

ALMA MATER STUDIORUM · UNIVERSITY OF BOLOGNA

---

---

Department of Physics and Astronomy

Department of Biological, Geological and Environmental Sciences

Master Degree in Science of Climate

# Hybrid Statistical–Dynamical Methods for Decadal Prediction of African Precipitation Variability

**Supervisor:**

**Dr. Paolo Ruggieri**

**Submitted by:**

**Valerio Ippoliti**

Academic Year 2024/2025

## Abstract

Decadal climate prediction (DCP) seeks to forecast climate evolution from years to a decade ahead, filling the critical gap between seasonal forecasts and long-term projections. Predictability arises from slowly evolving components of the Earth system, including ocean variability, greenhouse gases, aerosols, and solar forcing, which intermittently constrain chaotic atmospheric evolution. Despite advances, challenges such as model drift, underestimation of predictable signals, and the signal-to-noise paradox limit skill, particularly for precipitation.

Here, I explore methods to enhance decadal rainfall predictability across West, Central, and East Africa. For boreal winter (DJFM), a two-stage post-processing approach adjusts multimodel ensemble to retain members consistent with predictable North Atlantic Oscillation (NAO) signals, improving low-frequency circulation and precipitation representation. For boreal summer (JJAS), hybrid forecasting frameworks combine physics-based Earth System Models with machine learning. Ensemble subsetting guided by El Niño Southern Oscillation (ENSO) mode-matching and Convolutional Neural Network diagnostics of spatiotemporal teleconnections reduces noise and enhances skill.

Results show that these post processing methodologies gain some improvements regarding interannual and decadal rainfall predictability across the Sahel, Guinea Coast regions, Central Africa and East Africa, further providing an assessment of the impact of low-frequency mode of oceanic-atmospheric variability on it, highlighting the potential of DCP to contribute actionable multi-year climate information for water resources, extreme events, and climate adaptation.

# 1 Introduction

## 1.1 Foundations of Decadal Climate Prediction

Predictability in Decadal Climate Predictions (DCPs) crucially relies on both initial conditions and external forcings. Slowly evolving coupled interactions among Earth's components can serve to constrain the chaotic evolution of the atmosphere. Such crucial coupled processes are often intermittent and have regionally dependent impacts (Mariotti et al., 2020).

The intermittency of opportunities for skillful long-lead predictions is already evident from long-lead weather forecasts, with those initiated from some atmospheric flow configurations being more skillful than others. Intrinsic predictability levels of different atmospheric states can be estimated from forecast ensembles. In reliable and well-calibrated prediction systems, the spread among ensemble members quantifies forecast uncertainty and the underlying predictability. At times when the forecast uncertainties are relatively small we deduce that the atmosphere is more predictable; higher predictability means a window of opportunity for more skillful forecasts. If we can identify processes and conditions that lead to these predictable conditions, we can potentially provide more useful S2S (subseasonal-to-seasonal) and S2D (seasonal-to-decadal) forecasts (Figure 1) (Mariotti et al., 2020).

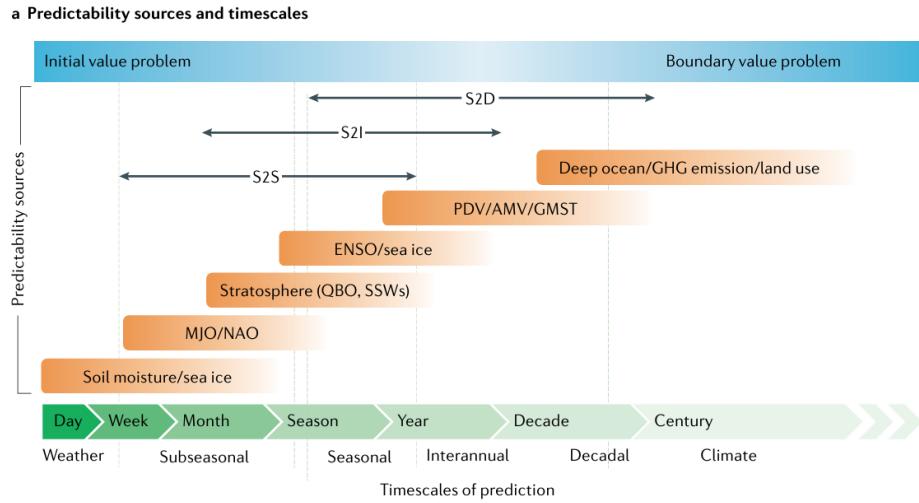


Figure 1: Predictability sources and timescales Meehl et al., 2021.

Although individual weather events cannot be predicted more than a couple of weeks ahead, slowly varying predictable components of the climate system, including ocean variability and changes in radiative forcing from greenhouse gases, aerosols and solar variations, influence the frequency, duration and intensity of weather events over the coming seasons to years. Hence some aspects of climate, for example, multi-year atmospheric circulation changes, the frequency of extreme weather events and total amounts of rainfall, are potentially predictable over the coming decade. Such forecasts will not be perfectly deterministic because uncertainties are inevitable due to the chaotic nature of

the atmosphere, and errors will be introduced by imperfect climate models and imperfect knowledge of the initial state of the climate system (Smith et al., 2019).

An important advantage of decadal over centennial predictions is that the likely skill can potentially be quantified in tests of past cases, or hindcast experiments. In these, the model is used to “forecast” a historical period, but only using data that would have been available prior to this period (though this is not quite an independent test because these data generally are used to develop, test, and tune the model). The accuracy of the model can then be assessed by comparing with what actually happened. A large set of such hindcasts is typically made in order to obtain a robust estimate of the likely skill and reliability of actual forecasts (Meehl et al., 2009) .

## 1.2 Hindcasts and Model Performance

The Coupled Model Intercomparison Project phases 5 and 6 (CMIP5 and CMIP6) with the Decadal Climate Prediction Project (DCPP, Boer et al., 2016) provide initialized hindcasts from multiple modeling centers, demonstrating skill for surface temperature and large-scale SST patterns. However, challenges persist in initialized models which often underestimate the amplitude of predictable signals, especially atmospheric circulation and precipitation.

## 1.3 Ensemble Generation and Initialization

Weather and climate predictions are well known to be sensitive to small perturbations in the initial state, and an ensemble of initial conditions and subsequent predictions is typically generated to investigate this. Ensemble generation can utilize different methodologies for perturbation, extending from ensemble Kalman filter-type assimilation methods, different start days for the initial state of the atmosphere around the time of the prediction, application of breeding methods to generate optimal initial perturbation, or variations on those methods. This diversity of approaches results from different philosophical approaches to the initialization and ensemble generation problem (Meehl et al., 2014).

Decadal climate variations are driven by both internal variability (e.g., ocean–atmosphere oscillations like AMV, NAO, PDO) and external forcings (e.g., greenhouse gases, aerosols, volcanic eruptions, solar changes). Decadal predictions are initialized with observations of the ocean and the atmosphere. This has two consequences: 1) It aligns the model’s internal state with that of the real climate system, so the model and observed internal variability “start from the same point”; this gives a chance to forecast the evolution of internal variability. 2) It may also help correct biases in how the model represents responses to external forcings, if those forcings leave fingerprints in the system that observation can detect and initialization can correct (Smith et al., 2019).

Volcanic eruptions are not predictable. Once they occur and their impact is captured in the initial conditions, they can enhance predictability across a variety of time scales. In addition, anthropogenic aerosols may also present a source of decadal predictability.

Long-term increases in greenhouse gas concentrations and associated warming trends are a major source of predictability for surface air temperature and other surface meteorological variables, affecting the skill of climate forecasts at S2S time scales and beyond (Mariotti et al., 2020).

One of the biggest technical challenges affecting the quality of decadal climate predictions is the initialization of the model from observations to start a decadal prediction (Goddard et al., 2013).

Modeling groups are actively exploring different techniques and methodologies for initializing decadal climate predictions. These include either partial or fully coupled assimilation of ocean and/or atmospheric observations, forcing the ocean with atmospheric observations, and, additionally, full-field or anomaly initialization. Evaluations of two of these methods, the three-dimensional initialization of the ocean with observations and the use of observed surface forcing to initialize the ocean (S. Yeager et al., 2012a) have shown that the latter may constitute a simple but successful alternative strategy for initialization, especially over the extratropical regions (Doblas-Reyes et al., 2011; Meehl et al., 2014).

Assessing the impact of initialization is useful, because it helps diagnose whether skill in decadal prediction comes more from whether skill is dominated by external factors or whether internal variability could be playing a role, thereby highlighting the most important factors to be considered in order to reduce uncertainties in forecasts (Smith et al., 2019). Initialization has the biggest impact where internal variability strongly influences climate. It is important to note that initialization may also improve the response to external factors where these leave fingerprints, especially in the ocean, that are imperfectly simulated by models but can be corrected with observations (Smith et al., 2019). In uninitialized forecasts, models generally capture the long-term warming trend driven by greenhouse gases but perform poorly in predicting the timing of regional fluctuations, as they do not incorporate the real ocean and atmosphere state. By contrast, initialized forecasts retain the ability to capture the long-term warming trend while also better representing internal variability, such as Atlantic shifts, the NAO, and regional rainfall swings. This results in improved skill for regional climate features, including Sahel rainfall and North Atlantic pressure. This reveals significant benefits of initialization including for temperature over Europe, Sahel rainfall, and North Atlantic pressure. However, the overall patterns of skill for temperature, precipitation and pressure are largely captured by the uninitialized simulations, and improvements from initialization generally occur in regions where the uninitialized simulations already have some skill. This could arise because: (1) improved skill arises from predicting internal variability in regions where there is an externally forced response; (2) externally forced skill in these regions is largely incidental and skill in initialized predictions arises mainly from internal variability; (3) the variability is predominantly externally forced and improved skill arises from correcting the modeled response to external factors. The first situation is particularly expected where there is a long term trend driven by slow variations in greenhouse gases. However, uninitialized simulations also capture aspects of the variability around the linear trend highlighting the need for improved understanding of the roles of other external factors,

including solar variations and volcanic and anthropogenic aerosols. (Smith et al., 2019).

## 1.4 Bias Correction and de-drifting

There are also various methods to deal with model drift away from the observed initial state. Full-field initialization brings the ocean model state close to observations and the model then drifts toward its systematic error state during the prediction and requires bias adjustment in predictions (discussed below). Anomaly initialization adds the anomalous component of the observed state to the model climatology to minimize the drift during the prediction. A comparison of the two methods shows that the former generally produces more skillful predictions on the seasonal time scale (Magnusson et al., 2013), though the latter provides more skillful predictions in hindcasts where it has been tested (Smith et al., 2012). (Hazeleger et al., 2013) find there is no significant difference between full-field and anomaly initialization in decadal prediction skill. However, the anomaly initialization method can produce mismatches between the observational anomalies and the model climatology in some regions (e.g., in sharp Gulf Stream gradient locations). A majority of modeling groups at present are leaning toward full-field initialization, but further evaluations are necessary with more models to draw definitive conclusions as to the best initialization technique (Meehl et al., 2014). Biases in decadal climate predictions that develop as a function of time come from a range of sources, including model drift from the observed initial state to its own preferred state, which is a product mainly of its own set of systematic errors that can happen quite rapidly (sometimes called initialization shock), inability to realistically simulate the natural modes of interannual-to multidecadal variability, uncertain future levels of radiative forcings (such as volcanic eruptions and aerosols), and insufficient and imperfect observations. Additionally, there is uncertainty from insufficient sampling of the natural variability owing to both the short history of the hindcasts as well as the limited number of predictions. The rate and spatial pattern of bias growth may give useful information about the physical processes that lead to prediction error and may allow targeted model improvements. Nevertheless, current predictions generally must attempt to remove this bias in order to be useful in predicting small signals. Some of the issues in adjusting prediction biases are given by WCRP (International CLIVAR Project Office, 2011). As shown in figure 2 (Meehl et al., 2014), the result for full-field bias adjustment is equivalent to calculating a predicted climatological average for each forecast range and considering the forecast anomalies obtained by subtracting the average. The same calculation is done for the observations, and the anomalies are compared. The model drifts from its observed initial states (dashed lines) toward its preferred climate state, which is closer to the uninitialized model state (gray line) with a stronger trend than is observed. Systematic model errors may be removed by subtracting the average rate of drift over all hindcasts. In this example, however, removal of the mean bias produces states that remain biased low early in the period and biased high later in the period, indicating the bias adjustments are too small early on, and too large later (solid line).

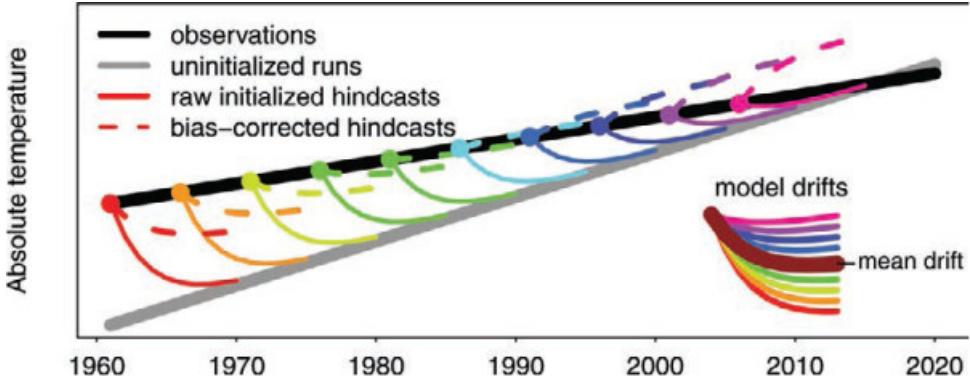


Figure 2: Full-field bias adjustment (Meehl et al., 2014)

(Smith et al., 2013) show that fully field initialization better captures precipitation anomalies in teleconnected regions compared to anomaly inintilaiztion, as initializing full observed fields starts from a state already consistent with ENSO events.

## 1.5 Spatial and Time Averaging

Although other criteria could be used, a balance between skill improvement and signal-to-noise retention suggests that  $5^{\circ}$ latitude  $\times$   $5^{\circ}$ longitude represents a reasonable scale for smoothing precipitation, and  $10^{\circ}$ latitude  $\times$   $10^{\circ}$ longitude for temperature (Räisänen, 2011). For time averages, the recommendation is to evaluate prediction skill for 2–5 and 6–9 year lead-time averages to represent interannual time scales (but not to include year 1 so as to exclude skill from the seasonal time scale) and for 2–9 year lead-time averages to assess subdecadal-scale predictions (Goddard et al., 2013). The reason to show different averaging periods is that one may be tempted to look at skill for decadal averages from these hindcasts and assume that level of quality applies throughout the period . However, the 4-year average forecasts (years 2–5 and 6–9) within the decade are likely to have lower skill, and there are potential differences between those also. Thus these four cases are a minimum to show skill dependence on averaging and lead time (Smith et al., 2007).

The first year of the prediction overlaps with currently available seasonal-to-interannual predictions and should be most predictable due to proximity to initial conditions. The year 2–5 average still represents the interannual timescale, but discards year 1; it is likely still dominated by year-to-year variability and less by the climate change signal.. The year 2–9 average represents decadal-scale climate and excludes the large contribution from year 1. In conclusion, the skill of 6–9 year average predictions is compared with 2–5 year averages to assess dependence on lead time (Goddard et al., 2013).

## 1.6 Signal-to-Noise Paradox

Climate models frequently exhibit a surprising behavior: they capture observed variability well, yet the amplitude of their predictable signals is systematically too weak. In the Atlantic sector, ensemble forecasts often correlate more strongly with observations than with their own ensemble members (Scaife and Smith, 2018). This signal-to-noise paradox

has been observed across multiple prediction systems and for diverse sources of climate variability (Scaife and Smith, 2018). The paradox does not arise from an overestimation of variability, indeed total model variance generally matches observations, but from an underrepresentation of the fraction that is predictable, often by a factor of two or more. In essence, models can reflect climate fluctuations yet fail to express the true strength of predictable signals, aligning with reality while appearing weakly predictive (Eade et al., 2014; Scaife and Smith, 2018). Potential causes include missing extratropical ocean–atmosphere coupling, unresolved eddy feedbacks, misrepresented teleconnections, and parameterization errors (W. Zhang, Kirtman, et al., 2021).

This underestimation has significant implications for the interpretation and use of climate forecasts. Metrics that rely on absolute errors or ensemble spread—such as RMSE, MSSS (Goddard et al., 2013), or probabilistic measures like reliability and Brier skill score—can systematically underestimate the skill that is potentially available. While postprocessing can amplify predictable signals and correct for these biases (Eade et al., 2014; Sansom et al., 2016), raw model outputs may give a misleading picture. Measures such as anomaly correlation coefficient, which are insensitive to signal amplitude, remain robust indicators when ensembles are sufficiently large.

The paradox also challenges conventional methods for estimating predictability. Comparing a single ensemble member to observations is often treated as an upper bound on skill, yet if models underestimate signal strength, this approach masks real-world predictability. The actual climate system may therefore be more predictable than suggested by models, and relying solely on ensemble means can create overconfidence in forecasts while underestimating uncertainty (Griffies and Bryan, 1997; Koenigk and Mikolajewicz, 2009). Event attribution studies face a similar risk: models that underestimate the response to external forcing, such as SST changes, can exaggerate the role of internal variability and diminish the perceived impact of anthropogenic influences (Branstator and Teng, 2010; Schaller et al., 2016; Stott et al., 2016). In regions like the North Atlantic, this effect is especially pronounced, potentially biasing conclusions about human influence on extreme events.

The paradox also has consequences for projections of regional climate change. Large ensembles frequently suggest that internal variability dominates future uncertainty (Deser et al., 2012), but if forced signals are underestimated, the true contribution of external forcing may be larger, leading to an overemphasis on natural variability. Paradoxically, the signal-to-noise mismatch simultaneously reveals hidden predictability. Even when signal amplitudes are weak, phenomena such as the North Atlantic Oscillation (NAO) can be forecasted skillfully (Scaife et al., 2014), provided that large ensembles are used to extract the signal and postprocessing amplifies it.

Addressing the paradox directly by adjusting models to match observed signal-to-noise characteristics could strengthen forced responses, improve the realism of teleconnections, reduce ensemble size requirements, and enhance forecast skill and reliability across multiple regions and variables. The signal-to-noise paradox thus highlights a fundamental insight: while models capture climate variability, they systematically underestimate predictable

signals. Recognizing and correcting this discrepancy is not only critical for accurate skill assessment, event attribution, and projections, but also opens a pathway to unlock latent predictability that models already contain.

