

Why does the signal-to-noise paradox exist in seasonal climate predictability?

Shivamurthy Yashas^{1,2}, Subodh Kumar Saha¹, Samir Pokhrel¹, Mahen Konwar¹, and Verma Utkarsh^{1,2}

¹Indian Institute of Tropical Meteorology, Pune, India

²Savitribai Phule Pune University, Pune, India

Correspondence: Subodh Kumar Saha (subodh@tropmet.res.in)

Abstract. Estimates of the potential predictability limit (PPL) for seasonal climate, typically based on a perfect model framework, sometimes encounter challenges of being paradoxical, as actual skill surpasses the PPL. The signal-to-noise paradox (SNP) gets its name from the use of model-based signal-to-noise ratios to estimate the PPL. Here, we study seasonal climate predictability in the tropical and subtropical regions during the boreal summer (June to September), with a focus on the SNP. We estimate PPL within the perfect model framework, only considering error growth from initial conditions. Signal and noise components display temporal non-orthogonality and a weak association between estimates of PPL and actual prediction skill, contradicting its intended purpose. A significant correlation between signal and noise violates the perfect-model framework. Moreover, paradoxical regions show no clear correspondence with signal–noise correlation, indicating that while accurate signal–noise separation is necessary, it is not sufficient to eliminate paradoxes.

We have also demonstrated that sub-seasonal components, which are building blocks of seasonal mean, substantially contribute to seasonal anomalies in association with major global predictors. The co-variability between sub-seasonal components and seasonal anomalies is wide-ranging and often skewed compared to observations, thereby influencing seasonal prediction skills and PPL. Therefore, a robust PPL estimation should consider errors from initial conditions and model-related factors such as physics, dynamics, and numerical methods. In this context, we present a simple diagnostic approach to estimate the maximum achievable seasonal prediction skill, which may be interpreted as an empirical upper bound of skill or the PPL.

1 Introduction

Seasonal climate prediction plays a pivotal role in making policy decisions, facilitating long-term planning, and implementing mitigation strategies, ultimately contributing to the development of a climate-resilient society. The emergence of the first computer in the 1950s initially fostered optimism for achieving precise weather and climate prediction. However, studies by Lorenz (1963, 1969) pointed out the inherent unpredictability of weather beyond a limited temporal horizon of only a few days. Consequently, this revelation led to the perception that the endeavour of seasonal prediction was considered unattainable. Nevertheless, a significant advancement in our understanding of long-lead predictability in the tropical climate was achieved by Charney and Shukla (1981), with further elaboration by Shukla (1998). In this context, the predictability of seasonal climate is attributed to the influence of slowly evolving boundary conditions, such as sea surface temperature (SST) and soil mois-

ture, which retain memory and significantly influence atmospheric instabilities on longer timescales. This breakthrough has established the scientific foundation for seasonal climate prediction.

While significant advancements have been made in improving the accuracy of weather forecasts over the past few decades (e.g., Bauer et al., 2015), the task of providing reliable seasonal predictions remains a formidable challenge. The Indian summer monsoon serves as an illustrative example, as attempts to predict it date back a century (Walker, 1924; Mooley and Parthasarathy, 1984), yet achieving consistently accurate seasonal forecast across all years remains challenging (Shukla, 2007; Jain et al., 2019). In addition to evaluating current prediction skills, it is of equal importance to ascertain the inherent constraints of seasonal climate predictability, denoted as the Potential Predictability Limit (PPL). It is commonly assumed that the PPL represents an upper limit on forecast skill, beyond which further enhancements become unattainable and, that a dynamical prediction system cannot surpass this limit (Kumar et al., 2005; Rajeevan et al., 2012). However, several studies have documented instances where a model's seasonal prediction skill surpasses the estimated PPL (Kumar et al., 2014; Scaife et al., 2014; Saha et al., 2016a; Scaife and Smith, 2018, e.g.,), giving rise to a signal-to-noise paradox (SNP), a situation where the proportion of predictable variance in models is too weak. This translates into actual skill (correlation skill between ensemble mean and observation) being greater than the estimated PPL.

In real-world forecasting, predictability is limited by two primary sources of error: uncertainties in the initial conditions, and model errors (arising from imperfect physics, dynamics, and numerical schemes), which grow nonlinearly with time. The perfect-model framework, however, neglects model errors by construction and considers only initial-condition uncertainty, thereby providing an idealized upper bound on predictability (i.e. PPL). Within this framework, inter-ensemble spread is assumed to arise solely from initial-condition uncertainty and is treated as the noise component, while deviations of the ensemble mean from the climatological mean are interpreted as the signal component. In practice, however, model's error can contribute to both signal and noise, which potentially could giving rise to SNP, where the realized forecast skill becomes greater than the estimated PPL.

In previous studies, attempts to estimate the potential predictability of seasonal climate were predominantly focused on shorter hindcasts and employed various estimation methods (Yang et al., 2012; Saha et al., 2016b, a, 2019; Scaife et al., 2014). However, recent studies have highlighted the pervasive challenge of the SNP, which extends across different climate models and temporal scales (Scaife and Smith, 2018; Strommen and Palmer, 2019; Zhang and Kirtman, 2019; Zhang et al., 2021; Sévellec and Drijfhout, 2019). Intriguingly, about 70 percent of the globe exhibits SNP in annual surface air temperature simulated by CMIP5 models (Sévellec and Drijfhout, 2019). This paradox is not solely attributed to issues with initialization processes (Zhang and Kirtman, 2019; Cottrell et al., 2024). Addressing the low signal-to-noise ratio present in models has been recognized as a potential solution to improve model future projections (Smith et al., 2020). Model's response to external radiative forcing and limited ensemble size may also be responsible for SNP (Klavans et al., 2021). Nonetheless, another study utilizing the Lorenz 1963 model (Lorenz, 1963) argues that the paradox can arise from the initialization process, wherein the initial ensemble spread (standard deviation) surpasses the observational spread (Mayer et al., 2021). Large ensemble sizes and the application of post-processing techniques also hold promise for enhancing prediction accuracy by reducing noise and enhancing forecast variance (Eade et al., 2014; Smith et al., 2019). Hu et al. (2021) have shown that in NCEP CFSv2, the

spatial pattern of the primary mode of signal and noise components for SST coincides over the tropical Pacific, yet their temporal variability remains uncorrelated. Moreover, the non-stationary nature of the climate system and smaller sample size (i.e. ensemble member and number of years) can confound the detection of the paradox (Weisheimer et al., 2018).

This study aims to explore the SNP in the context of seasonal prediction skill and sub-seasonal variability within tropical and subtropical regions, with a specific emphasis on South Asia and the central Pacific from June to September (JJAS). We estimate the PPL using traditional methods and assess correlation skill of rainfall, surface temperature, sea surface temperature, and mean sea level pressure (MSLP). Subsequently, we apply these methods to identify regions where the paradox exists. While the estimation methods of PPL are rooted in the idealized notion of a perfect model framework, where the model itself is considered flawless, but errors in the initial conditions result in divergence (i.e., simulations deviate with minor perturbations in initial conditions), it is essential to acknowledge that dynamical models are inherently imperfect, and the realization of a perfect model is improbable. Furthermore, estimates of PPLs are often very close to actual forecast skill and are expected to increase with advancements in the model. This raises critical questions regarding the overall usefulness of PPL. In this context, we posit that the signal and noise components in seasonal mean are not solely attributed to errors in the initial conditions but also to inadequate representation of physical processes and their associated non-linear feedback mechanisms within a dynamical model. We demonstrate a substantial contribution of sub-seasonal components to the overall variability and predictability of seasonal climate. Additionally, we identify a systematic error in the co-variability between sub-seasonal components and seasonal anomalies, which affects the accuracy of seasonal climate predictions and predictability.

Here, we introduce a diagnostic method to estimate the maximum attainable seasonal prediction skill within the available ensemble members. To account for the constraints imposed by a finite sample size (52 ensemble members over 41 years), we employ a bootstrap resampling technique to infer the true population statistics. The resulting correlation estimate may be interpreted as the upper limit of achievable prediction skill, or the PPL. Detailed data and methods are provided in section 2, while section 3 presents the study's results, and section 4 summarizes the findings.

2 Data and Methods

2.1 Model and Experiments

The modified version of the Climate Forecast System version 2 (CFSv2; Saha et al., 2014a, b; Hazra et al., 2017) of the National Centers for Environmental Prediction (NCEP) is used in this study. The CFSv2 consists of a spectral atmospheric model (i.e. Global Forecast System) at a horizontal resolution of about 1° (i.e. T126) with 64 hybrid vertical levels and the GFDL Modular Ocean Model version 4p0d (Griffies et al., 2004) with 40 vertical layers and $0.25 - 0.50^\circ$ horizontal grid spacing. The CFSv2 is coupled with a two-layer sea ice model (Wu et al., 1997; Winton, 2000) and land surface model Noah with four layers of soil and single layer snow scheme (Ek et al., 2003).

As CFSv2 shows a maximum (minimum) skill in Indian summer monsoon rainfall (ISMR) with February and April (March and May) initial conditions (e.g., Saha et al., 2016a; Pokhrel et al., 2016), based on NCEP's Climate Forecast System Reanalysis (CFSR; Saha et al., 2010), February initial conditions are used to generate ensemble re-forecasts/hindcasts for the years

1981–2021. The model is initialized every day (15 February to 27 February) and four cycles in a day (00, 06, 12, and 18 GMT).
 95 Therefore, 52 ensemble member simulations are performed, each year spanning 9-months.

