different sectors of the society.

Acknowledgments

We thank Ben Livneh for producing and supplying the gridded historical minimum and maximum temperature dataset, Dennis Lettenmaier for the VIC historical data of soil moisture, and David Pierce for supplying those datasets. We also thank the Hadley Center Sea Ice and Sea Surface Temperature for the SST reanalysis. Finally, we would like to thank also Amato Evan and David Pierce for useful discussions. This study forms a part of the Ph.D. dissertation of DFG, who was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the Grant

221222/2014-6 and by the California Energy Commission under Agreement Number PIR-15-005. Additional funding from the NSF (OCE1419306) and the National Oceanic and Atmospheric Administration (NOAA; NA17OAR4310106) is greatly appreciated.

Chapter 3, in full, is in revision for publication in *Journal of Climate*. Dias, D.F., A. Gershunov, D. Cayan and A.J. Miller (2020), The influence of sea surface temperature and soil moisture in seasonal predictions of air temperature over Western North America. The dissertation author was the primary investigator and author of this paper.

Chapter 4

A statistical benchmark for NMME: the case of seasonal predictions of minimum and maximum temperature over the Western North America

Abstract

We develop a linear empirical model, based on canonical correlation analysis (CCA), to predict seasonal daytime (Tmax) and nighttime (Tmin) temperatures over the Western North America. Sea Surface Temperature anomalies across North and Tropical Pacific are used as predictor, representing the large-scale climate variability that affects the variability of air temperature over land. The predictive skill of CCA is compared to that obtained from persistence and from the North American Multimodel Ensemble (NMME), which is an operational prediction system that uses state-of-art global climate models. We find that CCA skill outperforms persistence for almost every season predicted. Additionally, it is comparable and sometimes better than that from any NMME single-model as well as from NMME multimodel ensemble average. We also analyze whether CCA and NMME reproduce the observed trends in the diurnal temperature range and the asymmetry in Tmax and Tmin changes. While CCA captures the observed trends, as well as the change in the trends over time, NMME overestimates the increasing trend in Tmax. Considering that persistence is the widely-used measure when evaluating the predictive skill for dynamical models, we suggest the use of CCA as a much higher benchmark. Finally, these differences in the skill from a simple linear model and form a nonlinear dynamical model indicate that there is room for improvement in the forecasts.

Significance statement

The variability of minimum and maximum temperatures has distinct and specific impacts on agriculture, human health and energy usage. Therefore, different strategies are necessary to manage those impacts and predictions of these extremes, rather than the mean temperature value, are of great relevance. We evaluate the predictive skill of seasonal minimum and maximum temperature over the Western North America using two simple statistical models, persistence and canonical correlation analysis (CCA), and one state-of-art prediction system with complex dynamical models. We show that the skill from the CCA outperforms persistence and is comparable and sometimes better than from the dynamical models. We show that by using CCA a much higher benchmark for the predictions with the dynamical models.

4.1 Introduction

The diurnal cycle of land surface air temperature, defined by the maximum daytime (Tmax) and minimum nighttime (Tmin) temperatures, has been changing during the recent decades. Globally, Tmin has been warming faster than Tmax, causing a reduction in the diurnal temperature range (Karl et al., 1993; Alexander et al., 2006; Davy et al., 2017). In the Western North America (WNA), the accelerated warming of Tmin is changing the characteristics of the heat waves. They are becoming longer, more extensive and more humid, and their expression has been robustly seen in nighttime temperatures (Gershunov et al., 2009). In a region characterized by typically dry summer days and cool nights, more humid and warmer nights exacerbate the impacts on agriculture (Gershunov et al., 2010), human health (Ostro et al., 2009; Guirguis et al., 2014) and electrical energy usage. Therefore, the seasonal to interannual variability in Tmax and Tmin can have distinct and specific impacts, which may require different strategies to manage. Additionally, the mechanisms that control the variability and, consequently, the predictability during day and night may be very different. Because of that, seasonal climate predictions of these daily extremes separately, rather than mean temperature, are of great relevance.

Previous work exploring seasonal predictions of surface air temperature, with both dynamical and statistical models, have mainly focused on variations in the mean temperature rather than Tmin and Tmax (Barnston, 1994; Huang et al., 1996; Mo, 2003; Doblas-Reyes et al., 2013a; Becker et al., 2014; Zhang et al., 2019). A widely-used global prediction system is The North-American Multimodel Ensemble (NMME, Kirtman et al., 2014). It uses state-of-art global climate models (GCMs) to explore the improvement in forecast skill with the use multiple models with many ensembles. In fact, Becker et al. (2014) showed that the NMME multimodel ensemble mean predictive skill is equal or higher than that from any single model for different oceanic and atmospheric variables, including surface air temperature.

Statistical models are also widely-used to perform climate seasonal forecasts. Being relatively easy and much more computationally economical to perform, they can be used either as a predictive approach per se or as benchmarks when assessing the skill of the complex dynamical models (e.g., Alfaro et al., 2006; Ho et al., 2013; Newman and Sardeshmukh, 2017; Kapnick et al., 2018). Additionally, the statistical methods can complement the dynamical ones because advances in statistical predictions lead to an enhanced understanding of the system that is being predicted, which in turn leads to an improvement in the dynamical models (Doblas-Reyes et al., 2013a).There is a variety of statistical predictive models that can be used, such as simple persistence in the initial conditions, linear models, nonlinear methods and machine learning techniques (Penland and Sardeshmukh, 1995; Wilks, 2011; Ham et al., 2019).

Persistence has been the most common measure used to evaluate seasonal to decadal climate predictions performed with GCMs (Kirtman et al., 2014; Yeager et al., 2018). However, more sophisticated statistical models can have skill significantly better than persistence and comparable to that from the dynamical models, particularly for mean air temperature and sea surface temperature anomalies (Newman, 2012; Newman and Sardeshmukh, 2017; Huddart et al., 2017; Dias et al., 2018). This raises the question of whether a higher benchmark could be set for seasonal predictions of Tmin and Tmax using dynamical models, particularly the NMME.

We propose a new benchmark to be used for seasonal predictions of Tmin and Tmax over the WNA, based on canonical correlation analysis (CCA). This method finds linear combination between two sets of variables (for example, predictor and lagged-predictand) based on their covariability, seeking to match patterns in the predictor field with patterns in the predictand field (see *Materials and Methods*). For the WNA, using Pacific Sea Surface Temperature anomalies (SST) as predictors, CCA has provide skillful seasonal predictions of precipitation (Gershunov and Cayan, 2003), mean suraface air temperature (Barnett and Preisendorfer, 1987; Barnston, 1994), and Tmin and Tmax (Alfaro et al., 2006; Dias et al., 2020). The weather and climate over the WNA are directly affected by the teleconnections with the Tropical and Extratropical Pacific Ocean, in the form of El Nino Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) related variability (e.g, Gershunov and Barnett, 1998; Hidalgo and Dracup, 2003; Thompson and Wallace, 2001; Guzman-Morales et al., 2016; Gershunov et al., 2017; Guirguis et al., 2019). Therefore, Pacific SST anomalies are broadly representative of large-scale climate measures that vary over time scales that are relevant to seasonal air temperature fluctuations (Dias et al., 2020).

4.2 Nighttime and daytime temperature predictions over Western North America

We use three prediction approaches to evaluate the predictability of Tmin and Tmax over WNA. The first and simpler approach is the persistence of Tmin and Tmax anomalies. Persistence here is defined as the correlation of the initial month with the following season (for example, Tmin anomalies in January correlated with Tmin anomalies in FMA). The second approach is a linear statistical prediction model based on Canonical Correlation Analysis (CCA), that uses Pacific Sea Surface Temperature anomalies (SST) as predictor (see *Materials and Methods* for details). Finally, we evaluate Tmin and Tmax seasonal prediction skill of the NMME models, as well as the NMME multimodel ensemble average, based on bias-corrected hindcasts.

Figure 4.1 shows annual cycle of the field-averaged skill for all the approaches: persistence, CCA, individual NMME models, and NMME multimodel ensemble average (NMME_{avg}). This figure also shows the results of the randomized CCA experiments as the grey shaded area: skill values that are outside of this area are considered significant with 95% confidence, representing a result of deterministic processes between the predictors and the predictands. For this exercise, we use anomaly correlation coefficient (ACC) as a measure of the skill, since it evaluates the ability of the model to reproduce the phasing of the interannual variability of the predictand.

Considering that persistence is the measure usually used as a benchmark for dynamical forecast systems, given its simplicity, the first important question to address is whether CCA predictive skill is better than persistence. Figure 4.1 shows that CCA predictions outperforms persistence in almost every season and for both Tmin and Tmax. In the seasons that CCA does not do a better job than persistence, the values are comparable (for example, predictions initialized in March and April, correspondent to the seasons AMJ and MJJ, for both Tmax and Tmin). Bearing that in mind, and considering that CCA is a simple linear model, we can directly compare its skill with that from NMME.

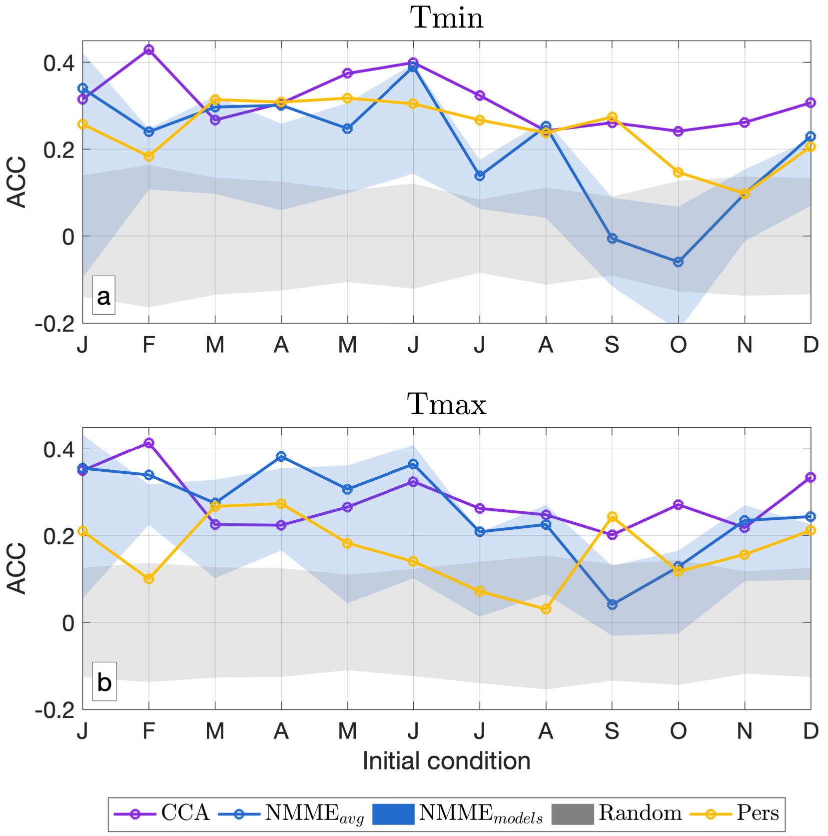


Figure 4.1: Annual cycle of (a) Tmin and (b) Tmax field-averaged anomaly correlation coefficient (ACC) for predictions run at one month lead time obtained from different models: CCA (purple line), persistence (yellow line), NMME multimodel ensemble mean (blue line), and individual NMME models (shaded blue area); x-axis shows the initial condition, for example, a model initialized in January (J) gives the skill for Tmin and Tmax predictions for the following season (FMA). Shaded grey areas are the confidence interval obtained from the randomized experiments: values that fall outside this shaded area have statistical significance greater than 95% (see Methods for details).

NMME skill presents high inter-model variability for different seasons, with differences between the models with the lowest and highest ACC values for one specific season ranging from 0.15 to 0.45. NMME multimodel ensemble average is significant almost throughout the entire year, except during late-fall and early-winter (OND and NDJ), for both Tmin and Tmax. NMME$_{avg}$ skill usually is very close to that from the highest single-model skill, except during those seasons where its skill values are not significant. This is not surprising since the multimodel ensemble average tends to have higher skill than any single-model (Becker et al., 2014). Considering the ability of NMME to predict Tmin or Tmax, there is not striking differences. The annual-cycle is very similar, with highest skill values during late-winter (FMA) and late-summer (ASO).

Regarding the CCA hindcasts skill, we observe from Figure 4.1 that there is significant skill for every season, for both Tmin and Tmax, with Spring (MAM) being the season with the highest skill values. Overall, the annual-cycle of Tmin and Tmax CCA skill is very similar, with the greatest differences from late-spring to late-summer, when CCA predicts Tmin better than Tmax. Because of that, CCA does a better job than NMME in predicting one-season ahead Tmin variability throughout the year. The skill is higher than or comparable to any NMME single-model as well as the NMME$_{avg}$, for all seasons (Figure 4.1a). This is also true for Tmax, with the exception for the seasons when CCA poorly predicts Tmax compared to Tmin (from late-spring to late-summer).

The field-averaged skill analysis shows patterns in the annual variability of the predictive skill for the western north America. However, it is important to consider the regional differences in the NMME skill and how they compare to that from CCA. We first explore those regional variations by calculating the difference between ACC obtained with CCA and with NMME$_{avg}$ ($\Delta ACC = ACC_{CCA} - ACC_{NMME}$). The results are shown in Figure 4.2 as maps of the ΔACC for selected seasons (DJF, MAM, JJA and SON). Since CCA has a better skill than persistence and NMME multimodel ensemble average has a comparable skill to that from any single-model, we show the maps using CCA as a benchmark for the NMME$_{avg}$.

For Tmin, CCA has higher ACC than NMME for the majority of the domain and for all the seasons, except during fall (SON), shown as red regions in Figure 4.2a-d. The improvement is more outstanding during seasons that the Tmin forecasts from CCA have the strongest ACC: spring (MAM) and summer (JJA), particularly for the California coast. This is a promising result, as it has been shown that NMME have a persistent deficit of seasonal temperature forecast skill over the Western coastal regions, particularly for warm seasons (Zhang et al., 2019). For Tmax, the differences between CCA and NMME skill aren't as remarkable as that from Tmin. Fall (SON) and winter (DJF) have very similar skill, whereas during spring (MAM) and summer (JJA) CCA outperforms NMME in the northeast part of the domain. However, for the southeast NMME does a better job in predicting Tmax, especially during summer. Interestingly, the skill improvement along the California coast for Tmin during summer is not seen for Tmax (Figures 4.2e-h).

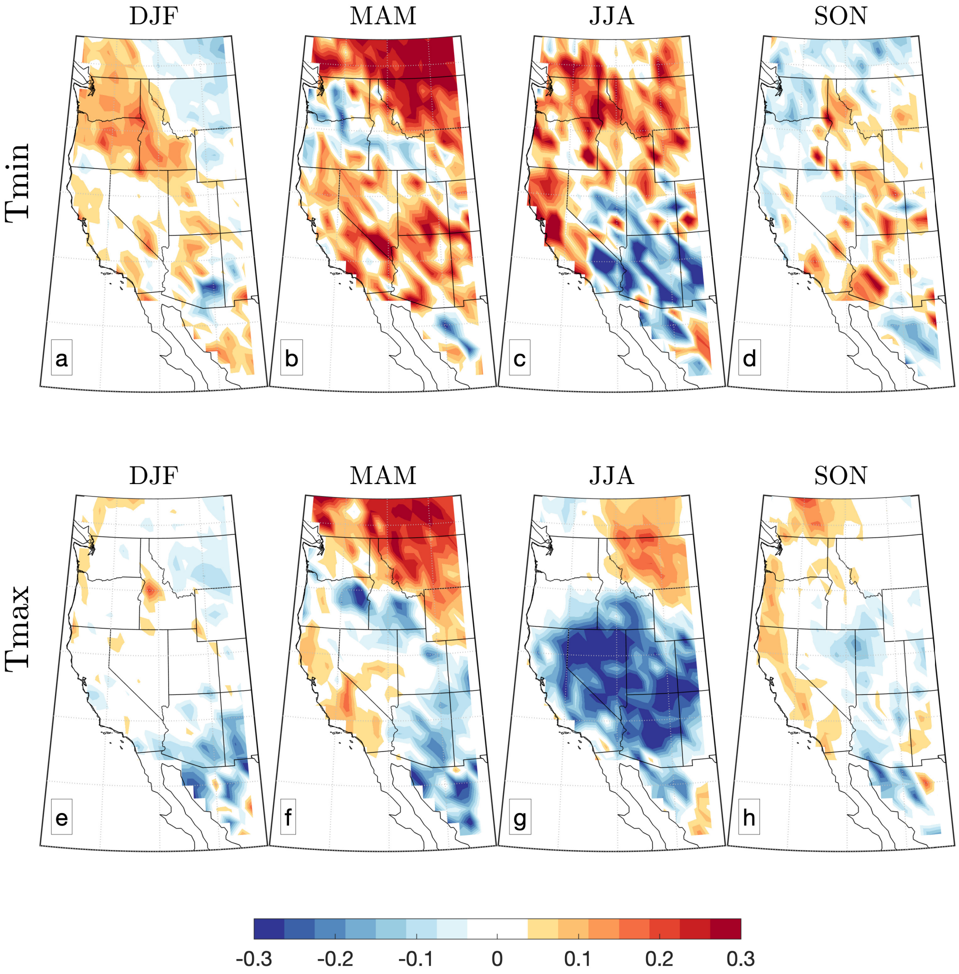


Figure 4.2: Difference between (a–d) Tmin ACC and (e–h) Tmax ACC obtained from CCA and from NMME$_{avg}$ for selected seasons. Positive (negative) values represent regions where CCA skill is better (worse) than NMME skill. Colors indicate regions where the differences are significant.

ACC is a great measure of skill to understand how the predicted and observed interannual variations in the anomalous field match. However, it does not capture errors in the magnitude of the predicted field. Therefore, it is important to complement the skill analysis with a measure that captures those errors. Here we calculate the Root-Mean Squared Error (RMSE) for CCA and NMME$_{avg}$ predictions, defined as:

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(pred_i - obs_i)^2}, \tag{4.1}$$

where *pred* is the predicted field (NMME$_{avg}$ and CCA), *obs* is the observed field, and $n = 29$ is the number of samples, i.e., number of years (1982 to 2010).

In order to compare NMME errors with that from CCA, we use a relative RMSE (Huddart et al., 2017), defined as:

$$RMSE_{rel} = \frac{RMSE_{NMME}}{RMSE_{CCA}}. \tag{4.2}$$

A relative RMSE bigger (smaller) than one means that CCA is performing better (worse) than NMME, whereas a relative RMSE of one means that both models have the same performance. Figure 4.3 shows the relative RMSE for the same selected seasons as Figure 4.2, for both Tmin and Tmax. Overall, the maps present relative RMSE bigger than one, for almost all seasons, locations and predicted variable, indicating that the magnitude of NMME predicted values are more spread with relation to the observations that those from CCA. For Tmin, CCA has smaller errors than NMME during the winter (DJF) and spring (MAM), particularly for the Northernmost part of the domain. Summer (JJA) and fall (SON) have less coherent pattern, with patches where CCA performs better and many areas where the differences between NMME and CCA are not significant, with RMSE$_{rel}$ closer to one (Figures 4.3c-d). On the other hand, the errors of the CCA predictions for Tmax are smaller than those from NMME$_{avg}$ for all the four seasons, especially in the

Northernmost part of the domain (Figures 4.3e-h).