## 1.7 Improving Decadal Precipitation Predictions through NAO-Based Post-Processing of Climate Models

Although the signal-to-noise paradox highlights a potentially serious problem with climate models, its discovery helps to reveal that skilful forecasts are now possible for some phenomena, including the NAO (North Atlantic Oscillation) (Scaife and Smith, 2018). Recent studies suggest that the relatively weak signal-to-noise ratios in current decadal prediction systems can be mitigated through the use of large ensemble approaches, which enhance skill via the noise-dampening effect (N. Dunstone et al., 2016; Eade et al., 2014; Scaife et al., 2014). (Scaife and Smith, 2018) note that a large ensemble is required in order to extract the maximum predictable signal and postprocessing is needed to boost its magnitude (Eade et al., 2014). The North Atlantic stands out as a key region for decadal predictability, where initialized predictions show notable improvements in near-surface air temperature compared to uninitialized simulations (Doblas-Reyes et al., 2013; Smith et al., 2010). This enhanced skill has been linked to the improved representation of the Atlantic Meridional Overturning Circulation (AMOC), which plays a central role in modulating heat transport and influencing climate variability across the basin (Pohlmann et al., 2013; Robson et al., 2014; S. Yeager et al., 2012b). Variations in Atlantic temperatures have also been shown to affect atmospheric circulation and rainfall patterns in surrounding regions, including mean sea level pressure in the tropical Atlantic and precipitation in the Sahel and West Africa (N. J. Dunstone et al., 2011; R. Zhang and Delworth, 2006). The research proposed by Smith et al., 2020 analysed six decades of climate model data and suggests decadal variations in North Atlantic atmospheric pressure patterns during boreal winter are highly predictable, enabling advanced warning of whether winters in the coming decade are likely to be stormy, warm and wet or calm, cold and dry. The authors demonstrate that decadal variations in the North Atlantic winter climate are more predictable than current models suggest. They identify that suffering from singal to noise paradox existing models underestimate the predictable signal of the NAO by an order of magnitude. Current models substantially underestimate the predictable signal of the NAO, requiring 100 times more ensemble members than perfect models to extract this signal. To address this, they implement a two-stage post-processing approach: variance adjustment of the ensemble-mean NAO forecast, followed by selection of ensemble members close to the adjusted mean. This method improves decadal predictions of winter climate over the region of interest. The results highlight the need to understand and correct the low signal-to-noise ratio in current models (Scaife and Smith, 2018) , offering the potential to reduce uncertainties in regional climate projections on decadal and longer timescales.

## 1.8 West African Monsoon Variability and Predictability

The West Africa Monsoon (WAM), a boreal summer (June to September, JJAS) monsoon system, exhibits one of the largest signals of global climatic variability (Martin et al., 2014) and is sensitive to both local and remote influences (Folland et al., 1986; Fontaine, 1996; Fontaine et al., 1998). Its variability spans a wide range of timescales, from intraseasonal to interdecadal, reflecting the complex interplay between oceanic, atmospheric, and land-surface processes.

Interannual variability is primarily associated with precipitation changes over the coastal regions of the Gulf of Guinea (approximately 5–10°N), for which equatorial Atlantic sea surface temperature (SST) anomalies—commonly referred to as the Atlantic Niño—constitute the dominant oceanic forcing (Janicot et al., 1998; Losada et al., 2012; Okumura and Xie, 2004). At decadal scales, the variability is most pronounced in the semiarid Sahel (approximately 10–20°N), located along the southern edge of the Sahara, where alternating wet and dry periods have been documented since the early twentieth century (Folland et al., 1986; Fontaine et al., 1995; Nicholson, 1993). These decadal fluctuations are closely linked to the Atlantic Multidecadal Variability (AMV), Indian Ocean decadal variability, and the Pacific Decadal Oscillation (PDO) or Interdecadal Pacific Oscillation (IPO) (Giannini et al., 2003, 2005; Hoerling et al., 2006; Mohino et al., 2011).

In addition to natural variability, significant contributions to decadal and long-term rainfall trends in the Sahel appear to be externally forced. Anthropogenic aerosol loadings (Biasutti and Giannini, 2006; Held et al., 2005) and greenhouse gas (GHG) emissions (Biasutti et al., 2008; Haarsma et al., 2005; Lu and Delworth, 2005) have been identified as key drivers influencing regional hydrological changes.

Climate models forced with observed SSTs are generally able to reproduce the main features of interannual rainfall variability along the Guinea Coast and decadal variability in the Sahel (Giannini et al., 2003, 2005; Lu and Delworth, 2005). These findings indicate that SST forcing represents the dominant driver of WAM rainfall variability. However, SSTs are not the sole factor controlling rainfall in the region. Three additional mechanisms play important roles: (i) atmospheric internal variability, which contributes significantly to decadal and multidecadal rainfall fluctuations (Caminade and Terray, 2010); (ii) land-surface processes and dust–vegetation feedbacks, which can amplify precipitation anomalies (Biasutti et al., 2008); and (iii) the influence of global warming, which affects both multidecadal SST variability and monsoonal circulation (Paeth and Hense, 2004).

The challenge of simulating WAM rainfall variability across interannual to decadal timescales is therefore particularly complex, given the concurrent influence of these mechanisms. As a result, there is still no consensus regarding the response of the Sahel hydrological system to increasing GHG concentrations during the twenty-first century (Druyan, 2010). Indeed, several studies have reported contradictory outcomes for African rainfall trends toward the end of the century (Caminade and Terray, 2010), while others highlight persistent uncertainties in the projection of WAM rainfall (Biasutti et al., 2008).

The Guinea Coast and the Sahel exhibit distinct characteristics in terms of rainfall variability and teleconnections with remote oceanic boundary conditions. Interannual rainfall variability is stronger along the Guinea Coast, whereas the Sahel is dominated by pronounced decadal wet and dry periods, notably the severe droughts of the 1970s and 1980s. On decadal timescales, rainfall anomalies in the two regions tend to co-vary, though the amplitude of anomalies is greater in the Sahel.

### 1.8.1 West African Monsoon Hybrid forecasting

Hybrid forecasting seeks to improve the predictability of hydroclimatic variables by integrating the strengths of both physics-based dynamical models and data-driven statistical or machine learning (ML) approaches. Dynamical models, such as Numerical Weather Prediction (NWP) systems and Earth System Models (ESMs), simulate physical processes governing the climate system, offering physically interpretable outputs. Meanwhile, data-driven models excel in extracting patterns from large observational datasets, providing fast and flexible prediction capabilities (Slater et al., 2023). A typical hybrid forecasting system adopts a statistical-dynamical structure. In this architecture, predictions from dynamical models (e.g., precipitation, temperature, sea surface temperature) serve as predictors within a statistical or ML framework. In such data-driven models, the hydroclimatological predictands can be regressed on a range of covariates, such as observed precipitation/temperature records, static variables (e.g., elevation, slope, and geology), initial hydrologic conditions, or large-scale predictors such as sea surface temperatures (SSTs), surface air temperature, geopotential height, meridional wind, sea ice extent, or modes of climate variability such as the El Niño–Southern Oscillation (ENSO)(Slater et al., 2023). At subseasonal to decadal timescales, dynamical models provide physically consistent forecasts, but often underestimate variability and suffer from low signal-to-noise ratios. Climate model predictions are often used to drive statistical or ML models. A simple example of a hybrid statistical–dynamical model is one that uses predictions of precipitation or temperature from a climate model as predictors within a regression model, where the predictand can be a hydroclimatic variable such as streamflow magnitude, as shown by (Slater et al., 2019), or flood duration, as demonstrated by (Neri et al., 2020). Seasonal forecasts of diverse hydroclimatic variables such as SSTs, sea level pressure, or large-scale climate indices have also been used in hybrid models to predict variables such as precipitation (Madadgar et al., 2016) and tropical cyclone activity (Murakami et al., 2016). For instance, (Madadgar et al., 2016) used atmosphere ocean teleconnections obtained from the NMME (North American Multimodel Ensemble) including the Pacific Decadal Oscillation (PDO), Multivariate ENSO Index (MEI), and Atlantic Multidecadal Oscillation (AMO) to successfully predict seasonal precipitation anomalies in the southwestern USA using a statistical Bayesian-based model.

A promising development for hybrid forecasting is the integration of machine learning to guide ensemble member selection based on their ability to reproduce relevant modes of climate variability. Mode-matching strategies identify ensemble members that align

with large-scale indices such as ENSO, NAO, or AMO ((Moulds et al., 2023; Smith et al., 2020)). Recent advances extend this idea by applying deep learning methods to detect spatial–temporal climate patterns directly from model output. For example, (Sun et al. (2023)) developed a convolutional neural network (CNN) designed to predict the zonal distribution of sea surface temperature anomalies (SSTA) associated with ENSO up to 18 months in advance. Unlike many ENSO prediction systems that focus only on the Niño3.4 index, their CNN-CD (Cosine Distance CNN) model successfully captures the full SSTA zonal pattern across the equatorial Pacific, significantly outperforming state-of-the-art dynamical models. Moreover, the interpretability analysis of the CNN revealed that the model relies on physically meaningful precursors, including the North Pacific Meridional Mode, the South Pacific Quadrupole, and tropical Atlantic SST anomalies at 10-month leads, and on the Pacific discharge–recharge cycle at longer leads. This capability has important implications for hybrid ensemble selection. CNN-based diagnostics can be used to identify which ensemble members most faithfully reproduce ENSO spatial patterns and teleconnections. The large ensemble size of CESM-DPLE (S. G. Yeager et al., 2018), in conjunction with that of CESM-LE, allows an unprecedented exploration of the sensitivity of the decadal prediction skill assessment to the level of noise reduction achieved through ensemble averaging. (Boer et al., 2013).The large ensemble size will also facilitate process-oriented conditional subsampling of the ensemble in order to develop a deeper understanding of the critical mechanisms at play in near-term prediction (S. G. Yeager et al., 2018). Subsetting large ensembles such as CESM-DPLE in this way enables hybrid forecasting systems to focus on dynamically credible members, thereby improving the reliability of West Africa Monsoon prediction on interannual to decadal timescales.

## 1.9 This thesis

Accurate prediction of hydroclimatic variability on decadal timescales remains one of the foremost challenges in climate science, largely due to the persistent signal-to-noise paradox in current dynamical models. This study investigates methods to enhance decadal precipitation predictability over a region that spans from West Africa to East Africa through the integration of post-processing techniques informed by the North Atlantic Oscillation (NAO) and hybrid statistical–dynamical forecasting frameworks applied to the West African Monsoon (WAM). Building upon evidence that decadal variations in the NAO are substantially more predictable than simulated by existing models during boreal winter (DJFM), and based on a recent a recent technique, I apply a two-stage post-processing approach that adjusts ensemble variance from multimodel ensemble and selectively retains ensemble members consistent with the predictable signal. This approach improves the representation of low-frequency atmospheric circulation patterns associated with decadal precipitation during boreal winter (DJFM) for some regions of West, Central and East Africa .

Extending the analysis to the WAM system during boreal summer (JJAS), the study explores hybrid forecasting architectures that combine physics-based Earth System Models

with machine learning (ML) techniques. Large-ensemble predictions from CESM-DPLE are used to identify dynamically credible ensemble members through mode-matching strategies that align with dominant mode of variability, specifically the El Niño Southern Oscillation (ENSO). Convolutional Neural Network (CNN) model is further employed to diagnose spatial-temporal teleconnection patterns and guide ensemble subsetting. Results demonstrate that hybrid ML-dynamical systems can reduce noise and enhance skill in reproducing both interannual and decadal rainfall variability and predictability across the Sahel and Guinea Coast and Central Africa.

Overall, this work provides a unified framework for improving near-term climate predictability by correcting weak signal amplitudes in dynamical models and leveraging machine learning to refine ensemble selection. The findings offer new insights into the mechanisms governing decadal climate variability and highlight pathways to more reliable regional climate projections for climate-sensitive regions. In ([Data and Methods](#)), data and methods used for the DJFM and JJAS seasons are described. In ([Results: Decadal Predictability of Precipitation through NAO-Based Post-Processing](#)), NAO-based post processed results are shown. In ([Results: Variability and Predictability of the West African Monsoon](#)), variability and predictability of WAM and hybrid forecasting results are shown. In ([Discussions and Conclusions](#)) results are discussed and the main conclusions are drawn out.

## 2 Data and Methods

### 2.1 Methods for the boreal Winter season

#### 2.1.1 Simulations

For boreal winter analysis (December to March, DJFM), as reproduced by (Smith et al., 2020), a large multi-model ensemble (169 members; Table 1) of decadal predictions from 13 modelling systems from CMIP5 (Taylor et al., 2012) and CMIP6 (Boer et al., 2016) using hindcasts starting each year from 1960 to 2005 is assessed. A rolling 8-yr boreal winter means is assessed (December to March) representing years 2–9 from each start date. The forecasting systems have different start months between 1 November and 1 January each year; taking the year-2–9 forecast to have a lead time of at least 11 months to focus on decadal timescales and avoid predictability arising from seasonal variability. Hence, the year-2–9 forecast from the 1960 start date is the average of December to March over the period 1 December 1961 to 31 March 1969 (Smith et al., 2020). Model biases and drifts are treated by computing anomalies relative to climatology for each model computed over all hindcasts, and comparing with observed anomalies computed over the same period. The anomalies are relative to the average over all year-2–9 hindcasts.

#### 2.1.2 Observations

Precipitation observations and sea level pressure observations are taken from the ERA5 reanalysis dataset. ERA5, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides comprehensive global atmospheric data with high spatial and temporal resolution, spanning from 1950 onwards. It assimilates vast amounts of observations into a state-of-the-art numerical weather prediction model to produce a consistent and reliable estimate of the historical climate (Hersbach et al., 2023). Nearest neighbor interpolation is applied, regridding the observations data to match forecast data's grid. The climatologies and anomalies are computed in the same way as for the simulations, considering the simulations' and observations' overlapping periods (Gaetani et al., 2013).

#### 2.1.3 Indices

The NAO index is calculated as the difference in mean sea-level pressure between two small boxes located around the Azores (28–20° W, 36–40° N) and Iceland (25–16° W, 63–70° N), with the average over the whole time series removed to create anomalies (N. Dunstone et al., 2016). Rainfall anomalies are averaged throughout the region (10°S–20°N, 20°W–60°E) that extends from west to East Africa, passing through the northern part of Central Africa.

Table 1: Forecast systems, model, ensemble sizes, model resolution, CMIP project

| Forecast Centre   | Model         | Atmosphere resolution <sup>a</sup>        | Ocean resolution <sup>b</sup>            | Ensemble size | CMIP  |
|---|---------------|---|--|---------------|-------|
| Barcelona Supercomputing Center, Spain                  | EC-Earth      | $0.7 \times 0.7 \times 91 \times 0.01$    | $1 \times 1 \times 0.3 \times 75$        | 10            | CMIP6 |
| Bjerknes Center for Climate Research, Norway            | NorCPM1       | $1.9 \times 2.5 \times 26 \times 3$       | $0.7 \times 1.125 \times 0.25 \times 53$ | 20            | CMIP6 |
| Canadian Centre for Climate Modelling and Analysis, CCC | CanCM4        | $2.8 \times 2.8 \times 35 \times 1$       | $0.94 \times 1.41 \times 40$             | 10            | CMIP5 |
|   | CanCM5        | $2.8 \times 2.8 \times 49 \times 1$       | $1 \times 1 \times 0.3 \times 45$        | 10            | CMIP6 |
| Geophysical Fluid Dynamics Laboratory, USA              | CM2.1         | $2 \times 2.5 \times 24 \times 3$         | $1 \times 1 \times 0.3 \times 50$        | 10            | CMIP5 |
| IPSL-EPOC, France                                       | IPSL-CM6A-LR  | $1.25 \times 2.5 \times 79 \times 0.005$  | $1 \times 1 \times 0.3 \times 75$        | 10            | CMIP6 |
| Met Office Hadley Centre, UK                            | HadCM3        | $2.5 \times 3.75 \times 19 \times 4.5$    | $1.25 \times 1.25 \times 20$             | 20            | CMIP5 |
|   | HadGEM3       | $0.55 \times 0.83 \times 85 \times 0.005$ | $0.25 \times 0.25 \times 75$             | 10            | CMIP6 |
| Max Planck Institute for Meteorology, Germany           | MPI-ESM1.0-LR | $1.9 \times 1.9 \times 47 \times 0.01$    | $1.5 \times 1.5 \times 40$               | 3             | CMIP5 |
|   | MPI-ESM1.2-HR | $0.9 \times 0.9 \times 95 \times 0.01$    | $0.4 \times 0.4 \times 40$               | 10            | CMIP6 |
| National Center for Atmospheric Research, USA           | CESM1.1       | $0.9 \times 1.25 \times 30 \times 2.26$   | $1 \times 1.125 \times 0.27 \times 60$   | 40            | CMIP6 |
| University of Tokyo, NIES, and JAMSTEC, Japan           | MIROC5        | $1.4 \times 1.4 \times 40 \times 3$       | $1.4 \times 1.4 \times 0.5 \times 49$    | 6             | CMIP5 |
|   | MIROC6        | $1.4 \times 1.4 \times 81 \times 0.004$   | $1 \times 1 \times 0.5 \times 62$        | 10            | CMIP6 |

<sup>a</sup> Atmosphere resolution: latitude  $\times$  longitude  $\times$  vertical levels  $\times$  timestep.

<sup>b</sup> Ocean resolution: latitude  $\times$  longitude  $\times$  vertical levels  $\times$  timestep.

#### 2.1.4 Forecast quality and uncertainty measures for the boreal winter analysis

Model biases and drifts are treated by computing anomalies relative to each model's own climatology (derived from all hindcasts) and comparing them with observed anomalies calculated over the same reference period.

The *anomaly correlation coefficient* (ACC) is defined as

$$\text{ACC} = \frac{\sum(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum(f_i - \bar{f})^2} \sqrt{\sum(o_i - \bar{o})^2}},$$

where  $f_i$  and  $o_i$  are the ensemble-mean forecast and observed values at time  $i$ , respectively, and overbars denote averages over the set of 8-year rolling mean periods. ACC measures the model's phase skill; how well the forecast captures the timing and sign of observed anomalies.

Errors in the signal-to-noise ratio can be quantified by comparing the predictable components (the predictable fraction of the total variability) in observations and models that can be represented by ratio of predictable components (Eade et al., 2014; Scaife and Smith, 2018; Siegert et al., 2016). The *ratio of predictable components* (RPC) is given by

$$\text{RPC} = \frac{\sigma_{\text{sig}}^o / \sigma_{\text{tot}}^o}{\sigma_{\text{sig}}^f / \sigma_{\text{tot}}^f}$$

=

$$\frac{\text{ACC}}{\sigma_{\text{sig}}^f / \sigma_{\text{tot}}^f},$$

with  $\sigma_{\text{sig}}^o$  and  $\sigma_{\text{sig}}^f$  denoting the standard deviations of the predictable signal in observations and forecasts, and  $\sigma_{\text{tot}}^o$  and  $\sigma_{\text{tot}}^f$  their respective total variability. For the forecasts  $\sigma_{\text{sig}}^f$  and  $\sigma_{\text{tot}}^f$  are computed from the ensemble mean and individual members, respectively. is not computed where ACC is negative, probably the equation of RPC represents a lower bound, because ACC would be expected to increase with improved models or more ensemble members and  $\sigma_{\text{sig}}^f$  would be expected to decrease with more ensemble members (Eade et al., 2014; Scaife and Smith, 2018). It quantifies the errors in the signal-to-noise ratio. RPC is expected to be 1 for a perfect forecasting system; values greater than 1 indicate where the signal-to-noise ratio is erroneously too small in models. It is theoretically possible for the multi-model RPC to be larger than that for individual models if time-dependent model biases or teleconnection errors reduce the model signal more than the correlation with observations. Assessing this thoroughly would require large ensembles of individual model hindcasts, which are not available. However, assessing the largest individual model ensemble that is available does not support this hypothesis: the NCAR RPC of 6.2 is not significantly different from the average RPC of multi-model ensembles of the same size (4.8, averaged over 1,000 random samples).

Finally, the *ratio of predictable signal* (RPS) compares the absolute magnitude of

predictable variance between observations and forecasts:

$$\text{RPS} = \frac{\sigma_{\text{sig}}^o}{\sigma_{\text{sig}}^f} = \text{RPC} \frac{\sigma_{\text{tot}}^o}{\sigma_{\text{tot}}^f}.$$

RPS is leveraged as scaling factor within NAO matching procedure to take into account the underestimated signal (see [Nao Matching](#)) .

### 2.1.5 Lagged Ensemble

Consecutive 8-yr means contain seven identical years. Hence, large interannual variations, as seen in 169-member ensemble-mean NAO forecasts (Figure 6, thin red curve), are not expected. They occur because the signal-to-noise ratio is too small in models and consecutive decadal predictions consist of independent model simulations that are dominated by different samples of the noise. Create a lagged ensemble by combining the required forecast with the previous three; for example, the year-2–9 forecasts starting in 1963 are combined with the year-2–9 forecasts starting in 1962, 1961 and 1960. Doing so yields a total of 676 members (169 members times four start dates).

