

Geophysical Research Letters[®]



RESEARCH LETTER

10.1029/2023GL105891

Process-Informed Subsampling Improves Subseasonal Rainfall Forecasts in Central America

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

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. M. Kowal,
katiemkowal@gmail.com

Citation:

Kowal, K. M., Slater, L. J., Li, S., Kelder, T., Hall, K. J. C., Moulds, S., et al. (2024). Process-informed subsampling improves subseasonal rainfall forecasts in Central America. *Geophysical Research Letters*, 51, e2023GL105891. <https://doi.org/10.1029/2023GL105891>







Received 11 AUG 2023
Accepted 20 DEC 2023

Author Contributions:

Conceptualization: Katherine M. Kowal, Louise J. Slater, Simon Moulds, Alan A. García-López, Christian Birkel
Data curation: Katherine M. Kowal
Formal analysis: Katherine M. Kowal
Investigation: Katherine M. Kowal
Methodology: Katherine M. Kowal, Louise J. Slater, Simon Moulds
Resources: Christian Birkel
Software: Katherine M. Kowal, Kyle J. C. Hall
Supervision: Louise J. Slater, Christian Birkel
Validation: Katherine M. Kowal
Visualization: Katherine M. Kowal

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Katherine M. Kowal¹ , Louise J. Slater¹ , Sihan Li², Timo Kelder³, Kyle J. C. Hall^{4,5} , Simon Moulds^{1,6} , Alan A. García-López⁷ , and Christian Birkel⁸ 

¹Department of Geography and the Environment, University of Oxford, Oxford, UK, ²Department of Geography, University of Sheffield, Sheffield, UK, ³Climate Adaptation Services, Bussum, The Netherlands, ⁴National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory, Boulder, CO, USA, ⁵Cooperative Institute for Research in Environmental Sciences, NOAA and University of Colorado Boulder, Boulder, CO, USA, ⁶School of GeoSciences, University of Edinburgh, Edinburgh, UK, ⁷Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA, ⁸Department of Geography, University of Costa Rica, San Jose, Costa Rica

Abstract Subseasonal rainfall forecast skill is critical to support preparedness for hydrometeorological extremes. We assess how a process-informed evaluation, which subsamples forecasting model members based on their ability to represent potential predictors of rainfall, can improve monthly rainfall forecasts within Central America in the following month, 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 multimodel ensemble based on their representation of both (a) zonal wind direction and (b) 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 in the following month, using Costa Rica and Guatemala as key test cases. We select members from five contributing models of the C3S multimodel 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.

1. Introduction

Rainfall forecasts can support preparedness for hydrometeorological extremes like droughts and floods (Braman et al., 2013; Domeisen et al., 2022; Merz et al., 2020; 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 & Robertson, 2018).

Atmospheric oceanic general circulation models (AOGCMs) provide one way to predict rainfall by generating dynamic predictions of the earth system (Bauer et al., 2015; Hagedorn et al., 2005; Stockdale et al., 2010). 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 et al., 2009). Traditionally, model evaluation is conducted based on aggregated mean

Writing – original draft: Katherine M. Kowal

Writing – review & editing: Katherine M. Kowal, Louise J. Slater, Kyle J. C. Hall, Simon Moulds, Alan A. García-López, Christian Birkel

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 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 direction). 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 months' lead time. To subsample members in a manner that could be operationally deployed, we identify better performing model members at the point of model initialization. We evaluate the predictions of SST and zonal wind direction, two variables that change more slowly over time, enabling us to use preceding 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 (e.g., Carrão et al., 2018; Hidalgo & Alfaro, 2012, 2015; Kowal et al., 2021, 2023; Maldonado, Alfaro, Amador, & Rutgersson, 2018; Maldonado, Alfaro, & Hidalgo, 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 et al., 2017, 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 examine whether our method improves subseasonal rainfall forecast skill in two distinct subregions: Guatemala and Costa Rica (Figure 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 skillful.

2. Data

We generated a MME using members from five AOGCMs (Table S1 in Supporting Information S1) 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, et al., 2018). We used Climate Hazards Group InfraRed Precipitation with Station 0.25° rainfall data (CHIRPS - Funk et al., 2014) to test the final predictions given CHIRPS has been found to be a reasonable data set within the region (e.g., Arciniega-Esparza et al., 2022). All variables tested (predictors and rainfall predictands) were converted into standardized anomalies prior to testing (Equation S1 in Text S1 in Supporting Information S1).

