Improving Subseasonal to Seasonal Rainfall Forecasts in Central America Using Dynamic Model Ensembles

UNIVERSITY OF OXFORD

Katherine McCarthy Kowal

School of Geography and the Environment
Hertford College
University of Oxford

Supervised by
Dr. Louise Slater, University of Oxford
Dr. Anne Van Loon, Vrije Universiteit Amsterdam

Submitted thesis for the degree Doctor of Philosophy
Oxford, United Kingdom
August, 2023

Acknowledgements

This thesis would not have been possible without my supervisors Louise Slater and Anne Van Loon. Thank you both for your unwavering support throughout this journey. Completing this dissertation during the covid pandemic could have been an isolating experience. Instead, I felt connected and engaged in large part due to your efforts. You both have built thriving research communities around you, and I feel honoured to have been included in them. Your insights kept me pushing farther, and I took more risks to learn new skills because of your encouragement.

A huge thank you to my co-authors. Collaborating with you has been one of the highlights of my thesis, and our discussions always picked up my spirits when I was feeling stuck. Without Christian Birkel, I also never would have made it to the University of Costa Rica (UCR) to complete a research exchange that deepened my understanding of the regional climate in Central America. Thank you for welcoming me into your community and sharing your time and insights. I am so grateful to your colleagues at UCR who also shared their knowledge with me when I was puzzling over how best to develop the subsampling methodology.

Thank you to the Rhodes Trust and Hertford College for funding me during my time in Oxford. These funds not only made it possible for me to complete my thesis on time, but also supported me to present my work at conferences and complete the research exchange at UCR.

This work would also not have been possible without my friends and family. Thank you for the laughter, the many meals, and the heartfelt hugs when I needed a pick-me-up. Oxford would not be home to me without the Australian Rules Football Club. Only you could have gotten me out of bed on those dark winter mornings to run around in the mud every week. A special thank you for my parents. My champions from day one who taught me the beauty of being silly, the power to stay curious in uncertainty, and the daring to risk failures over regrets.

Abstract

Hydrometeorological hazards such as droughts and floods can have devastating consequences. Central America is one region at risk of hydrometeorological extremes impacts, which will likely increase under human-induced climate change. Physically-based subseasonal to seasonal (S2S) rainfall forecasts from Atmospheric Oceanic General Circulation Models (AOGCMs) can be used to inform anticipatory actions from multiple weeks to several months ahead. Multiple AOGCMs can be combined into multi-model ensembles (MMEs) to generate ensemble forecasts, which often have higher skill than individual models. There are two leading MMEs that provide publicly available forecasts: the North American Multimodel Ensemble (NMME) and the European multi-model seasonal prediction system (C3S). The skill of these models' rainfall forecasts is critical. Uncertainties in AOGCM estimates affect preparedness planning, as they are used to drive regional rainfall and hydrological models. Rainfall forecasts from individual AOGCMs and MMEs need more evaluation over Central America at the S2S scale. Relatively few evaluations have compared multiple AOGCMs over the region, making it difficult for stakeholders to choose which models to use and to know when to trust the predictions. More clarity is needed on the performance of S2S rainfall forecasts spatially and temporally, and on optimal postprocessing techniques to enhance the detection of low and high rainfall extremes.

This thesis conducts a comparative assessment of S2S rainfall forecasts across ten AOGCMs that contribute to the NMME and C3S, using statistical and process-based evaluation techniques. AOGCMs are found to generally have higher skill in the late wet season (September and October) compared to the early wet season (May and June), possibly due to the increased strength of the El Niño Southern Oscillation (ENSO) teleconnection in the late wet season. The models often perform better on the Pacific side of the region, which experiences a more distinct wet season compared to the Caribbean. Low and high rainfall extremes are challenging to predict even in the late wet season when the ENSO teleconnection is stronger. Techniques to optimize AOGCM outputs for S2S rainfall prediction are then assessed, including using hybrid forecasts and subsampling. Hybrid forecasting methods use AOGCMs to generate forecasts by statistically relating their predictions of large-scale drivers, such as sea surface temperatures, to rainfall. Findings show some hybrid methods improve AOGCM forecasts at the seasonal scale, such as those using forecasts of Tropical North Atlantic (TNA) sea surface temperatures in the early wet season when TNA is more strongly associated with regional rainfall. A novel post-processing technique is also developed that improves subseasonal detection rates of low and high rainfall extremes by subsampling ensemble members that better represent key drivers of Central American rainfall. The post-processing technique is beneficial for operational forecasters who can leverage their expertise of relevant rainfall-generating processes to subsample better performing ensemble members for their region. These findings underscore the importance of evaluating AOGCM rainfall forecasts regionally due to the spatial and temporal variability in AOGCM skill and showcase how alternative ways to post-process or use these models can significantly improve S2S forecasts over Central America.

Contents

Abbreviations

AOGCM / GCM – Atmospheric Oceanic General Circulation Model / General Circulation Model

CADC – Central American Dry Corridor

CAPE – Convective Available Potential Energy

CCA – Canonical Correlation Analysis

CEPAL – Comisión Económica Para América Latina y El Caribe

CHIRPS - Climate Hazards Group InfraRed Precipitation with Station data

CLLJ – Caribbean Low Level Jet

CJ / CHOCO – Chorro del Occidente Colombiano Jet

CMCC – Centro Euro-Mediterraneo Sui Cambiamenti Climatici

COLA – Center for Ocean Land Atmosphere Studies

CRPS – Continuous Ranked Probability Score

CRPSS – Continuous Ranked Probability Skill Score

DWD – Deutscher Wetterdienst

ECCC – Environment and Climate Change Canada – Canadian Meteorological Center

ECMWF – European Centre for Medium-Range Weather Forecasts

GFDL – Geophysical Fluid Dynamics Laboratory

GMAO – Global Modelling and Assimilation Office

GPCC – Global Precipitation Climatology Centre

GPH – Geopotential Height

HSS – Heidke Skill Score

IICACR – Inter-American Institute for Cooperation on Agriculture Costa Rica

IMN – Instituto Meteorológico Nacional

INSIVUMEH – Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología

IRI – International Research Institute for Climate and Society at Columbia University

ITCZ – Intertropical Convergence Zone

LLJ – Low Level Jet

mASL – meters Above Sea Level

ML – Machine Learning

MME – Multi-model Ensemble

MSD – Mid-Summer Dry Period, sometimes referred to as Mid-Summer Drought

MSE – Mean Squared Error

MSE-SS – Mean Squared Error Skill Score

MSWEP – Multi-Source Multi-Weighted Ensemble Precipitation

NASA – National Aeronautics and Space Agency

NASH – North Atlantic Subtropical High

NCAR – National Center for Atmospheric Research

NCEP – National Center for Environmental Prediction

NMME – North American Multimodel Ensemble

NOAA – National Oceanic and Atmospheric Administration

NWP – Numerical Weather Prediction

O2R – Operations to Research

R2O – Research to Operations

SLP – Sea Level Pressure

SST – Sea Surface Temperature

S2S – Subseasonal to Seasonal

UCR – Universidad de Costa Rica

UKMO – UK Met Office

WHWP – Western Hemisphere Warm Pool

WMO – World Meteorological Organization

2AFC – Two Alternative Forced Choice Score

CHAPTER 1

Introduction

Planning for weather is a pressing issue, especially as extreme weather continues to threaten the safety and livelihoods of people across the world. Disasters have increased fivefold in the last fifty years (WMO, 2023), making disaster preparedness a more widespread challenge. Central America is one region that warrants significant attention, as it faces heightened vulnerability to extreme weather, which will likely increase with human-induced climate change (e.g. Giorgi, 2006; Taylor et al., 2012; Alfaro-Córdoba, Hidalgo and Alfaro, 2020; Almazroui et al., 2021). Early warning systems can support anticipatory action when they are both effectively communicated and provide reliable information (Domeisen et al., 2022; White et al., 2022). Dynamic model forecasts are generated using physically-based models that use numerical weather prediction to estimate future states of the earth system (Doblas-Reyes et al., 2013). This work focuses on the technical quality of these models (and their combined ensembles) to address some challenges in forecasting rainfall at the subseasonal to seasonal forecast scale (S2S – multiple weeks to several months ahead).

Three aims are addressed here: (1) further the understanding of how dynamic model-based ensembles of S2S rainfall forecasts compare spatially and temporally over Central America, including how well the models detect low and high rainfall extremes; (2) provide more insight into why model performance may vary by considering how key processes that drive regional rainfall relate to the models' rainfall predictions; and (3) identify opportunities to optimize regional rainfall forecasts from these models by investigating novel ways to use the model outputs to generate skilful rainfall forecasts (e.g. through hybrid forecasting methods and alternative post-processing techniques). These aims are addressed over the course of the three studies developed as part of this work.

1.1 Background and Motivation

Central America is vulnerable to the effects of climate change. To date, changes in Central America's climate have primarily been observed as increases in temperature, potential evapotranspiration, and the rate of precipitation extremes (Aguilar et al., 2005; Taylor et al., 2012; Anderson et al., 2019; Alfaro-Córdoba, Hidalgo and Alfaro, 2020). Reports include increased frequency of warmer days and nights on the Caribbean side of Central America (Peterson et al., 2002; Stephenson et al., 2014; Jones et al., 2016), and a significant increase in the duration of the mid-summer dry period (Anderson et al., 2019).

1.1 | Background and Motivation

Several studies have projected future declines in regional water availability across multiple time-scales over the 21st century (Giorgi, 2006; Rauscher et al., 2008; Maurer, Adam and Wood 2009; Hidalgo et al. 2013; Hannah et al., 2017; Imbach et al., 2018; Cook et al., 2020; Corrales-Suastegui, Fuentes-Franco and Pavia, 2020; Almazroui et al., 2021; Vichot-Llano et al., 2021). Giorgi (2006), for instance, identified Central America as the main tropical hotspot using the Regional Climate Change Index for the 2080-2099 period, primarily due to decreases in rainfall and increases in rainfall variability. Maurer, Adam and Wood (2009) also showed decreases in potential rainfall and reservoir levels in the largest river basin (Rio Lempa) in Central America. Nearer-term projections align with these trends. Imbach et al. (2018), for instance, also found that rainfall is likely reduced during the wet season within the 2021-2050 time-period.

Changes in rainfall can significantly affect critical infrastructure services within the region. Key impacts include hydropower (e.g. Meza, 2014; Ng et al., 2017), agriculture (Marengo et al., 2014; FAO, 2015b; Hannah et al., 2017; Stewart et al., 2022), and water and waste-water services, which become increasingly difficult to provide when water availability decreases or water quality is compromised through contamination (Rangel Soares et al., 2002; Bundschuh et al., 2010; Litter, Morgada and Bundschuh, 2010). In many Central-American catchments, impacts are closely tied with rainfall extremes. For instance, there is a relatively close relationship between rainfall deficits and hydrological droughts because groundwater processes do not dominate (Quesada-Montano et al., 2019). In the Central American Dry Corridor (CADC), a subregion that is particularly prone to drought (Gotlieb et al., 2019; Hidalgo et al., 2019), many smallholder livelihoods depend on rain-fed agriculture (e.g. Lopez-Ridaura et al. 2019), and has faced significant increases in drought frequency driven by decreases in rainfall (Stewart et al., 2022).

Early warnings based on forecasting provide one way to reduce vulnerability (Domeisen et al., 2022; Goddard et al., 2020; White et al., 2022). Climate services encompass a broad range of tools for climate preparedness, including communication of these warnings to decision-makers to enhance risk management for day-to-day variability and extreme weather (Vaughan and Dessai, 2014). S2S climate information can inform many aspects of society (Zebiak et al., 2015), including agriculture (de Sousa et al., 2018; Fernandes et al., 2020; Hansen et al., 2006; Pons et al., 2021), health (DiSera et al., 2020; Muñoz et al., 2017; 2020), and disaster mitigation (Coughlan de Perez et al., 2015; Merz et al., 2020). S2S predictions of rainfall can support planning for a variety of disasters, such as floods (Braman et al., 2013; Coughlan de Perez et al., 2016, 2017; Golding et al., 2019; Kreibich et al., 2017a; 2017b), droughts (Cook, 2016; Lange & Cook, 2015; Wilhite, 2016), and fires (Fernandes, Bell and Muñoz, 2022). Rainfall forecasts are one of many tools needed to enhance resilience. Drought and flooding impacts, for instance, are modulated by several factors, including other physical variables (e.g. soil moisture, temperature), and social and economic conditions (Alfaro Martínez et al., 2018; Kuzdas et al., 2015; Peréz-Briceño et al., 2016; Van Loon, 2015).

Several forecasting methods exist, including dynamic, statistical, and hybrid methods that combine the two (Hao, Singh and Xia, 2018; Troin et al., 2021). Dynamic methods use

physically based models of the earth system, such as coupled atmospheric and oceanic general circulation models (AOGCMs). Statistical methods find patterns using observed variables to make predictions (e.g. sea surface temperatures and rainfall), and hybrid methods combine outputs from AOGCMs with statistical methods to generate forecasts (Slater et al., 2022). Many AOGCMs are operationally deployed today in Central America. CFSv2 (an AOGCM produced by NOAA; Saha et al., 2014), for instance, currently provides inputs for the Weather Research and Forecasting (WRF) model, which is commonly deployed for seasonal rainfall forecasts by multiple countries in Central America (WRF Clima - https://wrf-clima.imn.ac.cr).

AOGCM outputs can also be combined into multi-model ensembles (MMEs) to generate forecasts. The comparative advantage of using MMEs compared to single model estimates has been demonstrated in several cases (e.g. Elvidge et al., 2023; Palmer et al., 2004; Wang et al., 2009). Two leading MMEs that produce S2S forecasts include the North American Multimodel Ensemble (NMME – Becker et al., 2020), and C3S, the European MME (Marsh and Penebad, 2016). These ensembles provide inputs into the seasonal rainfall outlooks produced by the Central American Climate Outlook Forum (CA-COF), an international institution that supports sector-specific and national decision-making across Central America (Alfaro et al., 2016; Donoso & Ramirez, 2001; Garcia-Solera & Ramirez, 2012).

Forecast evaluation sits within the research to operations (R2O) lifecycle (e.g. Buizza et al., 2017). Evaluation is a central part of R2O to ensure forecasts provide reliable information for decision-makers. Some evaluations of AOGCMs and combined MMEs have showcased their potential to provide rainfall and temperature forecasts over Central America and in nearby regions (e.g. Becker et al., 2020; Carrão et al., 2018; Gubler et al., 2020; Khajehei et al., 2018; Kirtman et al., 2014; Slater et al., 2016; Weisheimer & Palmer, 2014). A few studies have targeted subregions within Central America, evaluating dynamic forecasts of rainfall at lead times outside of the S2S scale (Hidalgo & Alfaro, 2012, 2015; Maldonado, Alfaro, & Hidalgo, 2018; Maldonado et al., 2018). There is a lack in evaluations, however, that compare S2S forecasts across multiple AOGCMs over Central America, which makes it more difficult to choose between models or know when and where to trust their S2S predictions. Furthermore, subregions like the CADC need more targeted forecast verification to support preparedness efforts.

Additionally, more analysis is needed on the underlying drivers of regional forecast skill. Enhanced understanding of the connections between regional climate mechanisms and AOGCM forecast skill can inform why some forecasts perform better than others and identify possibilities to improve performance. For instance, although the El Niño Southern Oscillation (ENSO) is a dominant climate mechanism affecting rainfall over Central America (Durán-Quesada, Gimeno and Amador, 2017), less than half of the drought-related impacts in the CADC occurred during an El Niño event in some historical analyses (Muñoz-Jiménez et al., 2019). As ENSO is also a main driver of AOGCM performance in the eastern Pacific near Central America (Scaife et al. 2014), further investigation of how AOGCM forecasts differ from relying on statistical forecasts that use ENSO alone to make predictions would be useful to establish their comparative value.

Besides ENSO, other processes are critical to rainfall variability over the region (Anderson et al., 2023; Durán-Quesada et al., 2020), and several large-scale processes and features modulate the strength of trade wind patterns and subsequent rainfall, including SSTs in the Atlantic Ocean (Taylor et al., 2002; Maldonado et al., 2017). The difference between SSTs in the two oceans is significant, as the magnitude of their gradient affects moisture transport across the region (Enfield & Mayer, 1997; Spence et al., 2004; Waylen & Quesada, 2001). Testing how models capture different phenomena that are important to rainfall highlight mechanisms that are important to represent in future model development.

A variety of approaches can also be deployed to post-process raw dynamic ensemble outputs and optimize their forecasts (e.g. Doblas-Reyes et al., 2005; Hemri et al., 2020; Manzanas et al., 2019). The Next-Gen operational approach to seasonal ensemble-based forecasting, for instance, uses canonical correlation analysis (CCA) to statistically calibrate models before combining them into an MME mean estimate (e.g. Fernandes et al., 2020; Muñoz et al., 2020b; Pons et al., 2021). Sophisticated combination techniques have also been developed because MME skill does not always significantly differ from using a calibrated single model, as the errors of contributing AOGCMs in an MME are often correlated (Weigel et al., 2009). To improve on the unweighted MME mean, some studies at the decadal to climate projection scale have combined AOGCMs using weighted means (e.g. Brunner et al., 2020; Herger et al., 2018; Knutti et al., 2017; Sanderson et al., 2017) to account for overlaps in structural components (Knutti, 2010). Some recent studies have subsampled mean estimates based on model member representation of key physical processes. For instance, a few groups have subsampled ensembles based on the ensemble members' abilities to represent the North Atlantic Oscillation (NAO; Hurrell et al. 2003) to improve temperature and precipitation forecasts in Europe (Dobrynin et al., 2018, 2022; Dusterhus, 2020; Smith et al., 2020).

Another way to potentially improve AOGCM forecasts is through using hybrid forecasting approaches to extract AOGCM forecasts of predictor variables (e.g. sea surface temperatures - SSTs) in zones that drive regional rainfall and statistically relate these outputs to rainfall estimates. Hybrid forecasting could potentially improve AOGCM rainfall estimates because these ensembles are more skilful at forecasting large-scale variability (e.g. oceanic variables like SSTs) than rainfall (Barnston et al., 2012, 2019; Saha et al., 2014). CFSv2, for instance, has high anomaly correlations with SST in the Niño 3.4 region (~0.82), while the average correlation with observed Northern Hemisphere precipitation rate over land is equal to 0.12 (Saha et al., 2014). Hybrid forecasts could therefore use AOGCMs for their strengths by forecasting SSTs (e.g. in the Niño 3.4 region as a proxy for ENSO) and then use a statistical relationship to generate the forecast rather than relying on the AOGCMs more variable rainfall estimates. Hybrid forecasts have been trialled with some successes in a few locations and times of year (e.g. Alfaro et al., 2018; Colman et al., 2020; Strazzo et al., 2019). Expanding on analyses like these could identify more opportunities to improve ensemble forecasts by capitalizing on important processes for regional rainfall.

1.2 Thesis Outline

This thesis is designed to support the R2O lifecycle of dynamic ensembles by evaluating their S2S rainfall forecasts over Central America. Comparing how the ensembles' S2S rainfall forecasts perform over the region, examining how some key rainfall generating processes relate to rainfall skill, and trialling out novel ways to post-process AOGCM forecasts can also further support operational choices between which forecasts to use and how best to optimize their results for preparedness decisions in Central America. The thesis is structured by first providing further background on the key regional drivers of rainfall over Central America, a review of how AOGCM forecasts are generated using numerical weather prediction, and an overview of the methodological framework used in the thesis to evaluate different S2S forecasting methods.

The subsequent chapters follow a publication-based format and address the following research questions:

1. How well do S2S models forecast monthly to seasonal rainfall, including low rainfall extremes, over the Central American Dry Corridor? **Chapters 4, 5**
2. How do leading S2S ensembles compare spatially and temporally when used for seasonal rainfall forecasts as deployed operationally today by institutions like the Central American Climate Outlook Forum? **Chapter 5**
3. How do alternative forecasting methods compare to using the S2S models to predict rainfall directly, including (1) a hybrid forecasting approach that uses the ensembles to predict rainfall indirectly with SSTs in known teleconnection zones, or (2) a statistical forecast based on ENSO? **Chapter 5**
4. Why do some models or their component members perform better than others at forecasting rainfall over Central America? **Chapters 4, 5, 6**
5. Can we using physical process-based evaluations to not only diagnose model performance, but to also select model members that represent key processes for rainfall to improve standard ensembling approaches? **Chapter 6**

CHAPTER 2

Literature Review

2.1 Central American Climate

Central America forms a relatively thin isthmus that connects North and South America, surrounded by the Caribbean Sea, Gulf of Mexico, and Easternmost Pacific Ocean, which is sometimes described as the Intra Americas Sea region (e.g. Amador et al., 2006, Figure 2.1.1 inset). The region's marked topography (Figure 2.1.1) gives rise to complex weather patterns, as moisture interacts with the Cordillera mountain range when it gets transported between ocean basins on either side of the isthmus (Taylor and Alfaro, 2005; Durán-Quesada, Gimeno and Amador, 2017; Muñoz-Jiménez et al., 2019; Durán-Quesada et al., 2020).

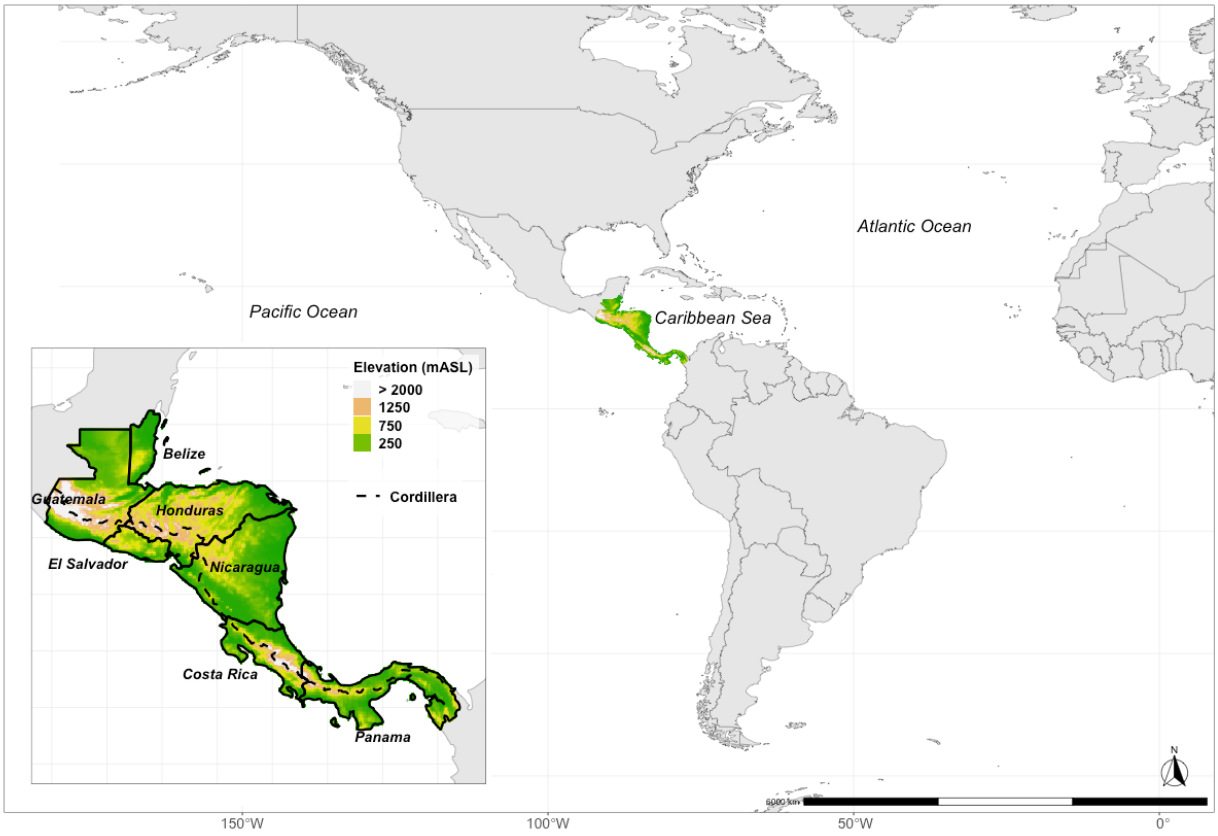


Figure 2.1.1 Map of Central America within Intra-Americas Sea region. Inset shows regional topography and labels Central American countries along with the Cordillera mountain range that divides the isthmus.

2.1 | Central American Climate

Rainfall is central to the Central American climate, which is often characterised by a wet and dry season (Figure 2.1.2a, Giannini, Kushnir and Cane, 2000; Taylor et al., 2002; Taylor & Alfaro, 2005). Seasonal rainfall patterns are often broadly categorized into Caribbean and Pacific rainfall subregions (e.g. Figure 2.1.2b) due to interactions between topography and

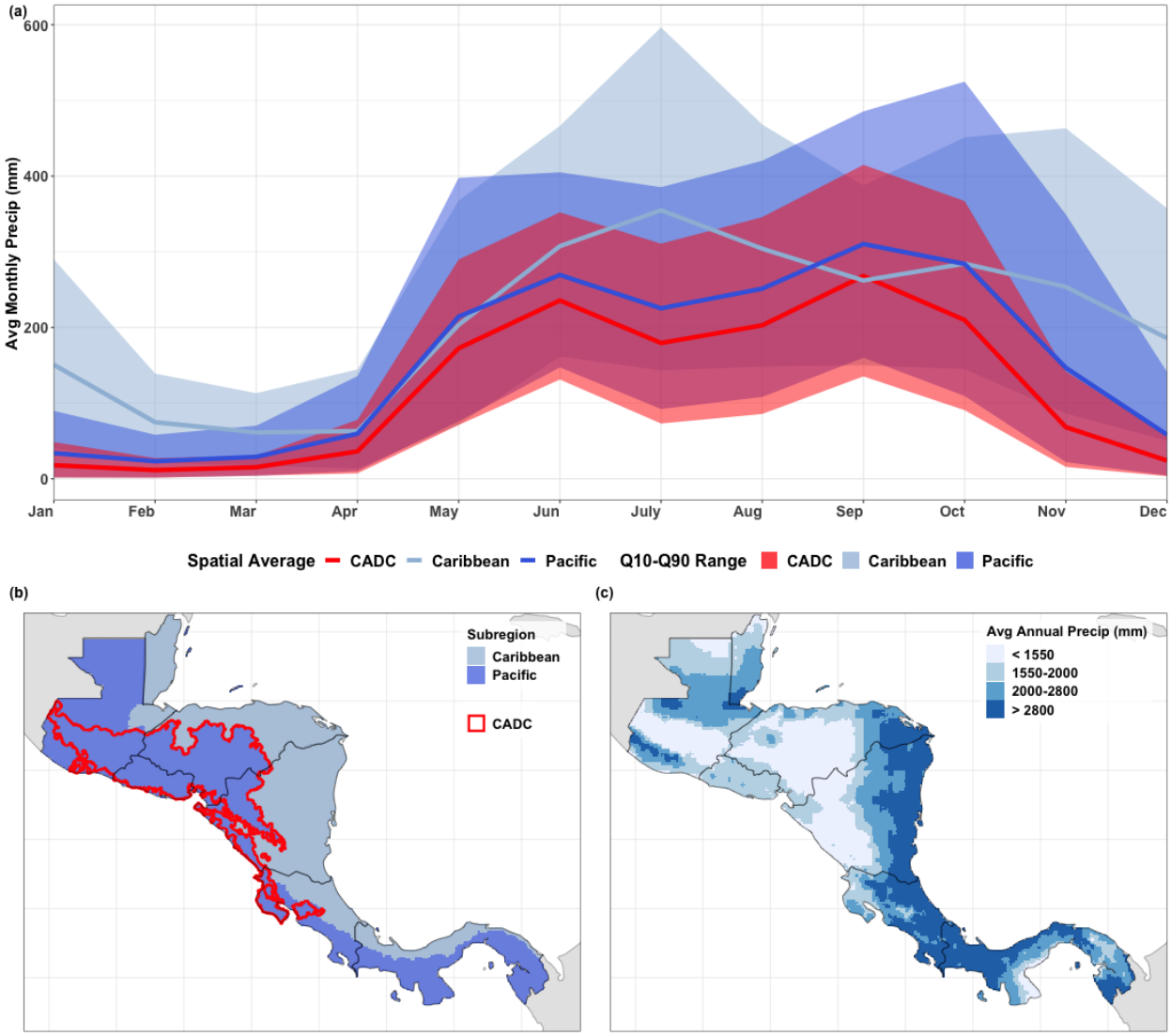


Figure 2.1.2 Central American climate. (a) average monthly variations in rainfall over the 35 years analysed in the thesis (CHIRPS, 1982-2016), using the spatial average of subregions defined in (b). Ribbons are plotted using the 10th and 90th percentiles within the time period for the spatially averaged monthly rainfall anomalies. (b) Although the Pacific, Caribbean, and CADC boundaries all vary (e.g. Muñoz-Jiménez et al. 2019; Quesada-Hernández et al. 2019), a loose approximation of the boundaries is shown to demonstrate the differences in monthly variability in (a). CADC boundary is delineated using the relatively contiguous locations in the CADC that have a dry season that lasts longer than four months (IICACR, 2014), excluding the dry arc of Panama and a couple of isolated locations in northern Guatemala. The Pacific is delineated to include regions that fall within the CADC, are on the Pacific side of the Cordillera, or are in the Pacific rainfall regime in Guatemala as delineated by the Guatemalan national weather service (e.g. INSIVUMEH, 2022a).

large-scale circulation patterns (García-Oliva & Pazos, 2021; Maldonado, et al., 2018; Portig, 1961; Taylor & Alfaro, 2005; Waylen et al., 1994), although these subregions are not always clearly distinct (Muñoz-Jiménez et al., 2019). The annual rainfall variance in most locations in Central America exhibits characteristics consistent with Pacific rainfall patterns, generally experiencing a wet season from May through October (Alfaro, 2002). This wet season is marked by a bimodal distribution with peaks in June and September and a trough in July and August, commonly referred to as the mid-summer dry period (MSD – Magaña et al., 1999, see Figure 2.1.2a). The CADC is another subregion within the Pacific region that experiences a longer MSD during the wet season compared to other locations (Gotlieb et al., 2019), frequent dry spells during the wet season (Peralta Rodríguez, Carrazón Alocén and Andrés Zelaya Elvir, 2012), and heightened vulnerability to droughts (Gotlieb et al., 2019).

The Caribbean rainfall region, in contrast, is often rainier (Figure 2.1.2.c) and experiences a longer wet season (Figure 2.1.2a), although the seasonal variability of Caribbean rainfall depends on the spatial and temporal scale of investigation (Martinez et al., 2019, 2020). Some Caribbean studies have shown the bimodal distribution of the wet season only exists in the northwestern Caribbean (Taylor and Alfaro, 2005; García-Oliva and Pazos, 2021), and other studies disagree on where and when the MSD exists (Alfaro, 2002; Curtis & Gamble, 2008; Gamble et al., 2008; Giannini, Kushnir and Cane, 2000). Several locations within the Caribbean regime near the Caribbean coasts of Honduras, Costa Rica and Panama have no defined dry season in some studies (Alfaro, 2002; Giannini, Kushnir and Cane, 2000). Martinez et al. (2019) provides one of the more recent comprehensive studies across the Caribbean using daily rainfall data, showcasing how the northern Caribbean coast of Central America still experiences a bimodal distribution in rainfall, but with a less distinct MSD.

2.2 Forecasting Across Timescales

Predictions can be generated across many timescales, from nowcasting at the very short range (hours ahead) up to climate projections multiple decades onwards (see Figure 2.2.1 for timeline of terminology). The type of information that can be effectively predicted changes as lead time increases, which affects the decision-making context, ranging from immediate mitigation efforts to longer-term adaptive and strategic choices, as illustrated in a schematic of decision-making steps using the agricultural sector as an example in Figure 2.2.1. Rainfall forecasts at the S2S timescale, for instance, can include lead times multiple weeks to several months ahead and can provide climate information about longer-term averages and deviations from mean conditions (e.g. average monthly rainfall, anomalies, below or above average rainfall, Stockdale et al., 2010). These types of forecasts can support routine planning and more proactive preparedness activities, including strategies for planting crops (e.g. Flohr et al., 2017, 2018) and transportation of resources to higher ground (Coughlan de Perez et al., 2016).

The processes that drive predictability change as lead time increases. Up to two weeks ahead, weather forecasts are primarily driven by the initial conditions of the atmosphere (Luo et al.,

2011). At much longer lead times, external boundary forces become increasingly important, as radiative forcing and other anthropogenic agents can affect the earth's climate (e.g. Solomon et al., 2007). S2S forecasts sit in between these time frames and their predictability is driven by a combination of atmospheric and oceanic forces. At the seasonal timescale, initial conditions of slower changing processes like soil moisture, sea-ice concentration, and SSTs dominate predictability (Doblas-Reyes et al., 2013), and predictability is primarily driven by the coupling of oceanic and atmospheric variability in the tropics (Luo et al., 2011). Subseasonal forecasts (e.g. two weeks to one month ahead) can face special prediction challenges, as atmospheric conditions lose most of their memory after several weeks and larger scale oceanic variability is not as dominant compared to longer lead times (Vitart and Robertson, 2018).

The El Niño Southern Oscillation (ENSO Trenberth, 1997) has the most important influence on S2S predictability globally and is the most skilfully predicted large-scale phenomenon at seasonal-to-interannual time scales (Stockdale et al., 2010). Predicting ENSO well may not be sufficient to generate skilful rainfall forecasts over Central America, however, as other climate phenomena drive regional rainfall (2.3). External boundary forces can also operate at the S2S timescale (Boer, 2010; Wulff, Vitart and Domeisen, 2022). Anthropogenic forcing, for instance, while sometimes thought to operate primarily beyond 10-year lead times (e.g. Cox & Stephenson, 2007), affects seasonal predictions of air surface temperature as early as three months ahead (Doblas-Reyes et al., 2006; Liniger et al., 2007; Boer, 2010).

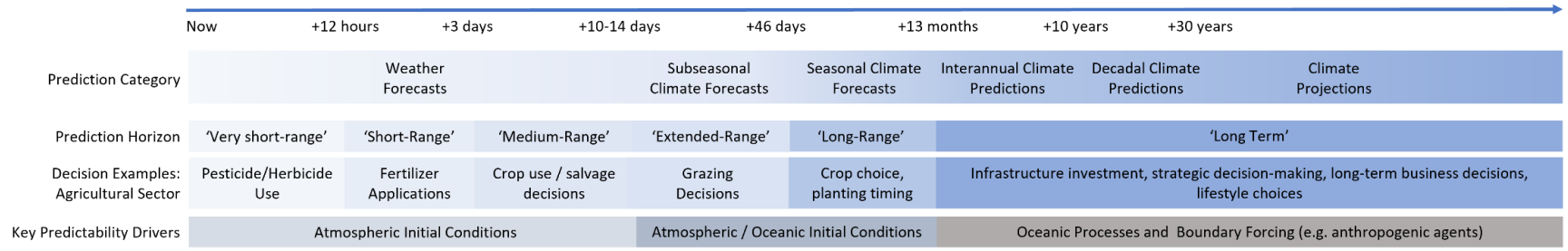


Figure 2.2.1 Forecast Terminology Timeline. Timeline of the terminology used, with key drivers of predictability labelled as the lead time increases from an initialization date. Terminology is generally agreed upon across forecasting centres but has some variation. For instance the UK Met Office cuts 'medium-range' off at 10 days while the ECMWF will define medium-range out to 15 days ahead (https://www.ecmwf.int/en/forecasts/documentation-and-support/medium-range-forecasts). The S2S scale applies as early as two weeks up to several months ahead (e.g. White et al. 2022). Decadal predictions are often used to describe forecasts up to 30 years ahead (e.g. Doblas-Reyes et al. 2013). Examples of types of decisions that can be informed in the agricultural sector are provided across the timeline, ranging from pesticide use that depends on whether rainfall will wash away a treatment in the next half hour, to salvage decisions about how to use planted crops in a drought prone season depending on whether rainfall will occur in the next couple of weeks, and crop choice and planting timing at the S2S scale up to several months ahead. More adaptive decision-making is possible using decadal predictions and climate projections by making longer-term strategic and investment decisions looking years onwards (e.g., Bodner et al., 2015; Flohr et al., 2017; Kamran et al., 2014). The key predictability drivers that affect forecasts are summarized at the bottom of the figure, see Lorenz (1975) and Luo et al. (2011) for further description of the transition between predictability drivers across timescales.

2.3 Key Drivers of S2S Rainfall over Central America

As a relatively thin isthmus between large ocean basins, Central America is subject to a complex interaction of weather patterns (Durán-Quesada et al., 2020). Several components of the earth system affect rainfall occurrence in Central America (Amador, et al., 2016a, 2016b). These components include energy reservoirs, available moisture in the air, atmospheric circulation structures, and moisture transport processes. For instance, dry winters in Central America are characterized by colder ocean temperatures in the Atlantic Ocean (energy reservoirs), limited atmospheric humidity (available moisture), and strong trade winds (moisture transport related to atmospheric circulation structures) (Taylor and Alfaro, 2005). These types of drivers and how they interact to modify regional rainfall are summarized in the following subsection.

2.3.1 Summary of Rainfall Drivers

(1) **energy reservoirs for potential convective and transport activity** are measured with temperature metrics (e.g. SSTs), energy budget metrics (e.g. convective available potential energy – CAPE), and heat exchange variables (e.g. latent and sensible heat fluxes that transfer energy when water vapor condenses or directly through conduction and convection between the earth surface and the atmosphere) (Amador, 2016a, 2016b). Warm water structures form on either side of Central America in locations that experience peak incoming short-wave radiation, often referred to as the Pacific and Atlantic components of the Western Hemisphere Warm Pool (WHWP; Wang et al., 2008; Wang & Enfield, 2001). These structures are often circumscribed by the 28.5 ºC isotherm, forming regions for convection and modifying moisture transport (Amador et al., 2006; Wang et al., 2008; Wang & Enfield, 2001, 2003).

(2) **how much moisture is currently available in the air** is measured with metrics like specific humidity. In low moisture conditions, the effects of transport processes are reduced, as seen in moisture flux divergence computations (e.g. Durán-Quesada et al., 2010, 2017).

(3) **how air will circulate within the atmospheric structure** is measured with pressure metrics at different heights of the atmosphere like geopotential height (GPH) and sea level pressure (SLP). The subtropical high of the North Atlantic (NASH), which forms the southern part of the North Atlantic Oscillation (NAO - Hurrell et al., 2003; Stephenson et al., 2003) around 1024 hPA (Taylor and Alfaro, 2005; Martinez et al., 2019), for instance, plays a dominant synoptic role in Central American rainfall, and its displacement will affect rainfall variability (Straffon, Zavala-Hidalgo and Estrada, 2020). The bimodal distribution of wet season rainfall over the Pacific regime, for instance, coincides with the seasonal migration of the NASH (Alfaro, 2002). The atmospheric circulation over the Pacific and Atlantic oceans also interacts with the regional topography, affecting rainfall formation over Central America (Taylor and Alfaro, 2005).

(4) **how fast air gets transported around the world** is measured by metrics like zonal (u) and meridional (v) wind components at different pressure levels, vertical velocity, and vorticity (Amador, 1998). Regional moisture transport processes across the isthmus are

connected to atmospheric circulation and play a key role in how much rainfall will occur over Central America (Durán-Quesada, Gimeno and Amador, 2017; Perdigón-Morales et al., 2021). Two low level jet (LLJ) streams are significant for regional moisture transport: the Caribbean Low Level Jet (CLLJ - Amador, 1998) and the Chorro del Occidente Colombiano (CHOCO Jet / CJ; Poveda & Mesa, 1999, 2000). The strength of the LLJs affects rainfall in part because it changes the amount of time clouds can precipitate over land, and weaker LLJs are often associated with greater amounts of rainfall, whereas strong LLJs often correspond to less total rainfall, especially on the Pacific side of the isthmus (Cook & Vizy, 2010; Hidalgo et al., 2015, 2019; Taylor & Alfaro, 2005). For instance, the significant periodicity of droughts over the CADC has been linked to increased NASH pressure and trade wind intensification (Hidalgo et al., 2019).

2.3.2 Driver Interactions and Effects on Seasonal/Interannual Rainfall

The CLLJ is connected to the NASH and is the main LLJ that transports moisture from the Caribbean Sea to the Pacific Ocean over Central America (Amador, 1998, 2008; Durán-Quesada et al., 2017; Hidalgo et al., 2015, 2019). The core of the CLLJ ranges from 925 to 700 hPA, its core crossing the middle of Central America around 12ºN and peaking in February (in part due to earth surface heating over South America Cook & Vizy, 2010) and July (in part due to NASH expansion; Cook & Vizy, 2010; Martinez et al., 2019), then reaching a minimum in October (García-Martínez and Bollasina, 2020). CLLJ effects vary spatially, as the CLLJ branches around the 10ºN latitude line (Hidalgo et al., 2015). Regional moisture transport in the northern half of Central America, for instance, is more affected by the summer peak of the CLLJ in July, while the southern half is more affected by the winter peak of the CLLJ in February (Muñoz et al., 2008).

The CHOCO jet transports moisture in the opposite direction from the Pacific to the Caribbean Sea over southern Central America and South America (Poveda and Mesa, 1999, 2000; Yepes et al., 2019; Mejía et al., 2021; Sierra et al., 2021), and its intensity is about half of the CLLJ (Durán-Quesada et al., 2010; Yepes et al., 2019). The CHOCO jet peaks in September and October when it also moves further north (Poveda and Mesa, 2000). This jet is secondary to the CLLJ over Central America due to large divergence in winds, lower wind intensities in the eastern equatorial Pacific region compared to the Caribbean Sea and marked topography interrupting the moisture flow (Durán-Quesada et al., 2010).

The convergence of these two jet streams is important, as Central America is located around the Hadley cells that move hot air from the equator towards the mid-latitudes. The positioning of the low pressure zone at the bottom of the ascending branch of the Hadley circulation, known as the Intertropical Convergence Zone (ITCZ - Adam et al., 2016; Adam, Schneider and Brient, 2018), is the area where the heaviest rainfall occurs in the earth system (Barry and Chorley, 2010). The CJ and CLLJ are two main components of the states of the ITCZ over the eastern Pacific (Mejía et al., 2021), and their convergence generates favourable conditions for Mesoscale Convective Systems, which generate large amounts of rainfall (Zuluaga and Houze, 2015) The seasonal migration of the ITCZ is key for regional rainfall generation (Alfaro, 2002; Sachs et al., 2009). Seasonal ITCZ migration, for instance, has been shown to contribute to the MSD formation (Small, de Szoeke and Xie, 2007). Each

of these drivers interact and affect rainfall formation over Central America. Over the Caribbean coast, for instance, the eastern Pacific ITCZ, interactions between the NASH and ITCZ, and CLLJ modifications from the Atlantic component of the WHWP will come together and affect rainfall (Martinez et al., 2019).

SSTs provide key energy reservoirs for rainfall over Central America. The oceans have different connections, however, as the eastern Pacific SSTs likely have more remote control over the ITCZ position in warm years (Taylor and Alfaro, 2005), whereas the Tropical North Atlantic (TNA - Enfield & Mayer, 1997) has more local forcing, increasing latent and sensible heat transfer over the troposphere in warmer conditions (Knaff, 1997; Giannini, Kushnir and Cane, 2000; Taylor et al., 2002) The TNA also modulates LLJ positioning and strength (Alfaro, 2002; Hidalgo et al., 2015; Maldonado et al., 2013; Taylor et al., 2002), and is more strongly associated with early wet season rainfall as compared to the late wet season (Taylor et al., 2002; Maldonado et al., 2017).

The El Niño Southern Oscillation (ENSO - Trenberth, 1997) in the eastern Pacific is the main mode of variability that affects rainfall over Central America (Durán-Quesada et al., 2020; Maldonado, Alfaro, & Hidalgo, 2018; Waylen et al., 1994), especially through its modification of CLLJ strength (Amador, 1998, 2008; Hidalgo et al., 2019), CJ and ITCZ position (Perdigón-Morales et al., 2021). El Niño events, for instance, have been shown to decrease annual Central American dam hydropower production (Ng, Turner and Galelli, 2017). ENSO affects Caribbean and Pacific rainfall in different ways though, as El Niño conditions are often associated with drier conditions in the Pacific and wetter conditions on the Caribbean side of the isthmus (Muñoz-Jiménez et al., 2019). ENSO is not the sole driver of rainfall though. For instance, only around one in three drought events in the CADC were related to an El Niño period over the period of 1950-2014 (Muñoz-Jiménez et al., 2019).

The relationship between SSTs on the Atlantic and Pacific sides of the isthmus is also important (Alfaro, 2002; Enfield & Alfaro, 1999; Giannini, Kushnir and Cane, 2000; Taylor et al., 2002) Anomalously wet years with increased rainfall and extended wet seasons are often associated with increased differences in temperature between the Pacific and Atlantic Oceans on either side of the isthmus (Enfield and Mayer, 1997; Enfield and Alfaro, 1999) Late wet season rainfall over the Caribbean, for instance, is strongly linked with the gradient between the Niño 3 region and the central equatorial Atlantic (0-15'W, 5'S-5'N) (Taylor et al., 2002). Successful rainfall forecasts over Central America therefore will account not only for individual changes in separate processes but be able to account for processes in both the Atlantic and Pacific oceans in addition to their interactions.

Other mechanisms act on a variety of timescales and affect regional rainfall. For instance, the Atlantic Multidecadal Oscillation (AMO - Enfield et al., 2001; Maldonado et al., 2016) and the Pacific Decadal Oscillation (PDO - Fallas López & Alfaro Martínez, 2012; Maldonado, Alfaro, & Hidalgo, 2018; Mantua et al., 1997) change more slowly over time and partially modulate tropical cyclone frequency just north over Mexico (Farfán, Alfaro and Cavazos, 2013; Martinez-Sanchez and Cavazos, 2014). The Madden Julian Oscillation (MJO - Madden & Julian, 1971) acts on shorter intraseasonal timescales (e.g. Barlow & Salstein, 2006; Barrett et al., 2013). Several other features can also affect rainfall, including direct

solar heating, wind-topography interactions, tropical waves, tropical cyclones, and other processes (Amador et al. 2016a, 2016b). Polar fronts from the midlatitudes (Nortes in Spanish), for instance, cause dry winters and early summer periods in northern Central America (Taylor and Alfaro, 2005), and tropical disturbances can propagate westward primarily affecting summer rainfall in the Caribbean (Taylor and Alfaro, 2005).

2.4 Dynamic Model Based Ensembles

Dynamic model-based ensembles use numerical weather prediction (NWP) to generate forecasts, which is the process of solving an initial-value problem: given what the earth system looks like today, they simulate how that system will evolve. Solutions require approximating physical processes based on a series of governing equations for how earth system variables will change in space and time (e.g. wind and pressure; Bauer, Thorpe and Burnet, 2015). The governing equations of atmospheric evolution were first referenced together in the early 1900's, as Vilhelm Bjerknes introduced them in his manifesto on NWP (Bjerknes, 1904). The governing equations include Newton's second law conserving momentum, the Navier-Stokes continuity equations, equations for the ideal gas states, the first law of thermodynamics, and the conservation equation for water mass (see full list of equations in Kalnay, 2003).

Applying these equations, however, is complex because the interactions between variables within the earth system are non-linear (Palmer, 2006). The non-linear aspect of the earth system creates instabilities, as Lorenz first discovered looking at how minuscule disturbances of a system can have large effects on its state (Lorenz, 1963, 1972, 1995). Lead time (time between when a model is initialized and a prediction is made) is important, as the internal variability of the earth system increasingly limits predictability as lead time increases (Stockdale et al., 2010; Lehner and Deser, 2023)

Modern AOGCMs couple dynamic atmospheric and oceanic models to approximate the states of many variables over space and time, such as wind speed, GPH, surface temperature, and arctic-sea ice concentration. Multi-model ensembles (MMEs) then combine AOGCMs to help address uncertainties in the earth system (different AOGCMs may have slightly different structural variations that affect how they describe the nonlinear dynamics of the earth system). MMEs ideally can offer a more robust prediction by sampling model uncertainty across AOGCMs (Palmer et al., 2004; Hagedorn, Doblas-Reyes and Palmer, 2005; Ferro, Richardson and Weigel, 2008). The comparative advantage of using MMEs compared to single model estimates has been demonstrated in several cases (e.g. Elvidge et al., 2023; Palmer et al., 2004; Wang et al., 2009). The difference is not always significant compared to using a calibrated single model, however, because the model errors in an MME are often correlated (Weigel et al., 2009).

To help address uncertainties in the observations, dynamic model-based ensembles generate probabilistic outputs comprised of multiple members that represent a distribution of guesses of the initial state of the earth system (Balmaseda et al., 2011). NWP has evolved as computational power has increased, creating the possibility to add more members and increase the spatial resolution of the models (Bauer, Thorpe and Brunet, 2015). In 1950, for instance, Charney, Fjortoft, and von Neuman made a one-day forecast using a barotropic

one-layer filtered model (assumes theoretical winds will be the same across heights) with some of the first electronic computers (Charney, Fjortoft and Neumann, 1950). By the 1980's the UK Met Office (UKMO) had operationalized a 15-level model that included oceanic wave components (Golding et al., 2004). Generating multiple members across initialized states then grew as computational capacity expanded in the 1990's (e.g. Tracton & Kalnay, 1993).

Methods to operationalize dynamic model-based ensembles are varied. Several choices can affect their predictions, including the number of members included in model initialization (e.g. Buizza, 2008; Leutbecher, 2019), bias adjustment and calibration techniques used to correct raw model outputs (Manzanas et al., 2019), and combination methods applied to aggregate models into MMEs (Hemri et al., 2020). Increasing the number of model members is often in tension with the spatial resolution of the model due to computational power constraints, which will also have a significant effect on skill, since models with higher spatial resolution tend to have higher skill (Leutbecher, 2019). As summarized in Table 2.4.1, the number of members in dynamic ensembles varies widely, ranging from as little as four (e.g. CCSM4) to much larger (e.g. 51 members in SEAS5).

Table 2.4.1 Summary of AOGCMs evaluated tagged by MME(s) to which they contribute, available hindcasts for statistical analysis, and forecasts for operational predictions.

