RESEARCH ARTICLE

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

Katherine M. Kowal[1] | Louise J. Slater[1] | Alan García López[2] | Anne F. Van Loon[3]

[1]Department of Geography and the Environment, University of Oxford, Oxford, UK

[2]Sección de Aplicaciones Climáticas en el Departamento de Investigación y Servicios Meteorológicos, Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología (INSIVUMEH), Guatemala City, Guatemala

[3]Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Katherine M. Kowal, Department of Geography and the Environment, University of Oxford, Oxford, UK.
Email: katherine.kowal@ouce.ox.ac.uk

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 multimodel 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 Niño3.4 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.

1 | INTRODUCTION

Central America has been identified as a climate change hotspot at high risk of hydrometeorological extremes (Giorgi, 2006; Taylor *et al.*, 2012; Almazroui *et al.*, 2021). 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 (Hidalgo *et al.*, 2013; Hannah *et al.*, 2017; Imbach *et al.*, 2018; Almazroui *et al.*, 2021), 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 *et al.*, 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; de Perez *et al.*, 2016; Kreibich *et al.*, 2017; Golding *et al.*, 2019; Goddard *et al.*, 2020; Domeisen *et al.*, 2022; 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 (Donoso and Ramirez, 2001; Garcia-Solera and Ramirez, 2012; Maldonado *et al.*, 2013; Alfaro *et al.*, 2016b). 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, 2022b).

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 (Amador and Alfaro, 2009;

Hao *et al.*, 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 *et al.*, 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 SST fields play in rainfall over Central America (Giannini *et al.*, 2000; Alfaro, 2007; Maldonado *et al.*, 2013; 2016; 2017; Alfaro *et al.*, 2016a), including for predictions of low- and high-rainfall events (e.g., rainfall above the 80th and below the 10th percentiles of the monthly climatology; Maldonado *et al.*, 2013). Hybrid methods then combine dynamic 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 (e.g., Alfaro *et al.*, 2018; Strazzo *et al.*, 2019; Colman *et al.*, 2020). Indirect forecasts could perform well, as forecasts of large-scale variability like SST over tropical domains are often a strength of the GCMs 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 high anomaly correlations with SST in the Niño3.4 region ($\sim$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 their direct rainfall forecast skill.

In this paper we compare the following methods: dynamic methods that take rainfall forecasts directly from GCMs (direct), statistical methods based on observed SST values months ahead of a season
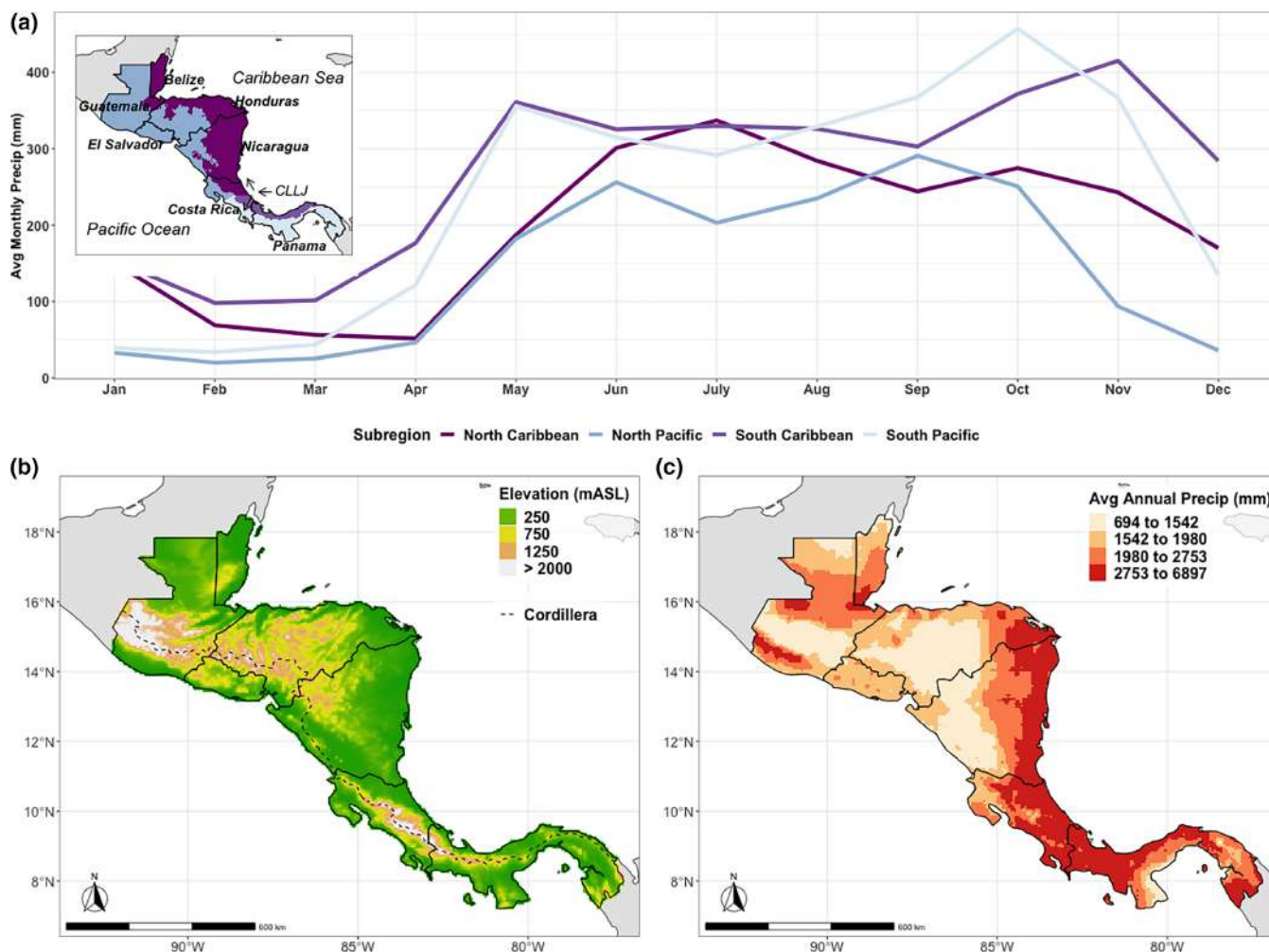
FIGURE 1 (a) Spatially averaged monthly rainfall between 1993 and 2016 (using CHIRPS) plotted by subregions, which are tagged in the inset map. The southern part of the isthmus is separated below 10°N when the CLLJ branches (Wang, 2007; Muñoz et al., 2008; Cook and Vizy, 2010; Hidalgo et al., 2015) and 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, 2022c), 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 to 2016 plotted spatially. [Colour figure can be viewed at wileyonlinelibrary.com]

(statistical), and hybrid methods that generate forecasts indirectly by using 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 forecast rainfall well over a given region (Hagedorn et al., 2005; Amador and Alfaro, 2009; Hidalgo and Alfaro, 2012; 2015; Doblas-Reyes et al., 2013; Kharin et al., 2017; Almazroui et al., 2021). 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; Khaje-hei et al., 2018; Slater et al., 2019; Becker et al., 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 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 models and methods 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 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 1a; Magaña et al., 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 et al., 1999). The seasonal distribution in rainfall also varies regionally with more extreme rainfall events often occurring on the Pacific slope during August–October compared to the Caribbean slope (Peterson et al., 2002; Taylor and Alfaro, 2005; Maldonado et al., 2018). The variability and intensity of seasonal rainfall is not uniform across the isthmus (e.g., Muñoz-Jiménez et al., 2019) (see Figure 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) compared to other locations (Figure 1a).

Central America is an important region for a regional forecast evaluation, as it 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 1b). One tropical teleconnection is the El Niño–Southern Oscillation (ENSO) in the eastern Pacific (Waylen et al., 1994; Spence et al., 2004; Amador et al., 2006; Poveda et al., 2006; Amador, 2008; Durán-Quesada et al., 2017; Maldonado et al., 2018; Mariotti et al., 2020). 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 et al., 2004; Durán-Quesada et al., 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 (Enfield and Alfaro, 1999; Poveda and Mesa, 1999; Wang

and Enfield, 2001; Amador et al., 2006; Wang, 2007; Amador, 2008; Hidalgo et al., 2015; Sori et al., 2015; Durán-Quesada et al., 2017; 2020).