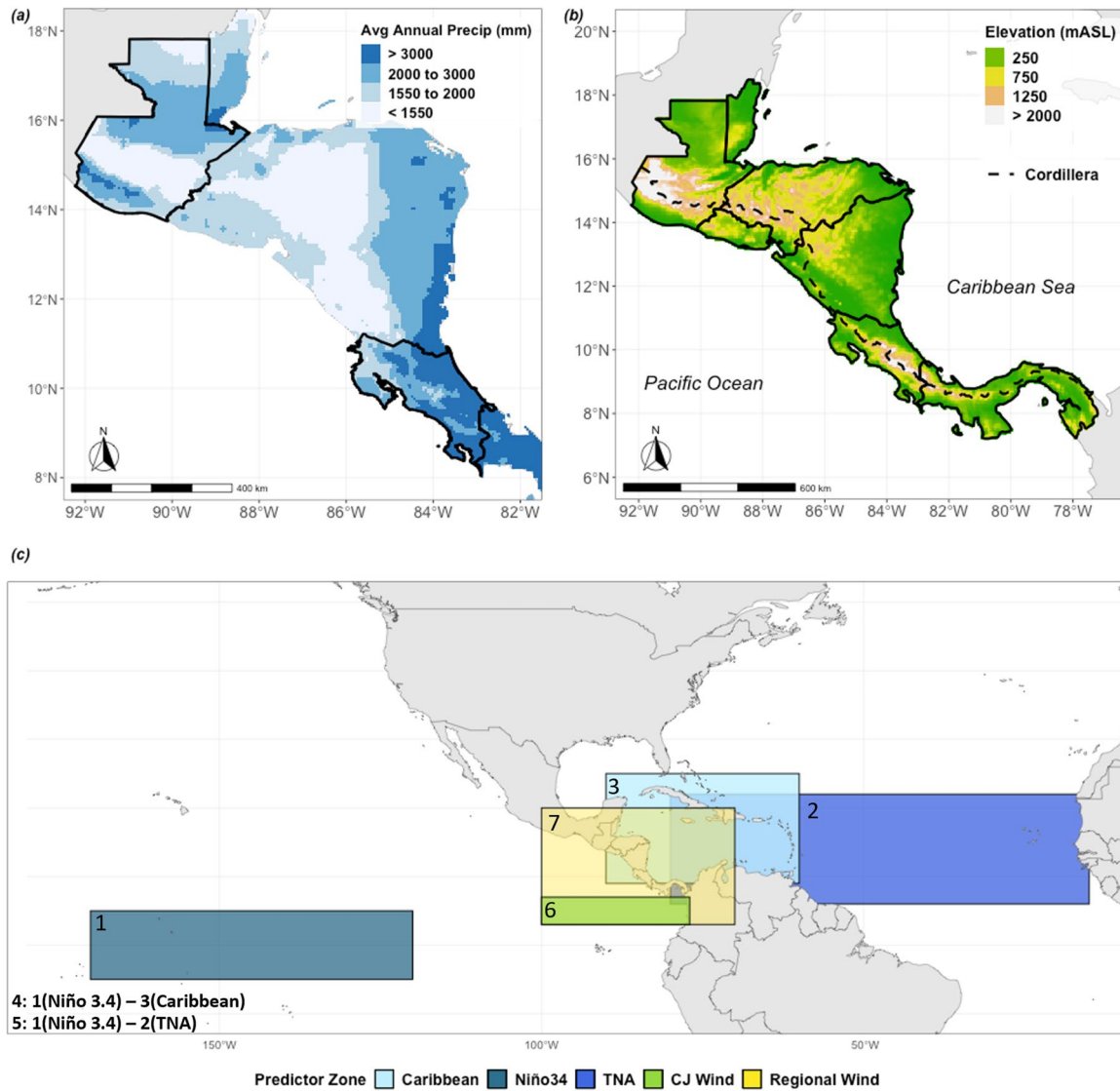


Figure 1. Illustration of Central American climate and tested predictor zones. (a) Total annual rainfall averaged over 1993–2016 using CHIRPS 0.25' (Funk et al., 2014) monthly rainfall data with Guatemala and Costa Rica outlined in black. (b) Elevation in Central America including the Cordillera. (c) Predictor locations used to test member performance plotted spatially (numbered as labeled in text with SST predictor zones plotted behind country outlines and wind predictor zones plotted in front). Zones include the Caribbean Sea (90W, 60W, 9N, 25N) which includes the core zone for moisture transport illustrated in Durán-Quesada et al. (2017), Niño 3.4 (170W, 120W, 5N, 5S Trenberth, 1997), and TNA (80W, 15W, 6N, 22N Enfield & Alfaro, 1999); and wind direction over the CJ (100W, 77W, 3N, 7N Poveda & Mesa, 1999) and a broader “Regional Wind” zone that circumscribes the eastern Pacific, core CJ and CLLJ locations (100W, 70W, 3N, 20N).