Forecasting Center	Ensemble	MME Contribution	Available Members (Hindcasts)	Available Members (Forecasts)	Reference
CMCC	System 35	C3S	40	50	(Gualdi et al., 2020)
DWD	System 21	C3S	30	50	(Fröhlich et al., 2021)
ECCC	CansipsIC3	C3S, NMME	20	20	(Merryfield et al., 2013; Lin et al., 2021)
ECMWF	SEAS5	C3S	25	51	(Johnson et al., 2018)
Meteo France	Systems 7,8	C3S	12	25	(Batté et al., 2021)
NASA	GEOSS2S	NMME	4	10	(Molod et al., 2020)
NCAR/COLA	CCSM4	NMME	10	10	(Gent et al., 2011)
NOAA GFDL	SPEAR	NMME	15	30	(Delworth et al., 2020)
NOAA NCEP	CFSv2	C3S, NMME	4/start date from CDS, 24 from IRI	4/day from CDS, 28 from IRI	(Saha et al., 2014)
UKMO	Glosea6	C3S	7/start time	2/day	(Davis et al., 2020)

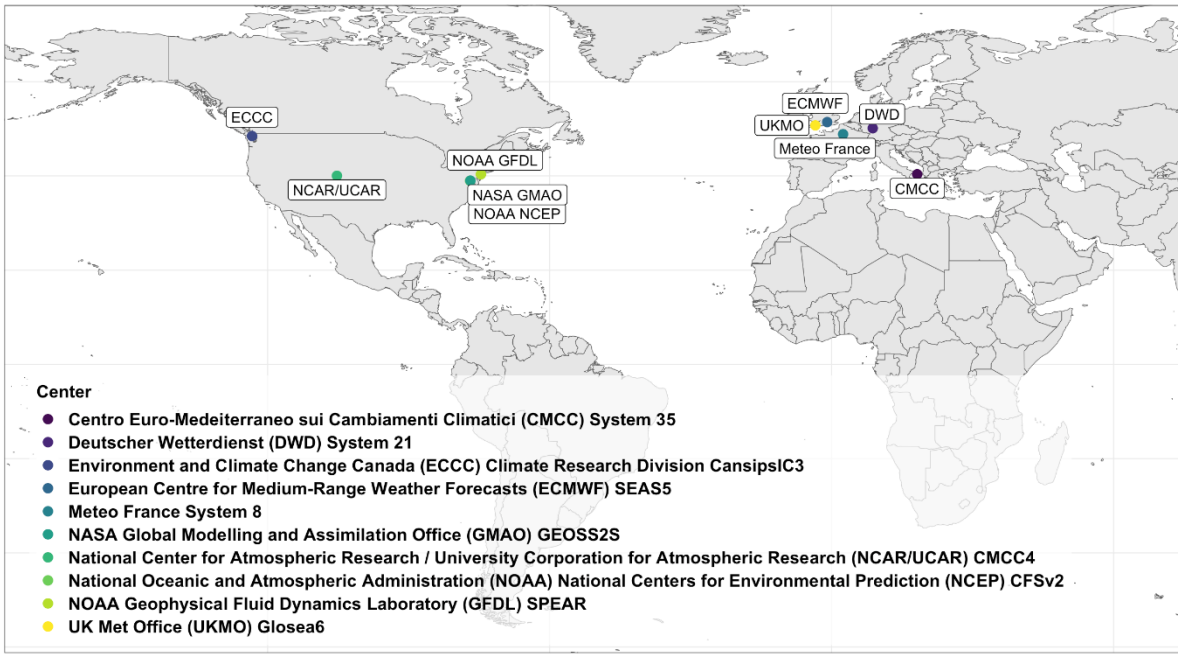
Bias adjustment and calibration approaches are varied and can range from simpler methods, such as mean variance adjustment (MVA, Hagedorn et al., 2005) and climate conserving recalibration (Weigel et al., 2009), to more sophisticated regression-based calibration techniques such as non-homogeneous Gaussian Regression (NGR - Gneiting et al., 2005). Canonical correlation analysis (CCA) has also been deployed to calibrate ensembles in the Next-Gen ensemble approach by fitting a correlation on principle components identified using empirical orthogonal functions over a larger zone that encompasses the predictand area (e.g. Acharya et al., 2021; Muñoz et al., 2019; Pons et al., 2021). The relative value of calibrating models using more complicated techniques than simple bias adjustment, however, is not always large (Manzanas et al., 2019).

Techniques to combine models into MMEs are also debated (e.g. Hemri et al., 2020; Troin et al., 2021). MMEs are often generated by first calibrating the individual model outputs and then combining the results using an unweighted mean (Acharya et al., 2021; Hemri et al., 2020; Muñoz et al., 2020; Pons et al., 2021). Using an unweighted mean to combine model outputs is debated, however, given that models are not necessarily independent from each other. Different models can have overlapping structural components and assumptions about how the earth system works, so they can have the same types of errors (Knutti, 2010).

Multiple efforts have been made to reduce interdependency issues when combining models, primarily at longer timescales (e.g. climate projection), including using model independence criteria to add relative weights to models before combining them or to select a subset of models (Delsole et al., 2013; Elvidge et al., 2023; Eyring et al., 2019; Giuntoli et al., 2021; Herger et al., 2018; Knutti et al., 2017; Lorenz, 2018; Sanderson et al., 2015, 2017; Slater et al., 2017; Weigel et al., 2010). Weighting means using interdependence criteria, however, adds complexity to the operationalization process, and the relative benefits of weighted means are not necessarily large enough to warrant the added complexity. Some findings have shown that hindcast samples at the seasonal scale, for instance, are too limited to see added benefits from combining models using a weighted mean (e.g. Mishra et al., 2019).

Two leading MMEs, the NMME (Kirtman et al., 2014) and C3S (Marsh and Penebad, 2016), are generated using model contributions from North America and Europe, as illustrated in Figure 2.4.1, which shows where the models that are evaluated in this thesis are located. While other forecasting centres also have AOGCMs that are publicly available (e.g. JMA/MRI-CPS1 from the Japanese Meteorological Agency is also a contributor to C3S, Takaya et al., 2017), those models' spatial resolution is often coarser than 1º. It was preferred to evaluate models that consistently offer at least 1º spatial resolution to keep the evaluation more focused on higher resolution applications (closer to 0.25º). A few dynamic ensemble systems are also produced in the southern hemisphere, such as the Predictive Ocean

Figure 2.4.1 Map of locations of evaluated AOGCMs in this thesis



Atmosphere Model for Australia (Australian Bureau of Meteorology). Latin America (Central and South America) has no unified MME designed for operational S2S forecasts, although Brazil has an AOGCM (BESM; Veiga et al., 2019) with an atmospheric component (BAM-v1.2) designed for operational seasonal climate predictions (Coelho et al., 2020).

Although the list of forecasting centres with dynamic AOGCMs plotted in Figure 2.4.1 is not comprehensive, these centres represent the main AOGCM options deployed by regional forecasting institutions in Central America (e.g. CA-COF seasonal climate outlooks; Garcia-Solera & Ramirez, 2012). Addressing research gaps in evaluations of dynamic model-based ensembles over Central America is important because they are not designed for the Central American context, but these models are often the best dynamic options available to be deployed today and their uncertainties can have cascading effects on regional outlooks and forecasting applications.

Forecast Evaluation Methodologies

CHAPTER 3

3.1 Forecast Purposes Affect Evaluation Techniques Used

Forecast evaluation can be approached in multiple ways depending on the motivation for analysis. Evaluation aims can include considering operational elements that pose barriers to forecast uptake (laws and norms, e.g. Feldman & Ingram, 2009; Rayner et al., 2005), and technical quality constraints (e.g. Stockdale et al., 2010). Within the technical forecast quality research context, evaluation purposes are often divided into three categories (Brier and Allen, 1951):

(1) **economic purposes** place the value of a forecast in monetary gains and are unique to the individual purpose of the forecast (e.g. how much more grain can a farmer produce from using a forecasting technique);

(2) **administrative purposes** can encompass a diverse range of needs but also include the need to communicate a forecast for non-specialist decisions (e.g. whether a flood warning is sufficiently likely to warrant an evacuation); and

(3) **scientific purposes** are more specifically aimed at the precision of a forecast for given states of the earth system.

The scientific value of a forecast often receives the most attention within the scientific community (Joliffe and Stephenson, 2012), which is useful within the model development context, but may not be as useful for decision-making as using an administrative lens (Mason and Weigel, 2009). In this thesis, scientific and administrative purposes are primarily considered to inform the evaluation methodology and metric choice.

3.2 Benefits of Statistical and Process-Based Evaluation Methods

Statistical methods are employed here alongside process-based analyses to further diagnose model performance. Statistical methods primarily focus on how models predict target variables (e.g. total monthly rainfall), using reforecasts over historical periods (often referred to as hindcasts) and comparing the mean and variance of model outputs against observations to test performance (e.g. Acharya et al., 2021; Coelho, 2013; Gubler et al., 2020; Hidalgo & Alfaro, 2012; Mishra et al., 2019; Pillai et al., 2018; Scaife et al., 2019; Slater et al., 2016; Walker et al., 2019). Statistical methods that assess predictions of target variables can answer some of the questions most relevant to administrative purposes, such as which model performs best for detecting high rainfall extremes in the early rainy season for a given location and provide insights into the spatial temporal variation in forecast performance for future model development.

Process-based methods assess model representation of potential predictors of target variables, such as moisture transport or SST (e.g. Baker et al., 2018; Oldenborgh et al., 2005; Stockdale et al., 2018). Process-based methods can inform future model development by shedding light on why a model is predicting a target variable well or poorly, and they can also support implementation efforts that use hybrid forecasting methodologies. Colman et al. (2020), for instance, create a hybrid forecast by using dynamic models to predict SSTs in important locations for generating rainfall and then statistically relate those outputs to a target variable. Many forecast evaluation studies perform a combination of the two methods. For instance, Scaife et al. (2019) identifies the regions in the world with the highest seasonal rainfall forecast skill using 14 dynamic model-based ensembles and examines how the models' abilities to predict the ENSO teleconnection relate to their rainfall forecast skill to inform opportunities for future model development.

3.3 Evaluation Metrics for Forecast Performance are Varied

Forecast evaluation metrics are diverse and will change depending on the purposes for analysis (e.g. administrative vs. scientific). Murphy (1993) outlines nine key attributes of forecast quality:

- **Bias** measures correspondence between the mean forecast and mean observation (e.g. does the forecast overpredict or underpredict rainfall);
- **Association** assesses the strength of the linear relationship between the forecasts and observations (e.g. Pearson correlation);
- **Skill** documents the relative accuracy of the forecast compared to a reference (e.g. a climatology, a random forecast, or persistence – most recent set of observations);
- **Reliability** compares the observed frequency of events with the predicted frequency of events (e.g. if a forecast system predicts 80% chance of above average rainfall for upcoming season, the system is statistically reliable if in 80% of cases, the observations are above normal in that season; Weisheimer & Palmer, 2014);
- **Resolution** is the degree to which the forecast can resolve events into subsets with different frequency distributions;
- **Sharpness** measures forecast tendency to predict extreme values (not necessarily correctly);
- **Discrimination** examines the forecast's ability to distinguish between observations, i.e., higher predictive frequency when the event occurs; and
- **Uncertainty** considers the variability of the observations.

Using more metrics is often better to support scientific purposes by informing different aspects of how a model represents the atmospheric system. Forecast verification papers often deploy of few of these metrics together, commonly citing forecast skill, association, and bias (e.g. Baker et al., 2018; Slater et al., 2016; Walker et al., 2019). Some skill scores can also be decomposed into component parts that represent different aspects of model quality like reliability and resolution (e.g. decomposition of Brier skill score; Young, 2010).

For administrative contexts, discrimination has been promoted as a useful aspect of model quality to test for decision-makers (Mason, 2018; Mason & Weigel, 2009; Weigel & Mason, 2011). In the CA-COF for instance, forecasts are currently communicated in tercile categories (Above Normal, Normal, Below Normal – Garcia-Solera & Ramirez, 2012). Mason and Weigel (2009) propose metrics like two alternative forced choice (2AFC) to showcase a model's ability to differentiate between different states (e.g. high rainfall and low rainfall cases versus normal conditions). 2AFC has been deployed in recent forecast evaluation efforts (e.g. Acharya et al., 2021; Pons et al., 2021) and is cited as a metric for forecast analysis in CA-COF meetings (discussion with CA-COF contributors, 28 September 2021).

Skill metrics like the Heidke Skill Score (HSS – Heidke, 1926) that test detection rates of extremes are also common (e.g. Higgins, Kim and Unger, 2004; Becker et al., 2012; Walker et al., 2019). Simpler metrics are also available for extremes detection (e.g. Hit Rate, how often a model identifies an extreme event correctly, and False Alarm Rate, how often a model misidentifies an extreme even when one did not occur). HSS is preferred for extremes detection here because it is equitable, so random forecasts (e.g. a forecast that always guesses an extreme will occur) always score as zero, meaning no skill (Hogan et al., 2010). Forecast evaluations here are primarily conducted on the raw ensemble outputs to understand their baseline forecast skill, which can provide insights into forecast potential, since positive skill is often easier to improve through post-processing (e.g. Slater & Villarini, 2018). Chapter 5 also uses calibration techniques that are used operationally today (e.g. using CCA as is used in the Next-Gen forecasting approach).

Ultimately, a range of forecast evaluation metrics and techniques are useful to test the trustworthiness of a forecast. Using statistical and process-based assessments is also beneficial to more comprehensively showcase both the spatial and temporal variability in forecast performance and some causal mechanisms that potentially drive AOGCM predictions (e.g. how ENSO representation relates to predictive skill). The following chapters (4-6) apply a range of probabilistic and deterministic metrics to assess both mean estimates and the range of ensemble member predictions for different months and locations.

To further support implementation efforts, the evaluation methodology used here is also structured around comparative evaluation. Comparative evaluation means framing the analysis within relative terms as compared to solely using absolute values to assess performance given that people often make decisions in terms of comparative losses and gains (Kahneman and Tversky, 1984). Quoting absolute forecast skill values alone is therefore less informative than showcasing how one model or method performs relative to another when helping decision-makers choose between forecast tools. This includes comparing how well AOGCM updates compare to their previous versions (e.g. SEAS5 versus S4) and how well AOGCMs and their combined MMEs compare against each other.

SEAS5 Skilfully Predicts Late Wet-Season Precipitation in Central American Dry Corridor Excelling in Costa Rica and Nicaragua

CHAPTER 4

Co-authors. Louise Slater, Anne F. Van Loon, and Christian Birkel
Published. 24 December 2021 https://doi.org/10.1002/joc.7514

Motivations

Low rainfall extremes are investigated in Chapter 4 due to drought risks in the CADC (Calvo-Solano et al., 2018; Gotlieb et al., 2019). Two of the five central research questions (RQs) are addressed in this chapter (RQs 1 and 4 in Thesis Outline, 1.2). SEAS5 was selected for this study because there were no evaluations that focused on that model's skill in Central America, especially over the CADC. SEAS5 was also of interest because of its potential to perform well in the region given evaluations in nearby regions (e.g. South America, Gubler et al., 2020) and evaluations of the previous version (S4) in Central America (e.g. Carrão et al., 2018).

Abstract

Better drought preparedness is critically needed in the Central American Dry Corridor (CADC). Seasonal forecasts can be used to build this preparedness but need localized evaluations to ensure they are relevant and useful. This study provides a CADC-focused assessment of the SEAS5 seasonal forecasting system produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). We evaluate SEAS5 predictions of the mean, variability, and extremes of precipitation across the CADC at one- to seven-month lead times. We assess differences in regional forecast quality across seasons and lead times by evaluating spatial and temporal associations with El Niño Southern Oscillation (ENSO) phase, topography, and continentality. Results show that SEAS5 precipitation forecasts often have better skill primarily during the mid to late wet season (July-October). In these months, low/normal precipitation forecasts outperform the climatological mean (1982-2016) up to five- or six-month lead times in some subregions. Forecast skill is often worse, however, for predicting precipitation during the early wet season, primarily in June. Forecast skill varies spatially across the region, with higher skill concentrated in the southeast (Costa Rica and Nicaragua). Forecast skill is significantly related to continentality and topography, and together these factors account for at least a quarter of the spatial variance in annual skill at all lead times. Forecast accuracy varies depending on ENSO phase: predictions are often worse in El Niño (warm ENSO) periods during the early wet season when ENSO also has a weaker association with cumulative precipitation relative to the later wet season. SEAS5 could be a particularly useful tool during the second half of the wet season in the southeast CADC, skilfully alerting of upcoming precipitation variability with over three-month lead times.

4.1 Introduction

Drought threatens many communities around the world. The Central American Dry Corridor (CADC) is one region that continues to face acute drought impacts with significant costs to health and livelihoods (Peréz-Briceno et al., 2016; Alfaro, Hidalgo, and Pérez Briceño, 2018; Calvo-Solano et al., 2018; Beveridge et al., 2019; Gotlieb et al., 2019). While most of Central America is composed of a relatively wet climate, the CADC is along the Pacific boundary (Figure 4.1) and experiences a longer dry season and higher risk of drought (Peralta Rodríguez, Carrazón Alocén and Andrés Zelaya Elvir, 2012). The agricultural sector in the CADC (a key income source for many locals – (Gotlieb et al., 2019) experienced significant drought-related impacts in at least eleven of the twenty years during 1997-2016 (Calvo-Solano et al., 2018). Going forward, water availability is projected to decrease across Central America with an increased frequency in drought (Hidalgo et al., 2013; Imbach et al., 2018; Cook et al., 2020; Almazroui et al., 2021). Almazroui et al. (2021) for instance, showed temperature is likely to increase while precipitation is projected to decrease between 10-40% over Central America by the end of the 21st Century. Therefore, exploring preparedness options is important since drought-related pressures in the CADC are widespread and unlikely to abate.

Seasonal forecasts can support ongoing preparedness efforts by informing decisions for the upcoming season, such as crop choice, hydroelectric power reliance, and water supply for human consumption. Several forecasting approaches are available to predict drought (e.g. statistical or dynamical; Hao, Singh and Xia, 2018). Dynamic seasonal forecasts based on ensembles of global circulation models combine multiple potential climate states into a prediction, and their quality continues to improve (Bauer, Thorpe and Brunet, 2015). These systems have shown promise near Central America in part due to the predictability of the El Niño Southern Oscillation (ENSO) teleconnection pattern in the eastern Pacific (Scaife et al., 2019). While these systems are often only publicly available at relatively low spatial resolutions (0.5º or 1º in some cases), they can warn of upcoming anomalies in drought-related indicators across the CADC and be followed by impact-based forecasting assessments that are tailored to more regional and sector-specific needs. Some dynamic

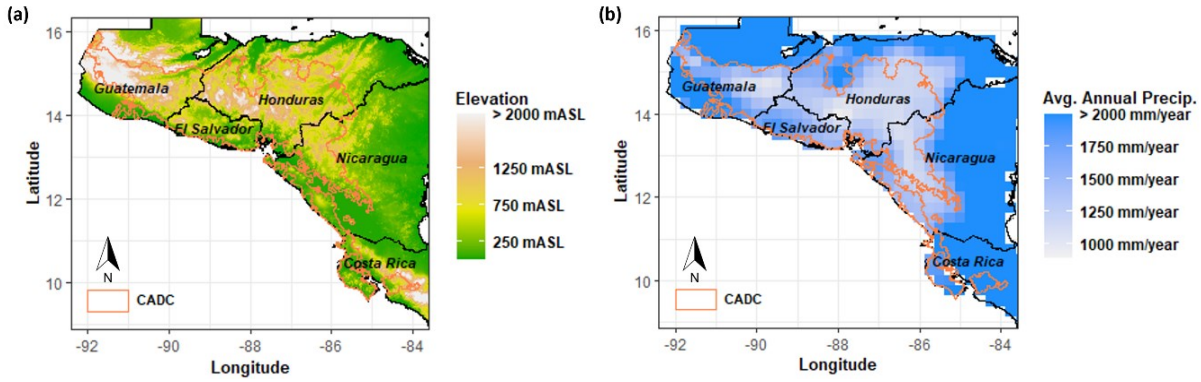


Figure 4.1 CADC characteristics. (a) Elevation measured in meters above sea level (mASL) across Central America with the CADC boundary outlined based on the climate risk index (IICACR, 2014), including relatively contiguous locations with a dry season that lasts longer than 4 months. (b) Average annual precipitation (1982–2016) using GPCC 0.25º resolution data (Schneider et al., 2018) over Central America including CADC boundary.

seasonal forecasting systems are in use today across Central America – the Central American Climate Outlook Forum (CA-COF) leverages multiple dynamic systems to inform three-month outlooks primarily for precipitation (Donoso & Ramirez, 2001; Garcia-Solera & Ramirez, 2012; Maldonado et al. 2013, 2018).

Central America is subject to complex interactions between climatic mechanisms (Durán-Quesada et al., 2020), which can limit drought predictability. ENSO, for instance, is key for regional precipitation, and its various modes modulate the strength of other phenomena (e.g. Caribbean Low Level Jet – CLLJ; Amador, 2008) that convey moisture across the isthmus (Durán-Quesada, Gimeno and Amador, 2017). ENSO alone, however, cannot explain regional drought occurrences (Hidalgo et al., 2019; Maldonado, Alfaro, & Hidalgo, 2018; Muñoz-Jiménez et al., 2019). Although drought risk in the CADC is connected to El Niño (warm ENSO) events, drought severity is not consistent between strong El Niño phases (Muñoz-Jiménez et al., 2019) and extreme droughts can equally occur in other phases of ENSO. During a La Niña episode in 2010, for instance, the CADC experienced drought-related losses in corn and bean crops worth over 85 million USD in El Salvador alone (Calvo-Solano et al., 2018). Other factors may overwhelm the influence of El Niño phases on precipitation deficits (e.g. Tropical North Atlantic and quasi decadal variability – Alfaro, 2007; Maldonado et al., 2017; Hidalgo et al., 2019), as both the Atlantic and Pacific Oceans play significant roles in weather formation over the thin land-bridge (Durán-Quesada et al., 2020). The CADC also presents unique forecasting challenges compared to the rest of Central America in part due to its complex topography. The region is mountainous with steep changes in elevation from sea level to over 2000 mASL with some mountains reaching anywhere from 3000-4000 mASL in some parts of the cordillera (Gotlieb et al., 2019, Figure 4.1a). Topography is important, as mountainous terrain creates special challenges for numerical weather prediction (Serafin et al., 2018; Chow et al., 2019), but the relationship between elevation and predictability could use further exploration in the CADC. A forecasting evaluation that focuses on the CADC is therefore useful because it can identify both the potential skill for regional applications and potential limits to predictability in the region.

This study provides a first look into dynamic seasonal forecast quality across the CADC using SEAS5, the 2017 update to one of the leading dynamic forecasting systems produced by the ECMWF (Johnson et al., 2019). Dynamic forecasting systems like SEAS5 require more localized evaluation in the CADC specifically, as relatively few evaluations have focused on how these systems perform across the region, and their quality may vary regionally. Some studies have addressed how the previous version of the ECMWF forecasting system (S4) performed over Central America more generally (Dutra et al., 2014; Weisheimer and Palmer, 2014; Carrão et al., 2018), and others have investigated SEAS5 performance in nearby regions (e.g. in South America – Gubler et al., 2020). For instance, (Weisheimer and Palmer, 2014) showed S4 had a perfect reliability score in predicting precipitation variability over Central America during June through August (JJA) and December through February (DJF) using initialization periods of two and four months in both relatively warm and cool years. Others have shown that S4 precipitation skill in South-Central America was higher when predicting moderate drought events, but the system's skill

suffered as the intensity of the drought increased (Carrão et al., 2018). To build on these analyses, we investigate SEAS5 in more months, including the entire wet season (May-October) when precipitation is most variable, explore sources of regional predictability that may affect SEAS5 skill, and consider how SEAS5 predictions may have changed relative to S4 – SEAS5 has upgrades compared to S4, including an increase in the resolution of the ocean models from 1.0 to 0.25 degrees, which enhances the representation of ENSO variability (Johnson et al., 2019) and could improve performance over Central America.

SEAS5 skill is assessed in different contexts, including precipitation mean and variability over the entire period of analysis (1982-2016), below-normal precipitation cases (using 10^{th} and 20^{th} percentiles of monthly precipitation anomalies), and historical drought events to showcase how SEAS5 could perform in years when residents would benefit from an early precipitation-deficit warning. The specific goals of this assessment are to (1) examine the spatiotemporal characteristics of SEAS5 forecasts across months and lead times within the CADC with a focus on low-precipitation anomalies, and (2) explore potential associations between forecast skill and regional drivers of predictability such as elevation, continentality, and ENSO variability.

4.2 Data and Methodology

4.2.1 CADC Boundary Delineation and Data Selection

The delineation of the CADC is imprecise. Some studies have circumscribed the CADC using the Climate Risk Index that identifies areas with a dry season lasting longer than four months, including parts of Guatemala, Honduras, Nicaragua, El Salvador, Costa Rica, and occasionally the Dry Arc of Panama (FAO, 2016; Gotlieb et al., 2019). Others have defined the "core" CADC to include locations where the aridity index is less than one (ratio between average annual precipitation to potential evaporation) (Hidalgo et al., 2019), or applied a flexible boundary definition to better account for annual variations in precipitation (Quesada-Hernández et al., 2019). Our research defines the CADC by the outline shown in Figure 4.1, which simplifies the boundary to a relatively contiguous region using the Climate Risk Index (IICACR, 2014), excluding the Dry Arc of Panama and a few isolated places in Guatemala.

SEAS5 total monthly precipitation 0.25º resolution hindcasts from all 26 available model members (51 members are available for real-time forecasts and 26 are available for hindcasts) were obtained over 1982-2016 across one- to seven-month lead times and assessed against the Global Precipitation Climatology Centre (GPCC) 0.25º dataset, a gridded gauge-based observational precipitation dataset (Schneider et al., 2018). While observational datasets all have their relative benefits and drawbacks, GPCC is preferred here because it is publicly available and has strict quality control requirements such as a minimum of ten uninterrupted years of run time for rain-gauge station inclusion (Schneider et al., 2018; Sun et al., 2018). Prior to evaluation, GPCC was also compared against Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS; Funk et al. 2014), which showed similar mean monthly precipitation estimates at 0.25º resolution. Decreasing support for rain gauges is also a known challenge (Sun et al., 2018), so the number of rain gauges in the GPCC dataset were examined over the evaluation time-period. After 2000, the total

number of gauges in the region dropped below 50, primarily in the central CADC (Honduras). Fewer rain gauges were also in operation over the evaluation time period near sea level and above 1250 mASL. Caution is therefore warranted in relying on the observations outside of intermediary elevations, as for many observational rain gauge datasets.

A few additional datasets (all publicly available) are leveraged to assess possible relationships between different phenomena and forecast skill. To examine topography, resampled Shuttle Radar Topography Mission elevation data (Hollister, 2021) is re-gridded to 0.25º resolution using bilinear interpolation. Continentality is examined by calculating distance from the coast using the Euclidean distance with Natural Earth data. Both the Multivariate El Niño Index (MEI) (Wolter and Timlin, 2011) and Oceanic Niño Index (ONI) (NOAA Climate Prediction Center, 2021) are used to assess the effects of ENSO phase. MEI is employed here to provide a more comprehensive ENSO estimate, as this index is based on multiple indicators such as sea surface temperature (SST) and sea level pressure (Wolter and Timlin, 2011). ONI only accounts for SST anomalies but is preferred for categorizing historical droughts by ENSO phase due to its clear delineations between phases in the U.S. National Oceanic and Atmospheric Administration (NOAA) database.

4.2.2 Low-Precipitation Anomaly Thresholds and Historical Drought Events

For this assessment, SEAS5 is evaluated for its predictions of monthly precipitation across the CADC with an emphasis on predictions of below-normal precipitation. Precipitation anomalies drive meteorological droughts (below-normal precipitation events) and they are a key factor in other forms of drought (Van Loon, 2015). In many Central-American catchments in particular, there is a relatively close relationship between precipitation deficits and hydrological droughts (low streamflow events) because groundwater processes do not dominate (Quesada-Montano et al., 2019). Precipitation is also relevant because it is prioritized in existing seasonal forecasting applications like those developed through the CA-COF (Maldonado et al. 2013). Although several drought-monitoring indices and metrics exist (WMO, 2016), monthly precipitation was chosen for evaluation because this metric is relatively flexible – it can be applied in threshold analyses for sector-specific uses or calculated into an index to monitor drought more broadly (e.g. Standardized Precipitation Index – SPI; Mckee, 1995). Low-precipitation thresholds are defined using monthly precipitation anomalies below the 10^{th} and 20^{th} percentile chance of occurrence aligned with Svoboda et al. (2002) and Peterson et al. (2002) definitions of moderate to severe drought and below-normal precipitation. Although the absolute value of the threshold for each month may vary depending on monthly precipitation accumulation (e.g. the driest 10% of May data will be a higher value than the driest 10% of January data), the number of events included in each threshold is consistent across months (e.g. 10% thresholds include about 4 of the 35 years).

SEAS5 accuracy during historical drought events is evaluated using three-month mean precipitation hindcasts to show how SEAS5 would have performed in an operational context. Calculating the three-month mean hindcast requires setting one initialized month and combining different lead times from that date to construct a hindcast for the entire three-

month period. For instance, a three-month mean hindcast that is initialized in June and predicts precipitation for August through October would be calculated as the average precipitation from a two-month lead time for August, a three-month lead time for September, and a four-month lead time for October. Events are identified from the CADC drought-related agricultural impact review by (Calvo-Solano et al., 2018). Agricultural impacts are emphasized because this sector is key to most livelihoods in the CADC (Gotlieb et al., 2019). Events are then categorized by ENSO phase to see how hindcast accuracy compares between events that occurred in different ENSO phases, including those not typically associated with drought (e.g. La Niña and Neutral phases – Table 4.1). The 2000 event is additionally included, because the annual precipitation anomalies were drier than one standard deviation below the climatology in multiple locations and worsened the impacts from the drought in 2001 (CEPAL, 2002).

Table 4.1 Historical drought events in the CADC from 1997-2016 with documented agricultural impacts (Calvo-Solano et al. 2018) are categorized by the ENSO phases when they primarily occurred (El Niño, La Niña, Neutral) as measured three-month mean SST anomalies in the Oceanic Niño Index (ONI – NOAA Climate Prediction Center, 2021).

Year	Agricultural Impacts – see Calvo Solano et al. (2018) for more event descriptions
El Niño, ONI ≥ 0.5	
1997	Loss of basic grains across CADC
2004	Loss of basic grains primarily in Guatemala
2009	Crop loss in Guatemala, Honduras, Nicaragua, and Costa Rica intensifies July-September
2015	Crop loss across CADC, Guatemala loses 80% first crops
La Niña, ONI ≤ -0.5	
2000	Dry year preceding another dry year in 2001, which exacerbated crop losses in 2001
2008	Major crop reductions in El Salvador
2010	Crop loss in El Salvador and Guatemala, losses worth over 85 million USD in El Salvador
2016	Corn and bean losses in Guatemala and Honduras
Neutral, -0.5 < ONI < 0.5	
2001	Loss of basic grains across CADC
2012	Primarily coffee crop losses, between 33-100% of harvest lost across the CADC
2013	Major coffee crop losses across Costa Rica and rice and bean crop losses in Guatemala
2014	Loss of basic grains primarily in Costa Rica, El Salvador, and Nicaragua

4.2.3 Evaluation Criteria

Several verification metrics are used to assess SEAS5 performance across the time series – both deterministically using the ensemble mean, and probabilistically using the range of the ensemble members (Table 4.2). To assess deterministic predictions of precipitation mean and variability, skill is examined using the Mean Squared Error Skill Score (MSE-SS) (e.g. (Gubler et al., 2020; Kelemen et al., 2019; Maldonado et al., 2018; Slater et al., 2016; Slater & Villarini, 2018). MSE-SS is calculated from annual and monthly standard anomalies (using the climatological mean and standard deviation from the hindcast period 1982-2016). The Heidke Skill Score (HSS; Heidke, 1926) is used to assess the deterministic skill of identifying low-precipitation anomalies. HSS is chosen for several reasons: the metric is equitable by giving equal weight to different types of random forecasts (Hogan et al., 2010), understandable (1 = perfect, 0 = skill of random forecast), and commonly used (Higgins, Kim and Unger, 2004; Becker et al., 2012; Walker et al., 2019). Probabilistic skill for normal precipitation variability is evaluated with the Continuous Ranked Probability Skill Score

(CRPSS; Matheson and Winkler, 1976). CRPSS is employed to assess the spread of the model members for normal precipitation variability and is preferred because the score is widely used (Mishra et al., 2019; Ratri et al., 2019; Slater & Villarini, 2018; Wang et al., 2019), clear to interpret (1 = perfect, 0 = skill of climatological mean), and applicable to raw ensemble outputs to showcase initial performance prior to post-processing the data into a probability distribution (Bröcker, 2012). CRPS (different from CRPSS because CRPS only measures differences between forecasts and observations and is not baselined against a reference forecast) is calculated to show how the accuracy of the system varies in the historical event cases (0 = forecast and observations are equal, larger values less accurate).

Multiple linear regressions are then used to assess the strength of the relationships between regional characteristics and forecast skill. Forecast skill (using MSE-SS) is regressed against elevation and continentality to understand the degree to which these factors relate to the spatial variance in skill. To better understand the degree to which SEAS5 skill is associated with ENSO variability, here we examined (1) how the strength of the relationship between ENSO variability and precipitation accumulation compares with the seasonality in forecast skill, and (2) how ENSO variability relates to forecast accuracy in different times of year. We tested the first relationship by regressing MEI against precipitation in different periods of the year and compared the strength of the association with the variations in forecast skill. To explore possible connections between ENSO phase and forecast quality, the forecast accuracy (using spatially averaged CRPS) is regressed against MEI for different periods of the wet season and across lead times. We primarily focused on the relationships between ENSO and forecast skill in the wet season because these months are when meteorological drought forecasts are most relevant, as agriculture in the region is highly dependent on precipitation during this period.

Table 4.2 Summary of forecast verification metrics with their relevant formulas

Metric	Summary	Formula
Deterministic Measures		
Mean Squared Error Skill Score (MSE-SS)	Average squared difference between ensemble mean and observations baselined against climatological mean (1 = perfect, 0 = skill of climatological mean) (Deque, 2012).	$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2$ $MSE - SS = 1 - \frac{MSE_f}{MSE_{clim}}$ $\hat{x} = observation$ $x = forecast$ $n = total\ number\ of\ instances$ $MSE_f = MSE\ forecasts$ $MSE_{clim} = MSE\ climatology$
Heidke Skill Score (HSS)	Skill of categorical event detection using the ensemble mean that calculates proportion of correct identifications and accounts for number of hits due to chance (1 = perfect, 0 = skill of random forecast) (Heidke, 1926).	$HSS = \frac{a + d - a_r - d_r}{n - a_r - d_r}$ $a_r = (a + b)(a + c)/n$ $d_r = (b + d)(c + d)/n$ $a = hits$ $b = false\ alarms$ $c = misses$ $d = correct\ rejections$ *n = total number of instances*
Probabilistic Measures		
Continuous Ranked	Error of the range of ensemble members (0 = perfect, larger values are less	$CRPS = \int_{-\infty}^{\infty}\{Fs(q) - FQ(q)\}dq$

Metric	Summary	Formula
Probability Score (CRPS)	accurate). Integrates squared difference between distributions of forecasts and observations (Weigel, 2012)	$Fs(q) =$ cumulative forecast distribution $FQ(q) =$ cumulative observation distribution
Continuous Ranked Probability Skill Score (CRPSS)	Skill of the range of ensemble members using CRPS compared to the climatology (1 = perfect, 0 = skill of climatological mean)	$CRPSS = \frac{\overline{CRPS_f} - \overline{CRPS_{ref}}}{CRPS_{perfect} - \overline{CRPS_{ref}}}$ $\overline{CRPS_f}$ = mean CRPS forecasts $\overline{CRPS_{ref}}$ = mean CRPS observations $CRPS_{perfect} = 0$

4.3 Results

4.3.1 Normal Precipitation Variability

SEAS5 forecasts of normal monthly precipitation are better in the mid to late wet season across most lead times (July-October, Figure 4.2). Skill is highest in July and August, commonly known as the mid-summer dry period, which is a relatively consistent

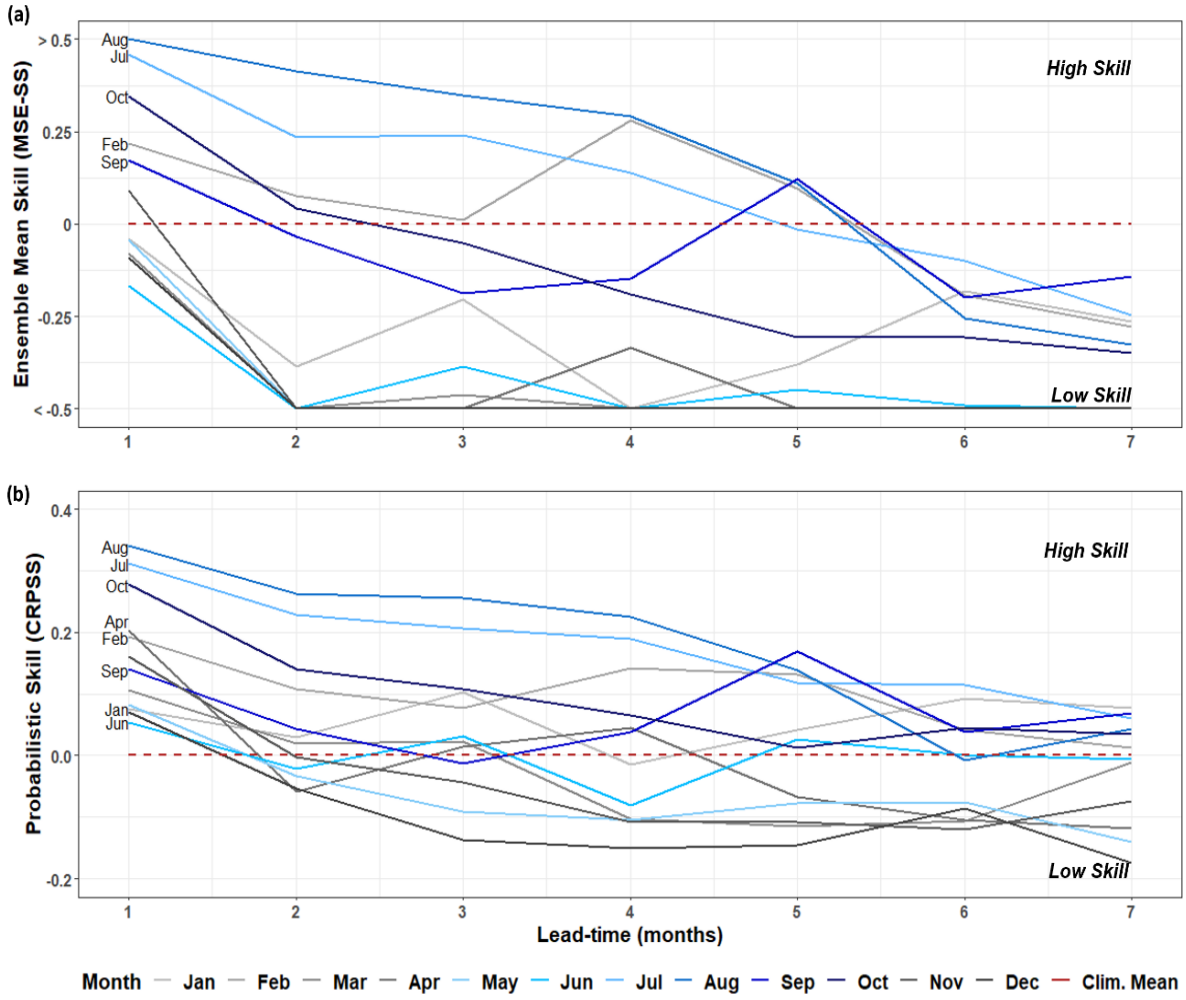


Figure 4.2 Spatially averaged skill (MSE-SS) is plotted over lead time and colour-coded by month using the ensemble mean (a) and using the range of the ensemble members with CRPSS (b). Skill of the climatological mean (1982–2016) is at zero, plotted with a horizontal dashed line. Months are labelled when the spatially averaged skill is positive for at least two lead times.

phenomenon in the CADC (Magaña, Amador and Medina, 1999) – MSE-SS (1 = perfect, 0 = skill of climatological mean) remains positive up to a four-month lead time in July and a five-month lead time in August, and CRPSS (1 = perfect, 0 = skill of climatological mean) is only negative in August at six-month lead time (Figure 4.2). SEAS5 forecasts also have higher skill during the late wet season (September-October) compared to May through June for normal precipitation variability. The spatially averaged skill (using MSE-SS), for instance, ranges from 0.19 to 0.35 in September through October at one-month lead time, whereas skill in the early wet season is never positive at any lead time (Figure 4.2).

The seasonality of SEAS5 skill is also apparent when the forecast skill (using CRPSS) is plotted spatially over months and lead time (Figure 4.3). In the best performing months (July

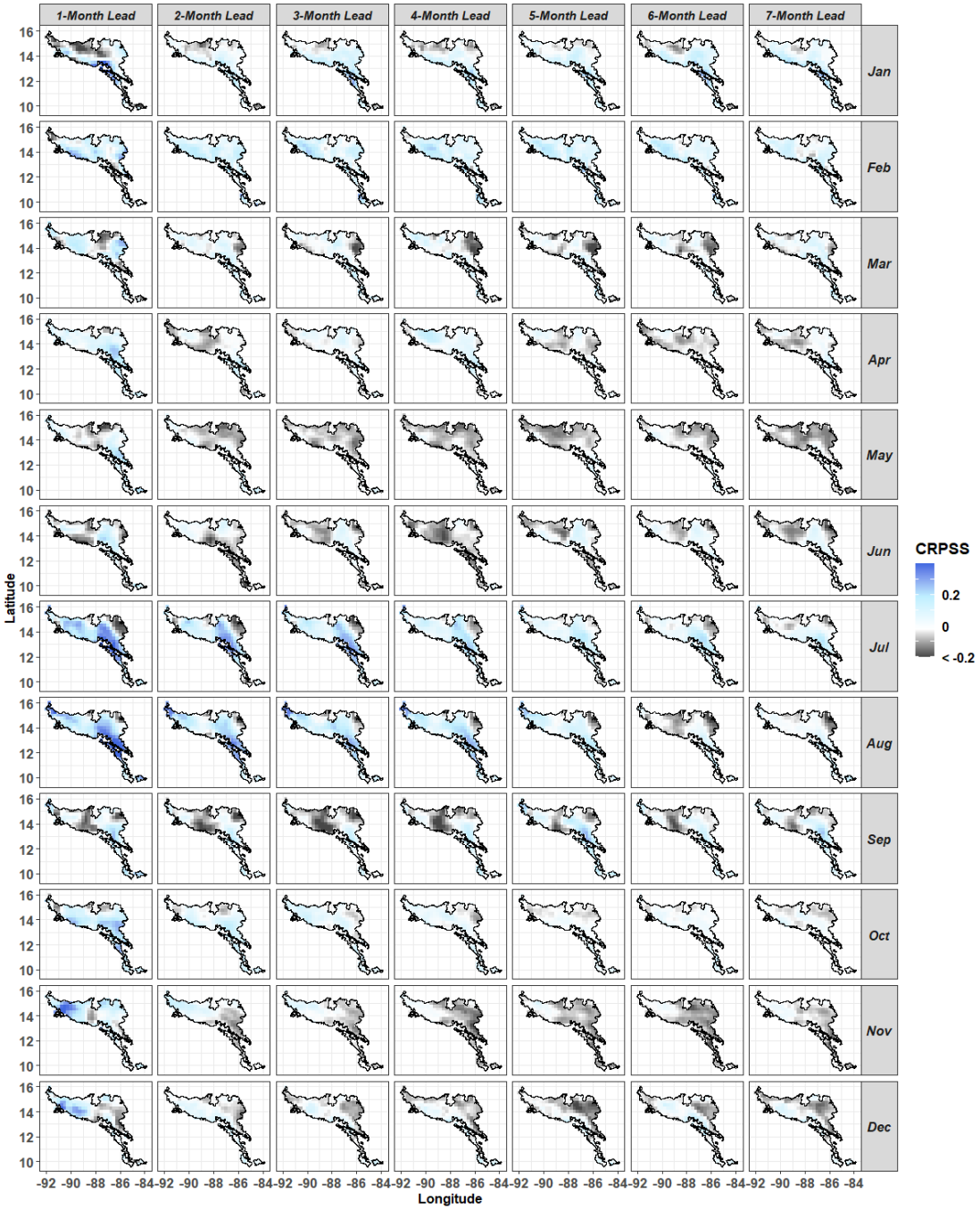


Figure 4.3. Skill plotted spatially (CRPSS) across months and lead times. Skill equal to using the climatological mean (1982–2016) is in white.

and August), skill is positive across almost the entire CADC with peaks in performance in the southeast (Nicaragua and Costa Rica). In the early wet season, positive skill is less widespread. Spatial skill maintains across multiple lead times in June primarily in the central CADC (Honduras), while skill in May is worse (close to zero or negative) across all lead times and locations except for in the southeast CADC (skill is positive at one-month lead time in Nicaragua and Costa Rica). In the late wet season, positive skill is concentrated in the southeast CADC over multiple lead times in September while skill better maintains across the region in October.

4.3.2 Low-Precipitation Anomalies

SEAS5 skill (HSS – 1 = perfect, 0 = skill of a random forecast) in predicting low-precipitation anomalies varies seasonally (Figure 4.4). Detections of the lowest 20% of monthly precipitation anomalies (Figure 4.4a) are best during July and August in the wet season and SEAS5 also achieves positive skill during May and October across multiple lead times for this threshold. In contrast, lower skill is more common in June and September (Figure 4.4a). When the low-precipitation events are constructed using the 0-10% and 10-20% range of low-precipitation anomalies (Figure 4.4b and 4.4c), SEAS5 skill is less seasonally consistent across lead times, but peaks and troughs in skill still occur in similar months. For the lowest 10% of anomalies especially, July, May and October stand out as high-performance months, whereas for the 10-20% range of monthly anomalies, there is less consistency across months and lead times (Figure 4.4b and 4.4c). Limited consistency in the 10% ranges is likely due to fewer values in the dataset being examined, as 10% of a 35-year period equates to only 4 events when the data is examined by month, whereas the lowest 20% of the data includes about 7 events per month.

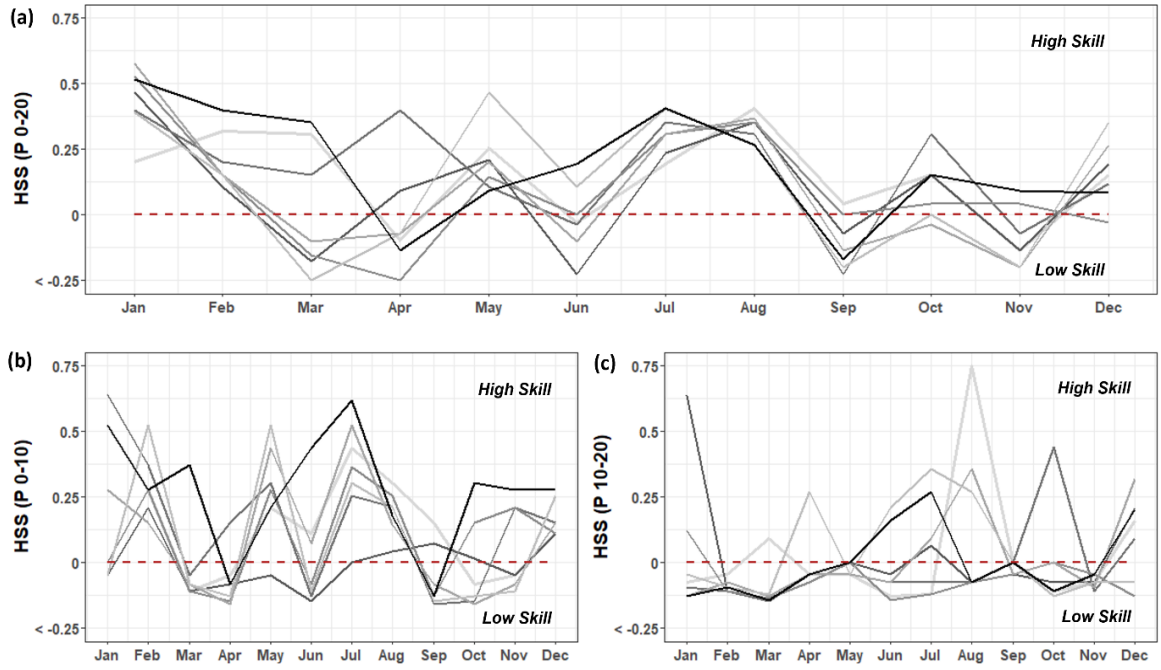


Figure 4.4. Spatially averaged skill (HSS) is plotted over months and shaded in lighter hues of grey as lead time increases for the lowest 20% of monthly anomalies (a), 0–10% of monthly anomalies (b), and 10–20% of monthly anomalies (c). Difference between the 10 and 20% monthly thresholds in absolute precipitation ranges from 14 to 24 mm/month in the wet season. One-month lead time is highlighted in black, and the skill of using a random forecast is plotted as a dashed horizontal line.

4.3.3 Historical Drought Events

Figure 4.5 highlights SEAS5 accuracy (CRPS – 0 = perfect, > 1 = inaccurate) in forecasting historical drought events that had documented drought-related impacts to the agricultural sector in the CADC. In agreement with the comparison across the entire time series, SEAS5 often performed worse (CRPS > 1 in multiple years, e.g. 1997, 2001, 2010) in the earlier wet season (April-June – AMJ, May-July – MJJ, June-August – JJA). In contrast, SEAS5 accuracy was higher (CRPS < 0.6) for multiple lead times in the mid to late wet season (July-September – JAS, August-October – ASO, September-November – SON). In addition, during the driest anomaly periods in the historic drought years (shaded in grey – Figure 4.5), SEAS5 inaccuracies (CRPS > 1) primarily occurred in the early wet season (e.g. in 1997 and 2013), whereas SEAS5 was more accurate across multiple lead times (CRPS < 0.4) when dry anomalies occurred in the late wet season during multiple historical drought years (e.g. in 2009, 2010, and 2014). Forecast quality during the historic drought events also varied across ENSO phases, and no single phase showed a marked improvement over the others. Another noteworthy aspect of Figure 4.5 is that not all reported years with drought-related impacts to the agricultural sector experienced precipitation deficits over one standard anomaly below the climatological mean, meaning other factors (e.g. temperature) were likely also contributing to drought severity during those times.

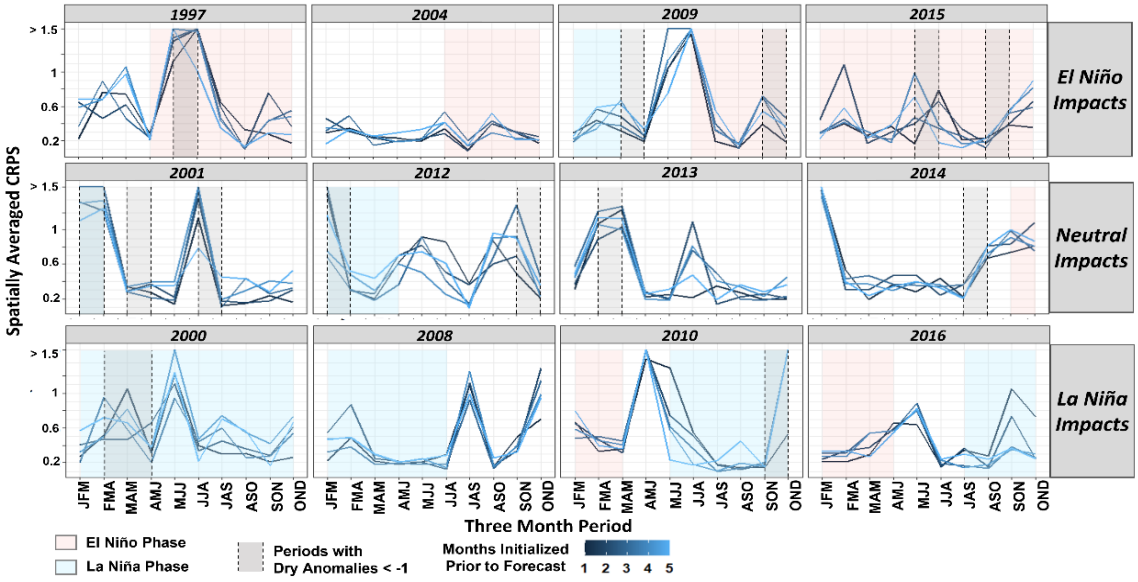


Figure 4.5 Spatially averaged accuracy during historic drought events (CRPS – 0 is perfect) plotted over 3-month periods and colour-coded by number of months SEAS5 is initialized prior to a hindcast. Drought years are categorized by the ENSO phase when impacts primarily occurred (impacts primarily in El Niño phases on top, in Neutral phases in the middle, and in La Niña phases on bottom). Changes in ENSO phase are shaded based on ONI anomalies-El Niño, La Niña, Neutral (no shading). Dry extremes (monthly precipitation < −1 below climatological mean from 1982 to 2016) are highlighted between vertical dashed lines.