Evaluating whether the GCMs can improve on an ENSO-based forecast is useful. ENSO is not only a known driver of regional rainfall but is also a known driver of skill of many GCMs (Yuan and Wood, 2013; Mo and Lyon, 2015; Scaife et al., 2019; Gubler et al., 2020), 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 European Centre for Medium-Range Weather Forecasting (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 (Enfield and Alfaro, 1999; Giannini et al., 2000; Waylen and Quesada, 2001; Taylor et al., 2002; Spence et al., 2004; Alfaro, 2007; Wang et al., 2008; Sori et al., 2015), as it modulates the magnitude of the mid-summer dry period (Wang et al., 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., 2016b). 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 (a) evaluating how the skill of direct, indirect, and statistical forecasts varies spatially and temporally during key wet season periods, (b) comparing the relative value of using the C3S and

TABLE 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[a]	20	Lin et al. (2021) and 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 Modelling and Assimilation Office; NOAA-NCEP, NOAA National Center for Environmental Prediction; NOAA-GFDL, NOAA Geophysical Fluid Dynamics Laboratory; CMCC, Fondazione Centro Euro-Mediterraneo Sui Cambiamenti Climatici; DWD, Deutscher Wetterdienst; ECMWF, European Centre for Medium-Range Weather Forecasts; UKMO UK Met Office.
[a]Cansips-IC3 is operationally treated as one but technically made up of the GEM5-NEMO and CanCM4i-IC3 models, each containing 10 members.

NMME ensembles for direct or indirect forecasts to identify cases when they are significantly different from each other and a purely statistical forecast, and (c) 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.

2 | DATA AND METHODS

2.1 | Seasonal forecast scope

We focus on seasonal rainfall forecasts during key wet season months (May–October) across Central America, using 3-month seasonal totals from the forecasting models initialized 1 month prior to the forecast, as is done operationally for rainfall outlooks produced by the CA-COF (Donoso and Ramirez, 2001; Garcia-Solera and Ramirez, 2012; Alfaro et al., 2016b). 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 3-month periods. For instance, a forecast of total seasonal rainfall from MJJ is initialized in April, essentially adding rainfall from a 1-month lead time for May, a 2-month lead time for June, and a 3-month lead time for July. Tercile deterministic forecasts, like the ones produced by the 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, comparing performance between the North Pacific, North Caribbean, South Pacific, and South Caribbean (as defined in Figure 1a).

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

timescales 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 timescales (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 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 (Becker et al., 2020; NOAA, 2020). Five C3S models are selected based on publicly available spatial resolution (1°) and consistent initialization (1st of every month), as summarized in Table 1. Here, C3S and NMME are compared using the overlap between their publicly available data, which includes monthly temporal resolution and hindcasts from 1993 to 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., 2016a; Hidalgo et al., 2017), covers the entire time period of analysis (1993–2016), and is available at high spatial resolution ∼0.05° (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 (Figure S1, Supporting Information), and the forecast skill shows similar regional and seasonal variability across observational datasets, so we prefer CHIRPS (Figure S2). 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.

2.3 | Forecast construction

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 are part of the NextGen methodology developed by IRI to improve climate services at subseasonal to seasonal timescales (Muñoz et al., 2019; WMO, 2020; Muñoz et al., 2020a). Here we provide a brief overview of the methodology, but other articles provide more details on the functionality (e.g., Muñoz et al., 2020a; Pons et al., 2021) and applications (e.g., Fernandes et al., 2020; 2022; WMO, 2020) of NextGen. Steps performed in pyCPT include (a) download raw 1° hindcasts of C3S and NMME GCMs from the IRI data library and (b) calibrate the hindcasts and transform them into 0.25° rainfall predictions over Central America using a Canonical Correlation Analysis (CCA) model (see Figure 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 Baddour (2008), which is designed to maximize the correlation of linear combinations between predictors and predictands, summarized as follows: (a) *correct biases* by applying integrated cosine latitude weighting to both predictor and predictands and transform them into normalized outputs; (b) *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); (c) *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; (d) *fit a CCA model* onto the two PCA time series—one for predictor and one for predictand; (e) *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 Muñoz et al., 2020b; Martinez et al., 2022).

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 (63°N–10°S, 170°–15°W), which is selected based on the
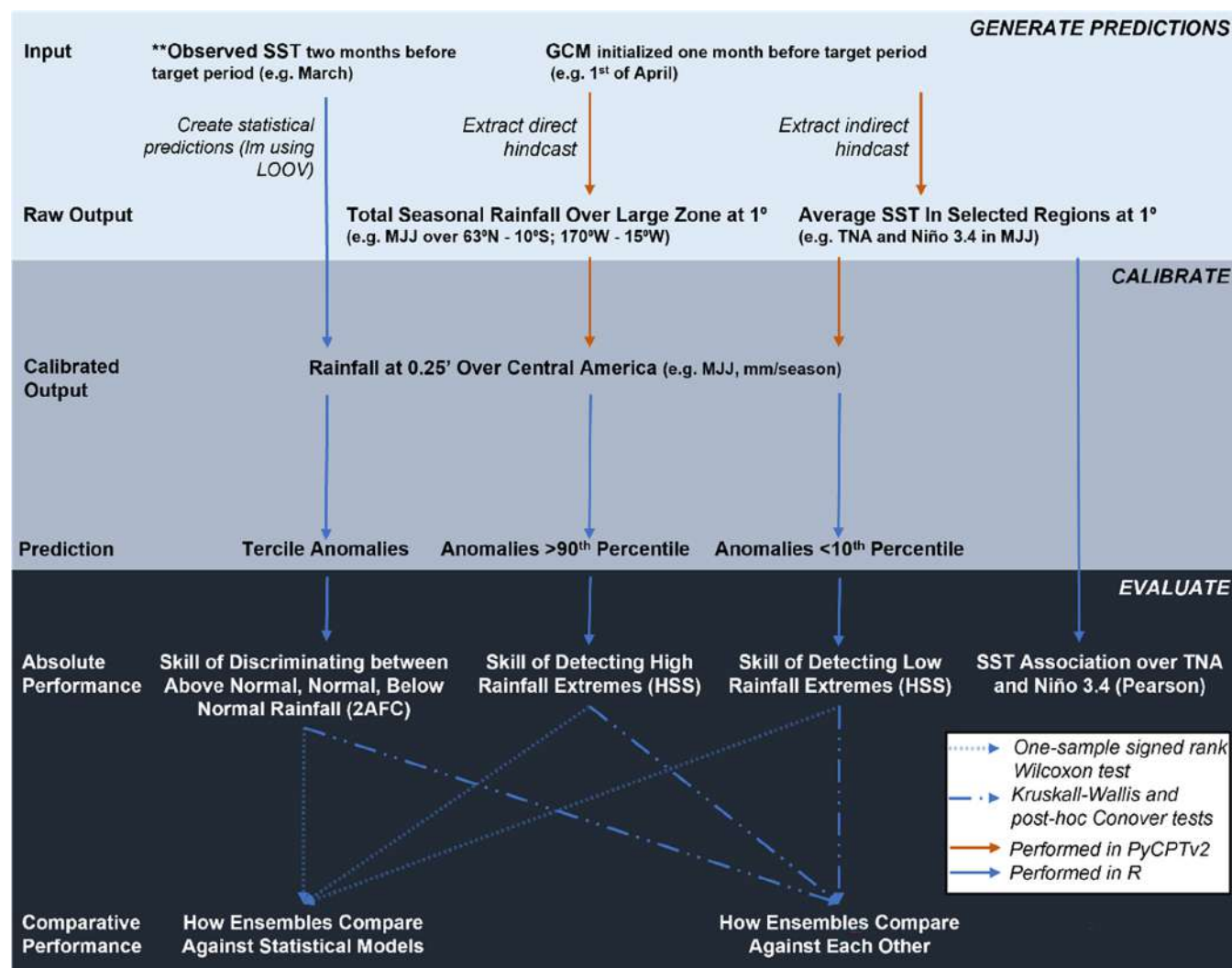
FIGURE 2 Diagram of methods employed to generate forecasts, calibrate them, and evaluate their performance. Statistical models: A linear model (lm) is generated using leave-one-out cross-validation (LOOV) from 1981 to 2016 to train the models to make predictions over 1993–2016 based on 2-month lagged 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. [Colour figure can be viewed at wileyonlinelibrary.com]

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 and Wickham, 2011; NCAR – Research Applications Laboratory, 2015; Wickham, 2016; Weigel, 2017; Pebesma, 2018; Pierce, 2019; Wickham *et al.*, 2019, 2021; Huang and Zhao, 2020; Kassambara, 2020; Wickham and Seidel, 2020; Hijmans, 2021; Mangiafico, 2022; Pedersen, 2022; Pohlert, 2022). 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 (Weigel *et al.*, 2010; Delsole *et al.*, 2013).