### 2.1.6 Nao Matching

At any location that is influenced by the NAO we can write

$$O = O_{\text{NAO}} + O_{\text{OTHER}} + \varepsilon_o, \quad (1)$$

$$F_k = F_{k,\text{NAO}} + F_{k,\text{OTHER}} + \varepsilon_k, \quad (2)$$

$$\hat{F} = \hat{F}_{\text{NAO}} + \hat{F}_{\text{OTHER}} + \hat{\varepsilon}, \quad (3)$$

where  $O$ ,  $F_k$  and  $\hat{F}$  are the observed, forecast ensemble member  $k$  and forecast ensemble-mean values of variable. The subscript “NAO” refers to the portion that is related to the NAO, the subscript “OTHER” refers to the portion related to other predictable drivers (including greenhouse gases and sea surface temperatures unrelated to the NAO) and  $\varepsilon$  is an unpredictable residual.

Because the predictable NAO signal is too small in models, the mean of a very large ensemble is required for skilful NAO predictions. However, the magnitude of the ensemble-mean NAO is much too small, so  $\hat{F}_{\text{NAO}}$  will be severely underestimated. One approach to overcome the model deficiencies replaces the underestimated  $\hat{F}_{\text{NAO}}$  with a more realistic  $F_{k,\text{NAO}}$  by selecting from the full ensemble a smaller set of members that have the required magnitude of the NAO. These members have NAO and related teleconnection magnitudes close to that of the required NAO, while retaining other influences.  $\hat{F}_{\text{NAO}}$  for this selected ensemble will be larger than that of the full ensemble, thereby increasing the signal, even if the noise reduction is weaker than in the full ensemble. This happens because the selection process shifts some of what would normally be treated as noise into the signal (since those variations are consistent with NAO) ([Dobrynin et al., 2018](#)). This is achieved by taking the required NAO to be the ensemble-mean forecast NAO, adjusting it to account

for the underestimation of the predictable signal. This is achieved by multiplying the ensemble-mean NAO by RPS. RPS is computed for each hindcast start date separately using a cross-validation approach in which the required hindcast and those on either side are omitted. The overall procedure is as follows.

For each start date  $i$ :

1. Compute the signal-adjusted NAO index (described above) of the ensemble mean,  $\text{NAO}_i$ .
2. Compute the NAO index for each ensemble member,  $\text{NAO}_i^k$ .
3. For each ensemble member, calculate the difference  $\Delta_i^k = \text{NAO}_i - \text{NAO}_i^k$ .
4. Select the  $M = 20$  ensemble members with the smallest absolute differences.

Then take the mean of this subset of  $M$  members and present ACC of this subsample. This method relies on models simulating realistic NAO teleconnections ( $F_{k,\text{NAO}}$ ). The further improvement that could enhance this approach is to select the best models in this respect with particular focus on the region of interest ( $10^\circ\text{S}$ – $20^\circ\text{N}$ ,  $20^\circ\text{W}$ – $60^\circ\text{E}$ ).

## 2.2 Methods for the boreal Summer season

### 2.2.1 Simulations

For boreal summer analysis (June to September, JJAS), the CESM (Community Earth System Model) Decadal Prediction Large Ensemble (DPLE) set of climate model simulations produced by the National Center for Atmospheric Research (NCAR) is assessed. The CESM DPLE is based on CESM, version 1.1, using the same model and component configuration as that used in the CESM-LE (Large Ensemble) that is the uninitialized counterpart of CESM DPLE (Kay et al., 2015). The DPLE provides a large sample of initialized decadal forecasts, enabling robust statistical analysis of near-term climate variability and predictability. The ensemble approach allows for quantification of uncertainty due to internal climate variability. For each start date, a 40-member ensemble was generated by randomly perturbing the atmospheric initial condition at the round-off level. The simulations were integrated forward for 122 months after initialization. Observation-based ocean and sea ice initial conditions for the 1954–2015 period were obtained from a reanalysis-forced simulation of the CESM ocean and sea ice models. The initial conditions for the atmosphere and land models were obtained from CESM Large Ensemble (LENS) simulations at corresponding historical times. In particular the study analyze sets of 10 year climate retrospective forecasts, also known as decadal reforecasts or hindcasts. The analysis focuses on the seasonal average precipitation during the monsoonal season June to September (JJAS) over the region ( $10^\circ\text{S}$ – $20^\circ\text{N}$ ,  $20^\circ\text{W}$ – $60^\circ\text{E}$ ) that extends from west to East Africa, passing through the northern part of Central Africa. For each model, monthly data from a set of 10-yr-long hindcast experiments initialized every year from 1960 to 2013 (10 experiments) are evaluated. Following (International CLIVAR Project

Office, 2011) to compute the climatology and anomaly of each hindcast lead year (i.e., lead years from 1 to 10), it is proceeded as follows:

The climatology for lead year  $l$  is defined as the mean of all hindcasts at that lead year, from 1960 to 2013:

$$C(l) = \frac{1}{N} \sum_{t=1960}^{2013} H(t, l),$$

where  $H(t, l)$  is the hindcast initialized in year  $t$  at lead year  $l$ , and  $N$  is the total number of start years (here,  $N = 2013 - 1960 + 1$ ).

Once the climatologies are calculated, the anomalies are obtained as the difference between the hindcast values and their corresponding climatology:

$$A(t, l) = H(t, l) - C(l).$$

For example, the anomalies of the 10-year hindcast starting from 1961 are computed as

$$A(1961, 1) = H(1960, 1) - C(1),$$

for lead year 1, and

$$A(1970, 10) = H(1970, 10) - C(10),$$

for lead year 10. This methodology is widely used assessment of decadal predictions (García-Serrano et al., 2012; Goddard et al., 2013; Kim et al., 2012). (García-Serrano et al., 2012) show that this approach systematically yields better ACC scores than the removal of the classic long-term mean climatology (Gaetani et al., 2013). Full field initialization was used for all component models, and so drift correction prior to analysis is generally recommended (International CLIVAR Project Office, 2011) see [Drift Correction for Decadal Climate Predictions](#)). According to guidance papers (Goddard et al., 2013; Meehl et al., 2014), which identify the forecast time averaging that has an actionable information scope, a 4 year average is performed upon these drift-corrected anomalies in order to retain interannual to decadal predictability (García-Serrano et al., 2012). So the average anomaly for lead years 1 to 4 is calculated by averaging the anomalies of lead years 1, 2, 3, and 4. This procedure is repeated for subsequent 4-year windows: lead years 2–5, 3–6, ..., up to 7–10.

### 2.2.2 Observations

Precipitation observations and sea level pressure observations are taken from the ERA5 reanalysis dataset. ERA5, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides comprehensive global atmospheric data with high spatial and temporal resolution, spanning from 1950 onwards. It assimilates vast amounts of observations into a state-of-the-art numerical weather prediction model to produce a consistent and reliable estimate of the historical climate (Hersbach et al., 2023). Nearest neighbor interpolation is applied, regridding the observations data to match forecast data's grid. The climatologies and anomalies are computed in the same way as for the

simulations, considering the simulations' and observations' overlapping periods (Gaetani et al., 2013).

### 2.2.3 Indices

Rainfall anomalies are averaged throughout the region ( $10^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ,  $20^{\circ}\text{W}$ – $60^{\circ}\text{E}$ ) that extends from west to East Africa, passing through the northern part of Central Africa and for more specific regions: Guinea Coast ( $5^{\circ}\text{N}$ – $10^{\circ}\text{N}$ ,  $8^{\circ}\text{E}$ – $15^{\circ}\text{E}$ ), West Sahel ( $5^{\circ}\text{N}$ – $18^{\circ}\text{N}$ ,  $18^{\circ}\text{W}$ – $8^{\circ}\text{W}$ ), and West Equatorial Africa ( $6^{\circ}\text{S}$ – $3^{\circ}\text{N}$ ,  $5^{\circ}\text{E}$ – $14^{\circ}\text{E}$ ). Niño3.4 index (SST anomalies averaged over  $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $170^{\circ}\text{W}$ – $120^{\circ}\text{W}$ ) (Ham et al. (2019)).

### 2.2.4 Drift Correction for Decadal Climate Predictions

Climate models are inherently imperfect and produce simulations that differ from observed climatologies. As a result, model predictions tend to *drift* toward the model's own climate over time. This drift can interfere with the accurate prediction of actual climate evolution. Therefore, drift correction is applied to make forecasts more comparable to observed data according the guideline proposed (International CLIVAR Project Office, 2011). The subtraction of the climatology is particularly necessary when comparing models to observations because models produce their own climatology, which can be quite different from the observed one (Gaetani et al., 2013).

Forecast and Climatological Averages

Let  $Y_{j\tau}$  denote a raw forecast at initialization  $j = 1, \dots, n$  and lead time  $\tau$ , and let  $X_{j\tau}$  represent the corresponding observed data.

The ensemble averages at lead time  $\tau$  are

$$Y^\tau = \frac{1}{n} \sum_{j=1}^n Y_{j\tau}, \quad X^\tau = \frac{1}{n} \sum_{j=1}^n X_{j\tau}.$$

These are *hindcast averages*, specific to the set of forecasts available. They should be distinguished from the longer-term climatological means of model and observations, denoted

$$\langle Y \rangle, \quad \langle X \rangle.$$

Full-field Initialization

In full-field initialization, the model state is initialized close to observed values:

$$Y_{j,\tau=0} \approx X_{j,\tau=0}.$$

Forecast ensembles are generated by perturbing the initial state. The perturbations are designed so that members are not identical but remain close to the observed initial conditions.

### 2.2.5 Drift Correction under Full-field Initialization

Model drift is estimated as the difference between the ensemble mean forecast and the observed mean at each lead time:

$$d_\tau = Y^\tau - X^\tau.$$

The drift-corrected forecast for member  $j$  is obtained by subtracting this drift:

$$\hat{Y}_{j\tau} = Y_{j\tau} - d_\tau = X^\tau + (Y_{j\tau} - Y^\tau).$$

Here: -  $Y_{j\tau} - Y^\tau$  represents the anomaly of the raw forecast relative to the forecast ensemble mean. - Adding  $X^\tau$  aligns the forecast anomalies with the observed climatology at lead time  $\tau$ .

This formulation ensures that the forecast expresses deviations from the model's own forecast climatology, but referenced against observed climatology. Drift correction should be applied in a cross-validated manner, i.e., the forecast being corrected must be excluded from the averages  $Y^\tau$  and  $X^\tau$  used to compute  $d_\tau$ .

### 2.2.6 Forecast quality and uncertainty measures for the boreal summer analysis

The *anomaly correlation coefficient* (ACC) is applied to the drift-corrected anomalies that were previously defined. After computing anomalies relative to each model's lead-time climatology, 4-year running averages are taken (lead years 1–4, 2–5, 3–6, ..., 7–10). The ACC is then evaluated separately for each of these averaging windows, measuring the temporal correlation between the ensemble-mean forecast anomalies and the observed anomalies over the common evaluation period.

Formally:

$$\text{ACC} = \frac{\sum(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum(f_i - \bar{f})^2} \sqrt{\sum(o_i - \bar{o})^2}},$$

where

- $f_i - \bar{f}$  = 4-year mean ensemble-mean forecast anomaly at period  $i$  (for a given averaging window),
- $o_i - \bar{o}$  = corresponding 4-year mean observed anomaly,
- overbars ( $\bar{f}, \bar{o}$ ) = averages over the full evaluation period (e.g., 1960–2013).

The ACC thus measures the model's *phase skill* in each window, that is, its ability to reproduce the timing and sign of the observed precipitation anomalies associated with the West African Monsoon, independent of systematic model biases and drift.

## 2.3 CNN-CD Model

To predict the zonal pattern of sea surface temperature anomalies (SSTA) across the equatorial Pacific, Sun et al. (2023) develop a convolutional neural network (CNN) model that incorporates the cosine distance as a core component of the loss function. This model is referred to as the CNN-CD model.

The equatorial Pacific domain is subdivided into 13 overlapping longitudinal regions (Figure 3). Each region spans 10 degrees in latitude (from 5°S to 5°N) and 20 degrees in longitude, with a 10-degree overlap between adjacent regions. Specifically, Region 1 covers 130°E–150°E, Region 2 covers 140°E–160°E, and so forth, up to Region 13, which spans 110°W–90°W. The CNN-CD model outputs area-averaged SSTA values for each of these regions, effectively predicting the entire SSTA zonal pattern. The architecture of the CNN-CD model (illustrated in Figure 4) is adapted from the CNN framework introduced by Ham et al. (2019), hereafter referred to as CNN-H19. The model inputs consist of global maps of SSTA and ocean heat content (OHC) anomalies, sampled three months prior to the forecast start time  $\tau$  (i.e., at  $\tau - 3$ ,  $\tau - 2$ , and  $\tau - 1$ ). The input spatial resolution is  $5.0^\circ \times 5.0^\circ$  in latitude and longitude.

The core of the network is composed of:

- Three convolutional layers
- Two max-pooling layers
- A fully connected dense layer

The nonlinear activation function employed is the hyperbolic tangent function. For optimization, using the Adam optimizer (Chen and McKay-Crites (2018)), enables adaptively to adjust learning rates during training.

To evaluate the similarity between the predicted and observed SSTA zonal patterns, the cosine distance as the loss function is adopted. This loss function emphasizes the alignment of the spatial patterns, rather than simply minimizing pointwise error.

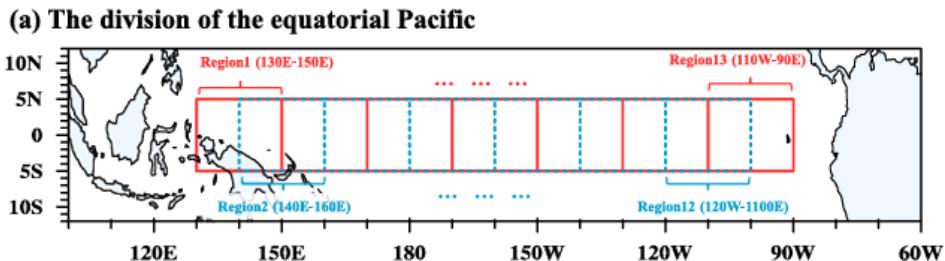


Figure 3: The division of Equatorial Pacific (Sun et al. (2023))

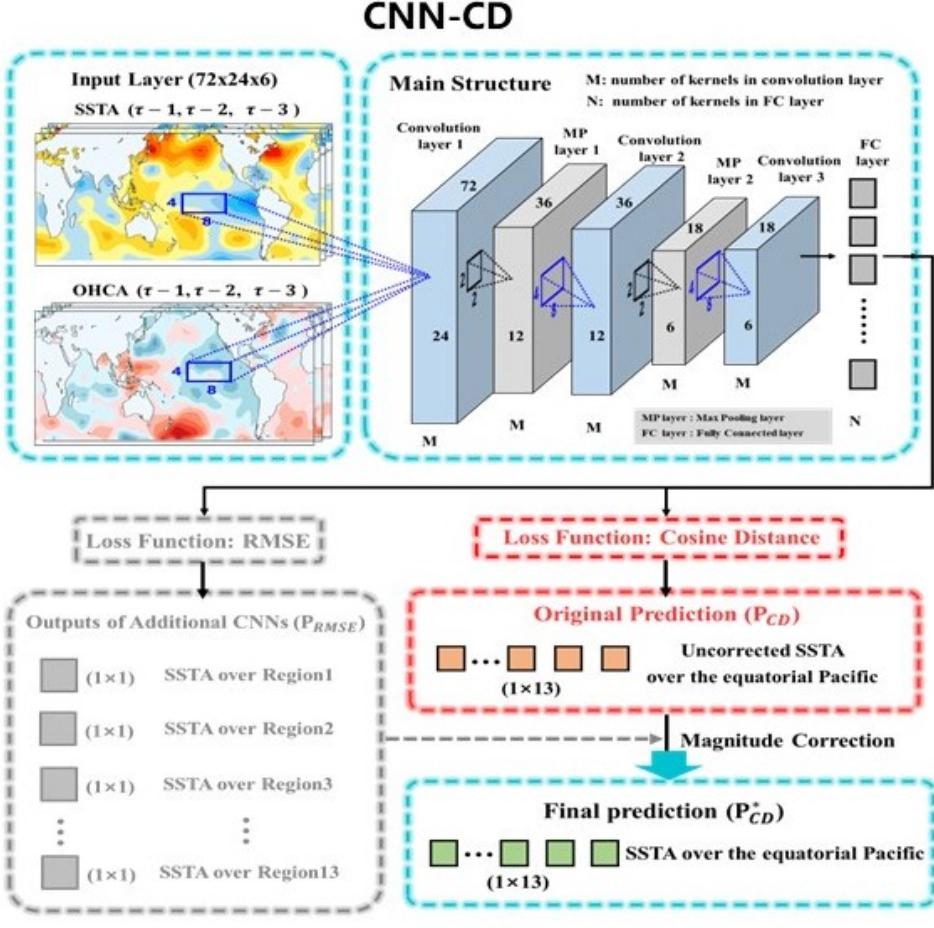


Figure 4: CNN architecture (Sun et al. (2023))

### 2.3.1 CMIP and reanalysis datasets

The transfer learning technique is applied to overcome the limitation due to the short length of the observation record when training the CNN-CD model. As shown in Table 2, the historical simulations (1850–2014) from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. (2016)) are used as the initial training set for CNN-CD. In total, 24 CMIP6 models with 43 members are used. The batch size for each epoch is set to 500, and we run a total of 200 epochs during the initial training. Then, to bridge the gap between the CMIP simulation and the observation, reanalysis datasets from 1871–1973 are used for fine-tuning the CNN-CD model, including SST from NOAA ERSSTv5 (Huang et al. (2017)) and SSH from SODA 2.2.4 (Carton and Giese (2008)). The batch size and epochs for fine-tuning are set to 20 and 20, respectively. During the hindcast period (1981–2016), SST from ERSSTv5 and SSH from GODAS (Behringer and Xue (2004)) are used as the validation set.

Table 2: List of climate models and their modeling groups.

| Model name      | Modeling group  |
|-----------------|---|
| ACCESS-CM2      | Commonwealth Scientific and Industrial Research Organisation, Australia; Australian Research Council Centre of Excellence for Climate System Science  |
| ACCESS-ESM1-5   | Commonwealth Scientific and Industrial Research Organisation, Australia; Australian Research Council Centre of Excellence for Climate System Science  |
| BCC-CSM2-MR     | Beijing Climate Center, China   |
| CanESM5         | Canadian Centre for Climate Modelling and Analysis  |
| CESM2           | National Center for Atmospheric Research, USA   |
| CNRM-CM6-1      | Centre National de Recherches Météorologiques, France; Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique  |
| CNRM-ESM2-1     | Centre National de Recherches Météorologiques, France; Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique  |
| EC-Earth3       | EC-Earth Consortium   |
| FGOALS-f3-L     | Chinese Academy of Sciences   |
| FGOALS-g3       | Chinese Academy of Sciences   |
| FIO-ESM-2-0     | First Institute of Oceanography, SOA, China; Qingdao National Laboratory for Marine Science and Technology, China   |
| GFDL-CM4        | NOAA Geophysical Fluid Dynamics Laboratory, USA   |
| GFDL-ESM4       | NOAA Geophysical Fluid Dynamics Laboratory, USA   |
| GISS-E2-1-H     | NASA Goddard Institute for Space Studies, USA   |
| HadGEM3-GC31-LL | Met Office Hadley Centre, UK  |
| IPSL-CM6A-LR    | Institut Pierre Simon Laplace, France   |
| MIROC6          | Japan Agency for Marine-Earth Science and Technology; Atmosphere and Ocean Research Institute, University of Tokyo; National Institute for Environmental Studies; RIKEN Center for Computational Science, Japan |
| MPI-ESM1-2-HR   | Max Planck Institute for Meteorology, Germany   |
| MPI-ESM1-2-LR   | Max Planck Institute for Meteorology, Germany   |
| MRI-ESM2-0      | Meteorological Research Institute, Japan  |
| NESM3           | Nanjing University of Information Science and Technology, China   |
| NorCPM1         | Norwegian Climate Centre, Norway  |
| NorESM2-LM      | Norwegian Climate Centre, Norway  |
| SAMO-UNICON     | Seoul National University, Republic of Korea  |

### 2.3.2 NMME

The North American Multimodel Ensemble (NMME; B. P. Kirtman et al. (2014)) is a multi-model forecasting system consisting of state of-the-art coupled models from American and Canadian centers. Eight NMME models providing seasonal forecasts from 1982–2016 are selected as benchmarks for CNN-CD performance (Table 3).