Figure 4.6 shows SEAS5 bias across the region in two of the historic drought years that occurred in different phases of ENSO – a strong El Niño phase (2015) and a Neutral phase (2001). These years were selected to illustrate how SEAS5 could have performed: (i) in a typical drought during a strong El Niño phase (traditionally associated with drought in the

CADC) that exhibited decreasing moisture in some subregions accompanied by strong easterly winds and a change in the position of the Intertropical Convergence Zone among other atmospheric anomalies (e.g. IMN, 2015; Sánchez-Murillo et al., 2017); and (ii) in a more atypical drought (2001), when mechanisms other than ENSO may have driven the event and the drought impacts manifested in part due to the irregular spatial distribution of the precipitation rather than a uniform deviation from the climatological normal (CEPAL, 2002).. Forecasting bias is reported when the system is initialized one month and four months prior to a forecast to illustrate how SEAS5 would have performed just prior to the season and at longer lead times (e.g. a one-month forecast initialized in March would combine a one-month lead time for April, a two-month lead time for May, and a three-month lead time for June – Figure 4.6).

As shown in Figure 4.6, SEAS5 predicted the dry anomalies in 2015 relatively well during the second half of the wet season (June-August and August-October), sometimes running even drier than the observations when the reference data was more than one standard anomaly below the climatological normal (e.g. in August-October). SEAS5 was less able to represent the dry anomalies in 2001, especially in the central and northwest CADC (Honduras and Guatemala) but showed minimal bias during the second half of the wet season at one-month initialization. During the early wet season (April-June), SEAS5 overpredicted precipitation in both cases, especially at the longer lead times. When precipitation returned to above normal conditions, the forecasting system also underpredicted precipitation and ran dry – for instance, SEAS5 continued to predict below-normal precipitation in the regions that returned to wetter than normal in August through October in 2015 (Figure 4.6). SEAS5 therefore would have been better at alerting people to a dry event in the mid to late wet season in both years but would have been less useful for strategically capitalizing on the sporadic above-normal precipitation conditions that manifested in these relatively dry periods.

Some subregions would have benefitted more than others from SEAS5 predictions during these years. In Honduras, for instance, the government declared a state of emergency in June due to continued precipitation deficits across the region (FAO, 2015b). SEAS5 demonstrates small bias in Honduras, slightly overrepresenting the magnitude of the dry conditions in this region during the wet season in 2015 (Figure 4.6). In El Salvador, locals also faced acute drought impacts due to irregular precipitation in 2015 (FAO, 2015a), but the SEAS5 hindcasts were less able to predict the precipitation deficit during the early wet season at longer lead times and then predicted below-normal precipitation during the late wet season (August-October) when El Salvador experienced a return to above-normal conditions. Overall, the spatial plots of these two years align with some of the other historic drought cases (Figure 4.5), which show SEAS5 was often less accurate primarily during the early wet season in the El Niño phase droughts and had a more sporadic performance across lead times and months for the Neutral phase droughts.

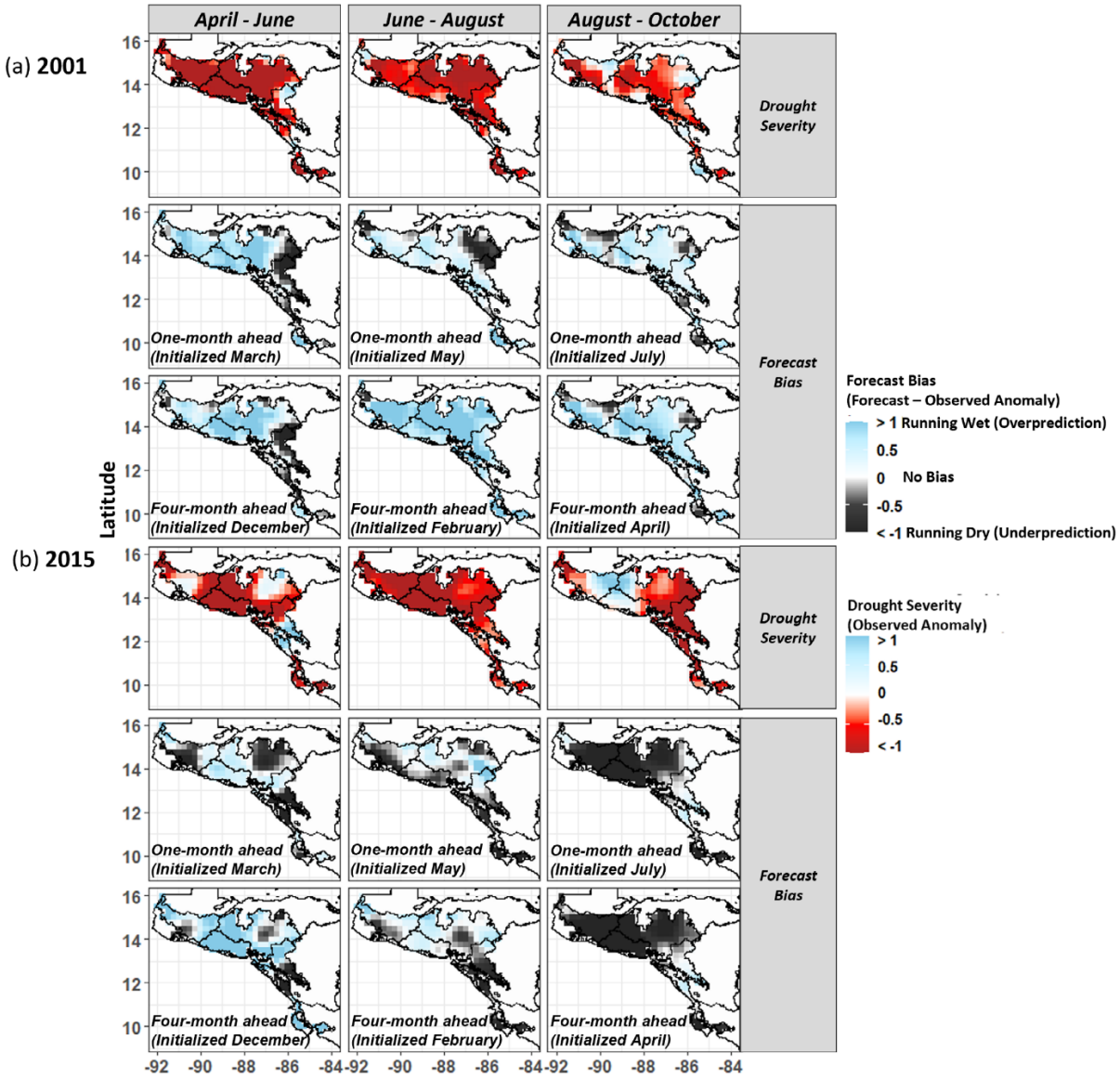


Figure 4.6. Bias in two historical drought events is plotted spatially over 3-month mean periods at different stages of the wet season (a) 2001, Neutral ENSO phase and (b) 2015, El Niño ENSO phase). Early wet season is on the left (April–June), mid wet season is in the middle (June–August), and late wet season is on the right (August–October). The drought severity is plotted using observed precipitation anomalies (based on 1982–2016 time period) on the top row of each panel and compared against the bias of the forecast system below using two initialization periods (row one of the forecast bias is at 1-month initialization and row two of the forecast bias is at 4-month initialization).

4.3.4 Spatial Associations with SEAS5 Skill: Elevation and Continentality

SEAS5 skill is concentrated in the northwest (Guatemala) and in the southeast (Nicaragua, Costa Rica, and part of Honduras) when the months are pooled (Figure 4.7). Neither continentality (distance from the coast) nor elevation play a decisive role in determining the spatial variance in skill. Skill is worse, for instance, both near the coast in El Salvador and further inland in Honduras. Skill is also better in the southeast, which is composed primarily of intermediate and lower elevations (<250 mASL to 750 mASL), and in the northwest, which has higher elevations (e.g. 1250-2000 mASL near -90', 14.25' coordinates). These differences are not necessarily due to differences in rain gauge support across the region. For

instance, while Honduras tended to experience more changes in rain gauge support over the 1982-2016 period relative to other countries, spatial skill varies across the country and the differences do not align with changes in rain gauge operations.

Together, continentality and elevation explain at least a quarter of the spatial variance in skill when skill (using MSE-SS) is regressed against the two factors (Table 4.3). The R-squared value of the linear model ranges between 0.26 at one-month lead time to a maximum of 0.41 at three-month lead time. Continentality has a small but significantly negative relationship with forecast skill across lead times, while elevation has an even smaller and more inconsistent relationship with skill that is negative at lead times beyond three months. Although relatively few rain gauges are in operation at the extreme ends of the elevation range across the CADC, especially above 2000 mASL, when the regression was performed solely for locations with elevations less than 2000 mASL, the findings had no significant changes, meaning the few gauges in operation above 2000 mASL are not biasing the rest of the sample.

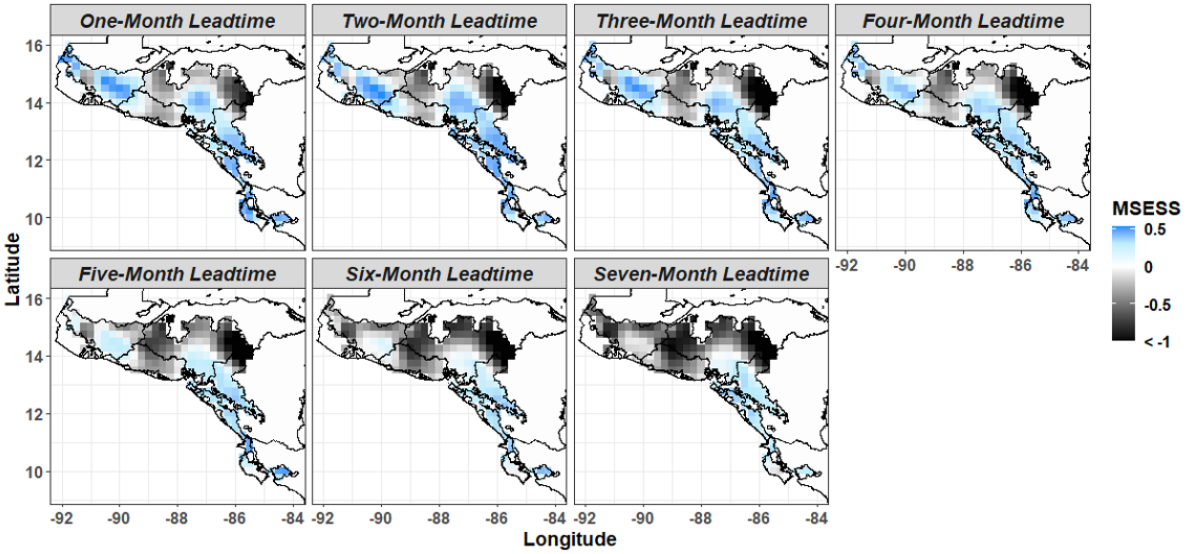


Figure 4.7. Annual skill (pooling 12 months) is plotted spatially at 1-month lead time each using MSE-SS. Positive skill is shaded in blue and negative skill is shaded in grey. Skill equal to using a climatological mean is in white.

Table 4.3. Regression of spatial variance in skill (MSE-SS) against elevation and continentality across lead times pooling months. R-Squared of multiple linear regression (Skill ~ Elevation and Continentality and R coefficients of individual linear regressions (e.g. Skill ~ Elevation) are reported. Significance p-value: *** < 0.001 ** < 0.01 *< 0.05

Lead time (months)	1	2	3	4	5	6	7
Model R-Squared (skill ~ elevation + continentality)	0.26***	0.35***	0.41***	0.38***	0.36***	0.32***	0.32***
Coefficient R of skill ~ elevation	0.00002***	0.000003***	0.00001***	-0.000009***	-0.0001***	-0.0002***	-0.0003***
Coefficient R of skill ~ continentality	-0.004***	-0.005***	-0.006***	-0.006***	-0.006***	-0.006***	-0.005***

4.3.5 ENSO Associations with SEAS5 Accuracy

The strength of the relationship between SEAS5 accuracy and ENSO variability varies seasonally (Table 4.4). In the mid to late wet season, MEI strength does not explain the accuracy of the forecasting system, as the model R-Squared remains below 0.10 across lead times in July through October (except for in July-August at five-month lead time). SEAS5 skill is therefore not significantly better or worse in any particular phase of ENSO in these months. May and June, however, stand out as one period when MEI is not only significantly correlated with skill, but as MEI becomes stronger, forecast accuracy also decreases (CRPS increases), meaning strong El Niño phases are significantly worse times for predictions in this period as compared to other months (Table 4.4). The traditionally recognized relationship between ENSO and precipitation (El Niño events lead to drought in the CADC – e.g. (Peralta Rodríguez, Carrazón Alocén and Andrés Zelaya Elvir, 2012; IMN, 2019) may not be as strong during these months, as MEI is also less correlated with monthly precipitation in May through June compared to other months in the wet season with a model R-squared of 0.09 compared to 0.27 (July-August) and 0.19 (September-October) (Table 4.4). These early wet season months also tend to occur during the transitional phases of ENSO, when ENSO is known for being more difficult to predict, commonly referred to as

Table 4.4. Associations between ENSO variability (using MEI) and the spatially averaged precipitation in different periods of the year (left-hand column); and regression of the spatially averaged forecast accuracy (using CRPS) against ENSO variability (using MEI) across different lead times. Significance p-value: *** < 0.001 ** <0.01 * <0.05

Bi-Month Period: Precipitation ~ MEI	**CRPS ~ MEI**	**Number of Months Model is Initialized Prior to Forecast**					
		1	**2**	**3**	**4**	**5**	**6**
January - February Model R-Squared: 0.12 R Coefficient: -0.32***	Model R-Squared	0.04	0.04	0.02	0.03	0.03	0.02
	R Coefficient	-0.09***	-0.09***	-0.06***	-0.07***	-0.07***	-0.06***
March - April Model R-Squared: 0.10 R Coefficient: 0.30***	Model R-Squared	0.00	0.02	0.01	0.01	0.00	0.01
	R Coefficient	0.02***	0.06***	0.04***	0.03***	0.02***	0.04***
May- June Model R-Squared: 0.09 R Coefficient: -0.30***	Model R-Squared	0.15	0.21	0.28	0.26	0.25	0.26
	R Coefficient	0.14***	0.19***	0.27***	0.27***	0.23***	0.25***
July - August Model R-Squared: 0.27 R Coefficient: -0.49***	Model R-Squared	0.01	0.02	0.03	0.07	0.18	0.12
	R Coefficient	-0.03***	-0.06***	-0.06***	-0.09***	-0.14***	-0.11***
September - October Model R-Squared: 0.19 R Coefficient: -0.41***	Model R-Squared	0.00	0.00	0.00	0.01	0.00	0.00
	R Coefficient	-0.02***	0.01	0.02***	0.04***	-0.02***	-0.02***
November - December Model R-Squared: 0.02 R Coefficient: -0.38***	Model R-Squared	0.00	0.00	0.01	0.00	0.02	0.01
	R Coefficient	0.02***	0.01*	0.04***	0.03***	0.07***	0.04***

the spring predictability barrier (Duan and Wei, 2013; Clarke, 2014; Lai, Herzog and Graf, 2018; Chen et al., 2020). Duan and Wei (2013) also showed that development phases into El Niño events also had worse predictability on average than for corresponding transitions into La Niña phases.

4.4 Discussion

4.4.1 Seasonal and Lead Time Variability in SEAS5 Skill

This analysis focused on the ensemble and mean of SEAS5 hindcasts. Cases when the system outperforms the skill of the climatological mean are noteworthy in part because they suggest that post-processing could effectively be employed to further heighten skill in the region (Slater and Villarini, 2018). A relatively consistent seasonal variation in skill maintains across normal and low precipitation periods (Figures 4.2 – 4.4). SEAS5 can be best used to inform decisions in the mid to late wet season (July – October) when skill is often positive up to four- or five- month lead times for normal precipitation variability and low-precipitation forecasts (Figures 2 and 4). September and May stand out as two exceptional cases from this pattern that show contrasting skill depending on the magnitude of precipitation predicted. In May, SEAS5 is positive across multiple lead times for low-precipitation variability (Figure 4.4) despite only achieving negative skill for normal precipitation variability (Figure 4.2). In September, the opposite occurs, and SEAS5 low-precipitation forecasts are much worse than its predictions of normal precipitation variability (Figures 4.2 and4. 4). The overall seasonal variability aligns with findings from other studies of the previous version (S4) – Carrao et al. (2018), for instance, demonstrated relatively lower skill in May and June relative to July and August across much of Central America when measuring S4 skill for predicting SPI at three-month lead time.

Regarding lead times, SEAS5 may maintain positive skill for longer periods than its predecessor. Dutra et al. (2014) illustrate a relatively steep drop in skill beyond three-month lead times when testing S4 predictions of SPI three-month mean values in September and October for the Central America region, whereas in this analysis, skill of monthly precipitation forecasts also declines for up to three-month lead times in these months but then recovers at five-month lead times in September (using both MSE-SS and CRPSS) and maintains positive skill up to seven-month lead times in October (using CRPSS) for predictions of normal precipitation variability (Figure 4.2). Although by no means is this a one-to-one comparison with S4 since three-month mean SPI and monthly precipitation are different metrics, positive spatially averaged skill extending beyond three-month lead times during most of the wet season is a good sign for progress in forecast system development. Regional variability also affects SEAS5 usability, however, as positive skill maintains across multiple months primarily in the southeast (Nicaragua and Costa Rica), whereas other regions (e.g. El Salvador) experience negative skill for most lead times and months (Figure 4.3).

4.4.2 Associations between SEAS5 Skill and Regional Climate Mechanisms

Understanding why SEAS5 performed well in some cases and not others is more complex. Peak wet-season skill during July and August is possibly tied to the occurrence of the mid-summer dry period, a regular break in wet-season precipitation that is sometimes referred to

as the Mid-Summer Drought (Magaña, Amador and Medina, 1999). This annual break in precipitation is a consistent phenomenon during July and August that many forecasting systems can represent regardless of their ability to model the underlying causal mechanisms (Rauscher et al., 2008). Going forward, further evaluations of forecast skill in this period may be worthwhile, however, as mid-summer dry-period consistency may not maintain in the future. (Anderson et al., 2019), for instance, found an increase in duration of the mid-summer dry period, and (Rauscher et al., 2008) predicted it will occur earlier in the wet season. Multiple causal mechanisms in both the Atlantic and Pacific Oceans likely drive the decrease in precipitation during this period (Mapes, Liu and Buenning, 2005; Rauscher et al., 2008; Karnauskas et al., 2013; Hidalgo et al., 2015; Maldonado, Alfaro and Hidalgo, 2018; Anderson et al., 2019), and changes in any one of them could affect predictability. The timing and location of both the Atlantic and Pacific components of the Western Hemisphere Warm Pool (WHWP - Wang and Enfield, 2001), for instance, affect precipitation over both eastern and western Central America (Durán-Quesada et al., 2017; Sori et al., 2015; Wang et al., 2008) by modulating the magnitude of the mid-summer dry period (Wang et al., 2008). Shifts in either of these warm pool components could affect the predictability of mid-summer dry-period precipitation and would be useful to monitor.

The role the ENSO teleconnection plays in SEAS5 skill is also an ongoing discussion. Previous studies have shown that the ENSO teleconnection often drives precipitation and drought predictive skill of dynamic forecasting systems like SEAS5 (e.g. (Yuan and Wood, 2013; Mo and Lyon, 2015; Scaife et al., 2019; Gubler et al., 2020). The ENSO teleconnection alone may not drive SEAS5 forecasting skill though – while SEAS5 forecasting skill was largely restricted to strong El Niño phases in the majority of South America, for instance, the system showed slight improvements over using a statistical ENSO model when predicting precipitation in multiple locations, meaning other sources of predictability were also driving forecasting skill in that region (Gubler et al., 2020). We wanted to explore how this teleconnection may relate to forecast skill in the CADC and identify if there are times of the year when the association may change. Prior to conducting this analysis, we hypothesized that SEAS5 would be more accurate at forecasting precipitation when ENSO variability and precipitation have a close association. We found that ENSO variability is less associated with precipitation during the early wet season than in the late wet season (Table 4.4), which is also when SEAS5 skill tends to be worse (Figures 4.2 – 4.4, and historical drought examples in Figures 4.5 and 4.6). These predictability challenges may be associated with the spring predictability barrier, which is a time when ENSO is often transitioning between phases and is more difficult to predict (Duan and Wei, 2013; Clarke, 2014; Chen et al., 2020).

Other sources of predictability than the ENSO teleconnection may need more emphasis, especially during the early wet season. During the early wet season (May-June), strong El Niño phases are also significantly negatively associated with forecast skill compared to other phases of ENSO (Table 4.4), meaning the traditional logic (El Niño phases lead to drought in the CADC) that typically drives forecast skill may not hold as strongly during this period and could even undermine the performance of the forecasting system if it relies too much on ENSO predictability during this time of year. Other regional mechanisms affect precipitation

and may need greater representation in the model to enhance skill. For instance, changing Atlantic SST causes changes to moisture transport into the northern Caribbean and Nicaraguan area (Enfield and Alfaro 1999). If there is less evaporation from the Caribbean Sea, air masses that reach this sector of the CADC experience a precipitation deficit and together with high temperatures enhance the soil moisture feedback leading to drought conditions (Barrantes 2019).

We also see the associations between regional topography and forecast skill potentially relating to the strength of ENSO's teleconnection effects on precipitation across the CADC. While regional topography and continentality only explain some of the spatial variance in skill (Table 4.3), this relationship may matter more for subregions of the CADC when these factors intermingle with the effects of ENSO. Annual spatial skill, for instance, is consistently poor across the majority of El Salvador (Figure 4.7). Moisture transport from the Caribbean Sea is lower and relatively constant over El Salvador, in part due to the local topography shielding against the easterly trade winds, while Honduras, Nicaragua, and Guatemala experience more intense moisture transport processes that are seasonally dependent and tied to the position and intensity of the CLLJ (Durán-Quesada, Gimeno and Amador, 2017). Because CLLJ strength plays a key role in moisture transport over the Pacific Slope (Hidalgo et al. 2015) and is modulated by ENSO variability (Durán-Quesada et al., 2020), the regions that experience the most intense fluctuations in precipitation due to CLLJ and are not as shielded by their topography, may be more predictable using the ENSO teleconnection.

To continue to improve seasonal forecasts, especially in the early wet season when skill is relatively low, other physical drivers originating in both the Atlantic and Pacific Oceans may need more emphasis relative to ENSO variability in the eastern Pacific. For instance, variations in the North Tropical Atlantic affect regional precipitation (Alfaro, 2007), especially in the early wet season (Maldonado et al., 2017). SST differentials are worth exploring further, as previous studies have enhanced precipitation predictability in Central America by accounting for differences in SST between the Pacific and Atlantic Oceans instead of focusing on predicting the variability of one ocean alone (Alfaro, 2007; Hidalgo et al., 2013; Alfaro et al., 2018). Alfaro et al. (2018), for instance, demonstrated positive skill when predicting normal precipitation variability and 20^{th} percentile precipitation anomalies during May through June at one- to two- month lead times in Central America using multiple predictors in both the Pacific and Atlantic Oceans. Exploring other associations between landcover and regional dynamics with forecast skill may also be worthwhile, as this study showed that the spatial variance in skill is only partially explained by elevation and continentality.

Examining precipitation characteristics at higher temporal resolution would also be useful. Martinez et al. (2019, 2020), for instance, used precipitation data at relatively high temporal resolution within the Intra Americas Seas region (including eastern Central America) to compare seasonal precipitation characteristics and found distinct climate subregions that exhibit different seasonal precipitation patterns, including differences in mid-summer dry period characteristics. Although eastern Central America and the broader Intra America Seas region are subject to a different interplay between climate mechanisms than western Central

America (in part due to closer proximity to the Caribbean and Atlantic Oceans – e.g. Martinez et al. 2019, 2020; Durán-Quesada et al. 2020), exploring daily precipitation data across the CADC and other parts of western Central America would also further the understanding of how the predictability of the mid-summer dry period and other seasonal precipitation characteristics varies within the region.

Considering SEAS5 within the broader suite of preparedness strategies will also be important for deployment possibilities. As shown by the analysis of historical drought events (Figure 4.5), not all drought-related impacts are caused by extreme precipitation anomalies. Other factors like temperature can modulate drought severity (Van Loon, 2015), and social and economic characteristics can heighten vulnerabilities to hazard exposure (Adger, 2006; O'Brien et al., 2007; Pérez Briceño et al., 2016; Alfaro, Hidalgo and Pérez Briceño, 2018). Peréz-Briceño et al. (2016), for instance, demonstrated how socioeconomic factors significantly affect the spatial distribution of hydrometeorological impacts in Central America. Precipitation forecasts alone will not ensure drought preparedness in the region, but they could feed into ongoing strategies (e.g. Red Cross Climate Centre Strategy for 2021-2025) that build resilience to drought and other extreme weather events at multiple levels. Considering their usability for different sector-specific cases will also be important to enhance broader drought preparedness across the CADC.

4.5 Conclusions

This study showcased the spatiotemporal variability of SEAS5 skill in predicting monthly precipitation variability with a focus on low-precipitation anomalies. SEAS5 demonstrates positive skill primarily in the mid to late wet season (July-October) out to at least four-month lead times in the southeast CADC (Nicaragua and Costa Rica). Precipitation forecasts matter during this time of year, as many applications including agriculture (a dominant sector in the region) rely on precipitation accumulation for success. The spatial variance in skill is partially explained by elevation and continentality, and lower skill in the early wet season (May-June) is linked with weaker associations between ENSO variability and precipitation relative to the second half of the wet season. Compared to its predecessor (S4), SEAS5 continues to show improvements in its predictive skill across lead times over the region. As an open-access seasonal forecasting resource, SEAS5 presents a potentially useful option for more regional water-based operations and it is worthwhile to continue to explore how the spatial and temporal accuracy of systems like SEAS5 vary over lead-times, relate to important regional drivers like ENSO, and compare against other drought prediction tools to support drought preparedness over the CADC.

A Comparison of Seasonal Rainfall Forecasts over Central America using Dynamic and Hybrid Approaches from C3S and NMME

CHAPTER 5

Co-authors. Louise Slater, Alan García-López, and Anne F. Van Loon
Published. 23 December 2022 https://doi.org/10.1002/joc.7969

Motivations

Several questions arose from the first assessment, including the need to further understand how AOGCMs compare across Central America. One of the challenges of interpreting dynamic ensemble evaluation studies is that they often use different metrics or time periods of analysis, and hindcast length can significantly affect forecast skill assessments (Tippett, Goddard and Barnston, 2005; Shi et al., 2015). This can make it difficult to choose which models to use for regional forecast applications. Examining multiple models together also enables investigation into the skill of the models when they are combined. Because MME skill can outperform the skill of using one AOGCM alone (Elvidge et al., 2023), it was also desired to test potential increases in forecast skill by combining models from both the C3S and NMME ensembles. This chapter uses a more operational lens to support existing regional forecast applications by testing the model forecasts as they are deployed today (e.g. by the CA-COF).

Four of the five central research questions (RQs) are addressed in Chapter 5 (RQs 1-4, Thesis Outline 1.2). The entirety of Central America is examined, including the CADC (RQ1), and MMEs are tested for their detection rates of high and low rainfall extremes using HSS (same metric as in Chapter 4). The AOGCMs and combined MMEs are compared against each other as they are used in current operational practice by the CA-COF (RQ2). The seasonal forecasting periods were chosen because of their operational value. Although the months overlap (e.g. May June July – MJJ and June July August – JJA), the forecasts are still distinct because they are initialized on different dates.

Additionally, this chapter contextualizes ensemble forecast value relative to other approaches (RQ3). A more relevant operational baseline was desired to test against the MMEs by using statistical forecasts based on ENSO (a well-known climate phenomenon over the region; e.g. IMN, 2015; 2019). Rather than testing models against a climatology or a random forecast, statistical forecasts are generated using ENSO and TNA to showcase when the MMEs provide comparative skill over using simpler forecasting techniques. Alternative forecasting techniques are also compared using hybrid forecasts. RQ4 is then also briefly addressed by examining how the forecasts vary over the wet season as compared to the relative strengths of ENSO and TNA.

Two points are worth noting regarding the format of this chapter. First, of the ten AOGCMs that are compared, CFSv2 and CanSIPS-IC3 are listed as contributors to NMME rather than C3S to avoid double counting the AOGCMs in the MMEs. Second, the nomenclature for lead times used in this chapter (0,1,2 months) differs from the other chapters in this thesis (1,2,3 months). The International Research Institute for Climate and Society (IRI) format is loosely followed here, which labels lead time months as 0.5, 1.5, etc…, and these lead times were shortened to the leading number, while the other chapters in this thesis follow the C3S approach (1,2,3, etc…).

Abstract

Seasonal rainfall forecasts provide information several months ahead to support decision-making. These forecasts may use dynamic, statistical, or hybrid approaches, but their comparative value is not well understood over Central America. This study conducts a regional evaluation of seasonal rainfall forecasts focusing on two of the leading dynamic climate ensembles: the Copernicus Climate Change Service seasonal forecasting system (C3S) and the North American Multimodel Ensemble (NMME). We compare the multi-model ensemble mean and individual model predictions of seasonal rainfall over key wet season periods in Central America to better understand their relative forecast skill at the seasonal scale. Three types of rainfall forecasts are compared: direct dynamic rainfall predictions from the C3S and NMME ensembles, a statistical approach using the lagged observed sea surface temperature (SST), and an indirect hybrid approach, driving a statistical model with dynamic ensemble SST predictions. Results show that C3S and NMME exhibit similar regional variability with strong performance in the northern Pacific part of Central America and weaker skill primarily in eastern Nicaragua. In the northern Pacific part of the region, the models have high skill across the wet season. Indirect forecasts can outperform the direct rainfall forecasts in specific cases where the direct forecasts have lower predictive power (e.g. eastern Nicaragua during the early wet season). The indirect skill generally reflects the strength of SST associations with rainfall. The indirect forecasts based on Tropical North Atlantic SSTs are best in the early wet season and the indirect forecasts based on Niño 3.4 SSTs are best in the late wet season when each SST zone has a stronger association with rainfall. Statistical predictions are competitive with the indirect and direct forecasts in multiple cases, especially in the late wet season, demonstrating how a variety of forecasting approaches can enhance seasonal forecasting.

5.1 Introduction

Central America has been identified as a climate change hotspot at high risk of hydrometeorological extremes (Almazroui et al., 2021; Giorgi, 2006; Taylor et al., 2012). Drought and flooding events have devastated local communities in recent years (CEPAL, 2002; Marengo et al., 2014; Calvo-Solano et al., 2018; Guevara-Murua et al., 2018; Beveridge et al., 2019; Gotlieb et al., 2019). These events are unlikely to abate (Almazroui et al., 2021; Hannah et al., 2017; Hidalgo et al., 2013; Imbach et al., 2018), which makes extreme weather preparedness an important challenge going forward. Drought and flooding impacts are modulated by many factors, including social and economic conditions (e.g. Perez-Briceno et al., 2016; Alfaro Martínez, Hidalgo León and Pérez Briceño, 2018). Early warnings based on forecasting do not provide a comprehensive solution to mitigate these impacts but can be one factor to reduce vulnerability when effectively communicated (Braman et al., 2013; Coughlan de Perez et al., 2016; Domeisen et al., 2022; Goddard et al., 2020; Golding et al., 2019; Kreibich, et al. 2017b; White et al., 2022). Forecasts at the seasonal scale are useful to inform planning for the upcoming season, including decisions like crop choice and water supply and demand measures. Multiple efforts are underway to enhance seasonal forecast efficacy over Central America, providing stakeholders with rainfall information several months ahead. The Central American Climate Outlook Forum (CA-COF), for instance, provides regional rainfall outlooks for national and sector-specific planning (Alfaro et al., 2016; Donoso & Ramirez, 2001; Garcia-Solera & Ramirez, 2012). These outlooks are disseminated through several public forums, such as the Mesa Técnicas Agroclimáticas (MTAs) in Guatemala, which are local technical agroclimatic forums where farmers can interact with scientists to learn more about upcoming seasonal rainfall (INSIVUMEH, 2022c).

Regional evaluations of seasonal forecasts can help stakeholders choose which models are most effective to be used in these forums and identify opportunities to improve forecast skill. There are multiple ways to generate seasonal rainfall forecasts, including dynamic, statistical, and hybrid approaches (Hao, Singh and Xia, 2018). Dynamic forecasting approaches employ the predictions of physically-based climate models (General Circulation Models - GCMs) that are initialized at different states of the climate system (Bauer, Thorpe and Brunet, 2015). These models are currently used by the CA-COF, which often uses GCMs from the North American Multimodel Ensemble (NMME - Kirtman et al., 2014). Other dynamic forecasting systems are also publicly available, including the European ensemble from the Copernicus Climate Change Service (C3S - **https://cds.climate.copernicus.eu/**).

Statistical methods have also been used to estimate regional rainfall. Some studies for instance use Canonical Correlation Analysis (CCA) to relate observed SST in the Atlantic and Pacific Oceans to regional rainfall, demonstrating the important role different SST fields play in rainfall over Central America (Alfaro, 2007; Alfaro et al. 2016; Giannini, Kushnir and Cane, 2000; Maldonado et al., 2013, 2016, 2017). Hybrid methods then combine dynamical and statistical methods (Slater et al., 2022). Some hybrid methods generate indirect rainfall forecasts by extracting related variables (e.g. SST) from the GCMs and then statistically translate those values to target variables like rainfall (e.g. Alfaro et al., 2018;

Strazzo et al., 2019; Colman, Graham and Davey, 2020). Indirect forecasts could perform well, as forecasts of large-scale variability like SST over tropical domains are often a strength of the GCMs as compared to their forecasts of more variable parameters like rainfall (Barnston et al., 2012, 2019; Saha et al., 2014). For example, Saha et al. (2014) found that one contributing model to NMME (CFSv2) had anomaly correlations with SST in the Niño 3.4 region of around 0.82, while the average correlation with observed Northern Hemisphere precipitation rate over land is equal to 0.12. Using the GCMs for their strengths (e.g. SST estimates) and statistically translating those to a rainfall forecast could potentially improve on their direct rainfall forecast skill.

In this paper we compare the following methods: dynamical methods that take rainfall forecasts directly from GCMs (direct), statistical methods based on observed SST values months ahead of a season (statistical), and hybrid methods that generate forecasts indirectly by using the GCMs to predict SST over a target period and then statistically translate those values to a rainfall forecast in that same period (indirect). Evaluation of seasonal rainfall predictions over Central America is needed because GCMs that perform well globally may not necessarily capture the climate mechanisms necessary to forecast rainfall well over a given region or the complex orography of regions like Central America, for instance, may require higher spatial resolution for improved skill. Some evaluations of GCMs have showcased their potential over Central America and in nearby regions (e.g. Kirtman et al., 2014; Weisheimer and Palmer, 2014; Carrão et al., 2018; Khajehei, Ahmadalipour and Moradkhani, 2018; Slater, Villarini and Bradley, 2019; Becker, Kirtman and Pegion, 2020; Gubler et al., 2020; Kowal et al., 2021). Very few of these studies target Central America specifically and they typically compare one model, ensemble, or forecasting method, using different forecast verification metrics and time-periods of evaluation. Furthermore, many of these studies baseline the model skill against a climatological mean or a random forecast, which is standard practice for many skill metrics. It is relatively easy to outperform the climatological mean or a random forecast, so although multiple studies may show models outperform the climatological mean, for instance, it is difficult to choose between different models and methods on offer without one evaluation that compares them using the same time frame and verification metrics.

Effective forecasts over Central America need to capture the region's seasonal rainfall cycle (Figure 5.1a). This cycle follows a bimodal distribution with peak rainfall primarily occurring in June and September typically separated by the mid-summer dry period in July and August (Figure 5.1a; Magaña, Amador, and Medina, 1999). This mid-summer dry period arises due to a decrease in nearby SST, which limits deep convection activity and enables trade winds to intensify (Magaña, Amador and Medina, 1999). The seasonal distribution in rainfall also varies regionally with more extreme rainfall events often occurring on the Pacific slope during August-October as compared to the Caribbean slope (Peterson et al., 2002; Taylor and Alfaro, 2005). The variability and intensity of seasonal rainfall is not uniform across the isthmus (e.g. Muñoz-Jiménez et al., 2019, see Figure 5.1a/c for plots of annual and monthly rainfall). For instance, the mid-summer dry period is more intense in the Central American Dry Corridor (CADC - a regional drought hotspot - Gotlieb et al., 2019), as compared to other locations on the isthmus (Figure 5.1a).

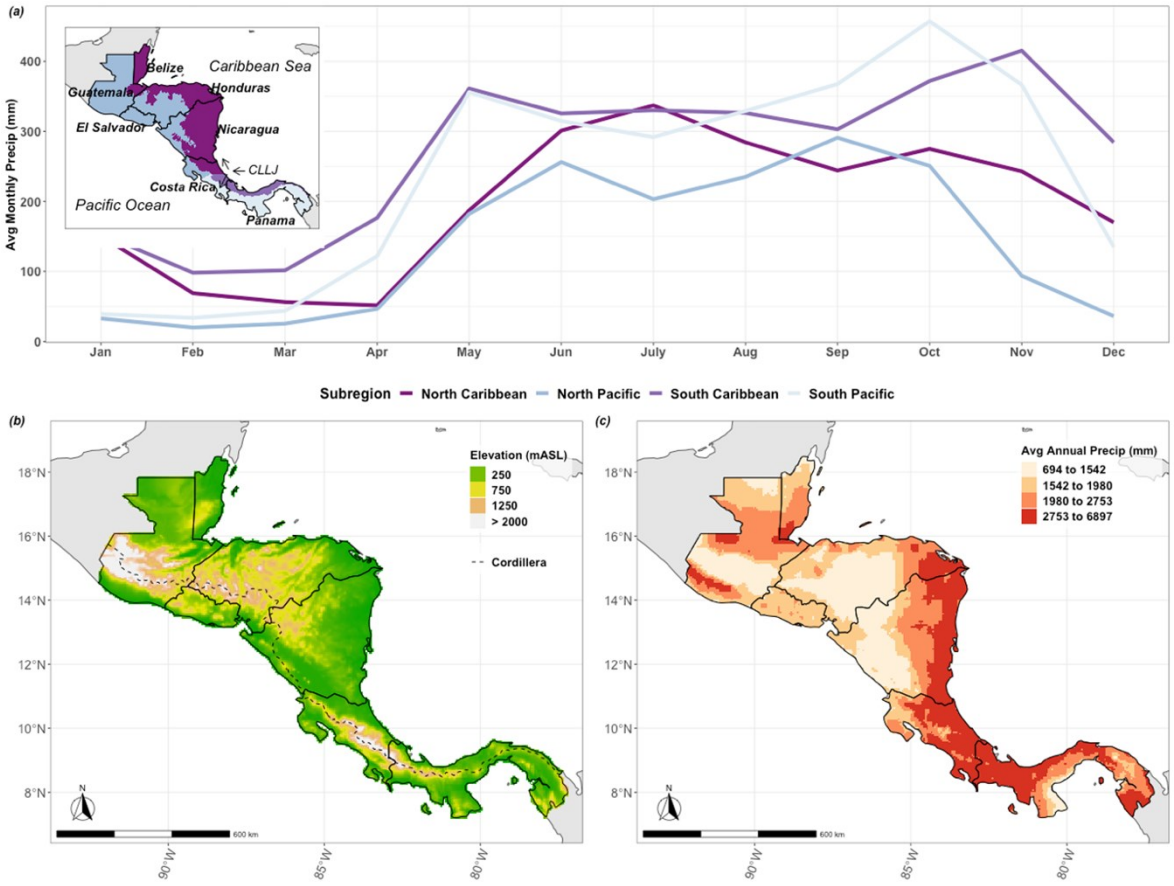


Figure 5.1. Central American climate. (a) Spatially-averaged monthly rainfall between 1993-2016 (using CHIRPS) plotted by subregions, which are tagged in the inset map. The southern part of the isthmus is separated below the 10ºN latitude line when the Caribbean Low Level Jet (CLLJ) branches (Cook & Vizy, 2010; Hidalgo et al., 2015; Muñoz et al., 2008; Wang, 2007) and divided into a Caribbean and Pacific regime. The North Pacific includes the Pacific side of the Cordillera, the locations in Guatemala not located in the Caribbean climate regime as delineated by INSIVUMEH based on the Thornthwaite index (e.g. INSIVUMEH, 2022a), and the CADC, all of which experience a similar monthly rainfall distribution over the wet season. The CADC boundary is delineated here based on the climate risk index that includes relatively contiguous locations with a dry season that lasts over 4 months (IICACR, 2014). The North Caribbean then lies on the Caribbean side of the Cordillera excluding the additional locations demarcated as the North Pacific. (b) Elevation measured in meters above sea level (mASL) plotted with the Cordillera mountain range. (c) Average annual rainfall (CHIRPS) from 1993-2016 plotted spatially.

Central America is an important region for a regional forecast evaluation, as it potentially can have both high forecast skill due to its location in the tropics, and low forecast skill because it is a relatively thin land bridge between two oceans and has relatively steep changes in terrain from the Cordillera (Figure 5.1b). One tropical teleconnection is the El Niño Southern Oscillation (ENSO) in the eastern Pacific (Amador, 2008; Amador et al., 2006; Durán-Quesada et al., 2017; Mariotti et al., 2020; Poveda et al., 2006; Spence et al., 2004; Waylen et al., 1994). ENSO is the dominant climate mechanism affecting rainfall over Central America through its modulation of moisture transport mechanisms like the Caribbean Low Level Jet (CLLJ) (Spence, Taylor and Chen, 2004; Durán-Quesada, Gimeno and Amador, 2017). Although ENSO is a well-known indicator of rainfall over Central America, and its phases are publicized by meteorological bulletins to inform stakeholders (e.g. IMN, 2019), this phenomenon is unlikely to provide the full rainfall picture. El Niño

phases of ENSO only explain some of the drought events in the CADC, for instance, and drought events differ in intensity between similar El Niño phases (Muñoz-Jiménez et al., 2019; Kowal et al., 2021). The isthmus is affected by a complex interaction of weather patterns that arise from processes in both the Pacific and Atlantic oceans (Durán-Quesada et al., 2020), including seasonal migration of the inter-tropical convergence zone (ITCZ), tropical cyclones, and movement of the Atlantic warm pool (Amador, 2008; Amador et al., 2006; Durán-Quesada et al., 2017, 2020; Enfield & Alfaro, 1999; Hidalgo et al., 2015; Poveda & Mesa, 1999; Sori et al., 2015; Wang, 2007; Wang & Enfield, 2001).

Evaluating whether the GCMs can improve on an ENSO-based forecast is useful. ENSO is a known driver of skill of many GCMs (Gubler et al., 2020; Mo & Lyon, 2015; Scaife et al., 2019; Yuan & Wood, 2013), and it is worthwhile to explore how the skill of the direct and indirect forecasts differ from a statistical forecast based on ENSO alone. Some assessments indicate ENSO alone does not drive GCM skill (e.g. Gubler et al., 2020; Zhao et al., 2021). In their evaluation of SEAS5, the seasonal forecasting system produced by the ECMWF, Gubler et al. (2020) used an ENSO-based statistical forecast to test against SEAS5 performance over South America and demonstrated that SEAS5 skill differed from the ENSO-based statistical forecast. Focusing on CFSv2, Zhao et al. (2021) also showed that the ENSO teleconnection contributed to some of the skill of CFSv2, especially in southern Central America in DJF, but that it was not a significant contributor to that model's skill in all parts of the isthmus, especially during other seasons when the ENSO teleconnection was less prominent. Both examples indicate that the GCMs can derive skill from other climatological mechanisms than ENSO, meaning they may outperform forecasts that only use this variable.

SST in the Atlantic is also important to regional rainfall (Alfaro, 2007; Enfield & Alfaro, 1999; Giannini, Kushnir and Cane, 2000; Sori et al., 2015; Spence et al., 2004; Taylor et al., 2002; Wang et al., 2008; Waylen & Quesada, 2001), as it modulates the magnitude of the mid-summer dry period (Wang, Lee and Enfield, 2008; Maldonado et al., 2016), among other influences over the region. Tropical North Atlantic anomalies (TNA; Enfield and Alfaro, 1999), for instance, are one of the inputs the CA-COF uses to inform rainfall outlooks (Alfaro et al., 2016). A systematic regional evaluation across multiple models and approaches compared to a statistical model that includes ENSO and TNA SSTs could inform how the skill of the GCM-based forecasts (direct and indirect) differs from the performance of a statistical forecast that uses common SST indices known to influence regional rainfall.

This study explores how two of the leading dynamic forecasting ensembles, NMME and C3S, compare at seasonal rainfall prediction over Central America, considers how best to use their outputs, and investigates some of the climate mechanisms that may drive skill variability over the region. We do this by (1) evaluating how the skill of direct, indirect, and statistical forecasts varies spatially and temporally during key wet season periods, (2) comparing the relative value of using the C3S and NMME ensembles for direct or indirect forecasts to identify cases when they are significantly different from each other and a purely statistical forecast, and (3) exploring why some GCMs may perform better than others by investigating the relationship between the models' rainfall skill and the skill of their predictions of SST in some key climate regions.

5.2 Data and Methods

5.2.1 Seasonal Forecast Scope

We focus on seasonal rainfall forecasts during key wet season months (May-October) across Central America, using three-month seasonal totals from the forecasting models initialized one month prior to the forecast, as is done operationally for rainfall outlooks produced by the CA-COF (Alfaro et al., 2016; Donoso & Ramirez, 2001; Garcia-Solera & Ramirez, 2012). We choose three periods of interest that are used by the CA-COF to assess different stages of the wet season: May-July (MJJ; early wet season), June-August (JJA; includes mid-summer dry period), and August-October (ASO; late wet season). Seasonal forecasts are generated by taking the total rainfall over three-month periods. For instance, a forecast of total seasonal rainfall from MJJ is initialized in April, essentially adding rainfall from a one-month lead time for May, a two-month lead time for June, and a three-month lead time for July. Tercile deterministic forecasts, like the ones produced by CA-COF where predictions are made for below normal, normal, and above normal rainfall, are assessed using 33% and 66% thresholds in the seasonal rainfall distribution. Predictions of total seasonal rainfall above the upper 90% and below the lowest 10% thresholds are also assessed to better understand how predictive skill may change for rainfall extremes. We then assess the forecasts spatially and by subregion (as defined in Figure 5.1a), comparing performance between the North Pacific, North Caribbean, South Pacific, and South Caribbean.

5.2.2 Dynamic Model Selection and Reference Data

This evaluation focuses on two multi-model ensembles: C3S and NMME. The European ensemble, C3S, has become publicly available as of 2018 and is hosted by the ECMWF at the Climate Data Store at daily and monthly time-scales with 1-6 month lead times (https://climate.copernicus.eu/seasonal-forecasts). The leading North American ensemble, NMME (Kirtman et al., 2014), is hosted by the International Research Institute for Climate and Society at Columbia University (IRI) and is publicly available up to 11 month lead times at monthly time scales (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). Several models contribute to both ensembles, of which 10 are selected for this evaluation (5 from NMME with 73 total members and 5 from C3S with 148 total members for hindcast analysis- see Table 5.1). The NMME models used in this study include multiple Phase II models that are operationally available and two newer updates since the 2019-2020 Phase II annual report – GFDL SPEAR and Cansips-IC3 (NOAA, 2020). Five C3S models are selected based on publicly available spatial resolution (1º) and consistent initialization (1st of every month), as summarized in Table 5.1. Here, C3S and NMME are compared using the overlap between their publicly available data, which includes monthly temporal resolution and hindcasts from 1993-2016. For the purposes of this evaluation, when we refer to forecasts, we use forecasts as a general term for predictions but are using hindcasts and historical years.

For the observed rainfall data, Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) total monthly rainfall data is used because it is a relatively common observational reference dataset in Central American studies (e.g. Alfaro et al., 2016; Hidalgo, Alfaro and Quesada-Montano, 2017), covers the entire time period of analysis (1993-2016),

and is available at high spatial resolution ~0.05º (Funk et al., 2014). This dataset is compared against Multi-Source Multi-Weighted Ensemble Precipitation (MSWEP - Beck et al., 2017), Global Precipitation Climatology Center (GPCC - Schneider et al., 2018), but the difference in observational datasets is often smaller than the range of the forecasts (Supplementary Information figure 5.A1), and the forecast skill shows similar regional and seasonal variability across observational datasets, so we prefer CHIRPS (Supplementary Information figure 5.A2). Optimum Interpolation Sea Surface Temperature data (OISSTv2 - blended dataset that combines a variety of inputs, e.g. buoys, satellites, Reynolds et al., 2007), is used as the observed SST dataset to validate the ensemble SST hindcasts and generate the statistical predictions.

Table 5.1. Summary of forecasting models in evaluation.

NMME			
Contributing Center	**Model**	**Hindcast Members**	**Reference**
NCAR-COLA/RSMAS	CCSM4	10	(Gent et al., 2011)
ECCC-CMC	Cansips-IC3*	20	(Lin et al., 2021; Merryfield et al., 2013)
NASA-GMAO	GEOSS2S, Version 2	4	(Molod et al., 2020)
NOAA-GFDL	SPEAR	15	(Delworth et al., 2020)
NOAA-NCEP	CFSv2	24	(Saha et al., 2014)
Total Models in Analysis: 5; Total Members in Analysis: 73			
C3S			
Contributing Center	**Model**	**Hindcast Members**	**Reference**
CMCC	System 35	40	(Gualdi et al., 2020)
DWD	System 21	30	(Fröhlich et al., 2021)
ECMWF	SEAS5	25	(Johnson et al., 2018)
Meteo France	System 8	25	(Batté et al., 2021)
UKMO	Glosea6	28	(Davis et al., 2020)
Total Models in Analysis: 5; Total Members in Analysis: 148			

Note: Contributing centre is listed next to model name, total number of members available for hindcasts, along with the reference for the model. Abbreviations: NCAR-COLA/RSMAS, U.S. National Center for Atmospheric Research Center for Ocean Land Atmosphere Studies/Rosenstiel School for Marine and Atmospheric Science at UMiami; ECCC, Environment and Climate Change Canada - Canadian Meteorological Centre; NASA-GMAO, NASA Global Modeling and Assimilation Office; NOAA-NCEP, NOAA National Center for Environmental Prediction; NOAA-GFDL, NOAA Geophysical Fluid Dynamics Laboratory; CMCC, Centro Euro-Mediterraneo Sui Cambiamenti Climatici; DWD, Deutscher Wetterdienst; ECMWF, European Centre for Medium-Range Weather Forecasts; UKMO UK Met Office. *Cansips-IC3 is operationally treated as one but technically made up of the GEM5-NEMO and CanCM4i-IC3 models, each containing 10 members.