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 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 S1 in Supporting Information S1).

We assessed the usefulness of this subsampling technique on monthly rainfall forecasts in the month after initialization (labeled 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 1 August 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 preceding 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 et al., 2000; Magaña et al., 1999; Taylor & Alfaro, 2005). 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 (3.1) for the Caribbean and Pacific sides of the region, for instance, vary in part due to how the Cordillera mountain range (Figure 1b) interacts with moisture transport from the Caribbean Sea to the Pacific Ocean (Durán-Quesada et al., 2017; Muñoz-Jiménez et al., 2019; Taylor & Alfaro, 2005).

3.1. Relevant Predictor Zones for Central America

We evaluated seven predictor types with well-documented connections to regional rainfall (Figure 1c), including (a) Niño 3.4 SSTs (Giannini et al., 2000; Sánchez-Murillo et al., 2017; Waylen et al., 1994), (b) Tropical North Atlantic (TNA) SSTs (Enfield & Mayer, 1997; Maldonado et al., 2017), and (c) Caribbean SSTs (Durán-Quesada et al., 2017, 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: (d) the maximum anomaly difference between the Niño 3.4 region and the Caribbean (Wang, 2007), and (e) 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 (f) using CJ alone, and (g) 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 S2 in Supporting Information S1).

3.2. Member Performance and Top Performing Subsamples Selection

We processed all model data with the XCast Python package (Hall & Acharya, 2022; Text S3 in Supporting Information S1) and evaluated predictor estimates from the raw 130-member ensemble at 1° spatial resolution. We regridded the 130-member ensemble rainfall predictions to 0.25° resolution using bilinear interpolation prior to subsampling. The 130 estimates of each predictor were ranked (Figure 2, step 1) using the average mean squared error (MSE; Equation S2 in Text S1 in Supporting Information S1) across the predictor zone for SST (or calculating one MSE estimate for the gradients with no spatial field). The average Percent Correct Score (Equation S3 in Text S1 in Supporting Information S1) 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.

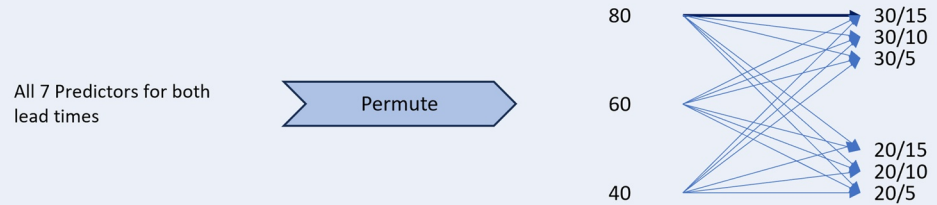
The “best” subsample is application dependent. Here we showcase a few subsamples that performed well on average over both countries using deterministic criteria over 1993–2016. We used multiple metrics to evaluate rainfall skill of the subsampled means against the all-member mean, including the Mean Squared Error Skill Score (MSE-SS—Deque, 2012, Equation S4 in Text S1 in Supporting Information S1) to assess performance across the entire rainfall distribution and the Heidke Skill Score (HSS - Heidke, 1926, Equation S5 in Text S1 in Supporting Information S1) to measure ensemble mean detection rate of low (high) rainfall extremes below the 10th (above the 90th) percentile of monthly anomalies.

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 2 step 2). Generating thousands of subsamples across a range of predictor permutations enabled us to compare subsamples that prioritized more well-rounded skill

Step 1: Calculate and rank member predictor scores for 1- and 2- month lead times

Average MSE in SST Zones	MSE of SST Gradients	Average Percent Correct of Wind Direction
Caribbean	Niño 3.4 / Caribbean	Choco Jet (CJ)
Niño 3.4	Niño 3.4 / TNA	Regional Wind
TNA		

Step 2: Subsample members based on their predictor scores



Step 3: Identify top subsample that exceeds all-member MSE-SS, HSS90, and HSS10 on average across country and select the subsample with the top HSS90 score in any month on average over hindcast years (1993-2016)

Figure 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 identify 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.

(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 2 step 3). We further explored the subsamples' sensitivity to filter size. Many filter sizes resulted in skillful subsamples (Text S4 in Supporting Information S1), and when filter size did matter, it often indicated the relative importance of different predictors (Figure S2 in Supporting Information S1).