2.3.2 | Indirect and statistical forecasts in known teleconnection zones

We generate indirect and statistical forecasts using two known teleconnection regions—Niño3.4 (170°–120°W, 5°N–5°S) for the ENSO teleconnection (Trenberth, 1997; Trenberth and Stepaniak, 2001) and Tropical North Atlantic (TNA; 55°–15°W, 5°–25°N) 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 (Spence *et al.*, 2004; Alfaro, 2007; Alfaro *et al.*, 2016a; Maldonado *et al.*, 2017), 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 Niño3.4 or the TNA region and then the same CCA method is used to transform the raw 1° GCM forecasts to 0.25° and statistically relate the SST predictors from a selected zone (e.g., Niño3.4) 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 Niño3.4 region, the TNA region, and using both regions as predictor variables. This process is done for each season individually over the entire overlapping period of available CHIRPS/OISSTv2 data through 2016 (1982–2016), giving the statistical models over 30 years for training. The predictions are created using a cross-validation leave-one-out-approach in R (Kuhn, 2022) as described in Mason and Baddour (2008). For example, to predict rainfall for 1993 with the Niño3.4 statistical model, observed rainfall estimates in MJJ are regressed against spatially averaged observed March Niño3.4 SST across 1982–2016 (without 1993), setting 1993 as the prediction year. We select observed SST data 2 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.

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; Heidke, 1926) for low- and high-rainfall detection (see Table 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 and Mason, 2011). HSS is preferred for extreme event detection because the metric is equitable (Hogan *et al.*, 2010), relatively common (e.g., Higgins *et al.*, 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 are 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 and 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 if all the statistical models perform worse than 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, that is, C3S vs. NMME, we use a Kruskal–Wallis test to assess if

TABLE 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 and 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 $n_{0,i}$ = ith forecast issued, event did not occur $n_{1,i}$ = jth 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 (Wilcoxon, 1945; Mann and Whitney, 1947)	Test statistic $(W_1) = \sum R_d^+$ $\mu W_1 = \frac{N_r(N_r+1)}{4}$ $\sigma W_1 = \sqrt{\frac{N_r(N_r+1)(2N_r+1)}{24}}$ $z\,\text{statistic} = \frac{W_1 - \mu W_1}{\sigma W_1}$ 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)
Kruskall–Wallis test	Nonparametric comparison of variance to test if multiple samples come from the same distribution (Kruskal and Wallis, 1952)	$H\ \text{statistic} = \frac{12}{N(N+1)}\sum \frac{R_i^2}{n_i} - 3(N+1)$ 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-square value < H) Computed with "stats" package in base R (R Core Team, 2021)
Conover-Iman post hoc test	Pairwise comparisons following rejection of null hypothesis in Kruskall–Wallis test	

TABLE 2 (Continued)

Metric	Summary	Formula
	(Conover and Iman, 1979; Conover, 1999)	Conover test statistic $(t) = |\bar{R}_i - \bar{R}_j| > t_{N-k}\sqrt{s^2\left[\frac{N-1-\hat{H}^*}{N-k}\right]\left[\frac{1}{n_i}+\frac{1}{n_j}\right]}$ $\hat{H}^*$ (tiecorrected statstic of KW test) $= \dfrac{\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3(N+1)}{1-\frac{\sum_{i=1}^{r}(t_i^3-t_i)}{N^3-N}}$ $s^2 = \frac{1}{N-1}\left[\sum R_i^2 - N\frac{(N+1)^2}{4}\right]$ Null hypothesis (H_0): Population medians are equal Alternative hypothesis (H_1): Population medians are not equal k = number of groups N = total number of samples n_i = sample number in the ith group R_i = total sum of ranks in ith group Computed with "PMCMRplus" and "rcompanion" R packages (Mangiafico, 2022; Pohlert, 2022)

there are any significant differences between the skill of the spread of the ensembles for a given seasonal period and subregion. Then we 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 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 Niño3.4 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.

3 | RESULTS

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 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 3). Combining the models into ensembles also 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 4). Looking across forecasting methods (e.g., direct vs. Niño3.4 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 (Niño3.4 indirect and direct) over eastern Nicaragua in MJJ, and the Niño3.4 indirect forecasts have highest skill in Guatemala during this period (Figure 4, middle group). While the direct ensemble skill tends to be highest in most other cases (Figure 4, left group), the Niño3.4 indirect forecasts also have similar skill in the late wet season (ASO; Figure 4, bottom row).
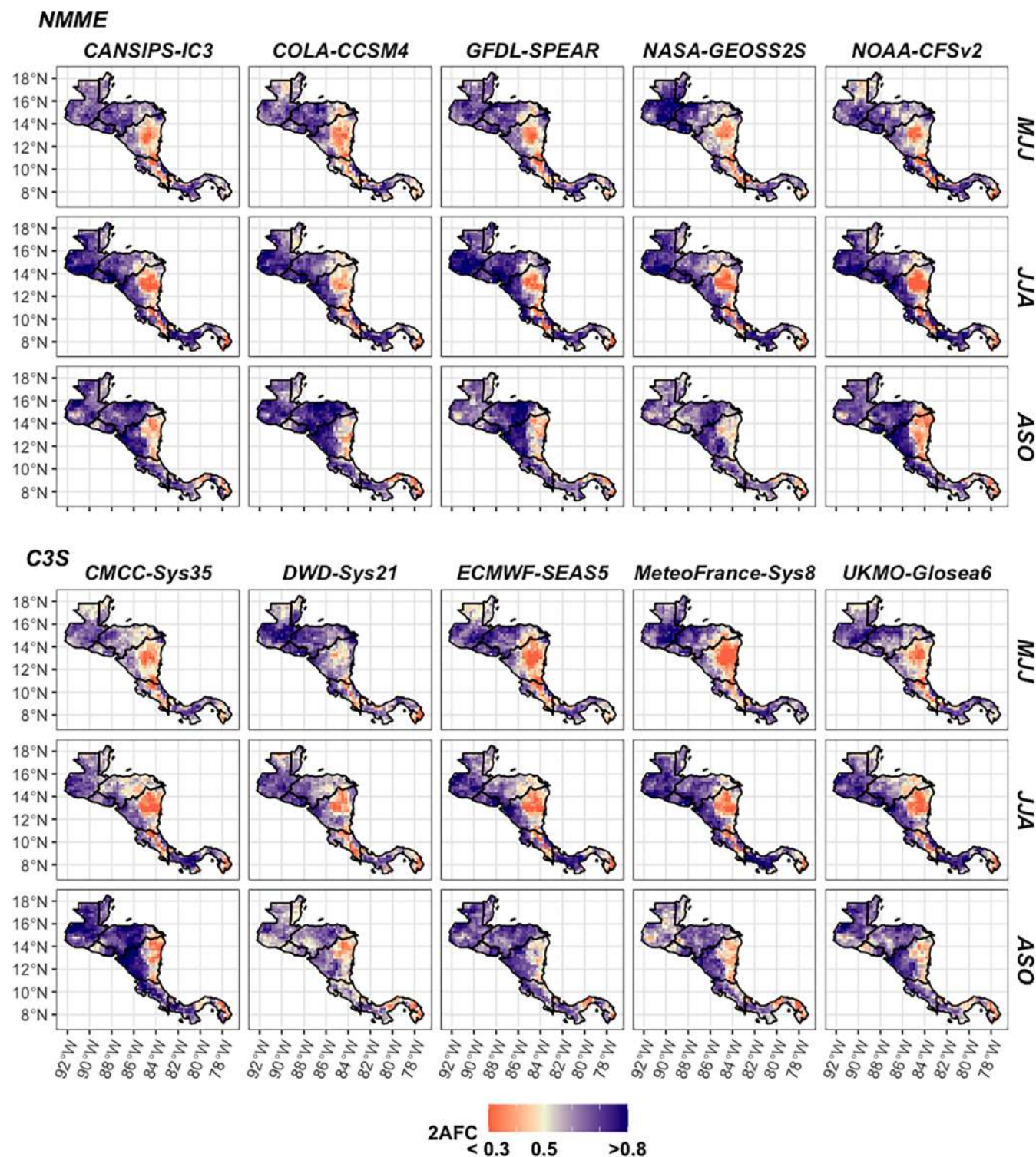
FIGURE 3 Rainfall skill for direct tercile forecasts using 2AFC across individual models in the wet season. 0.5 is equal to the skill of using a random forecast. [Colour figure can be viewed at wileyonlinelibrary.com]