5.2.3 Forecast Construction

5.2.3.1 Direct Forecasts from C3S and NMME using pyCPT

To test the ensembles in an operational-type setting, we evaluate the models and ensembles as they are produced with pyCPTv2 (Muñoz et al., 2019), a Python set of libraries that work as a wrapper for the Climate Predictability Tool developed by IRI (we used version 17.6.1 - Mason et al., 2022), and part of the NextGen methodology developed by IRI to improve climate services at subseasonal to seasonal time-scales (Muñoz et al., 2019; Muñoz et al., 2020; WMO, 2020). Here we provide a brief overview of the methodology, but other articles

provide more details on the functionality (e.g. Muñoz et al., 2020; Pons et al., 2021) and applications (Fernandes et al., 2020; WMO, 2020; Fernandes, Bell and Muñoz, 2022) of NextGen. Steps performed in pyCPT include (1) download raw 1º hindcasts of C3S and NMME GCMs from the IRI data library; and (2) calibrate the hindcasts and transform them into 0.25º rainfall predictions over Central America using a Canonical Correlation Analysis (CCA) model (see Figure 5.2 for a flowchart illustrating how we generate, calibrate, and evaluate the predictions). The calibration procedure follows the pattern-based CCA Model Output Statistics methodology described in Mason and Badour (2008) which is designed to

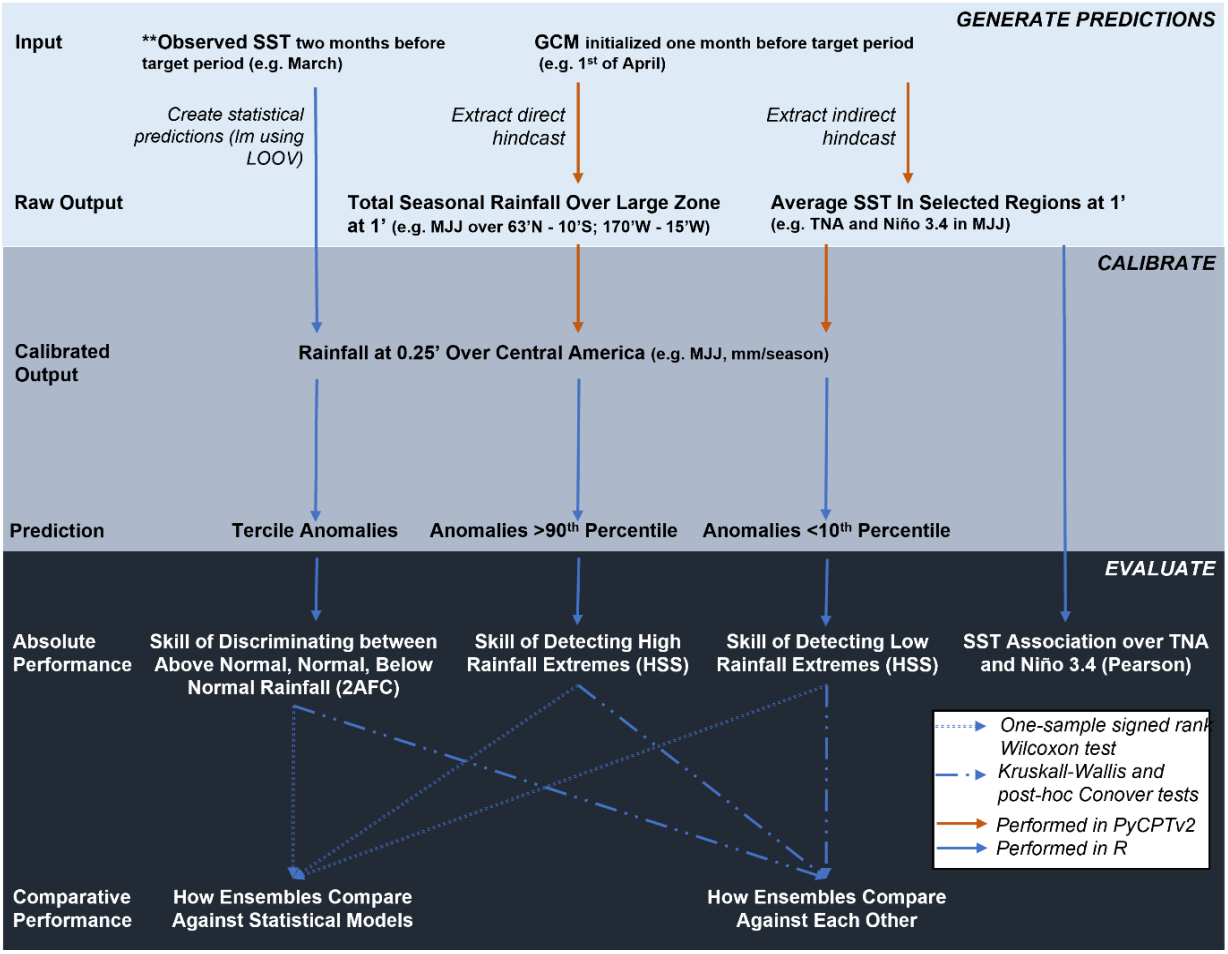


Figure 5.2. Diagram of methods employed to generate forecasts, calibrate them, and evaluate their performance. Statistical models: A linear model (lm) is generated using leave-one-out cross-validation (LOOV) from 1981 to 2016 to train the models to make predictions over 1993–2016 based on 2-month lagged SST for a specified region (e.g., Niño3.4 in March extracted to predict rainfall over May–July, MJJ). Direct models: Rainfall hindcasts are extracted at 1 from 1993 to 2016 over a larger region than Central America and then are bias-corrected and calibrated using a canonical correlation analysis (CCA) model that translates the 1 raw hindcasts into 0.25 hindcasts over Central America. Indirect models: SST hindcasts are extracted at 1 from 1993 to 2016 over a target region (e.g., Niño3.4) and then are bias-corrected and calibrated using a CCA model to then statistically translate the 1 raw SST hindcasts into 0.25 rainfall hindcasts over Central America. Once 0.25 raw predictions are made using each approach, the values are calculated as standardized anomalies and then percentiles of rainfall are extracted depending on the question (e.g., 0.33/0.66 for normal rainfall discrimination, 0.10 for low-rainfall extremes, 0.90 for high-rainfall extremes). Absolute skill is measured using Heidke skill score (HSS), 2 alternative forced choice (2AFC), and Pearson's correlation coefficient (Pearson). Comparative skill is tested by comparing the absolute skill between the models/forecasting approaches using Wilcoxon and Kruskal–Wallis/Conover statistical tests.

maximize the correlation of linear combinations between predictors and predictands, summarised as follows: (1) *correct biases* by applying integrated cosine latitude weighting to both predictor and predictands and transform them into normalized outputs; (2) *reduce dimensionality* of the lat/lon/time data sets into an n points by n years data set and handle missing datapoints (e.g. dropping data points with too many missing years of data and using K-Nearest Neighbours to fill in missing data); (3) *separately fit a Principle Components Analysis (PCA) time series* for predictors and predictands and calculate loadings by latitude and longitude to identify the number of Empirical Orthogonal Function (EOF) modes for predictor / predictand; (4) *fit a CCA model* onto the two PCA time series – one for predictor and one for predictand; (5) *produce a prediction* by calculating the PCA scores for predictor values, apply the CCA prediction with those scores as inputs, and use the CCA loadings to reconstruct the rainfall prediction (for more details on CCA calibration in pyCPT see Martinez et al., 2022; Muñoz et al., 2020).

In practice, these steps are simplified using the pyCPT libraries, where the generation and calibration functions only require identification of the predictor field, the predictor variable (e.g. total seasonal rainfall or mean seasonal SST), the predictand variable (e.g. total seasonal rainfall), the maximum number of EOF modes for both predictor and predictand, and the maximum number of CCA modes to constrain how many comparisons the model runs (we choose maximum EOFx modes = 6, maximum EOFy modes = 5, and maximum CCA modes = 5). For the direct forecasts, we set the predictor field to precipitation with a larger grid box than Central America (63ºN - 10ºS; 170ºW - 15ºW), which is selected based on the locations used for CCA in Maldonado et al. (2013) and includes both Pacific and Atlantic zones of variability. We export the calibrated models into R version 4.1.2 (R Core Team, 2021) to combine, evaluate, and visualize the results using several packages (Grolemund & Wickham, 2011; Hijmans, 2021; Huang & Zhao, 2020; Kassambara, 2020; Mangiafico, 2022; NCAR - Research Applications Laboratory, 2015; Pebesma, 2018; Pedersen, 2022; Pierce, 2019; Pohlert, 2022; Weigel, 2017; Wickham, 2016; Wickham et al., 2019, 2021; Wickham & Seidel, 2020).To create the ensembles, we use an unweighted combination of each calibrated model's mean to align with how the NextGen approach is currently deployed, and because the added benefit of a sophisticated post-processing method like Bayesian Model Averaging is often limited where sample periods are under 30 years (Delsole et al., 2013; Weigel et al., 2010).

5.2.3.2 Indirect and Statistical Forecasts in Known Teleconnection Zones

We generate indirect and statistical forecasts using two known teleconnection regions - Niño 3.4 (170ºW-120ºW, 5ºN – 5ºS) for the ENSO teleconnection (Trenberth, 1997; Trenberth and Stepaniak, 2001) and Tropical North Atlantic (TNA - 55ºW-15ºW, 5ºN-25ºN) for the TNA teleconnection (Enfield et al., 1999). TNA is selected partly because of its relatively strong association with Central American rainfall in the early wet season (Alfaro, 2007; Maldonado et al., 2017; Spence et al., 2004), a time of year when some GCMs have had relatively lower skill (e.g. Kowal et al., 2021) to see how this SST zone could potentially enhance their skill. The indirect forecasts are constructed in pyCPT like the direct rainfall forecasts, but SST is selected as the predictor for either the Niño 3.4 or the TNA region and then the same CCA method is used to transform the raw 1º GCM forecasts to 0.25º and statistically relate the

SST predictors from a selected zone (e.g. Niño 3.4) in a given period (e.g. MJJ) to rainfall over Central America in that period (e.g. MJJ).

For the statistical forecasts we run a linear regression of the observed rainfall against the spatial average of observed lagged (2 months prior) SST in the Nino 3.4 region, the TNA region, and using both regions as predictor variables. This process is done for each season individually and over the entire overlapping period of available CHIRPS/OISSTv2 data up until 2016 (1982-2016), giving the statistical models over 30 years of data for training. The predictions are created using a cross-validation leave-one-out-approach in R (Kuhn, 2022) as described in Mason and Badour (2008). For example, in the Niño 3.4 statistical model to predict precipitation for the year 1993, observed rainfall estimates in MJJ are regressed against the spatially-averaged observed March Niño 3.4 SST across 1982-2016 (without 1993), and 1993 is set as the year to predict. We select observed SST data two months prior to a forecast period to simulate operational forecasting conditions – because the GCMs are initialized on the first of the month (e.g. 1st of April), they do not benefit from the entirety of that month's data for the upcoming season (e.g. MJJ), so the observed SST values selected for the statistical forecast of MJJ rainfall, for instance, are averaged over March instead of April to not give the statistical model extra information the GCMs would not have had.

5.2.4 Evaluation Criteria

For this assessment, we focus on the models' discrimination between tercile categories of rainfall to understand the models as they are deployed today by the CA-COF. We also assess the models' identification rate of rainfall extremes above the 90th percentile and below the 10th percentile to better understand their potential for extreme event detection. We select the two alternative forced choice (2AFC) metric to evaluate tercile forecasts, and the Heidke Skill Score (HSS) for low and high rainfall detection (Table 5.2). 2AFC assesses forecast ability to discriminate between categories and is preferred in part because it is an equitable metric across ordinal categories (treats all random forecasts with the same score - Mason and Weigel, 2009), is relatively common (e.g. Alfaro et al., 2018), and has been identified as a good metric to support decision-making criteria (Weigel & Mason, 2011). HSS is preferred for extreme event detection because the metric is equitable (Hogan et al., 2010), relatively common (e.g. Higgins, Kim and Unger, 2004; Becker et al., 2012; Walker et al., 2019) and simple to understand (0 = skill of a random forecast, positive is better). Before the skill is tested, all the model and observational estimates are transformed into standardized anomalies by season across the hindcast period. For the MJJ period for instance, the mean and standard deviation of the MJJ forecasts from each model is calculated over 1993-2016 and then each estimate is subtracted from the mean and that value is divided by the standard deviation (as done in pyCPT).To assess comparative performance for each season and subregion, we use two nonparametric statistical tests: a one-sample signed rank Wilcoxon test (Wilcoxon, 1945), and a Kruskal-Wallis test (Kruskal and Wallis, 1952) combined with post-hoc testing using the Conover-Iman method (Conover & Iman, 1979; Conover, 1999), which is less well-known than Dunn but has more statistical power (Gilbert, 2019). We select nonparametric tests instead of their parametric counterparts (T-Test and ANOVA) because

the data sample sizes are limited and we do not assume normality or homogeneity of variance within each ensemble spread. To answer our first comparative question – do the ensembles significantly outperform the statistical models or a random forecast, we use the Wilcoxon test. For this test we first identify the skill of the top performing statistical model against the observations and check whether that model is better than a random forecast for a given period/subregion (e.g. North Pacific in MJJ). We then set the best performing statistical model skill as the value to beat (or use the skill of a random forecast) to see if the skill of the ensemble spread against the observations is significantly different from the best statistical model skill. To compare across distributions of ensembles – i.e. C3S vs. NMME, we use a Kruskal-Wallis test to identify if there are any significant differences between the skill of the spread of the ensembles for a given seasonal period and subregion, and then conduct post-hoc testing using Conover when the Kruskal-Wallis test shows there are significant differences warranting pair- wise comparisons between groups to identify top performers. Because we are only interested in identifying the top direct and indirect forecasts, we only perform Kruskal-Wallis/post-hoc testing to compare the ensembles that are found to be significantly better than the best

Table 5.2 Summary of evaluation metrics with their relevant formulas

Metric	Summary	Formula
Heidke Skill Score (HSS)	Skill of categorical event detection (1 = perfect, 0 = skill of random forecast) (Heidke, 1926)	$HSS = \frac{a + d - a_r - d_r}{n - a_r - d_r}$ $a_r = (a + b)(a + c)/n$ $d_r = (b + d)(c + d)/n$ $a = Hits$ $b = False\ alarms$ $c = Misses$ $d = Correct\ rejections$ Computed with 'verification' package in R
Two-Alternative Forced Choice (2AFC)	Discrimination between events, i.e. below normal, normal, above normal (1 = perfect, 0.5 = skill of random forecast) (Weigel & Mason, 2011)	$p2AFC = \frac{\sum_{i=1}^{m_{f-1}} \sum_{j=i+1}^{m_f} n_{0,i} n_{1,j} + 0.5 \sum_{k=1}^{m_f} n_{0,k} n_{1,k}}{n_0 n_1}$ $n_0 = nonevents$ $n_1 = events$ $m_f = number\ of\ forecast\ categories$ $n_{0,i} = i^{th}\ forecast\ issued, event\ did\ not\ occur$ $n_{1,j} = j^{th}\ forecast, event\ did\ occur$ $n_{0,k} = number\ of\ forecasts\ for\ category\ k, no\ event\ occurred$ $n_{1,k} = number\ of\ forecasts\ for\ category\ k, event\ did\ occur$ Computed with 'afc' R package (Weigel, 2017)
One-sample signed rank Wilcoxon test	Nonparametric test if sample median is significantly greater than a known value (Mann & Whitney, 1947; Wilcoxon, 1945)	$Test\ statistic\ (W_1) = \sum R_d^+$ $\mu W_1 = \frac{N_r(N_r + 1)}{4}$ $\sigma W_1 = \sqrt{\frac{N_r(N_r + 1)(2N_r + 1)}{24}}$ $z\ statistic = \frac{W_1 - \mu W_1}{\sigma W_1}$ Null hypothesis (H_0): $z \leq known\ value$ Alternative hypothesis (H_1): $z > known\ value$ R_d^+ $= Ranks\ corresponding\ to\ positive\ difference\ between\ sample\ and\ known\ value$ $N_r = Number\ of\ difference\ scores\ (sample - known\ value)\ not\ equal\ to\ zero$ Computed with 'stats' package in base R (R Core Team, 2021)
Kruskal-Wallis test	Nonparametric comparison of variance to test if multiple samples come from the same distribution (Kruskal & Wallis, 1952)	$H\ statistic = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N + 1)$ Test H statistic against critical chi-square value for N-1 degrees of freedom Null hypothesis (H_0): $population\ medians\ are\ equal\ (critical\ chi - square\ value \geq H)$ Alternative hypothesis (H_1): $population\ medians\ are\ not\ equal\ (critical\ chi - sqaure\ value < H)$ Computed with 'stats' package in base R (R Core Team, 2021)

compared statistical forecasts and a random forecast. We then assess differences in predictability between the ensembles by evaluating the prediction skill of SST in known teleconnection regions using Pearson's R. The SST locations of interest include both the Niño 3.4 region and the TNA region. We also test average association (using Pearson's R) between observed SST and rainfall over Central America over different wet season periods to compare with the SST forecast skill.

5.3 Results

5.3.1 Skill of Discrimination between Above Normal, Normal, and Below Normal Rainfall Categories

The 10 GCMs assessed show a similar seasonal and geographic variation in skill over Central America when forecasting rainfall directly (Figure 5.3). For instance, in eastern Nicaragua, the direct forecasts from the individual models have relatively low skill compared to the rest of the isthmus, while skill is often higher along the Pacific coast (Figure 5.3). Combining the models into ensembles shows similar variation in seasonal/geographic skill variability, and the ensembles tend to perform similarly within a given method (i.e. direct or Niño3.4/TNA indirect; Figure 4). Eastern Nicaragua is still a relatively lower skill zone, and the North Pacific is an area with high skill (Figure 5.4). Looking across forecasting methods (e.g. direct vs. Niño 3.4 indirect), the direct rainfall forecasts showcase higher skill than the indirect forecast skill in most cases except for the early wet season (MJJ) when the indirect forecasts demonstrate better performance compared to the direct forecasts in multiple locations (Figure 4 top row). The TNA indirect forecasts (Figure 4 right group), for instance, have relatively higher skill compared to other forecast methods (Niño 3.4 indirect and direct) over eastern Nicaragua in MJJ, and the Niño 3.4 indirect forecasts have highest skill in Guatemala during this period (Figure 5.4 middle group). While the direct ensemble skill tends to be highest in most other cases, the Niño 3.4 indirect forecasts also have similar skill in the late wet season (ASO) (Figure 5.4 bottom row).

The visual assessments of absolute skill (Figures 5.3 and 5.4) are confirmed when statistically testing the differences between the spatially averaged skill by subregion (Figure 5.5a) and tagging the top performer by location for each seasonal period (Figure 5.5b). On average, the North Pacific is the highest skill zone with the highest skill occurring in JJA (Figure 5a top left panel). The differences in spatially-averaged skill are smaller between the other sub regions, but similar differences still exist – e.g. direct forecasts have the lowest skill in the North Caribbean of all the subregions in JJA (Figure 5.5a top right panel), which is mirrored in the absolute skill plots of the direct forecasts (Figure 5.3). Between the direct and indirect ensembles, the direct forecasts also tend to have one of the highest skill scores in most cases, especially in the North Pacific (Figure 5.5a top left panel). Similar exceptions persist. For instance, a TNA indirect forecast tends to have the highest skill in MJJ in the North Caribbean (Figure 5.5a top right panel and 5.5b left panel). Although the spatially-averaged skill of the Niño 3.4 indirect forecasts are not best in the North Pacific generally in the early wet season (MJJ) (Figure 5a top left panel), the Niño 3.4 indirect forecast still demonstrates more targeted value over Guatemala as the top performer in MJJ (Figure 5.5b left panel). The Niño 3.4 indirect and direct forecasts also tend to have the highest skill in the

South Pacific and South Caribbean, often showing similar skill in ASO (Figure 5.5a bottom two panels).

The direct and indirect forecasts do not consistently outperform the statistical forecasts though for all periods. The forecasts are often better than a statistical forecast based on one teleconnection region alone (e.g. Niño 3.4 or TNA), but the combined statistical model skill is usually the best performer of the statistical models with peak skill in the late wet season (Figure 5.5a). In the North Pacific, for instance, the ensembles are better than the statistical forecasts in the early to middle wet season (MJJ/JJA), but their skill is not significantly better than using a combined statistical forecast based on TNA and Niño 3.4 SST in ASO (Figure 5.5a top left panel). Although direct forecast skill is high in the North Pacific in ASO, the comparative value of using the direct forecasts over this subregion is clearer in the early to middle wet season (MJJ and JJA) when the statistical model skill is lower.

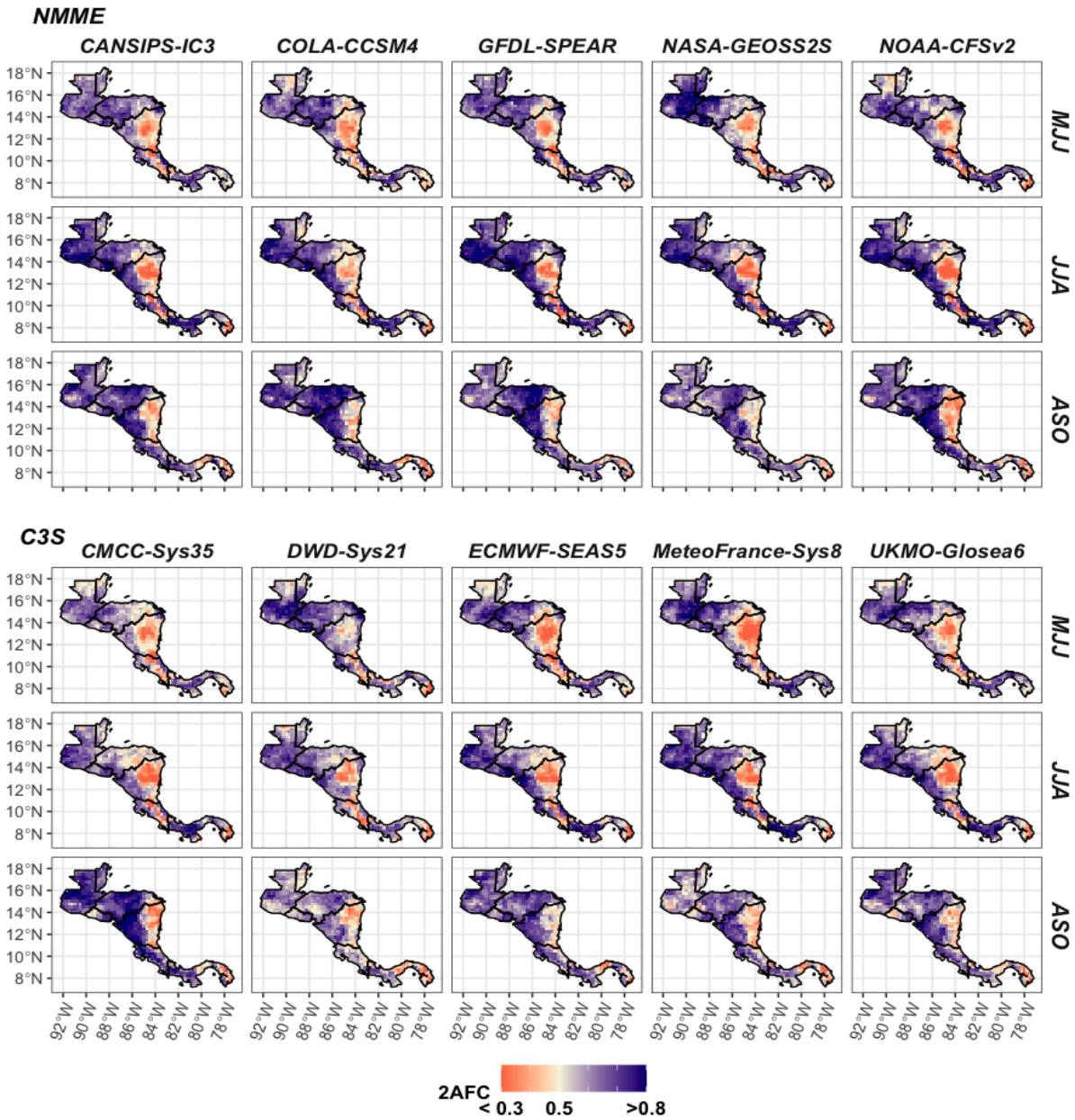


Figure 5.3. **Rainfall skill for direct tercile forecasts (2AFC) across individual models** in the wet season. 0.5 is equal to the skill of using a random forecast.

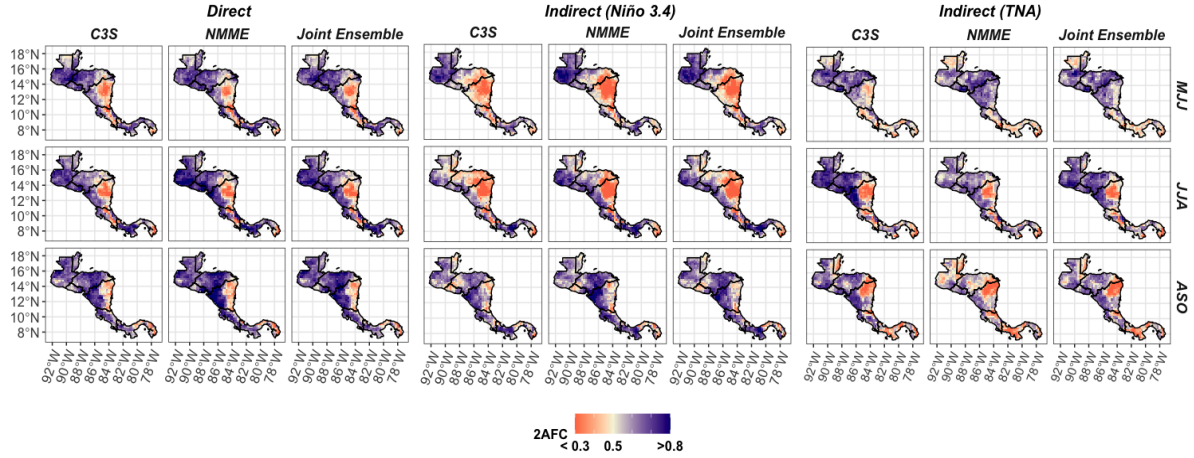


Figure 5.4. Rainfall skill of the ensemble means using C3S, NMME, and the Joint Ensemble.

5.3.1 | Skill of Discrimination between AN, N, and BN Rainfall Categories

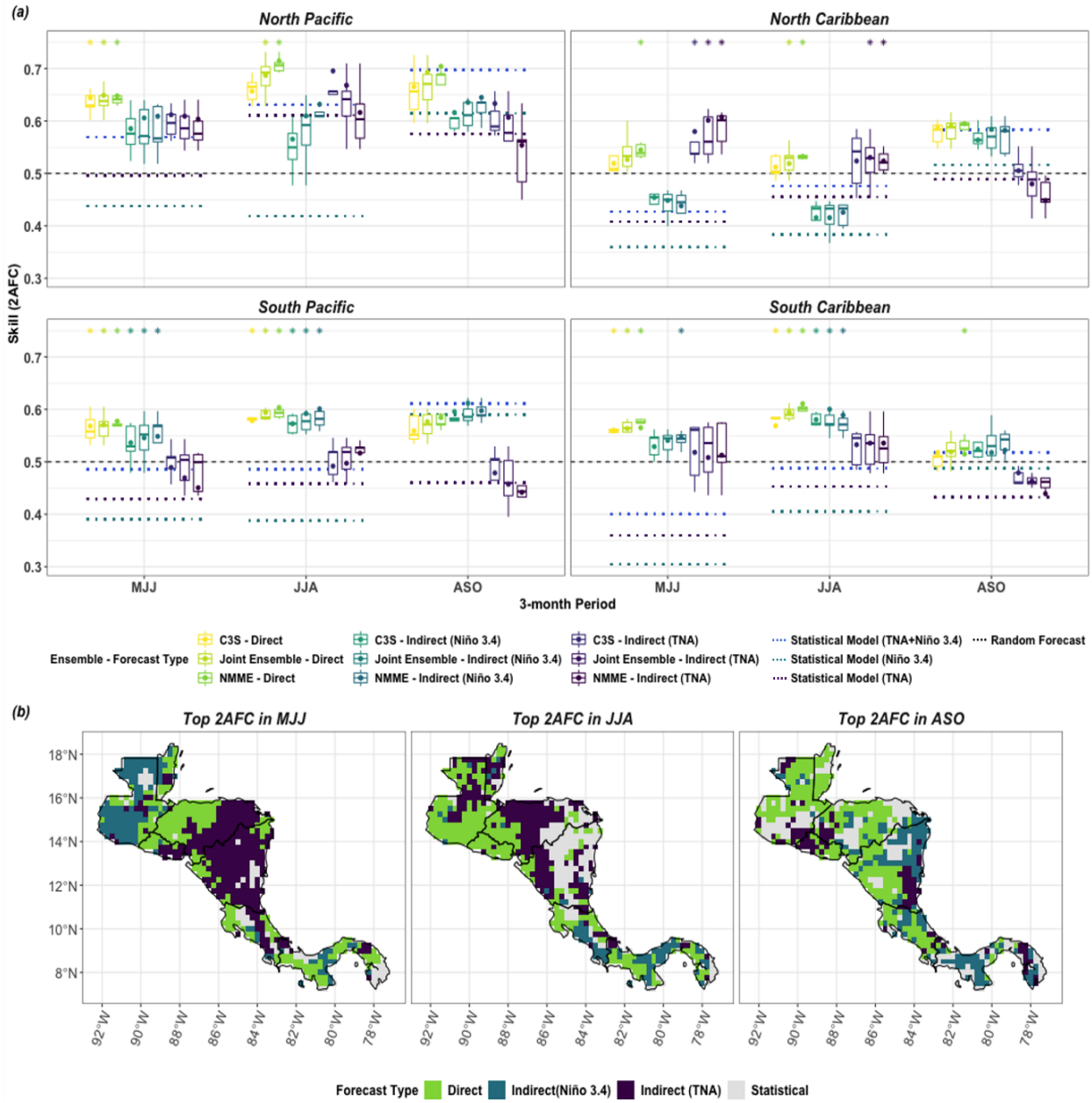


Figure 5.5. Statistical Comparison of Tercile Rainfall Discrimination. (a) Spatially averaged comparison of forecast skill using 2AFC across seasons and grouped by subregion as defined in Figure 1a. Boxplots are categorised by ensemble and show the spread of the individual model means contributing to each ensemble. Dots within boxplots show the skill of the ensemble mean across all contributing models using 2AFC. Dashed lines refer to statistical models (except 0.5 equal to skill of a random forecast). Top performing direct and indirect forecasts are given stars (above the relevant boxplot), which indicate if forecasts meet three criteria: they are significantly better than a random forecast, significantly better than the best statistical forecast, and are not significantly worse than another ensemble, e.g., an ensemble that beats a random and statistical forecast but is significantly worse than another ensemble will not receive a star. (b) Top performing forecasting approach using the 2AFC score grouped by season.

5.3.2. Skill of Detecting High ad Low Rainfall Extremes

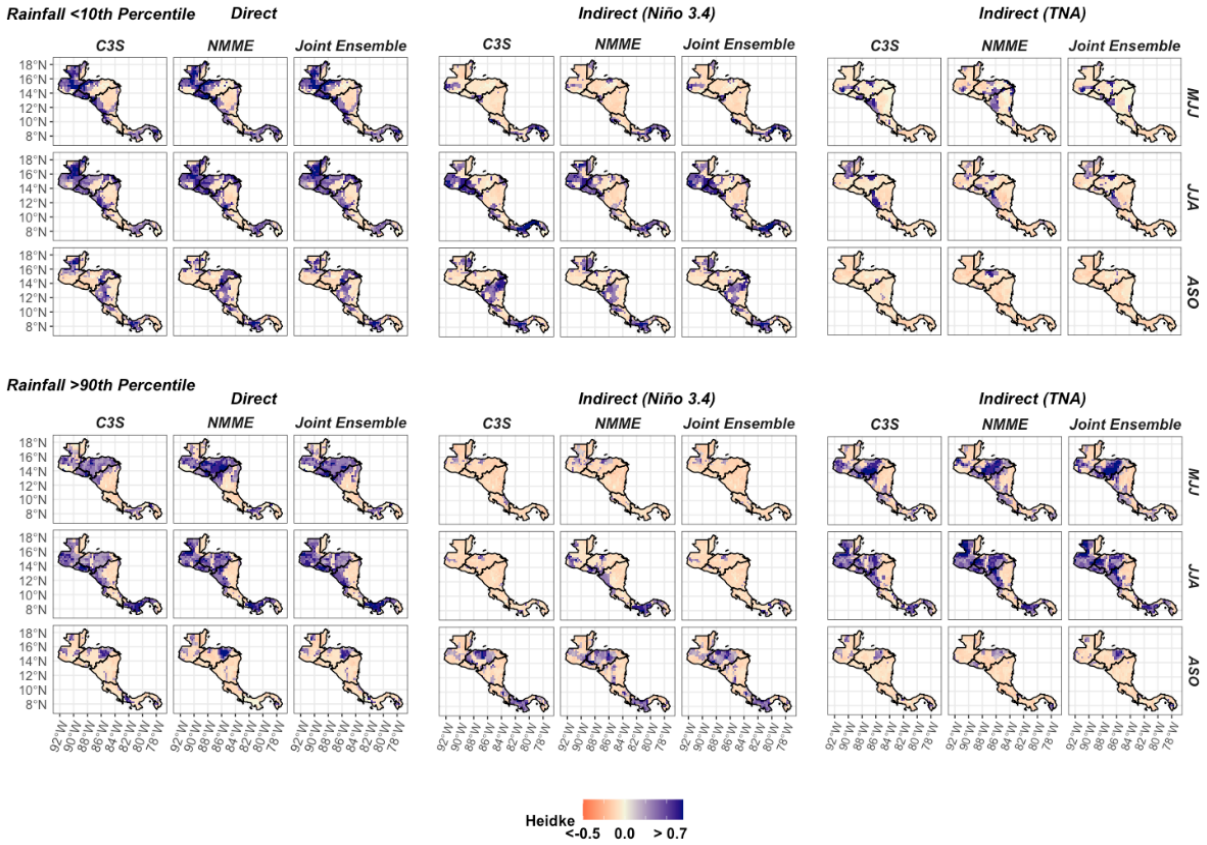


Figure 5.6. Ensemble mean skill for high- and low-rainfall extremes using HSS, where 0.0 is equal to the skill of a random forecast. Skill is plotted over separate parts of the wet season--MJJ, JJA, and ASO.

Ensemble skill is more limited when the models are used to predict extreme rainfall (Figure 5.6 – see Supplementary Information Figure 5.A3 for skill of direct forecasts of rainfall extremes plotted individually by model). In the late wet season (ASO), for instance, skill is almost always zero (Figure 5.6 bottom row of Rainfall <10th Percentile and Rainfall > 90th Percentile groups). Eastern Nicaragua is still a trouble spot for the direct and indirect forecasts (Figure 5.6), but direct forecast skill is still high in much of the North Pacific during the early to middle wet season (Figure 5.6 – left group, MJJ/JJA). When comparing the absolute skill of direct and indirect detection of extreme rainfall, the skill also does not have the same seasonal patterns as for tercile discrimination, emphasizing how the application matters when evaluating forecasting methods. For instance, although a direct forecast tends to have high spatial skill for the detection of low and high rainfall extremes in the early to middle wet season (MJJ/JJA, Figure 5.6 left group), a TNA indirect forecast has higher skill in some areas of Central America for detection of high rainfall extremes in JJA (Figure 5.6 bottom right group), a period when direct forecasts often performed better for tercile rainfall discrimination (Figure 5.5a).

The skill of the ensemble means is similar within forecasting methods (Figure 5.6), and the differences between the two ensembles are often not significant when they are both top performers (Figure 5.7a – e.g. both C3S and NMME direct forecasts get stars in most cases

when a direct method performs well), meaning method choice matters more than ensemble choice for extreme rainfall detection. The skill of extreme rainfall detection has higher variability within ensembles than for discrimination between rainfall terciles though, which makes it less easy to identify a top performing forecasting method for this application. The ensemble mean often outperforms the median of the contributing models (Figure 5.7a), but not as consistently as when the ensemble mean is used to discriminate between terciles (Figure 5.5a). The C3S skill variations exemplify this variability. The median of the C3S ensemble is sometimes higher than the NMME mean (e.g. MJJ in South Pacific Figure 5.7a top panel second from right), the top of the interquartile range of the C3S ensemble is often higher than the top of the interquartile range of the NMME for low and high rainfall extremes even when the mean is lower, and the interquartile spread of C3S is often larger than NMME, meaning some of the C3S contributors are dragging down the score of the C3S mean for this case while others are performing much better (Figure 5.7a).

When compared to the statistical models, the direct and Niño 3.4 indirect forecasts are often significantly better for detecting low rainfall extremes (Figure 7a top panels). The TNA indirect models show significantly better skill than the statistical models in multiple early and middle wet season periods (MJJ/JJA) when detecting high rainfall extremes but are never a standalone method that is significantly better than the others (any time a TNA indirect forecast gets a star, another forecast method also receives a star – Figure 5.7a bottom panels). Generally direct and indirect methods showcase more promise for detections of low over high rainfall extremes – the ensembles significantly outperform the statistical forecasts more often when detecting low rainfall extremes as compared to high rainfall extremes (Figure 5.7a); and the top performing forecast type has more spatial consistency for low rainfall extremes (Figure 5.7b – e.g. direct forecasts over Guatemala in MJJ).

5.3.3 | Predictive Skill of Key Teleconnection Regions: ENSO and TNA

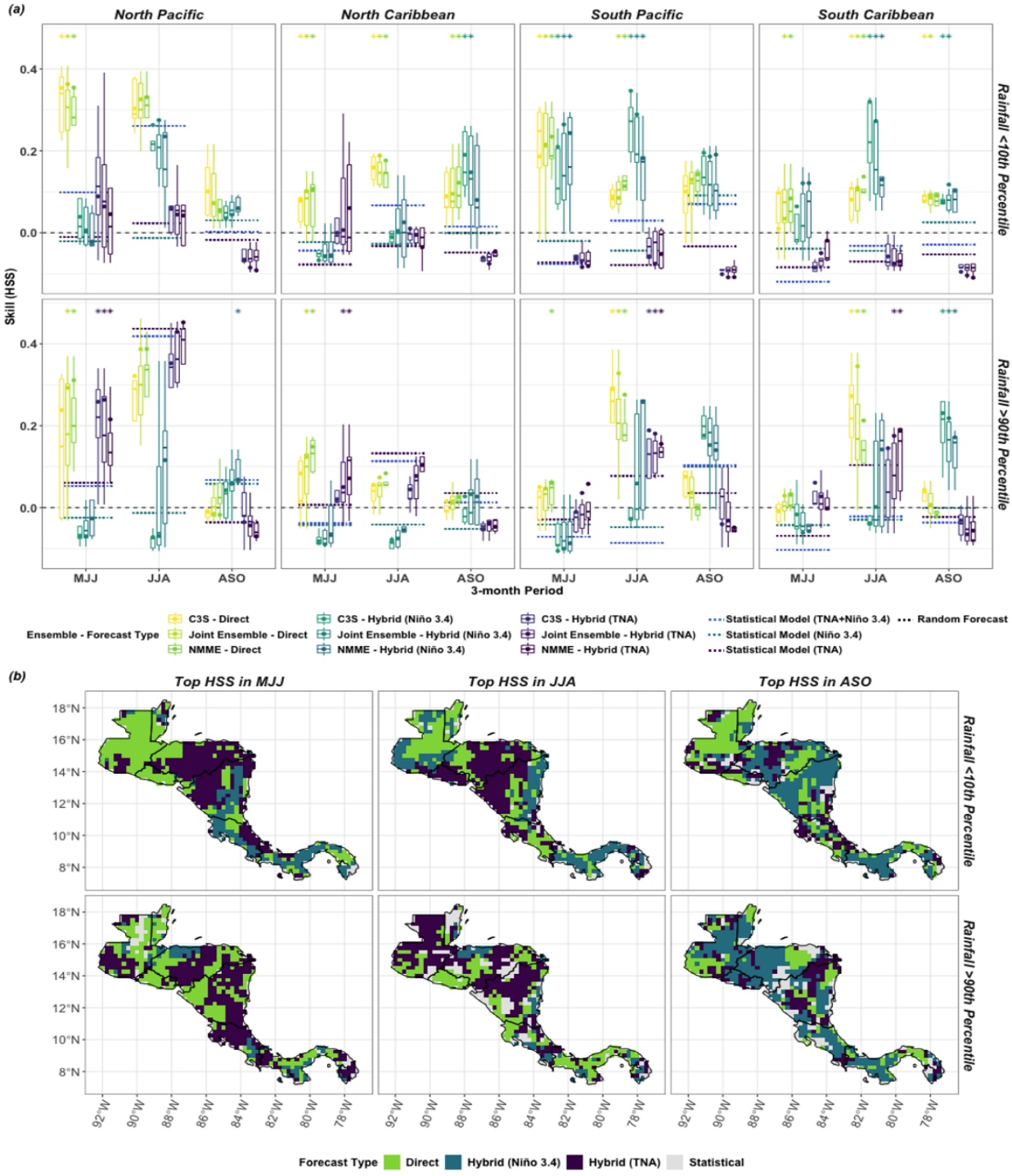


Figure 5.7 Statistical Comparison of Low and High Rainfall Extremes Detection. (a) Spatially averaged comparison of forecast skill for low- and high-rainfall extremes using HSS across seasons and grouped by subregion. Boxplots are categorized by ensemble and show the spread of the individual model means contributing to each ensemble. Dots within boxplots show the skill of the ensemble mean across all contributing models using HSS. Dashed lines refer to statistical models (except 0, skill of a random forecast). Top performing direct and indirect forecasts are given stars, which indicate if forecasts meet three criteria: They are significantly better than a random forecast, significantly better than the best statistical forecast, and are not significantly worse than another ensemble, e.g., an ensemble that beats a random and statistical forecast but is significantly worse than another ensemble will not receive a star. (b) Top performing forecasting approach using HSS grouped by season.

5.3.3 Predictive Skill of Key Teleconnection Regions: ENSO and TNA

The skill of the ensembles' predictions of SST over the Niño 3.4 region is consistently high. The association between each model and the observed Niño 3.4 SST exceeds R=0.7 for most models in the wet season periods, and the association of the ensemble means are between 0.85 and 0.98 for Niño 3.4 (Figure 5.8a). Differences in SST skill over Niño 3.4 therefore are not likely driving the differences in forecast skill over the region. SST skill is lower over the TNA region in the late wet season (Figure 5.8a), but this is also when the TNA region is least associated with rainfall, compared to the earlier months (Figure 8b). The stronger performance of the TNA indirect forecasts correlates with this association between TNA SST

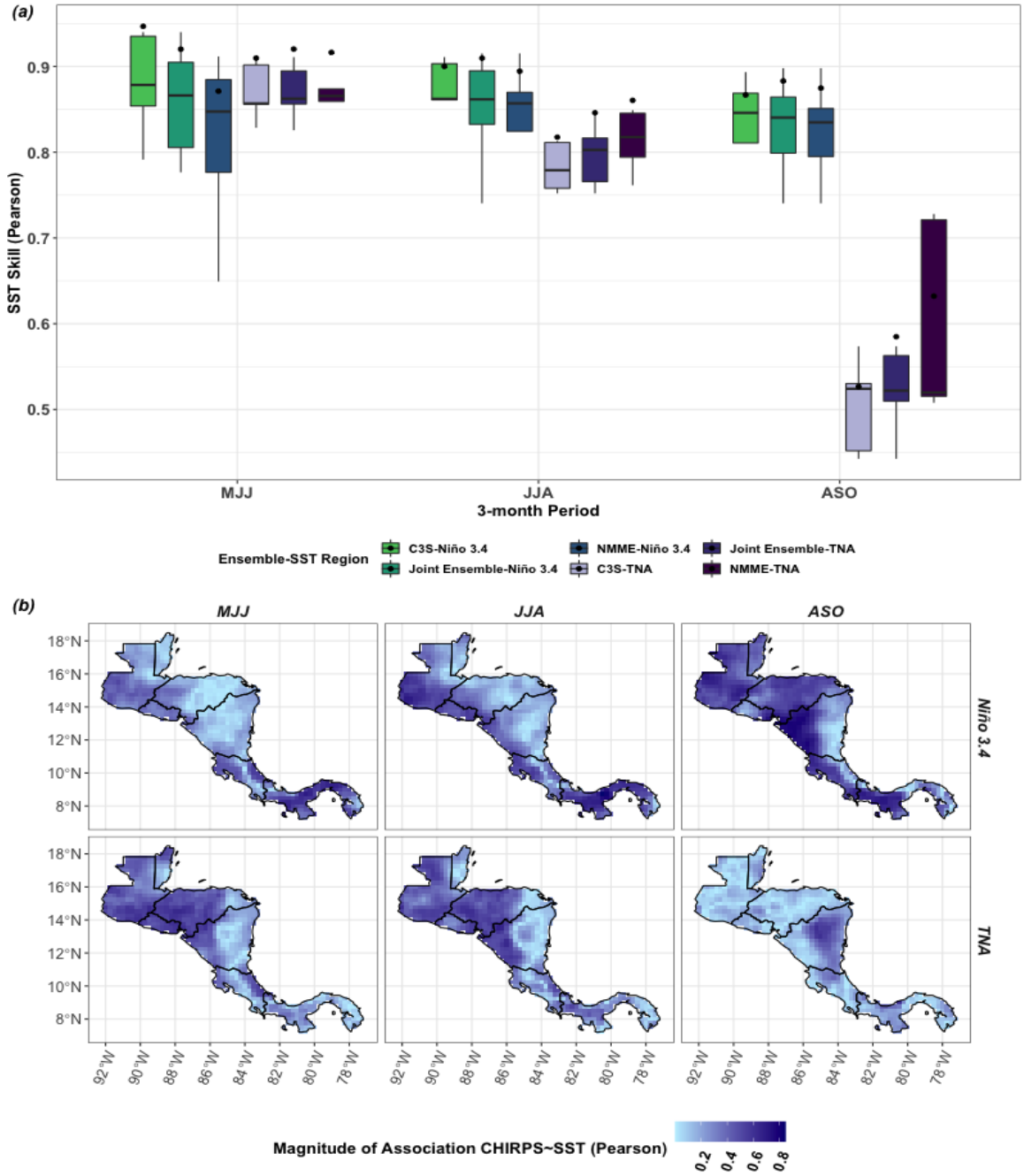


Figure 5.8 Relationship between ENSO, TNA, and Skill. (a) Ensemble SST skill measured using Pearson's R, plotted by season for the Niño3.4 and TNA SST regions. Boxplots represent the range of contributing models; points refer to the skill of each ensemble's mean. (b) Pearson's R between observed rainfall and SST in the Niño3.4 and TNA regions, plotted by

and Central American rainfall, as the TNA indirect forecasts tend to perform better earlier in the wet season (Figures 5.4-5.7). The relative strength of the associations between each SST zone and regional rainfall also somewhat aligns with the differences in rainfall skill. In the North Caribbean, for instance, the association between Niño 3.4 (TNA) SST and rainfall is more than 15% (20%) lower than in the North Pacific in the early to middle wet season (Figure 5.8b).

5.4 Discussion

No single model or method outperformed the others in all cases. Choosing between forecasting approaches is not quite as simple as picking a 'winner,' but our evaluation highlighted a few comparisons that could be useful for decision-making:

(1) ***The North Pacific is a relatively high skill zone*** for seasonal rainfall forecasts. This is true not only for tercile discrimination across the wet season but for low and high rainfall extremes detection in the early and middle wet season (MJJ/JJA). Highest skill occurs in JJA in the North Pacific for both tercile discrimination and rainfall extremes detection, which is when the most distinct and consistent mid-summer dry period occurs of all the subregions (see Figure 5.1a). These findings have useful implications for drought preparedness. The CADC (a drought hotspot – Gotlieb et al., 2019) is in the North Pacific, and high skill for low rainfall extremes detection could potentially help mitigate drought impacts in this region. High skill in the North Pacific in the early wet season is also important for the first planting period of the wet season. For corn and bean crops in Guatemala, for instance, planting occurs primarily in the first ten days of May and June (INSIVUMEH, 2022b), underscoring the importance of the MJJ forecasts for communicating rainfall information to farmers at the MTAs (INSIVUMEH, 2022c).

(2) ***Trouble spots for seasonal rainfall forecasts include eastern Nicaragua for most applications, and the late wet season for low and high rainfall extremes detection***. TNA indirect forecasts are relatively better than other forecasting methods over eastern Nicaragua during the early wet season (MJJ) for tercile rainfall detection, but no forecasting method has high skill when detecting high or low rainfall extremes over eastern Nicaragua. Limited skill across Central America in the late wet season for detection of high rainfall extremes is unfortunate but unsurprising, as this period is when high rainfall extremes most often occur over Central America, primarily on the Pacific slope (Maldonado et al., 2013; Taylor & Alfaro, 2005). High variability of rainfall during the late wet season is compounded by the hurricane season, which occurs in October, making rainfall more difficult to predict. This kind of variability is seen in the TNA SST skill (Figure 8a), which drops significantly in ASO, when storms are forming over the Atlantic.

(3) ***We saw larger differences between forecasting methods (e.g. direct vs. indirect) than between ensemble types (C3S vs. NMME).*** Similarities in skill across ensembles simplify decision-making, as the ensemble choice matters less than picking a forecasting method for a given location/season. Why the dynamically driven forecasts (direct and indirect) are so similar within a forecasting method could arise from a variety of reasons, including limited observational data that constrains how scientists understand the atmospheric

processes over a given location, regional predictability characteristics (e.g. eastern Nicaragua may be harder to predict because the rainfall could be more determined by atmospheric variability than oceanic influences, which are easier to simulate), or due to the nature of the models themselves including structural and initialization choices, but the exact reasoning behind the similarities lies outside the scope of this study. The similarities are unlikely due to observational reference choice, however, as the similarities persist across multiple observational datasets (see Supplementary Information).

(4) ***The comparative advantage of using direct, indirect, and statistical methods varies by application.***

- ***Tercile rainfall discrimination:*** direct and indirect forecasts are relatively better than statistical forecasts in the early to middle wet season (MJJ/JJA) as compared to the late wet season (ASO) when statistical skill is often high. Indirect forecasts have comparative advantages in three primary locations: a Niño 3.4 indirect forecast in Guatemala in the early wet season (MJJ), a TNA indirect forecast over eastern Nicaragua in the early wet season (MJJ), and a TNA indirect forecast centered over the 86ºW longitude line in the early to middle wet season (MJJ/JJA).
- ***Low rainfall extremes detection:*** direct and Niño 3.4 indirect forecasts significantly outperform statistical models in most cases, but in northern Central America in the early and middle wet season (MJJ/JJA), a TNA indirect forecast is the top performer.
- ***High rainfall extremes detection***: high rainfall extremes are more difficult for the models to detect than low rainfall extremes across the wet season. Top performers are less spatially and seasonally consistent, and the late wet season (ASO) is consistently a low skill period.