This selection procedure identifies subsamples that on average maintain better skill across the analyzed years. Because of the interannual variability in ensemble forecast skill, we performed a sensitivity test to evaluate the robustness of the approach (Text S5 in Supporting Information S1). Although the skill of the all-member and subsampled ensembles changes interannually (Figure S3 in Supporting Information S1), 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 S4 in Supporting Information S1). 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 analyzed.

4. Results and Discussion

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 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 1 row 1 right column). We found no subsamples

Table 1

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

Predictand month	Filter step	Costa Rica		Guatemala	
		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	MSE: -0.24	SSTs in TNA	MSE: -0.24
		1-month lead (60)	[-0.11, -0.59]	1-month lead (10)	[-0.12, -0.77]
	2	CJ Wind	%Correct: 21%		
September	1	SSTs in Niño 3.4	MSE: -0.02	SSTs in Niño 3.4	MSE: -0.02
		1-month lead (80)	[0.01, -0.06]	1-month lead (80)	[0.01, -0.06]
	2	SSTs in Niño 3.4	MSE: 0.06	SSTs in Niño 3.4	MSE: 0.05
October	1	Regional Wind	%Correct: 3%	SST Gradient Niño3.4/Caribbean	MSE: -0.54
		1-month lead (15)	[-1%, 6%]	2-month lead (10)	[-0.20, -1.12]
	2	SST Gradient Niño3.4/Caribbean	MSE: 0.28	SST Gradient Niño3.4/TNA	MSE: 0.70
October	2	2-month lead (20)	[0.70, -0.12]	2-month lead (20)	[1.63, 0.19]
		3	SST Gradient Niño3.4/Caribbean	MSE: -0.10	CJ Wind
	3	1-month lead (15)	[0.41, -0.49]	2-month lead (15)	[-12%, 10%]

Note. 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.

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 S1 in Supporting Information S1). While our method may constrain error that occurs in a large ensemble, any subsampling procedure is also subject to uncertainties in observational data sets. The usefulness of the method depends on these data sets 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 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 1 row 1) when the Caribbean and Atlantic have stronger correlations with regional rainfall (Durán-Quesada et al., 2017; Maldonado et al., 2017; Spence et al., 2004; Taylor et al., 2002). SST Gradients improve subsamples across the wet season, as they are important throughout (Enfield & Alfaro, 1999; Giannini et al., 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 et al., 2016), but this jet primarily affects rainfall south of Guatemala (Durán-Quesada et al., 2010, 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).

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 S6 in Supporting Information S1), not necessarily linked to the total number of members per AOGCM (Table S2 in Supporting Information S1).

4.2. Subsamples Improve Skill for Entire Rainfall Distribution and the Extreme Tails

In all three months, the subsamples improve MSE-SS across most of Costa Rica and Guatemala (Figure 3a/3c), 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 MSE-SS, 62% locations increase ≥ 0.3 HSS90, 5% locations increase ≥ 0.3 HSS10). The skill of predicting rainfall extremes improves more in some locations (Figure 3b/3d) 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 3c/3d top row). Filtering out splintered members is useful in Costa Rica in September, when the top subsample significantly improves skill (Figure 3a/3b middle row) despite similar average predictor scores as the all-member mean (Table 1). These results are similar at the raw 1° spatial resolution (Figure S5 in Supporting Information S1).

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 3a/3b). Pacific rainfall patterns are distinct from Caribbean rainfall often due to interactions between regional rainfall drivers and topography (Durán-Quesada et al., 2017; Muñoz-Jiménez et al., 2019). The top subsamples also show limited skill above 16°N in Guatemala in September and October (Figure 3c/3d). 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 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).

4.3. Models Fail to Represent Full Rainfall Distribution When Process Error Increases

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 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 other regions at longer lead times (e.g., Caribbean at 2-month lead), but the slope is steeper, meaning small process errors can have a greater influence on the models' ability to predict rainfall (Figure 4).

Despite some structural differences among the models (e.g., DWD System 2.1 has a smaller rainfall range than the other models, Figure 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