The visual assessments of absolute skill (Figures 3 and 4) are confirmed when statistically testing the differences between the spatially averaged skill by subregion (Figure 5a) and tagging the top performer by location for each seasonal period (Figure 5b). On average, the North

Pacific is the highest skill zone with 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
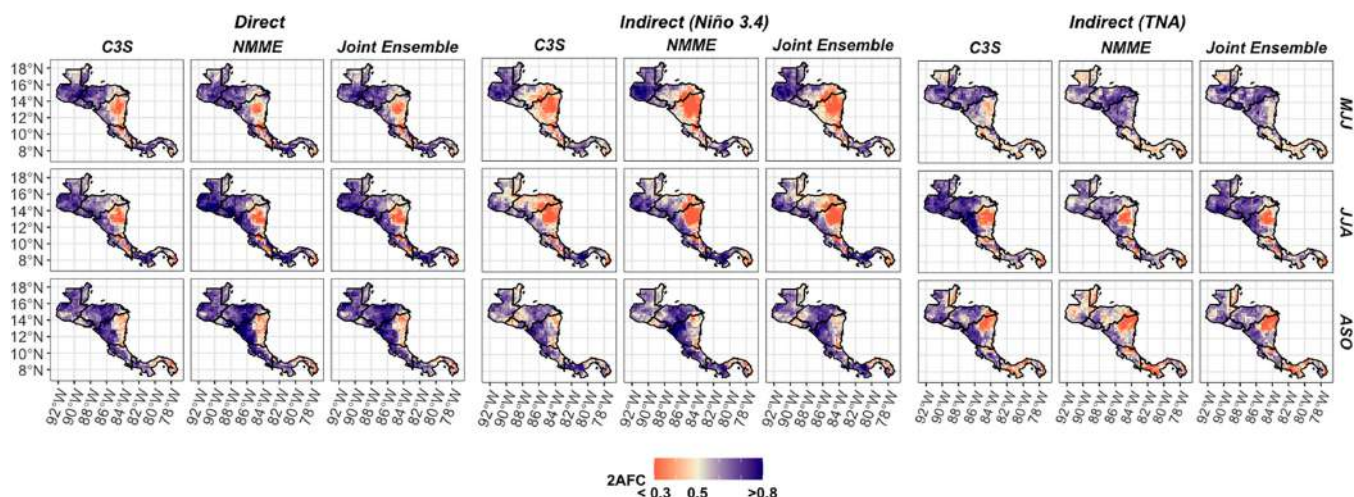
FIGURE 4 Rainfall skill of the ensemble means using C3S, NMME, and the Joint Ensemble. All models in the study are compared using their direct and hybrid forecasts with 2AFC skill. [Colour figure can be viewed at wileyonlinelibrary.com]

all the subregions in JJA (Figure 5a, top right panel), which is mirrored in the absolute skill plots of the direct forecasts (Figure 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 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 5a, top right panel and Figure 5b, left panel). Although the spatially averaged skill of the Niño3.4 indirect forecasts are not best in the North Pacific generally in the early wet season (MJJ) (Figure 5a, top left panel), the Niño3.4 indirect forecast still demonstrates more targeted value over Guatemala as the top performer in MJJ (Figure 5b, left panel). The direct and Niño3.4 indirect forecasts also tend to have the highest skill in the South Pacific and South Caribbean, often showing similar skill in ASO (Figure 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., Niño3.4 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 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 Niño3.4 SST in ASO (Figure 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.

3.2 | Skill of high- and low-rainfall extremes

Ensemble skill is more limited when the models are used to predict extreme rainfall (Figure 6 and see Figure S3 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 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 6), but direct forecast skill is still high in much of the North Pacific during the early to middle wet season (MJJ/JJA; Figure 6, left group). When comparing the absolute skill of direct and indirect extreme rainfall forecasts, 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 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 6, bottom right group), a period when direct forecasts often performed better for tercile rainfall discrimination (Figure 5a).

The skill of the ensemble means is similar within forecasting methods (Figure 6), and the differences between the two ensembles are often not significant
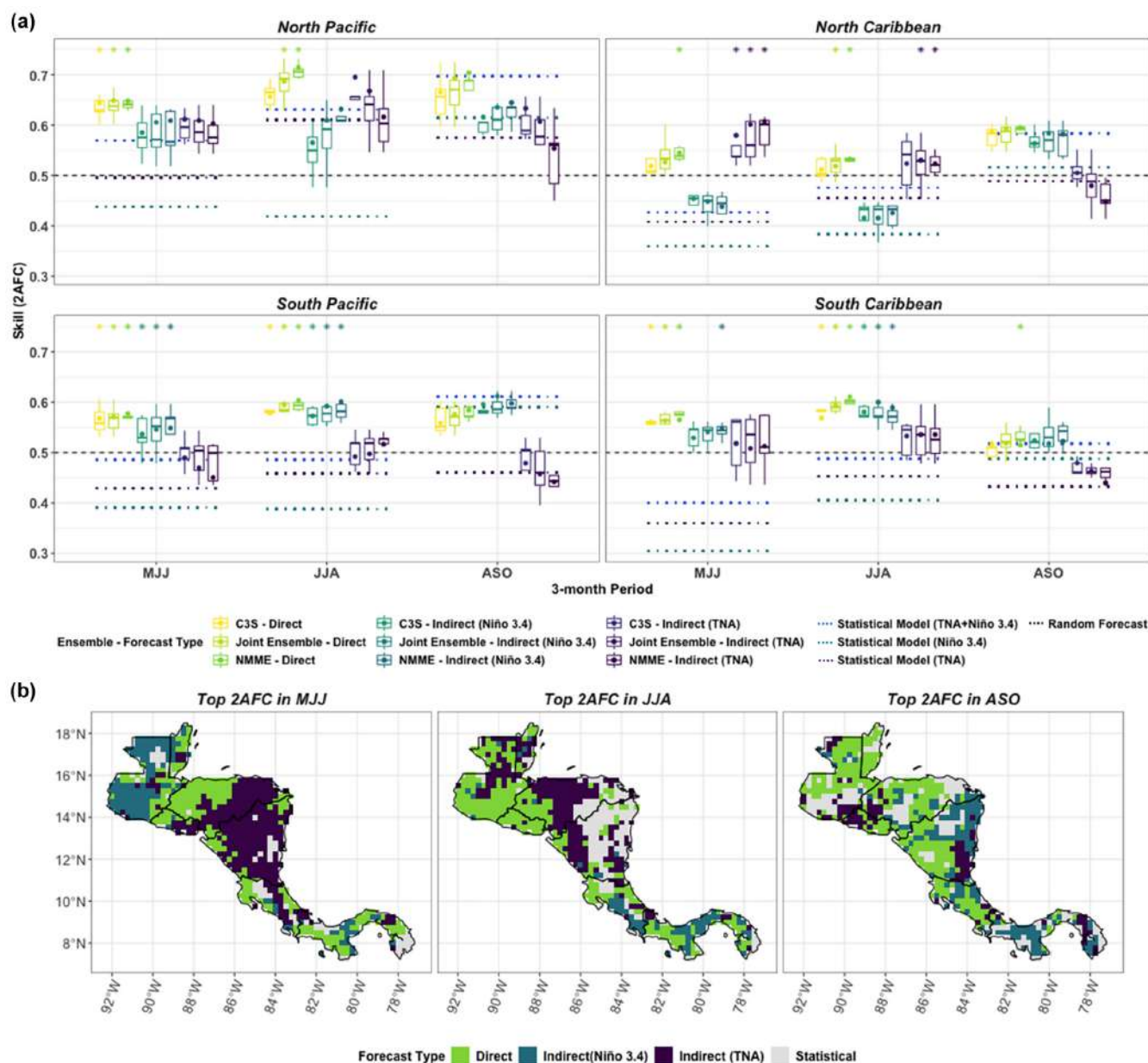
FIGURE 5 (a) Spatially averaged comparison of forecast skill using 2AFC across seasons and grouped by subregion as defined in Figure 1a. Boxplots are catogorized by ensemble and show the spread of the individual model means contributing to each ensemble. Dots within boxplots show the skill of the ensemble 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. [Colour figure can be viewed at wileyonlinelibrary.com]