The exact nature of the relationship between forecast skill and climatic drivers is still unclear. Seasonal differences in SST skill are unlikely to drive any differences in rainfall skill, as the SST skill is often high across the models, especially when SST for a given zone (e.g. Niño 3.4) is highly correlated with rainfall in Central America. The GCMs are relatively good at predicting Niño 3.4 SST (Figure 8a), which aligns with findings from other studies (e.g. Zhang et al., 2017). SST skill does drop over TNA in the late wet season, but rainfall also has a weaker relationship with SST in the TNA region during ASO as compared to earlier in the wet season (Figure 8b), so lower TNA SST skill may have a negligible effect on the skill of the rainfall predictions. The seasonal patterns of Niño 3.4 and TNA SST associations with rainfall also align with other studies, which find that ENSO has a stronger relationship with rainfall in the late wet season (e.g. Waylen and Quesada, 2001; Spence, Taylor and Chen, 2004), and TNA has a strong association with rainfall primarily in the early wet season (Alfaro, 2007; Maldonado et al., 2017; Spence et al., 2004). The comparative value of using dynamically driven forecasts primarily in the early to middle wet season (MJJ/JJA) in the North Pacific, is possibly due to the strong relationship between ENSO and regional rainfall during the late wet season, which increases the skill of the statistical forecasts (Figure 5.8b).

The regional variations in the strength of SST association with rainfall may also be related to rainfall forecast skill. In the Caribbean and the southern part of the isthmus, all forecast skill tends to be lower, with an over 10% drop in 2AFC and HSS scores for normal and extreme

rainfall as compared to their skill in the North Pacific (see Figures 5.5 and 5.7). SST associations with rainfall are also strongest in the North Pacific across much of the wet season (Figure 8b). As ENSO and TNA are not the only factors affecting rainfall over Central America, other influences are important to consider for improving rainfall predictions, especially in low skill zones (e.g. ITCZ migration, tropical cyclone formation, (Durán-Quesada et al., 2020; Hidalgo et al., 2015; Maldonado, Alfaro, & Hidalgo, 2018). The majority of eastern Nicaragua, for instance, has relatively low influence from the Pacific Ocean as compared to the rainfall in the North Pacific (Figure 8b) and may have stronger relationships with other drivers. For instance, the Caribbean Low Level Jet (CLLJ) branches at the 10ºN latitude line and travels north along Nicaragua and south across the southern isthmus (Hidalgo et al., 2015), and the models' abilities to capture this process and how it interplays with the Cordillera (since mountainous terrain is often a challenge for numerical weather prediction – e.g. Serafin et al., 2018) could play a significant role in rainfall forecast skill, which would be worth exploring in future work.

This study aimed to enhance the comparisons of C3S and NMME by including some basic statistical modelling for reference. The statistical models generated here show relatively high forecast skill and would be worth exploring further to see how their skill could be improved. For instance, two main drivers of rainfall were selected based on the literature, but predicting other drivers of rainfall (e.g. windspeed) or conducting an empirical analysis to identify potential hidden drivers of rainfall beyond traditional indices could further improve on the indirect and statistical modelling approaches (e.g. Kretschmer et al., 2016; Kretschmer, Runge and Coumou, 2017; Alfaro et al., 2018; di Capua et al., 2019; Renard and Thyer, 2019). Other study limitations include the limited time period of the hindcasts (1993-2016), which increases the uncertainty of a skill assessment (Tippett, Goddard and Barnston, 2005; Shi et al., 2015). As more models become publicly available for longer periods, it would be worth reconducting the analysis to see how top performers might change over the region. Furthermore, this work demonstrated spatial and seasonal variations in skill that are worth exploring further to understand how different processes may affect model skill. SST skill is relatively consistent across the GCMs (Figure 5.8a), but there is a need for further study of how they capture atmospheric processes, to make sure the models are predicting rainfall for the right reasons (e.g. modulating moisture transport mechanisms like the CLLJ correctly). Next steps would include process-based evaluations that examine other important variables like sea level pressure and wind speed that can affect moisture transport over Central America and test how the ability of the models to represent those process relates to the skill of their final outputs.

5.5 Conclusions

This study systematically compares seasonal rainfall forecasts from two multi-model ensembles relative to a statistical forecasting approach and examines different ways they could be deployed using direct and indirect forecasting techniques. Results demonstrate that both C3S and NMME have high skill when detecting below normal, normal, and above normal rainfall in the wet season, especially over the North Pacific region of Central America. Their comparative value is often highest in the early to middle wet season (MJJ/JJA) when the statistical forecast skill is relatively lower. Indirect forecasts can

improve upon direct rainfall forecasts when skill is low, primarily using a TNA indirect forecast during the early wet season period in the North Caribbean. The multi-model ensembles also show promise for detecting low rainfall extremes in drought hot spots like the CADC but are less reliable for detecting high rainfall extremes in the North Pacific where most high rainfall extremes occur. The variability of top performing forecasting combinations demonstrates how multiple tools can help improve forecast skill rather than selecting one forecasting method or model as the 'winner' for all cases. Continuing to explore drivers of skill over the region and identifying when and where to trust these forecasts is important to ensure they are as reliable as possible for seasonal rainfall planning.

5.A Appendix

5.A1 Observational Data Analysis

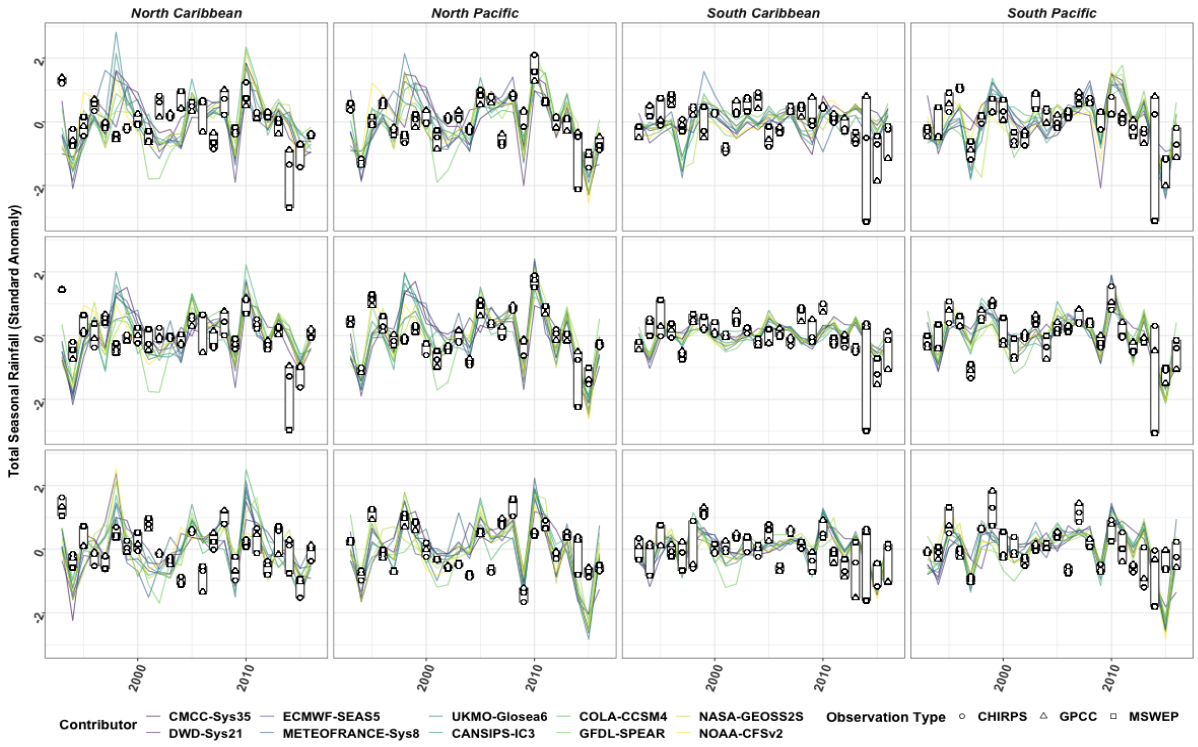


Figure 5.A1 Comparison of observed datasets over 1993-2016 plotted as box plots and points against direct rainfall forecasts from the compared models for each spatially-averaged compared zone and season.

We selected CHIRPS as the rainfall verification dataset for this study, but multiple observational datasets were considered for use in this evaluation. We compared rainfall at 0.25' monthly resolution using CHIRPS (Funk et al., 2014), the Global Precipitation Climatology Centre (GPCC - Schneider et al., 2018), and version 2 of Multi-Source Weighted-Ensemble Precipitation (MSWEP - Beck et al., 2017), which we selected because they include a variety of data inputs and their temporal coverage all span 1993-2016. There is a range in their anomalies of total seasonal rainfall over the time series, and MSWEP tends to lie on the extreme ends of the rainfall anomalies (Figure 5.A1). The range across the compared datasets, however, is often less than the spread of the models (on average less than one standard anomaly for the compared regions and seasons, Figure 5.A1).

In terms of implications for model skill, the correlation between the observed data and model skill is similar across GPCC, MSWEP, and CHIRPS (Figure 5.A2). The models perform similarly for each of the observed datasets, often demonstrating lower association with the observed data over the North Caribbean in Nicaragua, and higher association over the North Pacific over each of the seasons compared.

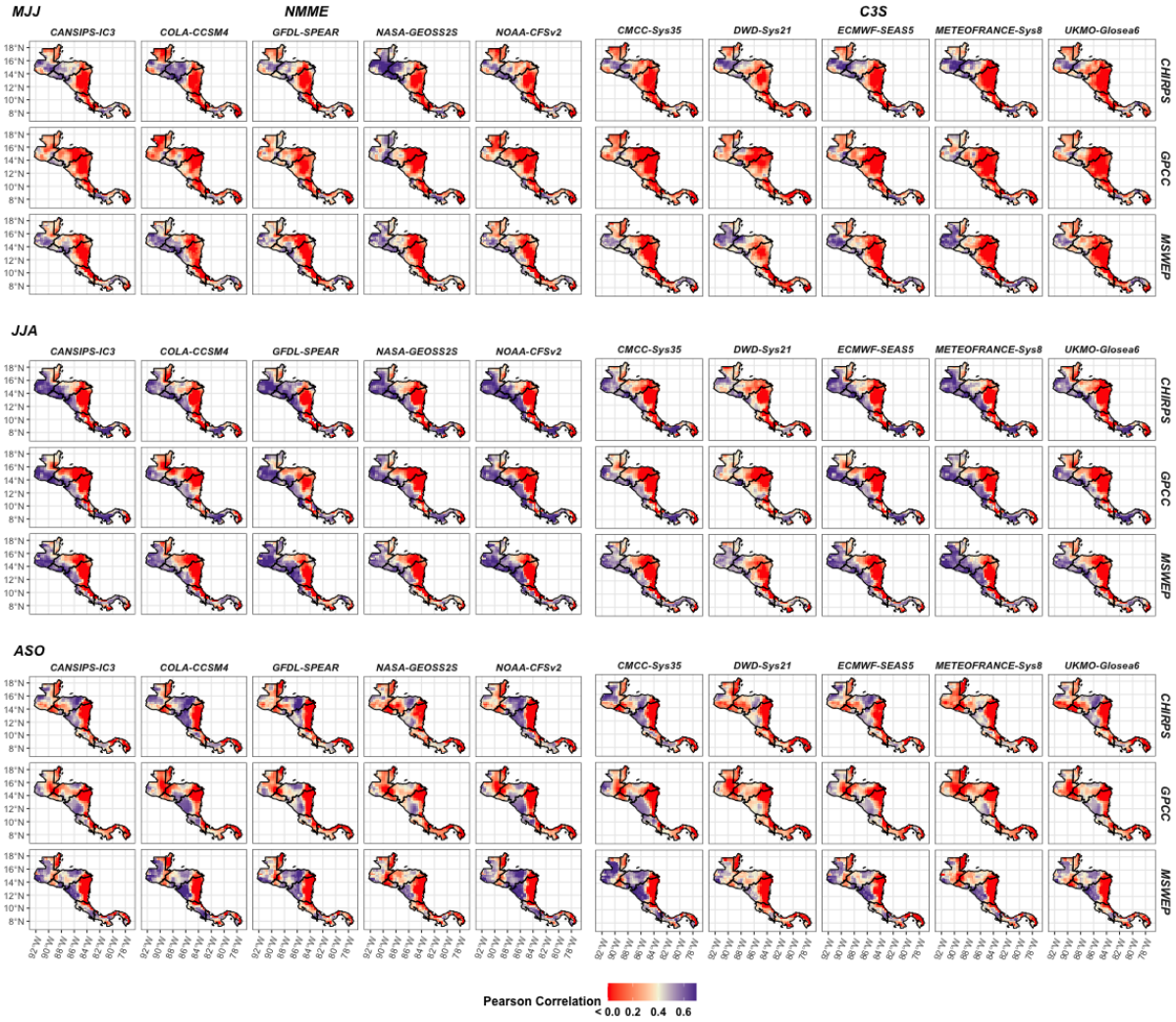


Figure 5.A2 Association (Pearson's R) is plotted spatially across observational datasets using direct model mean rainfall forecasts over May-July, June-August, and August-September.

5.A2 Individual Model Skill for Detecting Low and High Rainfall Extremes

Here we illustrate how individual model skill varies for detections of low and high rainfall extremes. There is more variability across models (Figure 5.A3) than between ensembles (Figure 5.6) – as might be expected, but similar geographic and seasonal variations occur. For instance, the North Pacific is still a relatively high skill zone and eastern Nicaragua is a relatively low skill zone. The variability of the individual model plots illustrates why forecast skill within ensembles has a larger range for detections of extremes. For example, DWD (System 21) has almost very limited skill for detecting high rainfall extremes across the wet season, as compared to Meteo France (System 8), which has high skill across most of the isthmus in the early to middle wet season (MJJ/JJA) (Figure 5.A3 bottom right group).

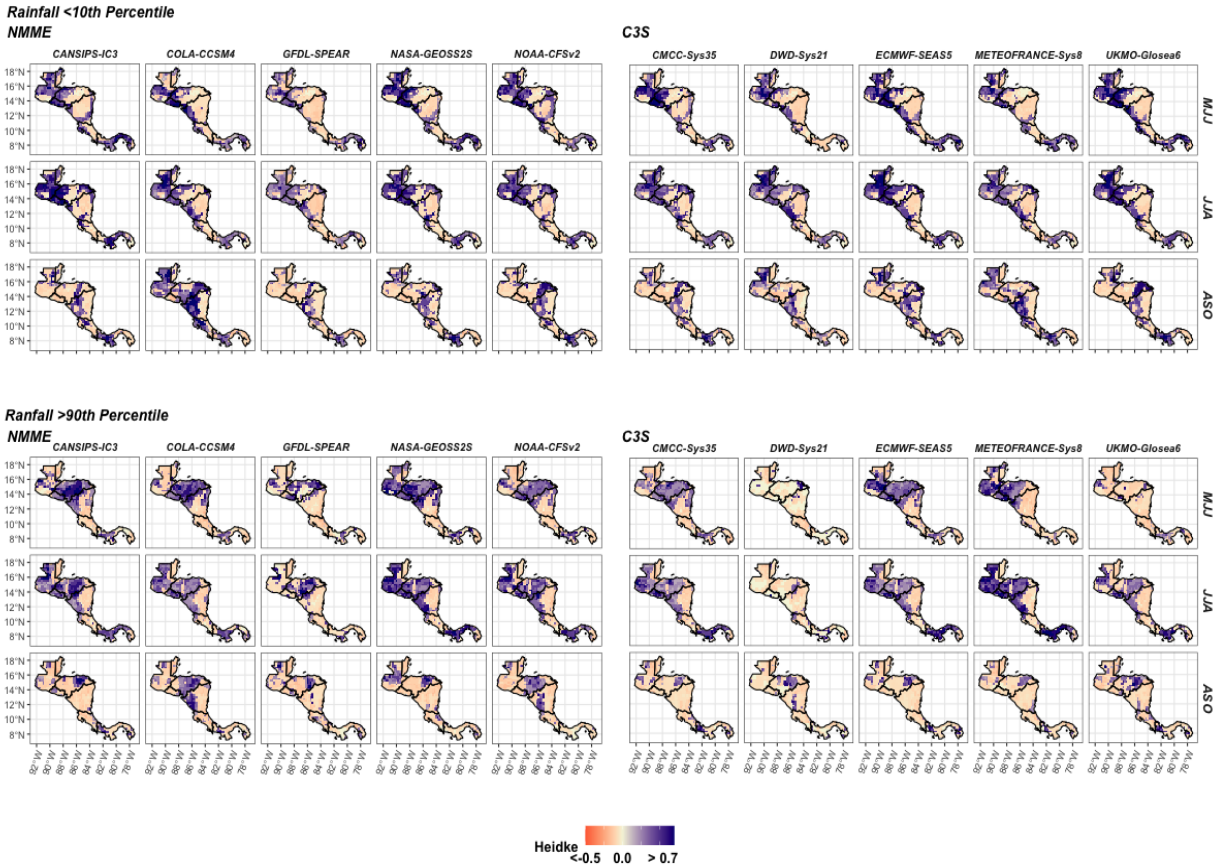


Figure 5.A3 Individual model skill (HSS) for direct forecasts of low and high rainfall extremes, plotted spatially by wet season period.

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

CHAPTER 6

Co-authors: Louise J. Slater, Sihan Li, Timo Kelder, Kyle J. C. Hall, Simon Moulds, Alan García-López, and Christian Birkel
Submitted 11th August 2023 to Geophysical Research Letters

Motivations

One of the challenges that emerged in Chapter 5 was identifying models that perform best for a given location. While Chapter 5 highlights the top forecasting method (Figures 5.5 and 5.7), choosing between models to use as part of an MME is unclear: there is no spatial or temporal consistency in the best performing model (Figure 6.0). Some models have the lowest skill across large areas in some seasons, but the lowest performing model often varies across seasons. Due to the seasonal and spatial inconsistencies in top and low scorers it becomes difficult to choose one or a subset of 'winning' models to use as part of an MME.

A key goal of Chapter 6 was finding opportunities to improve forecast skill. Chapter 4, for instance, demonstrates relatively low skill from SEAS5 in May and June. Chapter 5 reveals that the comparative advantage of using the dynamic ensemble forecasts is often in the early wet season (MJJ) because of the worse performance of the statistical models. If the dynamic ensembles could be improved during May and June, the comparative advantage of using

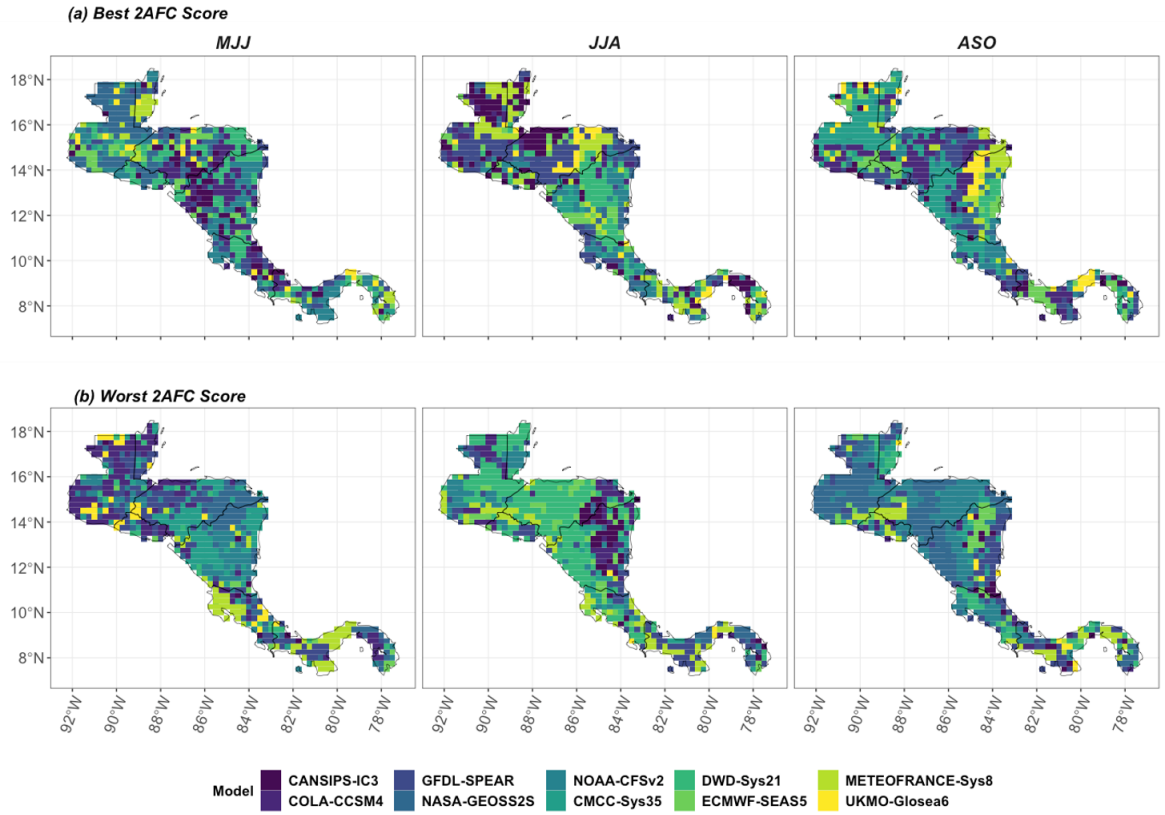


Figure 6.0 (a) best and (b) worst scoring models using 2AFC are plotted for three-month-mean seasonal forecasts initialized one month prior across seasons (panels).

these ensembles could possibly widen. Furthermore, while the ensembles often perform well in the late wet season (Figures 4.3, 5.5), they often fail to detect low and high rainfall extremes well, facing a steep drop in skill for the predictions that include September (Figures 4.4, 5.7).

Chapter 5 also aligns with previous studies that find MMEs have higher skill than individual models (e.g. Elvidge et al., 2023). The forecast skill of the MME mean is often higher than the entire interquartile range of individual model skill for tercile rainfall discrimination (Figure 5.5), demonstrating the added value of combining AOGCMs to create a forecast. However, the MME mean outperforms the interquartile range of the individual models less often for low and high rainfall extremes detections (Figure 5.7). This difference in the relative skill of the MME motivated Chapter 6's investigation into alternative ways to generate MMEs to improve detection rates of rainfall extremes.

Conducting a process-based evaluation was of interest partially due to the findings regarding the hybrid models in Chapter 5. The skill of the hybrid forecasts often improves when the teleconnection patterns they use (e.g. ENSO) are most related to rainfall (Figures 5.5, 5.7). There are some exceptions, however, as the Niño 3.4 hybrid forecast performs best in Guatemala at discriminating between terciles of rainfall in the early wet season (Figure 5.5), when Niño 3.4 is relatively less related to regional rainfall as compared to other regions like TNA (Figure 5.8). This exception raised the question about what drives rainfall forecast skill. Does the hybrid model perform better over Guatemala in that instance due to Niño 3.4 SSTs being a relatively easier location to model? If an ensemble could be optimized to better represent TNA SSTs, for instance, would a forecast that uses TNA outperform a forecast based on Niño 3.4 SSTs in the early wet season?

These questions further drove an interest in exploring how process representation in the models relates to their final skill (RQ4, Thesis Outline 1.2), and questions over whether their process representation could be capitalized on to improve MME skill (RQ5, Thesis Outline 1.2). Because the inconsistency in forecast skill across the models (Figure 6.0) does not indicate a clear 'best' or 'worst' performing model, the focus for skill improvement pivoted away from choosing between models towards developing a subsampling approach that selects the best members within models that by chance have the correct process representation at the point of model initialization.

A note on model inclusion in Chapter 6: contributing models to C3S are included partially because of consistent availability of predictor and target variables over the available hindcast period (1993-2016), but also because the C3S ensembles' interquartile range was much larger than that of the NMME (e.g. Figure 5.5), which might indicate a larger spread in predictions and potential errors in C3S.

Key Points

- Subsampling members using sea surface temperatures and zonal wind improves subseasonal ensemble rainfall forecasts in Central America.
- In multiple months and locations mean squared error skill increases by 0.4 and extreme rainfall skill improves by 0.5 (Heidke skill).
- Process-informed subsampling is useful because the models' representation of rainfall degrades as process error increases.

Abstract

Subseasonal rainfall forecast skill is critical to support hydrometeorological extremes preparedness. We assess how a process-informed evaluation, which subsamples forecasting models' members based on their ability to represent potential predictors of rainfall, can improve monthly rainfall forecasts within Central America one month ahead, using Costa Rica and Guatemala as test cases. We generate a constrained ensemble mean by subsampling 130 members from five dynamic forecasting models in the C3S multi-model ensemble based on their representation of both (1) zonal wind direction and (2) Pacific and Atlantic sea surface temperatures (SSTs), at the time of initialization. Our results show in multiple months and locations increased mean squared error skill by 0.4 and improved detection rates of rainfall extremes. This method is transferrable to other regions driven by slowly-changing processes. Process-informed subsampling is successful because it identifies members that fail to represent the entire rainfall distribution when wind/SST error increases.

Plain Language Summary

Subseasonal rainfall forecasts provide alerts multiple weeks ahead. These forecasts present an opportunity to facilitate anticipatory actions yet are often unreliable to use when preparing for extreme weather. We develop a method to optimize rainfall forecasts by selecting individual members from a large ensemble of dynamic forecasting model outputs based on their ability to represent potential predictors of rainfall. We test our method on monthly rainfall forecasts within Central America one month ahead, using Costa Rica and Guatemala as key test cases. We select members from five contributing models of the C3S multi-model ensemble using regional predictors, including wind direction and sea surface temperatures (SSTs). Our results show improvements in the detection of low and high rainfall extremes. This method is transferrable to other regions driven by slowly-changing processes like SSTs and is beneficial for operational forecasters who can leverage regional expertise of relevant rainfall-generating processes to subsample better performing ensemble members for their regions.

6.1 Introduction

Rainfall forecasts can support preparedness for hydrometeorological extremes like droughts and floods (Braman et al., 2013; Merz et al., 2020; Domeisen et al., 2022; White et al., 2022). At the subseasonal scale (14-60 days ahead), early warnings support proactive disaster mitigation activities such as strategies for planting crops (e.g. Flohr et al., 2017, 2018) and transporting resources to higher ground (De Perez et al., 2016). Subseasonal forecasts are challenging because after several weeks atmospheric conditions lose most of their memory and large scale oceanic variability often only provides a limited source of skill (Vitart and Robertson, 2018).

Atmospheric oceanic general circulation models (AOGCMs) provide one way to predict rainfall by generating dynamic predictions of the earth system (Hagedorn, Doblas-Reyes and Palmer, 2005; Stockdale et al., 2010; Bauer, Thorpe and Brunet, 2015). Several techniques can correct raw AOGCM outputs, often categorized into calibration (Manzanas et al., 2019) and combination methods (Hemri et al., 2020). Combining AOGCMs to generate multi model ensembles (MMEs) has been advantageous in several cases (e.g. Elvidge et al., 2023; Palmer et al., 2004; Wang et al., 2009), but does not always significantly improve forecast skill compared to calibrated single models because model errors in MMEs are often correlated (Weigel, Liniger and Appenzeller, 2009). Traditionally model evaluation is conducted based on aggregated mean error and variance. Such evaluation techniques, however, may obscure the possibility that models can be right for the wrong reasons (Eyring, 2016; Eyring et al., 2019; Nowack et al., 2020).

New approaches have been proposed to improve AOGCM temperature forecast skill in Europe at seasonal to decadal scales by capitalizing on process differences *within* AOGCM ensembles (e.g. Dobrynin et al., 2018, 2022; Dusterhus, 2020; Smith et al., 2020). Each model is comprised of multiple members, which when initialized represent a range of guesses of current day states to account for observational uncertainty (Balmaseda et al., 2011). Smith et al. (2020), for instance, improved North Atlantic Oscillation (NAO) representation by subsampling decadal predictions from CMIP5-6 based on each model member's proximity to the multimodel ensemble mean's NAO estimate. Process-informed subsampling approaches like these are useful not only because they can potentially improve forecast skill, but also because they can provide insights into causal mechanisms that affect model skill.

We expand on these ideas and propose a post-processing technique to help diagnose and improve AOGCM subseasonal rainfall forecasts in real-time. As subseasonal forecasts sit between weather and seasonal climate forecasting windows, we investigate process representation of both atmospheric and oceanic drivers of rainfall (sea surface temperature – SST – and zonal wind). We hypothesize that if through chance certain model members predict both SST and zonal wind well one to two months ahead, these same members will also have higher rainfall skill at two-month lead time. To subsamples members in a manner that could be operationally deployed, we identify better performing members at the point of model initialization. We evaluation the predictions of SST and zonal wind direction, two variables that change more slowly over time, enabling us to use lagged observations as a

reference to filter out members whose predictions have become unrealistic. We then subsample the ensemble to include only the top performing members, thereby improving the ensemble mean rainfall forecast.

We test our process-informed subsampling method in Central America, a region in need of further MME optimization where AOGCM skill varies across locations, times of year, and lead times (Carrão et al., 2018; Hidalgo & Alfaro, 2012, 2015; Kowal et al., 2021, 2023; Maldonado, Alfaro, & Hidalgo, 2018; Maldonado, Alfaro, Amador, et al., 2018). Central America presents prediction challenges in part due to the complex interaction of weather patterns originating from both the Pacific and Atlantic oceans and marked topography that moderates moisture transport over the region (Durán-Quesada, Gimeno and Amador, 2017; Durán-Quesada et al., 2020). Forecasts are therefore unlikely to improve by subsampling ensembles using one process alone. The El Niño Southern Oscillation (ENSO – Trenberth, 1997), for instance, is widely recognized as an important driver of ensemble forecast skill (e.g. Scaife et al., 2019) due to its teleconnections with regional rainfall (Durán-Quesada et al., 2020). ENSO alone, however, cannot always explain regional rainfall deficits (Muñoz-Jiménez et al., 2019).

Here, we subsample ensemble members based on their representation of multiple key regional rainfall-driving mechanisms, and examined whether our method improves subseasonal rainfall forecast skill in two distinct subregions: Guatemala and Costa Rica (Figure 6.1a). We then use this approach to assess which large-scale rainfall-generating processes need to be accurately captured by models for subseasonal forecasts to be skilful.

6.2 Data

We generated a MME using members from five AOGCMs (Table 6.A1) that contribute to the leading European seasonal forecasting system (C3S; Marsh & Penebad, 2016). These models' mean monthly estimates are available at 1˚ spatial resolution over a 24-year hindcast period (1993-2016) up to six months ahead with a total of 130 members initialized on the first day of every month. Each model contains SST, wind at 850 hPA, and rainfall; and they are commonly deployed for monthly to seasonal forecasts around the world (e.g. Colman et al., 2020; Mishra et al., 2019; Walker et al., 2019).

We used daily Optimum Interpolated Sea Surface Temperature data (OISSTv2 - Huang et al., 2021) from National Oceanic Atmospheric Administration (NOAA) as our observational reference data for the models' SST predictions and ERA5 reanalysis data for wind direction (Hersbach, 2020). We used Climate Hazards Group InfraRed Precipitation with Station 0.25º rainfall data (CHIRPS - Funk et al., 2014) to test the final predictions given CHIRPS has been found to be a reasonable dataset within the region (Arciniega-Esparza et al., 2022; Stewart et al., 2022). All variables tested (predictors and rainfall predictands) were converted into standardized anomalies prior to testing (Text 6.A1 Equation 1).

6.3 Methodology

To support operational feasibility, we used observations just prior to model initialization as reference data to select optimal members when the forecasts were issued. In this approach, we selected members with the best representation of SST and zonal wind direction at one-

and two- month lead times. SST and zonal wind were chosen because they are key drivers of regional rainfall (see 6.3.1) and typically have greater persistence over other variables. Although the magnitude of windspeed may change instantaneously, the low-level jet wind direction is often seasonally stable (Figure 6.A1).

We assessed the usefulness of this subsampling technique on monthly rainfall forecasts in the month after initialization (labelled two-month lead per C3S naming conventions). For example, to generate a subsample for September rainfall predictions, we extracted all available members from the models initialized on August 1st at two separate lead times: one-month lead (August) and two-month lead (September). We selected members that best represented our chosen predictors at one- and two-month lead times, using observations averaged over the last two weeks of July (reference period) and used only those top performing members in our final two-month lead rainfall prediction. Because this method relies on lagged observations to select members at one- and two-month lead times, the method will work better when SST and zonal wind direction change more slowly over our predictor zones.

We tested the subsampling method in May, June, September, and October to see how its ability to constrain ensemble predictions of rainfall varied temporally. The Central American climate is characterized by monthly rainfall variations, separated into a wet and dry season with peak rainfall often in May/June (early wet season) and September/October (late wet season) (Giannini, Kushnir and Cane, 2000; Taylor & Alfaro, 2005; Magaña et al. 1999). We evaluated the subsampling method over the entirety of Costa Rica and Guatemala respectively as key case studies and considered spatial patterns in subsample skill. Rainfall connections with relevant predictors (6.3.1) for the Caribbean and Pacific sides of the region, for instance, vary in part due to how the Cordillera mountain range (Figure 6.1b) interacts with moisture transport from the Caribbean Sea to the Pacific Ocean (Taylor and Alfaro, 2005; Durán-Quesada, Gimeno and Amador, 2017; Muñoz-Jiménez et al., 2019).

6.3.1 Relevant Predictor Zones for Central America

We evaluated seven predictor types with well-documented connections to regional rainfall (Figure 6.1c), including (1) Niño 3.4 SSTs (Waylen, Quesada and Caviedes, 1994; Giannini, Kushnir and Cane, 2000; Sánchez-Murillo et al., 2017), (2) Tropical North Atlantic (TNA) SSTs (Enfield and Mayer, 1997; Maldonado et al., 2017), and (3) Caribbean SSTs (Durán-Quesada, Gimeno and Amador, 2017; Durán-Quesada et al., 2020). We also evaluated representation of the SST gradient across the isthmus (Alfaro, 2007; Enfield & Alfaro, 1999; Enfield & Mayer, 1997; Hidalgo et al., 2015; Maldonado, Alfaro, & Hidalgo, 2018). We used two different SST gradients as predictors to account for the spatial and temporal variability in the relationship between SST gradients and regional rainfall: (4) the maximum anomaly difference between the Niño 3.4 region and the Caribbean (Wang, 2007), and (5) the maximum anomaly difference between Niño 3.4 and TNA SSTs (Enfield & Alfaro, 1999; Taylor et al. 2002). Because the Caribbean Low Level Jet (CLLJ – Amador, 2008) and Chorro del Occidente Colombiano Jet (CJ - Poveda & Mesa, 1999) are also well-documented moisture transport mechanisms across Central America (Cook & Vizy, 2010; Hidalgo et al., 2015; Muñoz et al., 2008; Poveda & Mesa, 1999; Wang, 2007), we assessed zonal wind direction (6) using CJ alone, and (7) within a region that circumscribes the core

CLLJ, Central America, and the eastern Pacific where the CJ crosses the isthmus (Figure 1c). CLLJ was not used alone as a predictor because there was not a large difference between member representation of the core CLLJ (Text 6.A2).

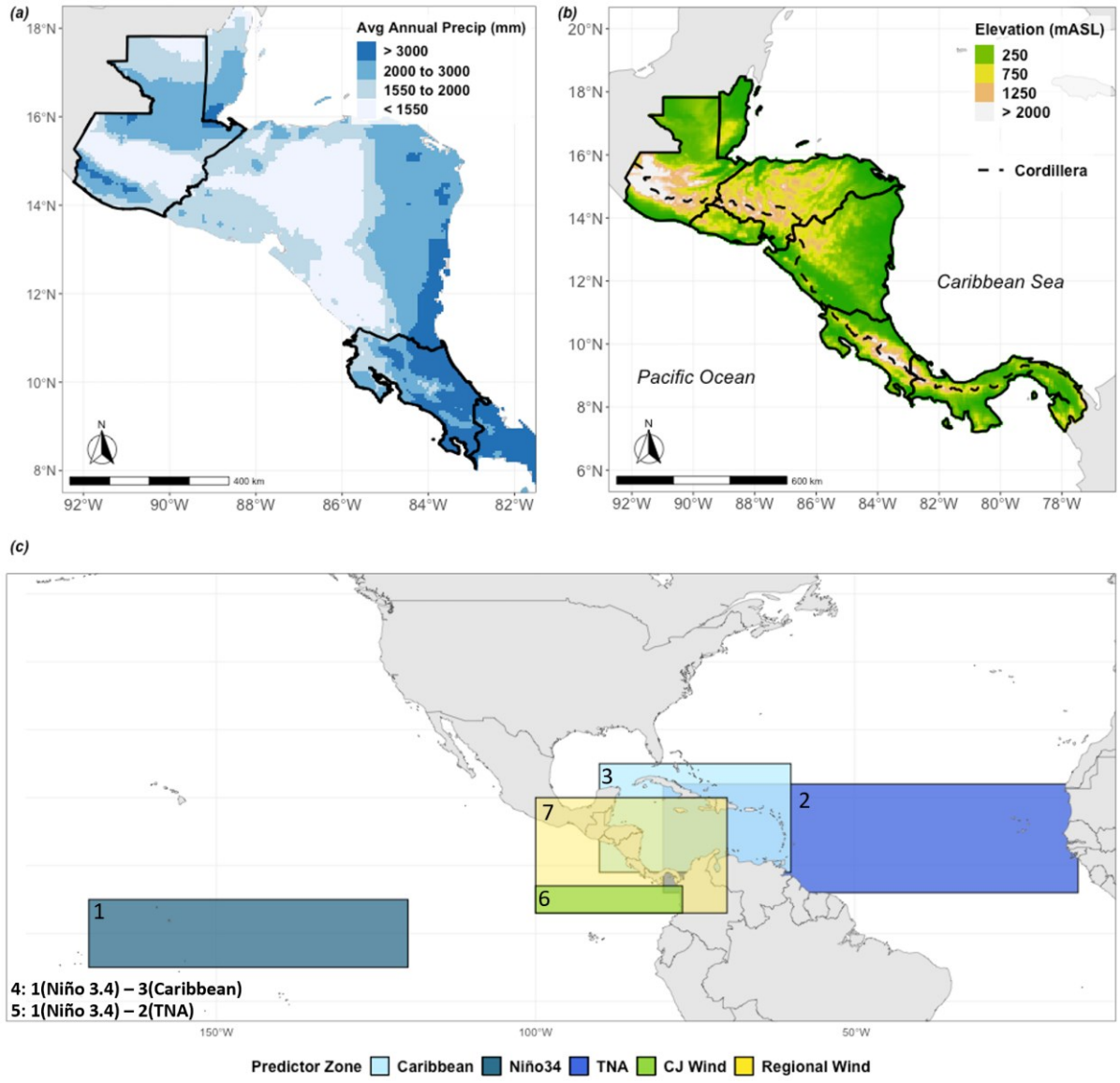


Figure 6.1. Illustration of Central American climate and tested predictor zones. (a) Total annual rainfall averaged over 1993-2016 using CHIRPS 0.25' (Funk et al. 2014) monthly rainfall data with Guatemala and Costa Rica outlined in black. (b) Elevation in Central America including the Cordillera. (c) Predictor locations used to test member performance plotted spatially (numbered as labelled in text with SST predictor zones plotted behind country outlines and wind predictor zones plotted in front). Zones include the Caribbean Sea (90ºW, 60ºW, 9ºN, 25ºN which includes the core zone for moisture transport illustrated in Durán-Quesada et al. 2017, Niño 3.4 (170ºW, 120ºW, 5ºN, 5ºS Trenberth 1997), and TNA (80ºW, 15ºW, 6ºN, 22ºN Enfield & Alfaro 1999); and wind direction over the CJ (100ºW, 77ºW, 3ºN, 7ºN Poveda & Mesa 2000) and a broader 'Regional Wind' zone that circumscribes the eastern Pacific, core CJ and CLLJ locations (100ºW, 70ºW, 3ºN, 20ºN).

6.3.2 Member Performance and Top Performing Subsamples Selection

We processed all model data with the XCast Python package (Hall & Acharya, 2022; Text 6.A3) and predictor estimates from the raw 130-member ensemble at 1º spatial resolution. We regridded the 130-member ensemble rainfall predictions to 0.25º resolution using bilinear interpolation prior to subsampling. The 130 estimates of each predictor were ranked (Figure 6.2, step 1) using the average mean squared error (MSE; Text 6.A1 Equation 2) across the predictor zone for SST (or calculating one MSE estimate for the gradients with no spatial field). The average Percent Correct Score (Text 6.A1 Equation 3) was used to rank the members' predictions of zonal wind direction within a predictor zone (e.g. CJ). We filtered members based on their relative performance to one another (e.g. top 10) to keep ensemble size consistent.

The 'best' subsample is application dependent. Here we showcase a few subsamples that performed well on average over both countries using deterministic criteria over 1993-2016. We used multiple metrics to evaluate rainfall skill of the subsampled means against the all-member mean, including the Mean Squared Error Skill Score (MSE-SS - Deque, 2012,Text 6.A1 Equation 4) to assess performance across the entire rainfall distribution, and the Heidke Skill Score (HSS - Heidke, 1926, Text 6.A1 Equation 5) to measure ensemble mean detection rate of low (high) rainfall extremes below the 10th (above the 90th) percentile of monthly anomalies.

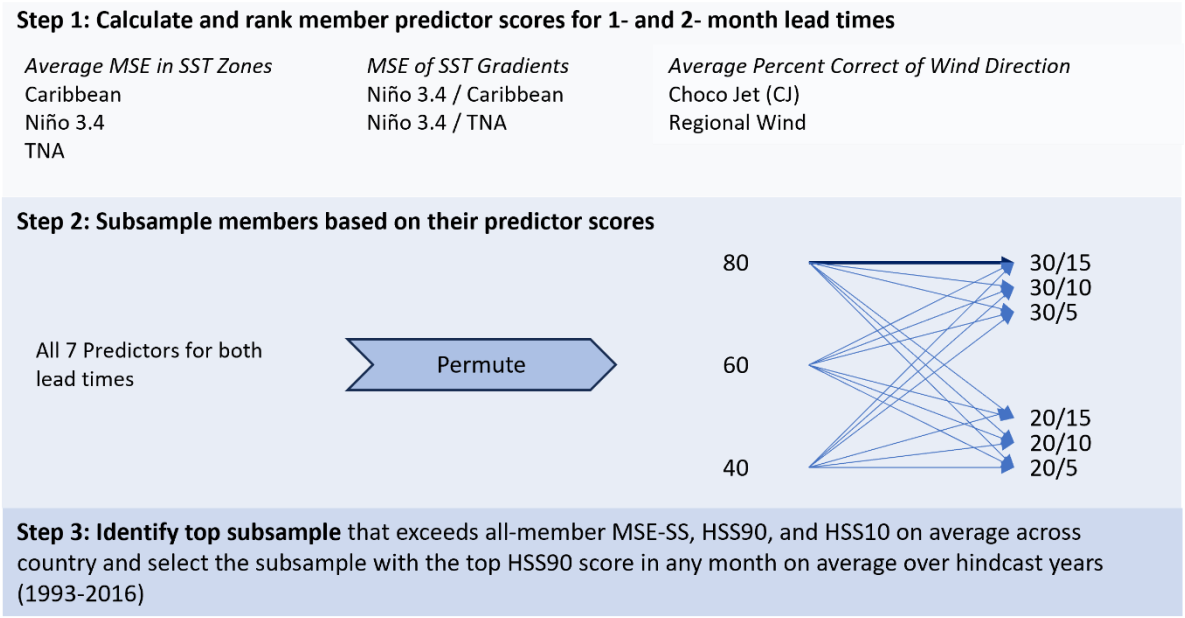


Figure 6.2 Schematic of method used to estimate how members represent physical processes. Step 1: rank members based on their predictor scores for all seven predictors. Step 2: Subsample members based on their predictor scores using permutations (order does not matter) of predictors. In this subsampling step, filter size options include 80/60/40 (filter 1); 30/20 (filter 2); or 15/10/5 (filter 3) members. For example (bolded arrow), select top 80 members that represent Niño 3.4 SSTs best at two-month lead then of those 80 members, select top 30 members that represent TNA SSTs best at one-month lead, then of those 30 members select top 15 members that represent CJ wind direction best at one-month lead. Do this for every possible order of predictors across the filter size options. Step 3: evaluate subsample skill scores using MSE-SS, HSS90, and HSS10 and top subsamples that scored as well as the all-member ensemble mean across all metrics on average over a country. Although many subsamples meet these criteria, for illustration purposes in the results section, top subsample is selected using HSS90.

6.3.2 | Member Performance and Top Performing Subsamples Selection

We generated subsamples using our ranked predictor scores to select members for permutations of one, two, and three predictors with several filter size options (Figure 6.2 step 2). Generating thousands of subsamples across a range of predictor permutations enabled us to compare the subsamples that prioritized more well-rounded skill (using multiple predictors to select members) against subsamples that prioritized representation of a particular predictor. We identified a subset of better performing subsamples in which rainfall predictions outperformed the all-member mean using all deterministic scores (MSESS, HSS90, and HSS10) in any month (May, June, September, or October) on average over a country for the entire assessed period (Figure 6.2 step 3). Many filter sizes resulted in skilful subsamples (Text 6.A4), and when filter size did matter, it often indicated the relative importance of different predictors (Figure 6.A4).

This selection procedure identifies subsamples that on average maintain better skill across the analysed years. Because of the interannual variability in ensemble forecast skill, we performed a sensitivity test to evaluate the robustness of the approach (Text 6.A5). Although the skill of the all-member and subsampled ensembles changes interannually (Figure 6.A3), randomly dropping five years from the analysis does not significantly change the difference in skill between the all-member mean and subsampled mean in almost any case tested (Figure 6.A4). To better understand what happens to the models' rainfall predictions when predictor error increases, we also examined the correlation between the model rainfall predictions and the estimates of the seven predictors for each month analysed.

6.4 Results and Discussion

6.4.1 Top Subsamples Align with Key Processes that Drive Regional Rainfall

The 'best' subsample depends on the application (e.g. many subsamples detect high rainfall extremes well but fail to consistently detect low rainfall extremes). Our results below summarize subsamples that outperform the all-member mean across all three metrics (MSESS, HSS90, and HSS10) on average within a country. Several subsamples outperform the all-member mean across all metrics for a given month, but the ones highlighted here have the top HSS90 skill for detections of high rainfall extremes (Table 6.1).

Subsamples that use three predictors to filter members tend to outperform subsamples with fewer predictors, except when using TNA alone in May over Guatemala (Table 6.1 row 1 right column). We found no subsamples in June that outperformed the all-member mean in Costa Rica across all metrics, while the ten member TNA subsample that performed best in Guatemala in May still performed well in June. More limited success in June may be related to the stronger influence of atmospheric drivers that affect moisture transport (Durán-Quesada et al., 2017), or higher predictor variability, as the lagged observations were poorly correlated with predicted months in that period (Figure 6.A1). While our method may constrain error that occurs in a large ensemble, any subsampling procedure is also subject to uncertainties in observational datasets. The usefulness of the method depends on these datasets being close enough approximations of 'real-world' conditions to help filter out unrealistic members.

Using process-informed criteria to select members from a large ensemble can help diagnose why model skill may vary (Eyring et al., 2019; Nowack et al., 2020). The top subsamples

Table 6.2. **Summary of top subsamples identified by country and months** where both countries have subsamples that outperformed the all-member mean on average across all metrics. The top subsamples for each month are listed next to the order of operations used to filter each predictor with number of members selected in each step in parentheses and the average [worst,best] change in predictor scores over 1993-2016 between the final subsample and the all-member mean. Improvements in predictor score are reported in terms of MSE for SST zones and SST gradients and Percent Correct for wind direction, as calculated for ranking members. Decreases in MSE (more negative) and increases in Percent Correct scores show improvements in predictor score. Final ensemble sizes and average changes in predictor errors are bolded.

		Costa Rica		Guatemala	
Predictand Month	**Filter Step**	**Predictors (Filter Size)**	**Average [worst,best] Change in Predictor Score**	**Predictors (Filter Size)**	**Average [worst,best] Change in Predictor Score**
May	**1**	SSTs in Caribbean 1-month lead (60)	**MSE: -0.24** [-0.11, -0.59]	SSTs in TNA 1-month lead **(10)**	**MSE: -0.24** [-0.12, -0.77]
	2	CJ Wind 2-month lead (20)	**%Correct: 21%** [10%, 33%]		
	3	SST Gradient Niño3.4/TNA 2-month lead **(5)**	**MSE: -0.63** [-0.17, -2.48]		
September	**1**	SSTs in Niño 3.4 1-month lead (80)	**MSE: -0.02** [0.01, -0.06]	SSTs in Niño 3.4 1-month lead (80)	**MSE: -0.02** [0.01, -0.06]
	2	SSTs in Niño 3.4 2-month lead (30)	**MSE: 0.06** [0.16, 0.01]	SSTs in Niño 3.4 2-month lead (30)	**MSE: 0.05** [0.16,-0.03]
	3	Regional Wind 1-month lead **(15)**	**%Correct: 3%** [-1%, 6%]	SST Gradient Niño3.4/Caribbean 2-month lead **(10)**	**MSE: -0.54** [-0.20, -1.12]
October	**1**	SSTs in Niño 3.4 2-month lead (40)	**MSE: -0.09** [-0.04, -0.19]	CJ Wind 1-month lead (60)	**%Correct: 12%** [3%, 30%]
	2	SST Gradient Niño3.4/Caribbean 2-month lead (20)	**MSE: 0.28** [0.70, -0.12]	SST Gradient Niño3.4/TNA 2-month lead (20)	**MSE: 0.70** [1.63, 0.19]
	3	SST Gradient Niño3.4/Caribbean 1-month lead **(15)**	**MSE: -0.10** [0.41, -0.49]	CJ Wind 2-month lead **(15)**	**%Correct: 5%** [-12%, 10%]

often relate to seasonal drivers of rainfall. In September, for instance, top subsamples primarily rely on Niño 3.4 SSTs, when ENSO drives rainfall more strongly, while in the early wet season, Caribbean and TNA SSTs are more important (Table 6.1 row 1) when the Caribbean and Atlantic have stronger correlations with regional rainfall (Taylor et al., 2002; Spence, Taylor and Chen, 2004; Durán-Quesada, Gimeno and Amador, 2017; Maldonado et al., 2017). SST Gradients improve subsamples across the wet season, as they are important throughout (Enfield & Alfaro, 1999; Giannini, Kushnir and Cane, 2000; Hidalgo et al., 2015; Maldonado, Alfaro, & Hidalgo, 2018).

The top subsample over Guatemala in October is less straightforward, as it is filtered primarily using CJ representation. CJ nears peak strength in October as the intertropical convergence zone is displaced northwards (Amador, Durán-Quesada, et al., 2016), but this jet primarily affects rainfall south of Guatemala (Durán-Quesada et al., 2010; Durán-Quesada, Gimeno and Amador, 2017). This predictor may indicate related model errors (e.g. if members poorly represent CJ, they may also poorly represent other drivers of Guatemalan rainfall in October).

6.4.1 | Top Subsamples Align with Key Processes that Drive Regional Rainfall

Average predictor representation is not always substantially different between the all-member mean and the top subsamples. The all-member mean typically includes members with mixed levels of skill – some with high skill, some with low skill, and some with splintered skill (high in some predictor metrics and low in others). The top subsample sometimes improves representation across all predictors in the analysis (e.g. subsamples in May for both countries), meaning the subsample filters out low-skill members. In other cases, average process representation of the top subsample is similar to (e.g. in September) or even worse than the all-member mean (e.g. average MSE of some SST gradient representations in October). In these cases, the top performing subsamples filter out splintered skill members from the all-member ensemble. In October, for instance, the subsamples select more well-rounded members that can better represent multiple predictors (e.g. CJ wind and SST gradients) compared to the filtered out splintered skill members that poorly represent certain predictors. Further, no single AOGCM consistently outperforms the others, and model representation within subsamples varies interannually (Text 6.A6), not necessarily linked to the total number of members per AOGCM (Table 6.A2).