2.2 Observed Data

In this study, we employ multiple observed/reanalysis data sets for analysis and comparisons with model simulations. Over land, 2m air temperature data from the Climatic Research Unit (CRU TS3.1; Harris et al. (2014)) is used. SST data is taken from EN4 reanalysis (Good et al., 2013). MSLP data is obtained from ERA5 reanalysis (Hersbach et al., 2020). Daily rainfall
 100 data from the Global Precipitation Climatology Project (GPCP Version 1.3, $1^\circ \times 1^\circ$; Huffman et al., 2001) for 1997-2021 and monthly rainfall data from GPCP version 2.3 (Adler et al., 2020) for 1979-2021 with $2.5^\circ \times 2.5^\circ$ horizontal resolution are employed.

2.3 Methodology

Here, we employ the Analysis of Variance (ANOVA) method based on the perfect model framework (e.g., Rowell et al., 1995;
 105 Rowell, 1998; Schneider and Griffies, 1999) to estimate PPL. The PPL serves as a measure of the model’s ability to predict the Earth’s weather and climate with the utmost skill, where prediction skill should always be lower than the PPL due to unavoidable errors in the initial conditions. However, it is found that the prediction skill of a model is higher than the PPL on several occasions and across the globe, which is termed as SNP. Furthermore, to verify the existence of SNP, the Ratio of Predictable Components (RPC) is used (Eade et al., 2014).

110 2.3.1 ANalysis Of VAriance (ANOVA) method

In this method, the total variance is split into signal (σ_{sig}^2) and noise (σ_{noise}^2) components. The variance decomposition framework of the analysis of variance (ANOVA) methodology introduced by Rowell et al. (1995, p. 699), which is used in several previous studies (e.g., Kang & Shukla, 2006; Scaife et al., 2014; Saha et al., 2016a; Scaife & Smith, 2018; Weisheimer et al., 2018), is employed here. The ratio of signal to noise variance is known as the signal-to-noise ratio (SNR). If x is the precipita-
 115 tion field of the model, i is the year of the model integration (total year ‘N’), and j is the number of ensemble simulations (total ensemble $n = 52$), then noise variance following Rowell et al. (1995), can be expressed as

$$\sigma_{noise}^2 = \frac{1}{N(n-1)} \sum_{j=1}^n \sum_{i=1}^N (x_{ij} - \bar{x}_i)^2 \quad (1)$$

where $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$ is the ensemble mean of the model for a year and the degrees of freedom is $N(n-1)$. The variance of ensemble mean (σ_{EV}^2) can be estimated as

$$120 \quad \sigma_{EV}^2 = \frac{1}{(N-1)} \sum_{i=1}^N (\bar{x}_i - \bar{x})^2 \quad (2)$$

where $\bar{x} = \frac{1}{Nn} \sum_{j=1}^n \sum_{i=1}^N x_{ij}$ is the average over all year and all ensemble. However, the variance of the ensemble mean is often an overestimation of signal variance (Scheffe, 1959). As the number of ensemble members is not very large (here 52), the ensemble mean contains residual noise component of the variability. Therefore, the signal variance may be estimated following Scheffe (1959) as

$$125 \quad \sigma_{sig}^2 = \sigma_{EV}^2 - \frac{1}{n} \sigma_{noise}^2 \quad (3)$$

Therefore, the total variance can be estimated as

$$\sigma_{TV}^2 = \sigma_{sig}^2 + \sigma_{noise}^2 \quad (4)$$

The signal variance arises due to slowly varying quasi-periodic or aperiodic boundary conditions, such as El-Niño Southern Oscillation (ENSO), which evolve on a longer time scale than the predictand (e.g. seasonal monsoon rainfall). In general, the power associated with quasi-periodic slowly evolving processes is higher than that of faster-evolving processes (Peixoto and Oort, 1992, , Figure 2.7). Hence, the slowly evolving process may dictate the evolution of the faster processes, contributing to predictability. The ratio of signal to noise variance (a quantitative measure of predictability) is known as the signal-to-noise ratio (or SNR). A perfect coupled model does not attest to a perfect seasonal rainfall forecast due to unavoidable errors in the initial conditions. Therefore, there will always be an upper limit to the seasonal predictability (e.g., Kang and Shukla, 2006; Westra and Sharma, 2010), which can be expressed in terms of SNR as

$$135 \quad PPL_{ANOVA} = \sqrt{\frac{SNR}{SNR+1}} \quad (5)$$

The values of PPL vary between 0 and 1 and show the maximum limit of correlation skill achievable by a model at a given lead time.

2.3.2 Ratio of Predictable Components (RPC)

140 RPC method (Eade et al., 2014) is used to identify the signal-to-noise paradox. The RPC is the ratio of predictable components in observation to that in the model. The predictable component in observation (PC_{obs}) is estimated directly from the fraction of the variance that can be explained by model forecasts (i.e. correlation). The predictable component in the model (PC_{model}) is estimated by the ratio of the variance of the ensemble mean to the variance of individual ensemble members. Ideally, RPC should be 1, as the observation and model should contain the same proportion of predictable variance, and the squared correlation should match the predictable proportion of variance in the model. Regions where the RPC index is greater than 1, show that the predictable component in the model is less than that we see in the real world (i.e. paradox). The RPC is given by

$$145 \quad RPC = \frac{PC_{obs}}{PC_{mod}} \geq \frac{r}{\sqrt{\sigma_{noise}^2 / \sigma_{TV}^2}} \quad (6)$$

2.3.3 Projection of sub-seasonal variance on seasonal anomaly

Unlike variables such as temperature or surface pressure, which vary smoothly in time, rainfall is inherently discrete and typically occurs in pulses (rain or no-rain) concentrated within preferred time bands (i.e., sub-seasonal bands). Furthermore, the amplitude of these events in the tropical monsoon region is often much larger than that of the annual cycle (Figure S1). As a result, variations in sub-seasonal rainfall can significantly modify the annual cycle or seasonal anomaly. This can be easily demonstrated by removing just one or two days of rainfall from daily time series over core monsoon region. Upon removal of two-day rainfall event (>80 mm/day; red bar in Figure S1), assuming it arises from sub-seasonal variability, the reconstructed annual cycle, becomes visibly weaker, with decrease in seasonal anomaly by 59% of its interannual standard deviation. It may be noted that only two single-day rainfall events are removed here; if a complete event is removed, as happens, the impact on the seasonal anomaly would be substantially larger. This highlights why sub-seasonal components are often termed the “building blocks” of the monsoon. The seasonal mean precipitation is simply the sum of precipitation during these events. Moreover, various sub-seasonal bands contributed differently to the seasonal anomaly (Figure S2). As the sum of these sub-seasonal events forms the seasonal mean, year-to-year variability in the characteristics of sub-seasonal components also contributes to seasonal anomalies. In other words, global predictors affect seasonal precipitation by modulating sub-seasonal components, such as changing the intensity and duration of these events (e.g. Saha et al., 2019, 2020, 2021).

Due to imperfections in model physics, accurately simulating sub-seasonal components remains a challenge. Therefore, their anomalous contributions to seasonal anomalies are problematic. For example, in the Indian summer monsoon, synoptic systems (i.e., lows and depressions) contribute about 45–55% of the seasonal mean precipitation (e.g. Yoon and Chen, 2005), and account for the maximum year-to-year variability (Saha et al., 2020). Therefore, errors in the co-variability between seasonal mean and sub-seasonal variance can also contribute to ensemble spread. To demonstrate this, we analyzed the seasonal variance of prominent sub-seasonal bands in the rainfall time series in relation to the seasonal anomaly. The seasonal variance of these sub-seasonal components serves as a measure of their vigour or strength in a season. It is important to note that errors in simulating precipitation arise from imperfections in various physical processes and their interactions. We chose precipitation to demonstrate the role of model physics in ensemble spread because it is a highly sought-after forecast parameter for society and, at the same time, has significant uncertainty. The time series of daily rainfall of a year (area average or a single point) can be represented by the following equation

$$x_T = x_c + x_a + \sum_f x_f \quad (7)$$

where, x_T is the total rain, x_c is the climatological mean annual cycle, x_a is the signal or anomalous annual cycle, x_f represents the rest sub-seasonal components consisting of all frequencies (f). Using harmonic analysis, the sum of the mean and the first three harmonics represents the ‘smooth annual cycle’ in the daily time series for a year. Here, x_c is the climatological mean of the ‘smooth annual cycle’, and x_a is the deviation of the ‘smooth annual cycle’ of a year from the climatological mean annual cycle. Therefore, after re-arrangement, the above equation can be written as

$$180 \quad (x_T - x_c) = x_a + \sum_f x_f \quad (8)$$

The left-hand term represents the total daily anomaly. In terms of seasonal variance, using daily June-to-September data (122 days) equation 8 for a particular season can be written as

$$\sum_{l=1}^{122} (x_T^l - x_c^l)^2 = \sum_{l=1}^{122} (x_a^l)^2 + \sum_{f=1}^K \sum_{l=1}^{122} (2x_a^l \cdot x_f^l) + \sum_{f=1}^K \sum_{l=1}^{122} (x_f^l)^2 \quad (9)$$

$$V_T = V_a + \sum_f (V_{cov}) + \sum_f (V_f) \quad (10)$$

185 where l represents the day, V_T is the total variance, V_a is the variance of the anomalous annual cycle, V_{cov} is the covariance among sub-seasonal and anomalous annual cycle, V_f represents the sub-seasonal variance, K is the number of sub-seasonal bands (e.g., synoptic, bi-weekly) in a season. However, due to orthogonality, the covariance term becomes negligible. In terms of seasonal anomaly, equation 10 can be written as

$$V_T' = V_a' + \sum_f V_f' \quad (11)$$

190 Where V_T' , V_a' and V_f' are seasonal anomalies of the total variance, variance of anomalous annual cycle, and sub-seasonal variance, respectively. Although the anomalous annual cycle (x_a) and sub-seasonal components (x_f) are orthogonal by construction, their seasonal variances (Equation 11) remain interlinked on interannual timescales (Figure S3), indicating a role for sub-seasonal variability in generating the seasonal anomaly. Let I' be the anomaly of seasonal mean rainfall then, the covariance between the seasonal rainfall anomaly and the anomaly of total variance can be written as

$$195 \quad \sum_i V_T' I' = \sum_i V_a' I' + \sum_i \sum_f V_f' I' \quad (12)$$

The left-hand term of equation 12 represents the interannual covariance between total sub-seasonal variance and seasonal anomaly. The first term on the right-hand side represents the covariance between the variance of the anomalous annual cycle and the seasonal mean, while the second term represents the covariance between the variance of the rest of the sub-seasonal components and the seasonal mean. It is important to note that the first term on the right-hand side explicitly does not contain
 200 new information on the building blocks of the seasonal mean, as both V_a' and I' are based on seasonal anomalies and are, therefore, not used in our analysis. On the other hand, the last term is of particular interest, as it represents the interannual covariance between seasonal rainfall anomaly and variance of sub-seasonal bands. Hence, the analysis of covariance within sub-seasonal bands may provide insights into their contributions and the causes of model bias. It may be noted that the variance

of sub-seasonal components, which reflects their energy or vigor, shows a strong correlation with the seasonal rainfall anomaly
205 in the observations (Figure S2).