when they are both top performers (Figure 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 7a),
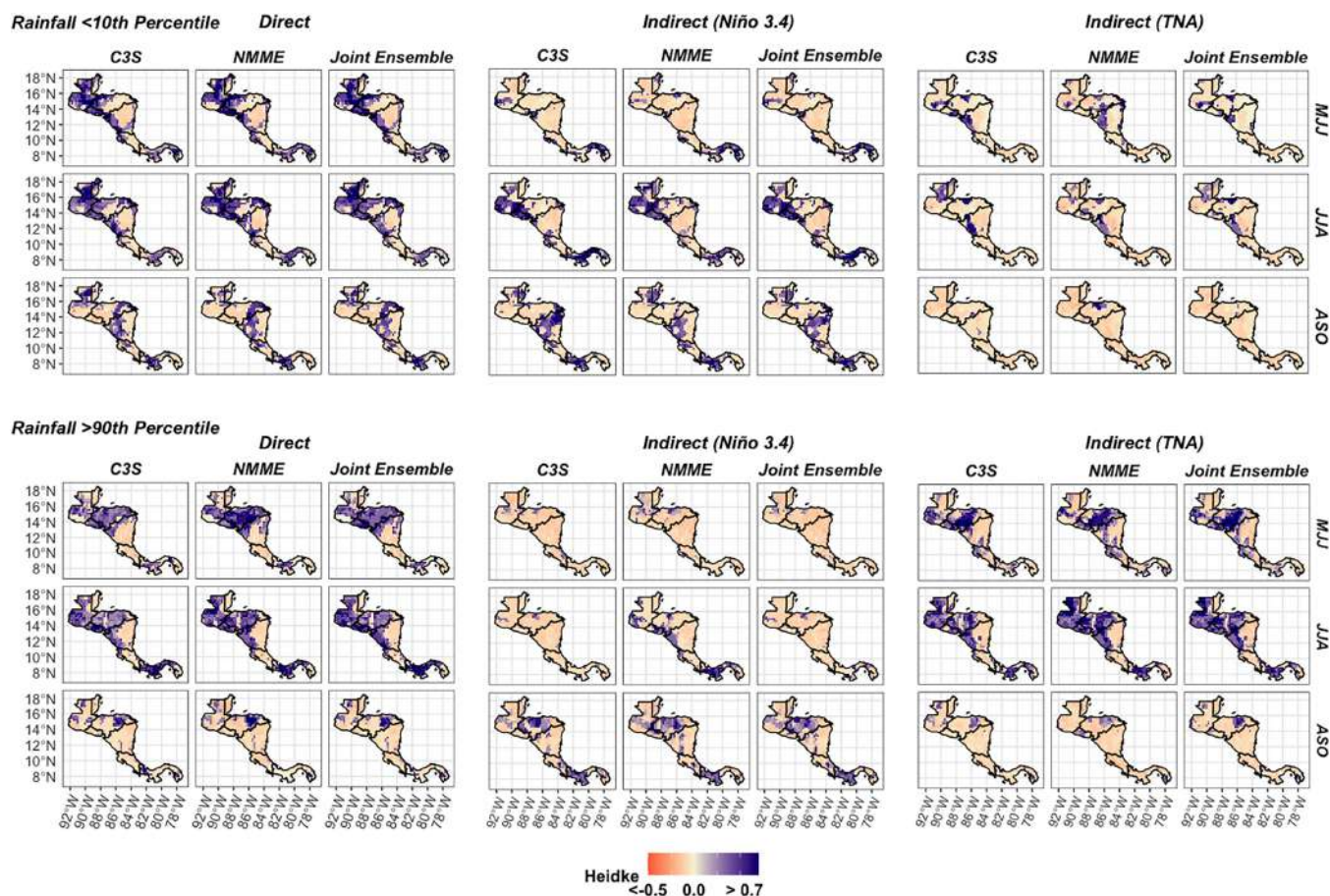
FIGURE 6 Ensemble mean skill for high- and low-rainfall extremes using HSS, where 0.0 is equal to the skill of a random forecast. Skill is plotted over separate parts of the wet season—MJJ, JJA, and ASO. [Colour figure can be viewed at wileyonlinelibrary.com]

but not as consistently as when the ensemble mean is used to discriminate between terciles (Figure 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 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 (e.g., MJJ in North Pacific; Figure 7a, left panels), and the interquartile spread of C3S is often larger than NMME, meaning some C3S contributors drag down the score of the C3S mean while others perform much better (Figure 7a).

When compared to the statistical models, the direct and Niño3.4 indirect forecasts are often significantly better at 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 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 7a); and the top performing forecast type has more spatial consistency for low-rainfall extremes (Figure 7b, e.g., direct forecasts over Guatemala in MJJ).

3.3 | Predictive skill of key teleconnection regions: ENSO and TNA

The skill of the ensembles' predictions of SST over the Niño3.4 region is consistently high. The association between each model and the observed Niño3.4 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 Niño3.4 (Figure 8a). Differences in SST skill over Niño3.4 therefore are not likely
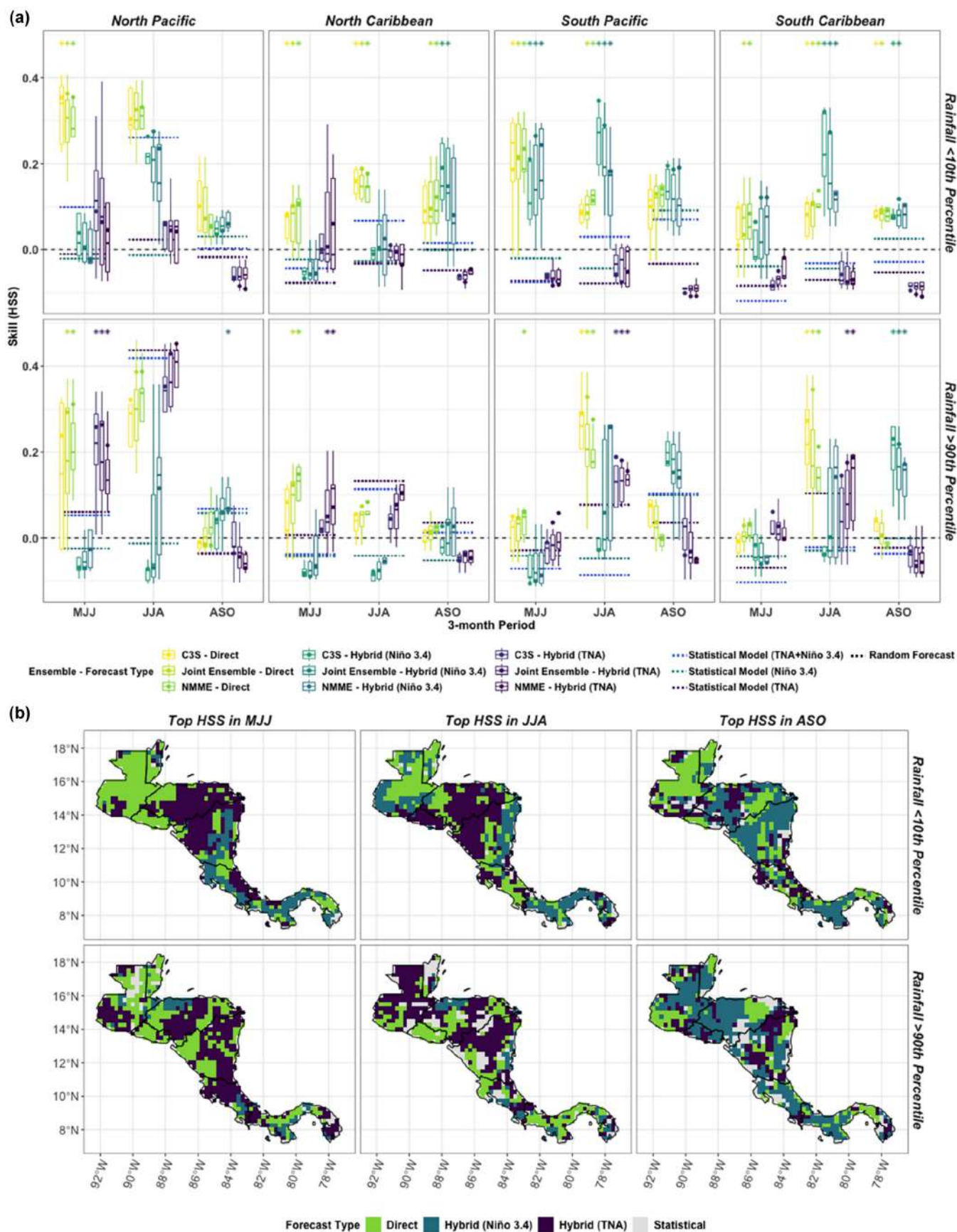
FIGURE 7   Legend on next page.

driving the differences in forecast skill over the region. SST skill is lower over the TNA region in the late wet season (Figure 8a), but this is also when the TNA region is least associated with rainfall, compared to the earlier months (Figure 8b, bottom right panel). The stronger performance of the TNA indirect forecasts correlates with this association between TNA SST and Central American rainfall, as the TNA indirect forecasts tend to perform better earlier in the wet season (Figures 4–7). 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 Niño3.4 (TNA) SST and rainfall is more than 15% (20%) lower than in the North Pacific in the early to middle wet season (Figure 8b).