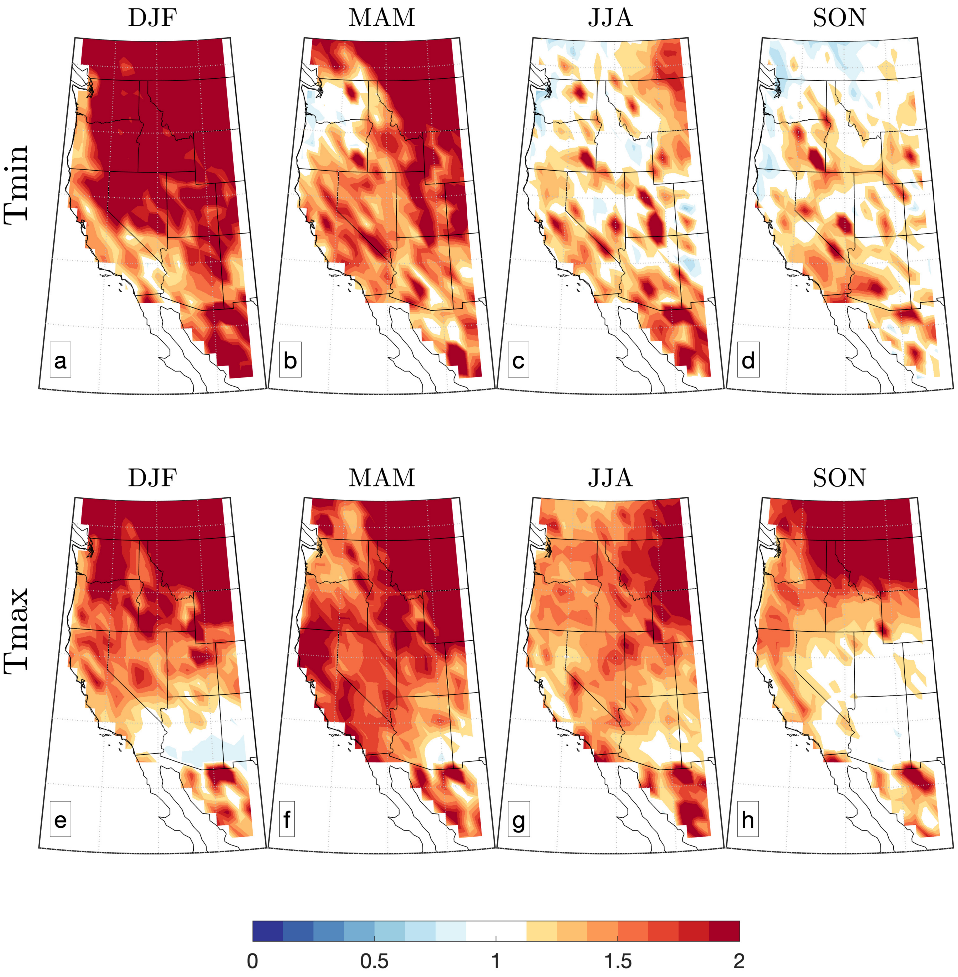


Figure 4.3: Relative NMME$_{avg}$ RMSE to that from CCA for selected seasons. Values greater (smaller) than 1 represent regions where NMME$_{avg}$ RMSE is greater (smaller) than CCA RMSE. Colors indicate regions where RMSE$_{rel}$ is significant.

Finally, we compile different skill metrics for all the seasons and for the CCA, NMME$_{avg}$, and individual NMME model hindcasts into a Taylor diagram (Taylor, 2001), shown in Figure 4.4. This diagram comprises a lot of information: we show the field-averaged skill for all those experiments, where the color represents the predictive model and the symbols represent one specific season. The angular coordinate represents the average correlation coefficient between the model hindcasts and the observed values (the correlation for seasons and models that originally presented negative correlation was set to zero to facilitate the visualization). The red dashed line represents the normalized standard deviation for the observations, and the distance from this line of each plotted point represents amount of the standard deviation that is estimate by each predictive model for each season. For example, a point located along the normalized standard deviation with value of 0.4 represents a model that predicts 40% of the observed standard deviation. Therefore, the plotted points that are (1) closer to the x-axis and (2) closer to the reference line depict the best predictive models.

It can be seen in Figure 4.4 that NMME has a high inter-model and inter-season variability with regards to the correlation values, as well as with ability of each model to reproduce the observed standard deviation, for both Tmin and Tmax. There isn't a clear pattern of predicted seasons from different models aggregating in the Taylor diagram space, particularly when considering the predicted standard deviation values. A noteworthy result from Figure 4.4 is the difference between the skill from NMME_{avg} and from each individual NMME models. Although the NMME_{avg} correlation coefficient is overall similar to that from the highest single-model correlation coefficient (as was shown in Figure 4.1), its predicted standard deviation values are underestimated, being much lower than the reference value (60% or less) and also lower than many of the individual NMME models.

There is a clear distinction between CCA and NMME skill. Although for some seasons the NMME models and the NMME_{avg} correlation coefficients are similar and sometimes better than those from CCA, CCA does a better job than the majority of the NMME models in reproducing the standard deviation from the observations. For both Tmin and Tmax, the predicted fields from CCA represents at least 60% of the observed standard deviation, with the vast majority of the seasons reproducing more than 80%. Therefore, it is possible to see from Figure 4.4 a clear aggregation closer to the reference line of the points representing the CCA.

Figure 4.4 allows us to verify which seasons are best predicted by each model. It is clear that the spring (MAM) is the season that has the highest skill obtained with the CCA model, followed by the summer (JAS), for both Tmin and Tmax: there is the highest correlation values and the predicted standard deviations are very close to those form the observations. This result agrees with previous studies that showed that spring is the season with the highest predictability for air temperature in the United States (Van den Dool et al., 2003; Alfaro et al., 2006), probably because this spring is the season with the highest warming trend and this trend accounts for an important part of the predictive skill for air temperatures, particularly Tmin (Dias et al., 2020). For NMME, the CanCM4 model is the one that has the best skill for late-spring (AMJ) and late-summer (JAS), with a combination of a relative high correlation value (bigger than 0.3) and a standard deviation predicted to be at least 80% of that observed. For the other cases, even if the correlation values are higher, the standard deviation values are underestimated.

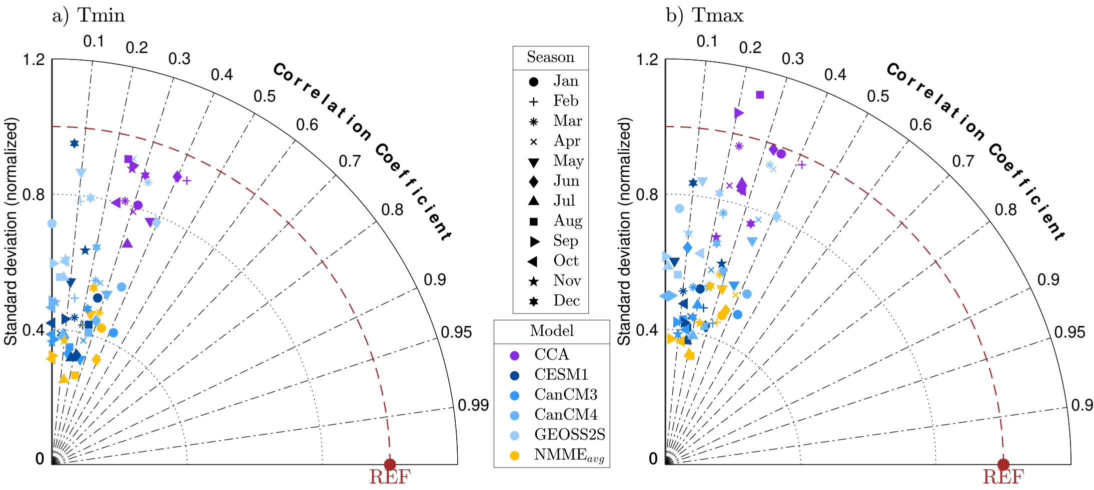


Figure 4.4: (a) Tmin and (b) Tmax Tyalor diagram. Each point in the graphic represents a combination of skill measurements for one specific model (depict in colors – CCA in purple, individual NMME models in blue tones and NMME$_{avg}$ in yellow) at one specific season (depict as different symbols. The month in the legend identifies the initialization month for the prediction made one-season ahead. For example, ''Jan" represents a prediction for FMA). The correlation coefficient and the normalized standard deviation are calculated as the average of all grid points. The red point and the red dashed line shows the reference value obtained from observations, with correlation and normalized standard deviation equal to 1. The models with the highest skill are those that are closer to the reference value in the x-axis, having higher correlation, and closer to the red dashed line, having a standard deviation similar to that from the observations.

4.3 Predictions of the diurnal temperature range trends

The CCA predictive skill for Tmin and Tmax over WNA comes from the ability of CCA to capture SST Pacific large-scale patterns at different time scales, such as PDO and ENSO (Dias et al., 2020). However, the increasing trend is also an important predictor that is captured by the canonical modes (Figures B.1 to B.6). Those figures show the time series and the correspondent spatial pattern of the leading canonical mode for both Tmin and Tmax as well as for SST. This leading mode correspond to an increasing trend in the temperatures over WNA throughout the year for Tmin (Figure B.3) and in the late winter to summer for Tmax (Figure B.6). Bearing that in mind, we investigate how well CCA and NMME capture the observed trend in the diurnal temperature range (DTR), defined as the difference between Tmax and Tmin.

Table 4.1 show the observed and predicted trends in DTR, Tmax and Tmin for two periods: from 1950 to 2010, which corresponds to the period of the observational data; and from 1982 to 2010, which corresponds to the period that the NMME hindcasts for Tmin and Tmax are performed. From Table 4.1, we observe that the trends in DTR have been changing. For the longer period (1950 to 2010), DTR has been decreasing because the Tmin has been warming faster than Tmax. However, if only the short period is considered (1982 to 2010), this trend changes, with Tmax warming faster than Tmin. CCA reproduces these changes, as well as the observed asymmetry in the Tmin and Tmax trends. NMME$_{avg}$ predicts a much stronger increasing trend in DTR than the observed values, because it overestimates the changes in Tmax. Since NMME hindcasts are not performed before 1982, it is not possible to evaluate whether NMME would have captured the changes in the DTR trends between those two periods. However, GCMs have not reproduced the observed trends from 1950 to 2010 in the global DTR because they have not captured the asymmetry in the changes in Tmin and Tmax. This deficiency may be related to the deficiencies in the models to reproduce regional feedback effects that affect Tmin and Tmax variability, such as cloud cover and land surface processes (Lewis and Karoly, 2013; Davy et al., 2017).

Table 4.1: Trends (^{o}C/decade) in diurnal temperature range (DTR), Tmax and Tmin over Western North America, calculated for the observations (Obs) and for the CCA and NMME$_{avg}$ one-month lag predictions. Trends were calculated for two different time periods, from 1950 to 2010 and from 1982 to 2010

	1950 – 2010			**1982 – 2010**		
	DTR	Tmax	Tmin	DTR	Tmax	Tmin
Obs	-0.08	0.08	0.16	0.05	0.19	0.13
CCA	-0.06	0.10	0.16	0.03	0.18	0.15
NMME	–	–	–	0.14	0.37	0.22

4.4 Conclusions

This study presents a new benchmark for the evaluation of the NMME seasonal predictions throughout the year of Tmin and Tmax over Western North America (WNA). The benchmark proposed is based on canonical correlation analysis (CCA) and it exploits relationships between Pacific SST variability and Tmin and Tmax patterns over WNA. We show that CCA, empirically built from 63 years of monthly-averaged SST and seasonally-averaged Tmin and Tmax, outperforms the predictive skill obtained from persistence, which has been a widely-used benchmark to evaluate predictions from GCMs. Therefore, considering that CCA is an easy and cheap model to run, the use of this model would set a much higher benchmark to evaluate the predictive ability for NMME.

In fact, the skill obtained with CCA is not only better than persistence, but it is also comparable and sometimes better than that from NMME. Skill here is evaluated with the anomaly correlation coefficient (ACC), which evaluates how the predicted and interannual variations in the anomalous fields match; and root-mean-squared error, which assess the spread of the predicted values relative to the observations. For both metrics, CCA has comparable performance to NMME. Additionally, we show in the Taylor diagram that the standard deviation predicted with CCA are similar to the observations, while the NMME models considerably underestimate these values.

Finally, we evaluate whether NMME and CCA can capture the observed trends in the diurnal temperature range (DTR), as well as the observed asymmetry between Tmin and Tmax changes. To that end, we calculate the trends for two periods, first for the longer observational record (from 1950 to 2010) and second for the NMME hindcast record (from 1982 to 2010). Consistently with the global trend, we find that DTR over WNA has been decreasing because Tmin has been warming faster than Tmax. However, for the shorter record, DTR shows a slightly increasing trend consistent with the faster warming in Tmax during the last 3 decades. CCA captures this change in the warming trend as well as the asymmetry between Tmin and Tmax changes. NMME also captures the increasing trend in DTR from 1982 to 2010, but the trend is overestimated, mostly because the warming in Tmax is augmented.

The fact that a simple statistical model has sometimes better skill than highly complex nonlinear models indicates that there is room for improvement of those dynamical forecasts. In the view of above, setting a higher benchmark for dynamical models can be very useful, particularly when developing forecast systems for societal applications.

Materials and Methods

The Canonical Correlation Analysis (CCA) explores and identifies the linear combination of two sets of variables that have the greatest correlation with each other. It condenses much of the spatial and temporal co-variability into a few modes. In a prediction sense, CCA can provide a simple and cheap way to predict one field of variables from another by matching patterns in the predictor field with patterns in the lagged-predictand field. Monthly-averaged Sea Surface Temperatures (SST) anomalies for the Pacific Basin (15°S to 60°N and 135°E to 110°W) were used as the predictor variable for seasonally-averaged minimum and maximum land surface temperature (Tmin and Tmax, respectively) over the Western North America (25°N to 53°N and 125°W to 105°W). While Tmax and Tmin are affected by several other variables such as winds, clouds, and topographic influences, those were not included here since the purpose is to explore linear prediction using a plausible, readily available, and manageable predictor set. Therefore, it is necessary to include a predictor that contain a measure of climate memory, so that it makes sense to use antecedent observations of this field. Pacific SST contain some measure of climate memory and it is broadly representative of large-scale climate measures that vary over time scales that are relevant to seasonal air temperature fluctuations.

The CCA model was built using training data from 1950 to 2013. The predictor and predictand fields were pre-filtered separately with the same number of p principal components (PCs), which are statistically orthogonal patterns of spatial and temporal variability, ordered by amount of variance explained. Those patterns were then related to each other using the q canonical correlates (CCs) extracted from the CCA analysis. The number of p PCs and q CCs were determined using a skill optimization scheme, in which the model complexity (i.e., the combination of PCs and CCs) used is the one that maximize the forecast skill (Gershunov and Cayan, 2003; Dias et al., 2020). Importantly, to avoid artificial skill from over-fitting, all results pertaining to CCA performance were cross-validated, wherein the year of a given prediction has been left out of the model development. To calculate the confidence intervals for our CCA forecasts, a randomized CCA experiments were performed: the time series of the predictor was randomized and used to build the CCA model to forecast air temperature. This procedure was then repeated 1,000 times and the forecast skill for each of those repetitions was calculated. Finally, using the skill obtained with this randomized experiments, for all these repetitions, the confidence intervals were calculated.

SST data was obtained from from the Hadley Center Sea Ice and Sea Surface Temperature (HadISST, Rayner et al., 2003), which has a resolution of 1° by 1° and it can be obtained at the following website: https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html. Tmin and Tmax data are from the Livneh CONUS near-surface gridded meteorological data, provided by the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (ESRL/NOAA). These datasets are gridded at a spatial resolution of 1/16° by 1/16° and are derived from daily temperature observations from approximately 20,000 NOAA Cooperative Observer (COOP) stations Livneh2015.

The CCA hindcast skill was compared to that to the NMME models that performed hindcasts of minimum and maximum temperature. This corresponds to four models and a total of 34 ensemble members (Table B.1). Before calculating the NMME hindcast skill, the bias from each ensemble of each model was removed as a function of calendar month and lead time, from 1982 to 2010, following (Barnston et al., 2015). A 3-month average was then calculated as a function of lead time for each individual bias-corrected ensemble member to be comparable to the seasonally-averaged CCA hindcasts. The skill was then calculated by comparing the hindcasts for these seasonal averages with observed data, for the ensemble mean of each individual model as well as the multimodel ensemble average (NMME$_{avg}$).

Acknowledgements

We thank Ben Livneh for producing and supplying the gridded historical minimum and maximum temperature dataset and the Hadley Center Sea Ice and Sea Surface Temperature for the SST reanalysis. We also thank the agencies that support the NMME system, and the participating climate modeling groups for producing and making their model output available. This study forms a part of the Ph.D. dissertation of DFG, who was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the Grant 221222/2014-6 and by the California Energy Commission under Agreement Number PIR-15-005.

Chapter 4, in full, is in preparation for submission for publication in the *Proceedings of the National Academy of Sciences of the United States of America* as: Dias, D.F., A. Gershunov and D. Cayan (2020), A statistical benchmark for NMME: the case of seasonal predictions of minimum and maximum temperature over the Western North America. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Exploring sources of errors in decadal prediction for the Pacific surface and subsurface ocean

Abstract

A recent 40-member large ensemble of decadal predictions using the Community Earth System Model (CESM-DPLE) shows good levels of skill in many fields in the ocean and in the atmosphere. However, some regions such as the tropical Pacific show very low skill for the surface ocean over the first few lead years. Here, we develop a Linear Inverse Model (LIM) to generate decadal forecasts for the surface and sub-surface ocean and compare its decadal forecast skill against that from the CESM-DPLE. The LIM was constructed using global observed seasonally-averaged anomalies over the period of 1958-2017. To be consistent with CESM-DPLE forecasts, LIM forecasts were initialized every November and integrated forward from one season out to 10 years. The results show that LIM forecast skill is comparable to and sometimes better than that from the

CESM-DPLE ensemble mean over many regions around the globe. In particular, LIM skill is better in the Eastern Subtropical Pacific for lead times from 1 to 5 years. These results indicate that the evolution of the system in those areas may not be not fully driven by unpredictable dynamics and that there may be some room for improvement in the CESM decadal forecasts, given that a low-dimensional linear model is able to generate better skill than the fully-coupled nonlinear model. Therefore, the low skill in the CESM may be related to the misrepresentation of processes in the ocean within and among those regions that exhibit low skill. We investigate possible sources of errors by comparing the LIM feedback matrix obtained from observations with that obtained from the CESM. This matrix show time-scale interactions between components, that in our case represent different subregions of the Pacific Ocean. Results show that the oceanic feedback matrix from CESM differs in several regions from that obtained with observations. These may indicate sources of error in CESM and therefore in its decadal prediction skill that merit focused attention.