Table 3: List of NMME models used as benchmarks for CNN-CD performance (1982–2016).

| Model               | Center/Institution |
|---------------------|--------------------|
| CFSv2               | NCEP               |
| CCSM4               | NCAR               |
| GFDL-CM2.1          | GFDL               |
| GFDL-CM2.5 FLOR-A06 | GFDL               |
| GFDL-CM2.5 FLOR-B01 | GFDL               |
| CanCM3              | CCCma              |
| CanCM4              | CCCma              |
| NASA-GMAO           | NASA               |

### 2.3.3 Methodology

#### Cosine Distance

To depict the similarity of the SSTA zonal pattern between the observation and the predictand, the cosine distance is used as the loss function in the CNN-CD model. Specifically, the cosine distance is calculated as follows:

$$C_D = 1 - \frac{O \cdot P_{CD}}{|O| |P_{CD}|}, \quad (4)$$

where  $C_D$  denotes the cosine distance.  $O$  is a one-dimensional vector containing thirteen values of the observed SSTA averaged from Region 1 to Region 13, and  $P_{CD}$  is the corresponding prediction vector from CNN-CD. The numerator  $O \cdot P_{CD}$  represents the inner (dot) product of  $O$  and  $P_{CD}$ , while  $|O|$  and  $|P_{CD}|$  denote their respective modulus.

The cosine distance  $C_D$  ranges from 0 to 2. When  $O$  and  $P_{CD}$  have identical zonal patterns,  $C_D = 0$ . When their patterns are completely opposite,  $C_D = 2$ . Thus, smaller values of  $C_D$  indicate more accurate predictions of the SSTA zonal pattern over the equatorial Pacific.

#### Magnitude Correction

Through employing the cosine distance as the loss function, raw output of the zonal Sea Surface Temperature Anomalies (SSTA) is obtained, denoted as  $P_{CD}$ . While the spatial pattern of  $P_{CD}$  may closely resemble the observations, its magnitude may still deviate. Therefore, we perform a post-processing step to correct the magnitude by introducing region-specific scale factors. For each region ( where  $i = 1, 2, \dots, 13$ ), we construct a CNN model with the same architecture as CNN-CD, but replace the loss function with root

mean square error (RMSE). The model is trained to predict the area-averaged SSTA for the corresponding region. By repeating this procedure for all 13 regions, we obtain a set of region-specific predictions that form a vector  $P_{\text{RMSE}}$ , representing the area-averaged SSTAs across Region1 to Region13.

Using  $P_{\text{RMSE}}$ , we apply a magnitude correction to the original  $P_{\text{CD}}$  as follows:

$$P_{\text{corr}} = P_{\text{cd}} \times \frac{\max(P_{\text{rmse}}) - \min(P_{\text{rmse}})}{\max(P_{\text{cd}}) - \min(P_{\text{cd}})} \quad (5)$$

where  $\max(P_{\text{RMSE}})$  and  $\min(P_{\text{RMSE}})$  denote the maximum and minimum values in  $P_{\text{RMSE}}$ , respectively. The final corrected prediction,  $P_{\text{CD}*}$ , represents the adjusted output from CNN-CD with improved magnitude consistency across the 13 predefined regions.

### Skill Metrics

The anomaly correlation coefficient (ACC) and the root mean square error (RMSE) are widely used metrics for evaluating prediction skill. The ACC denotes the correlation coefficient between the observational time series of a given index and its predicted counterpart. Its value ranges from  $-1$  to  $1$ , with a larger positive ACC indicating a more accurate prediction, and a negative ACC indicating poor or opposite prediction skill.

The RMSE measures the average magnitude of the differences between the predictand and the observation. The RMSE is non-negative, with smaller values corresponding to more accurate predictions.

These two metrics are calculated as follows:

$$ACC_l = \frac{\sum_{i=1}^N (O_{i,l} - \bar{O}_l)(P_{i,l} - \bar{P}_l)}{\sqrt{\sum_{i=1}^N (O_{i,l} - \bar{O}_l)^2 \sum_{i=1}^N (P_{i,l} - \bar{P}_l)^2}}, \quad (6)$$

$$RMSE_l = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_{i,l} - P_{i,l})^2}, \quad (7)$$

where  $N$  is the number of validation samples and  $l$  denotes the forecast lead month.  $O_{i,l}$  and  $P_{i,l}$  represent the observed and predicted values (e.g., Niño 3.4 index) for the  $i$ th sample at lead time  $l$ , while  $\bar{O}_l$  and  $\bar{P}_l$  are their averages across  $N$  samples.

The modified Taylor skill score (TSS) further evaluates spatial pattern and magnitude similarity(Taylor (2001)):

$$TSS_l = \frac{1}{N} \sum_{i=1}^N \frac{1 + r(O_{i,l}, P_{i,l})}{1 + \left( \frac{\sigma_{P_{i,l}}}{\sigma_{O_{i,l}}} - 1 \right)^2} \quad (8)$$

where:

- $N$  is the number of validation samples,
- $O_{i,l}$  and  $P_{i,l}$  represent the observed and predicted SSTA zonal patterns at sample  $i$  and lead time  $l$ ,

- $r(O_{i,l}, P_{i,l})$  is the pattern correlation coefficient between  $O_{i,l}$  and  $P_{i,l}$ ,
- $\sigma_{O_{i,l}}$  and  $\sigma_{P_{i,l}}$  are the spatial standard deviations of  $O_{i,l}$  and  $P_{i,l}$ , respectively.

The TSS ranges from 0 to 1, where a higher value indicates a closer match between predicted and observed SSTA patterns in terms of both spatial distribution (correlation) and magnitude (standard deviation).

#### Activation map analysis

To visualize the contribution of the input signals to the predictions made by the CNN-CD model, heat value in the activation map is calculated as follows:(B. Zhou et al. (2016); Ham et al. (2021))

$$h(x, y, \lambda) = \sum_{n=1}^N \left\{ \tanh \left( \sum_{m=1}^{M_n} W_{F,m,n}(x, y) v_{L,m}(x, y) + \frac{b_n^F}{X_L Y_L} \right) W_{O,n,\lambda} \right\} + \frac{b_\lambda^O}{X_L Y_L} \quad (9)$$

where  $h(x, y, \lambda)$  is the heat value at the grid point  $(x, y)$  in the activation map for predicted SSTA over Region  $\lambda$  ( $\lambda = 1, 2, \dots, 13$ ).  $M_n$  and  $N$  respectively denote the number of kernels in the last convolutional layer  $L$  and fully connected layer  $F$ .  $W_{F,m,n}(x, y)$  denotes the weight at the grid point  $(x, y)$  used to link the  $m$ -th feature map in  $L$  to the  $n$ -th kernel in  $F$ , while  $W_{O,n,\lambda}$  is the weight that links the  $n$ -th kernel in  $F$  to the  $\lambda$ -th output in the output layer  $O$ .  $b_n^F$  and  $b_\lambda^O$  respectively denote the bias of the  $n$ -th kernel in  $F$  and the bias of the  $\lambda$ -th output in  $O$ .  $v_{L,m}(x, y)$  is the value at the grid point  $(x, y)$  of the  $m$ -th feature map in  $L$ .  $X_L$  and  $Y_L$  are the dimensions of the activation map in  $L$ .

### 3 Results: Decadal Predictability of Precipitation through NAO-Based Post-Processing

#### 3.1 NAO variability and predictability

The NAO exhibits marked decadal variability (Hurrell and van Loon, 1997), with a strong increase from the 1960s to the 1990s and a decrease thereafter (Figure 5, red curve). The raw ensemble-mean forecast shows virtually no signal (blue curve) and the observations generally lie within the model uncertainties (shading shows the 5%-95% range diagnosed from the ensemble spread). The extreme values in the early 1960s and late 1980s are not well captured by models in agreement with other studies (Bracegirdle, 2018; Scaife et al., 2009). The 5–95% spread from the ensemble members shows the model’s estimate of uncertainty. The observations are still consistent with being one possible realisation of the ensemble. The small model signal and much larger spread implies little ability to predict the NAO and a large component of unpredictable internal variability. However, by comparing with observations, a significant correlation skill of the ensemble mean ( $ACC = 0.48$ ) is found. Skilful climate model predictions of the NAO are possible using the ensemble mean, but the signal-to-noise ratio is too small ( $RPC = 4.2$ ) and its variance must be calibrated to provide realistic forecasts (Smith et al., 2019).

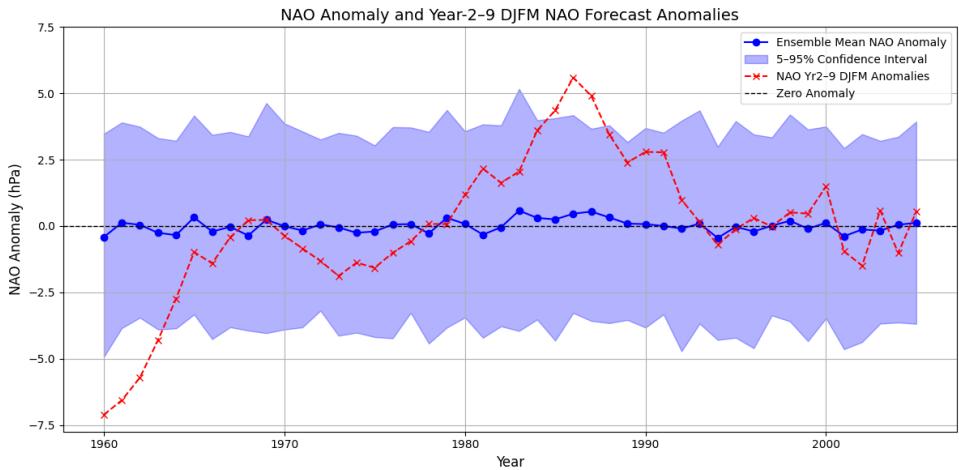


Figure 5: **ACC = 0.48, RPC = 4.2.** Observed (black) and model-forecast (years 2–9; blue) 8-year running means of the boreal winter (December–March) NAO index. The ensemble mean of 169 members is shown by the blue curve, and the red shading indicates the 5–95% confidence interval derived from individual ensemble members. Anomalies are relative to the average over all year-2–9 hindcasts (Smith et al., 2020).

Rescaling the ensemble-mean time series to have the same variance as the observations reveals that the predictions do capture the observed increase from the 1960s to 1990s and decrease thereafter (Figure 6). However, even with 169 ensemble members (Figure 6, thin red curve) there are large interannual variations that are not observed in 8-yr rolling means. Creating a larger, lagged ensemble by taking the average of the four latest forecasts available at each start date (giving 676 members; Figure 6, thick red curve). This reveals

that the NAO is highly predictable on decadal timescales ( $ACC = 0.79$ ) , in stark contrast to the lack of predictability implied by the standard interpretation of raw model output. Importantly, the signal-to-noise ratio is much too small in the models ( $RPC = 11$ ).

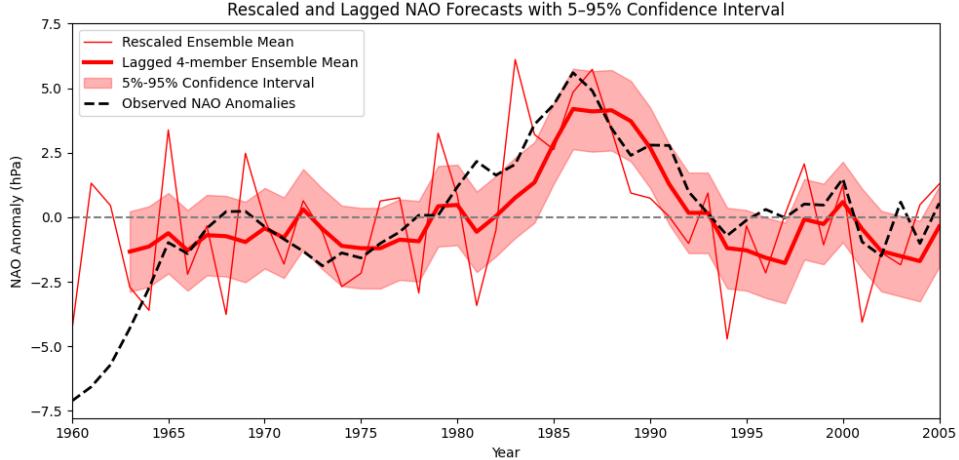


Figure 6: **ACC = 0.79, RPC = 11** Observed (black) and model-forecast (years 2–9; blue) 8-year running means of the boreal winter (December–March) NAO index. Ensemble-mean forecast rescaled to have the same variance as the observations (thin red curve), and then smoothed by taking the lagged average of the latest four forecasts at each start date (thick red curve; 676 members) (Smith et al., 2020).

The total 8-yr variability of the NAO in individual model members is not significantly different from the observations. Hence, the predictable signal is underestimated by an order of magnitude in the model ensemble. Because the standard error of the ensemble mean is reduced by the square-root of the ensemble size, the ensemble required to extract the signal is 100 times larger than it would be for perfect models. The fact that the NAO signal is much too weak in models implies that the effects of the NAO will be underestimated relative to other factors such as greenhouse gases. Hence, in regions influenced by the NAO, the ensemble mean will not reflect the true balance of driving factors and inflating its variance to be the same as observed will not correct the error. A potential solution is to post-process the model output by selecting a subset of ensemble members from the lagged ensemble (of 676 members), corresponding to the 20 members whose simulated NAO is closest in sign and magnitude to the ensemble-mean NAO, after adjusting it to take into account the underestimated signal. The simulated NAO for these members have similar magnitudes to that of the variance-adjusted forecast NAO, while retaining influences from greenhouse gases and other sources. This procedure is defined NAO matching (see [Nao Matching](#)).

### 3.2 NAO matching

Smith et al., 2020 suggest that further improvements in decadal prediction skill may be achieved by selecting and combining the best-performing models. In this study, the models MIROC6, MIROC5, CanCM5, CanESM5, and GFDL-CM2.1 selected by

individually evaluating the spatial distribution of the precipitation anomaly correlation coefficient (ACC) for the raw outputs of each model over the region of interest ( $10^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ,  $20^{\circ}\text{W}$ – $60^{\circ}\text{E}$ ). The ensemble mean of these selected models (46 members in total) is subsequently lagged to produce an expanded ensemble of 184 members. NAO matching procedure is then applied, using the signal-adjusted NAO index derived from the full 676-member ensemble as the target. This approach yields 20 members (out of 184) for each start date that best correspond to the desired NAO phase. The resulting ACC skill scores obtained after NAO matching are presented in Figure 7.

High skill in these experiments aligns with regions of known climate variability. Previous studies have shown that initialized decadal hindcasts exhibit enhanced predictability of near-surface air temperature (SAT) in the North Atlantic (Eade et al., 2014), primarily due to improved initialization of the Atlantic Meridional Overturning Circulation (AMOC), a major driver of multidecadal variability in the region (Doblas-Reyes et al., 2013; Pohlmann et al., 2013; Robson et al., 2014; Smith et al., 2010; S. Yeager et al., 2012b). Moreover, variability in Atlantic temperatures has been linked to fluctuations in mean sea level pressure (MSLP) in the tropical Atlantic and to precipitation anomalies in the Sahel and West Africa (N. J. Dunstone et al., 2011; R. Zhang and Delworth, 2006).

Model selection following the approach of Smith et al., 2020 leads to notable improvements in the spatial distribution of predictive skill relative to the multi-model ensemble used in that study. The results obtained through the NAO matching approach are consistent with the well-established North Atlantic–Sahel/West Africa climate teleconnection (Eade et al., 2014), reinforcing the influence of Atlantic variability on rainfall and atmospheric patterns across the region.

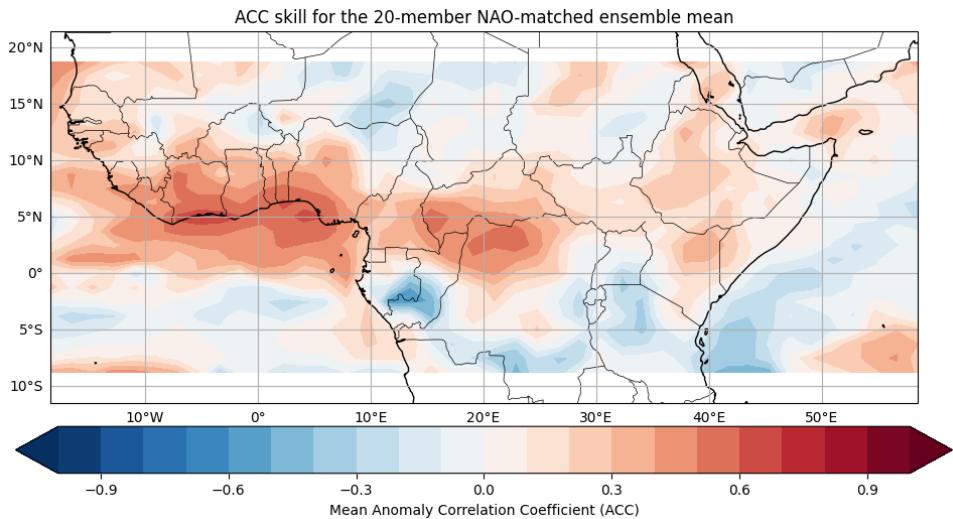


Figure 7: ACC skill for the 20-member NAO-matched ensemble mean.

The effect of NAO matching on skill is computed as the partial correlation between the observed and forecast residuals after regressing out the lagged ensemble-mean forecast, thereby focusing on the variability not already captured by the lagged ensemble mean (Smith et al., 2019) (Figure 8).

