Impact-based Skill Evaluation of Seasonal Precipitation Forecasts

Zahir Nikraftar[1], Rendani Mbuvha[1], Mojtaba Sadegh[2], and Willem Adolf Landman[3]

[1]Queen Mary University of London
[2]Boise State University
[3]Council for Scientific and Industrial Research

December 27, 2023

Abstract

Forecasting hydroclimatic extremes holds significant importance considering the increasing trends in natural cascading climate-induced hazards such as wildfires, floods, and droughts. This study evaluates the performance of five Copernicus Climate Change Service (C3S) seasonal forecast models (i.e., CMCC, DWD, ECCC, UK-Met, and Météo-France) in predicting extreme precipitation events from 1993 to 2016 using 28 extreme precipitation indices reflecting timing and intensity of precipitation in a seasonal timescale. We design indices using various precipitation thresholds to reflect model skill in capturing the distribution and intensity of precipitation over a season. We use percentage bias, the Kendall Tau rank correlation, and ROC scores for skill evaluation. We introduce an impact-based framework to evaluate model skill in capturing extreme events over regions prone to natural disasters such as floods and wildfires. The performance of models varies across regions and seasons. The model skill is highlighted primarily in the tropical and inter-tropical regions, while skill in extra-tropical regions is markedly lower. Elevated precipitation thresholds correlate with heightened model bias, revealing deficiencies in modelling severe precipitation events. The impact-based framework analysis highlights the superior predictive capabilities of the UK-Met and Météo-France models for extreme event forecasting across many regions and seasons. In contrast, other models exhibit strong performance in specific regions and seasons. These results advance our understanding of an impact-based framework in capturing a broad spectrum of extreme climatic events through the strategic amalgamation of diverse models across different regions and seasons, offering valuable insights for disaster management and risk analysis.

Impact-based Skill Evaluation of Seasonal Precipitation Forecasts

Zahir Nikraftar[1], Rendani Mbuvha[1,2], Mojtaba Sadegh[3,4], Willem A. Landman[5]

[1] Machine Intelligence and Decision Systems (MInDS) Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), London, UK, E1 4NS.
[2]School of Statistics and Actuarial Science, University of Witwatersrand, Johannesburg, South Africa
[3] Department of Civil Engineering, Boise State University, Boise, ID, USA.
[4] United Nations University Institute for Water, Environment and Health, Hamilton, ON, Canada.
[5] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa.

Corresponding Author: Rendani Mbuvha (r.mbuvha@qmul.ac.uk)

Abstract
Forecasting hydroclimatic extremes holds significant importance considering the increasing trends in natural cascading climate-induced hazards such as wildfires, floods, and droughts. This study evaluates the performance of five Copernicus Climate Change Service (C3S) seasonal forecast models (i.e., CMCC, DWD, ECCC, UK-Met, and Météo-France) in predicting extreme precipitation events from 1993 to 2016 using 28 extreme precipitation indices reflecting timing and intensity of precipitation in a seasonal timescale. We design indices using various precipitation thresholds to reflect model skill in capturing the distribution and intensity of precipitation over a season. We use percentage bias, the Kendall Tau rank correlation, and ROC scores for skill evaluation. We introduce an impact-based framework to evaluate model skill in capturing extreme events over regions prone to natural disasters such as floods and wildfires. The performance of models varies across regions and seasons. The model skill is highlighted primarily in the tropical and inter-tropical regions, while skill in extra-tropical regions is markedly lower. Elevated precipitation thresholds correlate with heightened model bias, revealing deficiencies in modelling severe precipitation events. The impact-based framework analysis highlights the superior predictive capabilities of the UK-Met and Météo-France models for extreme event forecasting across many regions and seasons. In contrast, other models exhibit strong performance in specific regions and seasons. These results advance our understanding of an impact-based framework in capturing a broad spectrum of extreme climatic events through the strategic amalgamation of diverse models across different regions and seasons, offering valuable insights for disaster management and risk analysis.

Key Points

- The C3S models' skill in predicting precipitation variability across seasons is highlighted in the tropical and inter-tropical regions.

- UK-Met and Météo-France consistently outperform other models, demonstrating better accuracy and reliability, aligning with findings from previous studies.

- An Impact-Based framework offers valuable insights for targeted risk assessments, particularly in regions prone to wildfire and floods.

Plain Language Summary

This study looks at how well five seasonal forecast models can predict extreme precipitation and related events such as floods and droughts. We assess the performance of models by looking at 28 different extreme precipitation indices that show when and how much it rains during different seasons. The results show that some models do better in certain areas and seasons. The study also introduces a new way of assessing model skill by looking at the impact of extreme events in areas prone to disasters such as floods and wildfires. We find that two models, UK-Met and Météo-France, are particularly good at predicting extreme events in various places and seasons. This information is important for better managing and understanding the risks of natural disasters.

1. Introduction

Precipitation plays a crucial role in momentum flux exchange at the ocean– atmosphere-land interface (Xue et al., 2020), and as such, is one of the primary outputs of weather and climate models (Tapiador et al., 2019). Numerous international initiatives such as the North American Multi-Model Ensemble (NMME)[1] and Copernicus Climate Change Service (C3S)[2] multi-system seasonal forecast predict precipitation, and other Météorological factors, at various spatiotemporal scales. Such forecasts are used for a variety of purposes, including extreme event early warning. Forecast models rely on the sources of atmospheric predictability, such as modes of variability including El Niño–Southern Oscillation (ENSO), Madden–Julian oscillation (MJO), Quasi-Biennial Oscillation (QBO), and Indian Ocean Dipole (IOD). Other sources of predictability include anomalies in the initial state of an Earth system component with a persistence time that aligns with the projected forecast duration (i.e., large-scale anomalies in upper ocean heat content, sea ice, snowpack, soil moisture), and external forcing (Assessment of Intraseasonal to Interannual Climate Prediction and Predictability, 2010; Baldwin et al., 2003; Committee on Developing a U.S. Research Agenda to Advance Subseasonal to Seasonal Forecasting et al., 2016; Lau & Waliser, 2012; Shukla et al., 2000; Zhang et al., 1997). Despite the significance of precipitation, numerical weather models face difficulties in predicting its spatial patterns, timing and intensity (Tapiador et al., 2019; Mallakpour et al. 2022). This is because the predictive capabilities of seasonal forecast models are constrained by the uncertainty in initial and boundary conditions, climate change-induced modifications of teleconnection patterns, imperfect parameterization schemes, and the variability in parameters (Villarini et al., 2011; Xu et al., 2021).

[1] https://www.cpc.ncep.no aa.gov/products /NMME/
[2] https://cds.climate .copernicus.eu/

Accurate precipitation predictions are of great importance in the formulation of mitigation and adaptation measures for climate and hydrological extreme events as well as minimizing impacts from their cascading hazards such as flood, drought, and wildfire (Gebrechorkos et al., 2022; Vitart & Robertson, 2018). Recently several environmental and Climate Forecasting Systems such as the hydrological forecasting system, the Canadian Forest Fire Weather Index System[3] , and the global drought forecasting system [4] have been developed, which use seasonal forecasts as input with the purpose of weather extremes risk assessment and response (Alfieri et al., 2013; Arheimer et al., 2020; Samaniego et al., 2019; Thielen et al., 2009). The accuracy and trustworthiness of such systems is highly dependent on the process representation and parametric accuracy of the weather forecast models that provide the forcing for risk assessment (Gebrechorkos et al., 2022; Wanders & Wood, 2016). The skill of seasonal forecast systems has a substantial spatial variability which could be due to factors like quality of observation systems, model biases, and inherent properties of the climate system (Kumar & Zhu, 2018). Also, seasonal forecast performance can vary spatially across regions due to the complex and region-specific interactions between climate drivers and local environmental factors (Hao et al., 2018). Therefore, it is essential to assess the performance of different forecasting models across diverse global regions, and specific to the impacts that forecast errors may induce to identify most the effective and reliable models.

While studies have evaluated the skill of seasonal forecast models in predicting total precipitation at the sub seasonal to seasonal scales (Becker et al., 2014; Gebrechorkos et al., 2022; Nobakht et al., 2021; Roy et al., 2020), there is a need for an alternative impact-based assessment of forecast models that can inform their applicability for specific target extremes (e.g., flood, wildfire). Traditional forecast model performance assessments conduct a top-down hazard information approach by mainly investigating the model's skill in capturing weather patterns in comparison to the reference datasets (De Andrade et al., 2019; Moron & Robertson, 2020; Vitart & Robertson, 2018). The shift towards impact-based assessment framework reflects the evolving landscape of climate science and its increasing relevance in the face of a changing climate (Rad et al. 2022). It emphasizes the importance of moving beyond traditional evaluation methods and towards a comprehensive understanding of how weather forecasts directly influence society, ecosystems, and infrastructure resilience (AghaKouchak et al., 2018; Khorshidi et al., 2020; Mallakpour et al., 2022; Modaresi Rad et al., 2022; Sadegh et al., 2018). Such a framework considers the vulnerability of the local environment to specific weather events and warns of the associated impacts. An instance of such an influence might involve a chain reaction of hazards, like flooding due to repeated heavy rainfall events (Sadegh et al. 2018) or wildfires resulting from consecutive days of no precipitation and increased temperature, which can create conditions conducive to ignition (Khorshidi et al. 2020). Impact-based assessment of seasonal precipitation forecasts involve assessing the effectiveness of models by considering the real-world impact of extreme

[3] http://cwfis.cfs.nrcan .gc.ca/en_CA/background/summary/fdr
[4] http://iridl. ldeo.columbia.edu/maproom/Global/Drought/Global/CPC_GOB/MME_pt_Persist.html

precipitation on various sectors and systems. It goes beyond assessing the mere accuracy of forecasted precipitation and aims to understand how well the forecasts translate into meaningful information for decision-making and risk management (AghaKouchak et al., 2023).

Our main objective in this study is to evaluate the skill of the five state-of-the-art seasonal prediction systems from the Copernicus Climate Change Service (C3S) multi-model at a global scale in predicting particular features of extreme events which could lead to cascading hazards. These indices are defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) and have been investigated in other studies (Chervenkov et al., 2019; Chervenkov & Slavov, 2019). These indices were designed to capture different aspects of precipitation such as timing and intensity and are useful in diagnosing the variability of precipitation at various timescales posing them as proper metrics for impact-based assessment. We refined these indices to capture weather patterns that could cause climate-induced hazards such as flood and wildfire. We conduct an evaluation of forecast models to assess their capability in discerning situations that result in the occurrence of a specific event from those that lead to its non-occurrence. As a related task, we perform a targeted analysis of model performance in regions with high risk of wildfires and floods. The following questions are answered in this study: Do seasonal forecast models have the capability to represent the variability of precipitation throughout the season? Is there a potential for combining various models to be an effective strategy for predicting extreme events, considering the variability of model performance across seasons and regions?

2. Methodology

This study examines the effectiveness of precipitation forecasts of five seasonal forecasting models from C3S project including the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC: version 35), Deutscher Wetterdienst (DWD: version 21), Environment and Climate Change Canada (ECCC: version 3), Météo France (Météo-France: version 8), and UK Met Office (UK-Met: version 601) models in accurately predicting extreme precipitation indices during the hindcast period spanning 1993 to 2016 (refer to Table S1). Validation was carried out using the fifth generation ECMWF reanalysis (ERA5) precipitation product.