Considering sub-seasonal components as building blocks of the seasonal rainfall, a global predictor (e.g., ENSO, AMO)
must modulate sub-seasonal variability to affect the seasonal rainfall (e.g. Saha et al., 2019; Borah et al., 2020). For a similar
reason, variability and predictability of the seasonal mean monsoon rainfall may be affected due to imperfect physics, dynamics
and numerical schemes through error in the simulation of sub-seasonal components. Fourier analysis is used to construct
210 a smooth annual cycle and total anomaly, while a Lanczos band-pass filter (Duchon, 1979) is applied to filter daily time
series data into synoptic (2-5 days), super-synoptic (10-20 days), and Monsoon Intra-Seasonal Oscillations/Madden-Julian
Oscillation(MISO/MJO) (20-60 days) bands.

3 Results

3.1 Regions of paradox

215 The assessment of PPL derived from ANOVA (equation 5) and the existence of SNP, which is also confirmed by RPC (equation
6), is the central focus of this study. This comprehensive analysis is conducted on seasonal (June-July-August-September)
averaged rainfall, SST, MSLP, and 2-metre air temperature (land region) data of 1981-2021, as illustrated in Figure 1. To find
out the potential presence of an SNP, we subtract correlation skill from the PPL estimated by ANOVA. Our analysis reveals
a conspicuous pattern across a substantial portion of the global tropical and sub-tropical regions, wherein the correlation
220 skill surpasses the estimated PPL (including the Indian region for rainfall and mean sea level pressure), thus manifesting a
paradoxical situation (stippled regions with black dots in Figure 1). Notably, these tropical regions exhibit a higher degree of
predictability, with the notable exception of rainfall, which consistently demonstrates the lowest level of predictability among
the considered parameters. The existence of SNP is further confirmed through RPC, represented by semi-transparent white
shading that predominantly overlaps the paradox regions identified by the ANOVA method. It is worth emphasizing that this
225 intriguing SNP in seasonal prediction has been recognized in several previous studies as well, including works by Scaife et al.
(2014); Kumar et al. (2014); Eade et al. (2014); Scaife and Smith (2018); Saha et al. (2019). However, the reasons for the
paradoxical behaviour in the seasonal forecast are not clear (e.g. Scaife and Smith, 2018).

Existing studies suggest that the SNP predominantly arises in the context of seasonal forecasting, which encompasses
timescales of a month and beyond, as opposed to weather or medium-range predictions. This distinction can be attributed
230 to the fundamental nature of weather forecasting, characterized as an initial value problem, where precision in the initial condi-
tions takes precedence. Conversely, when dealing with forecasts spanning from seasonal to decadal and even longer timescales,
slowly evolving boundary conditions (e.g., SST, soil moisture) and external forcings (e.g., solar variability, volcanic eruptions)
become more prominent. These factors are the sources of predictability of the second kind. As the forecast lead time increases,
the influence of initial errors wanes. In essence, in the realm of climate prediction, forecasts made several months or decades
235 ahead depend significantly on the accurate representation of external drivers (e.g., greenhouse gas concentrations, aerosol emis-
sions), the model's dynamics, its underlying physics, and the feedback mechanisms that operate within the model (i.e., internal

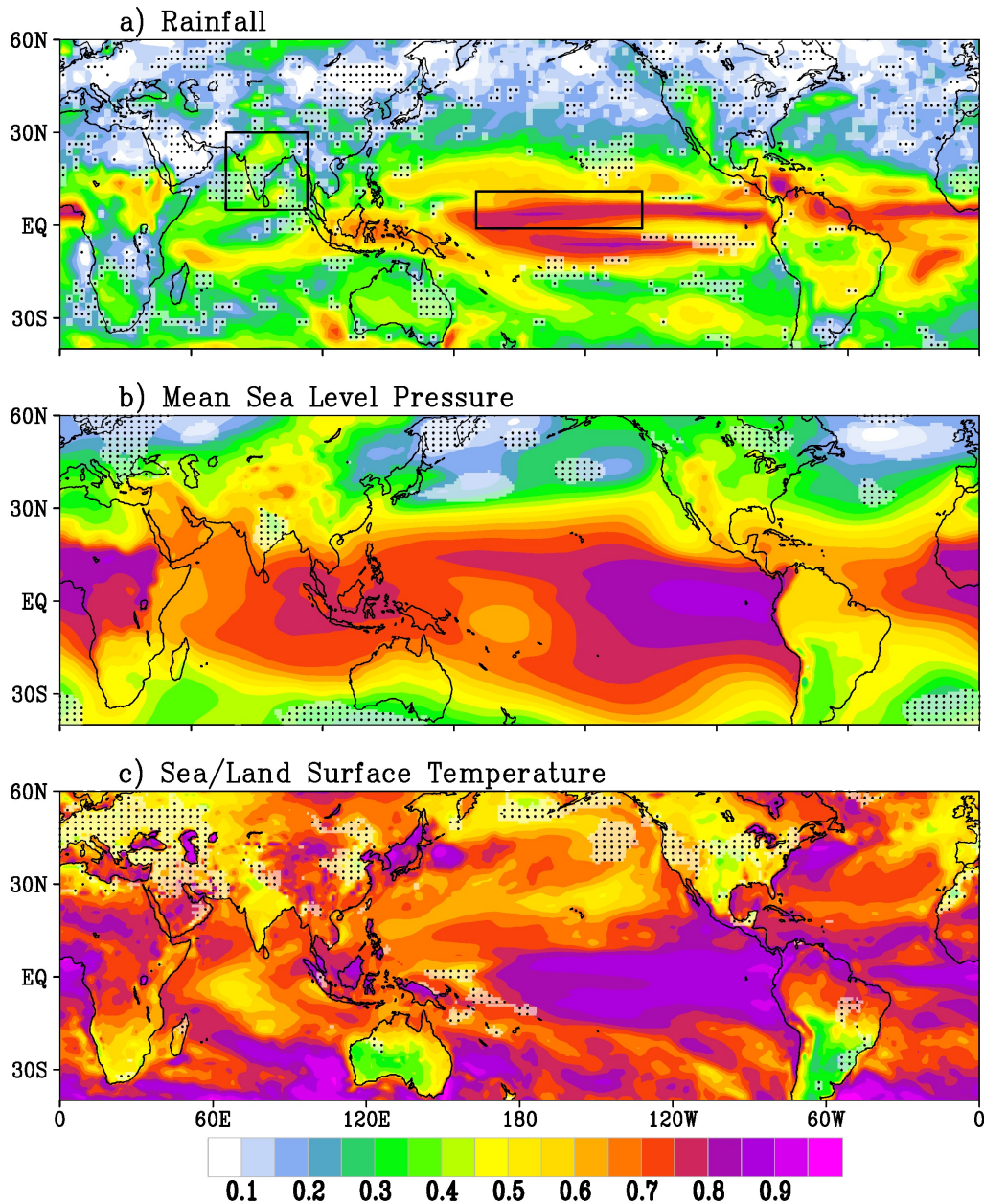


Figure 1. Potential predictability (shading) based on ANOVA method using equation no. 5. for JJAS averaged a) rainfall, b) mean Sea level pressure, c) Sea/land surface temperature using CFSv2 re-forecast of 41 years (1981-2021) and 52 ensemble members. Signal-to-noise paradox regions (where model correlation skill with observations is higher than the potential predictability) are stippled with black dots. White semi-transparent regions, mostly coinciding with stippled regions, represent $RPC > 1.0$.