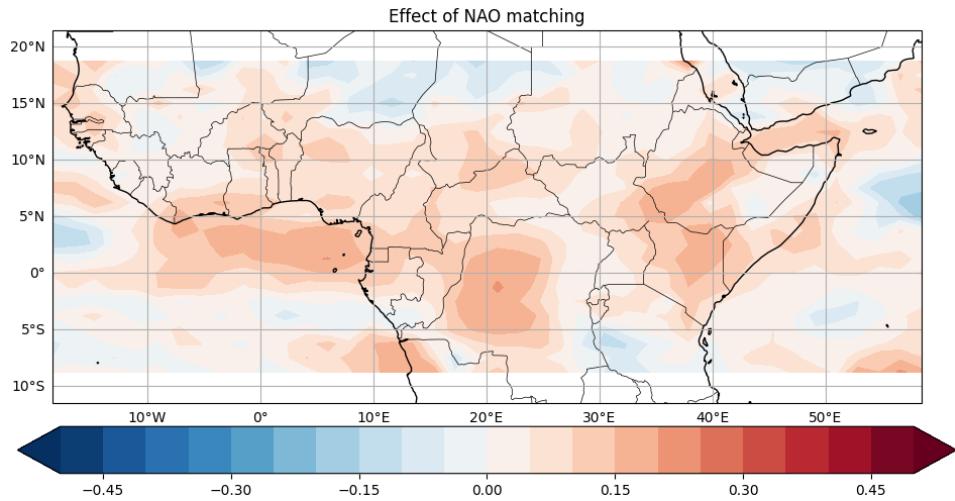


Figure 8: Effect of NAO matching

## 4 Results: Variability and Predictability of the West African Monsoon

### 4.1 Spatial distribution of West Africa Monsoon

The Figure 9( 10) shows interpolated JJAS precipitation climatologies from observations (model simulations) for lead Years 1–10. Each subplot shows the spatial distribution of mean summer precipitation (mm/day) for a specific lead year. In the Figure 11, each panel shows the difference between the forecast model climatology and the ERA5 observed climatology at a specific lead year, depicting the precipitation systematic error. Positive values (red) indicate where the model overestimates precipitation relative to observations, while negative values (blue) indicate underestimation. This translates into a clear overestimation of the tropical convection over the sea and a dry systematic error over the eastern Sahel due to a too southward location of the ITCZ, which goes along with the strong positive SST bias in the Gulf of Guinea during JAS (García-Serrano et al., 2013). Interestingly, a common feature of the rainfall systematic error in all forecast systems considered is the apparent underestimation of the monsoonal precipitation in the westernmost part of the continent, the region extending over Senegal, Guinea, and Liberia. Finally, it is worth noting that there is no correspondence between a better WAM representation (lesser drift/bias) and a higher skill (García-Serrano et al., 2013) as confirmed in the Figure 12 .

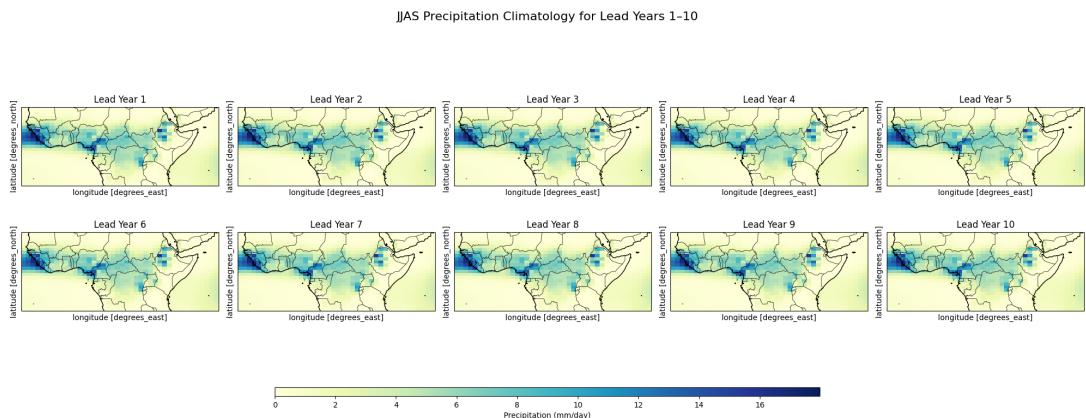


Figure 9: Observed JJAS Precipitation Climatology for Lead Years 1–10

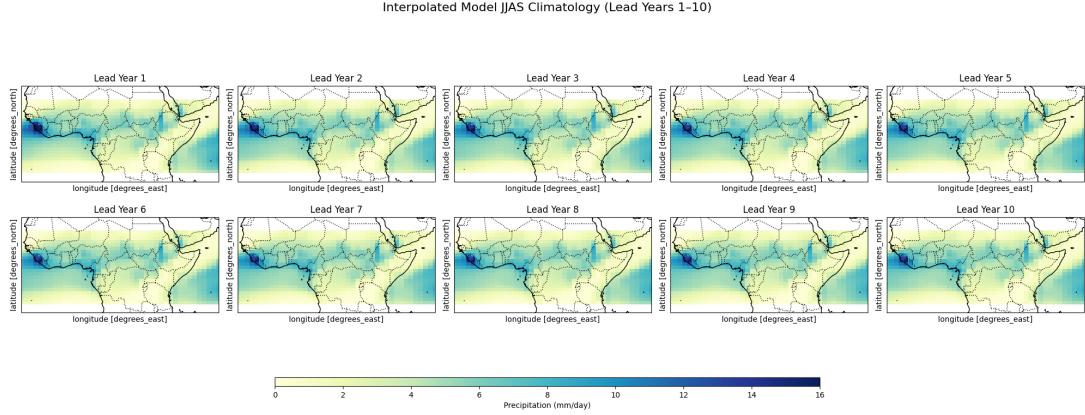


Figure 10: Simulated JJAS Precipitation Climatology for Lead Years 1–10

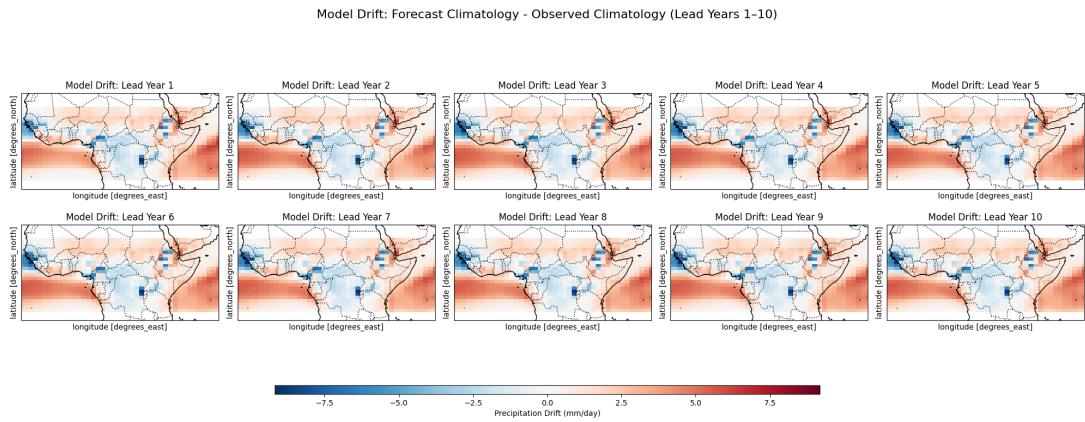


Figure 11: Model Drift: Forecast Climatology - Observed Climatology (Lead Years 1–10)

## 4.2 Spatial prediction skill of West Africa Monsoon

Figure 12 shows the spatial distributions of West Africa Monsoon prediction skill by the ensemble mean of all the available hindcast members. The window 2–5 drops the first year, representing interannual variability with diminished influence from initial conditions; some memory of the initial conditions is still there, but starting to fade. The window 6–9 represents longer lead times where predictive skill generally diminishes. The influence of the initial conditions is mostly gone. These predictions rely more on how the model simulates climate change and natural variability.

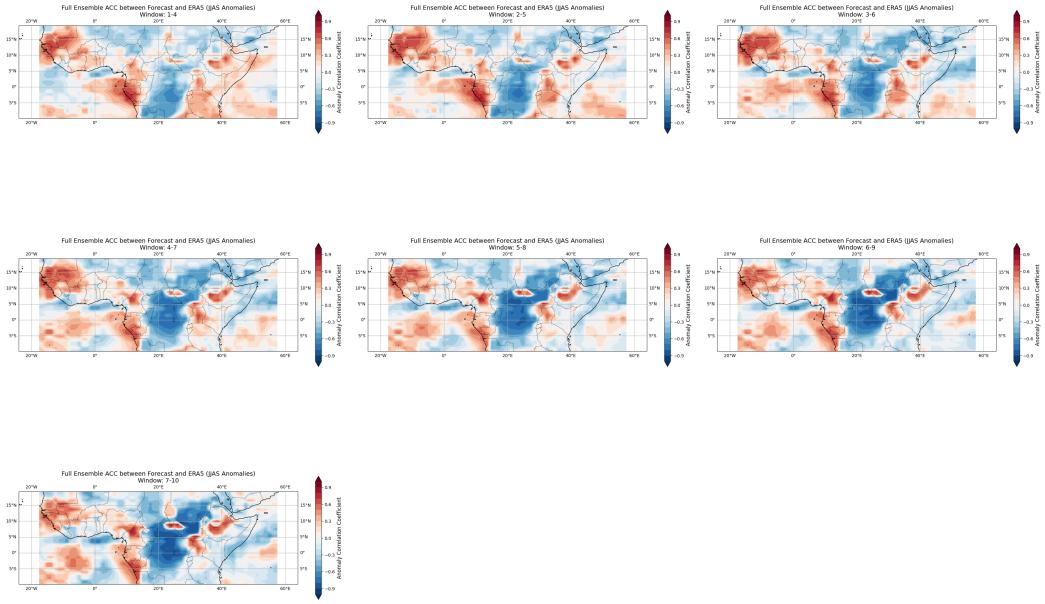


Figure 12: ACC at forecast years 1–4 to 7–10 between ERA5 and the precipitation fields from the decadal hindcasts.

Numerous studies have linked the multidecadal variability of NATL SST (often referred to as Atlantic multidecadal variability (AMV)) to seasonal climate fluctuations over Europe, Africa, Asia, and the Americas. It therefore seems likely that the CESM-DPLE surface climate skill in these regions stems in large part from (enhanced) skill at predicting NATL SST (S. G. Yeager et al., 2018). The skill improvement over the Sahel is also in line with our current understanding of AMV impacts over Africa (Otero et al., 2016) (S. G. Yeager et al., 2018) On multiyear time scales, a warmer NATL is associated with a northward shift in the intertropical convergence zone (ITCZ) and enhanced moisture supply for the West African monsoon (Green, 2017; Sheen et al., 2017).

The skill for regionally averaged JAS precipitation over the Sahel increases over the blue regions (where the skill improved with the increased of lead year) considerably from the window 2-5 to window 6-9 (Figure 13). (S. G. Yeager et al., 2018) speculate that this lead-time dependence is related to the changes in SST skill discussed, which result in notably higher ACC scores in the subtropical Atlantic when nontrend variability is isolated.

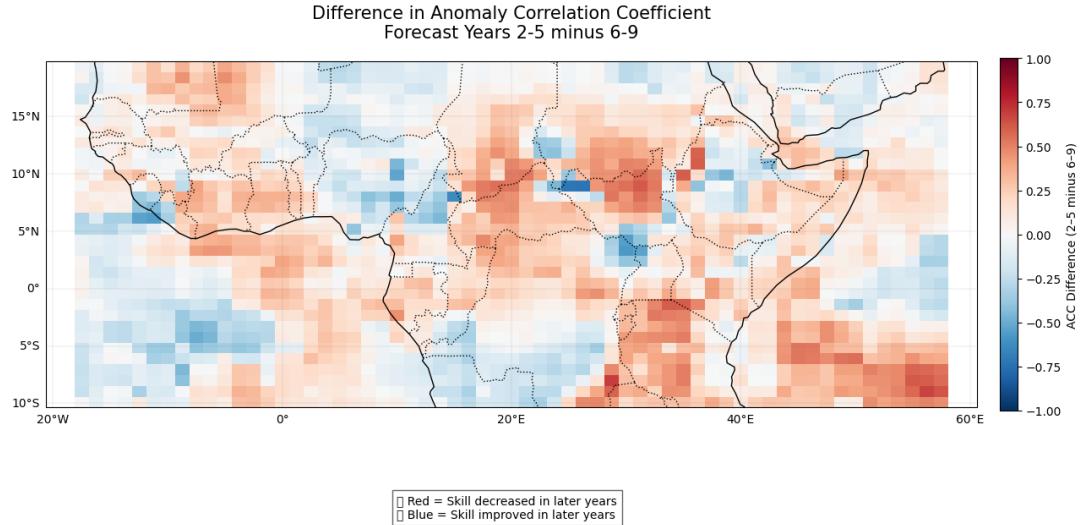


Figure 13: Difference in Anomaly Correlation Coefficient forecast years 2–5 and 6–9

Figure 14 illustrates the average Anomaly Correlation Coefficient (ACC) across different regions that shows  $\text{ACC} > 0.3$  over all the sequential 4 lead year periods. Each line represents a distinct regions: Guinea Coast ( $5^{\circ}\text{N}$ – $10^{\circ}\text{N}$ ,  $8^{\circ}\text{E}$ – $15^{\circ}\text{E}$ ), West Sahel ( $5^{\circ}\text{N}$ – $18^{\circ}\text{N}$ ,  $18^{\circ}\text{W}$ – $8^{\circ}\text{W}$ ), and West Equatorial Africa ( $6^{\circ}\text{S}$ – $3^{\circ}\text{N}$ ,  $5^{\circ}\text{E}$ – $14^{\circ}\text{E}$ ). For each region, the mean ACC was computed by averaging values from all grid points within the region, providing insight into regional forecast skill or climate model performance across time. The use of 4-year windows helps smooth short-term variability, highlighting broader temporal trends and regional disparities in ACC behavior.

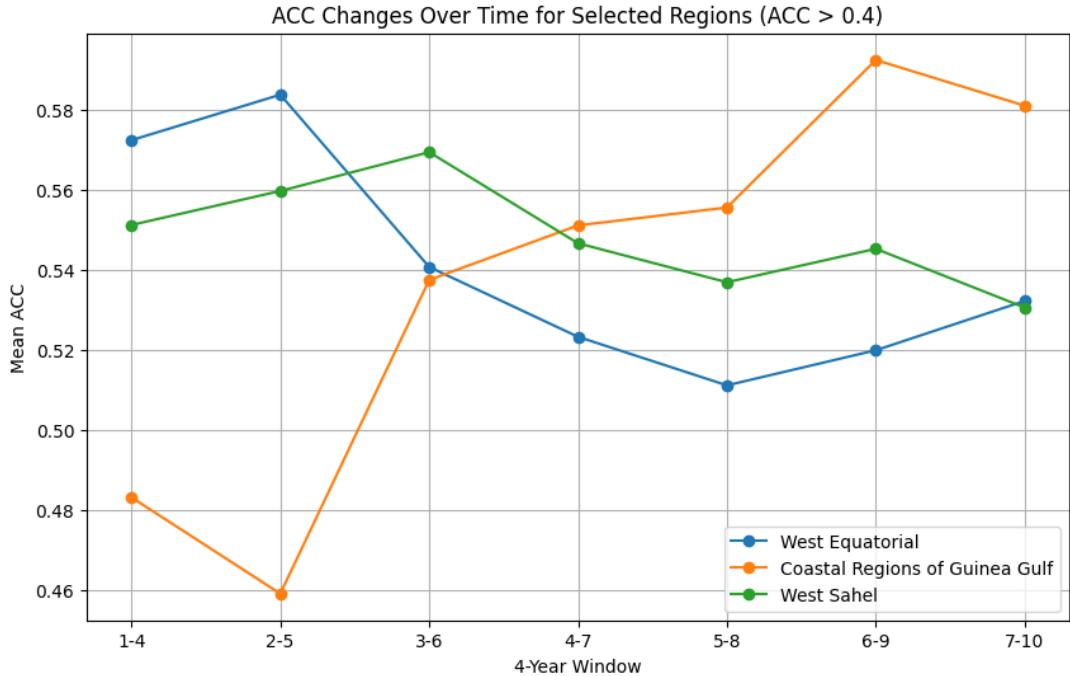


Figure 14: Mean Anomaly Correlation Coefficient (ACC) trends over time for selected regions with  $\text{ACC} >$  threshold. The plot highlights regional climate forecast skill evolution across West Equatorial Africa, the Coastal Guinea Gulf, and the West Sahel using 4-year moving windows.

### 4.3 CNN-Based ENSO Forecasts

ENSO (El Niño–Southern Oscillation) exerts its global impacts via ENSO teleconnection, and many studies pointed out the ENSO teleconnection and ENSO’s impacts greatly depend on the specific zonal distribution of ENSO-related SSTA over the equatorial Pacific, rather than the sole Niño3.4 index (Ashok et al. (2007); Jiang et al. (2019); W. Zhang, Mao, et al. (2021); W. Zhang et al. (2015)).

Over the past few decades, significant progress has been achieved in ENSO prediction; Jin et al. (2008); Barnston et al. (2012); B. Kirtman et al. (2001); Latif et al. (1998); Luo et al. (2016); Tippett et al. (2012)). Currently, both dynamical and statistical models can provide effective ENSO predictions at a lead time of 6–12 months (Barnston et al. (2019) ; Tang et al. (2018) ; Tippett et al. (2019)). Nevertheless, some challenges remain for long-lead-time ENSO predictions. Therefore, the accurate prediction of the SSTA zonal pattern over the equatorial Pacific is essential for improving global climate forecasts. However, current models have problems in predicting the SSTA zonal pattern. On one hand, dynamical models exhibit low prediction skills of the SSTA zonal pattern compared to their good performance in predicting the Niño3.4 index. As shown in Figure 15, eight North American Multimodel Ensemble (NMME; B. P. Kirtman et al. (2014)), dynamical models skillfully predict the December–February averaged Niño3.4 index at a lead time of 9–12 months. In contrast, the prediction skill of the SSTA zonal pattern over the equatorial Pacific in dynamical models quickly drops with increasing lead time (Figure 18 Sun et al. (2023), Figure 1b) compared to CNN-CD. Recently, deep learning (DL) methods have been widely applied to ENSO forecasts (Fang et al. (2022); Geng and Wang (2021); Ham et al. (2019); 2021; Hassanibesheli et al., 2022; He et al., 2019; Mu et al. (2021); Mu et al. (2022); Nooteboom et al. (2018); Saha and Nanjundiah (2020); Wang et al. (2021); Yan et al. (2020); L. Zhou et al. (2023)). Among various DL models, the convolutional neural network (CNN)-based models have achieved significant success in ENSO prediction. Ham et al. (2019) (hereafter H19) built a CNN.

Sun et al. (2023) introduce a convolutional neural network model utilizing cosine distance (CNN-CD) for predicting sea surface temperature anomalies (SSTA) in the equatorial Pacific.

#### 4.3.1 Prediction Skills of SSTA Over the Equatorial Pacific

In the CNN-CD model, the ACC exceeds 0.6 at a lead time of 14 months, surpassing the NMME dynamical models. The ACC in CNN-CD model can exceed 0.5 up to 17 months, indicating the CNN-CD model can skillfully predict Niño3.4 (Figure 16: Sun et al. (2023), Figure S1a in Supporting Information S1). In addition, the RSME in most dynamical models rises above one standard deviation (dashed lines) at a lead time of 10 months. In contrast, the RMSE of Niño3.4 predicted by the CNN-CD model stays below one standard deviation even at a lead time of 18 months and is lower than those in eight dynamical models and CNN-H19 at almost all lead times (Figure 17: Sun et al. (2023), Figure S1b in Supporting Information S1).