2.1 Data Preparation

For our evaluation, we employed eight distinct climate extreme indices, following Expert Team on Climate Change Detection and Indices (ETCDDI)[5] definitions, to encompass different aspects of precipitation extremes, such as event duration, intensity, and frequency. We used 1mm, 10mm, and 20mm precipitation thresholds (representing wet days, heavy precipitation days, and very heavy precipitation days, respectively) for the calculation of climate extreme indices. We also used 75th and 95th percentiles of each grid in the reference datasets as a secondary constraint in calculating indices to be more representative of the local environmental conditions. Combination of metrics and thresholds resulted in generating a

[5] https://www.wcrp-climate.org/data-etccdi

comprehensive set of 28 climate extreme indices (refer to Table 1). These indices were specifically designed to offer insights into the potential changes observed in precipitation extremes over time and their impacts on various aspects of Earth system processes (Dunn et al., 2022). The assessment of seasonal skill for the models was carried out using forecast initializations on the first day of February, May, August, and November with a 1-month lead time (i.e., March-May (MAM), June-August (JJA), September-November (SON), and December-February (DJF) seasons respectively).

Table 1. List of climate extreme indices extracted and used to conduct this study.

abbreviation	Index	abbreviation	Index
cdd1	maximum consecutive dry days 1mm	nwd20q95	number of wet days 20mm and 95th percentile
cdd10	maximum consecutive dry days 10mm	fr1q75	fraction of precipitation over 1mm and 75th percentile
cdd20	maximum consecutive dry days 20mm	fr1q95	fraction of precipitation over 1mm and 95th percentile
cwd1	maximum consecutive wet days 1mm	fr10q75	fraction of precipitation over 10mm and 75th percentile
cwd10	maximum consecutive wet days 10mm	fr10q95	fraction of precipitation over 10mm and 95th percentile
cwd20	maximum consecutive wet days 20mm	fr20q75	fraction of precipitation over 20mm and 75th percentile
int1	daily pr intensity 1mm	fr20q95	fraction of precipitation over 20mm and 95th percentile
int10	daily pr intensity 10mm	hpd	Heavy precipitation days
int20	daily pr intensity 20mm	vhpd	very Heavy precipitation days
nwd1q75	number of wet days 1mm and 75th percentile	h1dp	Highest 1-day precipitation amount
nwd1q95	number of wet days 1mm and 95th percentile	h5dp	Highest 5-day precipitation amount
nwd10q75	number of wet days 10mm and 75th percentile	propd1	Proportion of days with precipitation at or above 1mm
nwd10q95	number of wet days 10mm and 95th percentile	propd10	Proportion of days with precipitation at or above 10mm
nwd20q75	number of wet days 20mm and 75th percentile	propd20	Proportion of days with precipitation at or above 20mm

The analysis was conducted on the ensemble mean for each model. All forecast models and reference data were re-gridded to a consistent one-degree resolution. For spatial aggregation, we conducted our analysis over the Intergovernmental Panel on Climate Change (IPCC) regions shown in Table S2 (Iturbide et al., 2020). The IPCC divides the world into major regions, each of which includes a group of countries or territories that share similar climate characteristics, geographic features, and socio-economic factors. The IPCC regions, also known as the "IPCC Regional Reporting", are a set of geographical regions used by the IPCC as a framework for understanding how climate change affects different parts of the world and to facilitate the assessment of climate change impact, vulnerability, and adaptation strategies at the regional level.

2.2 Model Evaluation

For evaluation and comparison of forecast models against reference data, we employed percentage bias with an optimal value of zero. Here, zero signifies a perfect alignment between model predictions and observed data, and positive-bias/negative-bias signifies over-estimation/under-estimation (Eq.1).

$$PBIAS = 100\frac{\sum_{i=1}^{N}(S_i - O_i)}{\sum_{i=1}^{N} O_i} \qquad \text{Eq.1}$$

where S_i is the model simulation and O_i is the observed value at time i.

We also conducted a discriminant analysis to determine if the forecast skill varied in different sections of the precipitation distribution. To achieve this, we categorized the outcomes into three distinct groups using upper and lower terciles: one-third representing below-normal conditions (lower tercile), another third representing above-normal conditions (upper tercile), and the remaining third representing conditions falling between the lower and upper tercile. Utilizing a random forest classification, we conducted a classification based on 75 percent of the available datasets (training sets). Subsequently, we computed the area under the ROC curve which is often used as a summary measure of forecast discrimination for the test data. ROC score is an efficient way to analyze the overall discriminatory power of the forecasts. It takes values between 0 to1. Here we categorized the ROC score to level of discrimination (i.e., in our case ability to discriminate extreme events from non-extreme events). The values of ROC score between 0.0 to 0.6 is considered no discrimination, 0.6 to 0.7 is considered satisfactory discrimination, 0.7 to 0.8 is considered good discrimination, 0.8 to 0.9 is considered very good discrimination, and 0.9 to 1.0 is considered excellent discrimination (Mandrekar, 2010). Also, we conducted Kendall's Tau rank correlation analysis to measure the strength of the relationship between climatic indices extracted from forecast models and reference data aggregated over each IPCC regions (Sen, 1968). These metrics were utilized to assess the skill of the forecast models in capturing extreme events and their associated accuracy.

2.3 Impact based framework

The IPCC regions which were vulnerable to wildfire and flooding were identified by categorizing them based on the proportion of farmland (Friedl & Sulla-Menashe, 2022), proportion of burned areas (Chuvieco et al., 2018; Lizundia-Loiola et al., 2020), percentage of flood-affected zones (Tellman et al., 2021), proportion of built-up regions (Gong et al., 2020), and population density (Schiavina et al., 2023) (see Figure S1). We specifically focused on regions that not only had an elevated risk of wildfire (flood) exposure, but also had a significant built-up area and population density (substantial agricultural presence and population density). These selected regions were earmarked for additional analysis.

We carefully selected relevant extreme indices pertinent to the corresponding climate-induced hazard in each region. For the regions with wildfire as a prominent natural hazard we selected maximum consecutive dry days 1mm (cdd1), and proportion of days with precipitation at or above 1mm (propd1) indices as the relevant indicators of a weather condition conducive to wildfire. For regions with a high risk of flooding, we selected number of wet days with 10mm precipitation and 75th percentile of the reference data (nwd10q75) and heavy precipitation days (hpd) indices as the relevant indicators. We also determined the season with the highest occurrence rate for each specific hazard and regions. In every region, we identified the top-performing model in terms of predictive accuracy using the following selection criteria. Initially, we prioritized models with a combination of higher correlation (statistically significant) and lower bias. If multiple models demonstrated similar high performance compared to the others, we utilized a Taylor diagram to select models that aligned more closely with the reference data across various performance metrics. Our evaluation then focused on assessing the models’ forecast skill with respect to relevant indices in the identified hazard-prone seasons for each region. By adopting this impact-based approach, we aimed to pinpoint the most suitable models and indices for each climate-induced hazard, enabling more effective and tailored climate risk assessments.

3. Results

3.1 Global Analysis of Model Bias

The analysis of percentage bias at a global scale across the four seasons reveals a consistent tendency of underestimation in all forecast models with respect to reference data for most of the extreme wet and extreme dry precipitation indices (Figure 1 shows UK-Met model as an example, and Figures S2-S5 show other models). The cdd1 index, which measures the maximum consecutive dry days at 1mm threshold, shows negative bias for most but not all regions, implying that the models predict a lower number of dry days than observed. In the case of the cwd1 and propd1 indices, which measures the maximum consecutive wet days at 1mm threshold and proportion of days with at least 1mm precipitation, all models exhibit a positive bias, suggesting that they predict a higher number of wet days than observed data.

For wet indices, all models underestimate cwd10 and cwd20 indices (10- and 20-mm precipitation threshold, respectively) across the four seasons and most of the IPCC regions. The models exhibit lower bias for extreme indices that are defined based on a 10mm precipitation threshold. As the threshold increases, so does the bias. This pattern is consistent with the findings of regional C3S assessment studies, for example in Africa (Gebrechorkos et al., 2022). These observations underscore the limitations in forecast capabilities for accurately modelling persistent wet and dry periods. By introducing secondary constraints (i.e., 75th, and 95th percentiles of the reference data) to the indices, the bias increases signifying the models' limitation in capturing more severe extreme events.

Figure 1 and Figures S2-S5 show that the CMCC, DWD, and ECCC models demonstrate relatively lower ability to capture extreme rainfall events within the extratropical IPCC regions compared to the UK-Met and Météo-France models, particularly when 95th percentiles of the reference dataset are introduced as thresholds to the index. This observation is consistent across all four seasons. However, in the tropical and subtropical regions, all models (especially UK-Met and Météo-France models) exhibit relatively better performance (lower bias) in capturing extreme events, compared to extratropical regions, when 75th and 95th percentile thresholds were used in the indices as additional constraints. This is attributed to the model's predictive skill in grasping large-scale teleconnection patterns (Giuntoli et al., 2022).

Figure 2 shows the standardized precipitation anomalies of the five models across the four seasons. Standardized anomalies offer invaluable insights about the localized anomalies through the number of standard deviation departure of forecasts from observations. We normalized the precipitation anomalies against the climatological standard deviation in each grid. A notable tendency to produce a double Intertropical Convergence Zone with significant anomalies over tropical pacific is observed in all models (García-Franco et al., 2023). The pattern of anomalies differs substantially across seasons for all model with MAM, JJA, and DJF seasons showing larger anomalies in the tropical, subtropical, and equatorial regions, while SON season show lower levels of anomaly. In the northern hemisphere's extratropical land regions, seasonal forecasts reveal a larger negative anomaly in summer and larger positive anomalies in winter compared to spring and autumn respectively, which is likely due to the increased influence of local factors on summertime and wintertime precipitation. Uncertainties in these forecasts are largely attributed to model parameterization, including elements such as vegetation cover and cloud physics, which can significantly impact precipitation predictions (Borovikov et al., 2019).