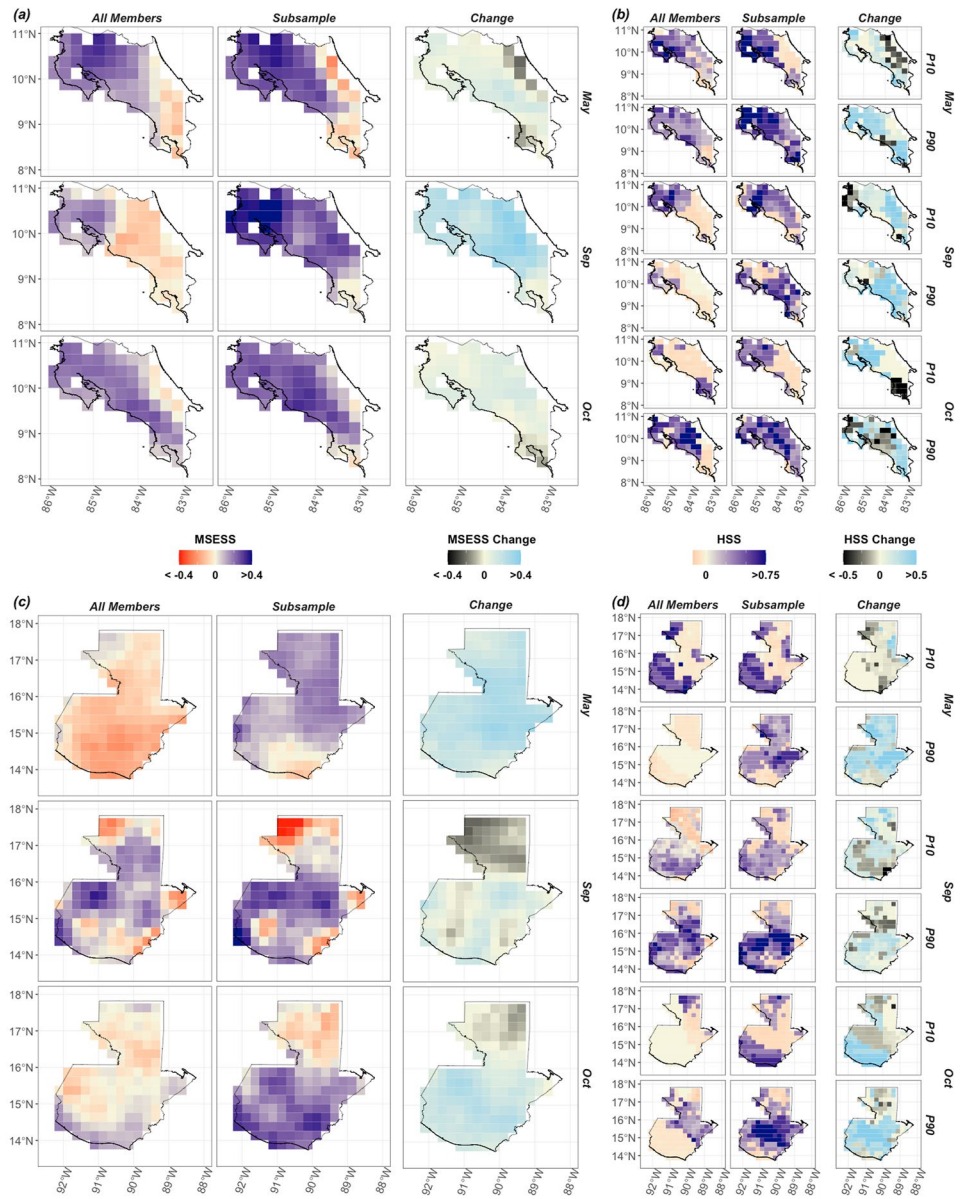


Figure 3. Spatial performance of the method in Costa Rica and Guatemala. (a) Spatial skill of all-member ensemble (left), top performing subsampled ensemble (center) and difference between the two (right) in Costa Rica. Skill is based on the entire rainfall distribution using MSE-SS. (b) Spatial skill 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). Panels (c) same as (a) but for Guatemala. Panel (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 2.

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 2.1), but by minimizing some key process errors, forecasters can increase the likelihood that a subsample better captures the entire rainfall distribution.

5. Conclusions

Subseasonal predictions present a gap between weather and seasonal climate forecasting that could benefit from optimization (Vitart & Robertson, 2018). Our subsampling method presents an opportunity when a forecast is