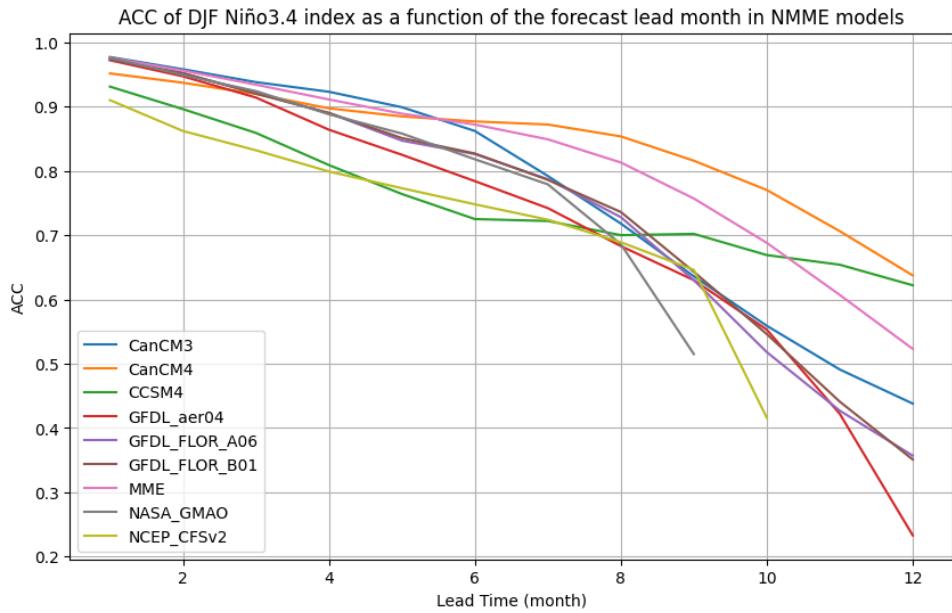


Figure 15: The ACC of DJF Niño3.4 index as a function of the forecast lead month in NMME models

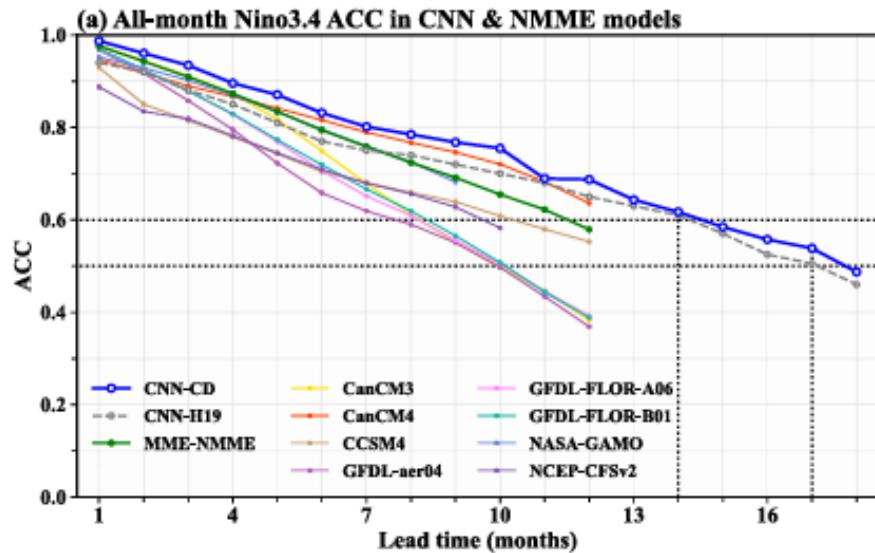


Figure 16: Prediction skills of the Niño3.4 index in CNN models and NMME models: ACC of three-month-running averaged Niño3.4 index as a function of forecast lead month in the CNN-CD (blue), CNN H19 (grey) and NMME models (other colors)

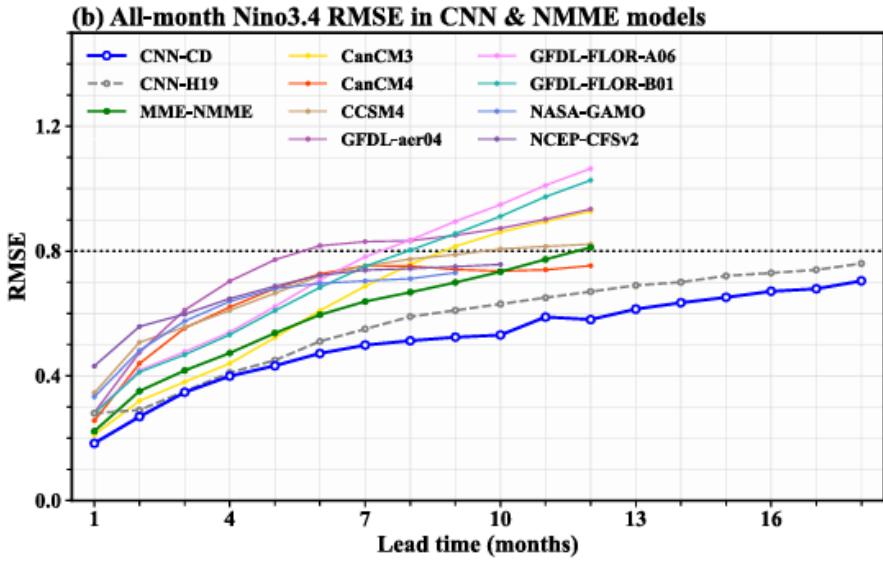


Figure 17: Prediction skills of the Niño3.4 index in CNN models and NMME models: RMSE of three-month-running averaged Niño3.4 index as a function of forecast lead month in the CNN-CD (blue), CNN H19 (grey) and NMME models (other colors)

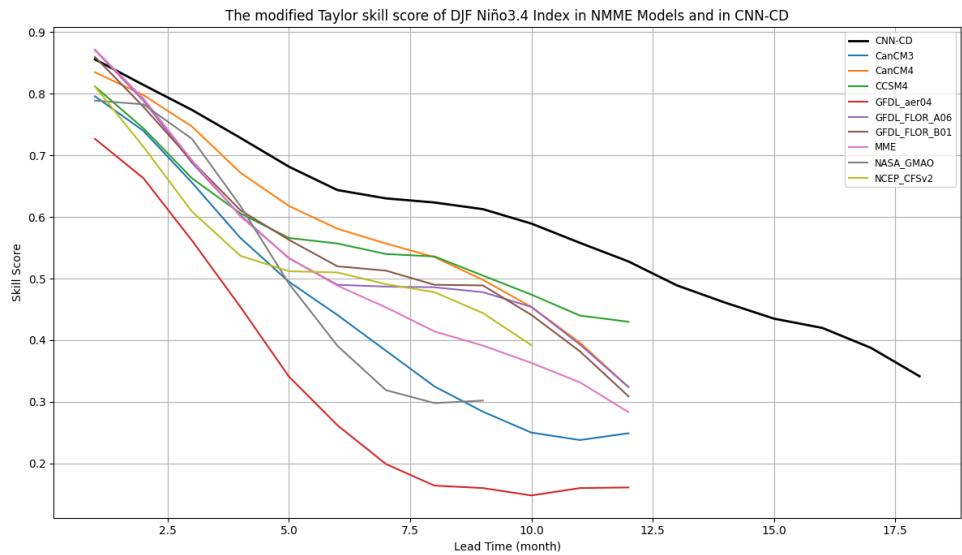


Figure 18: The modified Taylor skill score of DJF SSTA zonal pattern as a function of the lead month in CNN-CD model (blue) and NMME models(other colors)

At the 18-month lead, the CNN-CD still captures more than half of the ENSO events, although the magnitude of the predicted Niño3.4 is weaker than that of the observation (Figure 19). These results indicate the CNN-CD model can produce skillful forecasts of Niño3.4 index 1.5 years ahead, outperforming most current dynamical models.

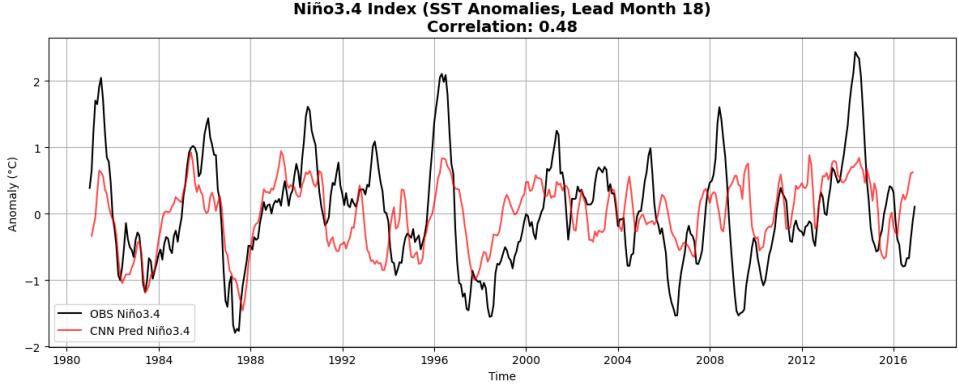


Figure 19: Niño3.4 Index (SST Anomalies, Lead Month 18) with correlation produced by observations (black) and by CNN-CD model

#### 4.3.2 Physical Interpretation of ENSO Forecasts at a 1.5 Year Lead

To interpret the physical basis behind the CNN-CD model’s forecasts, the model’s performance is examined at a 16-month lead time, with a particular focus on the precursory signals relevant to ENSO evolution. Since these signals may evolve slowly with time, understanding them at long lead times is critical. The results demonstrate that CNN-CD effectively learns physically meaningful precursors such as OHC evolution and uses them for long-lead ENSO forecasts. However, prediction challenges remain in cases of weak or irregular precursory signals and nonlinear influences from short-term atmospheric events. Successful El Niño predictions correspond to strong OHC anomalies consistent with a recharging phase, while failed forecasts show discharging or irregular patterns. For La Niña events, first-year occurrences are predicted more accurately than follow-up events. At lead month 16/18, recharge–discharge signals within the tropical Pacific dominate predictive skill (Sun et al. (2023)).

### 4.4 Mode Matching Technique

Process-oriented conditional subsampling of the ensemble is done in order to develop a deeper understanding of the critical mechanisms at play in near-term prediction (S. G. Yeager et al., 2018). (S. G. Yeager et al., 2018) suggest that DPLE precipitation skill is a lower bound that could be improved by reducing initialization shock in the Tropical Pacific. In the Sahel region ( $10^{\circ}$ – $20^{\circ}$ N,  $20^{\circ}$ W– $10^{\circ}$ E), even a 10-member ensemble shows clear benefits of initialization (Eade et al., 2014). With 40 members, the ratio of predictable components (RPC) reaches 2.05 for summer precipitation (JAS), indicating that real-world predictability is likely higher than suggested by the ensemble spread (Eade et al., 2014). The best members are chosen by selecting the ones that show the smallest absolute

differences between anomalies produced by CNN-CD model and anomalies produced by the CESM LENS project simulations (10 members for each year)(Figure 20).

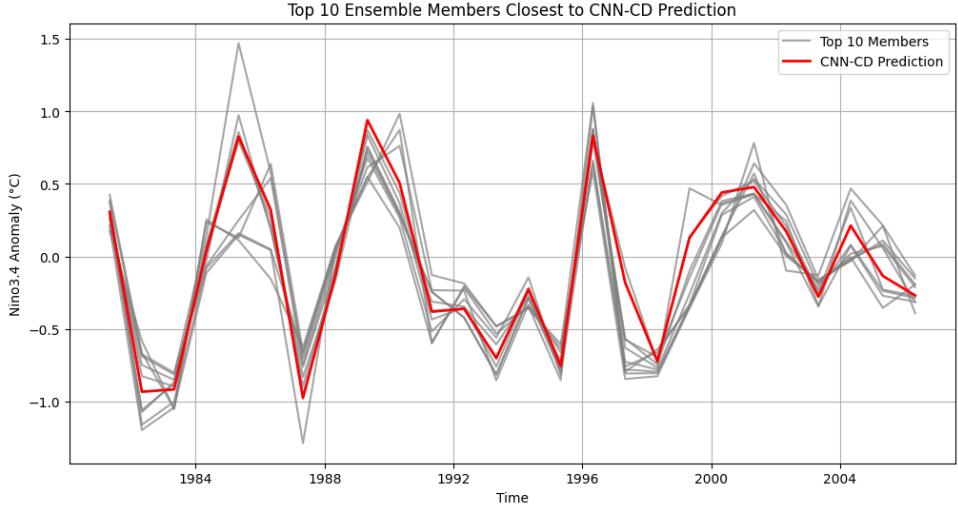


Figure 20: Niño 3.4 index in May by CNN-CD and by CESM LENS 10 selected ensemble members

By constraining the evaluation to these "matched" ensemble members, the goal is to:

- Quantify the extent to which West African rainfall variability is predictable given known ENSO evolution.
- Identify patterns of teleconnection strength and timing between the equatorial Pacific and the Sahel region.
- Improve seasonal-to-decadal climate forecasting for West Africa by leveraging high-skill ensemble selection.

For that selected members selected, the spatial prediction skill (ACC, see [Forecast quality and uncertainty measures for the boreal summer analysis](#)) is verified. Unbalanced initial conditions in the equatorial Pacific are hypothesized to give rise to spurious El Niño (La Niña) conditions that degrade the skill of initialized hindcasts at short lead times (Pohlmann et al., 2017; Teng et al., 2017). While the initialization shock in CESM-DPLE is much reduced compared to its predecessor (CCSM4-DP), there is still a significant improvement in SST skill in many tropical and extratropical regions as lead time increases. Despite this, skill still improves with lead time in many tropical and extratropical regions (S. G. Yeager et al., 2018). The early-lead-time skill degradation in western tropical Pacific may be related to the combination of spurious variability in the eastern tropical Pacific and model bias that extends ENSO activity too far to the west (van Oldenborgh et al., 2012). Looking at shorter lead time, the result (Figure 22) shows an improved skill over the Sahel region that can be interpreted as an improvement in the representation of the spurious variability in the eastern tropical Pacific and a model bias that extends ENSO activity too far to the west presented by the Full ensemble at shorter lead times (window

1-4, 2-5, 3-6) (van Oldenborgh et al., 2012). Indeed Sahel region is the region that shows a large initialization shock (S. G. Yeager et al., 2018), selecting the members that best represent the Niño3.4 partially solves this shock.

The influence of ENSO and the western Indian Ocean on Sahelian rainfall is largely confined to inter-annual timescales (Sheen et al. (2017)), in particular interannual rainfall variability is stronger along the Guinea Coast, than the Sahel that is dominated by pronounced decadal variability (Garcia 2013). As shown in (Figure 22), the skill along the Guinea coast is partly improved. Emmanuel (2022) suggest that strong El Niño episodes correspond to positive precipitation anomalies during the wet season (JJAS) in the Guinea region.

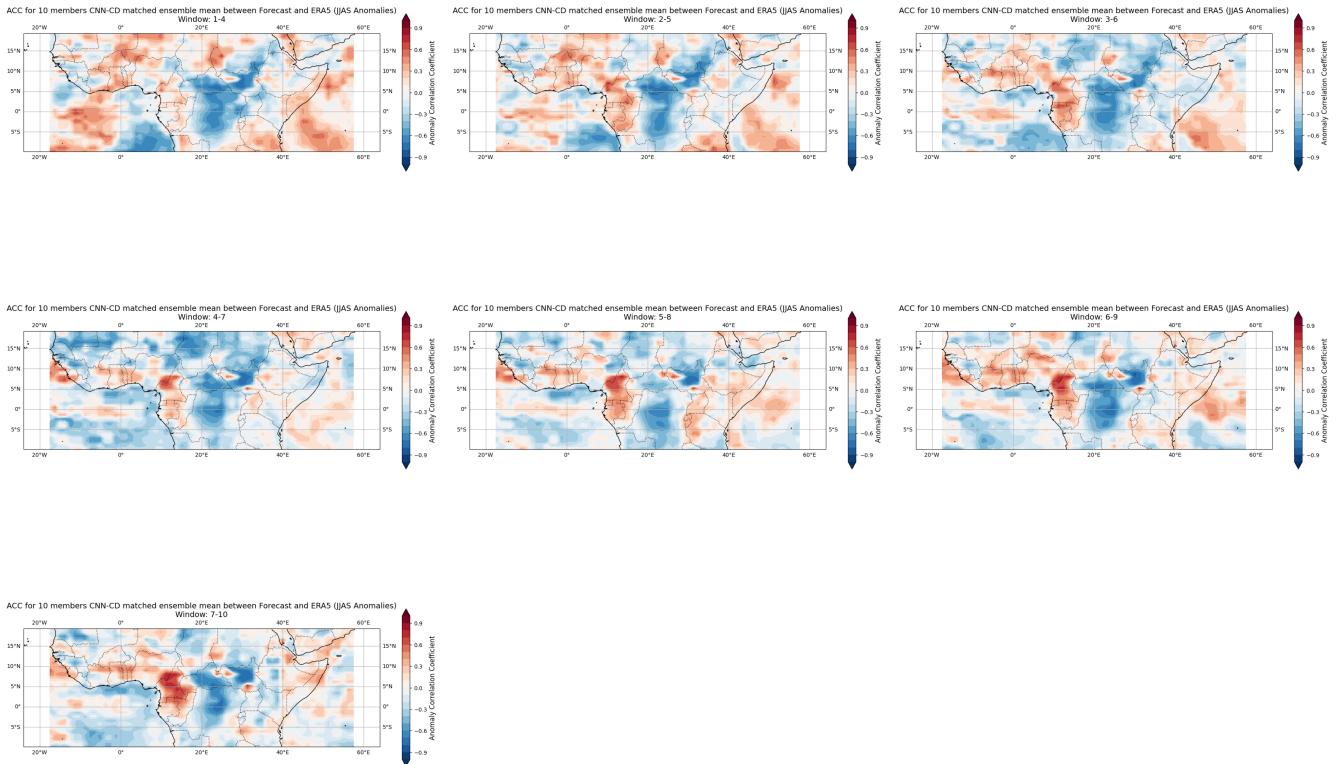


Figure 21: ACC for the 10 selected members JJAS anomalies

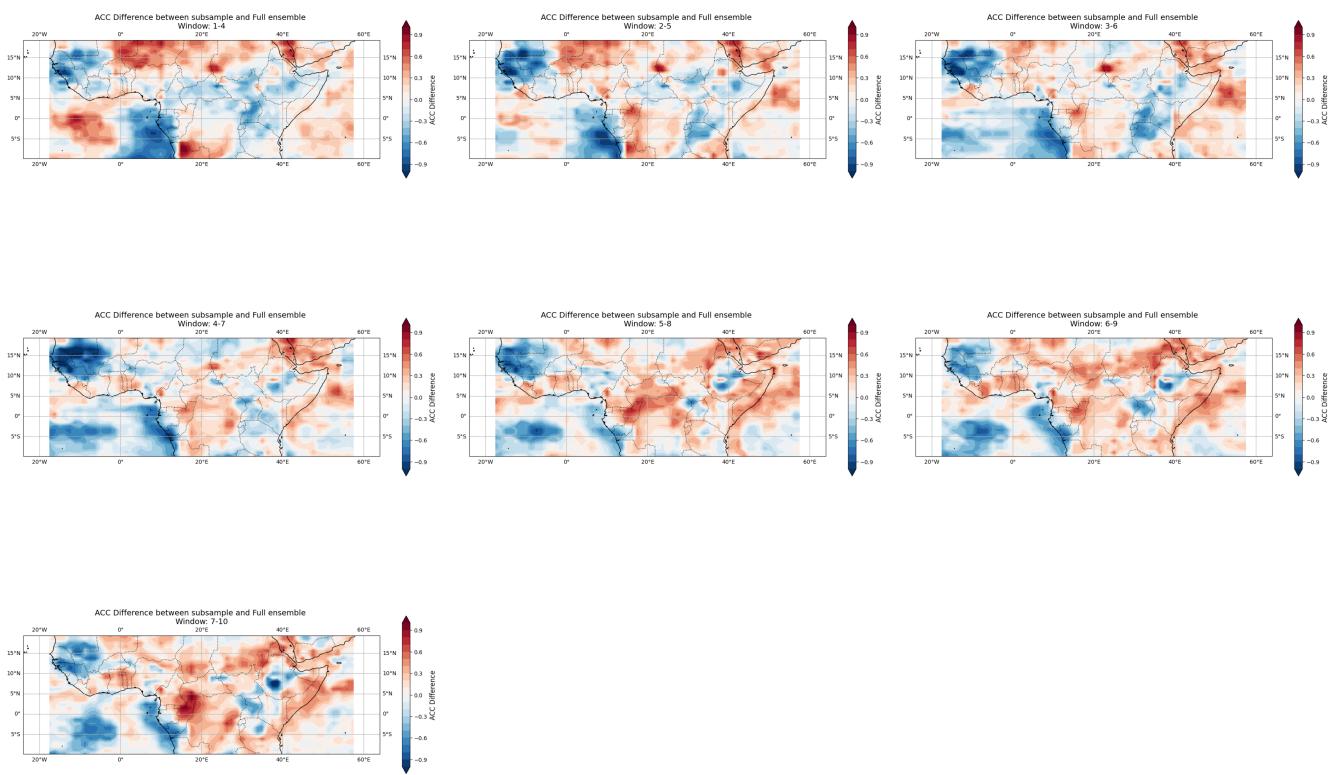


Figure 22: ACC Difference between subsample and Full ensemble

## 5 Discussions and Conclusions

High predictability of near-surface air temperature in the North Atlantic is largely linked to improved initialization of the Atlantic Meridional Overturning Circulation (AMOC). This predictability extends to the tropical Atlantic, Sahel and West Africa precipitations through established climate teleconnections (N. J. Dunstone et al., 2011; Eade et al., 2014; Pohlmann et al., 2013; Robson et al., 2014; S. Yeager et al., 2012b; R. Zhang and Delworth, 2006). In this study, further improvements to the research study proposed by (Smith et al., 2020), were achieved by selecting the most skillful models MIROC6, MIROC5, CanCM5, CanESM5, and GFDL-CM2.1 based on their precipitation anomaly correlation coefficient (ACC) over 10°S–20°N and 20°W–60°E. The lagged ensemble of 184 members and subsequent NAO-matching procedure enhanced prediction skill, particularly in regions linked to the North Atlantic–Sahel/West Africa teleconnection. These results suggest that targeted model selection and teleconnection-based ensemble refinement can improve decadal predictability in regions influenced by Atlantic variability. Selecting the best models as suggested by (Smith et al., 2020) shows some improvements in the spatial distribution of the skill (Figure 7) compared to (Smith et al., 2020) and indicates where effectively the NAO teleconnection affects the skill (Figure 8).