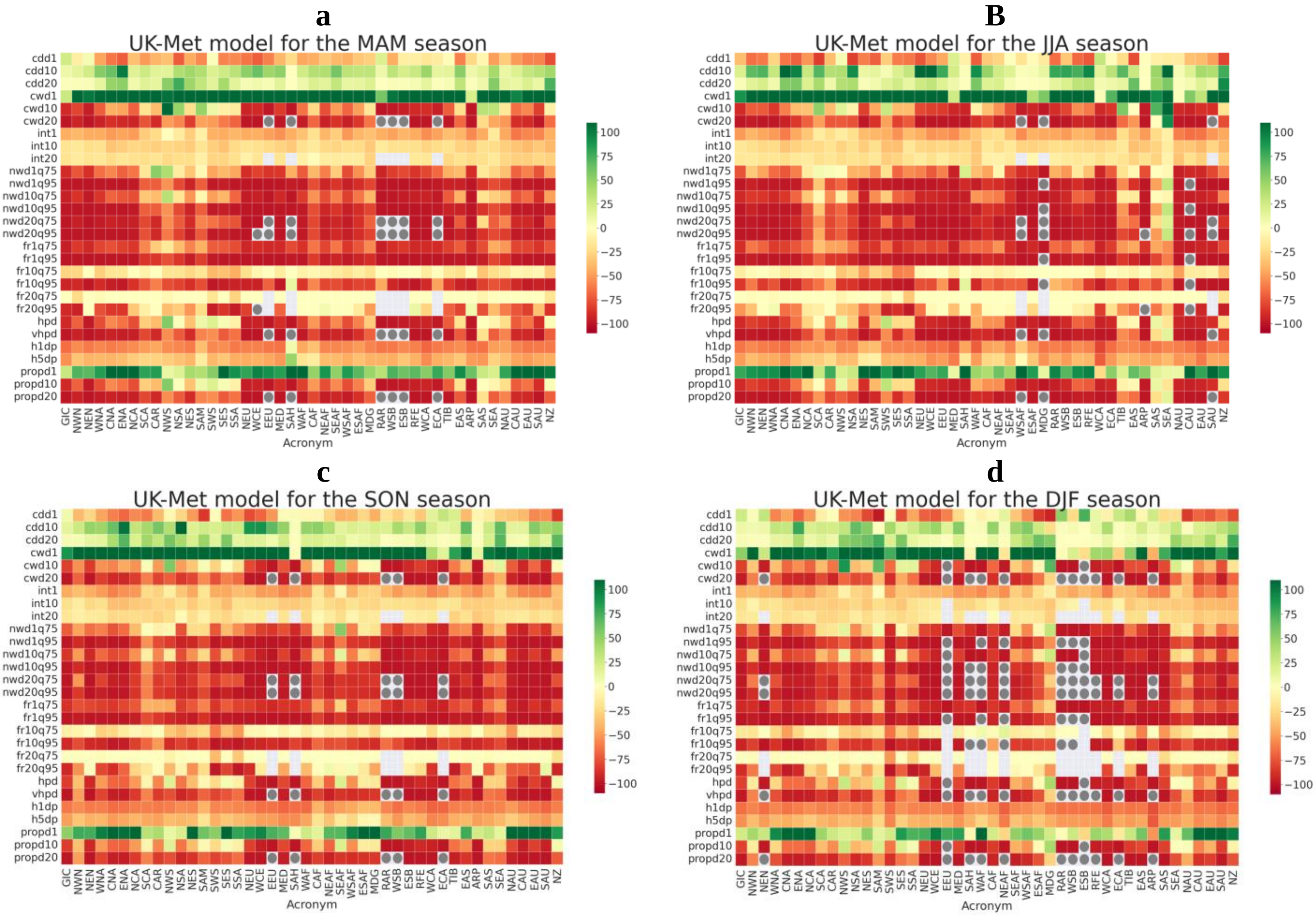
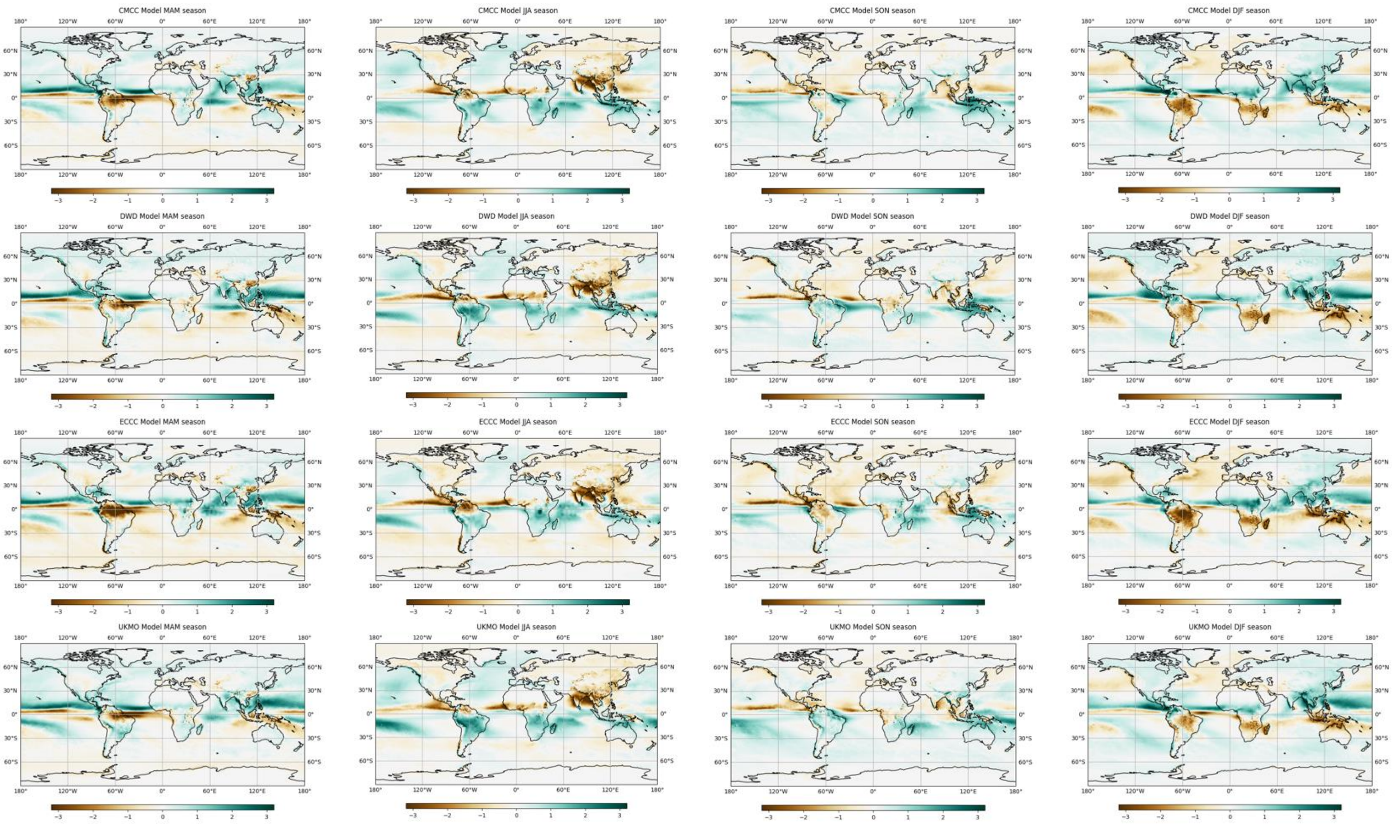


Figure 1. Percentage bias of UK-Met model for a) MAM, b) JJA, c) SON, and d) DJF seasons. Grids indicated in gray circles shows the regions where reference data indicated the existence of extreme events while the forecast models couldn't capture any events satisfying the index requirements. Grids with no values reflect regions that both reference data and forecasts did not capture any events satisfying the index requirements. The color bar is limited to the range of -100 to 100 for visualization purposes. This range was chosen to enhance visibility of variations in regions with small bias, which is of particular significance in this context. It should be noted that the actual values occasionally exceed this range but were truncated to facilitate better visual interpretation.

Overall, the precipitation anomalies are markedly larger in equatorial regions, and it decreases toward the northern and southern extra-tropical regions. Conversely, when assessing the percentage bias for climatic indices, we observe a higher bias in the extra-tropical regions and a lower bias in equatorial regions. In the extra-tropical regions, forecast models demonstrate a reasonable ability to predict total precipitation three months in advance but face challenges when estimating seasonal precipitation patterns and variation throughout the four seasons (for example estimating number of consecutive wet days/dry days). Although, in equatorial regions, the elevated levels of precipitation contribute to a higher anomaly in total precipitation, the models exhibit greater skill in predicting seasonal rainfall patterns, and consequently climatic indices, with a three-month lead time. This is partially attributable to more uniform precipitation patterns throughout the year in equatorial regions.

For indices measured in terms of the number of days, we observe larger bias compared to those representing total rainfall, indicating the models' limited ability to accurately replicate the variation of precipitation throughout the season (Figure 1). Indices that represent magnitude and intensity of precipitation (i.e., precipitation intensity, fraction of precipitation, highest 1-day precipitation amount, and highest 5-day precipitation amount) exhibit lower biases, suggesting that the model’s skill in simulating total seasonal precipitation. The UK-Met and Météo-France models exhibit higher capacity in capturing extreme events, demonstrating favorable performance across various regions when considering 75th and 95th percentile threshold levels. Moreover, even for indices not explicitly based on local thresholds, the biases for the UK-Met and Météo-France models remain lower compared to other models across the globe.



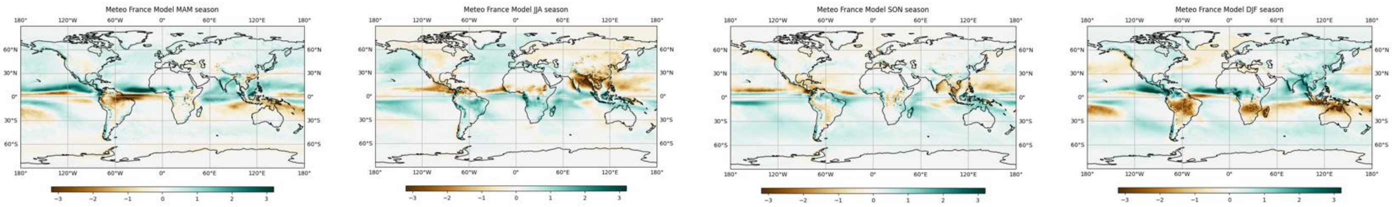


Figure 2. Standardized precipitation anomalies for CMCC (row 1), DWD (row 2), ECCC (row 3), UK-Met (row 4), and Météo-France (row 5) and during MAM (column 1), JJA (column 2), SON (column 3), and DJF (column 4). Anomalies are shown in terms of number of standard deviation departure of forecasts from observations in each grid.

3.2 Global Analysis of ROC Scores

While bias analysis assesses the systematic errors that exist in the forecasts, discrimination analysis is a useful to measure on how well the year-to-year variations in the forecasted values match those in the observations. Measurements of ROC score in Figure 3 and Figures S6-S9 show higher performance of forecast models in the intertropical regions located in Atlantic, Indian ocean, and west Pacific regions (Guimarães et al., 2021; Jie et al., 2017). The skill level varies across different models and seasons across Africa (refer to Table S2 for the complete names of IPCC regions). Notably, the Météo-France and UK-Met models exhibit superior performance during the SON and DJF seasons (i.e., indices with satisfactory, good, very good, and excellent discriminations are more frequent for these models). The exceptional performance of the Météo-France model in African regions has been the subject of discussion in prior studies (Gebrechorkos et al., 2022). Furthermore, when considering the ROC scores, the UK-Met model demonstrates a higher level of skill compared to the other four models in predicting extreme events in several Australian regions. This elevated skill of the UK-Met model is particularly pronounced during the MAM season whereas in JJA, SON, and DJF the skill drops dramatically. The lower performance of ACCESS-S1 forecast model (which is the same model used in UK-Met but with different ensemble generation scheme, ensemble size and the configuration of the system for operational forecasting) over Australia during southern hemisphere summer (DJF) is also concluded in other studies (King et al., 2020).

The prevalence of grids with no discrimination ROC categories is more pronounced in extratropical regions, possibly due to the inherent unpredictability of extratropical variations and limitations within the models when it comes to representing interactions between tropical and extratropical regions, as well as land surface processes (De Andrade et al., 2019). Notably, the CMCC, DWD, and ECCC models are associated with many regions where the models fail to detect any extreme event, as indicated by the absence of discrimination categories in Figures S6-S8. This disparity in extratropical regions is particularly conspicuous when compared to the UK-Met and Météo-France models. Specifically, the divergence is most apparent for wet day indices corresponding to the 75th and 95th percentiles of the reference data. This suggests that the CMCC, DWD, and ECCC models encounter challenges in accurately simulating extreme precipitation events exceeding the 75th and 95th percentiles of the reference dataset across a larger portion of the IPCC regions.

In certain tropical regions, there is a notably higher occurrence of indices falling within at least the satisfactory discrimination category when compared to other regions across all five models. Based on bias and ROC score analysis, the skill of C3S models vary markedly across regions. Hence, it is imperative to establish an impact-based framework for targeted selection of models tailored to address specific climate-related hazards in each region. This approach is crucial in ensuring effective and accurate responses to extreme weather events.