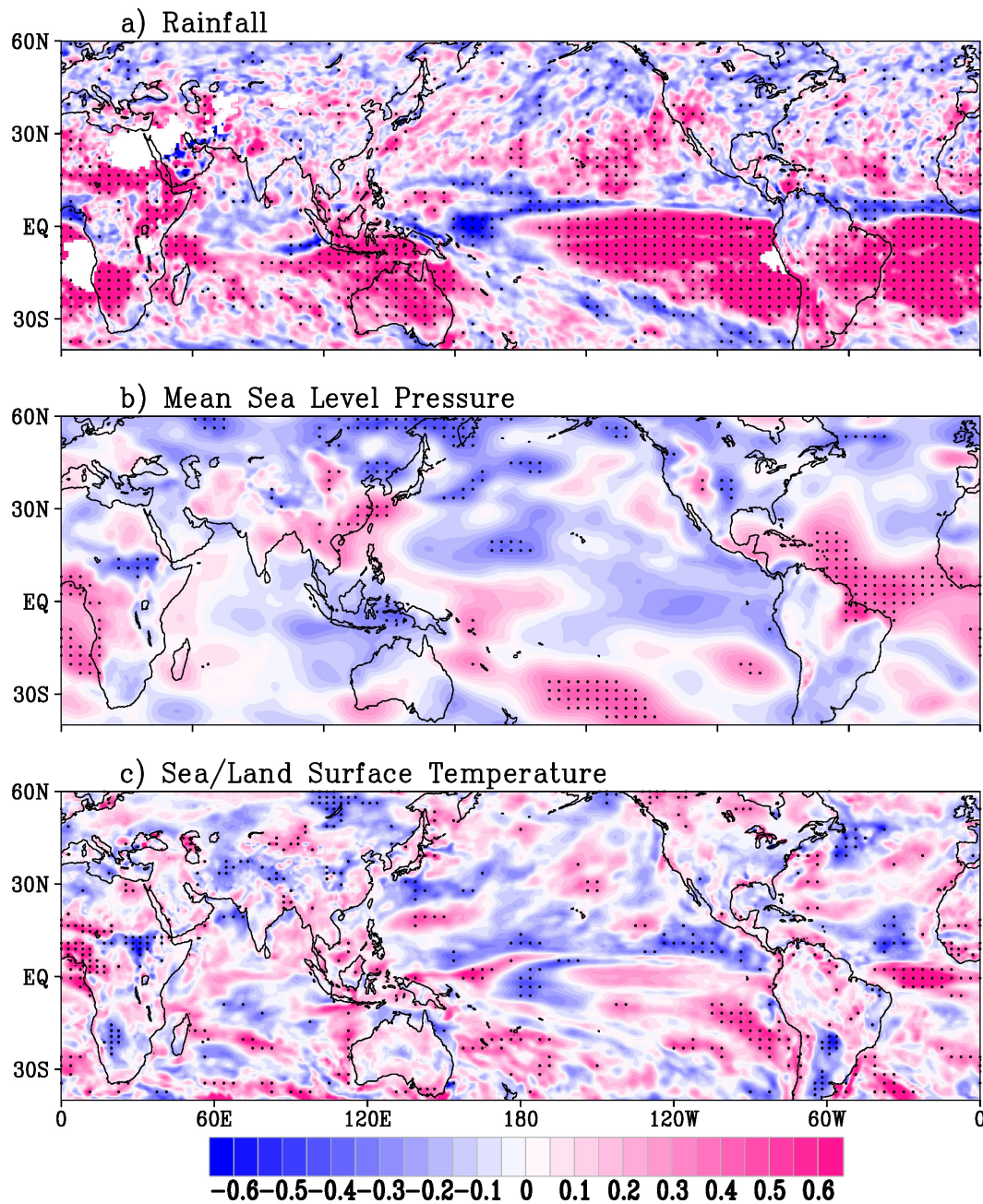


Figure 2. Test of orthogonality between signal and noise components. Correlations between signal and noise components of JJAS averaged a) rainfall, b) mean Sea level pressure, c) Sea/land surface temperature employing CFSv2 re-forecast of 41 years (1981-2021) and 52 ensemble members. Correlations significant at 95% (Student's two-tailed t-test) and above are stippled by black dots.

variability). It's worth noting that the methodology employed in this study, namely ANOVA, is based on the framework of the 'perfect model' paradigm. Here, the model is considered perfect or flawless, and any deviations in model simulations are attributed solely to errors in the initial conditions. This principle becomes particularly evident in large ensemble simulations, where even slight differences in initial conditions result in divergent model outcomes.

As estimates of PPL and RPC depend on the signal-to-noise ratio (section 2.3), the time series of the seasonal mean is partitioned into signal and noise components. Under the perfect-model assumption and within the ANOVA framework, the signal and noise components are expected to be clearly separated, that is, statistically independent of each other. Given that signal and noise inherently exhibit orthogonality, we examine the correlation between these components, aiming to ascertain the robustness of their partitioning. For rainfall, a substantial region in the tropics and sub-tropics exhibits statistically significant correlations between signal and noise components (Figure 2). Conversely, in the case of MSLP and SST/LST, the significant area is smaller but extends across tropical and sub-tropical regions. Weisheimer et al. (2018) hypothesized that paradox emerges due to statistical problems in estimating RPC. Bröcker et al. (2023) also shows that signal and noise are not orthogonal in time, further undermining the assumptions underlying standard RPC estimation. Importantly, it should be noted that the regions displaying the paradox do not necessarily align with those demonstrating high or significant correlations between signal and noise. The paradox also persists in alternative measures used to estimate the PPL that do not explicitly rely on signal and noise variability (Scaife and Smith, 2018). Hence, it is plausible that precise partitioning into signal and noise components alone may not be sufficient to eliminate paradoxical behaviour and, consequently, determine the true limit of potential predictability.

3.2 Signal-to-Noise Paradox in Relation to Model Skill

As the name suggests, PPL is the maximum achievable skill by a model. However, its estimate does not have any binding relationship with the observations (i.e. observations are not required to estimate PPL). Nevertheless, it is important to understand how strong is the association between actual skill and estimated PPL. ENSO, as the dominant mode of global climate variability and a primary driver of both the ISMR and tropical Pacific variability, provides a fundamental source of seasonal predictability in these regions. Therefore it is logical to investigate whether and to what extent the model's actual skill aligns with the estimated potential predictability by examining the predictability of the Niño3.4 index along with regional precipitation in both paradoxical and non-paradoxical areas. Here we use time series data involving a predictor (Niño3.4 SST) and rainfall within two delineated regions (as shown in Figure 1): one situated in the paradoxical region, denoted as the Indian Summer Monsoon Region (ISMR), and the other in a non-paradoxical region, referred to as the Pacific Region (PACR). These two contrasting regions are selected to investigate whether there are distinct variations in PPL with actual skill.

Furthermore, to test the association between PPL and actual skill, the ensemble members are arranged in ascending order in their absolute error in seasonal anomaly (with respect to GPCP data) for each year (i.e. ensemble members are arranged in a good-to-poor correlation skill order). We calculate predictability measures and correlations over a 21-ensemble member running window to see the general pattern. In the case of ISMR, although the PPL slightly decreases with diminishing skill, good ensemble members exhibit paradoxical behaviour while poor ensemble members do not (Figure 3). However, RPC varies with actual skill, as it is inherently a function of the correlation between model predictions and observations. For Niño3.4 and

PACR, PPL does not fluctuate with actual skill, and the actual skill consistently remains significantly lower than the PPL. This implies that PPL is not contingent on a model's predictive skill. In other words, a higher PPL does not necessarily indicate that a model has the potential to achieve greater skill, which contradicts the objective of estimating PPL. Ensemble members are arranged from best to worst (or good to poor) to clearly show the relationship between actual skill and PPL.

275 The estimated PPLs clearly do not represent the upper limit of seasonal prediction skill, raising the question of why such a discrepancy exists. A plausible explanation is that partitioning of seasonal mean into signal and noise components is problematic. It assumes, perhaps unrealistically, that the inter-ensemble spread in a year is only due to initial errors (see section 2.3). Subsequently, we compute correlations between signal and noise components using a 21-ensemble moving window of arranged ensembles (Figure 4). Remarkably, the results reveal a distinct pattern where correlations strengthen and become
280 statistically significant for ensemble members exhibiting greater bias in the seasonal anomaly. While the correlation between signal and noise becomes stronger (inverse correlation) as we move from good to poor ensembles (Figure 3), particularly for PACR, the PPL remains relatively constant (Figure 3c). We also note that for the ISMR, even though the paradox persists in the case of good ensemble members (as shown in Figure 3a), the correlation between the signal and noise components is relatively stronger for good ($r=0.2$) and poor ($r=-0.2$) ensemble members (Figure 4). Therefore, above all suggest that accurate
285 partitioning into signal and noise components is also not a sufficient criterion for obtaining the true estimates of PPL. This is consistent with the findings of Scaife and Smith (2018), who showed that paradoxes can arise even in PPL estimates that do not explicitly rely on the separation of signal and noise components.

Apart from PPL, it is also important to assess how skillful the model is in predicting both the predictors and the predictands. Given two contrasting regions for analysis of predictands (i.e. ISMR, PACR), we have evaluated the prediction skill for these
290 predictands and major global predictors (i.e. Niño3.4 index, Indian Ocean Dipole (IOD), Pacific Decadal Oscillation (PDO), and Atlantic Multi-decadal Oscillation (AMO)). It is noteworthy that a diverse range of correlation skills is evident among ensemble members (Figure 5). However, it is important to underscore that the skill exhibited by the ensemble mean (indicated by the red circle) consistently surpasses that of the average of the individual members (depicted by the blue circle). Furthermore, aside from the IOD, the correlation skill for all other predictors and both predictands remains statistically significant at or above
295 the 99% confidence level. This highlights the model's proficiency in forecasting both the predictors and predictands, which is essential for further analyzing their complex relationships with other components of the climate system, such as sub-seasonal components. However, the RMSE exceeds 100% of the observed standard deviation (Figure 5b), indicating that the model's errors are greater than the natural variability in the observed data.