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, and act as a starting point to tailor seasonal forecasts for use cases across the region:

1. ***The North Pacific i s 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 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 10 days of May and June (INSIVUMEH, 2022a), underscoring the importance of the MJJ forecasts for communicating rainfall information to farmers at the local technical agroclimatic forums (INSIVUMEH, 2022b).

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. Lower skill over this area may be a result of limited gauges over the Caribbean slope as compared to the Pacific slope (e.g., Maldonado et al., 2016), and although similar spatial variability in skill was seen against other gridded observed datasets that use multiple inputs such as satellites and gauges (Figure S2), it would be worth further exploring the spatial skill differences using gauge data alone in future studies. 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 (Taylor and Alfaro, 2005; Maldonado et al., 2013). 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 betwe en forecasting methods (e.g., direct vs. indirect) than between ensemble types (C3S vs. NMME).*** Similarities in skill across ensembles simplify decision making, as the

FIGURE 7 (a) Spatially averaged comparison of forecast skill for low- and high-rainfall extremes using Heidke (HSS) across seasons and grouped by subregion. Boxplots are catogorized 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 0.0, equal to 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. [Colour figure can be viewed at wileyonlinelibrary.com]
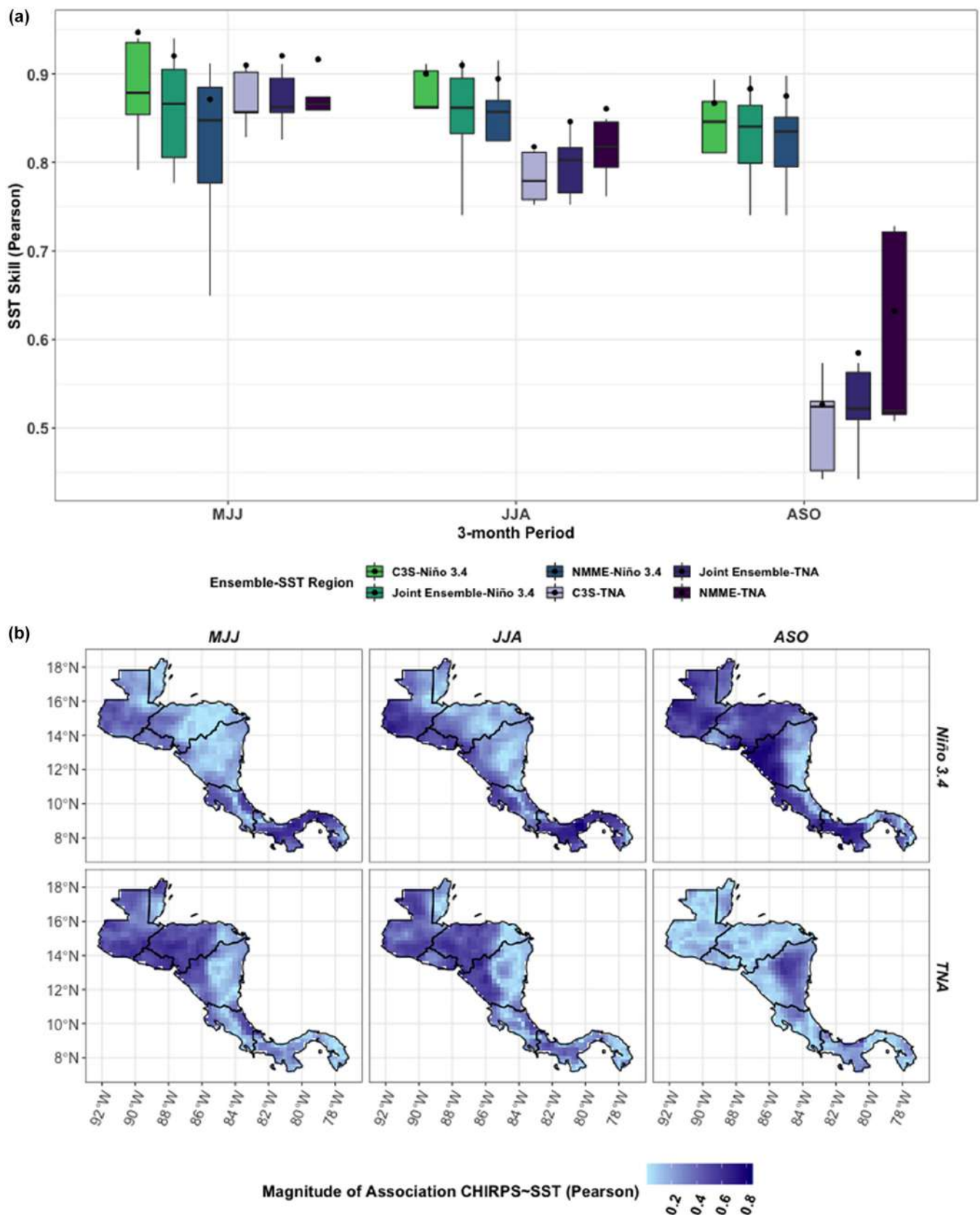
FIGURE 8 (a) Ensemble SST skill measured using Pearson's R, plotted by season for the Niño3.4 and TNA SST regions. Boxplots represent the range of contributing models; points refer to the skill of each ensemble's mean. (b) Pearson's R between observed rainfall and SST in the Niño3.4 and TNA regions, plotted by season [Colour figure can be viewed at wileyonlinelibrary.com]

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 Supporting 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 Niño3.4 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 centred over the 86^0 W longitude line in the early to middle wet season (MJJ/JJA).
- **Low-rainfall extremes detection:** Direct and Niño3.4 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 differences in rainfall skill, as SST skill is often high across the models, especially when SST for a given zone (e.g., Niño3.4) is highly correlated with rainfall in Central America. The GCMs are relatively good at predicting Niño3.4 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 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 Niño3.4 and TNA SST associations with rainfall also align with other studies, which show ENSO has a stronger relationship with rainfall in the late wet season (e.g., Waylen and Quesada, 2001; Spence *et al.*, 2004), and TNA has a strong association with rainfall in the early wet season (Spence *et al.*, 2004; Alfaro, 2007; Maldonado *et al.*, 2017). The comparative value of using dynamically driven forecasts 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 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 compared to their skill in the North Pacific (see Figures 5 and 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; Hidalgo *et al.*, 2015; Maldonado *et al.*, 2018; Durán-Quesada *et al.*, 2020). 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 CLLJ branches around 10^0 N 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 *et al.*, 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 *et al.*, 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 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 | CONCLUSIONS

This study systematically compares seasonal rainfall forecasts from two multimodel 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 multimodel 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.

CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT
The forecast data are publicly accessible via the Climate Data Store and the IRI data library. All reference data are publicly available as described in section 2.

ORCID
Katherine M. Kowal https://orcid.org/0000-0002-9792-2540
Louise J. Slater https://orcid.org/0000-0001-9416-488X
Anne F. Van Loon https://orcid.org/0000-0003-2308-0392

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. *Ingenieria y Competitividad*, 9(2), 33–48. https://doi.org/10.25100/iyc.v9i2.2486.

Alfaro, E.J., Chourio, X., Muñoz, A.G. and 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.

Alfaro, E. J., Hidalgo, H.G. and Mora, N. (2016a) Prediction of MJ rainfall season using CCA models. *Tópicos Meteorológicos y*

Oceanográficos, 15, 5–19. Available at: https://www.researchgate.net/publication/314865864.

Alfaro, E.J., Hidalgo, H.G., Mora, N.P., Prez, P. and Fallas, B. (2016b) Assessment of Central America regional climate outlook forum maps, 1998–2013. *Topicos Meteorologicos y Oceanograficos*, 15, 37–52.