5.1 Introduction

The field of decadal climate prediction has gained increased attention over the recent years. Decision-makers concerned with adaptation and resilience to climate variability and change are particularly interested in the coming decade and skillful and reliable predictions of the near-term climate can be very beneficial (Brasseur and Gallardo, 2016; Goddard, 2016; Kushnir et al., 2019). Near-term climate variations are caused by the response of the system to a forced component, such as greenhouse gases and aerosols, and its internally generated variability (Meehl et al., 2009; Solomon et al., 2011). Therefore, the predictions systems for the near-term climate consider not only present and projected anthropogenic forced response, but also the include information from the initial observed state. Major progress in the development of such predictions, performed with global climate models (GCMs) initialized with atmospheric and oceanic observations, has been made in the recent years. A number of modeling centers have carried out initialized decadal retrospective forecasts (also called as ”hindcasts“) together with their uninitialized counterpart. By comparing the initialized and the uninitialized experiments, it is possible to assess the impacts of the external forcing and the internal variability on the decadal predictive skill. Several studies over the last decade have been showing that, with a non-stationary climate, initial information based on observations can significantly enhance the predictive skill over the period of several years for many regions around the globe (e.g., Smith et al., 2007; Doblas-Reyes et al., 2013b; Yeager et al., 2018; Smith et al., 2019).

The benefits of initialization are more clear, in general, in regions in the ocean where the low-frequency variability dominates, such as the North Atlantic. However, a common problem in these initialized climate predictions is the persistent low predictive skill in the Tropical and Subtropical Central and Eastern Pacific. In fact, the predictive skill for surface and subsurface variables in these regions from initialized predictions are even lower than that from the uninitialized experiments (Mueller and Seneviratne, 2012; Doblas-Reyes et al., 2013b; Mignot et al., 2016; Yeager et al., 2018). This deficiency has been attributed to the unbalanced initial conditions in the equatorial Pacific that triggers spurious El Niño and La Niña events at the surface (Pohlmann et al., 2017; Teng et al., 2017). Considering the importance of these regions, particularly the Tropical Pacific, in influencing the regional climate throughout the globe through their teleconnections, it is important to explore other sources of errors in the simulations that can reduce the prediction skill.

Empirically-built statistical models have also been used in the decadal prediction problem (Newman, 2007, 2012; Ho et al., 2013). Being much less computationally expensive to run, such models can be used as a benchmark for GCMs to identify whether these GCMs have room for improvement. Additionally, the statistical models can also be used as a tool to identify deficiencies in the GCMs in representing important atmospheric and oceanic aspects of the global climate (Shin et al., 2010). Therefore, a two-fold approach of combining dynamical and empirical models can contribute to advance the problem of the near-term climate prediction.

In this paper, we evaluate global decadal prediction skill of sea surface temperature and sea surface height anomalies obtained with a linear inverse model (LIM, Penland and Sardeshmukh, 1995) compared to that obtained from a multi-ensemble GCM, the Community Earth System Model Decadal Predictions Large Ensemble (CESM-DPLE, Yeager et al., 2018). Considering the failure in GCMs to predict SST evolution in the Pacific, we use the LIM to identify possible sources of errors in the simulated feedbacks between different regions in the Pacific Ocean.

5.2 Data and Predictive model details

Retrospective decadal predictions of global Sea Surface Temperature and Sea Surface Height anomalies (SST and SSH) are performed and evaluated using two approaches: an empirical approach using linear inverse model (LIM, Penland and Sardeshmukh, 1995) and a dynamical approach using the Community Earth System Model Decadal Predictions Large Ensemble (CESM-DPLE, Yeager et al., 2018). Previous studies showed that LIMs provide a good approximation for observed Pacific SST anomalies evolution on time scales from months to years (Penland and Sardeshmukh, 1995; Newman, 2007; Alexander et al., 2008; Newman, 2012).

Linear Inverse Model is a stochastically forced multivariate linear model whose parameters can be estimate from the observed statistics of the system. It assumes that the evolution of a system can be separated into a linear deterministic part and a stochastic nonlinear part, represented by white noise fluctuation:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{L}\boldsymbol{x} + \boldsymbol{\xi}, \tag{5.1}$$

where $\boldsymbol{x}$ is the *n x 1* system state vector, $\boldsymbol{\xi}$ is the white stochastic forcing and $\boldsymbol{L}$ is the *n x n* linear operator matrix, representing the feedbacks among the components of the vector $\boldsymbol{x}$. $\boldsymbol{L}$ can be estimated from the statistics of the system

$$\boldsymbol{L} = \frac{1}{\tau_0} ln[C(\tau_0)C(0)^{-1}], \tag{5.2}$$

where $\boldsymbol{C}(\tau_0) = \langle \boldsymbol{x}(t+\tau_0)\boldsymbol{x}^T(t) \rangle$ and $\boldsymbol{C}(0) = \langle \boldsymbol{x}(t)\boldsymbol{x}^T(t) \rangle$ are the τ_0-lag and zero-lag covariance matrices, respectively and the angle brackets represent an average over all times. Forecast of $\boldsymbol{x}$ at the lead time $\boldsymbol{\tau}$, using the initial time t can be obtained with

$$\boldsymbol{x}(t+\tau) = \boldsymbol{B}(\tau)\boldsymbol{x}(t), \tag{5.3}$$

where $B(\tau) = exp(\boldsymbol{L}\tau)$ is the *n x n* forecast propagator matrix.

Here, two different configurations were used to build the state vector $\boldsymbol{x}$: one with only SST ($\boldsymbol{x}_{SST}$) included and the other with both SST and SSH ($\boldsymbol{x}_{SST+SSH}$). The choice of inclusion of SSH in the state vector follows results from previous studies showing that explicitly including some measure of oceanic heat content, such as thermocline depth or SSH, improves the ability of the LIM to capture subsurface ocean dynamics on seasonal times scales, particularly in the Tropical Pacific (Neelin et al., 1998; Newman et al., 2011; Newman and Sardeshmukh, 2017). Therefore, we tested whether this assumption also holds for decadal time scales by combining SST and SSH into $\boldsymbol{x}$. We used 3-month averaged anomalies of global SST and SSH filtered into an empirical orthogonal function (EOF) space. For $\boldsymbol{x}_{SST+SSH}$, a joint EOF was used. The time evolution of these EOFs, also called as principal components (PCs), were used as the components of $\boldsymbol{x}_{SST}$ and $\boldsymbol{x}_{SST+SSH}$. The leading 20 SST PCs (40 joint SST-SSH PCs) were used for this model, retaining 80% of the total SST variance (75% of the joint variance) and defining a 20-component (40-component) state vector $\boldsymbol{x}_{SST}$ ($\boldsymbol{x}_{SST+SSH}$):

$$\boldsymbol{x}_{SST} = \begin{bmatrix} PC1_{SST} \\ PC2_{SST} \\ ... \\ PC20_{SST} \end{bmatrix} \quad (5.4) \qquad \boldsymbol{x}_{SST+SSH} = \begin{bmatrix} PC1_{SST+SSH} \\ PC2_{SST+SSH} \\ ... \\ PC20_{SST+SSH} \end{bmatrix} \quad (5.5)$$

The SST field was obtained from the Hadley Center Sea Ice and Sea Surface Temperature (HadISST, Rayner et al., 2003) and the SSH field was obtained from the e European Centre for Medium-Range Weather Forecasting (ECMWF) Ocean Reanalysis System 4 (ORAS4, Balmaseda et al., 2013). Both fields were interpolated on a 1° by 1° grid from 80°S to 70°N for the period January 1958 to December 2017, corresponding to the period in which there is availability of SSH data from ORAS4.

The state vectors $\boldsymbol{x}_{SST}$ and $\boldsymbol{x}_{SST+SSH}$ from Eq. 5.4 and Eq. 5.5 were replaced in Eq. 5.1 and, following the procedure described above, the propagator matrices $\boldsymbol{B}_{SST}$ and $\boldsymbol{B}_{SST+SSH}$ were used to perform the forecasts of the SST PCs and the joint SST and SSH PCs. The forecasts were made by initializing the model each November from 1958 to 2007 (for a total of 50 start dates) and integrating it forward for 10 years. The initialization in November of each year was made to be consistent with the initialization scheme from the CESM-DPLE, then both models are directly comparable. Importantly, to avoid artificial skill from overfitting, the estimates of $\boldsymbol{B}$ and the forecast skill are made in a cross-validation sense, in which the 10 years that are being predicted are excluded from the training period. This process is repeated for every initialization date.

The LIM hindcasts were compared to that from the CESM-DPLE, which is a decadal prediction project composed of 40 ensemble members and based on the CESM version 1.1 (Hurrell et al., 2013). It uses the same code base, component model configurations, and historical and projected radiative forcing as its uninitialized counterpart, the CESM Large Ensemble (CESM-LE, Kay et al., 2015). CESM-DPLE was generated by initializing the full field each year on November, from 1954 to 2017. The predictions are then made by integrating the model forward for 10 years. The ocean and ice components were initialized from a forced ocean-sea ice reconstruction and the atmosphere and land components were initialized from a single member of CESM-LE. Further details on the CESM-DPLE can be found in Yeager et al. (2018).

To be consistent with the SSH observation period, the skill for LIM and CESM-DPLE hindcasts were calculated over the period 1958 – 2017. As a measure of skill, the anomaly correlation coefficient (ACC) was used. ACC was calculated by comparing LIM and CESM-DPLE predicted anomalies with the observations at each grid point for two lead times, corresponding to long-term averages: Years 1 to 5 and Years 3 to 7. "Year 1" corresponds to January to December average of the anomalies predicted in the first year. For example, for a prediction initialized in November 1958, "Year 1" is the 1959 January to December average; "Year 1–5" is the average over the period of January 1959 to December 1963; "Year 3–7" is the average over the period of January 1961 to December 1965. The statistical significance of the CESM-DPLE and LIM ACC is calculated using a Monte Carlo approach: the correlation between the predicted anomalies and a random time series is calculated and the process is repeated 1,000 times. ACC values that are above or below 3 times the standard deviation of this random distribution are considered significant with 99% of confidence.

5.3 Evaluation of LIM and DPLE decadal skill

We first assess the skill at predicting global SST using both LIM configurations described here: LIM trained with only SST data (LIM_{SST}) and the LIM that included

SSH together with SST ($LIM_{SST+SSH}$). We also evaluate the SSH predictive skill, given that this information can be useful for hazard management applications (e.g., Chowdhury and Chu, 2015). At shorter lead times, pentadal variations of both SST and SSH are well predicted in vast regions in every ocean basin (Figures 5.1a,c,e). However, LIM fails to predict these variations in SST and SSH along the Equator, particularly in the equatorial Pacific ocean, depicted as a band of lower and not significant ACC. The inclusion of SSH in the LIM yields an improvement in the ability to predict SST variations in every region. Although for the first pentadal lead time the ACC of the LIM_{SST} is significant in many regions (Figure 5.1a), the ACC values for SST of the $LIM_{SST+SSH}$ are much higher (Figure 5.1b). This improvement is more remarkable in the Tropical and Southern Pacific as well as along the Southern Ocean, where ACC values for the LIM_{SST} are not significant (black dots in Figure 5.1 indicates values statistically not significant).

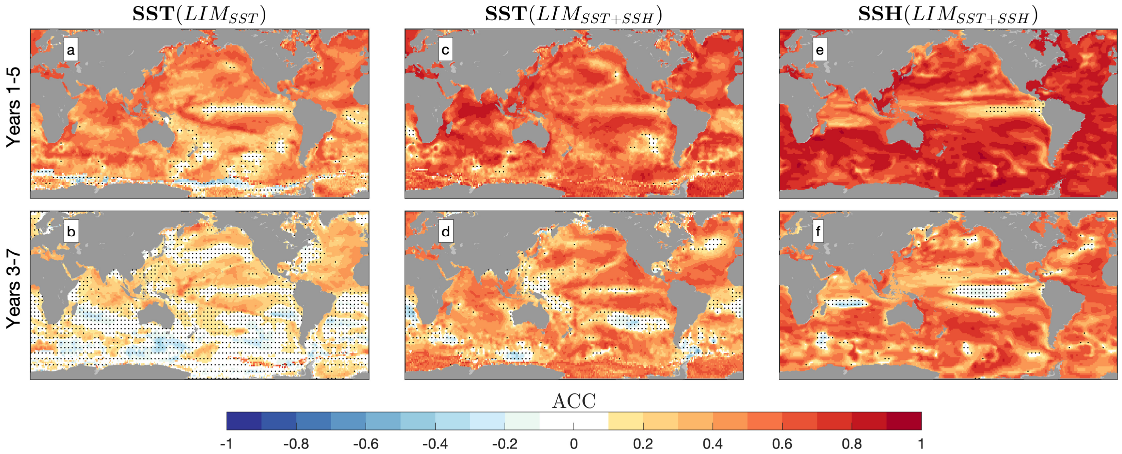


Figure 5.1: (a–b) ACC scores for SST anomalies from LIM_{SST} for lead times Years 1–5 and years 3–7, respectively; (c–d) same as (a) and (b), but for ACC scores from $LIM_{SST+SSH}$; (e–f) same as (c) and (d), but for SSH anomalies. The absence of black dots indicate values that have a statistical significance greater than 99%.

At longer lead times, SSH is even more important for the prediction of SST anomalies. For the LIM_{SST} the skill drops significantly in every location of all the ocean basins, particularly in the Southern Hemisphere, where ACC is insignificant (Figure 5.1b). On the other hand, when SSH is included, there is a loss of skill in some specific locations, such as the Equatorial and Western Pacific Oceans, the Extratropical South Pacific Ocean off the coast of Chile, and the North Atlantic, particularly the Subpolar region; however, the scores remains high (ACC > 0.6) and significant elsewhere (Figure 5.1d). For SSH, the most remarkable skill loss occurs in the Tropical Pacific and Tropical Indian Oceans (Figure 5.1d).

Since LIM can predict global SST and SSH on decadal time scales, we now explore how LIM skill compares to that obtained with CESM-DPLE. ACC scores differences between LIM and CESM-DPLE show that for shorter lead times LIM skill, particularly $LIM_{SST-SSH}$ skill, is comparable to and in some regions better than that from CESM-DPLE in predicting the pentadal SST and SSH variability (Figures 5.2a,c,e). The North Atlantic, the Western Tropical Pacific and the Indian Ocean are the regions where CESM-DPLE presents a better skill than LIM in predicting SST. These differences become stronger for longer lead times, when the ΔACC becomes significantly more negative (Figures 5.2b,d). The persistent high skill of CESM-DPLE over the Indian Ocean was attributed to the dominance of the external forcing, in the form of a warming trend, in the SST variance in this region (Yeager et al., 2018). LIM captures part of this variability, shown as significant ACC scores in the Indian Ocean in Figures 5.1c,d. However, at longer lead times, it fails to capture some of this warming trend, resulting in negative ΔACC scores over the Indian Ocean in Figure 5.2d. In the North Atlantic, the CESM-DPLE horseshoe-like pattern of higher SST skill, particularly for Years 3-7 (Figure 5.2d), resembles the canonical pattern of the Atlantic Multidecadal Variability (AMV, Sutton and Hodson, 2005). This was the region where the initialization conferred the greatest benefits for the SST skill. As for the Indian Ocean, LIM captured part of the SST variability in the North Atlantic, but there is a strong decrease with lead time.

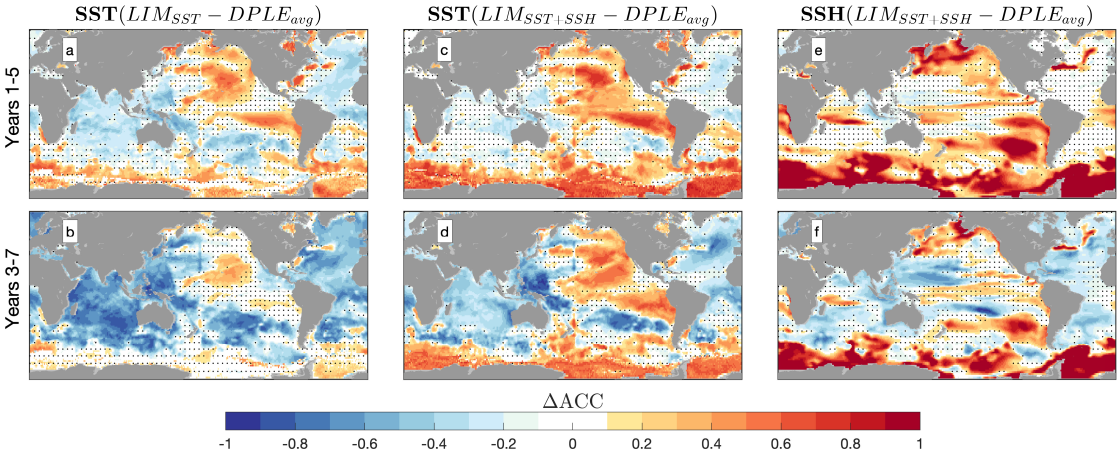


Figure 5.2: $\Delta ACC(ACC_{LIM} - ACC_{DPLE-avg})$ for SST anomalies from LIM_{SST} for lead times Years 1–5 and years 3–7 (a–b,respectively); (c–d) same as (a) and (b), but for ACC scores from $LIM_{SST+SSH}$; (e–f) same as (c) and (d), but for SSH anomalies. Red indicate regions where LIM has better skill than CESM-DPLE, blue are for regions where CESM-DPLE outperforms LIM and regions with black dots indicate that the differences are not statistically significant at 99% level.

The Southern Ocean, the Central Tropical Pacific Ocean, and the Northeastern and Southeastern Pacific Oceans stand out as regions where LIM outperforms CESM-DPLE, with ΔACC in general higher than 0.4. At longer lead times, while the ΔACC for LIM_{SST} are either not significant or are negative (meaning that CESM-DPLE has better skill than LIM_{SST}, Figure 5.2b), for $LIM_{SST+SSH}$ ΔACC remains high, emphasizing regions where CESM-DPLE presents low predictability (Figures 5.2c,e), as shown by Yeager et al. (2018). The Southern Ocean is the region where the skill is overall very low in both CESM-DPLE and CESM-LE, with skill that degrades with lead time, even when compared to simple persistence. It is noteworthy that LIM can predict the SST and SSH variability in this region even at longer lead times. The improvement in the SST skill over this region is a reflection of the good skill in predicting SSH, since LIM_{SST} cannot capture this variability, particularly for Years 3-7 (Figures 5.2b,d,e).

The skill improvement with LIM in the Central Tropical Pacific, in the Southeastern Pacific off the coast of Peru, and in the Northeastern Pacific off the coast of California is remarkable. The low prediction skill in this region, in particular for the Tropical Pacific, is a common problem not only for the CESM-DPLE, but also for many other decadal prediction systems. This failure in predicting the SST variability over this region has been attributed to the initialization shock, in which unbalanced initial conditions in the equatorial Pacific triggers artificial El Niño and La Niña events at the surface (Pohlmann et al., 2017; Teng et al., 2017; Yeager et al., 2018). The significant skill obtained with LIM for these regions indicates that the SST evolution on decadal time scales is, at least in part, driven by deterministic linear processes and is therefore predictable. Moreover, the fact that the LIM_{SST} does not show the same predictive ability as $LIM_{SST+SSH}$ indicates that processes in the subsurface ocean contributes to this predictive ability. Bearing that in mind, we use the LIM linear operator matrix to explore possible sources of errors in the CESM to reproduce the SST and SSH variability in the Pacific Ocean, as explained in next sections.