3.3 Role of Sub-seasonal Components

300 Although the advent of computers in the 1950s gave hope for accurate weather prediction, Lorenz's work (Lorenz, 1963) revealed the inherent unpredictability beyond a few days due to non-linearity in the systems. Subsequently, Charney and Shukla (1981) pioneered the concept of long-lead predictability in the tropics (e.g. monthly/seasonal mean) due to the influence of slowly varying boundary conditions (e.g., SST, soil moisture) on atmospheric instabilities. Recent studies have shown that sub-seasonal components of the Indian summer monsoon, in association with global predictors (e.g., ENSO, AMO, etc.), sig-

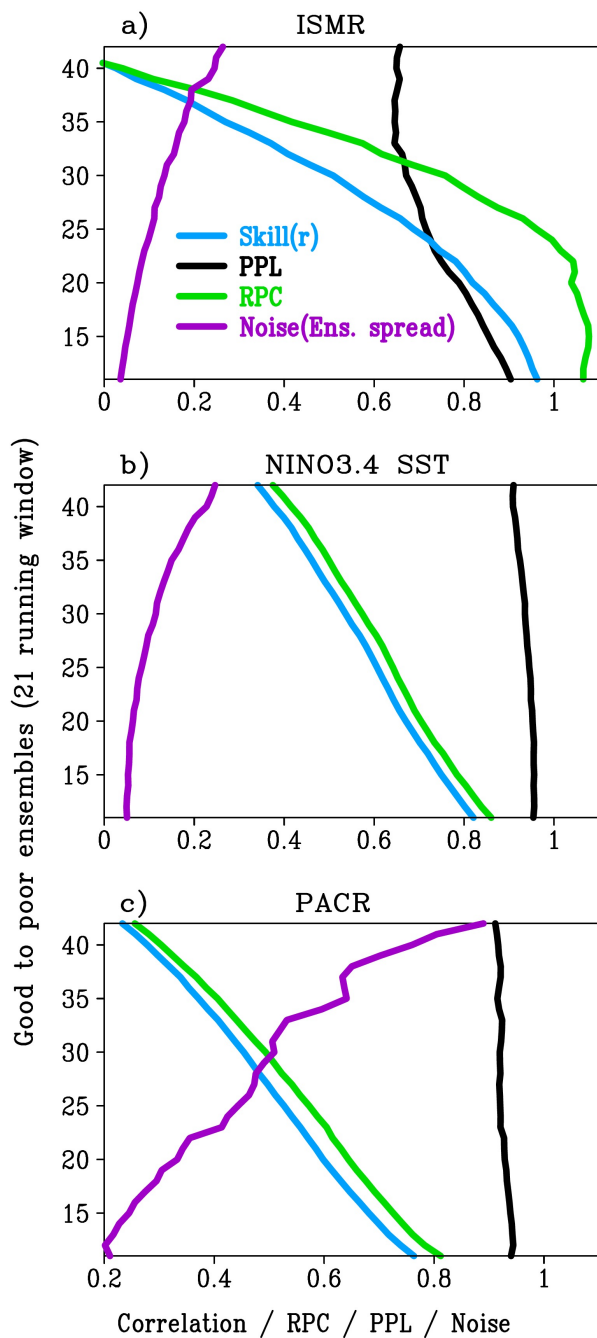


Figure 3. Correlation skill (blue), ANOVA based Potential predictability limit (PPL; black), Ratio of Predictable Component (RPC; green) and ensemble spread/‘noise’ component (purple) of arranged ensembles from good to poor and using a 21-ensemble moving window for a) ISMR rainfall, b) Niño3.4 SST, c) PACR rainfall. Correlation skill is defined as the correlation between the ensemble-mean re-forecast and observations. The Niño3.4 index is used as a reference predictor to provide a baseline for ENSO-driven predictability of regional precipitation, allowing comparison with the model’s full dynamical skill.

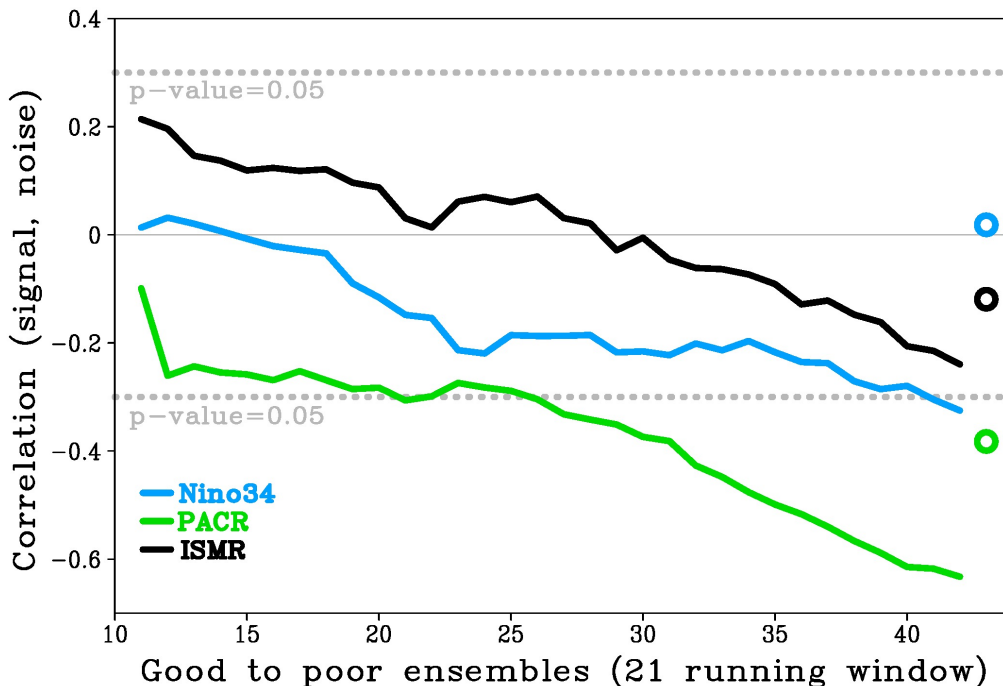


Figure 4. Correlations between signal and noise components estimated based on the ANOVA method. Correlations using all 52 ensemble members (open circles) and a 21-ensemble member moving window, where ensemble members are arranged from good to increasingly poor ensembles, are shown for ISMR (black), PACR (green), and Niño3.4 SST (blue). The grey dotted line indicates correlations that are significant at the 95% confidence level.

305 nificantly contribute to the inter-annual variability of the ISMR (Saha et al., 2019, 2020, 2021). The sub-seasonal components are the building blocks of the seasonal mean. Moreover, precipitation is a discrete phenomenon that comes as an event (either zero or a positive value). The annual/seasonal cycles of precipitation are composed of these discrete events. The sum of these events constitutes the seasonal, or annual mean. In principle, for any predictor affecting a predictand, the predictor communicates through the modulation of sub-seasonal components, which are then projected onto the seasonal/decadal anomaly of

310 the predictand. However, due to the inherent non-linear characteristics of sub-seasonal components, it is plausible that their variability may not be entirely predictable. In such cases, a portion of the sub-seasonal variability may remain associated with relevant predictors, contributing to overall seasonal predictability (illustrated through schematic in Figure 6). Given that sub-seasonal components constitute fundamental elements of seasonal rainfall, the seasonal prediction skill can be assessed through an examination of the co-variances between sub-seasonal components and seasonal rainfall/predictors (see section 2.3.3, equa-

315 tions 7-12). Moreover, the co-variance of sub-seasonal components with seasonal rainfall, in relation to PPL and ensemble spread, can likely be attributed to imperfect physics and can be demonstrated. To further elucidate this relationship, we have

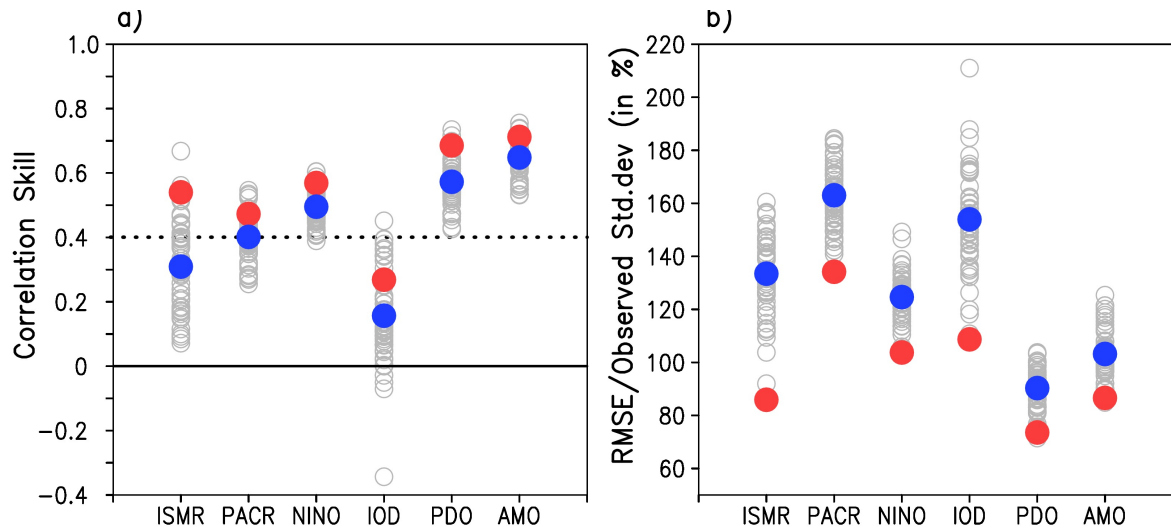


Figure 5. a) Prediction skill (correlation) of major predictors (NINO → Nino3.4 index; IOD → Indian Ocean dipole index; PDO → Pacific Decadal Oscillation; AMO → Atlantic Multi-decadal Oscillation) and predictands (ISMR → Indian Summer Monsoon Rainfall averaged over the box shown in Figure 1a; PACR → area-averaged rainfall over the Pacific box (Figure 1a)) in CFSv2 using 41 years of re-forecast (1981-2021) and observations. b) RMSE compared to observed standard deviation (i.e. RMSE/Obs. Std.). Grey open (red solid) circles represent individual (averaged) ensemble member skills. The blue solid circle represents the average skill of individual members (i.e., the average of the grey circles). The black dashed line represents the significance level of correlation values at the 99% confidence level.



Figure 6. A schematic diagram illustrates how slowly varying predictors influence seasonal rainfall anomaly by modulating sub-seasonal components. The sub-seasonal components are considered the building blocks of seasonal rainfall. Highlighting the dominant role of sub-seasonal variability in modulating seasonal monsoon precipitation (specific to tropical/monsoon regions).

organized ensemble members in ascending order from more accurate to less accurate, i.e., from good to poor ensembles, based on seasonal rainfall anomaly (ISMR and PACR) compared to observations (i.e. GPCP rainfall).