Alfaro Martínez, E.J., Hidalgo León, H.G. and 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.

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., Rashid, I.U., Sylia, M.B., Stephenson, T., Taylor, M., and 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. (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. and Alfaro, E.J. (2009) Métodos de reducción de escala: aplicaciones al tiempo, clima, variabilidad climática y cambio climático. *Revibec: Revista Iberoamericana de Economía Ecológica*, 11, 39–52.

Amador, J.A., Alfaro, E.J., Lizano, O.G. and 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.

Barnston, A.G., Tippett, M.K., L'Heureux, M.L., Li, S. and 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. and Barnston, A.G. (2019) Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, 53, 7215–7234. https://doi.org/10.1007/s00382-017-3603-3.

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

Bauer, P., Thorpe, A. and 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. and 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., Kirtman, B.P. and Pegion, K. (2020) Evolution of the North American Multi-Model Ensemble. *Geophysical Research Letters*, 47(9), e2020GL087408. https://doi.org/10.1029/2020GL087408.

Becker, E., Zhang, Q., van den Dool, H., Saha, S., Peng, P., Peña, M., Tripp, P. and Huang, J. (2012) Evaluation of the National Multi-Model Ensemble System for Seasonal and Monthly Prediction. In: Science and Technology Infusion Climate Bulletin, *36th NOAA Annual Climate Diagnostics and Prediction Workshop*, pp. 43–45. Fort Collins, Colorado: NOAA's Natinal Weather Service. Available at: https://www.nws.noaa.gov/ost/climate/STIP/37CDPW/37cdpw-ebecker1.pdf.

Beveridge, L., Whitfield, S., Fraval, S., van Wijk, M., van Etten, J., Mercado, L., Hammond, J., Davila Cortez, L., Gabriel Suchini, J. and 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.

Braman, L.M., van Aalst, M.K., Mason, S.J., Suarez, P., Ait-Chellouche, Y. and 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.

Calvo-Solano, O.D., Quesada-Hernández, L.E., Hidalgo, H. and 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., Dutra, E., Lavaysse, C. and 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. and 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.

Comisión Económica Para América Latina y El Caribe (CEPAL). (2002) *El impacto socioeconómico y ambiental de la sequía de 2001 en Centroamérica*. Available at: https://www.cepal.org/es/publicaciones/25574-impacto-socioeconomico-ambiental-la-sequia-2001-centroamerica.

Conover, W.J. (1999) Nonparametric methods. In: *Practical Nonparametric Statistics*, 3rd edition. New Jersey: Wiley.

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

Cook, K.H. and 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.

Davis, P., Ruth, C., Scaife, A.A. and Kettleborough, J. (2020) A large ensemble seasonal forecasting system: GloSea6. *American Geophysical Union, Fall Meeting*, 2020, A192-05. Available at: https://ui.adsabs.harvard.edu/abs/2020AGUFMA192...05D/abstract.

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. and 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.

Delsole, T., Yang, X. and 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., Pascale, S., Paynter, D., Rosati, A., Schwarzkopf, M.D., Shevliakova, E., Underwood, S., Wittenberg, A.T., Xiang, B., Yang, X., Zeng, F., Zhang, H., Zhang, L. and 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, e2019MS001895. https://doi.org/10.1029/2019MS001895.

di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R., van den Hurk, B., Vellore, R., Krishnan, R. and 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.

Doblas-Reyes, F.J., García-Serrano, J., Lienert, F., Biescas, A.P. and 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.

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., Gonzalez, P.L.M., Romgero (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. and 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*. Pretoria: IRICW, pp. 11–18.

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

Enfield, D.B. and 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. and 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.

Fernandes, K., Bell, M. and Muñoz, Á.G. (2022) Combining precipitation forecasts and vegetation health to predict fire risk at subseasonal timescale in the Amazon. *Environmental Research Letters*, 17, 074009. https://doi.org/10.1088/1748-9326/ac76d8.

Fernandes, K., Muñoz, A.G., Ramirez-Villegas, J., Agudelo, D., Llanos-Herrera, L., Esquivel, A., Rodriguez-Espinoza, J. and 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.

Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., Haak, H., Brune, S., Früh, B. and Baehr, J. (2021) The German Climate Forecast System: GCFS. *Journal of Advances in Modeling Earth Systems*, 13(2), e2020MS002101. 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. and Husak, G.J. (2014) *A Quasi-Global Precipitation Time Series for Drought Monitoring*. Reston, VA: U.S. Geological Survey. Available at: https://pubs.er.usgs.gov/publication/ds832.

Garcia-Solera, I. and Ramirez, P. (2012) *Central America's Seasonal Climate Outlook Forum*. Geneva: WMO. Available at: 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. and 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, K., Kushnir, Y. and Cane, M.A. (2000) Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. *Journal of Climate*, 13, 297–311.

Gilbert, G.E. (2019) Nonparametric tests used in simulation research. In: Nestel, D., Hui, J., Kunkler, K., Scerbo, M. and Calhoun, A. (Eds.) *Healthcare Simulation Research*. Cham: Springer, pp. 207–214. https://doi.org/10.1007/978-3-030-26837-4_28.

Giorgi, F. (2006) Climate change hot-spots. *Geophysical Research Letters*, 33, 8707. https://doi.org/10.1029/2006GL025734.

Goddard, L., González Romero, C., Muñoz, Á.G., Acharya, N., Shamsuddin, A., 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., García-López, A., ... (2020) Climate services ecosystems in times of COVID-19. *WMO Bulletin*, 69(2), 39–46.

Golding, N., Hewitt, C., Zhang, P., Liu, M., Zhang, J. and 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., Pérez-Briceo, P., Hidalgo, H.G. and Alfaro, E.J. (2019) The Central American Dry Corridor: a consensus statement and its background. *Revista Yu'am*, 3(5), 42–51.

Grolemund, G. and 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. and Cantelli, A. (2020) *The new CMCC operational seasonal prediction system*. Lecce: CMCC. Technical notes: TN0288 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. and 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. and 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. and 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.

Hannah, L., Donatti, C.I., Harvey, C.A., Alfaro, E., Rodriguez, D.A., Bouroncle, C., Castellanos, E., Diaz, F., Fung, E., Hidalgo, H.G., Imbach, P., Läderach, P., Landrum, J.P. and Solano, A.L. (2017) Regional modeling of climate change impacts on smallholder agriculture and ecosystems in Central America. *Climatic Change*, 141(1), 29–45. https://doi.org/10.1007/s10584-016-1867-y.

Hao, Z., Singh, V.P. and 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.

Hidalgo, H.G. and 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. https://doi.org/10.15517/RBT.V60I3.28340.

Hidalgo, H.G. and 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., Alfaro, E.J. and 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., Amador, J.A., Alfaro, E.J. and 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., Durán-quesada, A.M., Amador, J.A. and 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.

Higgins, R.W., Kim, H.K. and 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. and 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.

Huang, T.-Y. and 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.

Imbach, P., Chou, S.C., Lyra, A., Rodrigues, D., Rodriguez, D., Latinovic, D., Siqueira, G., Silva, A., Garofolo, L. and Georgiou, S. (2018) Future climate change scenarios in Central America at high spatial resolution. *PLoS One*, 13(4), e0193570. https://doi.org/10.1371/journal.pone.0193570.

INSIVUMEH. (2022b) *Mesas Agroclimaticas*. Available at: https://insivumeh.gob.gt/mesas-agroclimaticas/.

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

Instituto Meteorológico Nacional (IMN). (2019) *Boletin del Enos No.118: Fase actual: El Niño*. Available at: https://www.imn.ac.cr/en/boletin-enos.

Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología (INSIVUMEH). (2022a) *Fenología*. Available at: http://svsa.insivumeh.gob.gt/fenologia/.

Inter-American Institute for Cooperation on Agriculture Costa Rica (IICACR). (2014) *Corredor Seco Americano Boundary*. Costa Rica: ArcGIS. Available at: 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. and Monge-Sanz, B. (2018) SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development Discussions*, 154, 1–44. https://doi.org/10.5194/gmd-2018-228.

Kassambara, A. (2020) *ggpubr: "ggplot2" based publication ready plots* (R package version 0.4.0).

Khajehei, S., Ahmadalipour, A. and 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.