5.4 Diagnostic approach

The diagnostic approach for the Pacific SST and SSH interactions uses the method described in (Shin et al., 2010). These authors showed that several climate models do not reproduce well the observed SST interactions between different regions in the Tropical Pacific Ocean, in particular the interactions between Subtropical Pacific with the ENSO region and with the Western Pacific Warm Pool. Therefore, we aim to evaluate how well CESM simulates the linear feedbacks among different regions in the Pacific. In order to do so, we use the linear operator matrix, $\boldsymbol{x}$, from Eq. 5.2 to evaluate the observed and simulated interactions among the components in $\boldsymbol{L}$.

It is important to define a state vector $\boldsymbol{x}$ in which is possible to calculate these linear feedbacks. Since the focus is to evaluate interactions among different regions in the

Pacific Ocean, the state vector $\boldsymbol{x}$ for the diagnostic method is different from that of the predictive approach. Instead of truncating the datasets into an EOF space, we construct $\boldsymbol{x}$ based on a grid space. This is done because the dominant EOF modes are not necessarily geographically contained, accounting for different fraction of the SST and SSH variance at different locations. Therefore, we establish six different regions in the Pacific Ocean among which we investigate the observed and simulated SST and SSH interactions (Figure 5.3). We calculate the average of SST and SSH anomalies within each of those six regions and define a 12-component state vector $\boldsymbol{x}$ (6 regions and 2 variables):

$$\boldsymbol{x} = \begin{bmatrix} SST_{r1} \\ \vdots \\ SST_{r6} \\ SSH_{r1} \\ \vdots \\ SSH_{r6} \end{bmatrix} \tag{5.6}$$

where $r1$ to $r6$ represents each of the regions shown in Figure 5.3. With this definition of $\boldsymbol{x}$, our linear feedback matrix becomes:

$$\boldsymbol{L} = \begin{bmatrix} L_{SST-SST} & L_{SSH-SST} \\ L_{SST-SSH} & L_{SSH-SSH} \end{bmatrix} \tag{5.7}$$

where the components of each submatrix of $\boldsymbol{L}$,e.g. $\boldsymbol{L}_{SST-SST}$, are defined as:

$$\boldsymbol{L}_{SST-SST} = \begin{bmatrix} r1-r1 & \cdots & r6-r1 \\ \vdots & \ddots & \vdots \\ r1-r6 & \cdots & r6-r6 \end{bmatrix} \tag{5.8}$$

Each component of $\boldsymbol{L}$, L_{ij}, identifies a time scale for the influence of the anomalies in region j on the anomalies in region i. The submatrices $L_{SST-SST}$ and $L_{SSH-SSH}$ give the time scale interaction of SST and SSH with itself between the different regions, whereas submatrices $L_{SSH-SST}$ and $L_{SST-SSH}$ present the feedbacks between SST and SSH. Note that $L_{i,j}$ is different from $L_{j,i}$, as it shows the direction of the interaction. For example $L_{SST-SSH}$ brings the influence of SSH anomalies in the SST anomalies. In other words, it shows how SSH anomalies in one region drive changes in the SST anomalies in other region. As noted by Penland and Sardeshmukh (1995), even though $\boldsymbol{L}$ includes only SST and SSH anomalies, it implicitly accounts for the effect of other variables, including atmospheric conditions. Therefore, following Shin et al. (2010), we interpret the SST interactions ($L_{SST-SST}$) as atmospheric teleconnections among the different regions and the SSH interactions ($L_{SSH-SSH}$) as oceanic teleconnections.

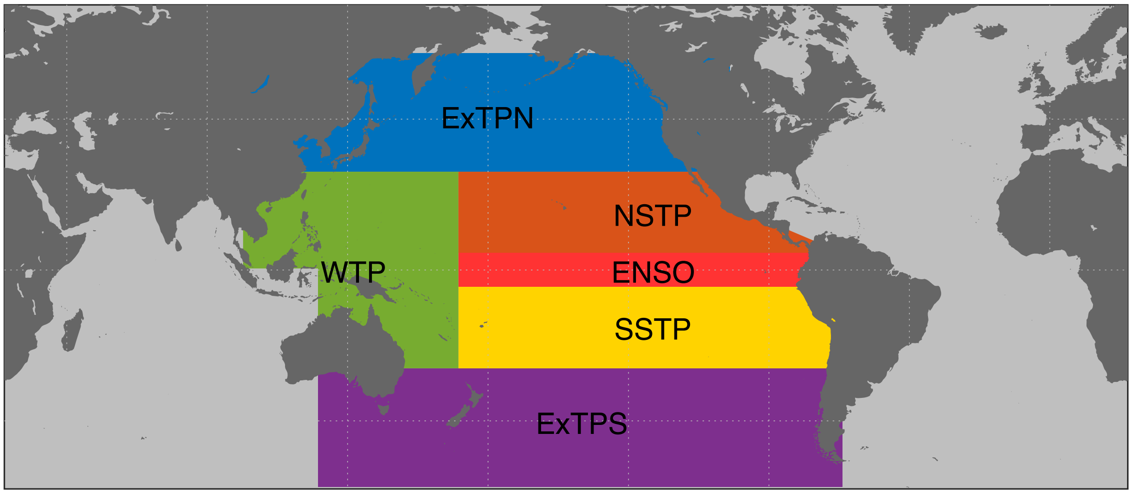


Figure 5.3: The six regions used to contruct the linear operator matrix $\boldsymbol{L}$ for the diagnostic analysis. Region 1 – ExTPN (Extratropical Pacific North); Region 1 – WTP (Western Tropical Pacific); Region 3 – NSTP (Northern Subtropical Pacific); Region 4 – ENSO (El Niño Southern Oscillation or Equatorial Pacific); Region 5 – SSTP (Southern Subtropical Pacific); and Region 6 – ExTPS (Extratropical Pacific South);

With these definitions, we use Eq. 5.2 and the state vector $\boldsymbol{x}$ from Eq. 5.6 to estimate the $\boldsymbol{L}$ from observations and from the CESM-LE. The observationally based $\boldsymbol{L}$ ($\boldsymbol{L}_{obs}$) was constructed with SST from HadISST and with SSH from ORAS4. We then compared $\boldsymbol{L}_{obs}$ with the linear operator matrices obtained from each individual ensemble of CESM-LE ($\boldsymbol{L}_{cesm-k}$, where $k = 1, 2, \cdots, 40$ is the ensemble member) and from the ensemble average ($\boldsymbol{L}_{cesm-avg}$). Therefore, we have a total of 42 $\boldsymbol{L}$ matrices. These matrices were estimated from monthly-average anomalies for the period 1958–2017, which is the period when there is availability for SSH data. The choice of using CESM-LE instead of CESM-DPLE is because we need a continuous run, over many decades, to estimate the feedbacks in the linear operator matrix and CESM-DPLE provides only 10 years of continous run. Finally, CESM-DPLE uses the same model and the same component configuration as that used in the CESM-LE. Therefore, they are directly comparable.

5.5 Errors in the feedback matrices

We first show a scatter plot of the components in the linear operator matrix calculated with the ensemble average of the CESM-LE ($\boldsymbol{L}_{cesm-avg}$) against the values obtained with the observed fields ($\boldsymbol{L}_{obs}$). This plot is shown in Figure 5.4. Each point represents the interaction between two of the regions in Figure 5.3 for different combination of variables, indicated by different colors. The values range from very close to zero (indicating that the interactions between regions are either very slow or non-existent) to around 0.4 months $^{-1}$, indicating that the damping (if negative value) or growth (if positive value) in the anomalies of one region by other region takes around 2.5 months. Ideally, the points would be closer to the black line, meaning that the observed and simulated feedbacks are similar. Despite the fact that some points are close to the black line, we observe a big spread in the values. Additionally, many of these simulated interactions have the opposite sign as the observed values, as shown by points located in the upper left and lower right corner of Figure 5.4. Therefore, many of these linear feedbacks among different regions are not well simulated by the CESM-LE.

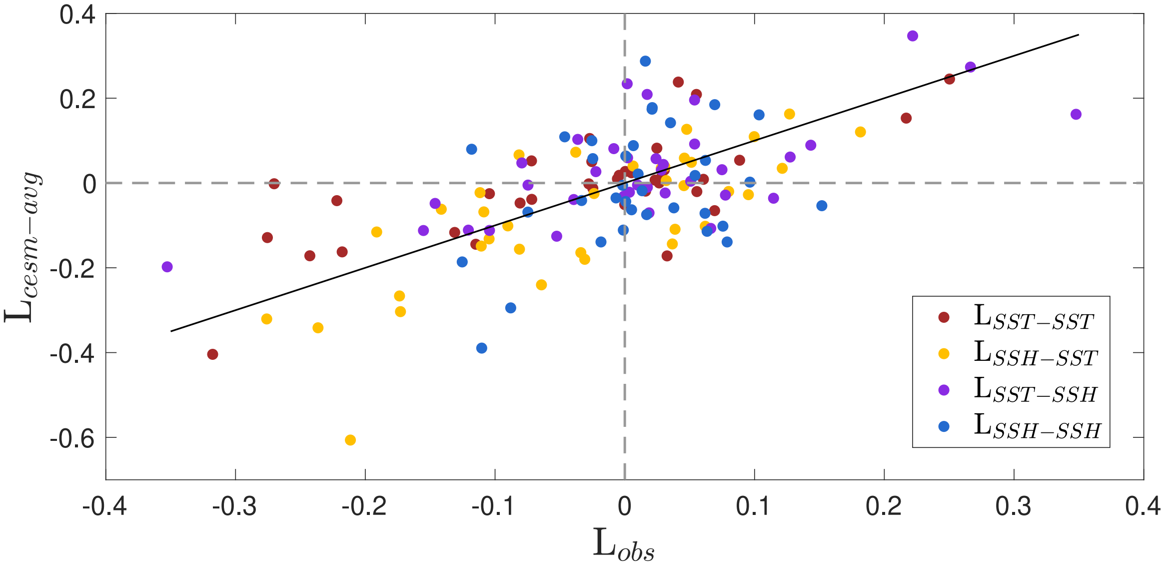


Figure 5.4: Components of the 12x12 CESM-LE ensemble average linear operator matrix ($\boldsymbol{L}_{cesm-avg}$) plotted against the components of $\boldsymbol{L}_{obs}$ (units: months^{-1}). Colors represent the submatrices of $\boldsymbol{L}_{cesm-avg}$ and $\boldsymbol{L}_{obs}$, indicated in the legend. Points that are close to the black line indicate that the simulated feedbacks obtained with CESM-LE are similar to that from the observations; points that are located in the upper left and lower right corners indicate that the CESM-LE simulates interactions with the opposite sign of the observed interactions.

To understand in which regions these interactions are not well represented, we show a comparison among all the components of the 42 linear operator matrices in Figures 5.5 and 5.6. Each element of these matrices show the feedback coefficients between two different regions (off-diagonal) and of a region with itself (diagonal), calculated for each individual ensemble member, which distribution is shown in the grey bars; for the ensemble mean, shown as the yellow line; and for the observation, shown as purple line. Figure 5.5 shows the interaction of SST with itself, being an approximation of the local atmospheric influence or the atmospheric influence of one region in another. Figure 5.6, in turn, shows these influences that occurs in the subsurface ocean.

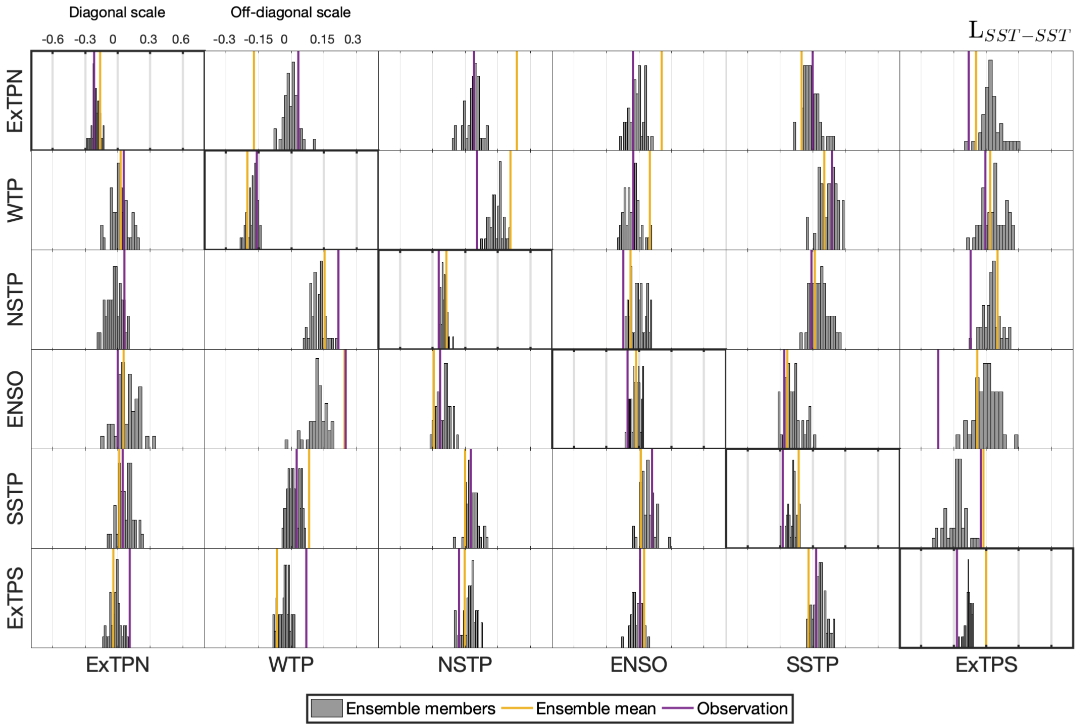


Figure 5.5: Components of each individual submatrix $\boldsymbol{L}_{SST-SST}$ from the linear operator matrix $\boldsymbol{L}$, estimated using observations and the CESM-LE (units: months^{-1}). For each matrix element (i,j), the grey bars represent the distribution of $\boldsymbol{L}_{i,j}$ estimated with each of the 40 CESM-LE ensemble members, yellow line show $\boldsymbol{L}_{i,j}$ estimated for CESM-LE ensemble mean value, and purple line show $\boldsymbol{L}_{i,j}$ for the observations. Diagonal elements $(i=j)$ show the local influence, hence, the local damping coefficients; off-diagonal elements show the influence of one region in another region. Note that the scale for the diagonal elements is different from that of the off-diagonal. Importantly, $\boldsymbol{L}_{i,j} \neq \boldsymbol{L}_{j,i}$ as it shows the direction of the interaction. For example, the influences *on* Extratropical Pacific North (ExTPN) are shown in the components $\boldsymbol{L}_{1,j}$, while the influeces *of* ExTPN on other regions are shown in the components $\boldsymbol{L}_{i,1}$.

The predominant negative values along the diagonal elements of Figure 5.5 indicates that the local SST anomalies in the Pacific are damped by local interactions with the atmosphere. The time scale of this local damping varies throughout the different regions. The fastest damping occurs in the Western Tropical Pacific (WTP-WTP, $\boldsymbol{L}_{2,2}$), with a time scale of 3 months. The Subtropics and Extratropics damping coefficient are very similar, but the South Pacific local SST damping is faster than that from the North Pacific (about 3.5 months for the South Subtropical Pacific – SSTP-SSTP, $\boldsymbol{L}_{5,5}$ – and for the Extropical

Pacific South – ExTPS-EXTPS, $\boldsymbol{L}_{6,6}$ –, and about 4.5 months for the North Subtropical Pacific – NSTP-NSTP, $\boldsymbol{L}_{3,3}$ – and for the Extropical Pacific South – ExTPN-EXTPN, $\boldsymbol{L}_{1,1}$). The longest local damping occurs in the ENSO region, with time scale of about 9.5 months($\boldsymbol{L}_{4,4}$). The ensemble mean values of the local damping coefficients reasonably agree with the observed values and the ensemble spread is usually small. However, there is a tendency of underestimating the time scale of these interactions by CESM-LE$_{mean}$, representing a weaker than observed local damping. The only exception is for WTP, where $\boldsymbol{L}_{2,2}$ for CESM-LE$_{mean}$ is stronger (i.e., faster) than the observed one. The greatest biases occur in the ENSO and ExTPS regions, where the local damping estimate by CESM-LE$_{mean}$ is significantly slower than that from the observations. This weaker local damping for the ENSO has been shown in previous studies using other coupled model simulations (Sun et al., 2006; Shin et al., 2010). The authors attributed this spurious result to the excessive coldness of the long-term mean SSTs in the ENSO region. For the ExTPS, the CESM-LE$_{mean}$ local damping is close to zero, indicating that the observed damping is not captured by the forced response of the CESM.

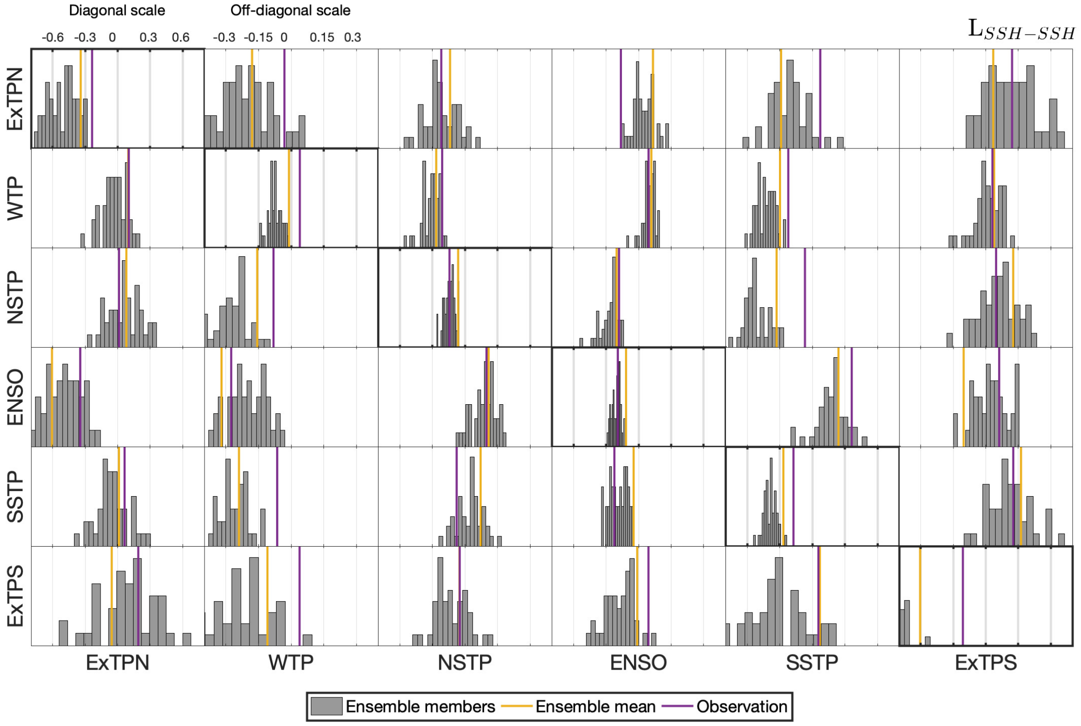


Figure 5.6: Same as Figure 5.5, but for the $\boldsymbol{L}_{SSH-SSH}$, showing the oceanic interactions.