During the boreal winter skill from initialized simulations over Sahel and West Africa are remarked as shown in (Smith et al., 2019). Improvements from initialization generally occur, but where uninitialized simulations already have some skill like in West Africa and Sahel, as shown in (Smith et al., 2019). This could arise because improved skill arises from predicting internal variability where there is an externally forced response and this particularly expected where there is a long term driven by slow variations in greenhouse gases. In addition, there is some evidence that the response of atmospheric circulation to Arctic sea-ice loss (Mori et al., 2019) and to external factors (Smith et al., 2019) such as volcanic eruptions, solar variations, anthropogenic aerosols and ozone changes are too weak in models, highlighting the need for improved understanding of the roles of other external factors.

CESM-DPLE is revealing significant and useful skill in predicting low frequency variations in hydroclimate over land such as over Africa, that appears to highlight the role of the ocean in modulating decadal climate variations. Large ensembles are needed to draw robust conclusions about the role of initialization in predictions of noisy atmospheric fields and to realize skill given what appear to be unrealistically small signal-to-noise ratios in the model (S. G. Yeager et al., 2018). The skill of CESM-DPLE over the Sahel and surrounding regions is consistent with the influence of North Atlantic SST variability (AMV) on seasonal and multiyear climate fluctuations (Otero et al., 2016; S. G. Yeager et al., 2018). On multiyear timescales, a warmer North Atlantic is linked to a northward shift of the ITCZ, enhancing moisture supply for the West African monsoon (Green, 2017; Sheen et al., 2017). These results align with the established teleconnections connecting North Atlantic variability to West African climate (Figure 12). The use of large ensembles facilitates process-oriented conditional subsampling, allowing a deeper

understanding of the mechanisms driving near-term climate predictability. Since the magnitude of teleconnection is often underestimated due to the signal-to-noise paradox, hybrid forecasting offers a pathway to combine ensemble members more effectively by weighting their varying information content. In particular, mode-matching or targeted sub-selection of ensemble members that best capture key modes of climate variability can enhance decadal forecast skill. The results presented here demonstrate that careful selection of high-skill ensemble members improves the predictability of Sahelian rainfall. By constraining members based on their Niño3.4 evolution produced by CNN-CD (Figure 20), the initialization shock in the equatorial Pacific is partially mitigated, leading to improved skill at short lead times (Figures 20, 22). This hybrid approach captures the influence of ENSO and the western Indian Ocean on interannual rainfall variability, particularly along the Guinea Coast, while also improving the representation of decadal variability in the Sahel (Emmanuel, 2022; García-Serrano et al., 2013; Sheen et al., 2017). It may also be noted that to confidently beat persistence in the Sahel, however, probably requires an initialized ensemble of 30 or more. In short, the confidence intervals for predictions of precipitation over land are large, particularly for the baseline skill associated with external forcing, which is very ill defined for UI (Uninitialized) ensembles of size 10 or less, which are not uncommon in the DP (Decadal Prediction) literature: even if the model shows some skill, the uncertainty around that estimate is so big that it is hard to say with confidence whether the skill is real (S. G. Yeager et al., 2018). The implication is that robust assessment of DP skill enhancement associated with initialization for fields such as precipitation may require much larger ensembles than current protocols recommend (Boer et al., 2013).

The growing demand for reliable climate information to support adaptation and risk management underscores the need for seamless hydroclimatic prediction systems capable of providing consistent forecasts from weeks to decades ahead. Reliable long-term predictions of climate impacts on floods and droughts remain limited because weather and climate prediction systems have traditionally been developed for distinct temporal and application domains—short-term forecasts (less than five years) depend on accurate initialization, while long-term projections rely mainly on external forcings such as greenhouse gas concentrations (Boer et al., 2016). One promising pathway toward seamless prediction is to constrain uninitialized climate projections using initialized decadal predictions, which better capture observed variability. D. J. Befort et al. (2020) showed that selecting climate projections consistent with decadal hindcasts enhances predictive skill even beyond the initialized forecast horizon, while D. Befort et al. (2022) demonstrated that bias-correction methods such as variance inflation can reduce inconsistencies between decadal predictions and long-term projections. Although challenges remain particularly for extreme hydroclimatic events these studies highlight the potential of hybrid modeling frameworks that merge initialized and uninitialized simulations to deliver more reliable, seamless forecasts of hydroclimatic variability and change.

These arguments concern the complexity of understanding the signal to noise paradox (SNP). Recently, (Mahmood, 2025) provide important insights into the persistence of the

SNP under idealized perfect-model conditions. The results demonstrate that the paradox can occur even when the model is internally consistent and perfectly initialized, suggesting that it is not solely a product of model errors or initialization imperfections. Notably, the perfect-model simulations exhibit regions with high RPC, implying that the model can appear to predict its own reference run more accurately than individual ensemble members.

This behavior arises primarily from the way ensemble predictions are processed and analyzed rather than from inherent model deficiencies. In particular, concatenating predictions initialized in consecutive but independent years into a single time series disrupts the natural temporal autocorrelation of each member. Such concatenation effectively combines independent trajectories that did not evolve continuously in time, yielding an artificially smoothed ensemble mean. Consequently, the correlation between individual members and the ensemble mean decreases, while the reference time series unaffected by discontinuities retains stronger temporal coherence. This imbalance inflates the RPC, making it appear that the ensemble mean is more reliable than its constituent members. Therefore, high RPC values should not automatically be interpreted as a sign of model inadequacy but rather as a statistical artifact stemming from ensemble concatenation and data-processing procedures.

## References

- Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El niño modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(C11), C11007. <https://doi.org/10.1029/2006JC003798> (cit. on p. 35).
- Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal enso model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, 93(5), 631–651. <https://doi.org/10.1175/BAMS-D-11-00111.1> (cit. on p. 35).
- Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019). Deterministic skill of enso predictions from the north american multimodel ensemble. *Climate Dynamics*, 53(12), 7215–7234. <https://doi.org/10.1007/s00382-017-3603-3> (cit. on p. 35).
- Befort, D., Brunner, L., Borchert, L., O'Reilly, C., Mignot, J., Ballinger, A., Hegerl, G., Murphy, J., & Weisheimer, A. (2022). Combination of decadal predictions and climate projections in time: Challenges and potential solutions. *Geophysical Research Letters*, 49, e2022GL098568. <https://doi.org/10.1029/2022GL098568> (cit. on p. 43).
- Befort, D. J., O'Reilly, C. H., & Weisheimer, A. (2020). Constraining projections using decadal predictions. *Geophysical Research Letters*, 47, e2020GL087900. <https://doi.org/10.1029/2020GL087900> (cit. on p. 43).
- Behringer, D., & Xue, Y. (2004). Evaluation of the global ocean data assimilation system at ncep: The pacific ocean. (Cit. on p. 22).
- Biasutti, M., & Giannini, A. (2006). Robust sahel drying in response to late 20th century forcings. *Geophysical Research Letters*, 33, L11706. <https://doi.org/10.1029/2006GL026067> (cit. on p. 9).
- Biasutti, M., Held, I. M., Sobel, A. H., & Giannini, A. (2008). Sst forcings and sahel rainfall variability in simulations of the twentieth and twenty-first centuries. *Journal of Climate*, 21(14), 3471–3486. <https://doi.org/10.1175/2007JCLI1896.1> (cit. on p. 9).
- Boer, G. J., Kharin, V. V., & Merryfield, W. J. (2013). Decadal predictability and forecast skill. *Climate Dynamics*, 41(7–8), 1817–1833. <https://doi.org/10.1007/s00382-013-1705-0> (cit. on pp. 11, 43).
- Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., & Eade, R. (2016). The decadal climate prediction project (dcpp) contribution to cmip6. *Geoscientific Model Development*, 9(10), 3751–3777. <https://doi.org/10.5194/gmd-9-3751-2016> (cit. on pp. 3, 13, 43).
- Bracegirdle, T. J. (2018). Do cmip5 models reproduce observed low-frequency variability in the north atlantic jet strength? *Geophysical Research Letters*, 45(24), 13, 409–13, 417. <https://doi.org/10.1029/2018GL078965> (cit. on p. 27).

- Branstator, G., & Teng, H. (2010). Two limits of initial-value decadal predictability in a cgcm. *Journal of Climate*, 23, 6292–6310. <https://doi.org/10.1175/2010JCLI3678.1> (cit. on p. 7).
- Caminade, C., & Terray, L. (2010). Twentieth century sahel rainfall variability as simulated by the arpege agcm, and future changes. *Climate Dynamics*, 35(1), 75–94. <https://doi.org/10.1007/s00382-009-0545-4> (cit. on p. 9).
- Carton, J. A., & Giese, B. S. (2008). A reanalysis of ocean climate using simple ocean data assimilation (soda). *Monthly Weather Review*, 136(8), 2999–3017. <https://doi.org/10.1175/2007MWR1978.1> (cit. on p. 22).
- Chen, X., & McKay-Crites, V. (2018). Adam: A method for stochastic optimization [Accessed: 2025-08-19]. (Cit. on p. 21).
- Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, 38(3–4), 527–547. <https://doi.org/10.1007/s00382-010-0977-x> (cit. on p. 7).
- Doblas-Reyes, F. J., Balmaseda, M. A., Weisheimer, A., & Palmer, T. N. (2011). Decadal climate prediction with the european centre for medium-range weather forecasts coupled forecast system: Impact of ocean observations. *Journal of Geophysical Research: Atmospheres*, 116, D19111. <https://doi.org/10.1029/2010JD015394> (cit. on p. 4).
- Doblas-Reyes, F. J., García-Serrano, J., Liennert, F., Biescas, A. P., & Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *WIREs Climate Change*, 4(3), 245–268. <https://doi.org/10.1002/wcc.217> (cit. on pp. 8, 29).
- Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., Düsterhus, A., Fröhlich, K., Pohlmann, H., & Baehr, J. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, 45(8), 3605–3614. <https://doi.org/10.1002/2018GL077209> (cit. on p. 16).
- Druyan, L. M. (2010). Studies of 21st-century precipitation trends over west africa. *International Journal of Climatology*, 31, 1415–1572. <https://doi.org/10.1002/joc.2180> (cit. on p. 9).
- Dunstone, N., Scaife, A., Hermanson, R., Eade, R., Robinson, N., Andrews, M., & Knight, J. (2016). Skilful predictions of the winter north atlantic oscillation one year ahead. *Nature Geoscience*, 9, 809–814. <https://doi.org/10.1038/ngeo2824> (cit. on pp. 8, 13).
- Dunstone, N. J., Smith, D. M., & Eade, R. (2011). Multi-year predictability of the tropical atlantic atmosphere driven by the high latitude north atlantic ocean. *Geophysical Research Letters*, 38, L14701. <https://doi.org/10.1029/2011GL047949> (cit. on pp. 8, 29, 42).
- Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41, 5620–5628. <https://doi.org/10.1002/2014GL061146> (cit. on pp. 7, 8, 15, 29, 38, 42).

- Emmanuel, I. (2022). Linkages between el niño–southern oscillation (enso) and precipitation in west africa regions. *Arabian Journal of Geosciences*, 15(7), 675. <https://doi.org/10.1007/s12517-022-09942-2> (cit. on pp. 40, 43).
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016> (cit. on p. 22).
- Fang, W., Sha, Y., & Sheng, V. S. (2022). Survey on the application of artificial intelligence in enso forecasting. *Mathematics*, 10(20), 3793. <https://doi.org/10.3390/math10203793> (cit. on p. 35).
- Folland, C. K., Palmer, T. N., & Parker, D. E. (1986). Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature*, 320, 602–607. <https://doi.org/10.1038/320602a0> (cit. on p. 9).
- Fontaine, B. (1996). Sea surface temperature fields associated with west african rainfall anomaly types. *Journal of Climate*, 9(11), 2935–2940. [https://doi.org/10.1175/1520-0442\(1996\)009<2935:SSTFAW>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<2935:SSTFAW>2.0.CO;2) (cit. on p. 9).
- Fontaine, B., Janicot, S., & Moron, V. (1995). Rainfall anomaly patterns and wind field signals over west africa in august (1958–1989). *Journal of Climate*, 8(6), 1503–1510. [https://doi.org/10.1175/1520-0442\(1995\)008<1503:RAPAWF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1503:RAPAWF>2.0.CO;2) (cit. on p. 9).
- Fontaine, B., Trzaska, S., & Janicot, S. (1998). Evolution of the relationship between near global and atlantic sst modes and the rainy season in west africa: Statistical analyses and sensitivity experiments. *Climate Dynamics*, 14, 353–368. <https://doi.org/10.1007/s003820050228> (cit. on p. 9).
- Gaetani, M., Mohino, E., Dosio, A., & Bader, J. (2013). Decadal prediction of the sahelian precipitation in cmip5 simulations. *Journal of Climate*, 26(19), 7708–7719. <https://doi.org/10.1175/JCLI-D-12-00635.1> (cit. on pp. 13, 18, 19).
- García-Serrano, J., Doblas-Reyes, F. J., & Coelho, C. A. S. (2012). On the assessment of near-surface global temperature and north atlantic multi-decadal variability in the ensembles decadal hindcast. *Climate Dynamics*, 39, 2025–2040. <https://doi.org/10.1007/s00382-012-1413-1> (cit. on p. 18).
- García-Serrano, J., Doblas-Reyes, F. J., & Coelho, C. A. S. (2013). Decadal prediction of the dominant west african monsoon. *Journal of Geophysical Research: Atmospheres*, 118(12), 6455–6468. <https://doi.org/10.1002/jgrd.50465> (cit. on pp. 31, 43).
- Geng, H., & Wang, T. (2021). Spatiotemporal model based on deep learning for enso forecasts. *Atmosphere*, 12(7), 810. <https://doi.org/10.3390/atmos12070810> (cit. on p. 35).
- Giannini, A., Saravanan, R., & Chang, P. (2003). Oceanic forcing of sahel rainfall on interannual to interdecadal time scales. *Science*, 302(5647), 1027–1030. <https://doi.org/10.1126/science.1089357> (cit. on p. 9).

- Giannini, A., Saravanan, R., & Chang, P. (2005). Dynamics of the boreal summer african monsoon in the nsipp1 atmospheric model. *Climate Dynamics*, 25, 517–535. <https://doi.org/10.1007/s00382-005-0056-x> (cit. on p. 9).
- Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., Mason, S. J., Kirtman, B. P., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C. A. T., Stephenson, D. B., Meehl, G. A., ... Delworth, T. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40(1-2), 245–272. <https://doi.org/10.1007/s00382-012-1481-2> (cit. on pp. 4, 6, 7, 18).
- Green, B. (2017). Twentieth century correlations between extratropical sea surface temperatures and the intertropical convergence zone. *Geophysical Research Letters*, 44(23), 11, 644–11, 652. <https://doi.org/10.1002/2017GL075044> (cit. on pp. 33, 42).
- Griffies, S. M., & Bryan, K. (1997). A predictability study of simulated north atlantic multidecadal variability. *Climate Dynamics*, 13, 459–487. <https://doi.org/10.1007/s003820050177> (cit. on p. 7).
- Haarsma, R. J., Selten, F. M., Weber, S. L., & Kliphuis, M. (2005). Sahel rainfall variability and response to greenhouse warming. *Geophysical Research Letters*, 32, L17702. <https://doi.org/10.1029/2005GL023232> (cit. on p. 9).
- Ham, Y.-G., Kim, J.-H., Kim, E.-S., & On, K.-W. (2021). Unified deep learning model for el niño/southern oscillation forecasts by incorporating seasonality in climate data. *Science Bulletin*, 66(13), 1358–1366. <https://doi.org/10.1016/j.scib.2021.03.009> (cit. on p. 26).
- Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature*, 573(7775), 568–572. <https://doi.org/10.1038/s41586-019-1559-7> (cit. on pp. 19, 21, 35).
- Hazeleger, W., Severijns, C., van Oldenborgh, G. J., Sterl, A., Drijfhout, S. S., Dijkstra, H. A., van den Hurk, B. J. J. M., Karspeck, A. R., van der Linden, E. C., van der Wiel, K., & Yiou, P. (2013). Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 40(9), 1794–1798. <https://doi.org/10.1002/grl.50355> (cit. on p. 5).
- Held, I. M., et al. (2005). Simulation of sahel drought in the 20th and 21st centuries. *Proceedings of the National Academy of Sciences*, 102, 17891–17896. <https://doi.org/10.1073/pnas.0509057102> (cit. on p. 9).
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023). *Era5 hourly data on single levels from 1940 to present* [(Accessed on DD-MMM-YYYY)]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.adbb2d47> (cit. on pp. 13, 18).