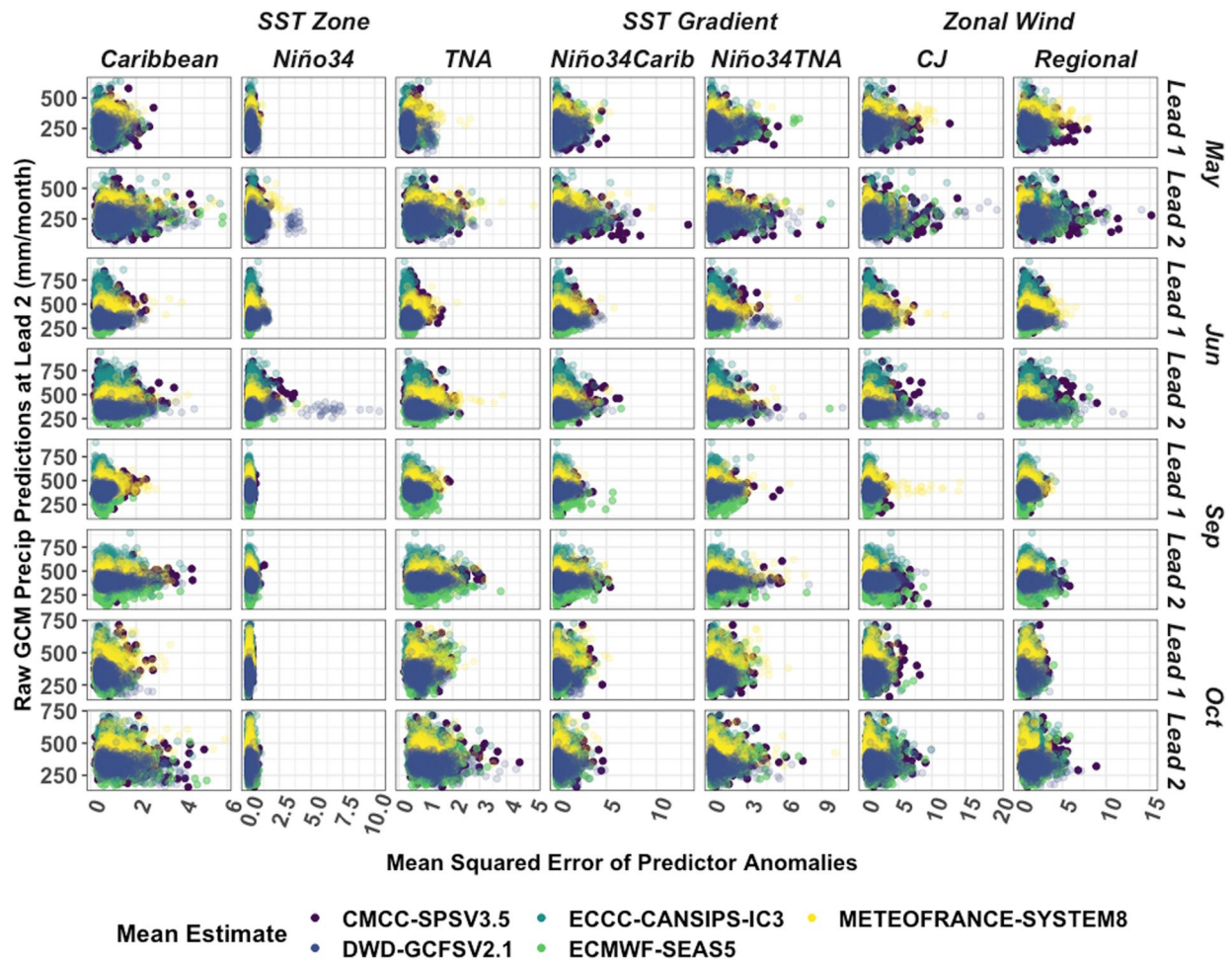


Figure 4. Relationship between the spatial average of the raw precipitation prediction at 2-month lead (y-axis) over 1993–2016 over each country and the MSE of the predictors (x-axis) for different zones/predictors (columns) grouped by lead time (one month ahead or two months ahead, rows) for different target months (rows), color-coded by the five models evaluated.

issued (initialized) 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 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 a key rainfall driver.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data is publicly available and free to use. C3S monthly hindcasts of total rainfall, SST, and zonal wind at 850 pressure can be accessed from the Climate Data Store (Copernicus Climate Change Service—CDS, 2018). ERA5 reanalysis data are also freely available at the Climate Data Store (Hersbach et al., 2018). CHIRPS data

(Funk et al., 2014) was accessed using the IRI Climate Data Library via the pyCPTv2 software package (Muñoz et al., 2019). OISSTv2 daily SST data (Huang et al., 2021) is hosted by the NOAA Physical Sciences Laboratory. All analysis was processed in XCast (Hall & Acharya, 2022) and visualized in R with ggplot2 and tidyverse (Wickham, 2016; Wickham et al., 2019) as described in Text S3 in Supporting Information S1. Code is available at Kowal (2023).

Acknowledgments

The authors would like to acknowledge Dr. Anne Van Loon for her support early in the research design process. This work was funded by the Rhodes Trust, Mortimer May Scholarship (Hertford College at Oxford University), UK Research and Innovation (UKRI; MR/V022008/1), and the University of Oxford Returning Carers' Fund.