The local SSH feedbacks indicate that, in general, the local SSH anomalies are also damped by its interaction with the ocean subsurface. Not surprisingly, the time scale for these interaction tends to be few months slower than that from the SST local feedbacks, showing that the ocean response is slower than the ocean-atmosphere interactions. The exception is for the WTP region, which presents positive feedback value. This indicates that the SSH anomalies grow because of the local interactions in the ocean. Additionally, WTP is one of the regions where the simulated local feedbacks have the greatest bias. All the ensemble members present negative values and the ensemble mean shows a negative value, but close to zero. This shows that, although the local SST interactions for this region are relatively well simulated by the CESM, the simulated local ocean feedbacks have the opposite sign. For the ENSO region, however, the local oceanic damping is well simulated, while the atmospheric damping is much weaker than the observed. Finally, the

ExTPS region also stands out as one of the greatest biases, as for the local SST interactions. However, while the simulated local SST damping is much weaker than the observed (and non-present for the ensemble mean), the local SSH damping simulated with CESM-LE is much faster (1.5 months for CESM-LE$_{avg}$ against 5 months for the observations).

As for the remote interactions between different regions (off-diagonal elements, $\boldsymbol{L}_{i \neq j}$, of Figures 5.5 and 5.6), some many model misrepresentations are evident, both in the remote SST and SSH feedbacks. For example, the Western Tropical Pacific contributes to the growth of the anomalies on both North and South Extratropics (ExTPN, $\boldsymbol{L}_{1,2}$, and ExTPS, $\boldsymbol{L}_{6,2}$), but the simulated values, especially the ensemble mean, indicate that a damping effect occurs instead. This error is consistent for both SST and SSH remote interactions. Additionally, there is also a misrepresentation of the effects of other regions on the WTP, such as the SST remote effects from NSTP and ENSO on WTP (Figure 5.5, $\boldsymbol{L}_{2,3}$ and $\boldsymbol{L}_{2,4}$).

Besides the interactions with the WTP region, Figures 5.5 and 5.6 suggests model misrepresentations of the time scale and direction of the interactions between other regions, such as the ExTPS influence on ENSO region ($\boldsymbol{L}_{6,4}$) and the ENSO influence on the ExTPN ($\boldsymbol{L}_{1,4}$). These errors are consistent for both SST and SSH interactions, suggesting that there are possible errors in representing the coupling between subsurface and surface ocean in the CESM simulation.

5.6 Summary and concluding remarks

In this study, we performed retrospective decadal predictions of global sea surface temperature (SST) and sea surface height (SSH) anomalies using empirically-built linear inverse models (LIM). Two LIM configurations were used to predict decadal variations in the global SST anomalies: one trained only with SST data and the other trained with SST and SSH data. We showed the inclusion of SSH significantly improves the LIM ability to predict the decadal variations in the SST anomalies, in particular at longer lead times. SSH, which is a measure for ocean heat content, increases the memory of the system and therefore improves the predictability at longer lead times. LIM skill was compared to that obtained with the Community Earth System Model Decadal Predictions Large Ensemble (CESM-DPLE). We found that the CESM-DPLE skill, for both SST and SSH does not notably exceed that from the LIM, particularly at shorter lead times. In fact, LIM skill proved to be better than that from CESM-DPLE in many regions around the globe. At longer lead times, CESM-DPLE skill exceeds that from the LIM in regions dominated by a strong warming trend, such as the Indian Ocean and the Western North Pacific. However, LIM notably exceeds CESM-DPLE skill in the Southern Ocean and in the Eastern Tropical and Subtropical Pacific Oceans.

Considering that the low prediction skill in the Eastern Tropical and Subtropical Pacific Oceans is a common problem for many decadal prediction systems, we explored possible sources of errors in predicting the SST and SSH evolution in the Pacific using the LIM linear operator matrix. We investigate the observed and simulated interactions among six broad regions in the Pacific, constructing a 12-component state vector and estimating a 12x12 linear operator matrix for observations, for each individual 40 ensemble members of CESM-LE, and for the ensemble mean. In general, we found that the local feedbacks for each region are well-represented in the simulations, with the exception in the ENSO region for SST and in the Extratropical South Pacific, for both SST and SSH. It was found that the observed local damping for these regions is much reduced in the simulations. Finally, we showed that the many remote influences are not well represented in the CESM, in particular the influence of the Western Tropical Pacific (WTP) on the other regions as well as the influence of other regions on the Western tropical Pacific. Since the WTP is a major heat and moisture source for the global atmospheric circulation (Godfrey et al., 1998; Kucharski et al., 2006; Sun et al., 2017), errors in simulating the remote effects can have implications for climate variability elsewhere. Many of these misrepresentations are coherent for the SST and SSH influences, indicating that there could be errors in the representation of the coupling between subsurface and surface ocean in the CESM simulations.

It is important to view the results of this paper with some caution. We understand that estimates of decadal predictability from 60 years of data is limited. When considering the errors estimate, it is also important to understand that the observation represents only one realization, while the CESM simulations give a wide-range of realizations. However, it is remarkable the improvement in the skill with the LIM in regions that the dynamical models have always had problems in simulating, such as Central Tropical Pacific. This is a promising result, showing that, at least in part, the decadal dynamics of SST and SSH are deterministic and there might be some room for improvement with the dynamical models.

Acknowledgements

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the National Science Foundation. We are grateful for the National Science Foundation (OCE1419306) and the National Oceanic and Atmospheric Administration (NOAA-MAPP; NA17OAR4310106) for funding that supported this research. DFD was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the grant 221222/2014-6.

Chapter 5, in part, is in preparation for submission for publication as: Dias, Daniela Faggiani; Yeager, Stephen; Miller, Arthur J.; Capotondi, Antonietta; Cayan, Daniel R.

Exploring sources of errors in decadal predictions of the Pacific surface and sub-surface ocean. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Conclusions

The Earth's weather and climate is intrinsically chaotic and an improvement in the state-of-art prediction systems challenges the best of our understanding and capabilities. In this dissertation, we showed that the use of a hierarchy of models of varying complexity can be very useful to explore untapped skill in the Earth's climate and to diagnose possible misrepresentations in complex GCMs of local and remote interactions between components of the climate system. These misrepresentations can, in turn, affect the ability of these complex models to perform skillful predictions at a variety of time scales.

In Chapter 2, we demonstrate the importance of temporal scale interactions in improving the forecast skill on decadal timescales. By separating the variability into different time scales and different regions, we identified how interactions between time scales affect predictability and how regions affect each other in driving predictable components linked to ENSO or intrinsic mid-latitude interactions. Specifically, we showed that the Extratropical Pacific enhanced the forecast skill for the Tropical Pacific on seasonal to interannual time scales. This enhanced predictive skill can be due the Pacific Meridional Mode (PMM) propagating the mid-latitude SST anomalies associated with the North Pacific Oscillation (NPO). On the other hand, the Tropical Pacific enhances the predictive skill of the extratropics when the decadal variability is considered. The atmospheric bridge mechanism, being one of the contributors to the PDO-related variability, can act to enhance the decadal predictive skill.

In Chapter 3, we showed that these slowing-varying components of the Pacific Ocean contribute to seasonal predictive ability of minimum and maximum temperature (Tmin and Tmax, respectively) over Western North America (WNA). Additionally, local effects between soil moisture and the overlying atmosphere seems to also be important in predicting Tmax during wintertime, coinciding with the season that soil moisture has the stronger memory (autocorrelation) in the region and therefore having more potential to influence the overlying atmosphere one season ahead.

In Chapter 4 and Chapter 5, we showed that the state-of-art global prediction systems for seasonal and decadal time scales have room for improvement in diverse aspects of the climate system. We showed that the predictive skill of linear statistical models is comparable to and sometimes better than that obtained from these dynamical hindcasts. We, therefore, suggest that a higher benchmark could be set for these dynamical forecasts - a baseline level of skill that complex dynamical models must aim to exceed. Additionally, the fact that simpler statistical models can generate some level of skill indicates that the dynamics of the system is, at least in part, driven by deterministic linear processes and is therefore predictable. Finally, in Chapter 5, we propose a comprehensive method to explore errors in the simulated local and remote interactions between different regions and components of the climate system. This method is based on the linear feedback matrix from the linear inverse modelling (LIM) approximation. Guided by the results showing that LIM skill for SST and SSH in the Eastern Tropical and Subtropical Pacific Oceans outperforms that from the CESM-DPLE, we investigate the observed and simulated interactions among six broad regions in the Pacific. In general, we found that the local feedbacks for each region are well-represented in the simulations, with the exception in the ENSO region for

SST and in the Extratropical South Pacific, for both SST and SSH. It was found that the observed local damping for these regions is much reduced in the simulations. Finally, we showed that the many remote influences are not well represented in the CESM, in particular the influence of the Western Tropical Pacific on the other regions as well as the influence of other regions on the Western tropical Pacific. Many of these misrepresentations are coherent for the SST and SSH influences, indicating that there could be errors in the representation of the coupling between subsurface and surface ocean in the CESM simulations.

Chapter 7

Bibliography

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research Atmospheres*, **111 (5)**, 1–22, doi:10.1029/2005JD006290.

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N. C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, **15 (16)**, 2205–2231, doi:10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Alexander, M. a., and C. Deser, 1995: A Mechanism for the Recurrence of Wintertime Midlatitude SST Anomalies. *Journal of Physical Oceanography*, **25**, 122–137, doi:10.1175/1520-0485.

Alexander, M. A., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. *Journal of Climate*, **21 (2)**, 385–402, doi:10.1175/2007JCLI1849.1.

Alexander, M. A., D. J. Vimont, P. Chang, and J. D. Scott, 2010: The Impact of Extratropical Atmospheric Variability on ENSO : Testing the Seasonal Footprinting Mechanism Using Coupled Model Experiments. *Journal of Climate*, **23 (11)**, 2885–2901, doi:10.1175/2010JCLI3205.1.

Alfaro, E. J., A. Gershunov, D. Cayan, E. J. Alfaro, A. Gershunov, and D. Cayan, 2006: Prediction of Summer Maximum and Minimum Temperature over the Central and Western United States: The Roles of Soil Moisture and Sea Surface Temperature. *Journal of Climate*, **19 (8)**, 1407–1421, doi:10.1175/JCLI3665.1.

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geoscience*, **8 (4)**, 278–283, doi:10.1038/ngeo2385.

Baldwin, M. P., and Coauthors, 2001: The Quasi-Biennial Oscillation. *Reviews of Geophysics*, **29 (2)**, 179–229.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, **139 (674)**, 1132–1161, doi:10.1002/qj.2063.

Barnett, T. P., and R. Preisendorfer, 1987: Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Monthly Weather Review*, **115 (9)**, 1825–1850, doi:10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2.

Barnston, A. G., 1994: Linear Statistical Short-Term Climate Predictive Skill in the Northern Hemisphere. *Journal of Climate*, **7 (10)**, 1513–1564.

Barnston, A. G., and R. E. Livezey, 1987: Classification, Seasonality an Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, **115 (6)**, 1083–1126.

Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *Journal of Climate*, **5 (11)**, 1316–1345, doi:10.1175/1520-0442(1992)005<1316:POEEUC>2.0.CO;2.

Barnston, A. G., M. K. Tippett, M. Ranganathan, and M. L. L'Heureux, 2017: Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, **0 (0)**, 1–20, doi:10.1007/s00382-017-3603-3.

Barnston, A. G., M. K. Tippett, H. M. van den Dool, and D. A. Unger, 2015: Toward an improved multimodel ENSO prediction. *Journal of Applied Meteorology and Climatology*, **54 (7)**, 1579–1595, doi:10.1175/JAMC-D-14-0188.1.

Battisti, D. S., and A. C. Hirst, 1989: Interannual Variability in a Tropical Atmosphere-Ocean Model: Influence of the Basic State, Ocean Geometry and Nonlinearity. *Journal of Atmospheric Sciences*, **46 (12)**, 1687–1712.

Becker, E., and H. van den Dool, 2016: Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment. *Journal of Climate*, **29 (8)**, 3015–3026, doi:10.1175/JCLI-D-14-00862.1.

Becker, E., H. Van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *Journal of Climate*, **27 (15)**, 5891–5906, doi:10.1175/JCLI-D-13-00597.1.

Bjerknes, J., 1966: A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, **18 (4)**, 820–829, doi:10.3402/tellusa.v18i4.9712.

Boer, G. J., and Coauthors, 2016: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, **9 (10)**, 3751–3777, doi: 10.5194/gmd-9-3751-2016.

Branstator, G., and H. Teng, 2010: Two limits of initial-value decadal predictability in a CGCM. *Journal of Climate*, **23 (23)**, 6292–6311, doi:10.1175/2010JCLI3678.1.

Brasseur, G. P., and L. Gallardo, 2016: Climate services: Lessons learned and future prospects. *Earth's Future*, **4 (3)**, 79–89, doi:10.1002/2015EF000338.

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

Capotondi, A., and P. D. Sardeshmukh, 2015: Optimal precursors of different types of ENSO events. *Geophysical Research Letters*, **42**, 9952–9960, doi:10.1002/ 2015GL066171.

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

Cavanaugh, N. R., T. Allen, A. Subramanian, B. Mapes, H. Seo, and A. J. Miller, 2014: The skill of atmospheric linear inverse models in hindcasting the Madden–Julian Oscillation. *Climate Dynamics*, **44 (3-4)**, 897–906, doi:10.1007/s00382-014-2181-x.

Ceballos, L. I., E. Di Lorenzo, C. D. Hoyos, N. Schneider, and B. Taguchi, 2009: North Pacific gyre oscillation synchronizes climate fluctuations in the eastern and western boundary systems. *Journal of Climate*, **22 (19)**, 5163–5174, doi:10.1175/2009JCLI2848.1.

Chang, P., L. Zhang, R. Saravanan, D. J. Vimont, J. C. Chiang, L. Ji, H. Seidel, and M. K. Tippett, 2007: Pacific meridional mode and El Niño - Southern oscillation. *Geophysical Research Letters*, **34 (16)**, 1–5, doi:10.1029/2007GL030302.

Chen, N., and A. J. Majda, 2016a: Simple dynamical models capturing the key features of the Central Pacific El Niño. *Proceedings of the National Academy of Sciences*, **113 (42)**, 11 732–11 737, doi:10.1073/pnas.1614533113.

Chen, N., and A. J. Majda, 2016b: Simple dynamical models capturing the key features of the Central Pacific El Niño. *Proceedings of the National Academy of Sciences*, **113 (42)**, 11 732–11 737, doi:10.1073/pnas.1614533113.

Cherkauer, K. A., 2003: Simulation of spatial variability in snow and frozen soil. *Journal of Geophysical Research*, **108 (D22)**, 8858, doi:10.1029/2003JD003575.

Chiang, J. C. H., and D. J. Vimont, 2004: Analogous Pacific and Atlantic Meridional Modes of Tropical Atmosphere – Ocean Variability. *Journal of Climate*, **17**, 4143–4158, doi:10.1175/JCLI4953.1.

Chowdhury, M. R., and P. S. Chu, 2015: Sea level forecasts and early-warning application expanding cooperation in the south pacific. *Bulletin of the American Meteorological Society*, **96 (3)**, 381–386, doi:10.1175/BAMS-D-14-00038.1.

Clayson, C. A., and D. Weitlich, 2007: Variability of tropical diurnal sea surface temperature. *Journal of Climate*, **20 (2)**, 334–352, doi:10.1175/JCLI3999.1.

Das, T., and Coauthors, 2009: Structure and Detectability of Trends in Hydrological Measures over the Western United States. *Journal of Hydrometeorology*, **10 (4)**, 871–892, doi:10.1175/2009JHM1095.1.

Davy, R., I. Esau, A. Chernokulsky, S. Outten, and S. Zilitinkevich, 2017: Diurnal asymmetry to the observed global warming. *International Journal of Climatology*, **37 (1)**, 79–93, doi:10.1002/joc.4688.

DeFlorio, M. J., D. W. Pierce, D. R. Cayan, and A. J. Miller, 2013: Western U.S. Extreme Precipitation Events and Their Relation to ENSO and PDO in CCSM4. *Journal of Climate*, **26 (12)**, 4231–4243, doi:10.1175/JCLI-D-12-00257.1.

DelSole, T., and M. K. Tippett, 2009: Average Predictability Time. Part I: Theory. *Journal of the Atmospheric Sciences*, **66 (5)**, 1172–1187, doi:10.1175/2008JAS2868.1.

Di Lorenzo, E., G. Liguori, N. Schneider, J. C. Furtado, B. T. Anderson, and M. A. Alexander, 2015: ENSO and meridional modes: A null hypothesis for Pacific climate variability. *Geophysical Research Letters*, **42**, 9440–9448, doi:10.1002/2015GL066281.

Di Lorenzo, E., and Coauthors, 2008: North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters*, **35 (8)**, L08 607, doi: 10.1029/2007GL032838.

Dias, D. F., D. Cayan, A. Gershunov, and A. J. Miller, 2020: The influence of sea surface temperature and soil moisture in seasonal predictions of air temperature over western north america. *Journal of Climate*, **under review**.

Dias, D. F., D. R. Cayan, and A. Gershunov, 2018: Statistical Prediction of Minimum and Maximum Air Temperature in California and Western North America. *California's Fourth Climate Change Assessment, California Energy Commission.*, doi:CCCA4-CEC-2018-011.

Ding, R., J. Li, F. Zheng, J. Feng, and D. Liu, 2016: Estimating the limit of decadal-scale climate predictability using observational data. *Climate Dynamics*, **46 (5-6)**, 1563–1580, doi:10.1007/s00382-015-2662-6.

Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. Rodrigues, 2013a: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, **4 (4)**, 245–268, doi:10.1002/wcc.217.

Doblas-Reyes, F. J., and Coauthors, 2013b: Initialized near-term regional climate change prediction. *Nature Communications*, **4**, 1–9, doi:10.1038/ncomms2704.

Duan, W., and C. Wei, 2013: The 'spring predictability barrier' for ENSO predictions and its possible mechanism: Results from a fully coupled model. *International Journal of Climatology*, **33 (5)**, 1280–1292, doi:10.1002/joc.3513.

Eccles, F., and E. Tziperman, 2004: Nonlinear effects on ENSO's period. *Journal of the Atmospheric Sciences*, **61 (4)**, 474–482, doi:10.1175/1520-0469.

Frankignoul, C., and K. Hasselmann, 1977: Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29 (4)**, 289–305, doi:10.3402/tellusa.v29i4.11362.

Gershunov, A., and T. P. Barnett, 1998: Interdecadal Modulation of ENSO Teleconnections. *Bulletin of the American Meteorological Society*, **79**, 2715–2725, doi:10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2.

Gershunov, A., T. P. Barnett, D. R. Cayan, T. Tubbs, and L. Goddard, 2000: Predicting and downscaling ENSO impacts on intraseasonal precipitation statistics in California: The 1997/98 event. *Journal of Hydrometeorology*, **1 (3)**, 201–210, doi:10.1175/1525-7541(2000)001<0201:PADEIO>2.0.CO;2.