- Hoerling, M., Hurrell, J., Eischeid, J., & Phillips, A. (2006). Detection and attribution of twentieth-century northern and southern african rainfall change. *Journal of Climate*, 19, 3989–4008. <https://doi.org/10.1175/JCLI3842.1> (cit. on p. 9).
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H.-M. (2017). Extended reconstructed sea surface temperature, version 5 (ersstv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205. <https://doi.org/10.1175/JCLI-D-16-0836.1> (cit. on p. 22).
- Hurrell, J. W., & van Loon, H. (1997). Decadal variations in climate associated with the north atlantic oscillation. *Climatic Change*, 36(3), 301–326. <https://doi.org/10.1023/A:1005314315270> (cit. on p. 27).
- International CLIVAR Project Office. (2011). *Data and bias correction for decadal climate predictions* (CLIVAR Publication Series No. 150). World Climate Research Programme. (Cit. on pp. 5, 17–19).
- Janicot, S., Harzallah, A., Fontaine, B., & Moron, V. (1998). West african monsoon dynamics and eastern equatorial atlantic and pacific sst anomalies (1970–88). *Journal of Climate*, 11(8), 1874–1882. [https://doi.org/10.1175/1520-0442\(1998\)011<1874:WAMDAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<1874:WAMDAE>2.0.CO;2) (cit. on p. 9).
- Jiang, F., Zhang, W., Geng, X., Stuecker, M. F., & Liu, C. (2019). Impacts of central pacific el niño on southern china spring precipitation controlled by its longitudinal position. *Journal of Climate*, 32(22), 7823–7836. <https://doi.org/10.1175/JCLI-D-19-0266.1> (cit. on p. 35).
- Jin, E. K., Kinter, J. L. I., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B. P., Kug, J.-S., Kumar, A., Luo, J.-J., Schemm, J., Shukla, J., & Yamagata, T. (2008). Current status of enso prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, 31(6), 647–664. <https://doi.org/10.1007/s00382-008-0397-3> (cit. on p. 35).
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P. J., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Muñoz, E., Neale, R., Oleson, K., ... Vertenstein, M. (2015). The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8), 1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1> (cit. on p. 17).
- Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Evaluation of short-term climate change prediction in multi-model cmip5 decadal hindcasts. *Geophysical Research Letters*, 39(10), L10701. <https://doi.org/10.1029/2012GL051644> (cit. on p. 18).
- Kirtman, B. P., Min, D., Infant, J. M., Kinter III, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Pena Mendez, M., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., ... Wood, E. F. (2014). The north american multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal

- prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1> (cit. on pp. 24, 35).
- Kirtman, B., Shukla, J., Balmaseda, M., Graham, N., Penland, C., Xue, Y., & Zebiak, S. (2001). *Current status of enso forecast skill: A report to the clivar working group on seasonal to interannual prediction* (tech. rep. No. ICPO Publication Series 56). International CLIVAR Project Office. Southampton, UK. (Cit. on p. 35).
- Koenigk, T., & Mikolajewicz, U. (2009). Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Climate Dynamics*, 32, 783–798. <https://doi.org/10.1007/s00382-008-0451-1> (cit. on p. 7).
- Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A., & Schneider, E. (1998). A review of the predictability and prediction of enso. *Journal of Geophysical Research: Oceans*, 103(C9), 14375–14393. <https://doi.org/10.1029/97JC03413> (cit. on p. 35).
- Losada, T., et al. (2012). Tropical sst and sahel rainfall: A non-stationary relationship. *Geophysical Research Letters*, 39, L21710. <https://doi.org/10.1029/2012GL052423> (cit. on p. 9).
- Lu, J., & Delworth, T. L. (2005). Oceanic forcing of the late 20th century sahel drought. *Geophysical Research Letters*, 32, L22706. <https://doi.org/10.1029/2005GL023316> (cit. on p. 9).
- Luo, J.-J., Yuan, C., Sasaki, W., Behera, S. K., Masumoto, Y., Yamagata, T., Lee, J.-Y., & Masson, S. (2016). Current status of intraseasonal–seasonal-to-interannual prediction of the indo-pacific climate. In T. Yamagata, J.-Y. Lee, & S. Masson (Eds.), *Indo-pacific climate variability and predictability* (pp. 63–107). World Scientific. [https://doi.org/10.1142/9789814696623\\_0003](https://doi.org/10.1142/9789814696623_0003) (cit. on p. 35).
- Madadgar, S., AghaKouchak, A., Shukla, S., Wood, A. W., Cheng, L., Hsu, K.-L., & Svoboda, M. (2016). A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern united states. *Water Resources Research*, 52(7), 5095–5110. <https://doi.org/10.1002/2015WR018547> (cit. on p. 10).
- Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, 41(9–10), 2393–2409. <https://doi.org/10.1007/s00382-012-1599-2> (cit. on p. 5).
- Mahmood, R. (2025). A perfect-model perspective on the signal-to-noise paradox in climate prediction. *Journal of Climate*, 38(12). <https://doi.org/10.1175/JCLI-D-24-0381.1> (cit. on p. 43).
- Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., & Albers, J. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, 101(5), E608–E625. <https://doi.org/10.1175/BAMS-D-18-0326.1> (cit. on pp. 2, 4).

- Martin, E. R., Thorncroft, C. D., & Chang, P. (2014). The impact of the amo on the west african monsoon annual cycle. *Quarterly Journal of the Royal Meteorological Society*, 140(680), 1609–1622. <https://doi.org/10.1002/qj.2107> (cit. on p. 9).
- Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Riendecker, M., ... Yeager, S. (2014). Decadal climate prediction: An update from the trenches. *Bulletin of the American Meteorological Society*, 95(2), 243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1> (cit. on pp. 3–6, 18).
- Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., & Stockdale, T. (2009). Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, 90(10), 1467–1485. <https://doi.org/10.1175/2009BAMS2778.1> (cit. on p. 3).
- Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., Kim, H., Kirtman, B. P., Kushnir, Y., Lovenduski, N. S., Mann, M. E., Merryfield, W. J., Nieves, V., Pégion, K., Rosenbloom, N., ... Xie, S.-P. (2021). Initialized earth system prediction from subseasonal to decadal timescales. *Nature Reviews Earth Environment*, 2, 340–357. <https://doi.org/10.1038/s43017-021-00155-x> (cit. on p. 2).
- Mohino, E., Janicot, S., & Bader, J. (2011). Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Climate Dynamics*, 37, 419–440. <https://doi.org/10.1007/s00382-010-0867-2> (cit. on p. 9).
- Mori, M., Kosaka, Y., Watanabe, M., Nakamura, H., & Kimoto, M. (2019). A reconciled estimate of the influence of arctic sea-ice loss on recent eurasian cooling. *Nature Climate Change*, 9, 123–129. <https://doi.org/10.1038/s41558-018-0379-3> (cit. on p. 42).
- Moulds, S., et al. (2023). Skillful decadal flood prediction. *Geophysical Research Letters*, 50(6), e2022GL100650. <https://doi.org/10.1029/2022GL100650> (cit. on p. 11).
- Mu, B., Qin, B., & Yuan, S. (2021). Enso-asc 1.0.0: Enso deep learning forecast model with a multivariate air-sea coupler. *Geoscientific Model Development*, 14(11), 6977–6999. <https://doi.org/10.5194/gmd-14-6977-2021> (cit. on p. 35).
- Mu, B., Qin, B., & Yuan, S. (2022). Enso-gtc: Enso deep learning forecast model with a global spatial-temporal teleconnection coupler. *Journal of Advances in Modeling Earth Systems*, 14(12), e2022MS003132. <https://doi.org/10.1029/2022MS003132> (cit. on p. 35).
- Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., & Gudgel, R. (2016). Statistical-dynamical seasonal forecast of north atlantic and us landfalling tropical cyclones using the high-resolution gfdl flor coupled model. *Monthly Weather Review*, 144(6), 2101–2123. <https://doi.org/10.1002/2015MS000607> (cit. on p. 10).

- Neri, A., Villarini, G., & Napolitano, F. (2020). Intraseasonal predictability of the duration of flooding above national weather service flood warning levels across the us midwest. *Hydrological Processes*, 34(24), 4505–4511. <https://doi.org/10.1002/hyp.13902> (cit. on p. 10).
- Nicholson, S. E. (1993). An overview of african rainfall fluctuations of the last decade. *Journal of Climate*, 6(7), 1463–1466. [https://doi.org/10.1175/1520-0442\(1993\)006<1463:AOOARF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1463:AOOARF>2.0.CO;2) (cit. on p. 9).
- Nooteboom, P. D., Feng, Q. Y., López, C., Hernández-García, E., & Dijkstra, H. A. (2018). Using network theory and machine learning to predict el niño. *Earth System Dynamics*, 9, 969–983. <https://doi.org/10.5194/esd-9-969-2018> (cit. on p. 35).
- Okumura, Y., & Xie, S.-P. (2004). Interaction of the atlantic equatorial cold tongue and the african monsoon. *Journal of Climate*, 17(18), 3589–3602. [https://doi.org/10.1175/1520-0442\(2004\)017<3589:IOTAEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3589:IOTAEC>2.0.CO;2) (cit. on p. 9).
- Otero, N., Mohino, E., & Gaetani, M. (2016). Decadal prediction of sahel rainfall using dynamics-based indices. *Climate Dynamics*, 47(9), 3415–3431. <https://doi.org/10.1007/s00382-015-2738-3> (cit. on pp. 33, 42).
- Paeth, H., & Hense, A. (2004). Sst versus climate change signals in west african rainfall: 20th-century variations and future projections. *Climatic Change*, 65(1-2), 179–208. <https://doi.org/10.1023/B:CLIM.0000037508.88115.8a> (cit. on p. 9).
- Pohlmann, H., Kröger, J., Greatbatch, R. J., & Müller, W. A. (2017). Initialization shock in decadal hindcasts due to errors in wind stress over the tropical pacific. *Proceedings of the 4th International Conference on Earth System Modelling (4ICESM)* (cit. on p. 39).
- Pohlmann, H., Smith, D. M., Balmaseda, M. A., Keenlyside, N. S., Masina, S., Matei, D., Muller, W. A., & Rogel, P. (2013). Predictability of the mid-latitude atlantic meridional overturning circulation in a multi-model system. *Climate Dynamics*, 41, 775–785. <https://doi.org/10.1007/s00382-013-1663-6> (cit. on pp. 8, 29, 42).
- Räisänen, J. (2011). How much should climate model output be smoothed in space? *Journal of Climate*, 24(3), 867–880. <https://doi.org/10.1175/2010JCLI3872.1> (cit. on p. 6).
- Robson, J. I., Sutton, R. T., & Smith, D. M. (2014). Decadal predictions of the cooling and freshening of the north atlantic in the 1960s and the role of ocean circulation. *Climate Dynamics*, 42, 2353–2365. <https://doi.org/10.1007/s00382-014-2115-7> (cit. on pp. 8, 29, 42).
- Saha, M., & Nanjundiah, R. S. (2020). Prediction of the enso and equinoo indices during june–september using a deep learning method. *Meteorological Applications*, 27(1), e1826. <https://doi.org/10.1002/met.1826> (cit. on p. 35).
- Sansom, P. G., Ferro, C. A. T., Stephenson, D. B., Goddard, L., & Mason, S. J. (2016). Best practices for postprocessing ensemble climate forecasts. part i: Selecting appropriate recalibration methods. *Journal of Climate*, 29(20), 7247–7264. <https://doi.org/10.1175/JCLI-D-15-0868.1> (cit. on p. 7).

- Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., ... Williams, A. (2014). Skillful long-range prediction of european and north american winters. *Geophysical Research Letters*, 41(7), 2514–2519. <https://doi.org/10.1002/2014GL059637> (cit. on pp. 7, 8).
- Scaife, A. A., Kucharski, F., Folland, C. K., Kinter, J., Brönnimann, S., Fereday, D., Fischer, A. M., Grainger, S., Jin, E. K., Kang, I. S., Knight, J. R., Kusunoki, S., Lau, N. C., Nath, M. J., Nakaegawa, T., Pegion, P., Schubert, S., Sporyshev, P., Syktus, J., ... Zhou, T. (2009). The clivar c20c project: Selected twentieth century climate events. *Climate Dynamics*, 33(5), 603–614. <https://doi.org/10.1007/s00382-008-0451-1> (cit. on p. 27).
- Scaife, A. A., & Smith, D. M. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, 1, 28. <https://doi.org/10.1038/s41612-018-0038-4> (cit. on pp. 6–8, 15).
- Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., ... Allen, M. R. (2016). Human influence on climate in the 2014 southern england winter floods and their impacts. *Nature Climate Change*, 6, 627–634. <https://doi.org/10.1038/nclimate2927> (cit. on p. 7).
- Sheen, K. L., Smith, D. M., Dunstone, N. J., Eade, R., Rowell, D. P., & Vellinga, M. (2017). Skilful prediction of sahel summer rainfall on inter-annual and multi-year timescales. *Nature Communications*, 8, 14966. <https://doi.org/10.1038/ncomms14966> (cit. on pp. 33, 40, 42, 43).
- Siegert, S., Stephenson, D. B., Sansom, P. G., Scaife, A. A., Eade, R., & Arribas, A. (2016). A bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is nao predictability? *Journal of Climate*, 29(3), 995–1012. <https://doi.org/10.1175/JCLI-D-15-0196.1> (cit. on p. 15).
- Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2023). Hybrid forecasting: Blending climate predictions with ai models. *Hydrology and Earth System Sciences*, 27(9), 1865–1889. <https://doi.org/10.5194/hess-27-1865-2023> (cit. on p. 10).
- Slater, L. J., Villarini, G., Bradley, A. A., & Vecchi, G. A. (2019). A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. *Climate Dynamics*, 53, 7429–7445. <https://doi.org/10.1007/s00382-017-3794-7> (cit. on p. 10).
- Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., & Yang, X. (2019). Robust skill of decadal climate predictions. *npj*

*Climate and Atmospheric Science*, 2. <https://doi.org/10.1038/s41612-019-0071-y> (cit. on pp. 3–5, 27, 29, 42).

- Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H., & Scaife, A. A. (2010). Skilful multi-year predictions of atlantic hurricane frequency. *Nature Geoscience*, 3, 846–849. <https://doi.org/10.1038/NGEO1004> (cit. on pp. 8, 29).
- Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., ... Zhang, L. (2020). North atlantic climate far more predictable than models imply. *Nature*, 583, 796–800. <https://doi.org/10.1038/s41586-020-2525-0> (cit. on pp. 8, 11, 13, 27–29, 42).
- Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, 41(11–12), 3325–3338. <https://doi.org/10.1007/s00382-013-1683-2> (cit. on p. 6).
- Smith, D. M., Murphy, J. M., Sexton, D. M. H., Jenkins, G. J., Boorman, P. M., Booth, B. B. B., Brown, S. J., Clark, R. T., Collins, M., Harris, G. R., Kendon, E. J., McCarthy, M. P., McDonald, R., Stephens, E., & Wallace, E. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, 317(5839), 796–799. <https://doi.org/10.1126/science.1139540> (cit. on p. 6).
- Smith, D. M., Scaife, A. A., & Kirtman, B. P. (2012). What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environmental Research Letters*, 7(1), 015602. <https://doi.org/10.1088/1748-9326/7/1/015602> (cit. on p. 5).
- Stott, P. A., Stone, D. A., Allen, M. R., Frame, D. J., Min, S.-K., Otto, F. E. L., van Oldenborgh, G. J., & Vautard, R. (2016). Attribution of extreme weather and climate-related events. *WIREs Climate Change*, 7(1), 23–41. <https://doi.org/10.1002/wcc.380> (cit. on p. 7).
- Sun, M., Chen, L., Li, T., & Luo, J.-J. (2023). Cnn-based enso forecasts with a focus on ssta zonal pattern and physical interpretation. *Geophysical Research Letters*, 50(20), e2023GL105175. <https://doi.org/10.1029/2023GL105175> (cit. on pp. 11, 21, 22, 35, 38).
- Tang, Y., Zhang, R.-H., Liu, T., Duan, W., Yang, D., Zheng, F., Ren, H., Lian, T., Gao, C., & Chen, D. (2018). Progress in enso prediction and predictability study. *National Science Review*, 5(6), 826–839. <https://doi.org/10.1093/nsr/nwy105> (cit. on p. 35).
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719> (cit. on p. 25).
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1> (cit. on p. 13).

- Teng, H., Branstator, G., Karspeck, A., Yeager, S., & Meehl, G. A. (2017). Initialization shock in ccm4 decadal prediction experiments. *CLIVAR Exchanges*, 22(2), 6–9 (cit. on p. 39).
- Tippett, M. K., Barnston, A. G., & Li, S. (2012). Performance of recent multimodel enso forecasts. *Journal of Applied Meteorology and Climatology*, 51(3), 637–654. <https://doi.org/10.1175/JAMC-D-11-093.1> (cit. on p. 35).
- Tippett, M. K., Ranganathan, M., L'Heureux, M., Barnston, A. G., & DelSole, T. (2019). Assessing probabilistic predictions of enso phase and intensity from the north american multimodel ensemble. *Climate Dynamics*, 53(12), 7497–7518. <https://doi.org/10.1007/s00382-017-3721-y> (cit. on p. 35).
- van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., & Hazeleger, W. (2012). Decadal prediction skill in a multi-model ensemble. *Climate Dynamics*, 38(7), 1263–1280. <https://doi.org/10.1007/s00382-012-1313-4> (cit. on pp. 39, 40).
- Wang, S., Mu, L., & Liu, D. (2021). A hybrid approach for el niño prediction based on empirical mode decomposition and convolutional lstm encoder-decoder. *Computers Geosciences*, 149, 104695. <https://doi.org/10.1016/j.cageo.2021.104695> (cit. on p. 35).
- Yan, J., Mu, L., Wang, L., Ranjan, R., & Zomaya, A. Y. (2020). Temporal convolutional networks for the advance prediction of enso. *Scientific Reports*, 10. <https://doi.org/10.1038/s41598-020-65070-5> (cit. on p. 35).
- Yeager, S. G., Danabasoglu, G., Bates, S. C., Phillips, A. S., Arblaster, J. M., Edwards, J. M., Holland, M. M., Kushner, P. J., Lamarque, J. F., Lawrence, D. M., Lindsay, K., Middleton, A., Muñoz, E., Neale, R., Oleson, K. W., Polvani, L. M., & Vertenstein, M. (2018). Predicting near-term changes in the earth system: A large ensemble of initialized decadal simulations. *Bulletin of the American Meteorological Society*, 99(9), 1867–1886. <https://doi.org/10.1175/BAMS-D-17-0098.1> (cit. on pp. 11, 33, 38–40, 42, 43).
- Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012a). A decadal prediction case study: Late 20th century north atlantic ocean heat content. *Journal of Climate*, 25, 5173–5189. <https://doi.org/10.1175/JCLI-D-11-00595.1> (cit. on p. 4).
- Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012b). A decadal prediction case study: Late 20th century north atlantic ocean heat content. *Journal of Climate*, 25, 5173–5189. <https://doi.org/10.1175/JCLI-D-11-00595.1> (cit. on pp. 8, 29, 42).
- Zhang, R., & Delworth, T. L. (2006). Impact of atlantic multidecadal oscillations on india/sahel rainfall and atlantic hurricanes. *Geophysical Research Letters*, 33, L17712. <https://doi.org/10.1029/2006GL026267> (cit. on pp. 8, 29, 42).
- Zhang, W., Kirtman, B., Siqueira, L., Clement, A., & Xia, J. (2021). Understanding the signal-to-noise paradox in decadal climate predictability from cmip5 and an eddying global coupled model. *Climate Dynamics*, 56, 2895–2913. <https://doi.org/10.1007/s00382-020-05621-8> (cit. on p. 7).

- Zhang, W., Mao, W., Jiang, F., Stuecker, M. F., Jin, F.-F., & Qi, L. (2021). Tropical indo-pacific compounding thermal conditions drive the 2019 australian extreme drought. *Geophysical Research Letters*, 48(2), e2020GL090323. <https://doi.org/10.1029/2020GL090323> (cit. on p. 35).
- Zhang, W., Wang, Y., Jin, F.-F., Stuecker, M. F., & Turner, A. G. (2015). Impact of different el niño types on the el niño/iod relationship. *Geophysical Research Letters*, 42(20), 8570–8576. <https://doi.org/10.1002/2015GL065703> (cit. on p. 35).
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929. <https://doi.org/10.1109/CVPR.2016.319> (cit. on p. 26).
- Zhou, L., Gao, C., & Zhang, R.-H. (2023). A spatiotemporal 3d convolutional neural network model for enso predictions: A test case for the 2020/21 la niña conditions. *Science Bulletin*, 68(5), 539–542. <https://doi.org/10.1016/j.scib.2023.01.004> (cit. on p. 35).