The sub-seasonal variances are computed within three distinct bands: 2-5 days (synoptic), 10-20 days (bi-weekly/super-synoptic), and 20-60 days (MISO, MJO). If sub-seasonal components indeed play a role in the prediction skill of seasonal rainfall, their co-variability should exhibit distinguishable patterns among good and poor ensembles. As global daily gridded

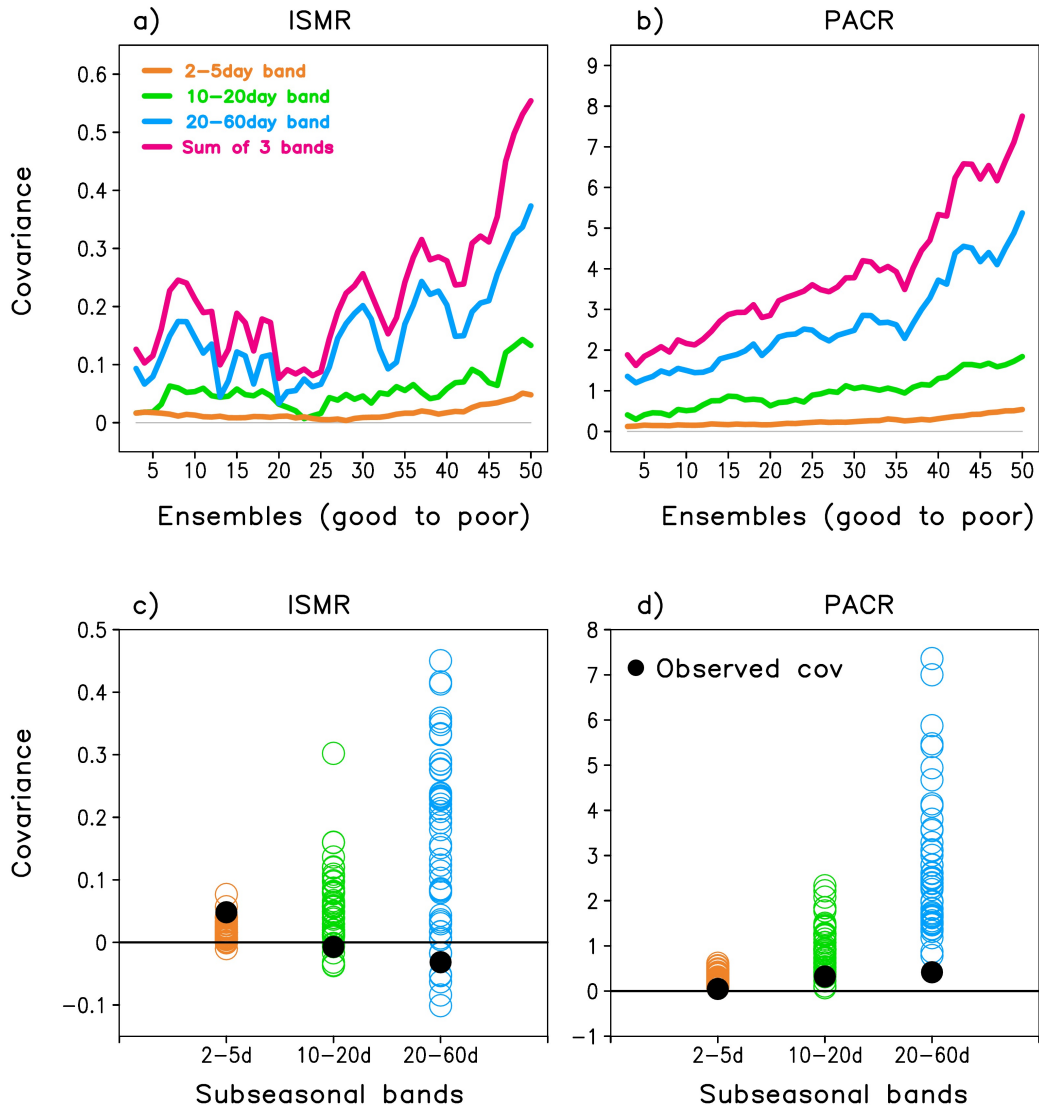


Figure 7. Inter-annual co-variance between sub-seasonal variances in three bands (2-5 days, 10-20 days, 20-60 days) and mean rainfall in the ISM and PAC domains (1997-2021). Line plots of co-variances arranged from good-to-poor ensemble members (compared to observed seasonal anomalies) in a 5-ensemble moving average window for a) the Indian summer monsoon domain and b) the Pacific domain. Co-variances in each ensemble member at three sub-seasonal bands, along with observations for c) ISMR and d) PACR.

rainfall observation data (GPCP) is available from October 1997 onward, our analysis is conducted for the period spanning from 1997 to 2021. It is worth noting that similar paradoxical behaviour is evident in the model during this time frame, as depicted in Figure S4. It is evident that the co-variance across these three bands gradually increases from good to poor ensembles in both the ISM and PAC regions (Figure 7a,b). Notably, for the ISM region, the 2-5 days variability exhibits a distinct contribution to seasonal prediction skill from good to bad ensembles, whereas contributions of the 10-20 days and 20-60 days bands are more significant for predicting the seasonal anomalies in the PAC region. Moreover, the model exhibits a considerable spread in the co-variances, particularly in the lower frequency bands (Figure 7c,d). These distributions of co-variances are also skewed in relation to the observations, a phenomenon that carries clear implications for model prediction skill. We also observe that the noise or ensemble spread increases (Figure 3) as the error in the co-variance grows from good to poor ensemble members (Figure 7a,b).

It may be noted that, in the case of seasonal prediction, the statistics of sub-seasonal components are important and not the actual timing of their occurrence. Moreover, the simulation of sub-seasonal statistics depends on the physics, dynamics, and their coupling with various components of the Earth's climate system (e.g., land, ocean, atmosphere). In general, global coupled models have serious difficulties in simulating synoptic systems, such as lows and depressions. Nevertheless, these synoptic events contribute significantly to the mean and variability of seasonal climate (e.g. Yoon and Chen, 2005; Saha et al., 2019). A disproportional contribution of sub-seasonal components to the seasonal mean is a big concern, as it can spoil the prediction skill, which is evident in Figure 7.

It is now evident that the model's ability to simulate the sub-seasonal contribution to seasonal anomalies significantly influences seasonal prediction skill. In addition to co-variability, it is equally important that the model accurately replicates the mean sub-seasonal components with high fidelity. Figure 8 provides an overview of the mean sub-seasonal variances within three temporal bands, both in absolute terms and as a percentage of the total variance. The model consistently underestimates synoptic (2-5 day band) variance in both the ISM and PAC regions. In the PAC region, both the 10-20 day and 20-60 day band variances are overestimated. When considering the total variance, the model overestimates it in both regions. Although the paradox has been shown to be most pronounced when the model's total variance closely matches that of observations (e.g., Scaife et al., 2014; Scaife and Smith, 2018), the model used here exhibits higher total variance than observed over the South Asian monsoon region (Section 3.2; Figs. 5b, 8c, 8d). This suggests that over-dispersion may contribute to the low signal-to-noise ratios diagnosed. Nevertheless, the persistence of paradoxical behavior even under these conditions underscores the influence of other model imperfections, including biases in sub-seasonal variability, in limiting predictability. Moreover, when expressed as a fraction of total variance, the synoptic-band variance is severely underestimated, accounting for only 2-7% of the total variance (Figure 8a,b) across ensemble members. It is intriguing to observe that in the observations, the synoptic variance contributes approximately 15% and 26% to the total variance over the ISM and PAC regions, respectively. The 20-60 day band is overestimated in the PAC region while slightly underestimated in the ISM region. Consequently, a discernible systematic trade-off emerges between the synoptic and MISO/MJO contributions within the model. This trade-off likely plays a role in explaining the model's comparatively lower prediction skill.