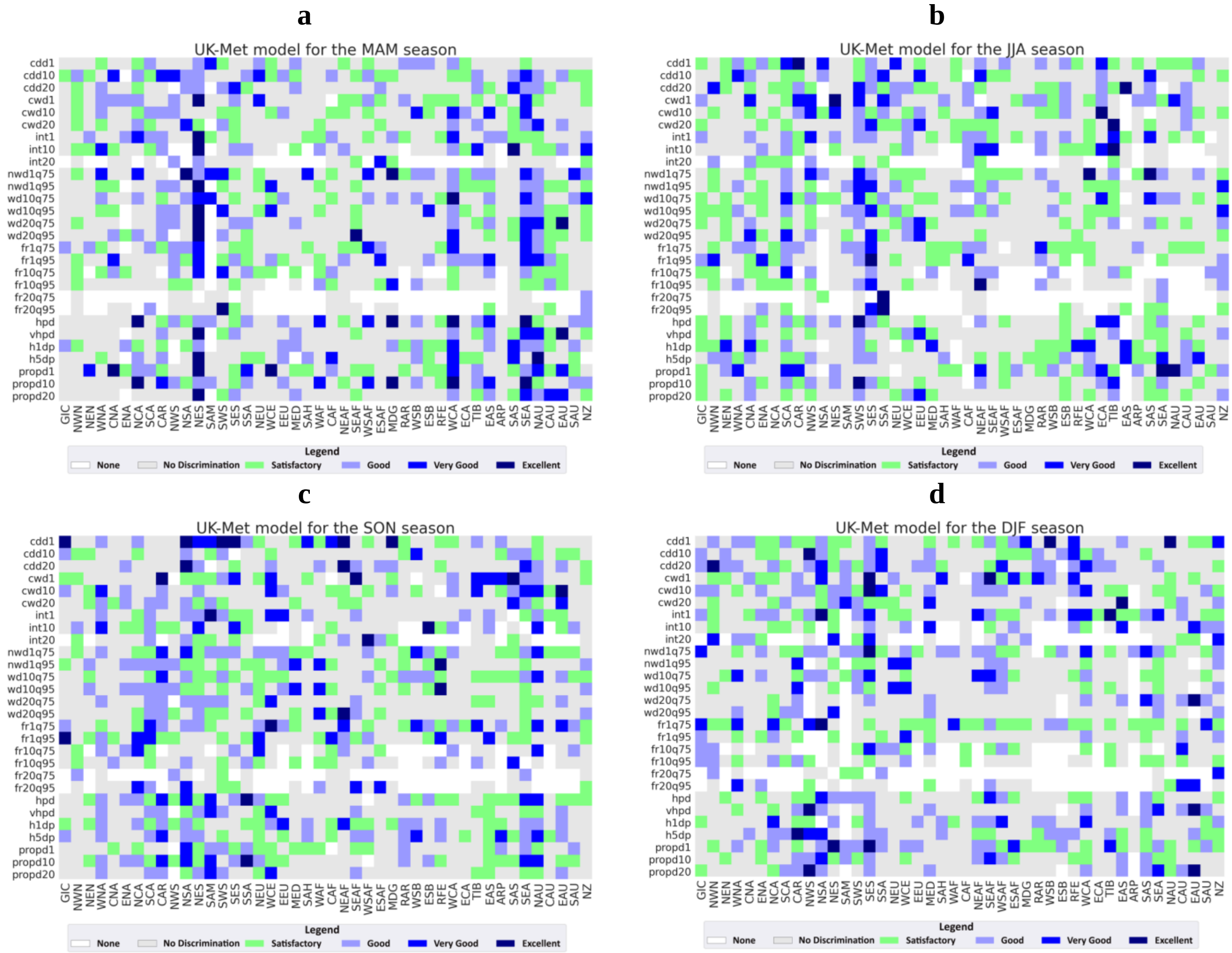


Figure 3. Discrimination levels using categorized ROC score for UK-Met model for a) MAM, b) JJA, c) SON, and d) DJF seasons. Grids that are shaded in white represent regions that either or both reference data and model did not capture any events satisfying the index requirements.

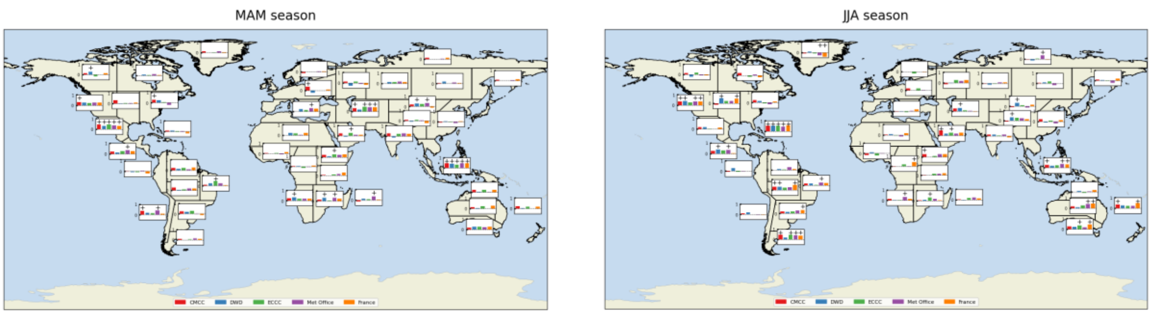
3.3 Global Analysis of Wildfire-related Indices

Many scientific investigations have underscored the notable influence of climatic patterns on the initiation of wildfires (Sharma et al., 2022; Turco et al., 2023). Extended periods characterized by elevated temperatures devoid of precipitation events establish an environment conducive to fire ignition and propagation, intensifying the combustibility of vegetative layers (Alizadeh et al., 2021, 2023). As the duration of consecutive dry days (days without rainfall or with rainfall below a specific threshold) extends, the moisture content of fuel diminishes, increasing its susceptibility to ignition (Abatzoglou & Williams, 2016). In this section, we investigate the cdd1 and propd1 indices to assess the predictive skill of models on a global scale and then within the IPCC region, where wildfires emerge a prominent natural hazard.

The evaluation results reveal that the performance of the models varies across different regions. This regional variation highlights the unique strengths and adaptability of each model, as they excel in response to the specific challenges and requirements posed by distinct geographic areas. The examination of global Kendall's Tau correlation scores based on the cdd1 index reveals that the models' predictive abilities are most pronounced in the intertropical and subtropical zones (Figure 4). While model performance varies notably across different seasons, at least one model displays a notable correlation with the reference datasets within each season and region. Same rule applies to the IPCC regions located in southern parts of Africa and Oceania continents. The predictive skill of the models is also pronounced in the West Central Asia (WCA), East Central Asia (ECA), Tibetan Plateau (TIB), and regions located in Australia and southern America.

Percentage bias results show that most of the models generally underestimate cdd1 across all seasons except for northern regions near the pole during the MAM and DJF seasons which show overestimation (Figure 5). The considerable bias in the southern hemisphere during the DJF season is particularly noteworthy, which can be linked to the shortfall in predictive accuracy of the forecast models in extratropical areas. In contrast, the reduced bias levels, and the significant correlation of models and observations such as south and southeast Asia, particularly during dry seasons, can be attributed to the influence of soil moisture memory on the predictive capabilities of the forecast models in this region (Zhou et al., 2021).

a b



c d

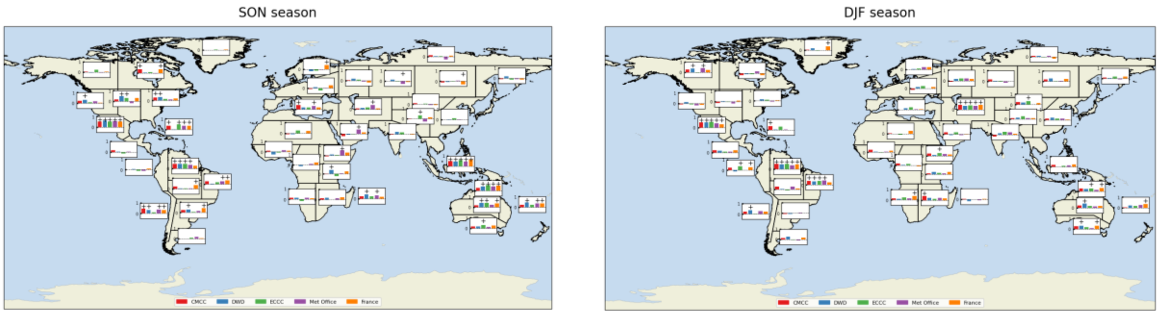


Figure 4. Kendall's Tau coefficients for the cdd1 index in IPCC regions during a) MAM, b) JJA, c) SON, and d) DJF. Models with a significant correlation coefficient at the 0.05 level are marked with a plus sign.

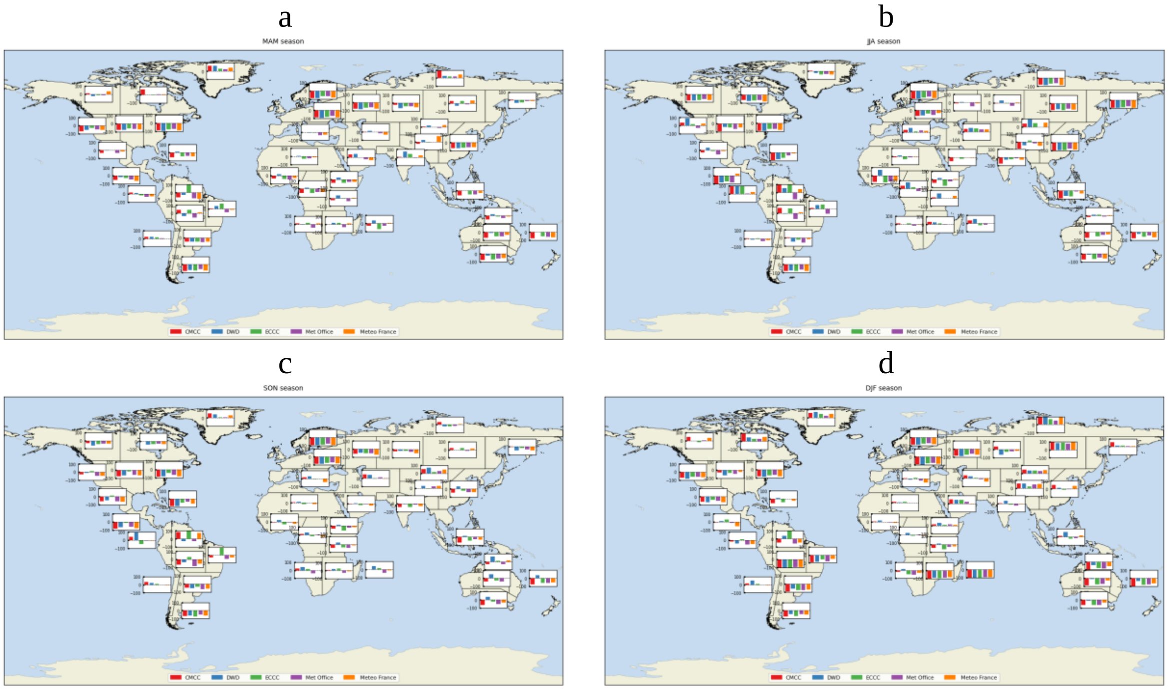


Figure 5. Percent bias for cdd1 index in IPCC regions during a) MAM, b) JJA, c) SON, and d) DJF.