References

- Alfaro, E. (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. *Ingeniería y Competitividad*, 9(2), 33–48. <https://doi.org/10.25100/iyc.v9i2.2486>
- 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. (2016). 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>
- 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. *Hydrology and Earth System Sciences*, 26(4), 975–999. <https://doi.org/10.5194/hess-26-975-2022>
- Balmaseda, M. A., Fujii, Y., Alves, O., Awaji, T., Behringer, D., Ferry, N., et al. (2011). Initialization for seasonal and decadal forecasts. In *Proceedings of Ocean Obs'09: Sustained ocean observations and information for society* (Vol. 2). ESA Publ. WPP-306. <https://doi.org/10.5270/OCEANOBS09.CWP.02>
- 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>
- 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>
- 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>
- 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>
- 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/10.1175/2009JCL3210.1>
- Copernicus Climate Change Service, Climate Data Store. (2018). Seasonal forecast monthly statistics on pressure levels [Dataset]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.0b79e7c5>
- De Perez, E. C., Van Den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., et al. (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>
- Deque, M. (2012). Deterministic forecasts of continuous variables. In I. Jolliffe & D. Stephenson (Eds.), *Forecast verification: A practitioner's guide in atmospheric science* (2nd ed., pp. 77–94).
- Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., et al. (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), e2021GL095063. <https://doi.org/10.1029/2021GL095063>
- Domeisen, D. I. V., White, C. J., Afargan-Gerstman, H., Muñoz, Á. G., Janiga, M. A., Vitart, F., et al. (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>
- 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., 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*, 115(5), D05103. <https://doi.org/10.1029/2009JD012455>
- 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>
- 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., & 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:TDCRO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2093:TDCRO>2.0.CO;2)
- 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: Oceans*, 102(1), 929–945. <https://doi.org/10.1029/96joc03296>
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., et al. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, 7(4), 813–830. <https://doi.org/10.5194/esd-7-813-2016>
- 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 & Pasture Science*, 69(6), 547–560. <https://doi.org/10.1071/CP18014>
- Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., et al. (2014). A quasi-global precipitation time series for drought monitoring [Dataset]. USGS <https://pubs.er.usgs.gov/publication/ds832> <https://iridl.ldeo.columbia.edu/SOURCES/UCSB/CHIRPS/>

- Giannini, A., Kushnir, Y., & Cane, M. A. (2000). Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. *Journal of Climate*, 13, 297–311. [https://doi.org/10.1175/1520-0442\(2000\)013<0297:ivocre>2.0.co;2](https://doi.org/10.1175/1520-0442(2000)013<0297:ivocre>2.0.co;2)
- 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.revistayam.com/volumen-2/numero-3/notas-de-divulgacion-cientifica/the-central-american-dry-corridor-a-consensus-statement-and-its-background/>
- 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 [Software]. *Frontiers in Climate*, 4, 953262. <https://doi.org/10.3389/FCLIM.2022.953262/BIBTEX>
- Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler*, 8, 301. <https://doi.org/10.2307/519729>
- Hemri, S., Bhend, J., Liniger, M. A., Manzanar, R., Siebert, S., David, S., et al. (2020). How to create an operational multi-model of seasonal forecasts? *Climate Dynamics*, 55(5–6), 1141–1157. <https://doi.org/10.1007/s00382-020-05314-2>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2018). ERA5 monthly averaged data on single levels from 1940 to present [Dataset]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.f170507>
- Hidalgo, H., & Alfaro, E. (2012). Global model selection for evaluation of climate change projections in the eastern tropical Pacific seascape. *Revista de Biología Tropical*, 60(S3), 67–81. <https://doi.org/10.15517/RBT.V60I3.28340>
- Hidalgo, H., 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. (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>
- Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2021). Improvements of the daily optimum interpolation sea surface temperature (OISST) Version 2.1 [Dataset]. *Journal of Climate*, 34(8), 2923–2939. <https://doi.org/10.1175/JCLI-D-20-0166.1>
- Kowal, K. (2023). Process-informed subsampling example (v0.0.0-beta). *Zenodo*. <https://doi.org/10.5281/zenodo.10063969>
- 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 skillfully predicts late wet-season precipitation in Central American Dry Corridor excelling in Costa Rica and Nicaragua. *International Journal of Climatology*, 42(9), 4953–4971. <https://doi.org/10.1002/JOC.7514>
- 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](https://doi.org/10.1175/1520-0442(1999)012<1577:TMDOMA>2.0.CO;2)
- Maldonado, T., Alfaro, E., & Hidalgo, H. (2018). Revision of the main drivers and variability of Central America Climate and seasonal forecast systems. *Revista de Biología Tropical*, 66(1), S153–S175. <https://doi.org/10.15517/rbt.v66i1.33294>
- 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 Biología Tropical*, 66(May), S231–S254. <https://doi.org/10.15517/rbt.v66i1.33303>
- Manzanar, R., Gutiérrez, J. M., Bhend, J., Hemri, S., Doblas-Reyes, F. J., Torralba, V., et al. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Climate Dynamics*, 53(3–4), 1287–1305. <https://doi.org/10.1007/s00382-019-04640-4>
- Marsh, K., & Penebad, E. (2016). Description of the C3S seasonal multi-system - Copernicus knowledge base - ECMWF confluence wiki. Retrieved from <https://confluence.ecmwf.int/display/CKB/Description+of+the+C3S+seasonal+multi-system>
- Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., et al. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, 58(4), e2020RG000704. <https://doi.org/10.1029/2020RG000704>
- 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/FIGURES/11>
- Muñoz, A. G., Robertson, A. W., Mason, S. J., Goddard, L., Pons, D., Chourio, X., et al. (2019). NextGen: A next-generation system for calibrating, ensembling and verifying regional seasonal and subseasonal forecasts [Software]. AGUFM, A23U-3024. <https://ui.adsabs.harvard.edu/abs/2019AGUFM.A23U3024M/abstract> <https://iri-pycpt.github.io/>
- 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>
- Nowack, P., Runge, J., Eyring, V., & Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. *Nature Communications*, 11(1), 1–11. <https://doi.org/10.1038/s41467-020-15195-y>
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Décluse, P., et al. (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>
- Poveda, G., & Mesa, O. J. (1999). La Corriente de chorro superficial del oeste “del Chocó”: Climatología y Variabilidad durante las fases del ENSO. In *Proc. V Congreso Colombiano de Meteorología, January*. Retrieved from <https://www.researchgate.net/publication/233857360>
- 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>
- Scaife, A. A., Ferranti, L., Alves, O., Athanasiadis, P., Baehr, J., Dequé, M., et al. (2019). Tropical rainfall predictions from multiple seasonal forecast systems. *International Journal of Climatology*, 39(2), 974–988. <https://doi.org/10.1002/joc.5855>
- Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North Atlantic climate far more predictable than models imply Check for updates. *Nature*, 583(7818), 796–800. <https://doi.org/10.1038/s41586-020-2525-0>
- 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>
- Stockdale, T., Alves, O., Boer, G., Deque, M., Ding, Y., Kumar, A., et al. (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>