6.4.2 Subsamples Improve Skill for Entire Rainfall Distribution and Extreme Tails

In all three months, the subsamples improve MSE-SS across most of Costa Rica and Guatemala (Figure 6.3a/c), with the largest change in September over Costa Rica (70% locations increase ≥0.2 MSE-SS, 55% locations increase ≥0.3 HSS90, 33% locations increase ≥0.3 HSS10) and in May over Guatemala (56% locations increase ≥0.2 MSESS, 62% locations increase ≥0.3 HSS90, 5% locations increase ≥0.3 HSS10). The skill of predicting rainfall extremes improves more in some locations (Figure 6.3b/d) with >0.5 changes in HSS for P10 and P90 detections. Filtering out members with low skill is clearly beneficial in Guatemala in May, when the top subsample filters members using TNA representation alone (Figure 6.3c/d top row). Filtering out splintered members is also useful in Costa Rica in September, when the top subsample significantly improves skill (Figure 6.3a/b middle row) despite similar average predictor scores as the all-member mean (Table 6.1). These results are similar at the raw 1˚ spatial resolution (Figure 6.A5).

Improving forecast skill across an entire country using one subsampling configuration is challenging due to nuances in how processes drive rainfall. Some improvements are spatially limited, highlighting locations where predictor relationships with rainfall may diverge. For instance, in May, the top Costa Rican subsample mostly improves skill on the Pacific side of the country (Figure 6.3a/b). Pacific rainfall patterns are distinct from Caribbean rainfall often due to interactions between regional rainfall drivers and topography (Durán-Quesada, Gimeno and Amador, 2017; Muñoz-Jiménez et al., 2019). The top subsamples also show limited skill above 16ºN in Guatemala in September and October (Figure 6.3c/d). This spatial pattern in skill somewhat corresponds to differences between rainier locations in the country that are closer to the Caribbean Sea, and drier locations in Guatemala with rainfall more dominated by the Pacific (Figure 6.1a), including the Guatemalan Dry Corridor (Gotlieb et al., 2019). This method could therefore benefit from using more locally optimized subsamples (e.g. over Caribbean vs. Pacific rainfall regions in Costa Rica, Dry Corridor in Guatemala).

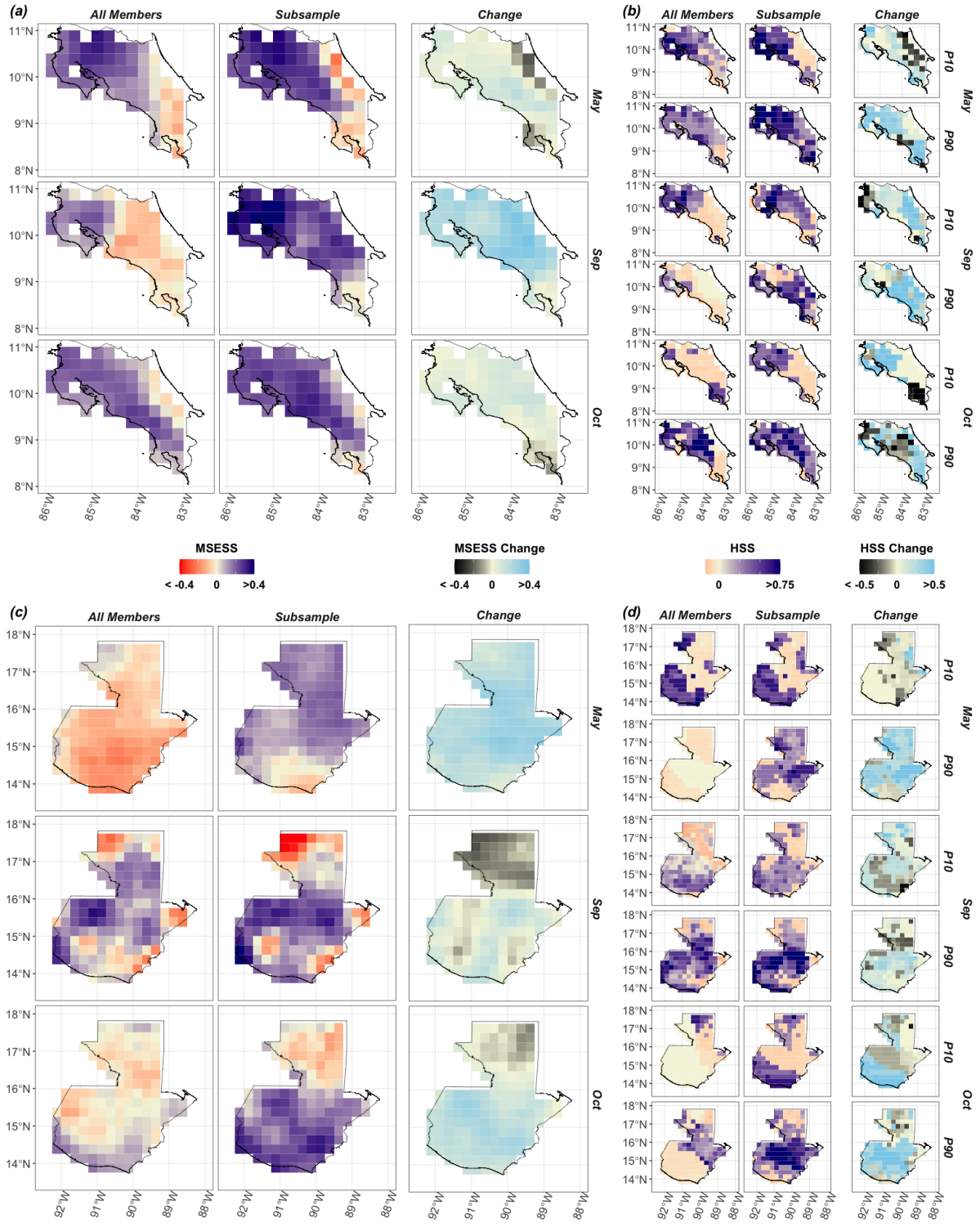


Figure 6.3 Spatial performance of the method in Costa Rica and Guatemala. (a) Spatial skill of all-member ensemble (left), top performing subsampled ensemble (center) and difference between the two (right) in Costa Rica. Skill is based on the entire rainfall distribution using MSE-SS. (b) Spatial skill for detection of low rainfall extremes (P10) and high rainfall extremes (P90) using HSS in Costa Rica, again for all-member (left), subsample (center), and their difference (right). (c) Same as (a) but for Guatemala. (d) same as (b) but for Guatemala. Table 1 summarizes the criteria used to generate the subsamples in each month that were selected for their ability to perform as well as the all-member mean across all three metrics (MSESS, HSS10, HSS90) and were the top performers for HSS90 in each month as illustrated in Figure 6.2.

6.4.3 Models Fail to Represent Full Rainfall Distribution when Process Error Increases

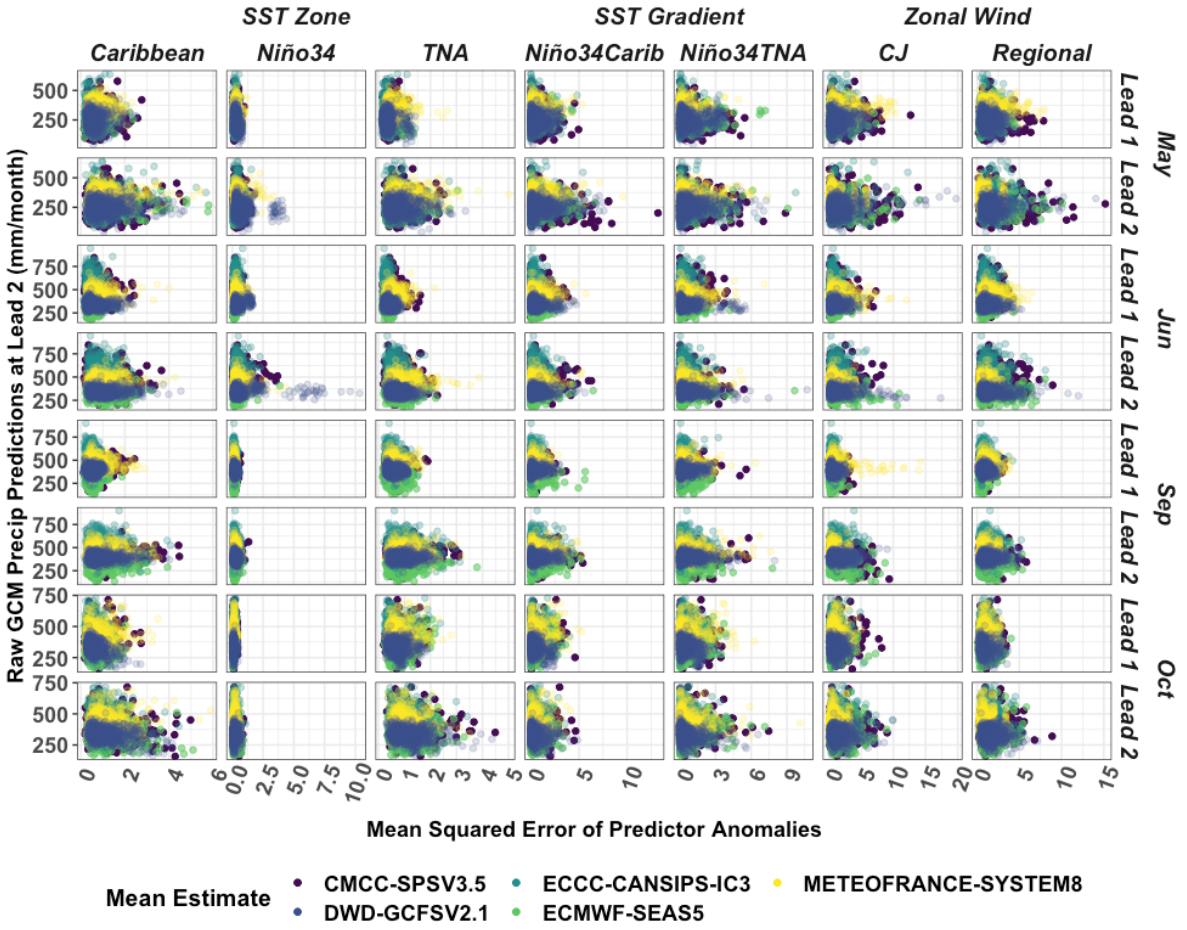


Figure 6.4 Process error versus rainfall estimates. Relationship between the spatial average of the raw *rainfall prediction* at 2-month lead (y-axis) over 1993-2016 over each country *and* the MSE of the predictors (x-axis) for different zones/predictors (columns*) grouped by lead times (one month ahead or two months ahead, rows)* for different target months (rows), color-coded by the *five* models evaluated.

We find when the models struggle to represent certain processes, as seen through estimates of their predictor error using MSE, they are less able to capture the full range of the rainfall distribution (Figure 6.4). This can be seen in the way the maximum rainfall predictions trend downward (y-axis) as predictor error increases (x-axis) across predictor type (columns) in each month tested (rows). Similarly, an upwards trend in the lowest rainfall predictions is also seen across predictors as error increases. For SST zones like Niño3.4 and for shorter lead times (one-month lead), the range in error is smaller than in other regions at longer lead times (e.g. Caribbean at two-month lead), but the slope is steeper, meaning small process errors can have a greater influence on the models' ability to predict rainfall (Figure 6.4).

Despite some structural differences among the models (e.g. DWD System 21 has a smaller rainfall range than the other models, Figure 6.4), the relationship between process error and the estimated rainfall range is consistent. This consistent cone-shaped pattern across models showcases the value of the process-informed subsampling method to support operational

forecasting. While a regional forecaster may not have as much familiarity as a model developer with the structural nuances of each model, they can leverage their expertise in regional rainfall drivers to select members that best represent important predictors across ensembles. Some members with low process error may still be selected from a more limited rainfall distribution that is unique to their parent model (e.g. in the case of DWD System 21), but by minimizing some key process errors, forecasters can increase the likelihood that a subsample better captures the entire rainfall distribution.

6.5 Conclusions

Subseasonal predictions present a gap between more well-established weather and seasonal climate forecast fields that could benefit from optimization (Vitart and Robertson, 2018). Our subsampling method presents an opportunity when a forecast is issued (initialization) to select the model members that by chance correctly represent key rainfall predictors to improve rainfall predictions a month ahead. We show the models are less able to capture the full range of the rainfall distribution when SST and zonal wind error increases (Figure 6.4). This method presents potential for better extreme rainfall preparedness planning by improving extreme rainfall predictions. Process-informed subsampling may improve MME rainfall skill especially when regional rainfall is driven by slowly-changing processes like SSTs. Given the interannual variability of key processes, future work could enhance our selection procedure by selecting subsamples based on representation of key processes that are empirically derived from a moving evaluation window. This technique is ideal for constraining regional forecasts based on knowledge of regional rainfall mechanisms and could especially benefit regions where SSTs are key rainfall drivers.

6.A Appendix - Texts

Text 6.A1 Formulas Used to Process and Evaluate the Ensemble Members

The skill scores were applied individually for each grid cell on data that had been converted from raw values into monthly standardized anomalies. Standardized anomalies were calculated monthly across the time series (i.e. months were calculated individually) for each grid cell individually following Equation 1:

$$\boldsymbol{Standardized\ Anomaly} = \frac{\mathbf{x} - \mu(\mathbf{x})}{\boldsymbol{\sigma}(\mathbf{x})} \quad \text{..... Equation 1}$$

$x = estimate\ (either\ a\ forecast\ or\ an\ observation)$
$\mu(x) = mean\ of\ estimate\ over\ time\ period\ (e.g. 1993 - 2016)$
$\sigma(x) = standard\ deviation\ of estimate\ over\ time\ period\ (e.g. 1993 - 2016)$

Predictor representation was scored using mean squared error (MSE, Equation 2) and a Percent Correct Score (Equation 3). MSE is a measure of average accuracy, calculated by adding the squared difference between the observations and the forecasts for every point in time and dividing by the total number of instances (every date). When calculated for an SST Gradient, the MSE score alone was used. When calculated for an SST zone, the average MSE score over the entire grid-box was calculated.

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\hat{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x}_{\boldsymbol{i}})^{\mathbf{2}} \quad \text{..... Equation 2}$$

$\hat{x} = observation$
$x = forecast$
$n = total\ number\ of\ instances$

Percent Correct Score was calculated for the binary zonal wind direction variables by calculating the number of times the model member predicted wind direction correctly divided by the total number of predictions of wind direction over a grid-box (e.g. CJ).

$$\boldsymbol{Percent\ Correct} = \frac{\boldsymbol{a}}{\boldsymbol{n}} * 100 \quad \text{..... Equation 3}$$

$a = correct\ detections$
$n = total\ number\ of\ instances$

The formulas of the skill-scores used to evaluate the all-member ensemble mean and subsample mean rainfall predictions are summarized below:

The MSE skill (MSE-SS) is a measure of how the average accuracy of the forecasts compares to using a climatology alone (average estimate across a time series). MSE-SS is calculated by dividing the MSE_f, which is MSE (Equation 1) calculated for the forecasts, with the MSE_{clim}, which is MSE (Equation 1) calculated for the mean estimate across the time series (climatology mean over 1993-2016). This ratio is subtracted from 1, meaning a perfect score will result in 1, while a negative score means the MSE_f is worse than the MSE_{clim}.

$$\boldsymbol{MSE - SS} = \boldsymbol{1} - \frac{\boldsymbol{MSE_f}}{\boldsymbol{MSE_{clim}}} \quad \text{..... Equation 4}$$

$MSE_f = MSE\ forecasts$
$MSE_{clim} = MSE\ climatology$

The Heidke Skill Score (HSS - Heidke, 1926) assesses a forecasts' ability to discriminate between different categories of events. In this study, HSS was applied to decile discrimination at the low and high ends of the rainfall distribution, essentially testing the forecasts' ability to detect high rainfall extremes (HSS90 - an event is defined as rainfall above the 90th percentile of the monthly climatology) and low rainfall extremes (HSS10 - an event is defined as rainfall below the 10th percentile). Unlike other scores, like hit rate (correct detection rate of an event), or false-alarm rate (incorrect detection rate by predicting an event occurs when there is none), the HSS is equitable (Hogan et al., 2010), meaning random or rigged forecasts will all score 0.0 (no skill). As summarized by the following formula, HSS is calculated based on a 2x2 contingency table of hits (correct detections - a), false alarms (predictions of an event when there is none – b), misses (failures to detect an event – c), and correct rejections (not predicting an event will occur when there is none).

$$\boldsymbol{HSS} = \frac{\boldsymbol{a+d-a_r-d_r}}{\boldsymbol{n-a_r-d_r}} \quad \text{..... Equation 5}$$

$a_r = (a+b)(a+c)/n$
$d_r = (b+d)(c+d)/n$
$a = correct\ detections$

$b = false\ alarms$
$c = misses$
$d = correct\ rejections$
n = *total number of instances*

Text 6.A2 Observational Correlations that Affect Strength of Subsampling Method

The R-squared values listed in the inset of Figure 6.A1 provide a measure of how quickly the observations change on average between the reference period and the months in which the predictors are tested, which affects the strength of the subsampling method. These R-squared values show the correlation between the temporal mean of the predictors in last two weeks just prior to model initialization for a given forecast month, and the average monthly mean of the predictors in the two months following model initialization (one- and two- month leads). The Niño 3.4 region has the most stable SSTs compared to other predictors used and is more persistent during the September and October forecasting periods relative to earlier months (Figure 6.A1 inset).

June is the least persistent observational forecasting period, with all observational correlations for wind direction equal to an R-squared of less than 0.1 and observational SST R-squared values below 0.5 for all zones tested (Figure 6.A1 inset). Although the observed correlation between the reference data and two-month lead is typically worse than for one-month lead (Figure 6.A1), two-month lead predictors were also tested as subsampling options because longer lead times give the ensemble members more time to diverge prior to evaluation.

Furthermore, although 925 hPA is at the peak intensity of the low level jets (Amador, 2008; Hidalgo et al., 2015), we used 850 hPA for wind direction because 850 hPA is available across all models evaluated and still falls within the range of the pressure levels of the low level jet streams. The model members were also consistently able to correctly identify the zonal wind direction within the maximum wind zone of the CLLJ (80W, 70W, 12.5N, 17.5N - Amador, 2008). There was more variability in the correct identification rate within the CJ and a broader zone that circumscribed the CLLJ and Central America, so we used a larger 'Regional Wind' box and CJ box as our final predictors to test how well the members identified zonal wind (Figure 6.A1).

Text 6.A3 Software

The subsamples were generated using XCast, a python software package that was developed to improve forecast post-processing and evaluation (https://github.com/kjhall01/xcast). The subsamples were processed using a combination of several other packages, including pysteps (Pulkkinen et al., 2019; Imhoff et al., 2023). We plotted all figures in R using tidyverse and ggplot2 (Huang & Zhao, 2020; Wickham, 2016; Wickham et al., 2019).

Text 6.A4 Sensitivity Test of Subsample Skill to Ensemble Size

We tested the sensitivity of the results to filter sizes in step four of the methodology, which refer to how many members are selected in each subsampling step. Figure 6.A2 shows two examples of sensitivity tests run on the top-performing subsamples: (1) using different numbers of members to constrain the top subsample in May over Guatemala (using

representation of TNA alone; Figure 6.A2a), and (2) constraining the top subsample in September over Costa Rica (using representation of Niño3.4 at one-month lead, Niño3.4 at two-month lead, and Regional Wind at one-month lead; Figure 6.A2b). Several studies have examined the effects of ensemble size on forecast skill (e.g. Buizza, 2008; Buizza et al., 1998; Buizza & Palmer, 1998; Leith, 1974; Leutbecher, 2019). Of these studies, many show increases in skill as size increases, but that the change in skill begins to saturate at ensemble sizes as small as eight (e.g. Buizza & Palmer, 1998; Leith, 1974). Similar to what these previous studies have shown, in the comparison in Figure 6.A2, skill often declines abruptly under 10 ensemble members, but the added value of selecting more members that best represent key predictors do not dramatically increase when the number of ensembles is at/above ten in the final subsample.

The example using May forecasts in Guatemala shows the effects of changes in ensemble size alone (Figure 6.A2a). The effect of selecting fewer ensemble members is illustrated most clearly when comparing MSE-SS (Figure 6.A2a bottom row). Selecting members that best represent TNA at one-month lead begins to show an effect on skill using the top 30 members or less, and skill drops off between the top-10 and top-5 member subsamples.

When multiple predictors are used, the change in ensemble skill is often an indicator of sensitivity to a particular predictor over ensemble size (Figure 6.A2b). For instance, in Costa Rica, filtering too strictly over Niño3.4 at one-month lead will limit skill (filtering by 60 members or less using Niño3.4 one-month lead representation in step 1 is associated with lower skill). Filtering more strictly for Niño3.4 at two-month lead is more useful (Figure 6.A2b). Looking at differences within the top 80 members, for instance (left-hand side of Figure 6.A2b), shows improved skill using the top 30 members that represent Niño3.4 SSTs at two-month lead over only selecting the top 40 members. In step three (still looking at left-hand side of 6.A4b), filtering more strictly for representation of Regional Wind at one-month lead within the top 30 or 20 members often performs better up to the top 10 members (subsamples that use top 10/15 members for Regional Wind at one-month lead representation often have higher skill than selecting top 20 members that represent Regional Wind at one-month lead).

The top subsample for Costa Rica in September that is highlighted in the main article text was identified using HSS90 skill (significance between this subsample and the all-member ensemble is plotted in Figure 6.A2b). This subsample represents a balance between filters, i.e. less strict filtering for the top members that represent Niño3.4 SSTs at one-month lead (top 80 members), more strict filtering for the top members that represent Niño3.4 SSTs at two-month lead (top 30 members), and additional filtering for the top 15 members that represent Regional Wind at one-month lead. While this subsample was highlighted in the main text, Figure 6.A2b shows how filtering for the top 20 members that represent Niño3.4 SST at two-month lead in step 2 and the 10 members of Regional Wind at one-month lead in step three also performs well on multiple metrics.

Text 6.A5 Sensitivity Test of Subsample Skill to Years Analyzed

The all-member and subsampled anomaly predictions change year to year, and more closely follow the observations in some years over others (Figure 6.A3). The subsampling method possibly has a greater effect on constraining the entire predicted range of the ensemble than

changing the mean estimates (Figure 6.A3). In May over Costa Rica (Figure 6.A4 top row, left panel), the ensemble is the most constrained, as it only contains five members, whereas in other months and regions, the average percent change in the range of the predicted estimate is around 30-40% for subsamples that contain 10-15 members (Figure 6.A3).

Although the constrained subsample range visually seems to follow the interannual pattern of the observations more closely than the all-member range, both forecasts completely miss the anomaly pattern in some years (e.g. May 2011 in Guatemala Figure 6.A3 left panel bottom row). Because of these misses, we see opportunities to refine the approach beyond using a constant set of process-based criteria to generate subsamples and more flexibly emphasize different predictors for particular years (e.g. prioritize filtering members that represent Niño 3.4 SSTs in strong El Niño years), which may be related to the changing strength of predictors like ENSO and their relative importance to regional rainfall.

Although the bias varies interannually (Figure 6.A3), the sensitivity analysis suggests the subsampling method is not necessarily sensitive to the time frame of analysis. Randomly dropping five years from the analysis has some effect on the skill of the all-member and top-subsample skill, but this does not often significantly affect the difference between the spread in skill of the all-member ensemble mean scores and the top subsample scores across each country (Figure 6.A4). The difference in skill between the all-member and top subsample case for a given country and month often are still within the same level of significance (e.g. $p < 0.001$) regardless of whether the entire time-series (1993-2016) or a random subset that drops five years is used (significance tests plotted in Figure 6.A4. Additionally, when a significance test was run between the 'All Years' and 'Drop 5' cases within each approach (e.g. between the top subsamples over Costa Rica), no significant differences were found between the boxplots except for MSESS in October (not plotted in 6.A4 because N.S. occurs for all cases excluding October MSESS).

While year-to-year variability is worth investigating further to see how particular years may be more difficult to predict or, alternatively provide windows of opportunity for improved ensemble-based forecasts and subsampling methods, the average performance of the top subsamples is not likely dependent on performing well in key years alone. It would be worth continuing to test subsampled skill over more years outside of the 1993-2016 window, however, to see how the relevance of the different predictors maintains over larger time-periods.

To further operationalize the subsampling approach, next steps would be to identify top subsamples over a moving testing window, and then use a separate verification window to assess how long the top-subsample criteria maintains outside the window of analysis. This step could be combined with comparing different lengths of testing windows with consideration for potential years of interest (e.g. Strong El Niño periods) to determine optimal subsampling criteria prior to applying a subsampling approach in a real operational forecasting scenario.

Text 6.A6. Average Model Representation in Subsamples

Subsample members come from different models depending on each models' performance in a given year. Models will vary from having zero representation in one year to making up

more than 80% of the subsample in another, depending on their members' predictor scores. Table 6.A2 highlights the average percentages of the members that are represented by a given model over the 24 year evaluation period (1993-2016) for each subsample type.

6.A Appendix- Figures

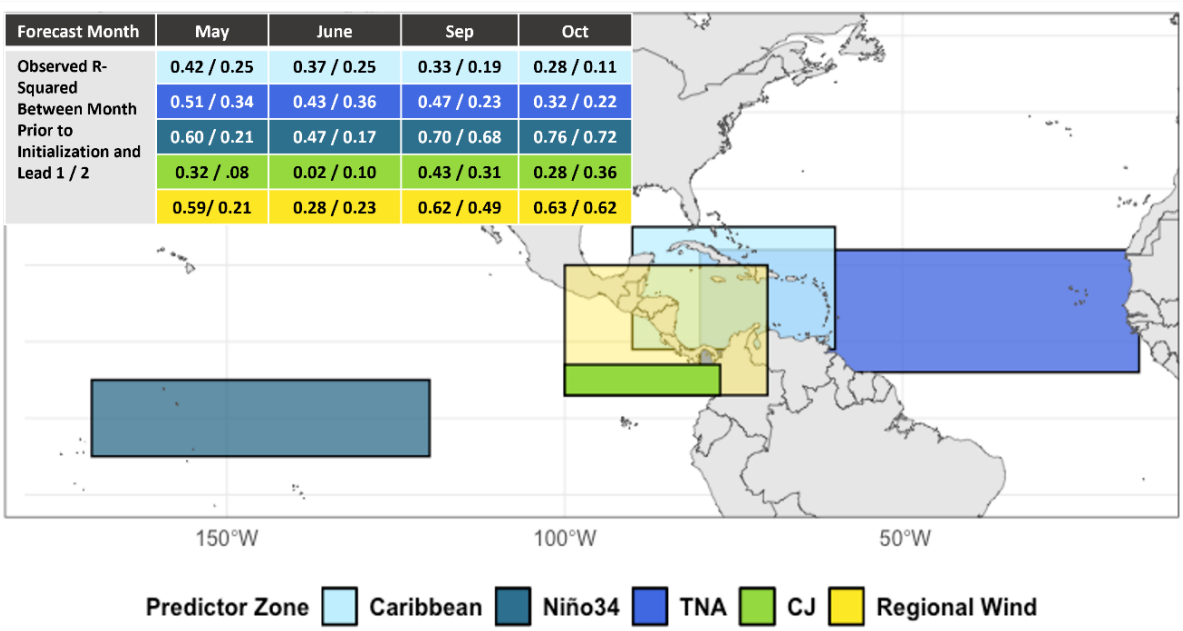


Figure 6.A1 Predictor locations used to test member performance plotted spatially (same as Figure 6.1c) with an additional inset that includes the R-squared values of the correlation between the observed reference period and the predictor period for each prediction evaluated (Lead 1/Lead 2). Colours indicate predictor zone ordered by column of forecast month at two-month lead, May, June, September, and October.

6.A Appendix - Figures

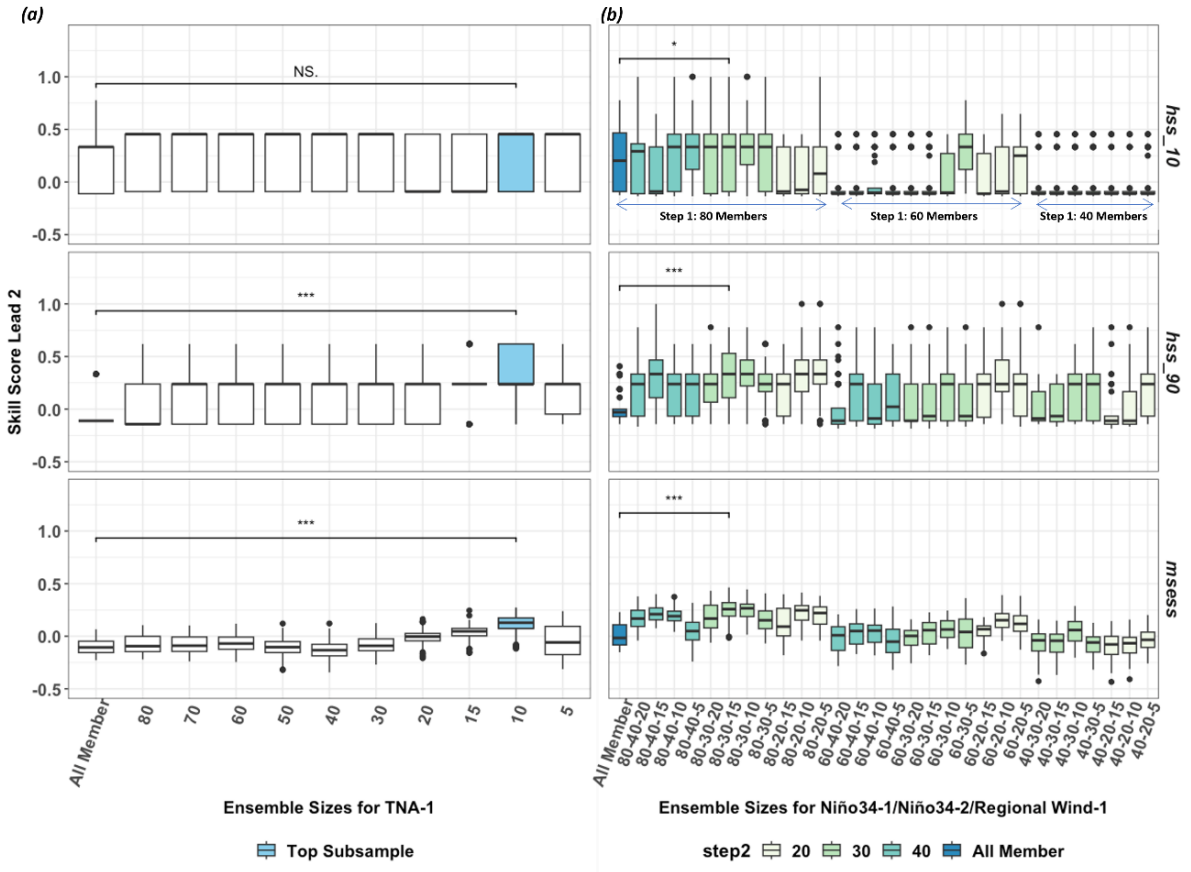


Figure 6.A2. Assessing the effect of filter size on subsample skill. Comparisons are performed using boxplots with significance tests plotted (using a t-test) between the top-performing subsamples and the all-member mean ensemble for each skill score (HSS10 top row, HSS90 middle row, MSE-SS bottom row). *** indicate $p < 0.001$; ** indicate $p < 0.01$; * indicate $p < 0.05$. NS indicates non-significant. (a) Skill at lead two is plotted for Guatemala in May using different filter sizes for the top-performing predictor (TNA at one-month lead). Top subsample is highlighted in shaded box. (b) Skill at lead two is plotted for Costa Rica in September using different filter sizes for the top-performing predictor combination (Niño3.4 at one-month lead; Niño3.4 at two-month lead; Regional Wind at one-month lead). To clarify the comparison between step differences across multiple predictors, filter sizes are ordered by step 1 (80, 60, 40) as labelled in the top right box; followed by step 2 (40;30;20), which is shaded across box-plots, and finally by step 3.

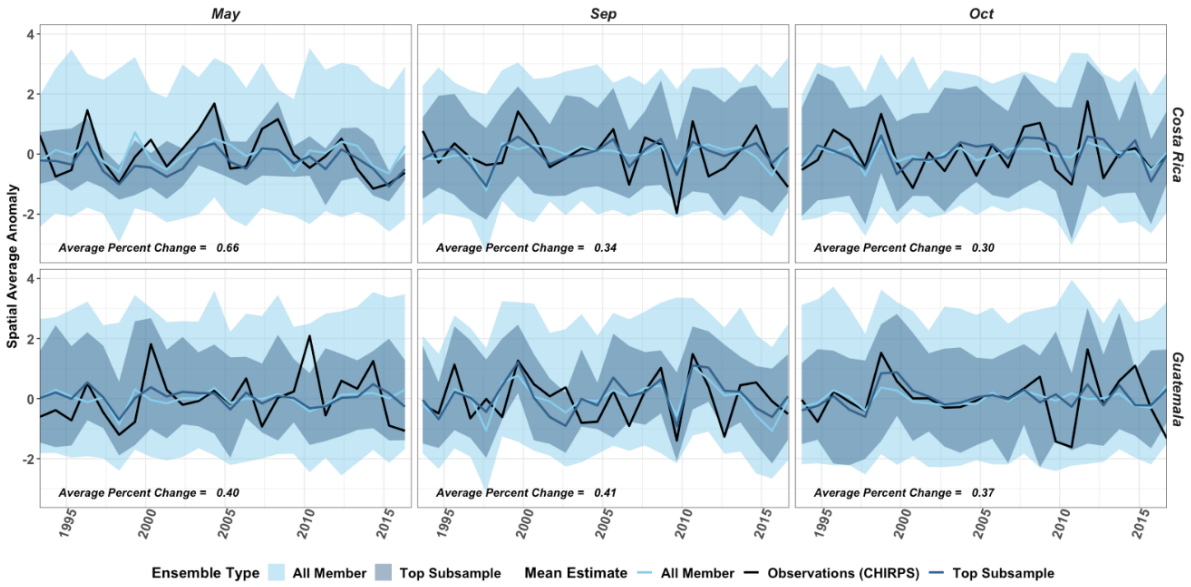


Figure 6.A3. Comparing time series of the all-member and subsampled ensembles. Time series of spatially averaged total monthly rainfall anomalies for Costa Rica and Guatemala (rows) and three months (columns). Observed rainfall is shown in black (CHIRPS). The all-member ensemble (light blue) and best performing subsample (grey) are shown using the mean (lines) and range (maximum and minimum) of members. Table 6.1 in main article summarizes which criteria are used to generate the subsamples in each month/country. Average percent change is calculated for each panel to show the average difference between the predicted range of the all-member ensemble and the predicted range of the top subsample over the time series.

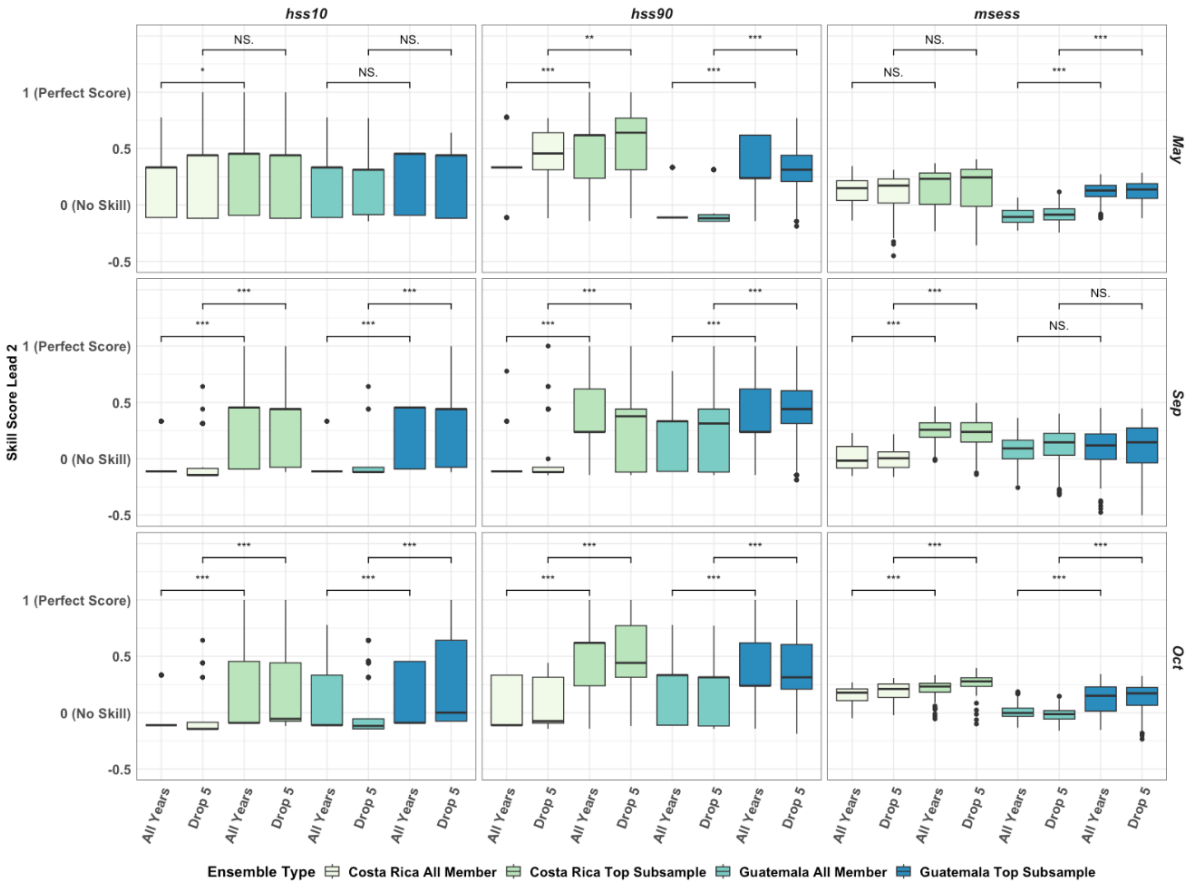


Figure 6.A4. Sensitivity of skill to years analyzed. Skill comparison between the baseline, using all years in the analysis (1993-2016), and randomly dropping five years from the dataset. Boxplots show the spread in skill using HSS10, HSS90, and MSE-SS metrics (columns) across Costa Rica (left four boxplots for every panel) and Guatemala (right four boxplots for every panel) for the all-member and top subsamples for May, September, and October (rows). Significance tests using t-tests are run between the all-member mean scores and the top subsamples for both the 'All Years' and 'Drop 5' cases in both countries (4 tests per panel). *** indicate $p < 0.001$; ** indicate $p < 0.01$; * indicate $p < 0.05$. NS indicates non-significant.

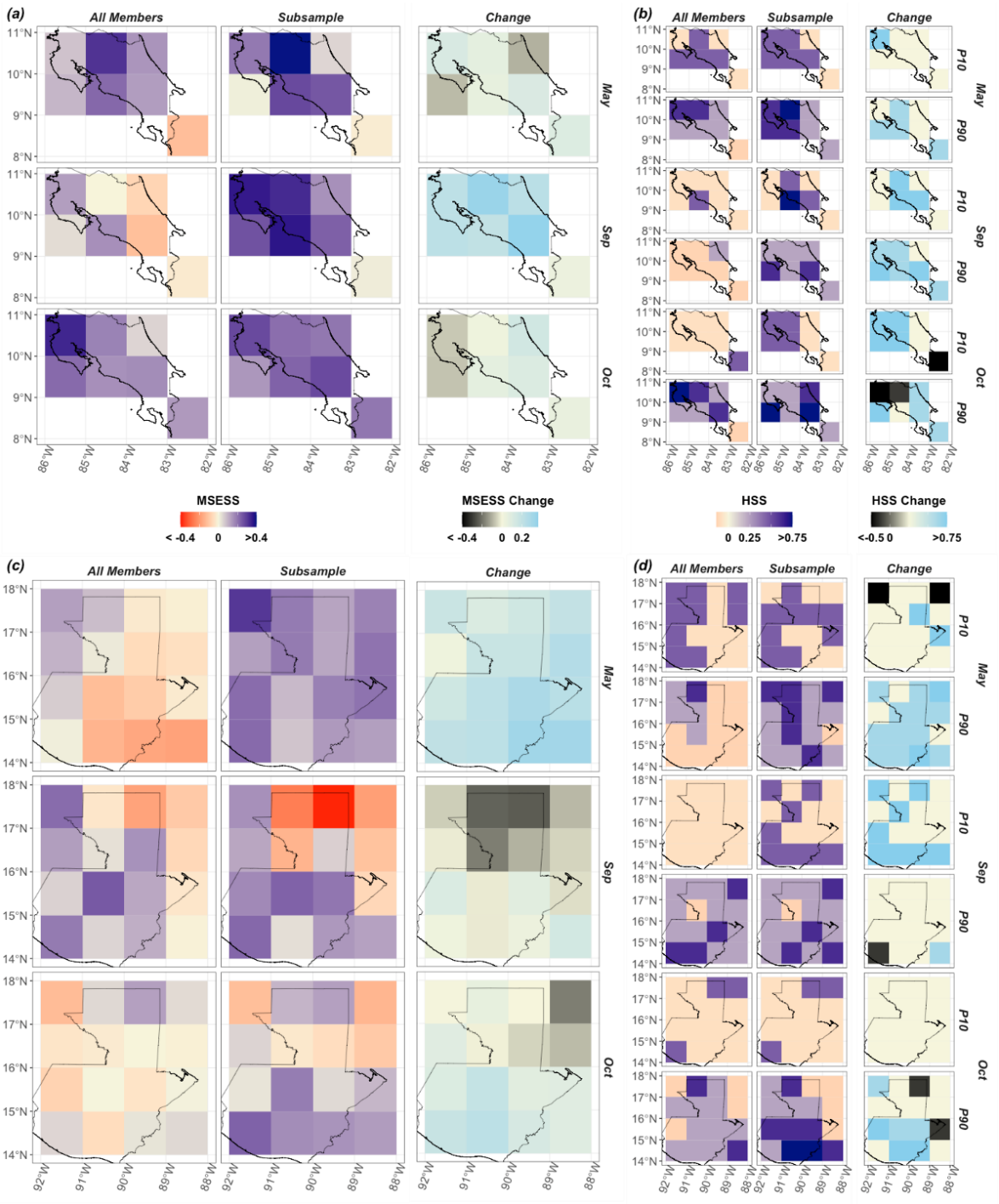


Figure 6.A5. Spatial skill (same as Figure 6.3) but plotted at the raw 1° resolution where the CHIRPS observations are bilinearly regridded to the raw model spatial resolution before the point-wise skill is calculated. (a) Spatial skill of all-member ensemble (left), top performing subsampled ensemble (center) and difference between the two (right) in Costa Rica. Skill is based on the entire rainfall distribution using MSE-SS. (b) Spatial skill for detection of low rainfall extremes (P10) and high rainfall extremes (P90) using HSS in Costa Rica, again for all-member (left), subsample (center), and their difference (right). (c) Same as (a) but for Guatemala. (d) same as (b) but for Guatemala. Table 6.1 in main text summarizes the criteria used to generate the subsamples in each month.

6.A Appendix - Tables

Table 6.A1. Summary of Models Used in Evaluation. Contributing center listed next to model name, total hindcast members used in analysis, and reference for the model. Abbreviations: CMCC, Fondazione Centro Euro-Mediterraneo Sui Cambiamenti Climatici; DWD, Deutscher Wetterdienst; ECCC, Environment and Climate Change Canada - Canadian Meteorological Centre; ECMWF, European Centre for Medium-Range Weather Forecasts.

Contributing Center	Model	Hindcast Members	Reference
CMCC	System 35	40	Gualdi et al. (2020)
DWD	System 2.1	30	Frölich et al. (2021)
ECCC	Cansips-IC3	10	Lin et al. (2021)
ECMWF	SEAS5	25	Johnson et al. (2018)
Meteo France	System 8	25	Batté et al. (2021)
Total Models in Analysis: 5 ; Total Members in Analysis: 130			

Table 6.A2 Average Model Representation in Subsamples over the hindcast period (1993-2016). Criteria used to select top subsamples for each month correspond to the criteria in Table 6.1.

Month	Model Name (Total Members)	Average Model Member Representation over (1993-2016) (%)	
		Top Costa Rica Subsample	Top Guatemala Subsample
May	CMCC (40)	42%	35%
	DWD (30)	40%	24%
	ECCC (10)	3%	18%
	ECMWF (25)	11%	17%
	METEOFRANCE (25)	5%	7%
September	CMCC (40)	29%	29%
	DWD (30)	9%	18%
	ECCC (10)	3%	10%
	ECMWF (25)	30%	20%
	METEOFRANCE (25)	29%	23%
October	CMCC (40)	28%	22%
	DWD (30)	30%	28%
	ECCC (10)	8%	14%
	ECMWF (25)	14%	24%
	METEOFRANCE (25)	20%	13%

CHAPTER 7

Discussion

Many approaches are available to predict rainfall at the S2S scale, and the method chosen can significantly affect forecast skill. The skill of statistical, dynamical and hybrid forecasting methods changes spatially and temporally. The following discussion identifies common themes that arose across the three main chapters (4-6) and identifies the comparative advantages of the different approaches evaluated.

7.1 Seasonal Variations in Dynamic Model-based Ensemble Skill

The seasonal patterns in skill largely align across the chapters. Skill is often highest in July and August, especially in locations with a consistent mid-summer dry period. For instance, SEAS5 has the highest skill over the CADC in these months (Figures 4.2, 4.3). When both months are included in a forecast (e.g. JJA), forecast skill is also often higher in locations with a distinct mid-summer dry period (e.g. North Pacific, Figure 5.3). The increased predictability of the mid-summer dry period may be due to its relative consistency compared to rainfall in the early and late wet season months, as AOGCMs can often approximate a seasonal decrease in rainfall despite not representing all the underlying causal mechanisms (Rauscher et al., 2008).

Forecast skill is also often high in the late wet season when examining model skill using the entire rainfall distribution (e.g. MSE-SS, CRPSS, or 2AFC metrics). For instance, the MSE-SS and CRPSS scores for SEAS5 are high in September and October (Figure 4.2/4.3), and both NMME and C3S have high 2AFC scores in ASO (Figure 5.5). This increased skill may be related to the ENSO teleconnection, as ENSO is more strongly correlated with rainfall in ASO compared to earlier months (Figure 5.8). In the late wet season, statistical models based on ENSO, or a combination of ENSO and TNA also have higher skill (Figures 5.5, 5.7).

The comparative advantage of using dynamic model-based ensembles may often be in the early wet season despite the relatively lower skill of the AOGCMs in May and June. June was identified as a particularly challenging month to predict rainfall. In June, SEAS5 rarely attained positive skill (e.g. Figure 4.2), and in Chapter 6 it was more difficult to find subsampling options that improved the ensemble-mean skill. The skill of the statistical and hybrid models, however, also suffers in early wet season, and the MMEs more frequently significantly outperform the other forecasting approaches during these months (Figure 5.5).

Detections of rainfall anomalies at the tails of the rainfall distribution are critical for extreme weather preparedness yet are often unreliable. As lead time increases, the low rainfall extremes detections are significantly worse in key months (e.g. most SEAS5 lead times have negative skill when detecting P0-10 and P10-20 rainfall events in October over the CADC; Figure 4.4). Skilful detections of low and high rainfall extremes are also more spatially and

seasonally limited within Central America (Figure 5.6). The seasonal variation in extremes detection rates is also opposite to the findings for predictions of rainfall variability generally. Extremes detections are often worse in the late wet season than the early wet season (Figure 5.6). For instance, SEAS5 detections of low rainfall extremes perform poorly in September (Figure 4.4), and ensemble skill for both low and high rainfall extremes detections in ASO is low using both the NMME and C3S MMEs (Figure 5.7).

7.2 Spatial Variations in Dynamic Model-based Ensemble Skill

Spatial patterns in skill also often align across the thesis chapters. The North Pacific is a relatively high skill zone for rainfall detection (Figures 5.3-5.7), which includes most of the southeast CADC (another high skill zone, Figure 4.3). Some of the spatial patterns change, however, when looking at monthly versus seasonal estimates of rainfall. El Salvador stands out within the North Pacific, for instance, as a country with relatively lower skill for SEAS5 in many months (e.g. in June and September, Figure 4.3). When the ensembles are used to generate three-month-mean seasonal forecasts, El Salvador no longer stands out as a particularly low skill zone compared to other subregions using SEAS5 alone or one of the combined MMEs (e.g. Figures 5.4, 5.6). This may be related to how forecasting over multiple months at a time changes the skill. El Salvador is not a relatively low skill zone in July, August, or October (Figure 4.3), and the predictability in these higher skill months may make up for relatively lower skill in June and September. This change in skill between monthly to seasonal forecasts highlights how seasonal rainfall forecast analysis can obscure some of the monthly variability in forecast performance, as forecasts can perform reasonably well at predicting three-month-means even if they fail to predict rainfall well in certain months. In most cases this may not be a concern, but the models' seasonal forecasts may underperform in years when less predictable months (e.g. June) are more anomalous and have a greater effect on the three-month-mean.

Eastern Nicaragua, however, stands out as a consistently low skill zone for both C3S and NMME (e.g. Figures 5.4, 5.6). There are a few possible reasons for this relatively low skill. Limited rain gauge availability presents a well-known challenge for observational rainfall datasets across Central and South America (Carvalho, 2020), and gauges are even more limited over the Caribbean slope as compared to the Pacific slope in Central America (e.g. Maldonado et al., 2016). Regional predictability characteristics may also affect AOGCM skill over eastern Nicaragua because the rainfall could be more determined by atmospheric variability than oceanic influences, which are easier to simulate, or the structural nature of the models themselves may also affect the spatial variability in skill (Chapter 5 discussion). An additional factor that is more unique to Nicaragua may be in play though. Nicaragua is currently under a dictatorship that has enabled massive deforestation in recent years on a scale that could significantly affect regional rainfall (e.g. Sesnie et al., 2017). The models may fail to account for this, which highlights the importance of including land surface dynamics in rainfall predictions in addition to modelling oceanic and atmospheric processes.

While topography remains a known challenge for numerical weather prediction (Serafin et al., 2018), topography alone does not provide a unifying explanation for spatial patterns in rainfall skill (e.g. Table 4.3). SEAS5 has more limited skill over El Salvador in particular months, however, which may be due to the interactions between the terrain and key

processes (e.g. mountainous terrain shielding El Salvador from the easterly trade winds and limiting the connection between the CLLJ and local rainfall, Durán-Quesada et al., 2017). Going forward, it would be worth continuing to examine the interactions between local terrain and regional mechanisms to test how these two factors combine to affect regional rainfall skill.