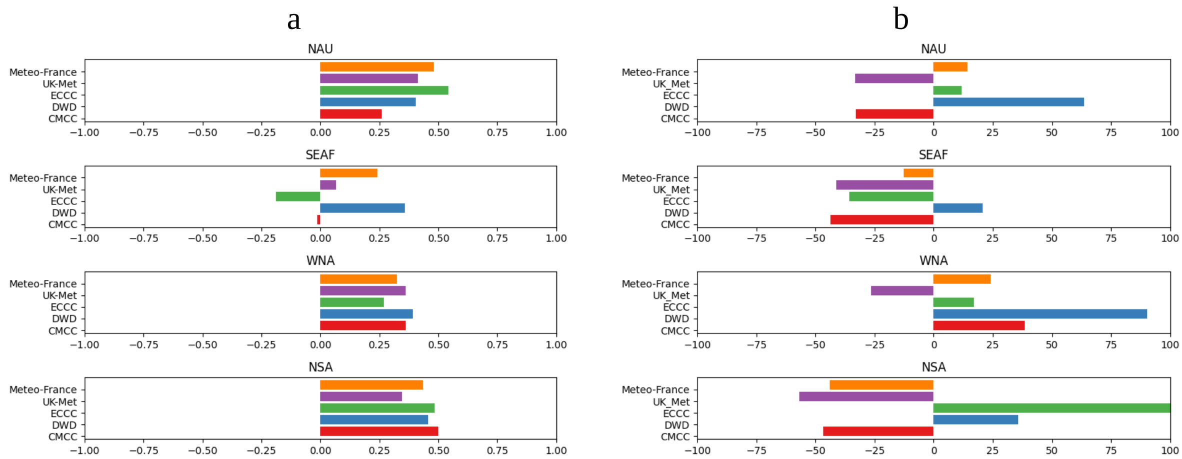
In the southern hemisphere, and in the JJA, SON, and DJF seasons, the correlation coefficient for almost all models on the propd1 index is highly significant, apart from southern Africa where the models show notable skill only in the DJF season (refer to Figure S10). Notably, the bias values for these southern hemisphere regions are both positive and large. This indicates that while the models can accurately represent the seasonal variations, they consistently overestimate the duration of wet days throughout the seasons (refer to Figure S13). In North American regions, the correlation coefficient of models fluctuates seasonally, with noteworthy performance during the JJA and SON seasons. In contrast, for areas across Asia, the models perform significantly well only in the winter season (i.e., DJF). All Models demonstrate strong predictive power throughout all seasons in the WCA region. It's important to note that although the bias is generally large and positive in most northern hemisphere regions, the bias is significantly lower in Asian regions during the DJF season compared to North America and Europe (refer to Figure S10-S11).

3.4 Wildfire-prone Regions: Targeted Forecast Performance Analysis

We now focus on the four regions where the wildfire is a prominent natural hazard: NAU, SEAF, Western North America (WNA), and NSA regions. Within each region, a particular season characterized by an elevated likelihood of wildfire incidence has been designated for subsequent analysis and processing.

In northern Australia, the peak period for wildfire aligns with the dry SON season. From August to December, many regions of Southern Africa experience the onset of their wildfire season therefore we selected the SON season for further analysis. In the United States, wildfire activity is a year-round concern, but the most severe wildfires arise during the summer months (JJA season), particularly in the western regions. In Latin America, the fire season typically commences at the end of January and extends through April (DJF season).

In Figure 6a, it is evident that all models except for CMCC demonstrate a notable correlation with the reference data for the maximum number of consecutive dry days in the NAU region. Notably, Météo-France and ECCC models exhibit the strongest correlation, positioning them as prominent contenders. Furthermore, in Figure 6b both Météo-France and ECCC display lower bias, demonstrating their predictive potential. However, in the Taylor diagram presented in Figure 6c, the ECCC model establishes its supremacy over Météo-France by exhibiting a standard deviation that is more closely aligned with the reference data. The overestimation of precipitation (and consequently underestimation of dry days) in CMCC and UK-Met models over NAU region is visible in Figure 7 where they exceed the 1mm threshold earlier and with steeper slope compared to other models resulting in the underestimation of cdd1 index.



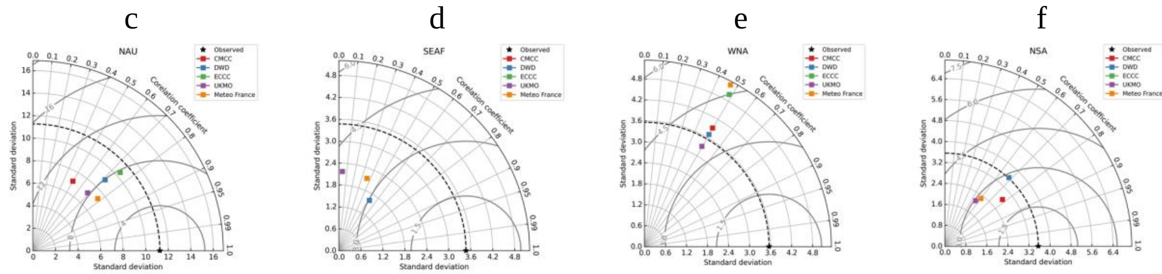


Figure 6. Performance metrics for the maximum consecutive dry days index with 1mm precipitation threshold (cdd1): a) Kendall's Tau coefficient, b) Percentage bias, and Taylor diagram for c) NAU, d) SEAF, e) WNA, and f) NSA regions respectively.

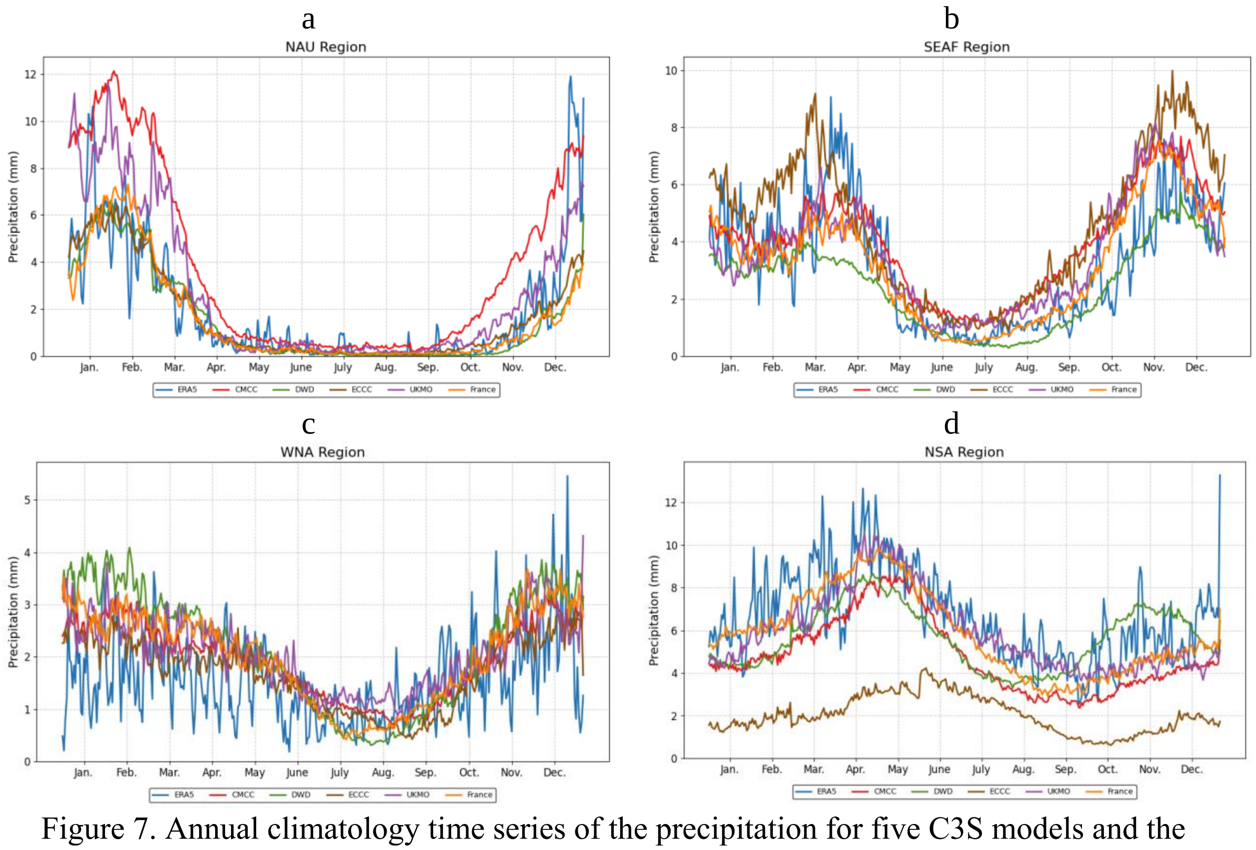


Figure 7. Annual climatology time series of the precipitation for five C3S models and the ERA5 datasets over NAU, SEAF, WNA, and NSA region.

This model selection framework is extended for other three region and the main findings reveal distinct model performance variations in different regions. The ECCC model is particularly strong in forecasting consecutive dry days in the NAU region and closely tracks reference data. In contrast, the DWD model emerges as the top performer in the SEAF and

NSA regions, exhibiting the highest correlation, lower bias, and lower root mean square error in these areas. The UK-Met model excels in the WNA region, demonstrating a close match with the reference dataset's standard deviation. These variations in model performance are attributed to their abilities to simulate significant large-scale climate variabilities such as ENSO, IOD, and north Australian SSTs, which play a crucial role in enhancing prediction skill during the SON season over Australia.

In the domain of predicting the proportion of wet days featuring precipitation exceeding 1mm (propd1), the Météo-France model exhibits superior overall performance in the NAU region, with high correlation values. Additionally, Météo-France demonstrates notably lower bias values, as shown in Figure S12a and S12b. Although Figure S12c's Taylor diagram reveals that Météo-France has a standard deviation that is slightly worse than other models, the bias metric indicates the superiority of the Météo-France model. In the SEAF region Météo-France and DWD model exhibit superior overall performance over other models according to correlation and bias values in Figure S12 and S12b. Based on the Taylor diagram, Météo-France has a standard deviation closer to the reference data. The substantial correlation difference and standard deviation indicates the superiority of Météo-France model. The overestimation of the propd1 index in Météo-France model compared to DWD model can be concluded from Figure 7b where the time series of precipitation of Météo-France is overestimating that of reference datasets during SON season.

For the WNA region, the ECCC model exhibits a higher correlation score but notably high bias values. The absence of substantial correlation values in other models results in the choice of the ECCC model as the most favorable option. In the NSA region, it's evident that the DWD model stands out as the top-performing model. It boasts a high correlation score, the smallest bias when compared to other models, and a standard deviation that closely aligns with the reference data (Figure S12a and S12b and S12f).

3.5 Global Analysis of Flood-related Indices

Consecutive occurrences of extreme precipitation over successive days can significantly elevate the probability of widespread flooding. Many investigations have documented instances of substantial flooding due to consecutive multi-day extreme precipitation incidents (Ávila et al., 2016; Du et al., 2022; Rivoire et al., 2023). To assess the capabilities of the C3S models across IPCC regions, where flooding is a predominant natural hazard, we employed the heavy precipitation days index (hpd) and number of wet days with 10mm precipitation threshold index exceeding the 75th percentiles of the reference dataset (nwd10q75).

According to Figure S13, the correlation of all models with the reference data, as assessed by the hpd index, is most evident in Central and South America, particularly during the JJA and SON seasons. Also, regions in central Africa shows noticeable predictive abilities in the DJF season over CMCC, and Météo-France, and for regions in Australia the CMCC model demonstrate noticeable results throughout the year, excluding the winter season (JJA). Percent bias values are large and negative for all models and across all seasons in northern hemisphere extratropical regions suggesting the inadequacy of models in capturing heavy precipitation days in these areas. In the tropical and sub-tropical regions bias values are markedly smaller especially during MAM, JJA and SON seasons. For the regions located in Australia across all seasons the model bias is large and negative. While the bias for north African regions is relatively smaller in DJF season, for other seasons the bias values are large and negative. The bias values in south American regions in extratropic is relatively lower than those of Africa and Australia across MAM, and DJF the seasons.