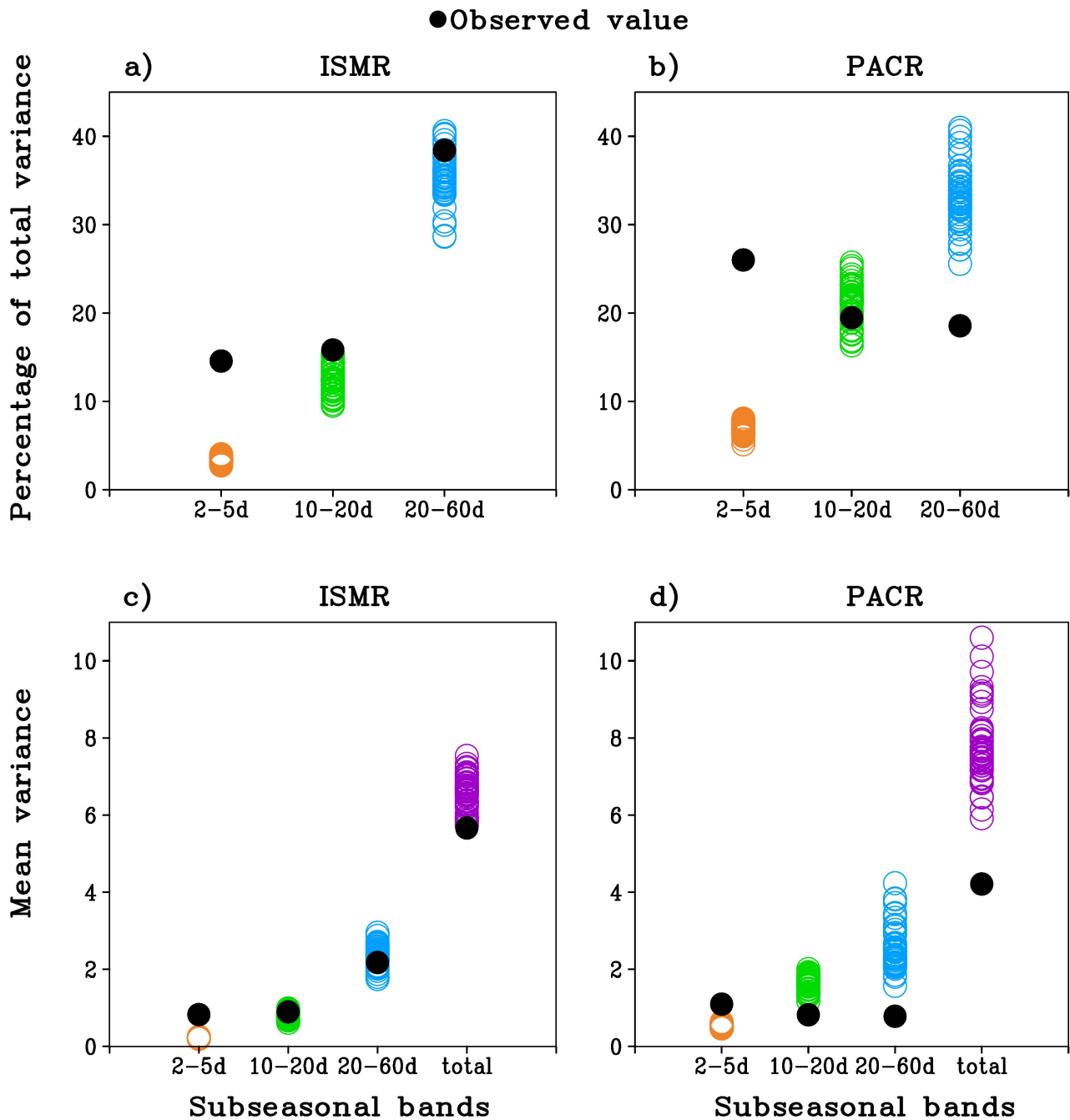


Figure 8. Mean sub-seasonal rainfall variances simulated by the model and observations (1997-2021). Variances in the 2-5, 10-20, and 20-60 day bands with respect to the total variance (in %) over a) the ISM domain and b) the Pacific domain. c) and d) are the same as a) and b), respectively, but show their actual values.

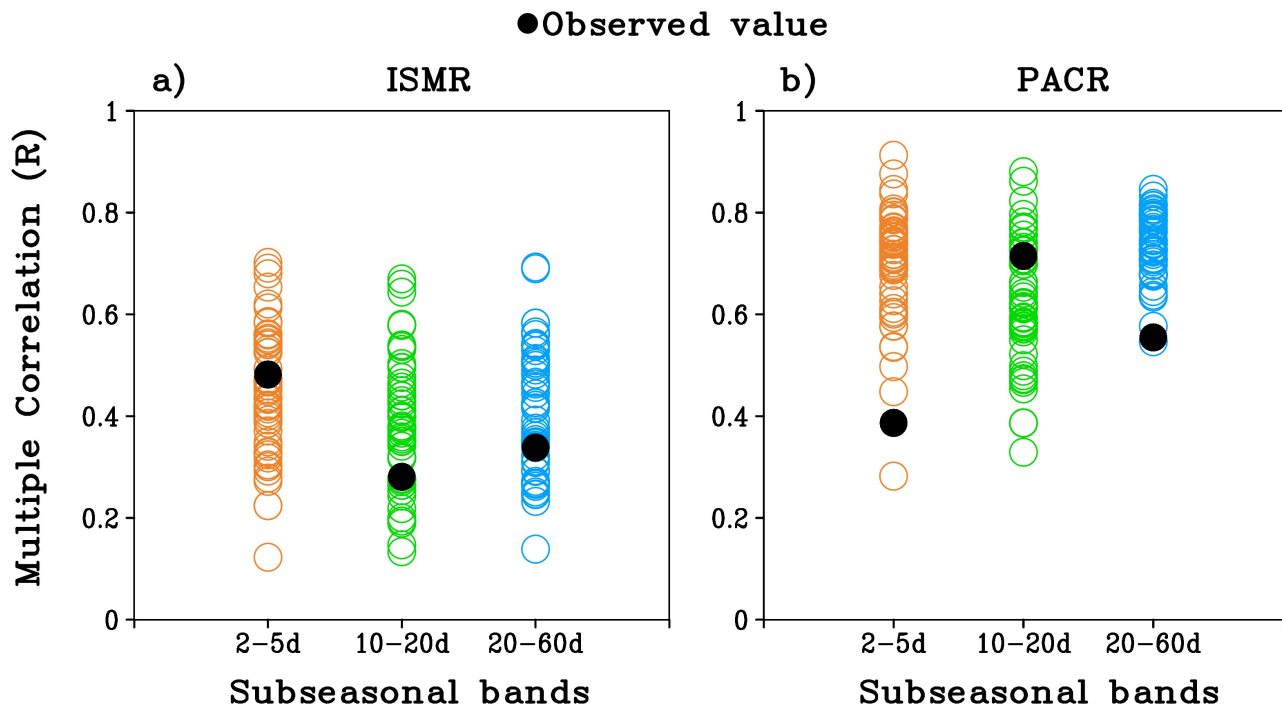


Figure 9. Multiple correlation analysis of sub-seasonal components (three bands) regressed with four global predictors (Niño3.4, IOD, PDO, AMO) for a) ISMR and b) PACR rainfall. Open circles represent individual ensemble members of the model, and black solid circles represent those in the observations.

Using multiple correlation analysis, we further investigate the impact of major global predictors, including ENSO, AMO, IOD, and PDO, on sub-seasonal components. Our analysis reveals significant patterns of association between these predictors and sub-seasonal bands (see Figure 9), reaffirming the findings of Saha et al. (2021). Observations show that, in the ISM region, the synoptic bands have the strongest association with these predictors. However, the 10-20 day and 20-60 day bands in the PAC region demonstrate more pronounced connections with the predictors. As expected, the model shows a wide range of associations between sub-seasonal components and predictors. A notable deviation is observed in the PAC region, where 2-5 day band and 20-60 day bands overestimate their associations with the predictors. Analogous to the patterns observed in Figure 7a and b, the multiple correlation distinctly separates good and poor ensemble members for PACR rainfall (Figure S5). However, for the ISMR, this distinction is not very prominent. Nonetheless, this analysis underscores the significance of major global predictors influencing sub-seasonal components, thereby elucidating the intricate dynamics that exert an influence on seasonal prediction skills.

Therefore, a significant association of sub-seasonal rainfall variance with the seasonal mean rainfall and four major global predictors is found in observations and, to a varying degree, in model simulations. Conversely, the model's shortcomings in

capturing interactions between fast (i.e. sub-seasonal) and slowly evolving components (e.g. ENSO) influence the seasonal mean and its variability. Consequently, the feasibility of estimating the PPL based on the ‘perfect model’ framework (or ANOVA) appears inadequate. The paradox we identify is therefore an expected consequence of the perfect-model assumption itself, which neglects real model errors and biases in sub-seasonal variability. We recall that the ‘perfect model’ framework assumes that the ensemble spread is due to an error in the initial conditions, and error growth due to imperfect physics or numerical scheme is not considered. It is crucial to acknowledge that real-world models are imperfect, and the realization of a truly perfect model is highly unlikely. Hence, the question arises: How can we obtain a reliable estimate of PPL that circumvents paradoxical behaviour?

3.4 Maximum achievable skill by model

Here, we explore the maximum skill, which can be achieved by the model. Forecast errors arise from various sources, including imperfect representations of physical processes, limitations in numerical methods, and inaccuracies in the initial conditions (ICs). Among these, uncertainties in the ICs can lead to substantial variations in forecast skill. Consequently, some ICs may result in more accurate predictions than others.

Thus, even when using the same forecasting model, different sets of ICs can produce ensemble forecasts with varying skills. The concept of the PPL aims to estimate the upper bound of prediction skill achievable by a forecast system, which is not a guaranteed outcome but the maximum possible skill under current model capabilities. A large ensemble of ICs can be used to generate a distribution of re-forecast skills for a given model. Since perfect initial conditions are practically unattainable, the best achievable forecast skill estimated from a sufficiently large ensemble will always remain below the idealized limit (i.e., correlation = 1.0). In this context, forecast skill is quantified as the correlation between observed and ensemble-mean ISMR or PACR. Accordingly, we define the maximum correlation skill obtained from a comprehensive set of ICs as the plausible PPL of the forecast system. The following points clarify the rationale behind this interpretation and ensure the concept remains logically consistent:

1. Predictability refers to the intrinsic potential for a system to be predicted, while the PPL represents the maximum achievable prediction skill by a specific forecast system.
2. The PPL is inherently model-dependent, varying with model physics, dynamics, and the characteristics of the initial conditions. It is, therefore, not a universal constant but specific to each forecast system.
3. Within a fixed model and data assimilation framework, variations in prediction skill arise solely from the different realizations of initial conditions derived from the same assimilation system.

It is thus essential to establish a robust method for determining the maximum achievable skill of a forecasting system at its current stage of development. This involves leveraging a wide range of initial conditions derived from advanced data assimilation systems. Among these, some ensemble combinations may exhibit notably high skill. Hence, the maximum skill attainable by the model, computed from the ensemble subsets can be considered its maximum achievable limit. To achieve

this, the maximum prediction skill is evaluated through correlations between observations and ensemble means, considering all possible combinations of ensemble members, i.e., nC_r , where n is the total number of ensemble members, and r is the number used in each subset. Similar ensemble subsampling distributions have been used to evaluate model predictive range (Saha et al., 2019), where improvements shift the skill envelope higher.