7.3 Relying on ENSO Alone is Not Sufficient for Predicting S2S Rainfall

One of the motivations throughout this thesis was to assess the added value of dynamic model-based ensemble forecasts compared to relying on well-known teleconnection patterns like ENSO. ENSO is a dominant rainfall driver partially due to its modulation of regional moisture transport (Durán-Quesada et al., 2020), but ENSO alone cannot explain cumulative rainfall deficits in many locations within Central America (Muñoz-Jiménez et al., 2019). AOGCM derived forecasts could be more skilful due to their more comprehensive simulations of the earth system, including atmospheric variability (Doblas-Reyes et al., 2013; Bauer, Thorpe and Brunet, 2015). It was unclear, however, how much more skilful these forecasts could be over the region given the ENSO teleconnection is also a key driver of AOGCM skill (e.g. Mo & Lyon, 2015; Scaife et al., 2019). Other studies have shown that while ENSO is connected to dynamic model-based forecast skill, AOGCMs also derive skill from other processes when predicting rainfall in nearby regions (Gubler et al., 2020; Zhao et al., 2021). By more comprehensively estimating the earth system using multiple processes, dynamic ensembles may outperform statistical models that solely rely on ENSO.

This work shows how although dynamic model-based ensemble skill may be affected by ENSO, the comparative advantage of using these forecasting systems may often be when the correlation between ENSO and regional rainfall is relatively weaker. Although all forecasting methods assessed tend to perform worse in the early wet season (Figure 5.5), the dynamic model-based ensembles better maintain their comparative edge over the statistical models that purely rely on this teleconnection (Figure 5.5). ENSO phases, however, also showcase the potential limitations of AOGCMs. Because AOGCMs are constructed to represent the current understanding of how the earth system works, these models may perform better in more typical conditions, and become less reliable in unexpected situations. In May and June, for instance, SEAS5 predictions tend to perform less well during strong El Niño phases (Table 4.4). ENSO onset is typically in the spring, developing in summer, and maturing in boreal winter (Kim and An, 2021). When a strong El Niño phase develops earlier than expected in the season, the AOGCMs may not account for this phenomenon as well.

Low and high rainfall extremes also present atypical situations when the ENSO teleconnection is not a sufficiently significant predictor to obtain a skilful prediction. For instance, multiple forecasting approaches struggle to detect low and high rainfall extremes in ASO, including the statistical models based on ENSO (Figure 5.7). Although this period is when ENSO is most strongly associated with regional rainfall (Figure 5.8), this teleconnection alone is not sufficient for extremes detection applications. There are alternative ways, however, to use the ENSO teleconnection to improve rainfall extremes detections. By still using the AOGCMs to predict rainfall extremes, an S2S forecast can benefit from their more comprehensive estimates of oceanic and atmospheric variability compared to relying on ENSO alone. S2S detections of extreme rainfall anomalies can then

be improved by selecting a subgroup of members that best represent SSTs in the Niño 3.4 region in the late wet season (e.g. over Costa Rica in September, Figure 7.2). This demonstrates how in times when ENSO is often a significant driver of regional rainfall (e.g. late wet season), predicting SSTs correctly in related locations (e.g. Niño 3.4) is especially important to accurately detect rainfall at the tails of the rainfall distribution.

7.4 Process-Informed Analysis can Improve Rainfall Predictions

Modelling key processes is important for predicting rainfall over Central America. ENSO for instance, is shown to be a key phenomenon that drives AOGCM skill, as MME predictions often improve in the late wet season when ENSO is more related to rainfall (Figure 5.5). Subsampling members based on their ability to represent key processes can further improve their skill. Selecting the top ten members that represent TNA SSTs in May, for instance, can improve skill over Guatemala (Table 6.1, Figure 6.3). Chapter 6 also demonstrates how the ensemble members' inabilities to represent key processes is correlated with their abilities to represent the entire rainfall distribution (Figure 6.4), which may explain why members with lower process error have higher skill when detecting rainfall extremes.

Representation of the entire rainfall distribution is important for multiple forecast applications. For instance, DWD System 21 is one of the only European models that was consistently identified as one of the worst models at discriminating between tercile categories of rainfall for JJA forecasts (Figure 6.0). This model also has a much more limited rainfall distribution than the other models in multiple months (Figure 6.4). Chapter 6, however, shows that even when a variety of models are subsampled, each with a range in rainfall distributions that are unique to their structure, the subsampling method can increase the probability of selecting members across a larger rainfall distribution. Regional forecasters do not have to be intimately familiar with each model structure to benefit from post-processing models using this subsampling approach. They can rely on their regional expertise of relevant rainfall-generating mechanisms and improve model skill so long as they pull members from a few models that have a varying range of rainfall distributions.

Using process-informed approaches to evaluate and optimize model skill may become increasingly important as the earth system changes in response to human-induced climate change. The intensity and duration of the mid-summer dry period, for instance, varies regionally (Alfaro, 2014; Corrales-Suastegui et al., 2020; García-Oliva & Pazos, 2021; Maurer et al., 2022), and its characteristics may change in the 21st century (Rauscher, Kucharski and Enfield, 2011; Anderson et al., 2019; Corrales-Suastegui, Fuentes-Franco and Pavia, 2020; Vichot-Llano et al., 2021; Maurer et al., 2022). Evaluating process-based representation is needed as a continued practice to improve the understanding of how different processes relate to rainfall and forecast skill.

Going forward, the subsampling technique could continue to be refined to diagnose process representation and optimize rainfall skill. This would include expanding the analysis to a moving testing and validation window to assess the interannual variability of climate mechanisms and exploring how the usefulness of the approach changes across lead times. While the subsampling method proposed in Chapter 6 is presented as an alternative post-processing technique on its own, there are also opportunities for this method to be combined

with the hybrid methods presented in chapter five. For instance, a hybrid forecast using TNA could be generated using only the members that best represent SSTs over that region.

7.5 Limitations

The research in this thesis centred on forecasts as they are operationally deployed today. The CA-COF was selected as one operational avenue to guide the forecasting analysis, which meant the forecasts tested in Chapter 5, for instance, used three-month-means at one-month initialization (same as CA-COF outlooks). The target periods MJJ and ASO were also selected due to their current use in CA-COF outlooks, but because these three-month estimates overlap with multiple phases of the wet season (e.g. early, mid-summer dry period, and late), seasonal variations in climate patterns and rainfall become more difficult to analyse. A few metrics were selected to evaluate the AOGCMs and MMEs in the thesis to keep the evaluations simple, clear, and aligned with metrics that are already used in operational settings like the CA-COF. While this can ease communication, additional metrics, including more probabilistic analyses, would be useful to compare the models' performance.

The time periods used in the analysis also limit some of the findings of the studies. The choice of hindcast years can significantly affect forecast skill assessments (Tippett, Goddard and Barnston, 2005; Shi et al., 2015; Risbey et al., 2021). Using hindcasts does not always allow like for like comparisons with forecasts in real time in part due to initialization differences (Risbey et al., 2021). For instance, some AOGCM hindcasts will be standardized on the first of the month, while their forecasts combine multiple initialized dates (e.g. CFSv2 in CDS, see Table 2.4.1). The length of the hindcast period can also significantly affect forecast skill (e.g. Shi et al., 2015). In this thesis, the longest hindcast period analysed was 35 years (Chapter 4), but Chapters 5 and 6 only include an evaluation period of 24 years due to the limited overlapping window between AOGCM hindcast periods. Due to this limited time window, some leave-one-out cross validation approaches were deployed when calibrating and testing the models (e.g. in Chapter 5), to make sure tested years are separated, but this cross-validation method does not account for potential changes in future years that could change the results. While hindcast analysis provides a streamlined way to evaluate models using a consistent evaluation framework, continuing to validate forecasts in real-time as they are deployed would be useful to test AOGCM skill and the usefulness of post-processing techniques like the process-based subsampling method developed in Chapter 6.

7.6 Towards a Cohesive Model Evaluation Workflow in the R2O Lifecycle

A continuous model evaluation workflow is needed to keep up with AOGCM updates to ensure they support regional applications. Cascading uncertainties from using AOGCMs as input data into other models has been identified as a significant challenge in climate projection studies (e.g. Maslin, 2013). This challenge also affects S2S rainfall forecasts, as rainfall forecasts from AOGCMs feed other models, and their uncertainties can affect regional climate forecasts, streamflow, and impact modelling. Although predicting rainfall anomalies is only one component of effective resilience, further study on how AOGCM

rainfall skill varies spatially and temporally is key to determine the trustworthiness of their predictions, which can support a holistic framework for preparedness strategies.

Continuing to comparatively evaluate dynamic model-based ensemble skill within other forecasting methods is useful to establish the comparative advantage of the different approaches, especially as new hybrid and machine-learning (ML) methods emerge. Comparative evaluation was pursued as a key pathway to assess dynamic model-based ensembles due to the power of framing effects on decision-making (i.e. it is easier to make choices using relative comparisons between approaches rather than deciding if an approach is good or bad using a numerical estimate in isolation, Kahneman & Tversky, 1984). Although dynamic model-based ensembles may reach certain skill thresholds (e.g. MSE-SS > 0.3), other approaches may perform even better. Canonical correlation analysis (CCA), for instance, has been demonstrated as a skilful statistical method to enhance the detection of extreme precipitation events in Central America and nearby regions (Alfaro et al., 2016; Maldonado et al., 2013; Muñoz et al., 2016). It would be worthwhile to continue exploring the relative performance of dynamic, statistical, and hybrid forecasting methods over different locations and months. Many forecasting centres are also exploring ML approaches. ECMWF, for instance, is developing a probabilistic ML-based forecasting system to complement their dynamic physically-based ensembles (https://www.ecmwf.int/en/about/media-centre/science-blog/2023/rise-machine-learning-weather-forecasting). Comparative evaluations will help decision-makers choose between forecasting tools and identify opportunities to leverage the value from each (e.g. using AOGCMs to help explain some of the predictive relationships generated by ML approaches).

Technical AOGCM evaluations can also potentially ease some operational barriers to forecast deployment. Within the R2O life cycle (Buizza et al., 2017), model evaluation plays a key role in clarifying the potential skill of a model. Some operational constraints studies have examined several barriers to forecast uptake, such as institutional mechanisms and cultural norms (e.g. Broad et al., 2007; Crochemore et al., 2016; Feldman & Ingram, 2009; Kirchhoff et al., 2013; Lange & Cook, 2015; Rayner et al., 2005). Lange and Cook (2015), for instance, examined how water managers navigated legal criteria that increased avoidance of using uncertain forecasts. Rayner, Lach, and Ingram (2005) illustrated how water managers' expectations to build their systems to survive any situation decreased their openness to using probabilistic forecasts. When technical forecast evaluations are done well, these evaluations can clarify the trustworthiness of a forecast, which can also further the dialogue on incorporating forecasting into decision-making. New decision-making frameworks are being developed to incorporate forecasts into resilience. Forecast-based-Action and Financing (FbA and FbF), for instance, have been approved and deployed as an early action protocol for tropical storm preparedness by the Red Cross Red Crescent in Central America (Red Cross Red Crescent, 2022a, 2022b, 2022c). Continuing to evaluate forecasts will be critical to ensure these protocols are successfully supporting planning efforts.

CHAPTER 8

Conclusions

Evaluations of dynamic model-based ensemble forecasts of rainfall over Central America at the S2S scale can support many regional applications. This thesis investigates the ability of several leading AOGCMs to forecast regional rainfall at the S2S scale with an emphasis on how these forecasts are operationally deployed today (e.g. in three-month-mean rainfall outlooks). By assessing AOGCMs that contribute to NMME and C3S, the work provides a comprehensive evaluation of the spatial and temporal skill of some of the leading dynamic S2S forecasts of rainfall over Central America, including their abilities to detect low and high rainfall extremes.

This work shows how AOGCMs and their combined MMEs have a comparative edge in the early wet season over simpler statistical approaches that rely on the ENSO teleconnection. While all forecasting methods assessed tend to improve in the late wet season, their skill does not maintain for detections of low and high rainfall extremes. Using a process-informed analysis can improve extreme rainfall detection rates by subsampling ensemble members that best represent key processes for regional rainfall. This technique is designed to support regional forecast operations by enabling forecasters to leverage their regional expertise of important rainfall-generating mechanisms to select ensemble members that will likely be more skilful at predicting rainfall. This work also identifies situations when hybrid forecasting techniques can improve seasonal rainfall skill, such as using TNA-driven forecasts in the early wet season over some locations in Central America.

These analyses provide a step forward in the understanding of how dynamic model-based ensembles can support planning for normal rainfall variability and hydrometeorological extremes in Central America. While the studies were developed for Central America, they also explore ideas that could be useful for other regions (e.g. using hybrid methods and process-based subsampling techniques). As part of the R2O lifecycle, model evaluation plays a key role in making sure forecasting tools adequately support stakeholder needs. Ongoing evaluation is an important aspect of model development, especially when models are being applied in regions outside of the context in which they were developed. Ideally, evaluations will also help strengthen the operations to research (O2R) workflow by identifying opportunities for future model development when models fail to meet operational needs. Continuing to strengthen the feedback loop between research and operations is critical going forward, as the risk of extreme weather is unlikely to abate. Anticipatory action can save lives and livelihoods, and making sure alerts meet stakeholder needs is critical to support communities around the world.

Appendix. Co-author Statements

I, Christian Birkel, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

SEAS5 Skilfully Predicts Late Wet-Season Precipitation in Central American Dry Corridor Excelling in Costa Rica and Nicaragua

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

Print Name: Christian Birkel

Signature:  Date: 14.08.2023

I Alan García López, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

A Comparison of Seasonal Rainfall Forecasts over Central America using Dynamic and Hybrid Approaches from C3S and NMME

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

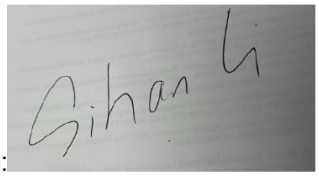
Print Name: Alan García

Signature:  Date: 08/14/2023

I Sihan Li, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

Print Name: Sihan Li

Signature: 

Date: 14/08/2023

I, Timo Kelder, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

Print Name: Timo Kelder

Signature: 

Date: 14/08/2023

I Kyle Hall, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

Print Name: Kyle Joseph Chen Hall

Signature: 

Date: 2023/08/14

I Simon Moulds, certify that Katherine Kowal completed the majority of the work in the following journal articles, which form part of her DPhil thesis:

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

Print Name: Simon Moulds

Signature: 

Date: 16/08/2023

References

Acharya, N., Ehsan, M. A., Admasu, A., Teshome, A., & Hall, K. J. C. (2021). On the next generation (NextGen) seasonal prediction system to enhance climate services over Ethiopia. Climate Services, 24. https://doi.org/10.1016/J.CLISER.2021.100272

Adam, O., Schneider, T., Brient, F., & Bischoff, T. (2016). Relation of the double-ITCZ bias to the atmospheric energy budget in climate models. Geophysical Research Letters , 43, 7670–7677. https://doi.org/10.1002/2016GL069465

Adam, O., Schneider, T., & Brient, F. (2018). Regional and seasonal variations of the double-ITCZ bias in CMIP5 models. Climate Dynamics, 51, 101–117. https://doi.org/10.1007/s00382-017-3909-1

Adger, W. N. (2006). Vulnerability. Global Environmental Change, 16(3), 268–281. https://doi.org/10.1016/j.gloenvcha.2006.02.006

Aguilar, E., Peterson, T. C., Obando, P. R., Frutos, R., Retana, J. A., Solera, M., Soley, J., García, I. G., Araujo, R. M., Santos, A. R., Valle, V. E., Brunet, M., Aguilar, L., Álvarez, L., Bautista, M., Castañón, C., Herrera, L., Ruano, E., Sinay, J. J., … Mayorga, R. (2005). Changes in precipitation and temperature extremes in Central America and northern South America, 1961-2003. Journal of Geophysical Research Atmospheres, 110(23), 1–15. https://doi.org/10.1029/2005JD006119

Alfaro, E. J. (2002). Some Characteristics of the Annual Precipitation Cycle in Central America and their Relationships with its Surrounding Tropical Oceans. Topics of Meteorology and Oceanography, 9(2), 88–103.

Alfaro, E. J. (2007). Uso del análisis de correlación canónica para la predicción de la precipitación pluvial en Centroamérica Use of the canonical correlation analysis for the prediction of rainfall in Central America. Ingenieria y Competitividad, 9(2), 33–48. https://doi.org/10.25100/iyc.v9i2.2486

Alfaro, E. J. (2014). Caracterización del “veranillo” en dos cuencas de la vertiente del Pacífico de Costa Rica, América Central. Revista de Biologia Tropical, 62, 1–15. https://doi.org/10.15517/RBT.V62I4.20010

Alfaro, E. J., Hidalgo, H., & Mora, N. (2016). Prediction of MJ rainfall season using CCA models. https://www.researchgate.net/publication/314865864

Alfaro, E. J., Hidalgo, H. G., Mora, N. P., Perez, P., & Fallas, B. (2016). Assessment of Central America Regional Climate Outlook Forum maps, 1998-2013. Topicos Meteorologicos y Oceanograficos, 15, 37–52. https://www.academia.edu/30424833/Assessment_of_Central_America_Regional_Climate_Outlook_Forum_maps_1998_2013

Alfaro, E. J., Chourio, X., Muñoz, Á. G., & Mason, S. J. (2018). Improved seasonal prediction skill of rainfall for the Primera season in Central America. Journal of Climatology, 38, e255–e268. https://doi.org/10.1002/joc.5366

References

Alfaro, E. J., Hidalgo H. G., & Pérez Briceño, P. (2018). Mapping Environmental and Socioeconomic impacts of hydrometeorological hazards across Central America. Study case: Honduras. Política Económica Para El Desarrollo Sostenible, 3(1), 20. https://doi.org/10.15359/peds.3-1.2

Alfaro-Córdoba, M., Hidalgo, H. G., & Alfaro, E. J. (2020). Aridity trends in central america: A spatial correlation analysis. Atmosphere, 11(4), 1–13. https://doi.org/10.3390/ATMOS11040427

Almazroui, M., Islam, M. N., Saeed, F., Saeed, S., Ismail, M., Ehsan, M. A., Diallo, I., O'Brien, E., Ashfaq, M., Martínez-Castro, D., Cavazos, T., Cerezo-Mota, R., Tippett, M. K., Gutowski, W. J., Alfaro, E. J., Hidalgo, H. G., Vichot-Llano, A., Campbell, J. D., Kamil, S., … Barlow, M. (2021). Projected Changes in Temperature and Precipitation Over the United States, Central America, and the Caribbean in CMIP6 GCMs. Earth Systems and Environment, 5(1), 1–24. https://doi.org/10.1007/s41748-021-00199-5

Amador, J. A. (1998). A climate feature of the tropical Americas: The trade wind easterly jet. Tópicos Meteorológicos y Oceanográficos, 5(2). https://kerwa.ucr.ac.cr/handle/10669/76623

Amador, J. A., Alfaro, E. J., Lizano, O. G., & Magaña, V. O. (2006). Atmospheric forcing of the eastern tropical Pacific: A review. Progress in Oceanography, 69(2–4), 101–142. https://doi.org/10.1016/j.pocean.2006.03.007

Amador, J. A. (2008). The Intra-Americas Sea Low-level Jet. Annals of the New York Academy of Sciences, 1146(1), 153–188. https://doi.org/10.1196/annals.1446.012

Amador, J. A., Durán-Quesada, A. M., Rivera, E. R., Mora, G., Sáenz, F., Calderón, B., & Mora, N. (2016a). The easternmost tropical Pacific. Part II: Seasonal and intraseasonal modes of atmospheric variability. Revista de Biología Tropical, 64(S1), 23–57. https://doi.org/10.15517/RBT.V64I1.23409

Amador, J. A., Rivera, E. R., Durán-Quesada, A. M., Mora, G., Sáenz, F., Calderón, B., & Mora, N. (2016b). The easternmost tropical Pacific. Part I: A climate review. Revista de Biología Tropical, 64(S1), 1–22. https://doi.org/10.15517/RBT.V64I1.23407

Anderson, T. G., Anchukaitis, K. J., Pons, D., & Taylor, M. (2019). Multiscale trends and precipitation extremes in the Central American Midsummer Drought. Environmental Research Letters, 14(12), 124016. https://doi.org/10.1088/1748-9326/ab5023

Anderson, T. G., McKinnon, K. A., Pons, D., and Anchukaitis, K. J. (2023). How exceptional was the 2015-2019 Central American Drought? 10.22541/essoar.168987131.17335510/v2

Arciniega-Esparza, S., Birkel, C., Chavarría-Palma, A., Arheimer, B., & Agustín Breña-Naranjo, J. (2022). Remote sensing-aided rainfall-runoff modeling in the tropics of Costa Rica. Hydrol. Earth Syst. Sci, 26, 975–999. https://doi.org/10.5194/hess-26-975-2022

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018). An Intercomparison of Skill and Overconfidence/Underconfidence of the Wintertime

North Atlantic Oscillation in Multimodel Seasonal Forecasts. Geophysical Research Letters, 45(15), 7808–7817. https://doi.org/10.1029/2018GL078838

Balmaseda, M.A., Fujii, Y., Alves, O., Awaji, T., Behringer, D., Ferry, N., Lee, T., Rienecker, M.M., Rosati, T., Stammer, D., Smith, D.M., & Molteni, F. (2011). INITIALIZATION FOR SEASONAL AND DECADAL FORECASTS. Environmental Science, 11–18. https://doi.org/10.5270/OCEANOBS09.CWP.02

Barlow, M., & Salstein, D. (2006). Summertime influence of the Madden-Julian Oscillation on daily rainfall over Mexico and Central America. Geophysical Research Letters, 33(21), 21708. https://doi.org/10.1029/2006GL027738

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of Real-Time Seasonal ENSO Model Predictions During 2002–11: Is Our Capability Increasing? Bulletin of the American Meteorological Society, 93(5), ES48–ES50. https://doi.org/10.1175/bams-d-11-00111.2

Barnston, A. G., Tippett, M. K., Ranganathan, M., L'heureux, M. L., & Barnston, A. G. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. 53, 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Barrett, B. S., Esquivel Longoria, M. I., Barrett, B. S., & Esquivel Longoria, M. I. (2013). Variability of precipitation and temperature in Guanajuato, Mexico. Atmo, 26(4), 521–536. https://doi.org/10.1016/S0187-6236(13)71093-2

Barry, R. G., & Chorley, R. J. (2010). Chapter 11: Tropical Weather and Climate. In Atmosphere, Weather, and Climate: Vol. 9th Edition (p. 327). https://books.google.co.uk/books?id=V959pgO08jwC&printsec=frontcover#v=onepage&q&f=false

Batté, L., Dorel, L., Ardilouze, C., & Guérémy, J.-F. (2021). Documentation of the METEO-FRANCE seasonal forecasting system 8. http://www.umr-cnrm.fr/IMG/pdf/system8-technical.pdf

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567), 47–55. https://doi.org/10.1038/nature14956

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences, 21(1), 589–615. https://doi.org/10.5194/hess-21-589-2017

Becker, E., Zhang, Q., Dool, H. Van Den, Saha, S., Peng, P., Peña, M., Tripp, P., & Huang, J. (2012). Evaluation of the National Multi-Model Ensemble System for Seasonal and Monthly Prediction. 36th NOAA Annual Climate Di7agnostics and Prediction Workshop, October, 43–45. https://www.nws.noaa.gov/ost/climate/STIP/37CDPW/37cdpw-ebecker1.pdf

Becker, E., Kirtman, B. P., & Pegion, K. (2020). Evolution of the North American Multi-Model Ensemble. Geophysical Research Letters, 47(9). https://doi.org/10.1029/2020GL087408

References

Beveridge, L., Whitfield, S., Fraval, S., van Wijk, M., van Etten, J., Mercado, L., Hammond, J., Davila Cortez, L., Gabriel Suchini, J., & Challinor, A. (2019). Experiences and Drivers of Food Insecurity in Guatemala's Dry Corridor: Insights From the Integration of Ethnographic and Household Survey Data. Frontiers in Sustainable Food Systems, 3, 65. https://doi.org/10.3389/fsufs.2019.00065

Bjerknes, V. (1904). Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (The problem of weather prediction, considered from the viewpoints of mechanics and physics) – Meteorol. Z. 21, 1–7. (translated and edited by VOLKEN E. and S. BRONNIMANN ¨ . – Meteorol. Z. 18 (2009), 663–667). Meteorologische Zeitschrift, 18(6), 663–667. https://doi.org/10.1127/0941-2948/2009/416

Bodner, G., Nakhforoosh, A., & Kaul, H. P. (2015). Management of crop water under drought: a review. Agronomy for Sustainable Development, 35(2), 401–442. https://doi.org/10.1007/S13593-015-0283-4

Boer, G. J. (2010). Climate trends in a seasonal forecasting system. Atmosphere-Ocean, 47(2), 123–138. https://doi.org/10.3137/AO1002.2009

Braman, L. M., van Aalst, M. K., Mason, S. J., Suarez, P., Ait-Chellouche, Y., & Tall, A. (2013). Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008. Disasters, 37(1), 144–164. https://doi.org/10.1111/J.1467-7717.2012.01297.X

Brier, G. W., & Allen, R. A. (1951). Verification of Weather Forecasts. Compendium of Meteorology, 841–848. https://doi.org/10.1007/978-1-940033-70-9_68

Broad, K., Pfaff, A., Taddei, R., Sankarasubramanian, A., Lall, U., & de Assis de Souza Filho, F. (2007). Climate, stream flow prediction and water management in northeast Brazil: Societal trends and forecast value. Climatic Change, 84(2), 217–239. https://doi.org/10.1007/s10584-007-9257-0

Bröcker, J. (2012). Evaluating raw ensembles with the continuous ranked probability score. Quarterly Journal of the Royal Meteorological Society, 138(667), 1611–1617. https://doi.org/10.1002/qj.1891

Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., Coppola, E., Vries, H. De, Harris, G., Hegerl, G. C., Knutti, R., Lenderink, G., Lowe, J., Nogherotto, R., O'Reilly, C., Qasmi, S., Ribes, A., Stocchi, P., & Undorf, S. (2020). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework. Journal of Climate, 33(20), 8671–8692. https://doi.org/10.1175/JCLI-D-19-0953.1

Buizza, R., & Palmer, T. (1998). Impact of Ensemble Size on Ensemble Prediction. Monthly Weather Review, 126(9). https://journals.ametsoc.org/view/journals/mwre/126/9/1520-0493_1998_126_2503_ioesoe_2.0.co_2.xml

Buizza, R., Petroliagis, T., Palmer, T., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., & Wedi, N. (1998). Impact of model resolution and ensemble size on the performance of an Ensemble Prediction System. Quarterly Journal of the Royal

Meteorological Society, 124(550), 1935–1960. https://doi.org/10.1002/QJ.49712455008

Buizza, R. (2008). Comparison of a 51-Member Low-Resolution (TL399L62) Ensemble with a 6-Member High-Resolution (TL799L91) Lagged-Forecast Ensemble. Monthly Weather Review, 136(9), 3343–3362. https://doi.org/10.1175/2008MWR2430.1

Buizza, R., Andersson, E., Forbes, R., & Sleigh, M. (2017). The ECMWF research to operations (R2O) process. http://www.ecmwf.int/en/research/publications

Bundschuh, J., Litter, M., Ciminelli, V. S. T., Morgada, M. E., Cornejo, L., Hoyos, S. G., Hoinkis, J., Alarcón-Herrera, M. T., Armienta, M. A., & Bhattacharya, P. (2010). Emerging mitigation needs and sustainable options for solving the arsenic problems of rural and isolated urban areas in Latin America - A critical analysis. Water Research, 44(19), 5828–5845. https://doi.org/10.1016/j.watres.2010.04.001

Calvo-Solano, O. D., Quesada-Hernández, L. E., Hidalgo, H., & Gotlieb, Y. (2018). Impactos de las sequías en el sector agropecuario del Corredor Seco Centroamericano. Agronomía Mesoamericana, 29(3), 695. https://doi.org/10.15517/ma.v29i3.30828

Carrão, H., Naumann, G., & Barbosa, P. (2016). Drought forecasting for Latin America Desertification, Land Degradation and Drought (DLDD), and bio-physical modelling for crop yield estimation in Latin America under a changing climate. https://ec.europa.eu/jrc/en/publication/drought-forecasting-latin-america

Carrão, H., Naumann, G., Dutra, E., Lavaysse, C., & Barbosa, P. (2018). Seasonal Drought Forecasting for Latin America Using the ECMWF S4 Forecast System. Climate, 6(2), 48. https://doi.org/10.3390/cli6020048

Carvalho, L. M. V. (2020). Assessing precipitation trends in the Americas with historical data: A review. Wiley Interdisciplinary Reviews: Climate Change, 11(2), e627. https://doi.org/10.1002/WCC.627

CEPAL. (2002). El impacto socioeconómico y ambiental de la sequía de 2001 en Centroamérica. https://repositorio.cepal.org/handle/11362/25574

Charney, J. G., Fjortoft, R., & Neumann, J. (1950). Numerical Integration of the Barotropic Vorticity Equation. Tellus, 2(4), 237–254. https://doi.org/10.1111/j.2153-3490.1950.tb00336.x

Chen, H. C., Tseng, Y. H., Hu, Z. Z., & Ding, R. (2020). Enhancing the ENSO Predictability beyond the Spring Barrier. Scientific Reports, 10(1), 1–12. https://doi.org/10.1038/s41598-020-57853-7

Chow, F., Schär, C., Ban, N., Lundquist, K., Schlemmer, L., & Shi, X. (2019). Crossing Multiple Gray Zones in the Transition from Mesoscale to Microscale Simulation over Complex Terrain. Atmosphere, 10(5), 274. https://doi.org/10.3390/atmos10050274

Clarke, A. J. (2014). El Niño physics and El Niño predictability. Annual Review of Marine Science, 6, 79–99. https://doi.org/10.1146/annurev-marine-010213-135026

References

Coelho, C. A. S. (2013). Comparative skill assessment of consensus and physically based tercile probability seasonal precipitation forecasts for Brazil. Meteorological Applications, 20(2), 236–245. https://doi.org/10.1002/MET.1407

Coelho, C. A. S., Dayana, De Souza, C., Kubota, P. Y., Costa, S. M. S., Menezes, L., Guimarães, B. S., Figueroa, S. N., Bonatti, J. P., Iracema, ·, Cavalcanti, F. A., Gilvan Sampaio, ·, Klingaman, N. P., Baker, J. C. A., & Coelho, C. A. S. (2020). Evaluation of climate simulations produced with the Brazilian global atmospheric model version 1.2. Climate Dynamics. https://doi.org/10.1007/s00382-020-05508-8

Colman, A. W., Graham, R. J., & Davey, M. K. (2020). Direct and indirect seasonal rainfall forecasts for East Africa using global dynamical models. International Journal of Climatology, 40(2), 1132–1148. https://doi.org/10.1002/JOC.6260

Conover, W. (1999). Nonparametric Methods. In Practical Nonparametric Statistics (3rd ed., p. Chapter 6). https://www.wiley.com/en-us/Practical+Nonparametric+Statistics%2C+3rd+Edition-p-9780471160687

Conover, W., & Iman, R. (1979). Multiple-comparisons procedures. Informal report. https://doi.org/10.2172/6057803

Cook, B. I., Mankin, J. S., Marvel, K., Williams, A. P., Smerdon, J. E., & Anchukaitis, K. J. (2020). Twenty-First Century Drought Projections in the CMIP6 Forcing Scenarios. Earth's Future, 8(6). https://doi.org/10.1029/2019EF001461

Cook, C. (2016). Drought planning as a proxy for water security in England. Current Opinion in Environmental Sustainability, 21, 65–69. https://doi.org/10.1016/j.cosust.2016.11.005

Cook, K. H., & Vizy, E. K. (2010). Hydrodynamics of the Caribbean low-level jet and its relationship to precipitation. Journal of Climate, 23(6), 1477–1494. https://doi.org/10.1175/2009JCLI3210.1

Corrales-Suastegui, A., Fuentes-Franco, R., & Pavia, E. G. (2020). The mid-summer drought over Mexico and Central America in the 21st century. International Journal of Climatology, 40(3), 1703–1715. https://doi.org/10.1002/JOC.6296

Coughlan de Perez, E., Van Den Hurk, B., Van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. Natural Hazards and Earth System Sciences, 15(4), 895–904. https://doi.org/10.5194/NHESS-15-895-2015

Coughlan de Perez, E. C., Van Den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., De Suarez, J. M., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., & Zsoter, E. (2016). Action-based flood forecasting for triggering humanitarian action. Hydrology and Earth System Sciences, 20(9), 3549–3560. https://doi.org/10.5194/HESS-20-3549-2016

Coughlan de Perez, E. C., Stephens, E., Bischiniotis, K., Van Aalst, M., Van Den Hurk, B., Mason, S., Nissan, H., & Pappenberger, F. (2017). Should seasonal rainfall forecasts be used for flood preparedness? Hydrology and Earth System Sciences, 21(9), 4517–4524. https://doi.org/10.5194/HESS-21-4517-2017

References

Cox, P., & Stephenson, D. (2007). A changing climate for prediction. Science, 317(5835), 207–208. https://doi.org/10.1126/SCIENCE.1145956/ASSET/F7CFD661-68C2-44CA-B277-FBFAE5962472/ASSETS/GRAPHIC/207-2.GIF

Crochemore, L., Ramos, M.-H., Pappenberger, F., Andel, S. J. van, & Wood, A. W. (2016). An Experiment on Risk-Based Decision-Making in Water Management Using Monthly Probabilistic Forecasts. Bulletin of the American Meteorological Society, 97(4), 541–551. https://doi.org/10.1175/BAMS-D-14-00270.1

Curtis, S., & Gamble, D. W. (2008). Regional variations of the Caribbean mid-summer drought. Theoretical and Applied Climatology, 94(1–2), 25–34. https://doi.org/10.1007/S00704-007-0342-0/METRICS

Davis, P., Ruth, C., Scaife, A. A., Kettleborough, J., Davis, P., Ruth, C., Scaife, A. A., & Kettleborough, J. (2020). A Large Ensemble Seasonal Forecasting System: GloSea6. AGUFM, 2020, A192-05. https://ui.adsabs.harvard.edu/abs/2020AGUFMA192...05D/abstract

de Sousa, K., Casanoves, F., Sellare, J., Ospina, A., Suchini, J. G., Aguilar, A., & Mercado, L. (2018). How climate awareness influences farmers' adaptation decisions in Central America? Journal of Rural Studies, 64, 11–19. https://doi.org/10.1016/j.jrurstud.2018.09.018

Delsole, T., Yang, X., & Tippett, M. K. (2013). Is unequal weighting significantly better than equal weighting for multi-model forecasting? Quarterly Journal of the Royal Meteorological Society, 139(670), 176–183. https://doi.org/10.1002/QJ.1961

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., Ginoux, P., Gudgel, R., Hallberg, R. W., Harris, L., Harrison, M. J., Johnson, N., Kapnick, S. B., Lin, S.-J., Lu, F., Malyshev, S., Milly, P. C., Murakami, H., Naik, V., … Zhao, M. (2020). SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. Journal of Advances in Modeling Earth Systems, 12. https://doi.org/10.1029/2019MS001895

Deque, M. (2012). Deterministic Forecasts of Continuous Variables. In I. Joliffe & D. Stephenson (Eds.), Forecast verification : a practitioner's guide in atmospheric science (2nd ed., pp. 77–94).

di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R. v., van den Hurk, B., Vellore, R., Krishnan, R., & Coumou, D. (2019). Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors. Weather and Forecasting, 34(5), 1377–1394. https://doi.org/10.1175/WAF-D-19-0002.1

DiSera, L., Sjödin, H., Rocklöv, J., Tozan, Y., Súdre, B., Zeller, H., & Muñoz, Á. G. (2020). The Mosquito, the Virus, the Climate: An Unforeseen Réunion in 2018. GeoHealth, 4(8), e2020GH000253. https://doi.org/10.1029/2020GH000253

Doblas-Reyes, F. J., Hagedorn, R., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting-II. Calibration and combination. 57, 234–252. https://doi.org/10.3402/tellusa.v57i3.14658

References

Doblas-Reyes, F. J., Hagedorn, R., Palmer, T. N., & Morcrette, J. J. (2006). Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. Geophysical Research Letters, 33(7). https://doi.org/10.1029/2005GL025061

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

Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., Düsterhus, A., Fröhlich, K., Pohlmann, H., & Baehr, J. (2018). Improved Teleconnection-Based Dynamical Seasonal Predictions of Boreal Winter. Geophysical Research Letters, 45(8), 3605–3614. https://doi.org/10.1002/2018GL077209

Dobrynin, M., Düsterhus, A., Fröhlich, K., Athanasiadis, P., Ruggieri, P., Müller, W. A., & Baehr, J. (2022). Hidden Potential in Predicting Wintertime Temperature Anomalies in the Northern Hemisphere. Geophysical Research Letters, 49(20). https://doi.org/10.1029/2021GL095063

Domeisen, D. I. V., White, C. J., Afargan-Gerstman, H., Muñoz, Á. G., Janiga, M. A., Vitart, F., Wulff, C. O., Antoine, S., Ardilouze, C., Batté, L., Bloomfield, H. C., Brayshaw, D. J., Camargo, S. J., Charlton-Pérez, A., Collins, D., Cowan, T., Del Mar Chaves, M., Ferranti, L., Gómez, R., … Tian, D. (2022). Advances in the Subseasonal Prediction of Extreme Events: Relevant Case Studies across the Globe. Bulletin of the American Meteorological Society, 103(6), E1473–E1501. https://doi.org/10.1175/BAMS-D-20-0221.1

Donoso, M., & Ramirez, P. (2001). Latin America and the Caribbean: Report on the Climate Outlook Forums for Mesoamerica. In Coping with the climate: A step Forward, Workshop Report: A multi-stakeholder review of Regional Climate Outlook Forums (pp. 11–18). IRICW.

Duan, W., & Wei, C. (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. https://doi.org/10.1002/joc.3513

Durán-Quesada, A. M., Gimeno, L., Amador, J. A., & Nieto, R. (2010). Moisture sources for Central America: Identification of moisture sources using a Lagrangian analysis technique. Journal of Geophysical Research Atmospheres, 115(5). https://doi.org/10.1029/2009JD012455

Durán-Quesada, A. M., Gimeno, L., & Amador, J. (2017). Role of moisture transport for Central American precipitation. Earth System Dynamics, 8(1), 147–161. https://doi.org/10.5194/esd-8-147-2017

Durán-Quesada, A. M., Sorí, R., Ordoñez, P., & Gimeno, L. (2020). Climate perspectives in the Intra-Americas seas. Atmosphere, 11(9), 1–32. https://doi.org/10.3390/ATMOS11090959

Dusterhus, A. (2020). Seasonal statistical-dynamical prediction of the North Atlantic Oscillation by probabilistic post-processing and its evaluation. Nonlinear Processes in Geophysics, 27(1), 121–131. https://doi.org/10.5194/NPG-27-121-2020

References

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., Barbosa, P., Vogt, J., & Pappenberger, F. (2014). Global meteorological drought-Part 2: Seasonal forecasts. Hydrol. Earth Syst. Sci, 18, 2669–2678. https://doi.org/10.5194/hess-18-2669-2014

Elvidge, S., Granados, S. R., Angling, M. J., Brown, M. K., Themens, D. R., & Wood, A. G. (2023). Multi-Model Ensembles for Upper Atmosphere Models. Space Weather, 21(3), e2022SW003356. https://doi.org/10.1029/2022SW003356

Enfield, D. B., & Mayer, D. A. (1997). Tropical atlantic sea surface temperature variability and its relation to El Niño-Southern Oscillation. Journal of Geophysical Research C: Oceans, 102(1), 929–945. https://doi.org/10.1029/96jc03296

Enfield, D. B., & Alfaro, E. J. (1999). The dependence of Caribbean rainfall on the interaction of the tropical Atlantic and Pacific Oceans. Journal of Climate, 12(7), 2093–2103. https://doi.org/10.1175/1520-0442(1999)012<2093:TDOCRO>2.0.CO;2

Enfield, D. B., Mestas-Nuñez, A. M., Mayer, D. A., & Cid-Serrano, L. (1999). How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? Journal of Geophysical Research: Oceans, 104(C4), 7841–7848. https://doi.org/10.1029/1998JC900109

Enfield, D. B., Mestas-Nunez, A. M., & Trimble, P. J. (2001). The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophysical Research Letters, 28(10), 2077–2080. https://www.researchgate.net/publication/236608411_The_Atlantic_multidecadal_oscillation_and_its_relation_to_rainfall_and_river_flows_in_the_continental_US

Eyring, V. (2016). Towards improved and more routine Earth system model evaluation in CMIP. Earth Syst. Dynam., 7, 813–830. https://esd.copernicus.org/articles/7/813/2016/

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Williamson, M. S. (2019). Taking climate model evaluation to the next level. Nature Climate Change 2019 9:2, 9(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Fallas López, B., & Alfaro Martínez, E. J. (2012). Uso de herramientas estadísticas para la predicción estacional del campo de precipitación en América Central como apoyo a los Foros Climáticos Regionales. 1: Análisis de tablas de contingencia. Revista de Climatología, ISSN-e 1578-8768, No. 12, 2012, Págs. 61-79, 12, 61–79. https://dialnet.unirioja.es/servlet/articulo?codigo=7416858&info=resumen&idioma=ENG

FAO. (2015a). global information and early warning system on food and agriculture GIEWS GIEWS Update Central America -Crop Calendar. September. https://www.fao.org/3/I4926E/I4926E.pdf

FAO (2015b). Disaster Risk Programme to strengthen resilience in the Dry Corridor in Central America : FAO in Emergencies.

https://www.fao.org/fileadmin/user_upload/emergencies/docs/Corredor_Seco_Breve_EN.pdf

FAO (2016). Dry Corridor Central America - Situation Report. https://www.fao.org/3/br092s/br092s.pdf

Farfán, L. M., Alfaro, E. J., & Cavazos, T. (2013). Characteristics of tropical cyclones making landfall on the Pacific coast of Mexico: 1970-2010. Atmósfera, 26(2), 163–182. https://doi.org/10.1016/S0187-6236(13)71070-1

Feldman, D. L., & Ingram, H. M. (2009). Making science useful to decision makers: Climate forecasts, water management, and knowledge networks. Weather, Climate, and Society, 1(1), 9–21. https://doi.org/10.1175/2009WCAS1007.1

Fernandes, K., Muñoz, A. G., Ramirez-Villegas, J., Agudelo, D., Llanos-Herrera, L., Esquivel, A., Rodriguez-Espinoza, J., & Prager, S. D. (2020). Improving seasonal precipitation forecasts for agriculture in the orinoquía Region of Colombia. Weather and Forecasting, 35(2), 437–449. https://doi.org/10.1175/WAF-D-19-0122.1

Fernandes, K., Bell, M., & Muñoz, Á. G. (2022). Combining precipitation forecasts and vegetation health to predict fire risk at subseasonal timescale in the Amazon. Environ. Res. Letters, 17. https://doi.org/10.1088/1748-9326/ac76d8

Ferro, C. A. T., Richardson, D. S., & Weigel, A. P. (2008). On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorological Applications, 15(1), 19–24. https://doi.org/10.1002/MET.45

Flohr, B. M., Hunt, J. R., Kirkegaard, J. A., & Evans, J. R. (2017). Water and temperature stress define the optimal flowering period for wheat in south-eastern Australia. Field Crops Research, 209, 108–119. https://doi.org/10.1016/J.FCR.2017.04.012

Flohr, B. M., Hunt, J. R., Kirkegaard, J. A., Evans, J. R., & Lilley, J. M. (2018). Genotype × management strategies to stabilise the flowering time of wheat in the south-eastern Australian wheatbelt. Crop and Pasture Science, 69(6), 547–560. https://doi.org/10.1071/CP18014

Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., Haak, H., Brune, S., Früh, B., & Baehr, J. (2021). The German Climate Forecast System: GCFS. Journal of Advances in Modeling Earth Systems, 13(2). https://doi.org/10.1029/2020MS002101

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., & Husak, G. J. (2014). A quasi-global precipitation time series for drought monitoring. https://pubs.er.usgs.gov/publication/ds832

Gamble, D. W., Parnell, D. B., & Curtis, S. (2008). Spatial variability of the Caribbean mid-summer drought and relation to north Atlantic high circulation. International Journal of Climatology, 28(3), 343–350. https://doi.org/10.1002/JOC.1600

García-Martínez, I. M., & Bollasina, M. A. (2020). Sub-monthly evolution of the Caribbean Low-Level Jet and its relationship with regional precipitation and atmospheric circulation. Climate Dynamics, 54(9–10), 4423–4440. https://doi.org/10.1007/S00382-020-05237-Y/FIGURES/10

References

García-Oliva, L. C., & Pazos, E. (2021). The mid-summer drought spatial variability over Mesoamerica. Atmosfera, 34(2), 227–232. https://doi.org/10.20937/ATM.52790

Garcia-Solera, I., & Ramirez, P. (2012). Central America's Seasonal Climate Outlook Forum. https://www.climate-services.org/wp-content/uploads/2015/09/CRRH_Case_Study.pdf

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z. L., & Zhang, M. (2011). The Community Climate System Model Version 4. Journal of Climate, 24(19), 4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Giannini, Kushnir, & Cane. (2000). Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. Journal of Climate, 13. https://www.researchgate.net/publication/242102582_Giannini_A_Kushnir_Y_Cane_MA_Interannual_variability_of_Caribbean_rainfall_ENSO_and_the_Atlantic_Ocean_Journal_of_Climate

Gilbert, G. E. (2019). Nonparametric Tests Used in Simulation Research. Healthcare Simulation Research, 207–214. https://doi.org/10.1007/978-3-030-26837-4_28

Giorgi, F. (2006). Climate change hot‐spots. Geophys. Res. Lett, 33, 8707. https://doi.org/10.1029/2006GL025734

Giuntoli, I., Prosdocimi, I., & Hannah, D. (2021). Going Beyond the Ensemble Mean: Assessment of Future Floods From Global Multi-Models. Water Resources Research, 57. https://doi.org/10.1029/2020WR027897

Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133(5), 1098–1118. https://doi.org/10.1175/MWR2904.1

Goddard, L., González Romero, C., Muñoz, Á. G., Acharya, N., Ahmed Shamsuddin, Baethgen, W., Blumenthal, B., Braun, M., Campos, D., Chourio, X., Cousin, R., Cortés, C., Curtis, A., del Corral, J., Dinh, D., Dinku, T., Fiondella, F., Furlow, J., Alan García-López4, … Vu-Van, T. (2020). Climate Services Ecosystems in times of COVID-19. WMO Bulletin, 69(2). https://public.wmo.int/en/resources/bulletin/climate-services-ecosystems-times-of-covid-19

Golding, B., Mylne, K., & Clark, P. (2004). The history and future of numerical weather prediction in the Met Office. Weather, 59(11), 299–306. https://doi.org/10.1256/WEA.113.04

Golding, N., Hewitt, C., Zhang, P., Liu, M., Zhang, J., & Bett, P. (2019). Co-development of a seasonal rainfall forecast service: Supporting flood risk management for the Yangtze River basin. Climate Risk Management, 23, 43–49. https://doi.org/10.1016/J.CRM.2019.01.002

Gotlieb, Y., Perez-Briceno, P., Hidalgo, H. G., & Alfaro, E. J. (2019). The Central American Dry Corridor: a consensus statement and its background. Revista Yu'am, 3(5), 42–51. https://www.revistayuam.com/volumen-2/numero-3/notas-de-divulgacion-

cientifica/the-central-american-dry-corridor-a-consensus-statement-and-its-background/

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1–25.

Gualdi, S., Borrelli, A., Davoli, G., Masina, S., Navarra, A., Sanna, A., Tibaldi, S., & Cantelli, A. (2020). The new CMCC Operational Seasonal Prediction System Issue TN0288 CMCC Technical Notes. https://doi.org/10.25424/CMCC/SPS3.5

Gubler, S., Sedlmeier, K., Bhend, J., Avalos, G., Coelho, C. A. S., Escajadillo, Y., Jacques-Coper, M., Martinez, R., Schwierz, C., de Skansi, M., & Spirig, C. (2020). Assessment of ECMWF SEAS5 seasonal forecast performance over South America. Weather and Forecasting, 35(2), 561–584. https://doi.org/10.1175/WAF-D-19-0106.1

Guevara-Murua, A., Williams, C. A., Hendy, E. J., & Imbach, P. (2018). 300 years of hydrological records and societal responses to droughts and floods on the Pacific coast of Central America. Climate of the Past, 14(2), 175–191. https://doi.org/10.5194/cp-14-175-2018

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting — I. Basic concept. Tellus A: Dynamic Meteorology and Oceanography, 57(3), 219–233. https://doi.org/10.3402/tellusa.v57i3.14657

Hall, K. J. C., & Acharya, N. (2022). XCast: A python climate forecasting toolkit. Frontiers in Climate, 4, 953262. https://doi.org/10.3389/FCLIM.2022.953262/BIBTEX

Hannah, L., Donatti, C. I., Harvey, C. A., Alfaro, E., Andres Rodriguez, D., Bouroncle, C., Castellanos, E., Diaz, F., Fung, E., Hidalgo, H. G., Imbach, P., Läderach, P., Landrum, J. P., & Lucía Solano, A. (2017). Regional modeling of climate change impacts on smallholder agriculture and ecosystems in Central America. Climatic Change, 141, 29–45. https://doi.org/10.1007/s10584-016-1867-y

Hansen, J., Challinor, A., Ines, A., Wheeler, T., & Moron, V. (2006). Translating climate forecasts into agricultural terms: advances and challenges. RESEARCH Clim Res, 33, 27–41. https://doi.org/10.3354/cr033027ï

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects. Reviews of Geophysics, 56(1), 108–141. https://doi.org/10.1002/2016RG000549

Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im SturmwarnungsdienstBerechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst. Geografiska Annaler, 8, 301. https://doi.org/10.2307/519729

Hemri, S., Bhend, J., Liniger, M. A., Manzanas, R., Siegert, S., David, ·, Stephenson, B., Gutiérrez, J. M., Brookshaw, A., Francisco, ·, & Doblas-Reyes, J. (2020). How to create an operational multi-model of seasonal forecasts? Climate Dynamics, 55, 1141–1157. https://doi.org/10.1007/s00382-020-05314-2

References

Herger, N., Abramowitz, G., Knutti, R., Angélil, O., Lehmann, K., & Sanderson, B. M. (2018). Selecting a climate model subset to optimise key ensemble properties. Earth System Dynamics, 9(1), 135–151. https://doi.org/10.5194/ESD-9-135-2018

Hersbach, H., Bell, B., Berrisford, B., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., …, and Thépaut, J. N. (2020). The ERA5 global reanalysis. Q J R Meteorol Soc. 146: 1999–2049. https://doi.org/10.1002/qj.3803

Hidalgo, H. G., & Alfaro, E. J. (2012). Global Model selection for evaluation of Climate Change projections in the Eastern Tropical Pacific Seascape. Revista de Biología Tropical, 60(S3), 67–81–67–81. https://doi.org/10.15517/RBT.V60I3.28340

Hidalgo, H. G., Amador, J. A., Alfaro, E. J., & Quesada, B. (2013). Hydrological climate change projections for Central America. Journal of Hydrology, 495, 94–112. https://doi.org/10.1016/j.jhydrol.2013.05.004

Hidalgo, H. G., & Alfaro, E. J. (2015). Skill of CMIP5 climate models in reproducing 20th century basic climate features in Central America. International Journal of Climatology, 35(12), 3397–3421. https://doi.org/10.1002/joc.4216

Hidalgo, H. G., Durán-Quesada, A. M., Amador, J. A., & Alfaro, E. J. (2015). The caribbean low-level jet, the inter-tropical convergence zone and precipitation patterns in the intra-americas sea: a proposed dynamical mechanism. Geografiska Annaler: Series A, Physical Geography, 97(1), 41–59. https://doi.org/10.1111/geoa.12085

Hidalgo, H. G., Alfaro, E. J., & Quesada-Montano, B. (2017). Observed (1970–1999) climate variability in Central America using a high-resolution meteorological dataset with implication to climate change studies. Climatic Change, 141(1), 13–28. https://doi.org/10.1007/s10584-016-1786-y

Hidalgo, H. G., Alfaro, E. J., Amador, J. A., & Bastidas, Á. (2019). Precursors of quasi-decadal dry-spells in the Central America Dry Corridor. Climate Dynamics, 53(3–4), 1307–1322. https://doi.org/10.1007/s00382-019-04638-y

Higgins, R. W., Kim, H. K., & Unger, D. (2004). Long-lead seasonal temperature and precipitation prediction using tropical Pacific SST consolidation forecasts. Journal of Climate, 17(17), 3398–3414. https://doi.org/10.1175/1520-0442(2004)017<3398:LSTAPP>2.0.CO;2

Hijmans, R. J. (2021). raster: Geographic Data Analysis and Modeling (R package version 3.4-13).