- 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., 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*, 107(C9), 10–11. <https://doi.org/10.1029/2001JC001097>
- Trenberth, K. E. (1997). The definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12), 2771–2777. [https://doi.org/10.1175/1520-0477\(1997\)078<2771:TDOENO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2)
- 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*, 1(1), 1–7. <https://doi.org/10.1038/s41612-018-0013-0>
- 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*, 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., et al. (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. (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>
- 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>
- 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>
- White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., et al. (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* [Software]. Springer-Verlag. <https://doi.org/10.1007/978-3-319-24277-4>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse [Software]. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

References From the Supporting Information

- Batté, L., Dorel, L., Ardilouze, C., & Guérémy, J.-F. (2021). Documentation of the METEO-FRANCE seasonal forecasting system 8. Retrieved from <http://www.umr-cnrm.fr/IMG/pdf/system8-technical.pdf>
- 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., & Palmer, T. (1998). Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126(9), 2503–2518. [https://doi.org/10.1175/1520-0493\(1998\)126<2503:ioesoe>2.0.co;2](https://doi.org/10.1175/1520-0493(1998)126<2503:ioesoe>2.0.co;2)
- Buizza, R., Petrolagiis, T., Palmer, T., Barkmeijer, J., Hamrud, M., Hollingsworth, A., et al. (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>
- Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., et al. (2021). The German climate forecast system: GCFS. *Journal of Advances in Modeling Earth Systems*, 13(2), e2020MS002101. <https://doi.org/10.1029/2020MS002101>
- Gualdi, S., Borrelli, A., Davoli, G., Masina, S., Navarra, A., Sanna, A., et al. (2020). The new CMCC operational seasonal prediction system issue TN0288 CMCC technical notes. <https://doi.org/10.25424/CMCC/SPS3.5>
- Huang, T.-Y., & Zhao, B. (2020). tidyfst: Tidy verbs for fast data manipulation [Software]. *Journal of Open Source Software*, 5(52), 2388. <https://doi.org/10.21105/joss.02388>
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2018). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development Discussions*, 1–44. <https://doi.org/10.5194/gmd-2018-228>
- 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](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., et al. (2021). The Canadian seasonal to interannual prediction system version 2.1. (CanSIPsv2.1).