Gershunov, A., and D. R. Cayan, 2003: Heavy daily precipitation frequency over the contiguous United States: Sources of climatic variability and seasonal predictability. *Journal of Climate*, **16 (16)**, 2752–2765, doi:10.1175/1520-0442(2003)016<2752:HDPFOT>2.0.CO;2.

Gershunov, A., D. R. Cayan, and S. F. Iacobellis, 2009: The Great 2006 Heat Wave over California and Nevada : Signal of an Increasing Trend. *Journal of Climate*, **22 (2008)**, 6181–6203, doi:10.1175/2009JCLI2465.1.

Gershunov, A., D. R. Cayan, and B. Retornaz, 2010: California heat waves with impacts on wine grapes. *The Ocean, the Wine, and the Valley: The Lives of Antoine Badan, EG Pavia, J. Sheinbaum and J. Candela (Eds.)*, 205–224.

Gershunov, A., and H. E. Douville, 2008: Extensive summer hot and cold extremes under current and possible future climatic conditions: Europe and north america. *Climate Extremes and Society*, **9780521870 (March)**, 74–98, doi:10.1017/CBO9780511535840.007.

Gershunov, A., T. Shulgina, M. Ralph, D. A. Lavers, and J. J. Rutz, 2017: Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters*, **(44)**, 7900–7908, doi:10.1002/2017GL074175.

Goddard, L., 2016: From science to service: Climate services are crucial for successful adaptation to current and future climate conditions. *Science*, **353 (6306)**, 1366–1367, doi:10.1126/science.aag3087.

Godfrey, J. S., R. a. Houze, R. H. Johnson, R. Lukas, and R. Weller, 1998: Coupled Ocean-Atmosphere Response Experiment (COARE): An interim report. *Journal of Geophysical Research*, **103 (C7)**, 14 395–14 450.

Grimm, A. M., and R. G. Tedeschi, 2009: ENSO and extreme rainfall events in South America. *Journal of Climate*, **22 (7)**, 1589–1609, doi:10.1175/2008JCLI2429.1.

Guirguis, K., A. Gershunov, R. Schwartz, and S. Bennett, 2011: Recent warm and cold daily winter temperature extremes in the Northern Hemisphere. *Geophysical Research Letters*, **38 (17)**, L17 701, doi:10.1029/2011GL048762.

Guirguis, K., A. Gershunov, T. Shulgina, R. E. S. Clemesha, and F. M. Ralph, 2019: Atmospheric rivers impacting Northern California and their modulation by a variable climate. *Climate Dynamics*, **52 (11)**, 6569–6583, doi:10.1007/s00382-018-4532-5.

Guirguis, K., A. Gershunov, A. Tardy, and R. Basu, 2014: The impact of recent heat waves on human health in California. *Journal of Applied Meteorology and Climatology*, **53 (1)**, 3–19, doi:10.1175/JAMC-D-13-0130.1.

Guzman-Morales, J., A. Gershunov, J. Theiss, H. Li, and D. Cayan, 2016: Santa Ana Winds of Southern California: Their climatology, extremes, and behavior spanning six and a half decades. *Geophysical Research Letters*, **43 (6)**, 2827–2834, doi:10.1002/2016GL067887.

Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573 (7775)**, 568–572, doi:10.1038/s41586-019-1559-7.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2007: Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the western United States. *Journal of Climate*, **20 (8)**, 1468–1486, doi:10.1175/JCLI4051.1.

Hartmann, D. L., 2015: Pacific sea surface temperature and the winter of 2014. *Geophysical Research Letters*, **42**, 1894–1902, doi:10.1002/2015GL063083.

Henley, B. J., J. Gergis, D. J. Karoly, S. Power, J. Kennedy, and C. K. Folland, 2015: A Tripole Index for the Interdecadal Pacific Oscillation. *Climate Dynamics*, **45 (11-12)**, 3077–3090, doi:10.1007/s00382-015-2525-1.

Hervieux, G., M. A. Alexander, C. A. Stock, M. G. Jacox, K. Pegion, E. Becker, F. Castruccio, and D. Tommasi, 2017: More reliable coastal SST forecasts from the North American multimodel ensemble. *Climate Dynamics*, **0 (0)**, 1–16, doi:10.1007/s00382-017-3652-7.

Hidalgo, H. G., and J. a. Dracup, 2003: ENSO and PDO Effects on Hydroclimatic Variations of the Upper Colorado River Basin. *Journal of Hydrometeorology*, **4 (1)**, 5–23, doi:10.1175/1525-7541(2003)004<0005:EAPEOH>2.0.CO;2.

Ho, C. K., E. Hawkins, L. Shaffrey, and F. M. Underwood, 2013: Statistical decadal predictions for sea surface temperatures: A benchmark for dynamical GCM predictions. *Climate Dynamics*, **41 (3-4)**, 917–935, doi:10.1007/s00382-012-1531-9.

Horel, J. D., and J. >. Wallace, 1981: Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation. *Monthly Weather Review*, **109**, 813–829.

Huang, J., H. Van Den Dool, and K. P. Georgakakos, 1996: Analysis of Model-Calculated Soil Moisture over the United States (1931 - 1993) and Applications to Long-Range Temperature Forecasts. *Journal of Climate*, **9**, 1350–1362.

Huddart, B., A. Subramanian, L. Zanna, and T. Palmer, 2017: Seasonal and decadal forecasts of Atlantic Sea surface temperatures using a linear inverse model. *Climate Dynamics*, **49 (5-6)**, 1833–1845, doi:10.1007/s00382-016-3375-1.

Hurrell, J. W., and Coauthors, 2013: The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, **94 (9)**, 1339–1360, doi:10.1175/BAMS-D-12-00121.1.

Jacox, M. G., M. A. Alexander, C. A. Stock, and G. Hervieux, 2017: On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. *Climate Dynamics*, **0 (0)**, 1–15, doi:10.1007/s00382-017-3608-y.

Jin, E. K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Climate Dynamics*, **31 (6)**, 647–664, doi:10.1007/s00382-008-0397-3.

Kao, H. Y., and J. Y. Yu, 2009: Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *Journal of Climate*, **22 (3)**, 615–632, doi:10.1175/2008JCLI2309.1.

Kapnick, S. B., and Coauthors, 2018: Potential for western US seasonal snowpack prediction. *Proceedings of the National Academy of Sciences of the United States of America*, **115 (6)**, 1180–1185, doi:10.1073/pnas.1716760115.

Karl, T. R., and Coauthors, 1993: A New Perspective on Recent Global Warming : Asymmetric Trends of Daily Maximum and Minimum Temperature. *Bulletin of the American Meteorological Society*, **74 (6)**, 1007–1024.

Kay, J. E., and Coauthors, 2015: The community earth system model (CESM) large ensemble project : A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, **96 (8)**, 1333–1349, doi:10.1175/BAMS-D-13-00255.1.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453 (7191)**, 84–88, doi:10.1038/nature06921.

Kirtman, B., S. B. Power, M. Kimoto, G. Vecchi, J. John, A. Slater, and F. Zwiers, 2013: Near-term Climate Change: Projections and Predictability. Climate Change 2013 - The Physical Science Basis. *Contribution to the Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ Press, Cambridge, UK, 953–1028, doi:10.1017/CBO9781107415324.023.

Kirtman, B. P., and Coauthors, 2014: The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, **95 (4)**, 585–601, doi:10.1175/BAMS-D-12-00050.1, `/dx.doi.org/10.1175/BAMS-D-12-00050.1`.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophysical Research Letters*, **33 (17)**, 2–5, doi:10.1029/2006GL026242.

Koster, R. D., M. J. Suarez, and M. Heiser, 2000: Variance and Predictability of Precipitation at Seasonal-to-Interannual Timescales. *Journal of Hydrometeorology*, **1**, 26–46, doi:10.1175/1525-754128200029001.

Kucharski, F., F. Molteni, and A. Bracco, 2006: Decadal interactions between the western tropical Pacific and the North Atlantic Oscillation. *Climate Dynamics*, **26 (1)**, 79–91, doi:10.1007/s00382-005-0085-5.

Kushnir, Y., and Coauthors, 2019: Towards operational predictions of the near-term climate. *Nature Climate Change*, **9 (2)**, 94–101, doi:10.1038/s41558-018-0359-7.

Lewis, S. C., and D. J. Karoly, 2013: Evaluation of historical diurnal temperature range trends in CMIP5 models. *Journal of Climate*, **26 (22)**, 9077–9089, doi:10.1175/JCLI-D-13-00032.1.

L'Heureux, M. L., M. K. Tipett, and A. G. Barnston, 2015: Characterizing enso coupled variability and its impact on north american seasonal precipitation and temperature. *Journal of Climate*, **28**, 4321–4245, doi:10.1175/JCLI-D-14-00508.1.

Li, S., L. Zhang, and L. Wu, 2017: Decadal potential predictability of upper ocean heat content over the twentieth century. *Climate Dynamics*, **49 (9-10)**, 3293–3307, doi:10.1007/s00382-016-3513-9.

Liang, X., D. P. Lettenmaier, E. F. Wood, and J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, **99 (D7)**, 14 415–14 428, doi:10.1029/94JD00483.

Linkin, M. E., and S. Nigam, 2008: The North Pacific Oscillation-West Pacific teleconnection pattern: Mature-phase structure and winter impacts. *Journal of Climate*, **21 (9)**, 1979–1997, doi:10.1175/2007JCLI2048.1.

Livneh, B., T. J. Bohn, D. W. Pierce, F. Munoz-arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive , hydrometeorological data set for Mexico , the U . S ., and Southern Canada 1950 – 2013. *Nature Scientific Data*, **2 (150042)**, 1–12, doi:10.1038/sdata.2015.42.

Lopez, H., and B. P. Kirtman, 2014: Journal of Geophysical Research : Atmospheres. *Journal of Geophysical Research: Atmospheres*, **119 (10)**, 114–138, doi: 10.1002/2014JD021908.

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences*, **20**, 130–141.

MacMynowski, D. G., and E. Tziperman, 2008: Factors Affecting ENSO's Period. *Journal of the Atmospheric Sciences*, **65 (5)**, 1570–1586, doi:10.1175/2007JAS2520.1.

Majda, A. J., C. Franzke, and D. Crommelin, 2009: Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences of the United States of America*, **106 (10)**, 3649–3653, doi:10.1073/pnas.0900173106.

Majda, a. J., I. Timofeyev, and Vanden Eijnden E, 1999: Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences of the United States of America*, **96 (26)**, 14 687–91, doi:10.1073/pnas.96.26.14687.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society*, **78 (January)**, 1069–1079.

Mcewen, G. F., 1914: Peculiarities of the California Climate. *Monthly Weather Review*, **42**, 14–25, doi:https://doi.org/10.1175/1520-0493.

McPhaden, M. J., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical Research Letters*, **30 (9)**, 1995–1998, doi: 10.1029/2003GL016872.

Meehl, G. A., and H. Teng, 2012: Case studies for initialized decadal hindcasts and predictions for the Pacific region. *Geophysical Research Letters*, **39 (22)**, 1–6, doi: 10.1029/2012GL053423.

Meehl, G. A., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, **90 (10)**, 1467–1485, doi:10.1175/2009BAMS2778.1.

Meehl, G. A., and Coauthors, 2014: Decadal climate prediction an update from the trenches. *Bulletin of the American Meteorological Society*, **95**, 243–267, doi:10.1175/BAMS-D-12-00241.1.

Mignot, J., J. García-Serrano, D. Swingedouw, A. Germe, S. Nguyen, P. Ortega, E. Guilyardi, and S. Ray, 2016: Decadal prediction skill in the ocean with surface nudging in the IPSL-CM5A-LR climate model. *Climate Dynamics*, **47 (3-4)**, 1225–1246, doi: 10.1007/s00382-015-2898-1.

Miller, A. J., and N. Schneider, 2000: Interdecadal climate regime dynamics in the North Pacific Ocean: Theories, observations and ecosystem impacts. *Progress in Oceanography*, **47 (2000)**, 355–379, doi:10.1016/S0079-6611(00)00044-6.

Mo, K. C., 2003: Ensemble canonical correlation prediction of surface temperature over the United States. *Journal of Climate*, **16 (11)**, 1665–1683, doi:10.1175/1520-0442(2003)016<1665:ECCPOS>2.0.CO;2.

Mueller, B., and S. I. Seneviratne, 2012: Hot days induced by precipitation deficits at the global scale. *Proceedings of the National Academy of Sciences*, **109 (31)**, 12 398–12 403, doi:10.1073/pnas.1204330109.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, 1998: ENSO Theory. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 261–14 290.

Newman, M., 2007: Interannual to Decadal Predictability of Tropical and North Pacific Sea Surface Temperatures. *Journal of Climate*, **20**, 2333–2356, doi:10.1175/JCLI4165.1.

Newman, M., 2012: An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies. *Journal of Climate*, **26**, 5260–5269, doi:10.1175/JCLI-D-12-00590.1.

Newman, M., 2013: An empirical benchmark for decadal forecasts of global surface temperature anomalies. *Journal of Climate*, **26 (14)**, 5260–5269, doi:10.1175/JCLI-D-12-00590.1.

Newman, M., M. A. Alexander, and J. D. Scott, 2011: An empirical model of tropical ocean dynamics. *Climate Dynamics*, doi:10.1007/s00382-011-1034-0.

Newman, M., G. P. Compo, and M. A. Alexander, 2003: ENSO-forced variability of the Pacific decadal oscillation. *Journal of Climate*, **16 (23)**, 3853–3857, doi:10.1175/1520-0442(2003)016<3853:EVOTPD>2.0.CO;2.

Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*, **44 (16)**, 8520–8529, doi:10.1002/2017GL074088.

Newman, M., P. D. Sardeshmukh, and C. Penland, 2009: How important is air-sea coupling in ENSO and MJO evolution? *Journal of Climate*, **22 (11)**, 2958–2977, doi:10.1175/2008JCLI2659.1.

Newman, M., and Coauthors, 2016: The Pacific Decadal Oscillation , Revisited. *Journal of Climate*, **29**, 4399–4427, doi:10.1175/JCLI-D-15-0508.1.

Ostro, B. D., L. A. Roth, R. S. Green, and R. Basu, 2009: Estimating the mortality effect of the July 2006 California heat wave. *Environmental Research*, **109 (5)**, 614–619, doi:10.1016/j.envres.2009.03.010.

Penland, C., 1989: Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. 2165–2185 pp., doi:10.1175/1520-0493(1989)117<2165:RFAFUP>2.0.CO;2.

Penland, C., and T. Magorian, 1993: Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, **6 (6)**, 1067–1076, doi:10.1175/1520-0442(1993)006<1067:PONSST>2.0.CO;2.

Penland, C., and L. Matrosova, 1994: A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation. *Journal of Climate*, **7 (9)**, 1352–1372, doi:10.1175/1520-0442(1994)007<1352:ABCFSN>2.0.CO;2.

Penland, C., and P. D. Sardeshmukh, 1995: The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, **8 (8)**, 1999–2024, doi:10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke, 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *Journal of Climate*, **22 (14)**, 3926–3938, doi:10.1175/2009JCLI2535.1.

Pohlmann, H., J. Kröger, R. J. Greatbatch, and W. A. Müller, 2017: Initialization shock in decadal hindcasts due to errors in wind stress over the tropical Pacific. *Climate Dynamics*, **49 (7-8)**, 2685–2693, doi:10.1007/s00382-016-3486-8.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, **15 (5)**, 319–324, doi:10.1007/s003820050284.

Qiu, B., 2003: Kuroshio Extension Variability and Forcing of the Pacific Decadal Oscillations: Responses and Potential Feedback. *Journal of Physical Oceanography*, **33 (12)**, 2465–2482, doi:10.1175/2459.1.

Quesada, B., R. Vautard, P. Yiou, M. Hirschi, and S. I. Seneviratne, 2012: Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nature Climate Change*, **2 (10)**, 736–741, doi:10.1038/nclimate1536.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, **108 (D14)**, 1–37, doi:10.1029/2002JD002670.

Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate system of period 65-70 years. *Nature*, **367 (6463)**, 532–8, doi:10.1038/350055a0, `nature.vol.342.30nov1989`.

Schneider, N., and B. D. Cornuelle, 2005: The forcing of the Pacific Decadal Oscillation. *Journal of Climate*, **18 (21)**, 4355–4373, doi:10.1175/JCLI3527.1.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture-climate interactions in a changing climate:

A review. *Earth-Science Reviews*, **99 (3-4)**, 125–161, doi:10.1016/j.earscirev.2010.02.004, `Web of Science`.

Shin, S. I., P. D. Sardeshmukh, and K. Pegion, 2010: Realism of local and remote feedbacks on tropical sea surface temperatures in climate models. *Journal of Geophysical Research Atmospheres*, **115 (21)**, 1–9, doi:10.1029/2010JD013927.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282 (5389)**, 728–731, doi:10.1126/science.282.5389.728.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model. *Science*, **317**, 796–799, doi:10.1126/science.1139540.

Smith, D. M., and Coauthors, 2013: Real-time multi-model decadal climate predictions. *Climate Dynamics*, **41**, 2875–2888, doi:10.1007/s00382-012-1600-0.

Smith, D. M., and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate and Atmospheric Science*, **2 (1)**, 1–10, doi:10.1038/s41612-019-0071-y.

Solomon, A., and Coauthors, 2011: Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction. *Bulletin of the American Meteorological Society*, **92 (2)**, 141–156, doi:10.1175/2010BAMS2962.1.

Sun, C., F. Kucharski, J. Li, F. F. Jin, I. S. Kang, and R. Ding, 2017: Western tropical Pacific multidecadal variability forced by the Atlantic multidecadal oscillation. *Nature Communications*, **8 (May)**, 1–10, doi:10.1038/ncomms15998.

Sun, D. Z., and Coauthors, 2006: Radiative and dynamical feedbacks over the equatorial cold tongue: Results from nine atmospheric GCMs. *Journal of Climate*, **19 (16)**, 4059–4074, doi:10.1175/JCLI3835.1.

Sutton, R. T., and D. L. Hodson, 2005: Ocean science: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309 (5731)**, 115–118, doi:10.1126/science.1109496.

Taguchi, B., S. P. Xie, N. Schneider, M. Nonaka, H. Sasaki, and Y. Sasai, 2007: Decadal variability of the Kuroshio Extension: Observations and an eddy-resolving model hindcast. *Journal of Climate*, **20 (11)**, 2357–2377, doi:10.1175/JCLI4142.1.

Tanahashi, S., 2003: Diurnal variations of sea surface temperature over the wide-ranging ocean using VISSR on board GMS. *Journal of Geophysical Research*, **108 (C7)**, 3216, doi:10.1029/2002JC001313.

Tang, B., W. W. Hsieh, A. H. Monahan, and F. T. Tangang, 2000: Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. *Journal of Climate*, **13 (1)**, 287–293, doi:10.1175/1520-0442(2000)013<0287:SCBNNA>2.0.CO;2.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106 (D7)**, 7183–7192.

Teng, H., G. A. Meehl, G. Branstator, S. Yeager, and A. Karspeck, 2017: Initialization Shock in CCSM4 Decadal Prediction Experiments. *Past Global Changes Magazine*, **25 (1)**, 41–46, doi:10.22498/pages.25.1.41.

Thomas, E. E., and D. J. Vimont, 2016: Modeling the mechanisms of linear and nonlinear ENSO responses to the pacific meridional mode. *Journal of Climate*, **29 (24)**, 8745–8761, doi:10.1175/JCLI-D-16-0090.1.