Kharin, V.V., Merryfield, W.J., Boer, G.J. and Lee, W.S. (2017) A postprocessing method for seasonal forecasts using temporally and spatially smoothed statistics. *Monthly Weather Review*, 145(9), 3545–3561. https://doi.org/10.1175/MWR-D-16-0337.1.

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., Rienecker, M., Suarez, M., Li, Z.E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B. and 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.

Kowal, K.M., Slater, L.J., van Loon, A.F. and 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*, 42, 4953–4971. https://doi.org/10.1002/JOC.7514.

Kreibich, H., Müller, M., Schröter, K. and Thieken, A.H. (2017) 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. and 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. and 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. and Wallis, W.A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621. https://doi.org/10.2307/2280779.

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

Lin, H., Muncaster, R., Diro, G.T., Merryfield, W., Smith, G., Markovic, M., Erfani, A., Kharin, S., Lee, W.-S., Parent, R., Pavlovic, R. and Charron, M. (2021) *The Canadian seasonal to interannual prediction system version 2.1 (CanSIPSv2.1)*. Canadian Meteorological and Environmental Prediction Centre. Technical note.

Magaña, V., Amador, J.A. and 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. and 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., Alfaro, E. and 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.

Maldonado, T., Alfaro, E., Rutgersson, A. and 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., Rutgersson, A., Alfaro, E., Amador, J. and 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.

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

Mann, H.B. and Whitney, D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1), 50–60. https://doi.org/10.1214/Aoms/1177730491.

Marengo, J.A., Chou, S.C., Torres, R.R., Giarolla, A., Alves, L.M. and Lyra, A. (2014) *Climate change in Central and South America: recent trends, future projections, and impacts on regional agriculture*. Available at: 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. and 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.

Martinez, C., Muñoz, Á.G., Goddard, L., Kushnir, Y. and Ting, M. (2022) Seasonal prediction of the Caribbean rainfall cycle. *Climate Services*, 27, 100309. https://doi.org/10.1016/J.CLISER.2022.100309.

Mason, S.J. and Baddour, O. (2008) Statistical modelling. In: Troccoli, A., Harrison, M., Anderson, D.L.T., Mason, S.J. (Eds.) *Seasonal Climate: Forecasting and Managing Risk*. Dordrecht, Netherlands: Springer, pp. 163–201.

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

Mason, S.J. and 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.

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

Mo, K.C. and 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., Longo, K.M., Nakada, K. and 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, e2019JD031767. https://doi.org/10.1029/2019JD031767.

Muñoz, Á.G., Chourio, X., Rivière-Cinnamond, A., Diuk-Wasser, M.A., Kache, A., Mordecai, A., Harrington, L. and Thomson, M.C. (2020a) 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, Á.G., Chourio, X., Rivière-Cinnamond, A., Diuk-Wasser, M.A., Kache, P.A., Mordecai, E.A., Harrington, L. and Thomson, M.C. (2020b) AeDES: a next-generation monitoring and forecasting system for environmental suitability of Aedes-borne disease transmission. *Scientific Reports*, 10(1), 12640. https://doi.org/10.1038/s41598-020-69625-4.

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

Muñoz, E., Busalacchi, A.J., Nigam, S. and 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. and 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.

National Center for Atmospheric Research (NCAR) - Research Applications Laboratory. (2015) *verification: weather forecast verification utilities* (R package version 1.42).

National Oceanic and Atmospheric Administration (NOAA). (2020) *The North American Multi-Model Ensemble (NMME) Operational Phase 2019–2020 Annual Report*. Silver Spring, Maryland: NOAA. https://doi.org/10.25923/TF81-0S82.

Pebesma, E. (2018) Simple features for R: standardized support for spatial vector data. *R Journal*, 10(1), 439–446.

Pedersen, T.L. (2022) *patchwork: the composer of plots* (R package version 1.1.2).

Pérez-Briceo, P.M., Alfaro, E., Hidalgo, H. and Jimenez, F. (2016) Distribución espacial de impactos de eventos hidrometeorológicos en América Central. *Revista de Climatologia*, 16, 63–75.

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. and Gleason, B. (2002) Recent changes in climate extremes in the Caribbean region. *Journal of Geophysical Research: Atmospheres*, 107(21), ACL16-1–ACL16-9. 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).

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. and 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*, 36, 2021-2038.. https://doi.org/10.1175/WAF-D-20-0133.1.

Poveda, G. and Mesa, O.J. (1999) *La Corriente de chorro superficial del oeste "del Chocó": Climatología y Variabilidad durante las fases del ENSO*. Paper presented at Proc. V Congreso Colombiano de Meteorología. Available at: https://www.researchgate.net/publication/233857360.

Poveda, G., Waylen, P.R. and 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.

R Core Team. (2021) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Renard, B. and Thyer, M. (2019) Revealing hidden climate indices from the occurrence of hydrologic extremes. *Water Resources Research*, 55, 7662–7681. https://doi.org/10.1029/2019WR024951.

Reynolds, R., Smith, T., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G. (2007) Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20, 5473–5496.

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. and 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.

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., Smith, D., Takaya, Y., Vecchi, G. and 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, A., and 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. and Zardi, D. (2018) Exchange processes in the atmospheric boundary layer over mountainous terrain. *Atmosphere*, 9(3), 102. https://doi.org/10.3390/atmos9030102.

Shi, W., Schaller, N., Macleod, D., Palmer, T.N. and 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.

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. and Zappa, M. (2022) Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models. *Hydrology and Earth System Sciences Discussions*, 2022, 1–35. https://doi.org/10.5194/HESS-2022-334.

Slater, L.J., Villarini, G. and Bradley, A.A. (2019) Evaluation of the skill of North American Multi-Model Ensemble (NMME) global climate models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dynamics*, 53(12), 7381–7396. https://doi.org/10.1007/s00382-016-3286-1.

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

Spence, J.M., Taylor, M.A. and 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.

Strazzo, S., Collins, D.C., Schepen, A., Wang, Q.J., Becker, E. and 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.

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

Taylor, M.A., Enfield, D.B., Chen, A.A., Taylor, M.A., Enfield, D.B. and 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.

Taylor, M.A., Stephenson, T.S., Chen, A.A. and Stephenson, K.A. (2012) Climate change and the Caribbean: review and response. *Caribbean Studies*, 40(2), 169–200.

Tippett, M.K., Goddard, L. and 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.

Trenberth, K.E. (1997) The definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12), 2771–2777.

Trenberth, K.E. and Stepaniak, D.P. (2001) Indices of El Niño evolution. *Journal of Climate*, 14(8), 1697–1701.

Walker, D.P., Birch, C.E., Marsham, J.H., Scaife, A.A., Graham, R.J. and 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, 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. and 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., Lee, S.K. and 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.

Waylen, P. and 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.

Waylen, P.R., Quesada, M.E. and 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. (2017). *afc: generalized discrimination score* (R package version 1.4.0).

Weigel, A.P., Knutti, R., Liniger, M.A. and 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. and 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.

Weisheimer, A. and 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., García-Solórzano, A.M., Gibson, P. B., Goddard, L., Romero, C.G., Graham, R.J., Graham, R.M., Grams, C.M., Halford, A., Katty Huang, W.T., Jensen, K., Kilavi, M., Lawal., K.A., Lee, R.W., MacLeod, D., Manrique-Suñén, A., Martins, E.S.P.R., Maxwell, C.J., Merryfield, W.J., Muñoz, Á.G., Olaniyan, E., Otieno, G., Oyedepo, J.A., Palma, L., Pechlivanidis, I.G., Pons, D., Martin Ralph, F., Reis, D.S., Remenyi, T.A., Risbey, J.S., Robertson, D.J.C., Robertson, A.W., Smith, S., Soret, A., Sun, T., Todd, M.C., Tozer, C.R., Vasconcelos, F.C., Vigo, I., Waliser, D. E., Wetterhall, F. and 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*. New York: Springer.

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., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and 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. and Müller, K. (2021) *dplyr: a grammar of data manipulation* (R package version 1.0.7).

Wickham, H. and 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.

World Meteorological Organization (WMO). (2020) *Guidance on operational practices for objective seasonal forecasting*. Available at: https://library.wmo.int/doc_num.php?explnum_id=11073.

Yuan, X. and 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.

Zhang, W., Villarini, G., Slater, L., Vecchi, G.A. and 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. and Chen, X. (2021) Correspondence relationship between ENSO teleconnection and anomaly correlation for GCM seasonal precipitation forecasts. *Climate Dynamics*, 58, 633–649. https://doi.org/10.1007/S00382-021-05925-3.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.