In Central and Northern Europe, ECCC, and DWD models show significant predictive abilities during the winter season (DJF). For Europe, the predictive skill of models for extreme precipitation events is higher in DJF; for the other seasons, predictive skill is poor as reported in other studies (Rivoire et al., 2023). The models are generally weak in their predictions for IPCC regions in Asia, apart from the southern and southeastern areas, including WCA, TIB, SAS, and East Asia, where CMCC, UK-Met and Météo-France models exhibit noteworthy predictive skill except for the JJA season. In North America, the models' predictive capacities are lacking across all seasons. Meanwhile, there is a noticeable negative bias in the predictive skill of all models, especially in extratropical regions of both the Southern and Northern Hemispheres as shown in Figure S14. These biases are comparatively smaller in the Western Pacific and Atlantic equatorial regions.

Regarding the nwd10q75 index, its characteristics mirror those of the hpd index, apart from the 75th percentile constraint in the reference data (as seen in Figures S15 and S16). In essence, the patterns of different models across various seasons and regions are nearly identical to those observed for the hpd index. Models that have some, albeit weak, predictive abilities based on the hpd index generally lose their predictive strength when evaluated using the nwd10q75 index.

3.6 Flood-prone Regions: Targeted Forecast Performance Analysis

In the South-East Asia (SEA) monsoon region during JJA (flood season), the UK-Met demonstrates a superior performance compared to other models, exhibiting notably high correlation and lower bias values (Figure 8.a and 8b). Although, the Taylor diagram indicates that UK-Met exhibits a larger standard deviation value compared to other models but the markedly lower bias values make this model the optimal choice here (Figure 8c). This is also evident in the Figure 9a where Météo-France shows overestimation of precipitation, ECCC shows underestimation, while UK-Met, DWD, and CMCC follow the reference precipitation very closely. Overall, in this region the prediction skill is mostly highlighted in the pre-monsoon (April–May) and post-monsoon (October–November), while during monsoon seasons (JJA) skill is poorer because of the monsoon influences on precipitation predictability (Wanthanaporn et al., 2023).

Based on Figure 8, in the Western and Central Europe (WCE) region during the DJF flooding season, the ECCC model exhibits higher significant correlation values compared to other models. However, all models struggle to adequately capture reference data variations, as indicated by high RMSE values and low correlation coefficients. In the South Asia (SAS) region during the JJA season, the UK-Met and CMCC models demonstrate higher correlation values, with the UK-Met model showing positive but statistically insignificant correlation. The UK-Met model outperforms others by exhibiting smaller bias, particularly in comparison to the CMCC model. Therefore, the UK-Met model is favored for the SAS region. In the Central North America (CNA) region during the JJA season, both the UK-Met and Météo-France models exhibit significant correlation coefficients, while all models display large bias values. Once again, the UK-Met model stands out due to its lower bias compared to the other models.

Upon eliminating the constraint associated with the 75th percentile of the reference data in predicting heavy precipitation days, there is an observable reduction in bias values across SEA and SAS regions for the hpd index (Figure S17 and S17b). The correlation values are almost like those of nwd10q75 index. Simultaneously, the standard deviation values become more aligned with the reference data (Figure S17c and S17f). However, it is important to note that despite these changes, the hierarchy of model selection remains consistent. This highlights that the models are relatively less effective at capturing anomalies linked to an increase in the impact of local factors.

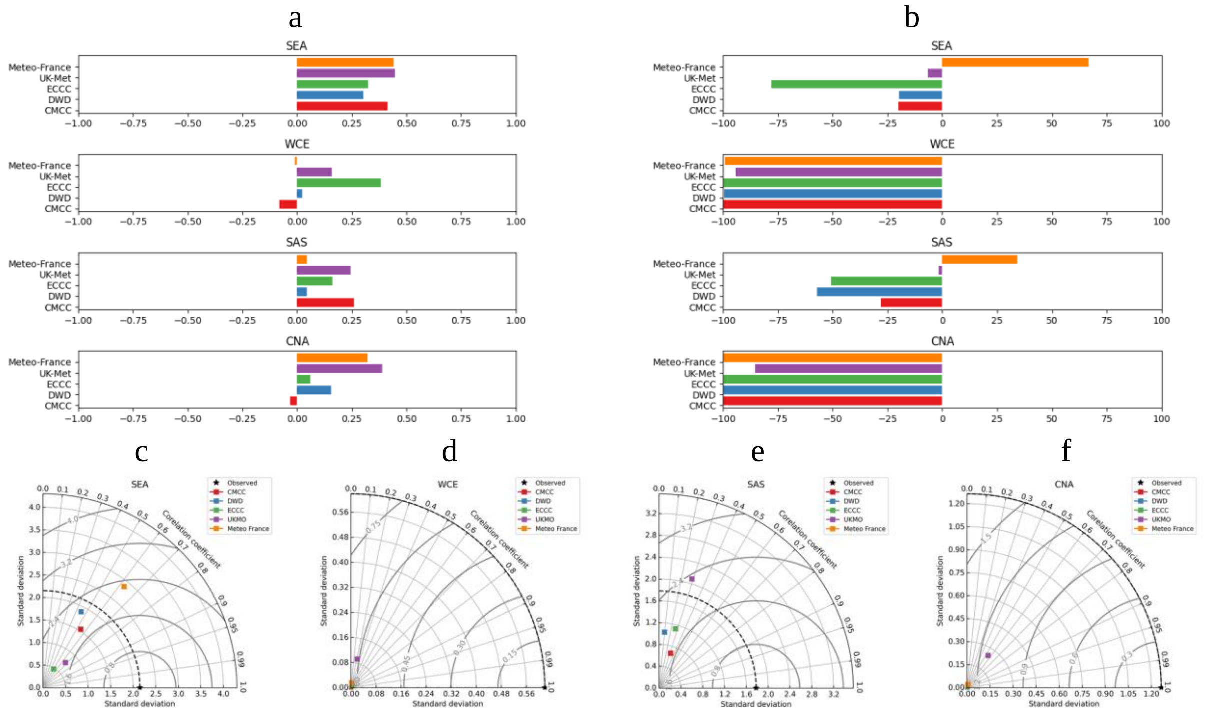


Figure 8. Model performance with respect to the number of heavy precipitation days exceeding 10 mm and the 75th percentiles of the reference dataset (nwd10q75): a) Kendall's Tau coefficient, b) Percentage bias, and Taylor diagram for c) SEA, d) WCE, e) SAS, and f) CNA regions, respectively.

a b

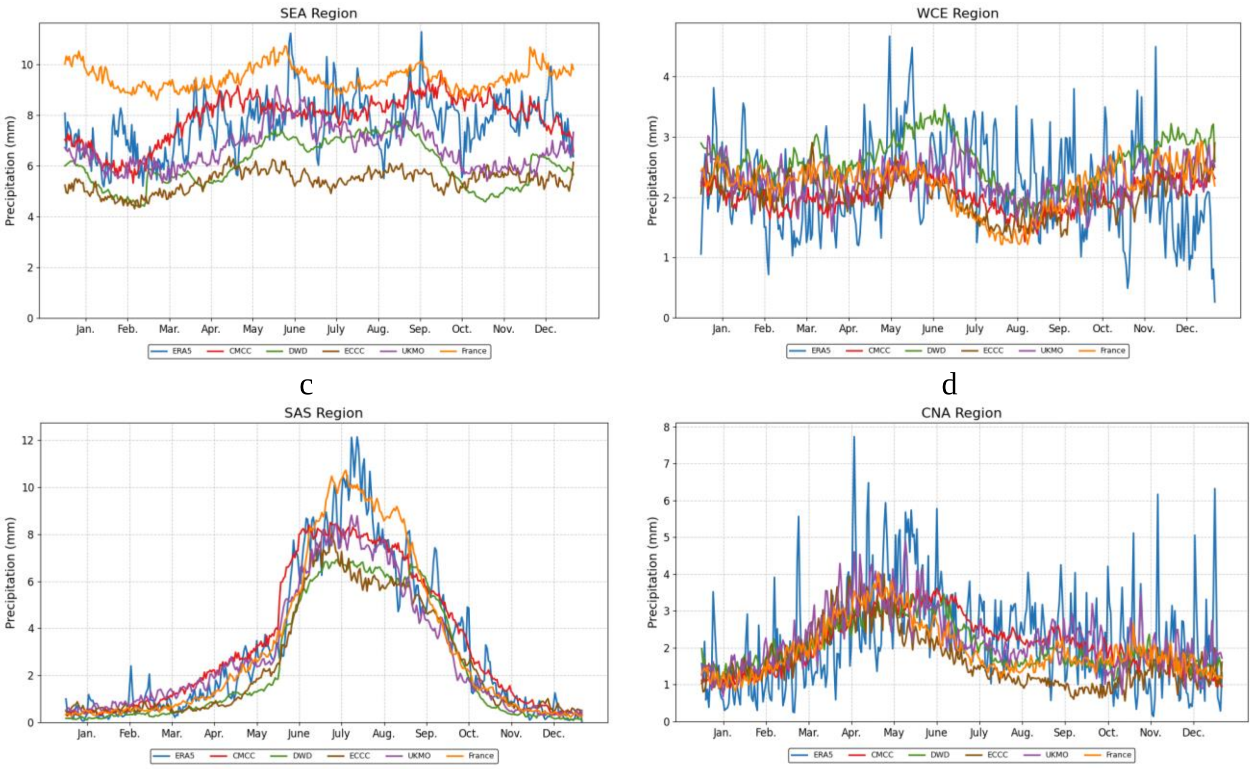


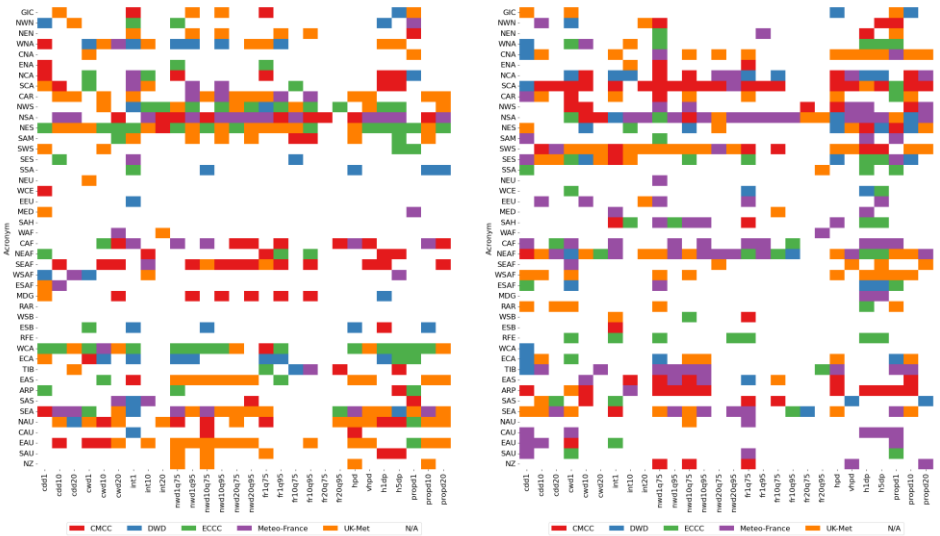
Figure 9. Annual climatology time series of the precipitation for five C3S the ERA5 datasets over a) SEA, b) WCE, c) SAS, and d) CNA region.