Thompson, D. W. J., and J. M. Wallace, 2001: Regional Climate Impacts of the Northern Hemisphere Annular Mode. *Science*, **293 (July)**, 85–90.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 291–14 324, doi:10.1029/97JC01444.

Trenberth, K. E., and J. W. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, **9**, 303–319.

Van den Dool, H., J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to us soil moisture over 1981–2001. *Journal of Geophysical Research: Atmospheres*, **108 (D16)**.

Van den Dool, H. M., 2007: *Empirical methods in short-term climate prediction*. Oxford University Press, 215 pp., doi:10.1029/2005GL023422, `arXiv:1011.1669v3`.

Vimont, D. J., M. Alexander, and A. Fontaine, 2009: Midlatitude excitation of tropical variability in the pacific: The role of thermodynamic coupling and seasonality. *Journal of Climate*, **22 (3)**, 518–534, doi:10.1175/2008JCLI2220.1.

Vimont, D. J., M. A. Alexander, and M. Newman, 2014: Optimal growth of Central and East Pacific ENSO events. *Geophysical Research Letters*, **41**, 4027–4034, doi:10.1002/2014GL059997.1.

Vimont, D. J., D. S. Battisti, and A. C. Hirst, 2001: Footprinting: A seasonal connection between the tropics and mid-latitudes. *Geophysical Research Letters*, **28 (20)**, 3923–3926, doi:10.1029/2001GL013435.

Vimont, D. J., J. M. Wallace, and D. S. Battisti, 2003: The seasonal footprinting mechanism in the Pacific: Implications for ENSO. *Journal of Climate*, **16 (16)**, 2668–2675, doi:10.1175/1520-0442(2003)016<2668:TSFMIT>2.0.CO;2.

Weatherhead, E., S. Gearheard, and R. G. Barry, 2010: Changes in weather persistence: insight from inuit knowledge. *Global Environmental Change*, **20 (3)**, 523–528.

WEBSTER, P. J., and S. YANG, 1992: Monsoon and Enso - Selectively Interactive Systems. *Quarterly Journal of the Royal Meteorological Society*, **118 (507, B)**, 877–926, doi:10.1002/qj.49711850705.

Wilks, D. S., 2008: Improved statistical seasonal forecasts using extended training data. *International Journal of Climatology*, **28**, 1589–1598, doi:10.1002/joc.1661.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Academic press.

Wing, G. Y. K., L. Sushama, and G. T. Diro, 2016: The intraannual variability of land-atmosphere coupling over North America in the Canadian Regional Climate Model (CRCM5). *Journal of Geophysical Research: Atmospheres*, **121 (23)**, 13–859, doi:10.1002/2016JD025423.

Xoplaki, E., J. F. González-Rouco, J. Luterbacher, and H. Wanner, 2003: Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs. *Climate Dynamics*, **20 (7-8)**, 723–739, doi:10.1007/s00382-003-0304-x.

Xue, Y., M. Chen, A. Kumar, Z. Z. Hu, and W. Wang, 2013: Prediction skill and bias of tropical pacific sea surface temperatures in the NCEP climate forecast system version 2. *Journal of Climate*, **26 (15)**, 5358–5378, doi:10.1175/JCLI-D-12-00600.1.

Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century north Atlantic Ocean heat content. *Journal of Climate*, **25 (15)**, 5173–5189, doi:10.1175/JCLI-D-11-00595.1.

Yeager, S. G., and Coauthors, 2018: Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*, **(September)**, BAMS–D–17–0098.1, doi:10.1175/BAMS-D-17-0098.1.

Yeh, S. W., B. P. Kirtman, J. S. Kug, W. Park, and M. Latif, 2011: Natural variability of the central Pacific El Niño event on multi-centennial timescales. *Geophysical Research Letters*, **38 (2)**, 2–6, doi:10.1029/2010GL045886.

Yu, J.-Y., H.-Y. Kao, and T. Lee, 2010: Subtropics-Related Interannual Sea Surface Temperature Variability in the Central Equatorial Pacific. *Journal of Climate*, **23 (11)**, 2869–2884, doi:10.1175/2010JCLI3171.1.

Yu, J. Y., and S. T. Kim, 2011: Relationships between extratropical sea level pressure variations and the central Pacific and eastern Pacific types of ENSO. *Journal of Climate*, **24 (3)**, 708–720, doi:10.1175/2010JCLI3688.1.

Zanna, L., 2012: Forecast skill and predictability of observed atlantic sea surface temperatures. *Journal of Climate*, **25 (14)**, 5047–5056, doi:10.1175/JCLI-D-11-00539.1.

Zhang, J., W.-c. Wang, and L. R. Leung, 2008: Contribution of land-atmosphere coupling to summer climate variability over the contiguous United States. *Journal of Geophysical Research*, **113 (August)**, 1–15, doi:10.1029/2008JD010136.

Zhang, L., T. L. Delworth, and L. Jia, 2017: Diagnosis of decadal predictability of Southern Ocean sea surface temperature in the GFDL CM2.1 Model. *Journal of Climate*, **30 (16)**, 6309–6328, doi:10.1175/JCLI-D-16-0537.1.

Zhang, R., R. Sutton, G. Danabasoglu, Y. O. Kwon, R. Marsh, S. G. Yeager, D. E. Amrhein, and C. M. Little, 2019: A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts. *Reviews of Geophysics*, **57 (2)**, 316–375, doi:10.1029/2019RG000644.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900-93. *Journal of Climate*, **10 (5)**, 1004–1020, doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zhao, M., H. H. Hendon, O. Alves, G. Liu, and G. Wang, 2016: Weakened Eastern Pacific El Nï¿½o predictability in the early Twenty-First Century. *Journal of Climate*, **29 (18)**, 6805–6822, doi:10.1175/JCLI-D-15-0876.1.

Appendix A

Supplementary information for Chapter 3

Table A.1: Optimum number of Principal Components (PCs) and Canonical Correlates (CCs) for all combinations of predictors and predictands

SST–Tmin

	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF	JFM
PCs	4	7	5	5	7	14	7	12	15	17	10	7
CCs	4	4	5	5	4	11	3	5	14	7	9	4

SST–Tmax

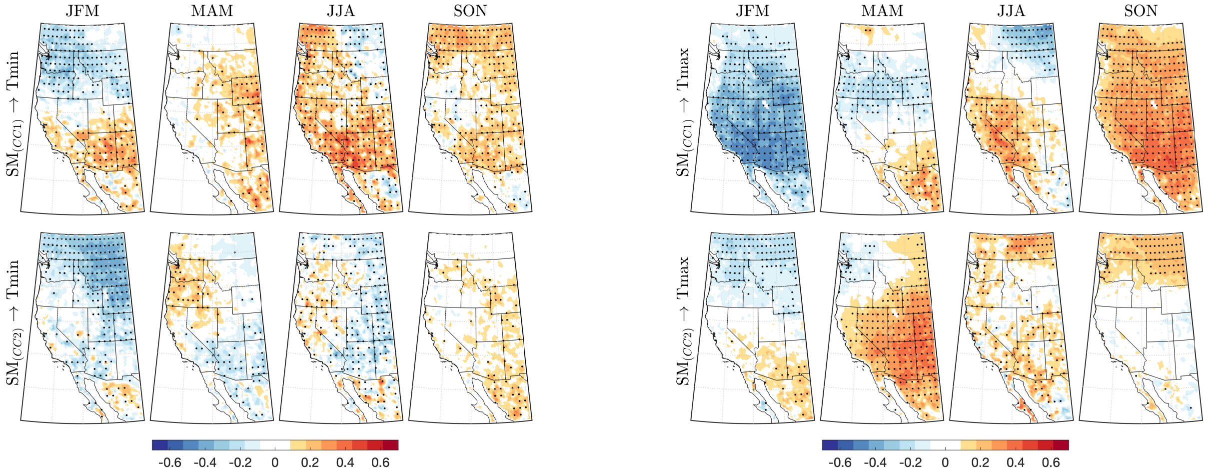
	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF	JFM
PCs	6	6	11	7	9	13	9	15	11	18	9	7
CCs	4	4	8	7	8	5	3	11	9	10	4	6

SM–Tmin

	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF	JFM
PCs	13	8	12	13	10	7	6	3	2	9	12	7
CCs	6	4	5	11	6	5	5	3	2	7	5	2

SM–Tmax

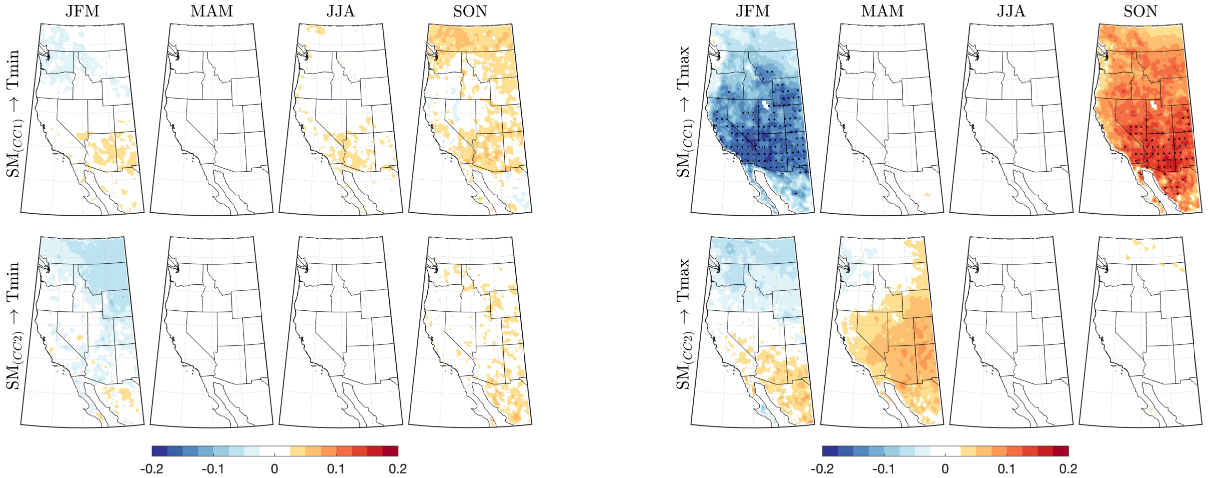
	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF	JFM
PCs	8	5	14	12	10	5	16	5	15	16	12	3
CCs	3	4	11	6	8	3	8	2	7	6	4	2



(a) Tmin

(b) Tmax

Figure A.1: First (upper panels) and second(lower panels) soil moisture canonical modes for selected seasons (JFM, MAM, JJA and SON) correlated with (a) Tmin and (b) Tmax time series. Black dots indicate values that have a statistical significance greater than 99%



(a) Tmin

(b) Tmax

Figure A.2: Same as Figure A.1, but with the the projected value weighted by the variance explained by each of soil moisture canonical mode.

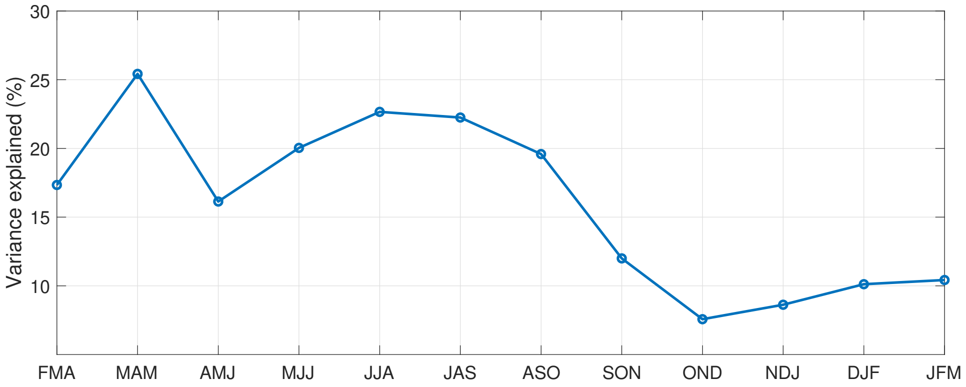


Figure A.3: Variance explained by the leading mode of the CCA model using SST as a predictor and Tmin as a predictand, for the 12 seasons analyzed

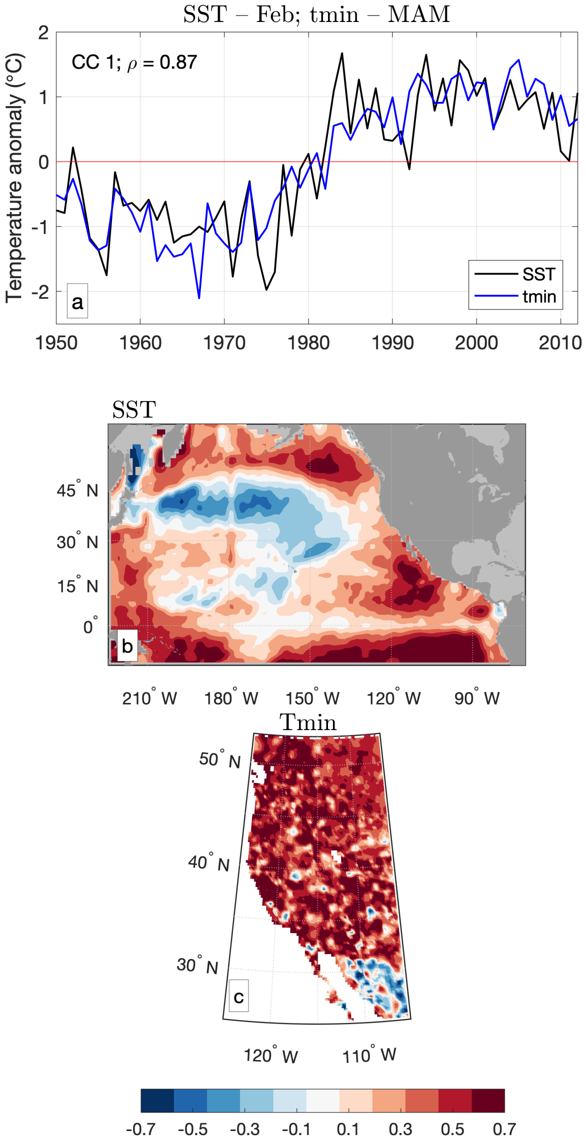


Figure A.4: Leading canonical mode for the SST-Tmin model for MAM. (a) Time series related to this leading canonical mode. The value ρ corresponds to the correlation between these two time series; (b) corresponding spatial pattern for the predictor (SST) and (c) for the predictand (Tmin).

Appendix B

Supplementary information for Chapter 4

Table B.1: The NMME models whose predictive skill was analyzed in this study. Those models were selected based on the availability of hindcasts for Tmin and Tmax.

Model name	Modeling Center	Number of ensembles
CanCM3	Third Generation Canadian Coupled Global Climate Model	10
CanCM4	Fourth Generation Canadian Coupled Global Climate Model	10
CESM1	National Center for Atmospheric Research (NCAR), Community Earth System Model, Version 1	10
GEOSS2S	Goddard Earth Observing System, Subseasonal to Seasonal Project, Version 2	4

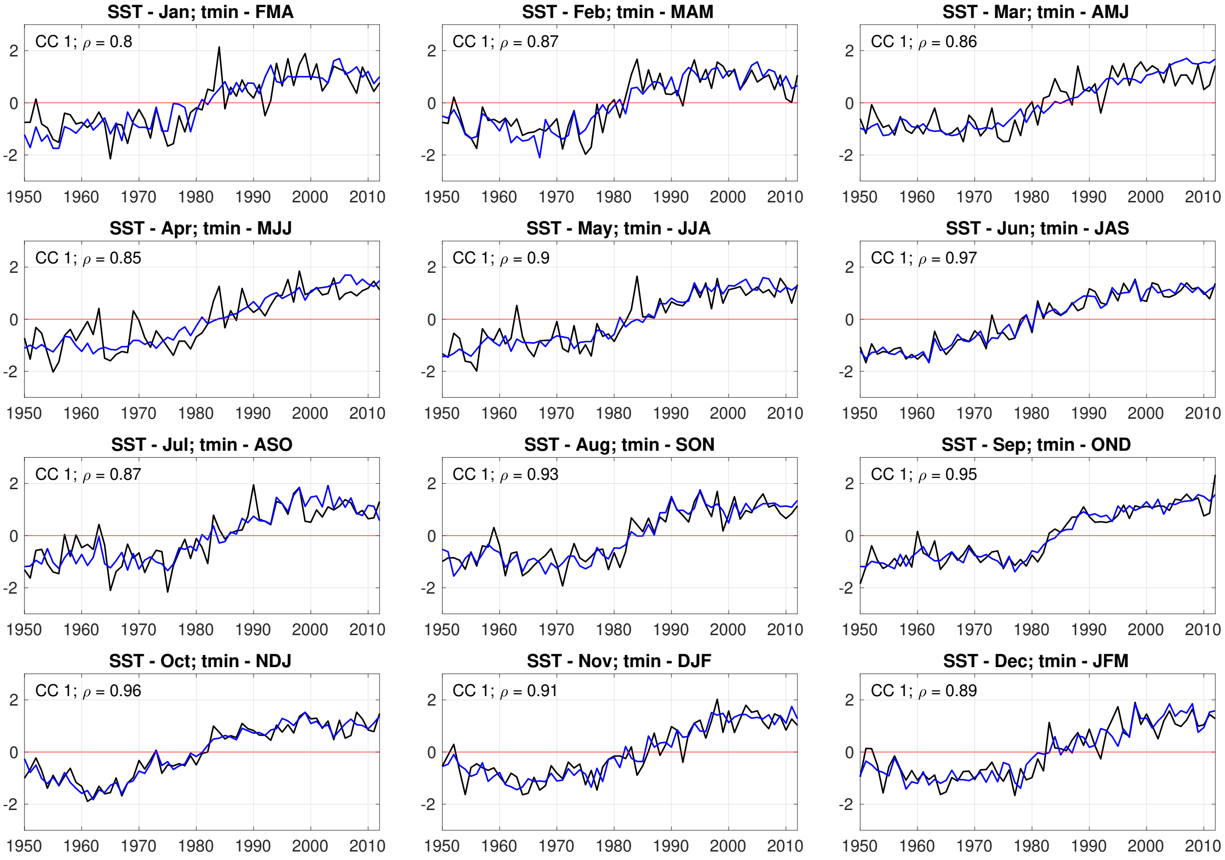


Figure B.1: Time series of the leading canonical mode for the SST-Tmin model for the 12 seasons analyzed. The value ρ corresponds to the correlation between the time series of the predictor (SST, in blue) and the predictand (Tmin, in black).