Hogan, R. J., Ferro, C. A. T., Jolliffe, I. T., & Stephenson, D. B. (2010). Equitability Revisited: Why the "Equitable Threat Score" Is Not Equitable. Weather and Forecasting, 25(2), 710–726. https://doi.org/10.1175/2009WAF2222350.1

Hollister, J. (2021, January 21). Accessing elevation data in R with the elevatr package. Cran.R-Project. https://cran.r-project.org/web/packages/elevatr/vignettes/introduction_to_elevatr.html

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., & Zhang, H. M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface

Temperature (DOISST) Version 2.1. Journal of Climate, 34(8), 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1

Huang, T.-Y., & Zhao, B. (2020). tidyfst: Tidy Verbs for Fast Data Manipulation. Journal of Open Source Software, 5(52), 2388. https://doi.org/10.21105/joss.02388

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003). An overview of the north atlantic oscillation. Geophysical Monograph Series, 134, 1–35. https://doi.org/10.1029/134GM01

Imbach, P., Chou, S. C., Lyra, A., Rodrigues, D., Rodriguez, D., Latinovic, D., Siqueira, G., Silva, A., Garofolo, L., & Georgiou, S. (2018). Future climate change scenarios in Central America at high spatial resolution. PLoS ONE, 13(4). https://doi.org/10.1371/journal.pone.0193570

Imhoff, R. O., De Cruz, L., Dewettinck, W., Brauer, C. C., Uijlenhoet, R., van Heeringen, K. J., Velasco-Forero, C., Nerini, D., Van Ginderachter, M., & Weerts, A. H. (2023). Scale-dependent blending of ensemble rainfall nowcasts and numerical weather prediction in the open-source pysteps library. Quarterly Journal of the Royal Meteorological Society, 149(753), 1335–1364. https://doi.org/10.1002/QJ.4461

IMN. (2015). Boletín Meteorológico Mensual: Mayo 2015. Boletín Meteorológico Mensual, 47. https://doi.org/ISSN 1654’0465

IMN. (2019). BOLETIN DEL ENOS No.118: Fase actual: El Niño. https://www.imn.ac.cr/en/boletin-enos

INSIVUMEH. (2022a). Perspectiva Climática Mensual Abril 2022. https://insivumeh.gob.gt/wp-content/uploads/2022/04/perspectivaClimatica_abr2022_insivumeh.pdf

INSIVUMEH. (2022b). Fenología. http://svsa.insivumeh.gob.gt/fenologia/

INSIVUMEH. (2022c). Mesas Agroclimaticas. https://insivumeh.gob.gt/mesas-agroclimaticas/

IICACR. (2014). Corredor Seco Americano boundary. ArcGIS. https://www.arcgis.com/home/webmap/viewer.html?webmap=fde0c585ebcc4a3098c08c8062dfba06

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., & Monge-Sanz, B. (2018). SEAS5: The new ECMWF seasonal forecast system. Geoscientific Model Development Discussions, 1–44. https://doi.org/10.5194/gmd-2018-228

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., & Monge-Sanz, B. M. (2019). SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), 1087–1117. https://doi.org/https://doi.org/10.5194/gmd-12-1087-2019

References

Joliffe, I., & Stephenson, D. (eds.). (2012). Forecast Verification: A Practitioner's Guide. In I. T. Jolliffe & D. B. Stephenson (Eds.), Atmospheric Science, Second Edition (2nd ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119960003

Jones, P. D., Harpham, C., Harris, I., Goodess, C. M., Burton, A., Centella-Artola, A., Taylor, M. A., Bezanilla-Morlot, A., Campbell, J. D., Stephenson, T. S., Joslyn, O., Nicholls, K., & Baur, T. (2016). Long-term trends in precipitation and temperature across the Caribbean. International Journal of Climatology, 36(9), 3314–3333. https://doi.org/10.1002/joc.4557

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames (Vol. 39, Issue 4).

Kalnay, E. (2003). Chapter 2: The continuous equations. In Atmospheric Modeling, Data Assimilation and Predictability (Vol. 129). Cambridge University Press. https://doi.org/10.1017/CBO9780511802270

Kamran, A., Iqbal, M., & Spaner, D. (2014). Flowering time in wheat (Triticum aestivum L.): A key factor for global adaptability. Euphytica, 197(1), 1–26. https://doi.org/10.1007/S10681-014-1075-7

Karnauskas, Seager, Giannini, & Busalacchi. (2013). A simple mechanism for the climatological midsummer drought along the Pacific coast of Central America. Atmósfera, 26(2), 261–281. https://doi.org/10.1016/S0187-6236(13)71075-0

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots (R package version 0.4.0.).

Kelemen, F. D., Primo, C., Feldmann, H., & Ahrens, B. (2019). Added value of atmosphere-ocean coupling in a century-long regional climate simulation. Atmosphere, 10(9). https://doi.org/10.3390/atmos10090537

Khajehei, S., Ahmadalipour, A., & Moradkhani, H. (2018). An effective post-processing of the North American multi-model ensemble (NMME) precipitation forecasts over the continental US. Climate Dynamics, 51(1–2), 457–472. https://doi.org/10.1007/S00382-017-3934-0

Kim, S. K., & An, S. Il. (2021). Seasonal Gap Theory for ENSO Phase Locking. Journal of Climate, 34(14), 5621–5634. https://doi.org/10.1175/JCLI-D-20-0495.1

Kirchhoff, C. J., Lemos, M. C., & Engle, N. L. (2013). What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S. Environmental Science and Policy, 26, 6–18. https://doi.org/10.1016/j.envsci.2012.07.001

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., Dewitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., … Wood, E. F. (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. https://doi.org/10.1175/BAMS-D-12-00050.1

References

Knaff, J. A. (1997). Implications of Summertime Sea Level Pressure Anomalies in the Tropical Atlantic Region. Journal of Climate, 10(4), 789–804. https://doi.org/10.1175/1520-0442(1997)010

Knutti, R. (2010). The end of model democracy? Climatic Change, 102(3), 395–404. https://doi.org/10.1007/S10584-010-9800-2/METRICS

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., Eyring, V., Knutti, C. :, Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. Geophysical Research Letters, 44(4), 1909–1918. https://doi.org/10.1002/2016GL072012

Kowal, K. M., Slater, L. J., García López, A., & Van Loon, A. F. (2023). A comparison of seasonal rainfall forecasts over Central America using dynamic and hybrid approaches from Copernicus Climate Change Service seasonal forecasting system and the North American Multimodel Ensemble. International Journal of Climatology, 43(5), 2175–2199. https://doi.org/10.1002/JOC.7969

Kowal, K. M., Slater, L. J., Van Loon, A. F., & Birkel, C. (2021). SEAS5 skilfully predicts late wet-season precipitation in Central American Dry Corridor excelling in Costa Rica and Nicaragua. International Journal of Climatology. https://doi.org/10.1002/JOC.7514

Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C. J. H., Apel, H., Aronica, G. T., Arnbjerg-Nielsen, K., Bouwer, L. M., Bubeck, P., Caloiero, T., Chinh, D. T., Cortès, M., Gain, A. K., Giampá, V., Kuhlicke, C., Kundzewicz, Z. W., Llasat, M. C., Mård, J., Matczak, P., … Merz, B. (2017a). Adaptation to flood risk: Results of international paired flood event studies. Earth's Future, 5(10), 953–965. https://doi.org/10.1002/2017EF000606

Kreibich, H., Müller, M., Schröter, K., & Thieken, A. H. (2017b). New insights into flood warning reception and emergency response by affected parties. Natural Hazards and Earth System Sciences, 17(12), 2075–2092. https://doi.org/10.5194/NHESS-17-2075-2017

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. Journal of Climate, 29(11), 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1

Kretschmer, M., Runge, J., & Coumou, D. (2017). Early prediction of extreme stratospheric polar vortex states based on causal precursors. Geophysical Research Letters, 44(16), 8592–8600. https://doi.org/10.1002/2017GL074696

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. Source: Journal of the American Statistical Association, 47(260), 583–621.

Kuhn, M. (2022). caret: Classification and Regression Training (R package version 6.0-93).

Kuzdas, C., Warner, B., Wiek, A., Yglesias, M., Vignola, R., & Ramírez-Cover, A. (2015). Identifying the potential of governance regimes to aggravate or mitigate local water conflicts in regions threatened by climate change.

Http://Dx.Doi.Org/10.1080/13549839.2015.1129604, 21(11), 1387–1408. https://doi.org/10.1080/13549839.2015.1129604

Lai, A. W. C., Herzog, M., & Graf, H. F. (2018). ENSO forecasts near the spring predictability barrier and possible reasons for the recently reduced predictability. Journal of Climate, 31(2), 815–838. https://doi.org/10.1175/JCLI-D-17-0180.1

Lange, B., & Cook, C. (2015). Mapping a Developing Governance Space: Managing Drought in the UK. Current Legal Problems, 68(1), 229–266. https://doi.org/10.1093/clp/cuv014

Lehner, F., & Deser, C. (2023). Origin, importance, and predictive limits of internal climate variability. Environmental Research: Climate, 2(2), 023001. https://doi.org/10.1088/2752-5295/ACCF30

Leith, C. (1974). Theoretical Skill of Monte Carlo Forecasts. Monthly Weather Review, 102(6), 409–418. https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2

Leutbecher, M. (2019). Ensemble size: How suboptimal is less than infinity? Quarterly Journal of the Royal Meteorological Society, 145(S1), 107–128. https://doi.org/10.1002/QJ.3387

Lin, H., Muncaster, R., Diro, G. T., Merryfield, W., Smith, G., Markovic, M., Erfani, A., Kharin, S., Lee, W.-S., Parent, R., Pavlovic, R., & Charron, M. (2021). The Canadian Seasonal to Interannual Prediction System version 2.1 (CanSIPSv2.1). https://iridl.ldeo.columbia.edu/documentation/Models/NMME/CanSIPS-IC3/technote.pdf

Liniger, M. A., Mathis, H., Appenzeller, C., & Doblas-Reyes, F. J. (2007). Realistic greenhouse gas forcing and seasonal forecasts. Geophysical Research Letters, 34(4). https://doi.org/10.1029/2006GL028335

Litter, M. I., Morgada, M. E., & Bundschuh, J. (2010). Possible treatments for arsenic removal in Latin American waters for human consumption. Environmental Pollution (Barking, Essex : 1987), 158(5), 1105–1118. https://doi.org/10.1016/J.ENVPOL.2010.01.028

Lopez-Ridaura, S., Barba-Escoto, L. Reyna C., Hellin, J., Gerard, B., and van Wijk, M. (2019). Food security and agriculture in the Western Highlands of Guatemala. Food Sec, 11, 817-833. https://doi.org/10.1007/s12571-019-00940-z

Lorenz, E. N. (1963). The predictability of hydrodynamic flow. Trans. New York Acad. Sci., , 2(26), 409–432. https://doi.org/10.1111/j.2164-0947.1963.tb01464.x

Lorenz, E. N. (1972, December 29). Predictibility: Does the flap of a Butterfly's Wings set off a Tornado in Texas? AAAS Section on Environmental Sciences New Approaches to Global Weather: GARP (The Global Atmospheric Research Program). https://fermatslibrary.com/s/predictability-does-the-flap-of-a-butterflys-wings-in-brazil-set-off-a-tornado-in-texas

References

Lorenz, E. N. (1975). The physical bases of climate and climate modelling. Climate Predictability. Global Atmospheric Research Program (GARP) Publication Series, 16, 132–136. https://library.wmo.int/index.php?lvl=notice_display&id=6943

Lorenz, E. N. (1995). The Essence of Chaos: Chapter 3 Our Chaotic Weather. https://uwapress.uw.edu/book/9780295975146/the-essence-of-chaos

Lorenz, R. (2018). Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. J. Geophys. Res. Atmos., 123, 4509–4526. https://doi.org/10.1029/2017JD027992

Luo, J. J., Behera, S. K., Masumoto, Y., & Yamagata, T. (2011). Impact of global ocean surface warming on seasonal-to-interannual climate prediction. Journal of Climate, 24(6), 1626–1646. https://doi.org/10.1175/2010JCLI3645.1

Lyra, A., Imbach, P., Rodriguez, D., Chan Chou, S., Georgiou, S., Garofolo, L., Donatti, C. I., & Hannah, L. (2017). Projections of climate change impacts on central America tropical rainforest. Climatic Change, 141, 93–105. https://doi.org/10.1007/s10584-016-1790-2

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. Journal of the Atmospheric Sciences, 28(5). https://journals.ametsoc.org/view/journals/atsc/28/5/1520-0469_1971_028_0702_doadoi_2_0_co_2.xml

Magaña, V., Amador, J. A., & Medina, S. (1999). The Midsummer Drought over Mexico and Central America. Journal of Climate, 12(6), 1577–1588. https://doi.org/10.1175/1520-0442(1999)012<1577:TMDOMA>2.0.CO;2

Maldonado, T., Alfaro, E., Fallas-López, B., & Alvarado, L. (2013). Seasonal prediction of extreme precipitation events and frequency of rainy days over Costa Rica, Central America, using Canonical Correlation Analysis. Advances in Geosciences, 33, 41–52. https://doi.org/10.5194/adgeo-33-41-2013

Maldonado, T., Rutgersson, A., Alfaro, E. J., Amador, J., & Claremar, B. (2016). Interannual variability of the midsummer drought in Central America and the connection with sea surface temperatures. Advances in Geosciences, 42, 35–50. https://doi.org/10.5194/adgeo-42-35-2016

Maldonado, T., Alfaro, E., Rutgersson, A., & Amador, J. A. (2017). The early rainy season in Central America: the role of the tropical North Atlantic SSTs. International Journal of Climatology, 37(9), 3731–3742. https://doi.org/10.1002/joc.4958

Maldonado, T., Alfaro, E. J., Amador, J. A., & Rutgersson, A. (2018). Regional precipitation estimations in central America using the weather research and forecast model. Revista de Biologia Tropical, 66(May), S231–S254. https://doi.org/10.15517/rbt.v66i1.33303

Maldonado, T., Alfaro, E., & Hidalgo, H. (2018). Revision of the main drivers and variability of Central America Climate and seasonal forecast systems. Revista de Biologia Tropical, 66(Suppl.1), S153–S175. https://www.researchgate.net/publication/324900632_Revision_of_the_main_drivers_and_variability_of_Central_America_Climate_and_seasonal_forecast_systems

References

Mangiafico, S. (2022). rcompanion: Functions to Support Extension Education Program Evaluation (R package version 2.4.18).

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society, 78, 1069–1079. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2

Manzanas, R., Gutiérrez, · J M, Bhend, · J, Hemri, · S, Doblas-Reyes, · F J, Torralba, · V, Penabad, · E, & Brookshaw, · A. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. 53, 1287–1305. https://doi.org/10.1007/s00382-019-04640-4

Mapes, B. E., Liu, P., & Buenning, N. (2005). Indian monsoon onset and the Americas midsummer drought: Out-of-equilibrium responses to smooth seasonal forcing. Journal of Climate, 18(7), 1109–1115. https://doi.org/10.1175/JCLI-3310.1

Marengo, J. A., Chou, S. C., Torres, R. R., Giarolla, A., Alves, L. M., & Lyra, A. (2014). Climate Change in Central and South America: Recent Trends, Future Projections, and Impacts on Regional Agriculture. www.ccafs.cgiar.org

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., & Albers, J. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. Bulletin of the American Meteorological Society, 101(5), E608–E625. https://doi.org/10.1175/BAMS-D-18-0326.1

Marsh, K., & Penebad, E. (2016, February 16). Description of the C3S seasonal multi-system - Copernicus Knowledge Base - ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/CKB/Description+of+the+C3S+seasonal+multi-system

Martinez, C., Goddard, L., Kushnir, Y., & Ting, M. (2019). Seasonal climatology and dynamical mechanisms of rainfall in the Caribbean. Climate Dynamics, 53(1–2), 825–846. https://doi.org/10.1007/s00382-019-04616-4

Martinez, C., Kushnir, Y., Goddard, L., & Ting, M. (2020). Interannual variability of the early and late-rainy seasons in the Caribbean. Climate Dynamics, 55(5–6), 1563–1583. https://doi.org/10.1007/s00382-020-05341-z

Martinez, C., Muñoz, Á. G., Goddard, L., Kushnir, Y., & Ting, M. (2022). Seasonal prediction of the Caribbean rainfall cycle. Climate Services, 27. https://doi.org/10.1016/J.CLISER.2022.100309

Martinez-Sanchez, J. N., & Cavazos, T. (2014). Eastern Tropical Pacific hurricane variability and landfalls on Mexican coasts. Climate Research, 58(3), 221–234. https://www.jstor.org/stable/24896142

Maslin, M. (2013). Commentary Cascading uncertainty in climate change models and its implications for policy. https://doi.org/10.1111/j.1475-4959.2012.00494.x

References

Mason, S. J., & Weigel, A. P. (2009). A Generic Forecast Verification Framework for Administrative Purposes. Monthly Weather Review, 137(1), 331–349. https://doi.org/10.1175/2008MWR2553.1

Mason, S. J. (2018). Guidance on Verification of Operational Seasonal Climate Forecasts. In World Meteorological Organization (Issue WMO-No. 1220). http://www.seevccc.rs/SEECOF/SEECOF-10/SEECOF-LRF-TRAINING/November 13th 2013/CCl verification recommendations.pdf

Mason, S. J., Tippett, M. K., Song, L., & Muñoz, Á. G. (2022). Climate Predictability Tool 17.6.1.

Matheson, J. E., & Winkler, R. L. (1976). Scoring Rules for Continuous Probability Distributions. Management Science, 22(10), 1087–1096. https://www.jstor.org/stable/2629907

Maurer, E. P., Adam, J. C., & Wood, A. W. (2009). Climate model based consensus on the hydrologic impacts of climate change to the Rio Lempa basin of Central America. Hydrology and Earth System Sciences, 13(2), 183–194. https://doi.org/https://doi.org/10.5194/hess-13-183-2009

Maurer, E. P., Stewart, I. T., Joseph, K., & Hidalgo, H. G. (2022). The Mesoamerican mid-summer drought: The impact of its definition on occurrences and recent changes. Hydrology and Earth System Sciences, 26(5), 1425–1437. https://doi.org/10.5194/HESS-26-1425-2022

Mckee, T. B. (1995). Drought monitoring with multiple time scales. Proceedings of 9th Conference on Applied Climatology, Boston.

Mejía, J. F., Yepes, J., Henao, J. J., Poveda, G., Zuluaga, M. D., Raymond, D. J., & Fuchs-Stone, Ž. (2021). Towards a Mechanistic Understanding of Precipitation Over the Far Eastern Tropical Pacific and Western Colombia, One of the Rainiest Spots on Earth. Journal of Geophysical Research: Atmospheres, 126(5). https://doi.org/10.1029/2020JD033415

Merryfield, W. J., Lee, W., Wang, W., Chen, M., & Kumar, A. (2013). Multi-system seasonal predictions of Arctic sea ice. Geophys. Res. Letters, 40, 1551–1556. https://doi.org/10.1002/grl.50317

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., Domeisen, D. I. V., Feser, F., Koszalka, I., Kreibich, H., Pantillon, F., Parolai, S., Pinto, J. G., Punge, H. J., Rivalta, E., Schröter, K., Strehlow, K., Weisse, R., & Wurpts, A. (2020). Impact Forecasting to Support Emergency Management of Natural Hazards. Reviews of Geophysics, 58(4), e2020RG000704. https://doi.org/10.1029/2020RG000704

Meza, C. (2014). A review on the Central America electrical energy scenario. Renewable and Sustainable Energy Reviews, 33, 566–577. https://doi.org/10.1016/J.RSER.2014.02.022

Mishra, N., Prodhomme, C., & Guemas, V. (2019). Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. Climate Dynamics, 52(7–8), 4207–4225. https://doi.org/10.1007/s00382-018-4404-z

References

Mo, K. C., & Lyon, B. (2015). Global meteorological drought prediction using the North American multi-model ensemble. Journal of Hydrometeorology, 16(3), 1409–1424. https://doi.org/10.1175/JHM-D-14-0192.1

Molod, A., Hackert, E., Vikhliaev, Y., Zhao, B., Barahona, D., Vernieres, G., Borovikov, A., Kovach, R. M., Marshak, J., Schubert, S., Li, Z., Lim, Y.-K., Andrews, L. C., Cullather, R., Koster, R., Achuthavarier, D., Carton, J., Coy, L., Friere, J. L. M., … Pawson, S. (2020). GEOS-S2S Version 2: The GMAO High-Resolution Coupled Model and Assimilation System for Seasonal Prediction. Journal of Geophysical Research: Atmospheres, 125. https://doi.org/10.1029/2019JD031767

Muñoz, Á. G., Goddard, L., Mason, S. J., & Robertson, A. W. (2016). Cross–Time Scale Interactions and Rainfall Extreme Events in Southeastern South America for the Austral Summer. Part II: Predictive Skill. Journal of Climate, 29(16), 5915–5934. https://doi.org/10.1175/JCLI-D-15-0699.1

Muñoz, Á. G., Pons, D., Giraldo-Mendez, D., Adamo, S. B., de Sherbinin, A. M., & Goddard, L. M. (2019). Can We Predict 'Climate Migrations'? The 2018 Guatemalan Case. AGUFM, 2019, GC13E-1213. https://ui.adsabs.harvard.edu/abs/2019AGUFMGC13E1213M/abstract

Muñoz, Á. G., Robertson, A. W., Turkington, T., Mason, S. J., & contributors. (2019). PyCPT: a Python interface and enhancement for IRI's Climate Predictability Tool.

Muñoz, Á. G., Chourio, X., Rivière-Cinnamond, A., Diuk-Wasser, M. A., Kache, A., Mordecai, A., Harrington, L., & Thomson, M. C. (2020). American Health Organization AeDeS: a next generation monitoring and forecasting system for environmental suitability of Aedes borne disease transmission. World Health Organization, 10, 12640. https://doi.org/10.1038/s41598-020-69625-4

Muñoz, E., Busalacchi, A. J., Nigam, S., & Ruiz-Barradas, A. (2008). Winter and Summer Structure of the Caribbean Low-Level Jet. Journal of Climate, 21(6), 1260–1276. https://doi.org/10.1175/2007JCLI1855.1

Muñoz-Jiménez, R., Giraldo-Osorio, J. D., Brenes-Torres, A., Avendaño-Flores, I., Nauditt, A., Hidalgo-León, H. G., & Birkel, C. (2019). Spatial and temporal patterns, trends and teleconnection of cumulative rainfall deficits across Central America. International Journal of Climatology, 39(4), 1940–1953. https://doi.org/10.1002/joc.5925

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather & Forecasting, 8(2), 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2

NCAR - Research Applications Laboratory. (2015). verification: Weather Forecast Verification Utilities (R package version 1.42).

Ng, J. Y., Turner, S. W. D., & Galelli, S. (2017). Influence of El Niño Southern Oscillation on global hydropower production. Environmental Research Letters, 12(3). https://doi.org/10.1088/1748-9326/aa5ef8

NOAA. (2020). The North American Multi-Model Ensemble (NMME) Operational Phase 2019-2020 Annual Report. https://doi.org/10.25923/TF81-0S82

References

NOAA. (2022). OOPC | State of the ocean climate | Surface indices | Atlantic | TNA. https://stateoftheocean.osmc.noaa.gov/sur/atl/tna.php

NOAA Climate Prediction Center. (2021). Historical El Nino / La Nina episodes (1950-present). https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

Nowack, P., Runge, J., Eyring, V., & Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. Nature Communications 2020 11:1, 11(1), 1–11. https://doi.org/10.1038/s41467-020-15195-y

O'Brien, K., Eriksen, S., Nygaard, L. P., & Schjolden, A. (2007). Why different interpretations of vulnerability matter in climate change discourses. Climate Policy, 7(1), 73–88. https://doi.org/10.1080/14693062.2007.9685639

Oldenborgh, G. J. van, Balmaseda, M. A., Ferranti, L., Stockdale, T. N., & Anderson, D. L. T. (2005). Evaluation of Atmospheric Fields from the ECMWF Seasonal Forecasts over a 15-Year Period. Journal of Climate, 18(16), 3250–3269. https://doi.org/10.1175/JCLI3421.1

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., Déqué, M., Díez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J. F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., … Thomson, M. C. (2004). DEVELOPMENT OF A EUROPEAN MULTIMODEL ENSEMBLE SYSTEM FOR SEASONAL-TO-INTERANNUAL PREDICTION (DEMETER). Bulletin of the American Meteorological Society, 85(6), 853–872. https://doi.org/10.1175/BAMS-85-6-853

Palmer, T. N. (2006). Predictability of weather and climate: From theory to practice. In Predictability of Weather and Climate (Vol. 9780521848824, pp. 1–29). Cambridge University Press. https://doi.org/10.1017/CBO9780511617652.002

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. . The R Journal , 10(1), 439–446.

Pedersen, T. L. (2022). patchwork: The Composer of Plots. (R package version 1.1.2).

Peralta Rodríguez, O., Carrazón Alocén, J., & Andrés Zelaya Elvir, C. (2012). Buenas prácticas para la seguridad alimentaria y la gestión de riesgos.

Perdigón-Morales, J., Romero-Centeno, R., Ordoñez, P., Nieto, R., Gimeno, L., & Barrett, B. S. (2021). Influence of the Madden-Julian Oscillation on moisture transport by the Caribbean Low Level Jet during the Midsummer Drought in Mexico. Atmospheric Research, 248, 105243. https://doi.org/10.1016/J.ATMOSRES.2020.105243

Peréz-Briceño, P. M., Alfaro, E. J., Hidalgo, H. G., & Jimenez, F. (2016). Distribución espacial de impactos de eventos hidrometeorológicos en América Central. Revista de Climatologia, 16, 63–75. https://www.researchgate.net/publication/307629313_Distribucion_espacial_de_impactos_de_eventos_hidrometeorologicos_en_America_Central

Peterson, T. C., Taylor, M. A., Demeritte, R., Duncombe, D. L., Burton, S., Thompson, F., Porter, A., Mercedes, M., Villegas, E., Fils, R. S., Tank, A. K., Martis, A., Warner, R.,

Joyette, A., Mills, W., Alexander, L., & Gleason, B. (2002). Recent changes in climate extremes in the Caribbean region. Journal of Geophysical Research Atmospheres, 107(21). https://doi.org/10.1029/2002JD002251

Pierce, D. (2019). ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files (R package version 1.17).

Pillai, P. A., Rao, S. A., Ramu, D. A., Pradhan, M., & George, G. (2018). Seasonal prediction skill of Indian summer monsoon rainfall in NMME models and monsoon mission CFSv2. International Journal of Climatology, 38, e847–e861. https://doi.org/10.1002/JOC.5413

Pohlert, T. (2022). PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended (R package version 1.9.6).

Pons, D., Muñoz, Á. G., Meléndez, L. M., Chocooj, M., Gómez, R., Chourio, X., & González Romero, C. (2021). A Coffee Yield Next-Generation Forecast System for Rain-fed Plantations: the Case of the Samalá Watershed in Guatemala. Weather and Forecasting. https://doi.org/10.1175/WAF-D-20-0133.1

Portig, W. H. (1961). Some Climatological Data of Salvador, Central America. Weather, 16(4), 103–112. https://doi.org/10.1002/j.1477-8696.1961.tb01900.x

Poveda, G., & Mesa, O. J. (1999). La Corriente de chorro superficial del oeste “del Chocó”: Climatología y Variabilidad durante las fases del ENSO. Proc. V Congreso Colombiano de Meteorología, January. https://www.researchgate.net/publication/233857360

Poveda, G., & Mesa, O. J. (2000). On the existence of Lloró (the rainiest locality on Earth): Enhanced ocean-land-atmosphere interaction by a low-level jet. Geophysical Research Letters, 27(11), 1675–1678. https://doi.org/10.1029/1999GL006091

Poveda, G., Waylen, P. R., & Pulwarty, R. S. (2006). Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica. Palaeogeography, Palaeoclimatology, Palaeoecology, 234(1), 3–27. https://doi.org/10.1016/j.palaeo.2005.10.031

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., & Foresti, L. (2019). Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0). Geoscientific Model Development, 12(10), 4185–4219. https://doi.org/10.5194/GMD-12-4185-2019

Quesada-Hernández, L. E., Calvo-Solano, O. D., Hidalgo, H. G., Pérez-Briceño, P. M., & Alfaro, E. J. (2019). Dynamical delimitation of the Central American Dry Corridor (CADC) using drought indices and aridity values. Progress in Physical Geography: Earth and Environment, 43(5), 627–642. https://doi.org/10.1177/0309133319860224

Quesada-Montano, B., Wetterhall, F., Westerberg, I. K., Hidalgo, H. G., & Halldin, S. (2019). Characterising droughts in Central America with uncertain hydro-meteorological data. Theoretical and Applied Climatology, 137(3–4), 2125–2138. https://doi.org/10.1007/s00704-018-2730-z

References

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Rangel Soares, L. C., Griesinger, M. O., Dachs, J. N. W., Bittner, M. A., & Tavares, S. (2002). Inequities in access to and use of drinking water services in Latin America and the Caribbean. Revista Panamericana de Salud Publica = Pan American Journal of Public Health, 11(5–6), 386–396. https://doi.org/10.1590/S1020-49892002000500013

Ratri, D. N., Whan, K., & Schmeits, M. (2019). A comparative verification of raw and bias-corrected ECMWF seasonal ensemble precipitation reforecasts in Java (Indonesia). Journal of Applied Meteorology and Climatology, 58(8), 1709–1723. https://doi.org/10.1175/JAMC-D-18-0210.1

Rauscher, S. A., Giorgi, F., Diffenbaugh, N. S., & Seth, A. (2008). Extension and Intensification of the Meso-American mid-summer drought in the twenty-first century. Climate Dynamics, 31(5), 551–571. https://doi.org/10.1007/s00382-007-0359-1

Rauscher, S. A., Kucharski, F., & Enfield, D. B. (2011). The role of regional SST warming variations in the drying of Meso-America in future climate projections. Journal of Climate, 24(7), 2003–2016. https://doi.org/10.1175/2010JCLI3536.1

Rayner, S., Lach, D., & Ingram, H. (2005). Weather forecasts are for wimps: Why water resource managers do not use climate forecasts. Climatic Change, 69(2–3), 197–227. https://doi.org/10.1007/s10584-005-3148-z

Red Cross Red Crescent (2022a). First-ever early action protocols triggered in Central America for TS Julia. October. https://www.climatecentre.org/9363/first-ever-early-action-protocols-triggeredin-central-america-for-ts-julia/

Red Cross Red Crescent (2022b). Guatemala: Floods associated with Tropical Cyclones Early Action Protocol summary. August https://adore.ifrc.org/Download.aspx?FileId=569393

Red Cross Red Crescent (2022c) Honduras: Tropical Storms Early Action Protocol summary. June. https://adore.ifrc.org/Download.aspx?FileId=547660

Renard, B., & Thyer, M. (2019). Revealing Hidden Climate Indices from the Occurrence of Hydrologic Extremes. Water Resources Research. https://doi.org/10.1029/2019WR024951

Reynolds, R., Smith, T., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Surface Temperature. Journal of Climate. https://www.jstor.org/stable/26259893

Risbey, J. S., Squire, D. T., Black, A. S., DelSole, T., Lepore, C., Matear, R. J., Monselesan, D. P., Moore, T. S., Richardson, D., Schepen, A., Tippett, M. K., & Tozer, C. R. (2021). Standard assessments of climate forecast skill can be misleading. Nature Communications 2021 12:1, 12(1), 1–14. https://doi.org/10.1038/s41467-021-23771-z

References

Sachs, J. P., Sachse, D., Smittenberg, R. H., Zhang, Z., Battisti, D. S., & Golubic, S. (2009). Southward movement of the Pacific intertropical convergence zone AD 1400-1850. Nature Geoscience, 2(7), 519–525. https://doi.org/10.1038/ngeo554

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2. Journal of Climate, 27(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Sánchez-Murillo, R., Durán-Quesada, A. M., Birkel, C., Esquivel-Hernández, G., & Boll, J. (2017). Tropical precipitation anomalies and d-excess evolution during El Niño 2014-16. Hydrological Processes, 31(4), 956–967. https://doi.org/10.1002/hyp.11088

Sanderson, B. M., Knutti, R., & Caldwell P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. J. Clim., 28, 5171–5194. https://doi.org/10.1175/JCLI-D-14-00362.1

Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. Geoscientific Model Development, 10(6), 2379–2395. https://doi.org/10.5194/gmd-10-2379-2017

Scaife, A. A., Ferranti, L., Alves, O., Athanasiadis, P., Baehr, J., Dequé, M., Dippe, T., Dunstone, N., Fereday, D., Gudgel, R. G., Greatbatch, R. J., Hermanson, L., Imada, Y., Jain, S., Kumar, A., MacLachlan, C., Merryfield, W., Müller, W. A., Ren, H.-L., … Yang, X. (2019). Tropical rainfall predictions from multiple seasonal forecast systems. International Journal of Climatology, 39(2), 974–988. https://doi.org/10.1002/joc.5855

Schneider, U., Becker, A., Finger, P., & Meyer-Christoffer, Anja Ziese, M. (2018). GPCC Full Data Monthly Product Version 2018 at 0.25’: Monthy Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data. https://doi.org/10.5676/DWD_GPCC/FD_M_V2018_025

Serafin, S., Adler, B., Cuxart, J., De Wekker, S., Gohm, A., Grisogono, B., Kalthoff, N., Kirshbaum, D., Rotach, M., Schmidli, J., Stiperski, I., Večenaj, Ž., & Zardi, D. (2018). Exchange Processes in the Atmospheric Boundary Layer Over Mountainous Terrain. Atmosphere, 9(3), 102. https://doi.org/10.3390/atmos9030102

Sesnie, S. E., Tellman, B., Wrathall, D., McSweeney, K., Nielsen, E., Benessaiah, K., Wang, O., & Rey, L. (2017). A spatio-temporal analysis of forest loss related to cocaine trafficking in Central America. Environmental Research Letters, 12(5), 054015. https://doi.org/10.1088/1748-9326/AA6FFF

Shi, W., Schaller, N., Macleod, D., Palmer, T. N., & Weisheimer, A. (2015). Impact of hindcast length on estimates of seasonal climate predictability. Geophysical Research Letters, 42(5), 1554–1559. https://doi.org/10.1002/2014GL062829

Sierra, J. P., Arias, P. A., Durán-Quesada, A. M., Tapias, K. A., Vieira, S. C., & Martínez, J. A. (2021). The Choco low-level jet: past, present and future. Climate Dynamics 2021 56:7, 56(7), 2667–2692. https://doi.org/10.1007/S00382-020-05611-W

References

Slater, L. J., Villarini, G., & Bradley, A. A. (2016). Evaluation of the skill North-America Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. Climate Dynamics. https://doi.org/10.1007/s00382-016-3286-1

Slater, L. J., Villarini, G., & Bradley, A. A. (2017). Weighting of NMME temperature and precipitation forecasts across Europe. Journal of Hydrology, 552, 646–659. https://doi.org/10.1016/J.JHYDROL.2017.07.029

Slater, L. J., & Villarini, G. (2018). Enhancing the Predictability of Seasonal Streamflow With a Statistical-Dynamical Approach. Geophysical Research Letters, 45(13), 6504–6513. https://doi.org/10.1029/2018GL077945

Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2022). Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models. Hydrology and Earth System Sciences Discussions, 1–35. https://doi.org/10.5194/HESS-2022-334

Small, R. J. O., de Szoeke, S. P., & Xie, S. P. (2007). The Central American Midsummer Drought: Regional Aspects and Large-Scale Forcing. Journal of Climate, 20(19), 4853–4873. https://doi.org/10.1175/JCLI4261.1

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., … Zhang, L. (2020). North Atlantic climate far more predictable than models imply Check for updates. Nature, 583. https://doi.org/10.1038/s41586-020-2525-0

Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M. M. B., LeRoy Miller, H. J., & Chen, Z. (2007). The Physical Science Basis: Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). http://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html

Sori, R., Drumond, A., & Nieto, R. (2015). Moisture contribution of the Atlantic warm pool to precipitation: A Lagrangian analysis. Frontiers in Environmental Science, 3(MAR), 1–11. https://doi.org/10.3389/fenvs.2015.00022

Spence, J. M., Taylor, M. A., & Chen, A. A. (2004). The effect of concurrent sea-surface temperature anomalies in the tropical Pacific and Atlantic on Caribbean rainfall. International Journal of Climatology, 24(12), 1531–1541. https://doi.org/10.1002/JOC.1068

Stephenson, D. B., Wanner, H., Brönnimann, S., & Luterbacher, J. (2003). The history of scientific research on the north atlantic oscillation. Geophysical Monograph Series, 134, 37–50. https://doi.org/10.1029/134GM02

Stephenson, T. S., Vincent, L. A., Allen, T., Van Meerbeeck, C. J., Mclean, N., Peterson, T. C., Taylor, M. A., Aaron-Morrison, A. P., Auguste, T., Bernard, D., Boekhoudt, J. R. I., Blenman, R. C., Braithwaite, G. C., Brown, G., Butler, M., Cumberbatch, C. J. M.,

Etienne-Leblanc, S., Lake, D. E., Martin, D. E., … Trotman, A. R. (2014). Changes in extreme temperature and precipitation in the Caribbean region, 1961-2010. International Journal of Climatology, 34(9), 2957–2971. https://doi.org/10.1002/joc.3889

Stewart, I. T., Maurer, E. P., Stahl, K., & Joseph, K. (2022). Recent evidence for warmer and drier growing seasons in climate sensitive regions of Central America from multiple global datasets. International Journal of Climatology, 42(3), 1399–1417. https://doi.org/10.1002/JOC.7310

Stockdale, T., Alves, O., Boer, G., Deque, M., Ding, Y., Kumar, A., Kumar, K., Landman, W., Mason, S., Nobre, P., Scaife, A., Tomoaki, O., & Yun, W. T. (2010). Understanding and Predicting Seasonal-to-Interannual Climate Variability-The Producer Perspective. Procedia Environmental Sciences, 1, 55–80. https://doi.org/10.1016/j.proenv.2010.09.006

Stockdale, T., Balmaseda, M., Johnson, S., Ferranti, L., Molteni, F., Magnusson, L., Tietsche, S., Vitart, F., Decremer, D., Weisheimer, A., Roberts, C., Balsamo, G., Keeley, S., Mogensen, K., Zuo, H., Mayer, M., & Monge-Sanz, B. (2018). Technical Memorandum 835: SEAS5 and the future evolution of the long-range forecast system. http://www.ecmwf.int/en/research/publications

Straffon, A., Zavala-Hidalgo, J., & Estrada, F. (2020). Preconditioning of the precipitation interannual variability in southern Mexico and Central America by oceanic and atmospheric anomalies. International Journal of Climatology, 40(8), 3906–3921. https://doi.org/10.1002/JOC.6434

Strazzo, S., Collins, D. C., Schepen, A., Wang, Q. J., Becker, E., & Jia, L. (2019). Application of a Hybrid Statistical–Dynamical System to Seasonal Prediction of North American Temperature and Precipitation. Monthly Weather Review, 147(2), 607–625. https://doi.org/10.1175/MWR-D-18-0156.1.10.1175/MWR-D-18-0156.1

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. Reviews of Geophysics, 56(1), 79–107. https://doi.org/10.1002/2017RG000574

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., & Stephens, S. (2002). The Drought Monitor. Drought Mitigation Center Faculty Publications. https://digitalcommons.unl.edu/droughtfacpub/166/

Takaya, Y., Yasuda, T., Fujii, Y., Matsumoto, S., Soga, T., Mori, H., Hirai, M., Ishikawa, I., Sato, H., Shimpo, A., Kamachi, M., Ose, T., Takaya, Y., Yasuda, T., Fujii, Y., Matsumoto, S., Soga, T., Mori, H., Hirai, M., … Ose, T. (2017). Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 1 (JMA/MRI-CPS1) for operational seasonal forecasting. ClDy, 48(1–2), 313–333. https://doi.org/10.1007/S00382-016-3076-9

Taylor, M. A., Enfield, D. B., Chen, A. A., Taylor, M. A., Enfield, D. B., & Chen, A. A. (2002). Influence of the tropical Atlantic versus the tropical Pacific on Caribbean rainfall. Journal of Geophysical Research: Oceans, 107(C9), 10–11. https://doi.org/10.1029/2001JC001097

References

Taylor, M. A., & Alfaro, E. J. (2005). Central America and the Caribbean, Climate of. Encyclopedia of Earth Sciences Series, 183–189. https://doi.org/10.1007/1-4020-3266-8_37/COVER/

Taylor, M. A., Stephenson, T. S., Chen, A. A., & Stephenson, K. A. (2012). Climate Change and the Caribbean, Review and Response. Caribbean Studies, 40(2), 169–200.

Tippett, M. K., Goddard, L., & Barnston, A. G. (2005). Statistical-dynamical seasonal forecasts of central-southwest Asian winter precipitation. Journal of Climate, 18(11), 1831–1843. https://doi.org/10.1175/JCLI3371.1

Tracton, S. M., & Kalnay, E. (1993). Operational Ensemble Prediction at the National Meteorological Center: Practical Aspects. Weather and Forecasting, 8, 379–398. https://doi.org/https://doi.org/10.1175/1520-0434(1993)008<0379:OEPATN>2.0.CO;2

Trenberth, K. E. (1997). The Definition of El Niño. Bulletin of the American Meteorological Society, 78(12). https://doi.org/https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of El Niño Evolution. Journal of Climate, 14(8). https://doi.org/https://doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. L. (2021). Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. Water Resources Research, 57(7), e2020WR028392. https://doi.org/10.1029/2020WR028392

Van Loon, A. F. (2015). Hydrological drought explained. Wiley Interdisciplinary Reviews: Water, 2(4), 359–392. https://doi.org/10.1002/wat2.1085

Vaughan, C., & Dessai, S. (2014). Climate services for society: Origins, institutional arrangements, and design elements for an evaluation framework. Wiley Interdisciplinary Reviews: Climate Change, 5(5), 587–603. https://doi.org/10.1002/wcc.290

Veiga, S. F., Nobre, P., Giarolla, E., Capistrano, V., Baptista, M., Marquez, A. L., Nilo Figueroa, S., Bonatti, J. P., Kubota, P., & Nobre, C. A. (2019). The Brazilian Earth System Model ocean-atmosphere (BESM-OA) version 2.5: Evaluation of its CMIP5 historical simulation. Geoscientific Model Development, 12(4), 1613–1642. https://doi.org/10.5194/GMD-12-1613-2019

Vichot-Llano, A., Martinez-Castro, D., Bezanilla-Morlot, A., Centella-Artola, A., & Giorgi, F. (2021). Projected changes in precipitation and temperature regimes and extremes over the Caribbean and Central America using a multiparameter ensemble of RegCM4. International Journal of Climatology, 41(2), 1328–1350. https://doi.org/10.1002/JOC.6811

Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. Npj Climate and Atmospheric Science 2018 1:1, 1(1), 1–7. https://doi.org/10.1038/s41612-018-0013-0

References

Walker, D. P., Birch, C. E., Marsham, J. H., Scaife, A. A., Graham, R. J., & Segele, Z. T. (2019). Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. Climate Dynamics 2019 53:7, 53(7), 4911–4935. https://doi.org/10.1007/S00382-019-04835-9

Wang, B., Lee, J. Y., Kang, I. S., Shukla, J., Park, C. K., Kumar, A., Schemm, J., Cocke, S., Kug, J. S., Luo, J. J., Zhou, T., Wang, B., Fu, X., Yun, W. T., Alves, O., Jin, E. K., Kinter, J., Kirtman, B., Krishnamurti, T., … Yamagata, T. (2009). Advance and prospectus of seasonal prediction: Assessment of the APCC/ CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004). Climate Dynamics, 33(1), 93–117. https://doi.org/10.1007/S00382-008-0460-0/FIGURES/19

Wang, C., & Enfield, D. B. (2001). The tropical western hemisphere warm pool. Geophysical Research Letters, 28(8), 1635–1638. https://doi.org/10.1029/2000GL011763

Wang, C., & Enfield, D. B. (2003). A Further Study of the Tropical Western Hemisphere Warm Pool. Journal of Climate, 16(10), 1476–1493. https://doi.org/10.1175/1520-0442(2003)016<1476:AFSOTT>2.0.CO;2

Wang, C. (2007). Variability of the Caribbean Low-Level Jet and its relations to climate. Climate Dynamics, 29(4), 411–422. https://doi.org/10.1007/s00382-007-0243-z

Wang, C., Lee, S. K., & Enfield, D. B. (2008). Climate response to anomalously large and small Atlantic warm pools during the summer. Journal of Climate, 21(11), 2437–2450. https://doi.org/10.1175/2007JCLI2029.1

Wang, Q. J., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., & Pappenberger, F. (2019). An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. In Environmental Modelling and Software (Vol. 122). https://doi.org/10.1016/j.envsoft.2019.104550

Waylen, P. R., Quesada, M. E., & Caviedes, C. N. (1994). The effects of el niño-southern oscillation on precipitation in san josé, costa rica. International Journal of Climatology, 14(5), 559–568. https://doi.org/10.1002/JOC.3370140506

Waylen, P. R., & Quesada, M. (2001). The effect of Atlantic and Pacific sea surface temperatures on the mid-summer drought of Costa Rica. Cuadernos de Investigación Geográfica, 27, 193–205. https://doi.org/10.18172/CIG.1123

Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2009). Seasonal Ensemble Forecasts: Are Recalibrated Single Models Better than Multimodels? Monthly Weather Review, 137(4), 1460–1479. https://doi.org/10.1175/2008MWR2773.1

Weigel, A. P., Knutti, R., Liniger, M. A., & Appenzeller, C. (2010). Risks of Model Weighting in Multimodel Climate Projections. Journal of Climate, 23(15), 4175–4191. https://doi.org/10.1175/2010JCLI3594.1

Weigel, A. P., & Mason, S. J. (2011). The generalized discrimination score for ensemble forecasts. Monthly Weather Review, 139(9), 3069–3074. https://doi.org/10.1175/MWR-D-10-05069.1

Weigel, A. P. (2017). afc: Generalized Discrimination Score (R package version 1.4.0).

References

Weisheimer, A., & Palmer, T. N. (2014). On the reliability of seasonal climate forecasts. Journal of The Royal Society Interface, 11(96), 20131162. https://doi.org/10.1098/rsif.2013.1162

White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., Balogun, A. A., Bertram, D., Bluhm, S., Brayshaw, D. J., Browell, J., Büeler, D., Charlton-Perez, A., Chourio, X., Christel, I., Coelho, C. A. S., DeFlorio, M. J., Monache, L. D., Giuseppe, F. Di, … Wilson, R. G. (2022). Advances in the Application and Utility of Subseasonal-to-Seasonal Predictions. Bulletin of the American Meteorological Society, 103(6), E1448–E1472. https://doi.org/10.1175/BAMS-D-20-0224.1

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation (R package version 1.0.7).

Wickham, H., & Seidel, D. (2020). scales: Scale Functions for Visualization. (R package version 1.1.1.).

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Bulletin, 1(6), 80–83.

Wilhite, D. A. (2016). Managing drought risk in a changing climate. Drought Mitigation Center Faculty Publications, 156. https://doi.org/10.3354/cr01430

WMO. (2016). Integrated Drought Management Programme Handbook of Drought Indicators and Indices. https://www.droughtmanagement.info/literature/GWP_Handbook_of_Drought_Indicators_and_Indices_2016.pdf

WMO. (2020). Guidance on Operational Practices for Objective Seasonal Forecasting. https://library.wmo.int/index.php?lvl=notice_display&id=21741

WMO (2023). Early Warnings For All Initiative scaled up into action on the ground. WMO Media Press Release. https://public.wmo.int/en/media/press-release/early-warnings-all-initiative-scaled-action-ground#:~:text=The%20Early%20Warnings%20For%20All%20Initiative%20(EW4All)%20was%20formally%20launched,by%20 the%20end%20of%202027

Wolter, K., & Timlin, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). Int. J. Climatol, 31, 1074–1087. https://doi.org/10.1002/joc.2336

Wulff, C. O., Vitart, F., & Domeisen, D. I. V. (2022). Influence of trends on subseasonal temperature prediction skill. Quarterly Journal of the Royal Meteorological Society, 148(744), 1280–1299. https://doi.org/10.1002/QJ.4259

References

Yepes, J., Poveda, G., Mejía, J. F., Moreno, L., & Rueda, C. (2019). CHOCO-JEX: A Research Experiment Focused on the Chocó Low-Level Jet over the Far Eastern Pacific and Western Colombia. Bulletin of the American Meteorological Society, 100(5), 779–796. https://doi.org/10.1175/BAMS-D-18-0045.1

Young, R. M. B. (2010). Decomposition of the Brier score for weighted forecast-verification pairs. Quarterly Journal of the Royal Meteorological Society By, 136, 1364–1370. https://doi.org/10.1002/qj.641

Yuan, X., & Wood, E. F. (2013). Multimodel seasonal forecasting of global drought onset. Geophysical Research Letters, 40(18), 4900–4905. https://doi.org/10.1002/GRL.50949

Zebiak, S. E., Orlove, B., Muñoz, Á. G., Vaughan, C., Hansen, J., Troy, T., Thomson, M. C., Lustig, A., & Garvin, S. (2015). Investigating El Niño-Southern Oscillation and society relationships. Wiley Interdisciplinary Reviews: Climate Change, 6(1), 17–34. https://doi.org/10.1002/WCC.294

Zhang, W., Villarini, G., Slater, L., Vecchi, G. A., & Allen Bradley, A. (2017). Improved ENSO Forecasting Using Bayesian Updating and the North American Multimodel Ensemble (NMME). Journal of Climate, 30(22), 9007–9025. https://doi.org/10.1175/JCLI-D-17-0073.1

Zhao, T., Chen, H., Pan, B., Ye, L., Cai, H., Zhang, Y., & Chen, X. (2021). Correspondence relationship between ENSO teleconnection and anomaly correlation for GCM seasonal precipitation forecasts. Climate Dynamics. https://doi.org/10.1007/S00382-021-05925-3

Zuluaga, M. D., & Houze, R. A. (2015). Extreme Convection of the Near-Equatorial Americas, Africa, and Adjoining Oceans as seen by TRMM. Monthly Weather Review, 143(1), 298–316. https://doi.org/10.1175/MWR-D-14-00109.1