3.7 Model Effectiveness Across Seasons and Regions

Analysis of forecasts of extreme precipitation indices over all five models reveals that with increase in the precipitation threshold the model's bias increases, suggesting a lack of skill in modelling severe precipitation events. Correlation scores are lower in extratropical regions as compared to the tropical regions, Likely due to the inherent unpredictability of extratropical variability and model limitations in replicating land surface processes and tropical-extratropical interactions, including the Pacific-South American (PSA) pattern and the Pacific-North American (PNA) pattern, both of which can be influenced by ENSO and the MJO (De Andrade et al., 2019). This is illustrated in Figure 10 where in the extratropical regions models were unable to meet the selection criteria (i.e., having statistically significant correlation while showing low bias) for most of the indices. This figure highlights the superiority of UK-Met and Météo-France for all the four seasons. In the MAM, and JJA season ECCC model has been selected frequently in some regions. The CMCC model is also an effective model after UK-Met and Météo-France by showing higher skill than the rest of models for a considerable number of indices and regions. The frequency of selecting each model at each region and over the 28 climate indices is illustrated in the Figure S18. As an example, in the MAM season at the SEA region for 10 of the indices Météo-France model is selected as a superior model (i.e., having significant correlation while a lower bias compared to other models). For the SON season UK-Met model is the superior model over most of the indices. Another noteworthy example would be NSA region where combination of UK-Met and CMCC models are skilful in predicting extreme events. In MAM season UK-Met meets the selection criteria for 13 of the indices and CMCC pass the selection criteria for 9 of the indices. In the JJA season UK-Met model meets the selection criteria for 13 of the indices while during DJF season the CMCC meets the selection criteria for 12 of the indices.

Over the SON season, UK-Met meet the selection criteria for 12 indices and CMCC for 7 of the remaining indices. It is evident that using these two models over NSA region provides the ability to capture large portion of extreme events. These results highlight the effectiveness of our impact-based framework in capturing variety of extreme climatic events by combination of different models in different season.

a b



c d

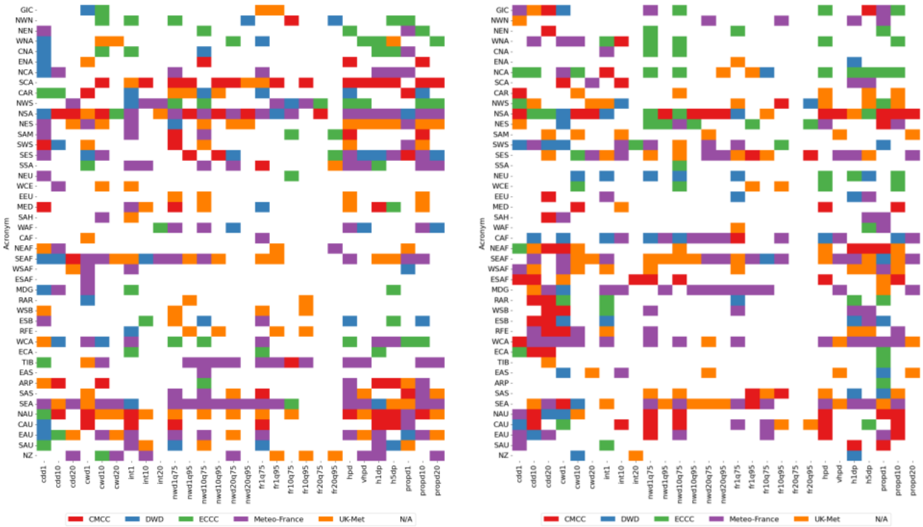


Figure 10. Heat map of the selected models based on statistically significant Kendall's Tau at 0.05 level and percentage bias over IPCC regions and across 28 climate extreme indices for a) MAM, b) JJA, c) SON, and d) DJF seasons.

4. Summary and Discussion

This study's primary objective is to assess the performance of five C3S seasonal forecast models in predicting extreme precipitation events spanning the period from 1993 to 2016. To achieve this, the study involves the extraction of 28 extreme precipitation indices, as defined by the ETCCDI group. These indices are established based on specific precipitation thresholds of 1mm, 10mm, and 20mm, as well as the 75th and 95th percentiles of reference data.

Furthermore, the study employs performance metrics, including Percent Bias and the Kendall Tau Rank Correlation Score, to gauge the models' accuracy in predicting extreme weather events. Also, we evaluate the discrimination capacity of models in discerning extreme events from non-events. To provide a more comprehensive assessment, the research introduces an impact-based framework. This framework is designed to evaluate the models' effectiveness in predicting extreme weather events that have the potential to instigate hazardous conditions such as floods and wildfires. The ERA5 reanalysis precipitation dataset is used as reference. The goal is to identify the most reliable models for targeted impact-based precipitation risk assessments.

A key finding is the consistent underestimation of bias across most extreme climate indices for all models, particularly evident when thresholds for precipitation are high. Notably, the UK-Met and Météo-France models are found to perform better, which has been reported in other studies (De Andrade et al., 2019; McAdam et al., 2022). Despite the prevalent bias, statistically significant correlation are found in tropical and subtropical regions, indicating that the models can reasonably capture the variability of events even if they miss the actual magnitude of the extremes (Vitart et al., 2017).

We develop an impact-based framework to assess the abilities of climate models in detecting extreme events in regions susceptible to cascading natural disasters like wildfires and floods. To evaluate performance in areas prone to wildfires, we employed indices such as cdd1, and propd1. For flood-prone zones, we used nwd10q75 and hpd as primary indices.

In the context of wildfire risk analysis, notable differences in predictive capacities are observed, with specific models showcasing their powers in different regions and for different extreme precipitation indices. In the Northern Australia region (NAU), Météo-France and ECCC models display robust performance in predicting consecutive dry days. In the Southern Africa region (SEAF), the DWD model emerged as a frontrunner for predicting extreme precipitation events. The UK-Met model shows promising results for Western North America (WNA). Lastly, the DWD model shows good performance for the North South America (NSA) region. The analysis reveals models' relative strengths and weaknesses in predicting various precipitation characteristics, providing valuable insights for wildfire-related risk assessments.

For flood-prone regions, the UK-Met model demonstrates superior predictive capabilities in the South-East Asia (SEA) monsoon season (JJA), marked by high correlation and low bias. In Western and Central Europe during the flood season (DJF), the ECCC model excel with notable correlation and comparable bias, despite challenges in capturing reference data variations. In South Asia (SAS) during the JJA season, the UK-Met and CMCC models excel, with the UK-Met showing favourable correlation and low bias.

Our analysis of extreme precipitation indices across multiple models reveals that higher precipitation thresholds correspond to increased model bias, indicating a lack of skill in modelling severe precipitation events. Lower correlation scores in extratropical regions can be attributed to the inherent unpredictability of extratropical variability and the errors stemming from model deficiencies in representing teleconnections (De Andrade et al., 2019). The superiority of UK-Met and Météo-France models throughout all four seasons is emphasized, with ECCC also performing well in specific regions. The ECCC and CMCC models demonstrate effectiveness, following UK-Met and Météo-France, across specific indices and regions. The combined use of models emerges as a successful approach for predicting extreme events across different seasons. These findings underscore the efficacy of the impact-based framework in comprehensively capturing a wide range of extreme climatic events through a strategic combination of diverse models across different seasons.

Data Availability Statement

The data used in this study were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) Copernicus Climate Change Service, specifically from the ERA5 reanalysis dataset and C3S seasonal forecasts. These datasets are publicly available through the Copernicus Climate Data Store (CDS) at https://cds.climate.copernicus.eu under an Open Data Commons Attribution 4.0 International (ODC-BY 4.0) license. To access the data, users can register for a free account on the Copernicus Climate Data Store platform and follow the provided guidelines for data retrieval. The specific seasonal model version numbers used in this study are detailed in the Section 2 of the paper.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, *113*(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

AghaKouchak, A., Huning, L. S., Chiang, F., Sadegh, M., Vahedifard, F., Mazdiyasni, O., Moftakhari, H., & Mallakpour, I. (2018). How do natural hazards cascade to cause disasters? *Nature*, *561*(7724), 458–460. https://doi.org/10.1038/d41586-018-06783-6

AghaKouchak, A., Huning, L. S., Sadegh, M., Qin, Y., Markonis, Y., Vahedifard, F., Love, C. A., Mishra, A., Mehran, A., Obringer, R., Hjelmstad, A., Pallickara, S., Jiwa, S., Hanel, M., Zhao, Y., Pendergrass, A. G., Arabi, M., Davis, S. J., Ward, P. J., … Kreibich, H. (2023). Toward impact-based monitoring of drought and its cascading hazards. *Nature Reviews Earth & Environment*, *4*(8), 582–595. https://doi.org/10.1038/s43017-023-00457-2

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS – global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Alizadeh, M. R., Abatzoglou, J. T., Adamowski, J., Modaresi Rad, A., AghaKouchak, A., Pausata, F. S. R., & Sadegh, M. (2023). Elevation-dependent intensification of fire danger in the western United States. *Nature Communications*, *14*(1), 1773. https://doi.org/10.1038/s41467-023-37311-4

Alizadeh, M. R., Abatzoglou, J. T., Luce, C. H., Adamowski, J. F., Farid, A., & Sadegh, M. (2021). Warming enabled upslope advance in western US forest fires. *Proceedings of* *the National Academy of Sciences*, *118*(22), e2009717118. https://doi.org/10.1073/pnas.2009717118

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2020). Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, *24*(2), 535–559. https://doi.org/10.5194/hess-24-535-2020

Assessment of Intraseasonal to Interannual Climate Prediction and Predictability (p. 12878). (2010). National Academies Press. https://doi.org/10.17226/12878

Ávila, A., Justino, F., Wilson, A., Bromwich, D., & Amorim, M. (2016). Recent precipitation trends, flash floods and landslides in southern Brazil. *Environmental Research Letters*, *11*(11), 114029. https://doi.org/10.1088/1748-9326/11/11/114029

Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, *301*(5633), 636–640. https://doi.org/10.1126/science.1087143

Becker, E., Den Dool, H. V., & Zhang, Q. (2014). Predictability and Forecast Skill in NMME. *Journal of Climate*, *27*(15), 5891–5906. https://doi.org/10.1175/JCLI-D-13-00597.1

Borovikov, A., Cullather, R., Kovach, R., Marshak, J., Vernieres, G., Vikhliaev, Y., Zhao, B., & Li, Z. (2019). GEOS-5 seasonal forecast system. *Climate Dynamics*, *53*(12), 7335–7361. https://doi.org/10.1007/s00382-017-3835-2

Chervenkov, H., & Slavov, K. (2019). STARDEX and ETCCDI Climate Indices Based on E-OBS and CARPATCLIM: Part Two: ClimData in Use. In G. Nikolov, N. Kolkovska, & K. Georgiev (Eds.), *Numerical Methods and Applications* (Vol. 11189, pp. 368–374). Springer International Publishing. https://doi.org/10.1007/978-3-030-10692-8_41

Chervenkov, H., Slavov, K., & Ivanov, V. (2019). STARDEX and ETCCDI Climate Indices Based on E-OBS and CARPATCLIM: Part One: General Description. In G. Nikolov, N. Kolkovska, & K. Georgiev (Eds.), *Numerical Methods and Applications* (Vol. 11189, pp. 360–367). Springer International Publishing. https://doi.org/10.1007/978-3-030-10692-8_40