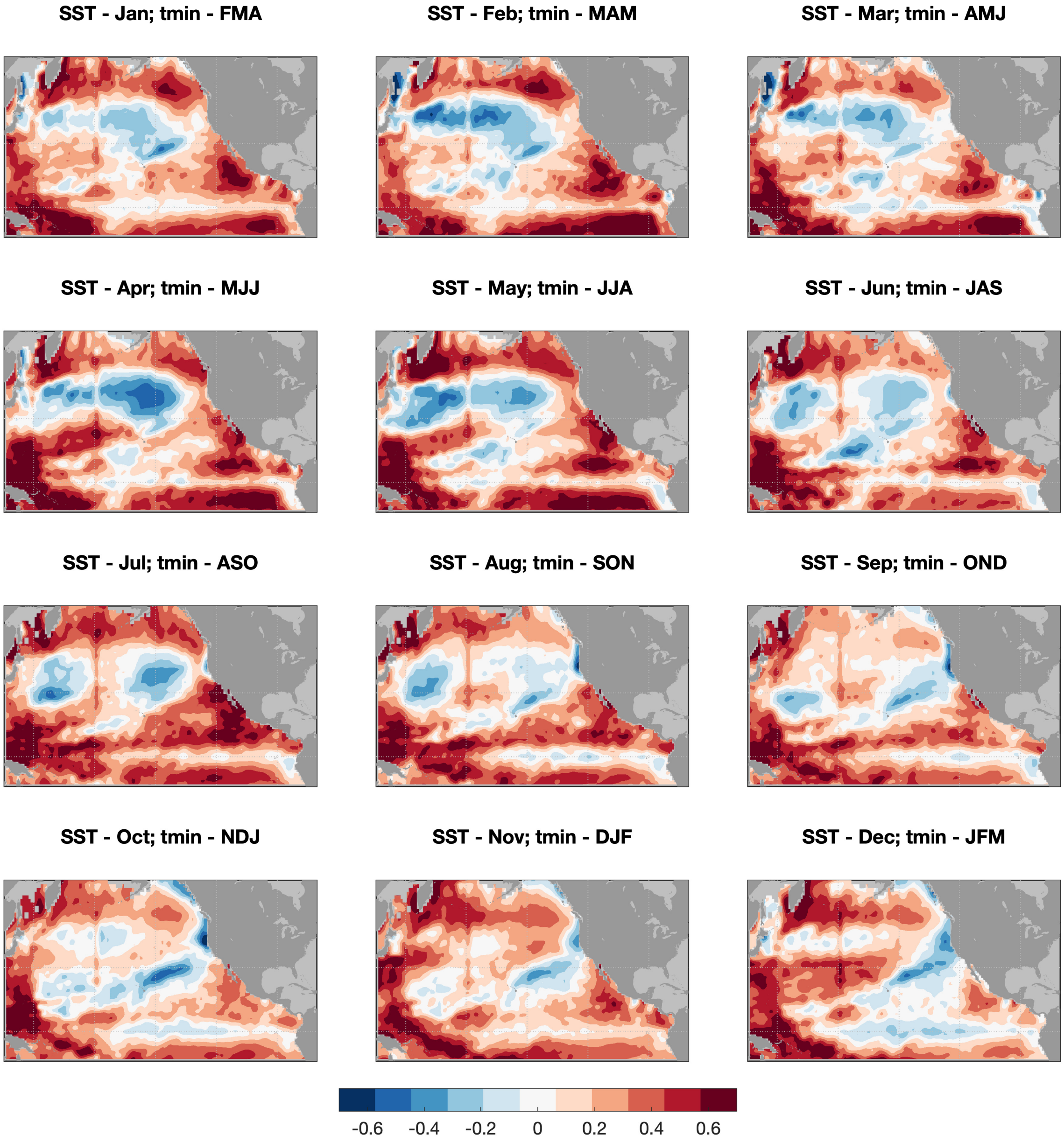
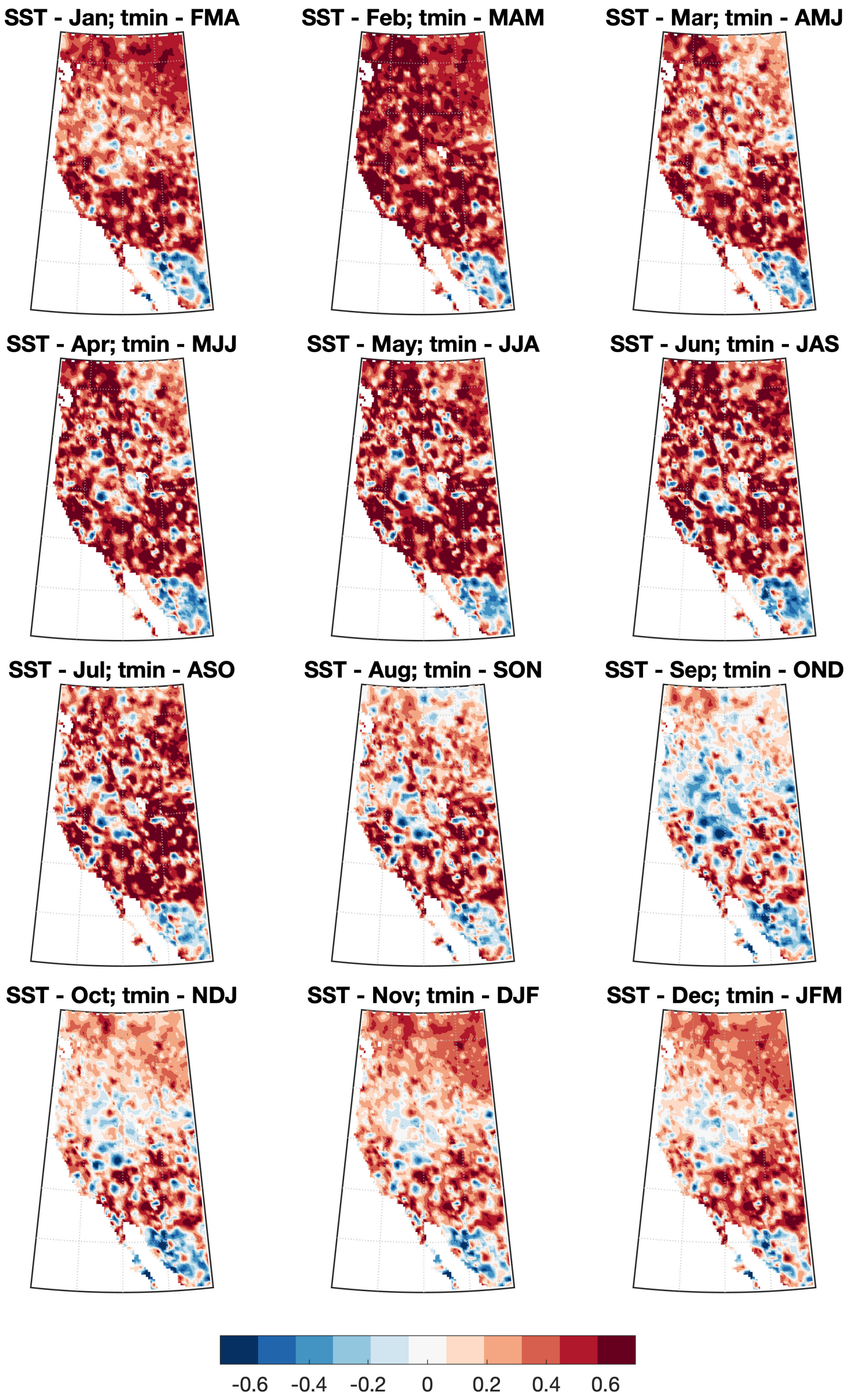


Figure B.2: Predictor (SST) spatial pattern of the leading canonical mode for the SST-Tmin model for the 12 seasons analyzed.



SST - Jan; tmin - FMA

SST - Feb; tmin - MAM

SST - Mar; tmin - AMJ

SST - Apr; tmin - MJJ

SST - May; tmin - JJA

SST - Jun; tmin - JAS

SST - Jul; tmin - ASO

SST - Aug; tmin - SON

SST - Sep; tmin - OND

SST - Oct; tmin - NDJ

SST - Nov; tmin - DJF

SST - Dec; tmin - JFM

Figure B.3: Predictand (Tmin) spatial pattern of the leading canonical mode for the SST-Tmin model for the 12 seasons analyzed.

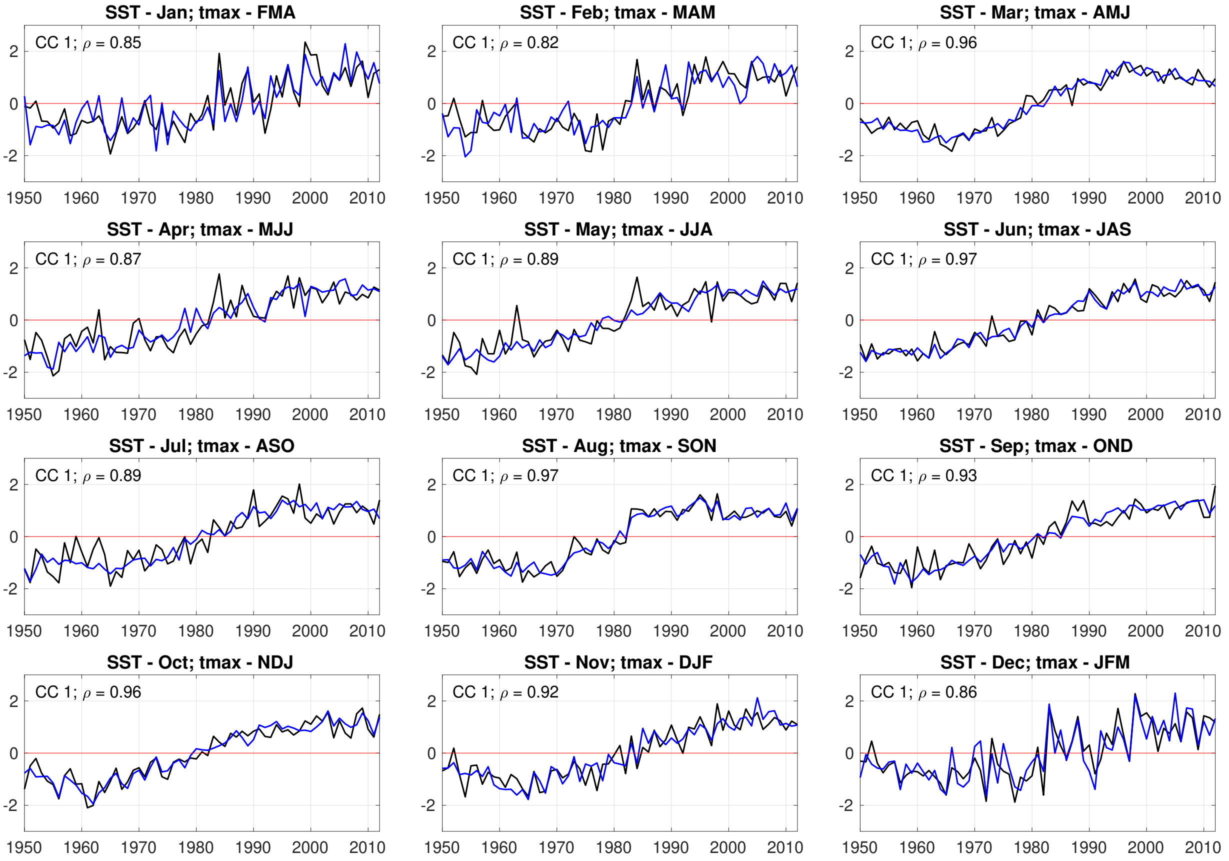
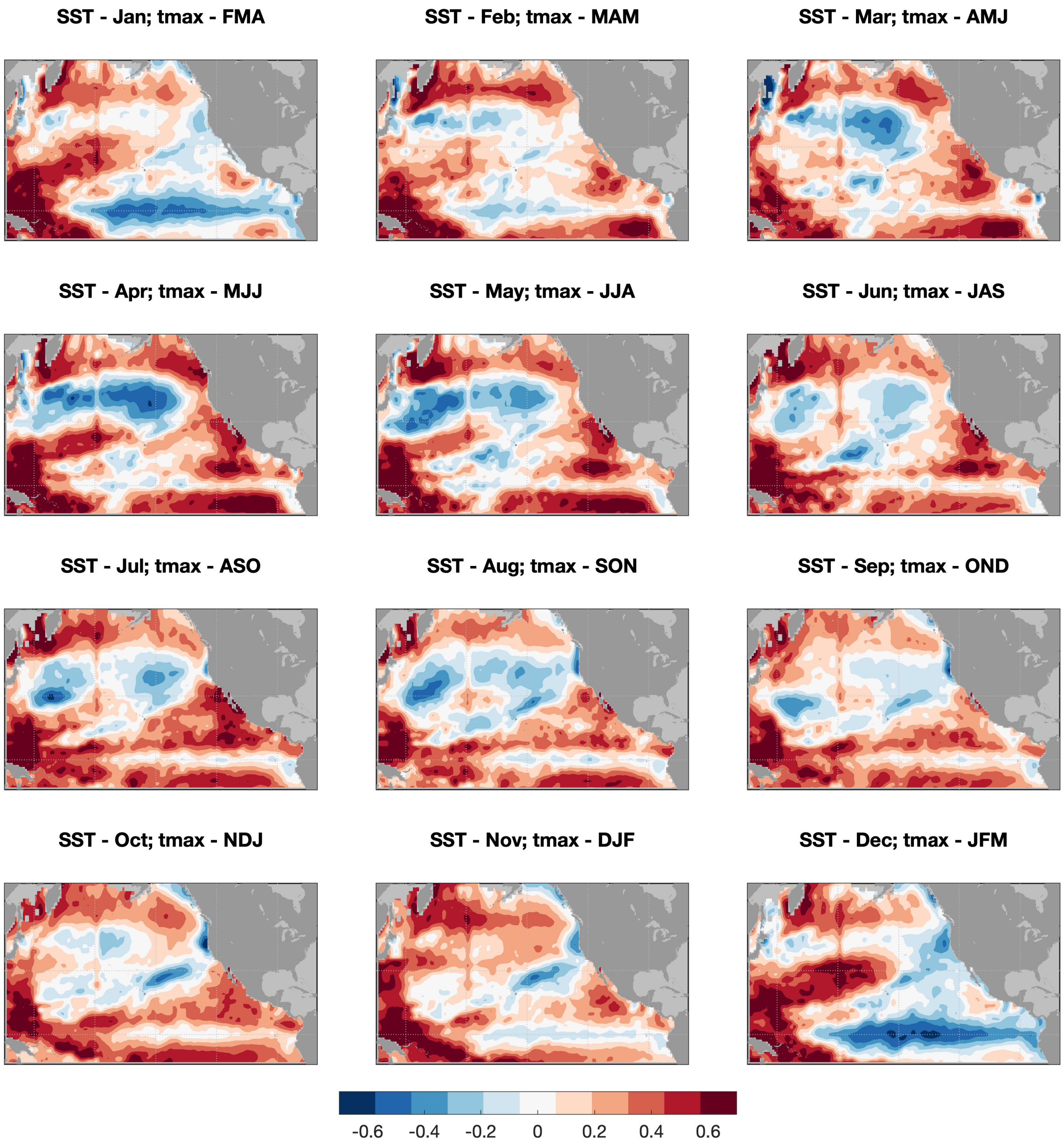


Figure B.4: Same as Figure B.1, but with Tmax as predictand



SST - Jan; tmax - FMA

SST - Feb; tmax - MAM

SST - Mar; tmax - AMJ

SST - Apr; tmax - MJJ

SST - May; tmax - JJA

SST - Jun; tmax - JAS

SST - Jul; tmax - ASO

SST - Aug; tmax - SON

SST - Sep; tmax - OND

SST - Oct; tmax - NDJ

SST - Nov; tmax - DJF

SST - Dec; tmax - JFM

Figure B.5: Same as Figure B.2, but with Tmax as predictand

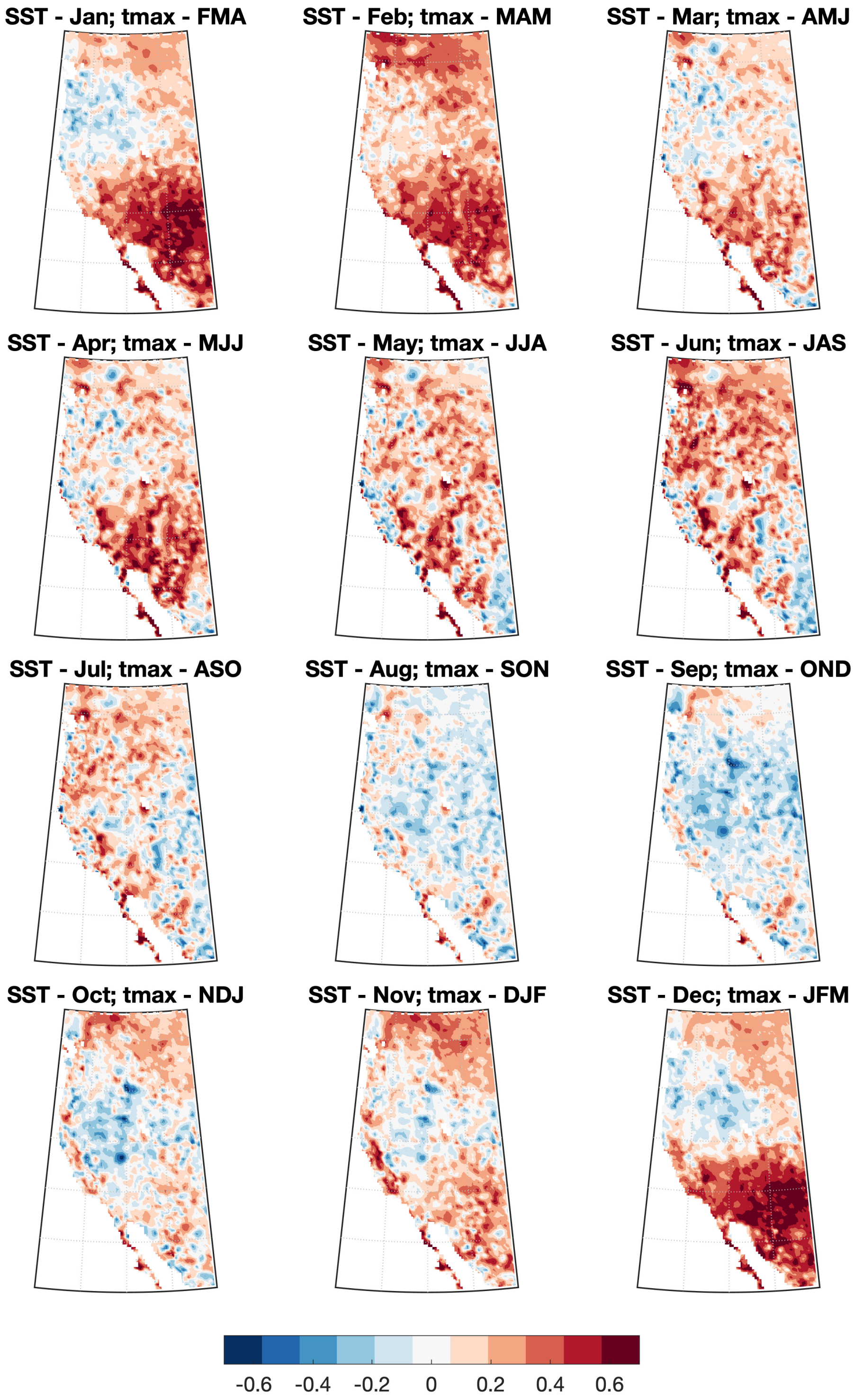


Figure B.6: Same as Figure B.3, but with Tmax as predictand