Chuvieco, E., Pettinari, M. L., Lizundia-Loiola, J., Storm, T., & Padilla Parellada, M. (2018). *ESA Fire Climate Change Initiative (Fire_cci): MODIS Fire_cci Burned Area Pixel product, version 5.1* (3.1) [Application/xml]. Centre for Environmental Data Analysis (CEDA). https://doi.org/10.5285/58F00D8814064B79A0C49662AD3AF537

Committee on Developing a U.S. Research Agenda to Advance Subseasonal to Seasonal Forecasting, Board on Atmospheric Sciences and Climate, Ocean Studies Board, Division on Earth and Life Studies, & National Academies of Sciences, Engineering, and Medicine. (2016). *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts* (p. 21873). National Academies Press. https://doi.org/10.17226/21873

De Andrade, F. M., Coelho, C. A. S., & Cavalcanti, I. F. A. (2019). Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, *52*(9–10), 5451–5475. https://doi.org/10.1007/s00382-018-4457-z

Du, H., Donat, M. G., Zong, S., Alexander, L. V., Manzanas, R., Kruger, A., Choi, G., Salinger, J., He, H. S., Li, M.-H., Fujibe, F., Nandintsetseg, B., Rehman, S., Abbas, F., Rusticucci, M., Srivastava, A., Zhai, P., Lippmann, T., Yabi, I., … Wu, Z. (2022). Extreme Precipitation on Consecutive Days Occurs More Often in a Warming Climate. *Bulletin of the American Meteorological Society*, *103*(4), E1130–E1145. https://doi.org/10.1175/BAMS-D-21-0140.1

Dunn, R. J. H., Donat, M. G., & Alexander, L. V. (2022). Comparing extremes indices in recent observational and reanalysis products. *Frontiers in Climate*, *4*, 989505. https://doi.org/10.3389/fclim.2022.989505

Friedl, M., & Sulla-Menashe, D. (2022). *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061* [dataset]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD12Q1.061

García-Franco, J. L., Lee, C.-Y., Camargo, S. J., Tippett, M. K., Kim, D., Molod, A., & Lim, Y.-K. (2023). Climatology of tropical cyclone precipitation in the S2S models. *Weather and Forecasting*. https://doi.org/10.1175/WAF-D-23-0029.1

Gebrechorkos, S. H., Pan, M., Beck, H. E., & Sheffield, J. (2022). Performance of State-of-the-Art C3S European Seasonal Climate Forecast Models for Mean and Extreme Precipitation Over Africa. *Water Resources Research*, *58*(3). https://doi.org/10.1029/2021WR031480

Giuntoli, I., Fabiano, F., & Corti, S. (2022). Seasonal predictability of Mediterranean weather regimes in the Copernicus C3S systems. *Climate Dynamics*, *58*(7–8), 2131–2147. https://doi.org/10.1007/s00382-021-05681-4

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., & Zhou, Y. (2020). Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment*, *236*, 111510. https://doi.org/10.1016/j.rse.2019.111510

Guimarães, B. D. S., Coelho, C. A. D. S., Woolnough, S. J., Kubota, P. Y., Bastarz, C. F., Figueroa, S. N., Bonatti, J. P., & De Souza, D. C. (2021). An inter-comparison performance assessment of a Brazilian global sub-seasonal prediction model against four sub-seasonal to seasonal (S2S) prediction project models. *Climate Dynamics*, *56*(7–8), 2359–2375. https://doi.org/10.1007/s00382-020-05589-5

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

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., … Vera, C. S. (2020). An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data*, *12*(4), 2959–2970. https://doi.org/10.5194/essd-12-2959-2020

Jie, W., Vitart, F., Wu, T., & Liu, X. (2017). Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database: Simulations of Asian Summer Monsoon in the S2S Database. *Quarterly Journal of the Royal Meteorological Society*, *143*(706), 2282–2295. https://doi.org/10.1002/qj.3085

Khorshidi, M. S., Dennison, P. E., Nikoo, M. R., AghaKouchak, A., Luce, C. H., & Sadegh, M. (2020). Increasing concurrence of wildfire drivers tripled megafire critical danger days in Southern California between1982 and 2018. *Environmental Research Letters*, *15*(10), 104002. https://doi.org/10.1088/1748-9326/abae9e

King, A. D., Hudson, D., Lim, E., Marshall, A. G., Hendon, H. H., Lane, T. P., & Alves, O. (2020). Sub-seasonal to seasonal prediction of rainfall extremes in Australia. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2228–2249. https://doi.org/10.1002/qj.3789

Kumar, A., & Zhu, J. (2018). Spatial Variability in Seasonal Prediction Skill of SSTs: Inherent Predictability or Forecast Errors? *Journal of Climate*, *31*(2), 613–621. https://doi.org/10.1175/JCLI-D-17-0279.1

Lau, W. K.-M., & Waliser, D. E. (2012). *Intraseasonal Variability in the Atmosphere-Ocean Climate System*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13914-7

Lizundia-Loiola, J., Otón, G., Ramo, R., & Chuvieco, E. (2020). A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sensing of Environment*, *236*, 111493. https://doi.org/10.1016/j.rse.2019.111493

Mallakpour, I., Sadeghi, M., Mosaffa, H., Akbari Asanjan, A., Sadegh, M., Nguyen, P., Sorooshian, S., & AghaKouchak, A. (2022). Discrepancies in changes in precipitation characteristics over the contiguous United States based on six daily gridded precipitation datasets. *Weather and Climate Extremes*, *36*, 100433. https://doi.org/10.1016/j.wace.2022.100433

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McAdam, R., Masina, S., Balmaseda, M., Gualdi, S., Senan, R., & Mayer, M. (2022). Seasonal forecast skill of upper-ocean heat content in coupled high-resolution systems. *Climate Dynamics*, *58*(11–12), 3335–3350. https://doi.org/10.1007/s00382-021-06101-3

Modaresi Rad, A., Kreitler, J., Abatzoglou, J. T., Fallon, K., Roche, K. R., & Sadegh, M. (2022). Anthropogenic stressors compound climate impacts on inland lake dynamics: The case of Hamun Lakes. *Science of The Total Environment*, *829*, 154419. https://doi.org/10.1016/j.scitotenv.2022.154419

Moron, V., & Robertson, A. W. (2020). Tropical rainfall subseasonal-to-seasonal predictability types. *Npj Climate and Atmospheric Science*, *3*(1), 4. https://doi.org/10.1038/s41612-020-0107-3

Nobakht, M., Saghafian, B., & Aminyavari, S. (2021). Skill Assessment of Copernicus Climate Change Service Seasonal Ensemble Precipitation Forecasts over Iran. *Advances in Atmospheric Sciences*, *38*(3), 504–521. https://doi.org/10.1007/s00376-020-0025-7

Rivoire, P., Martius, O., Naveau, P., & Tuel, A. (2023). Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over Europe. *Natural Hazards and Earth System Sciences*, *23*(8), 2857–2871. https://doi.org/10.5194/nhess-23-2857-2023

Roy, T., He, X., Lin, P., Beck, H. E., Castro, C., & Wood, E. F. (2020). Global Evaluation of Seasonal Precipitation and Temperature Forecasts from NMME. *Journal of Hydrometeorology*, *21*(11), 2473–2486. https://doi.org/10.1175/JHM-D-19-0095.1

Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasni, O., Sanders, B., Matthew, R., & AghaKouchak, A. (2018). Multihazard Scenarios for Analysis of Compound Extreme Events. *Geophysical Research Letters*, *45*(11), 5470–5480. https://doi.org/10.1029/2018GL077317

Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., Wood, E. F., Prudhomme, C., Rees, G., Houghton-Carr, H., Fry, M., Smith, K., Watts, G., Hisdal, H., Estrela, T., Buontempo, C., Marx, A., & Kumar, R. (2019). Hydrological Forecasts and Projections for Improved Decision-Making in the Water Sector in Europe. *Bulletin of the American Meteorological Society*, *100*(12), 2451–2472. https://doi.org/10.1175/BAMS-D-17-0274.1

Schiavina, M., Freire, S., & MacManus, K. (2023). *GHS-POP R2023A - GHS population grid multitemporal (1975-2030)* [dataset]. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall’s Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sharma, A. R., Jain, P., Abatzoglou, J. T., & Flannigan, M. (2022). Persistent Positive Anomalies in Geopotential Heights Promote Wildfires in Western North America. *Journal of Climate*, *35*(19), 6469–6486. https://doi.org/10.1175/JCLI-D-21-0926.1

Shukla, J., Marx, L., Paolino, D., Straus, D., Anderson, J., Ploshay, J., Baumhefner, D., Tribbia, J., Brankovic, C., Palmer, T., Chang, Y., Schubert, S., Suarez, M., & Kalnay, E. (2000). Dynamical Seasonal Prediction. *Bulletin of the American Meteorological Society*, *81*(11), 2593–2606. https://doi.org/10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2

Tapiador, F. J., Roca, R., Del Genio, A., Dewitte, B., Petersen, W., & Zhang, F. (2019). Is Precipitation a Good Metric for Model Performance? *Bulletin of the American Meteorological Society*, *100*(2), 223–233. https://doi.org/10.1175/BAMS-D-17-0218.1

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w

Thielen, J., Bartholmes, J., Ramos, M.-H., & De Roo, A. (2009). The European Flood Alert System – Part 1: Concept and development. *Hydrology and Earth System Sciences*, *13*(2), 125–140. https://doi.org/10.5194/hess-13-125-2009

Turco, M., Abatzoglou, J. T., Herrera, S., Zhuang, Y., Jerez, S., Lucas, D. D., AghaKouchak, A., & Cvijanovic, I. (2023). Anthropogenic climate change impacts exacerbate summer forest fires in California. *Proceedings of the National Academy of Sciences*, *120*(25), e2213815120. https://doi.org/10.1073/pnas.2213815120

Villarini, G., Vecchi, G. A., Knutson, T. R., Zhao, M., & Smith, J. A. (2011). North Atlantic Tropical Storm Frequency Response to Anthropogenic Forcing: Projections and Sources of Uncertainty. *Journal of Climate*, *24*(13), 3224–3238. https://doi.org/10.1175/2011JCLI3853.1

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., … Zhang, L. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *Npj Climate and Atmospheric Science*, *1*(1), 3. https://doi.org/10.1038/s41612-018-0013-0

Wanders, N., & Wood, E. F. (2016). Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters, 11*(9), 094007. https://doi.org/10.1088/1748-9326/11/9/094007

Wanthanaporn, U., Supit, I., Van Hove, B., & Hutjes, R. W. A. (2023). *Analysis of seasonal climate and streamflow forecasts performance for Mainland Southeast Asia* [Preprint]. Water Resources Management/Modelling approaches. https://doi.org/10.5194/hess-2023-56

Xu, L., Chen, N., Chen, Z., Zhang, C., & Yu, H. (2021). Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. *Earth-Science Reviews*, *222*, 103828. https://doi.org/10.1016/j.earscirev.2021.103828

Xue, P., Malanotte-Rizzoli, P., Wei, J., & Eltahir, E. A. B. (2020). Coupled Ocean-Atmosphere Modeling Over the Maritime Continent: A Review. *Journal of Geophysical Research: Oceans*, *125*(6). https://doi.org/10.1029/2019JC014978

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

Zhou, Y., Zaitchik, B. F., Kumar, S. V., Arsenault, K. R., Matin, M. A., Qamer, F. M., Zamora, R. A., & Shakya, K. (2021). Developing a hydrological monitoring and sub-seasonal to seasonal forecasting system for South and Southeast Asian river basins. *Hydrology and Earth System Sciences*, *25*(1), 41–61. https://doi.org/10.5194/hess-25-41-2021