405 However, the use of a large number of ensembles can impose significant computational demands. For instance, with a set of 52 ensembles, there are approximately 4.9×10^{14} sub-sets of 26 ensemble members (without repetition of any ensemble member in a subset). An alternative approach is to generate a limited number of random sub-sets and determine the maximum skill. To illustrate this feasibility, we have selected the best 40 out of 52 ensembles for 41 years (1981-2021) and computed correlations between ensemble-averaged ISMR rainfall and observations for all possible combinations (up to a maximum of
410 1.378×10^{11} subsets). Figure 10a depicts the maximum, minimum, and mean correlations across various combinations (ranging from subsets of 2 to 40 ensemble members). Notably, it becomes apparent that a combination of 5-10 ensemble members yields a maximum correlation skill of approximately 0.76. It is also conceivable that an equal distribution of positive and negative skill across ensemble members could yield a high maximum attainable skill while the ensemble-mean skill remains insignificant or close to zero. Nevertheless, we show here that both the maximum and minimum attainable skills are positive, indicating that
415 there is no possibility of achieving a high maximum skill while the realized prediction skill is zero.

Now, we examine the possibility of using a reduced number of random combinations (up to a maximum of 10^6) while ensuring that ensemble members do not repeat within a sub-set or combination. Remarkably, similar patterns emerge in terms of maximum, minimum, and mean correlations when using randomly selected subsets from the full spectrum of possible combinations (Figure 10a). It is noteworthy that the mean, minimum, and maximum correlations tend to converge as the
420 number of combinations increases. This phenomenon is associated with the increase in the number of combinations from 2 to 20 and subsequent decrease, ultimately culminating in only one combination at 40 ensemble members. Nonetheless, the adoption of random combinations proves to be a cost-effective approach that does not compromise the quality of results.

Sampling limitations often hinder accurate estimation of population-level statistics, as available model or observational data may not fully represent the true statistical characteristics of the entire population. To address this, bootstrapping, a resampling
425 technique, offers an effective approach to derive more accurate population statistics from such sub-sampled data. In this study, we utilize 52 ensemble members over 41 years of re-forecasts, which may be insufficient to capture the model's full predictive potential. Since our goal is to estimate the maximum forecast skill, bootstrapping serves as a practical alternative to exhaustive combination methods.

Unlike fixed combinations where ensemble members are non-repeating, bootstrapping allows repetition, enabling the cre-
430 ation of millions of subsets of ensemble members. We employed 10 million such subsets, varying in size from 2 to 52 members, and calculated the maximum, minimum, and mean correlation skill (Figure 10b). Interestingly, the highest skill observed is approximately 0.76 for subsets of 5 members and around 0.66 for the full set of 52 members.

The same methodology can be extended to the grid level, enabling the estimation of the global pattern of maximum achiev-
435 able skill. Although the maximum skill for ISMR rainfall is achieved with approximately a 5-ensemble member subset (correlation coefficient ~ 0.76), we have chosen to use a 10-ensemble member subset to ensure a sufficient number of ensembles

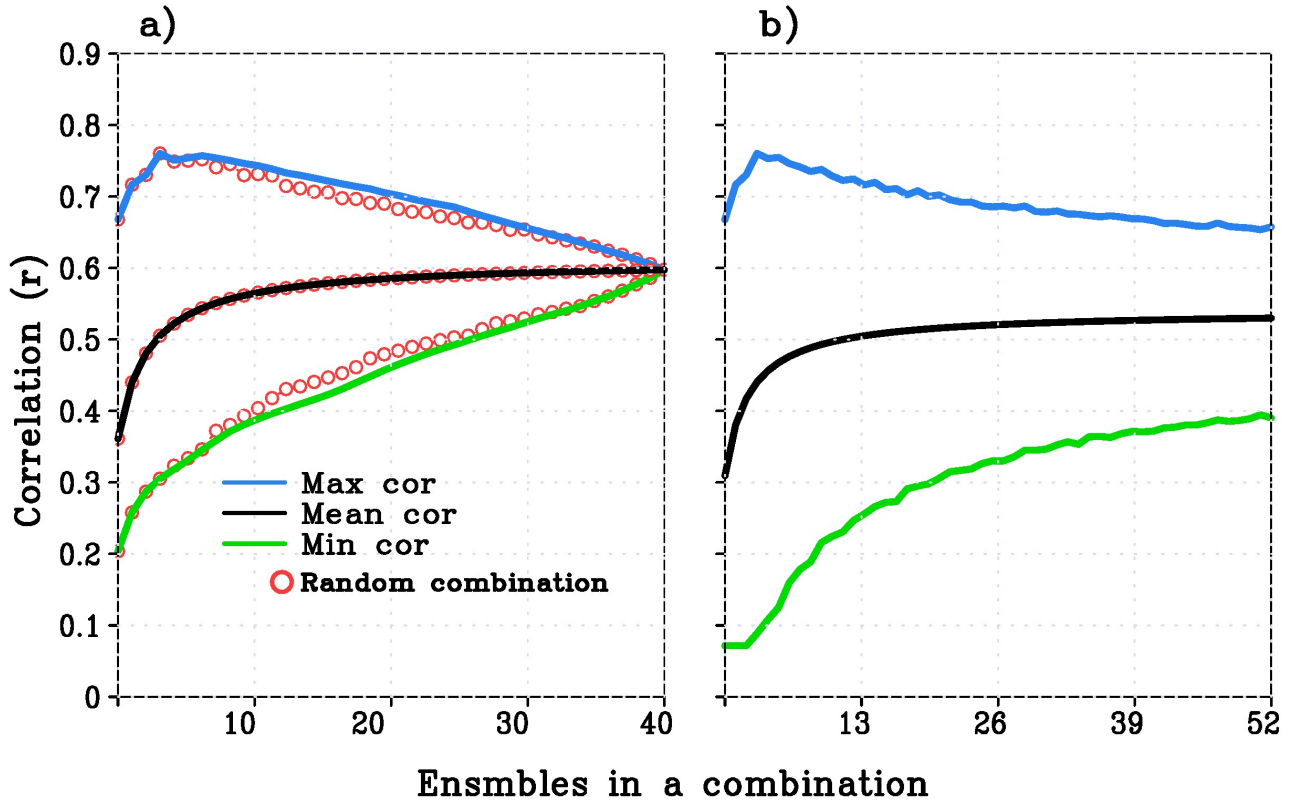


Figure 10. a) The maximum (blue solid line), mean (black solid line), and minimum (green solid line) correlation skill using all combinations of n ensemble-averaged ISMR rainfall (i.e., ${}^{40}C_n$, where n varies from 1 to 40). Similarly, the red open circles show the correlation skill using 1 million random combinations of n ensemble-averaged ISMR rainfall. b) The maximum (blue solid line), mean (black solid line), and minimum (green solid line) correlation skill using the Bootstrapping method with 10 million random combinations of n member ensemble-averaged (n varies from 1 to 52).

for PPL estimation at each grid point. Maximum correlation skill values are recorded at every grid point, derived from 10^5 combinations or subsets of 10 out of the 52 ensemble members using the bootstrap method. Consistent with the PPL estimates obtained through the ANOVA methods, the same three variables (rainfall, MSLP, SST/LST) have been employed in this analysis (Figure 11). It is clear that the highest attainable skill for rainfall is situated over the equatorial Pacific region, reaching a maximum of approximately 0.9. This is consistent with previous studies (e.g. Shukla, 1998) that have consistently demonstrated maximum prediction skill in this area compared to other regions of the globe. Regarding MSLP, skill appears to be more pronounced in the South Asian monsoon and eastern Pacific regions. A substantial portion of Europe, the Middle East, and northeastern Asia exhibits notable skill in 2-meter air temperature, with correlation coefficients ranging from approximately 0.8 to 0.9. It's worth noting that this method does not suffer from the paradox problem, as the maximum correlation value obtained by this technique is higher than the correlation skill, and this holds true for all grid points. Areas showing poor maximum skill or PPL can be targeted for further improvements, given that most ensemble members or their combinations demonstrate bad performance.

4 Summary and discussions

In this study, we employ a data set comprising 52 ensemble member re-forecasts spanning a 41-year period from 1981 to 2021, as simulated by the IITM CFS (Saha et al., 2019). The primary objective of this study is to elucidate the underlying factors contributing to the signal-to-noise paradox in seasonal climate prediction within the tropical and subtropical regions. We observe that many regions exhibit a signal-to-noise paradox when estimated by subtracting correlation skill from potential predictability estimated by the ANOVA method (Figure 1). This region also corresponds to areas where the ratio of predictable component (RPC) exceeds one. RPC is an alternative measure of predictability, with values surpassing one being deemed regions exhibiting paradoxical behaviour. Several prior studies have highlighted the presence of paradoxes in PPL estimates within the 'perfect model' framework (i.e. ANOVA). However, the underlying causes remain a subject of ongoing debate within the scientific community. As paradoxical behaviour is rare in weather time scale but prevalent on seasonal to decadal and beyond time scale, assessment of the underlying hypothesis for the method of estimating PPL is warranted. Therefore, to delve deeper into the intricacies of this paradoxical behaviour, we examine rainfall patterns in two distinct tropical regions: the Indian summer monsoon region (ISMR), exhibiting paradoxical behaviour over half of the region, and the central Pacific region (PACR), without any paradox.

While correlation skill in the time series of seasonal ISMR and PACR rainfall decreases from relatively good to poor ensemble members (arranged with respect to observations), the PPL exhibits minimal variation (Figure 3). This suggests that PPL is not closely associated with the actual skill of a model, contradicting the fundamental purpose of estimating it. In the ANOVA method, the separation of a forecast into 'signal' and 'noise' components is based on the assumption that initial error gives rise to the 'noise', while the model is considered to be perfect. Consequently, there should be no temporal relationship between 'noise' and 'signal' components. Conversely, it is observed that a significant and robust correlation exists between

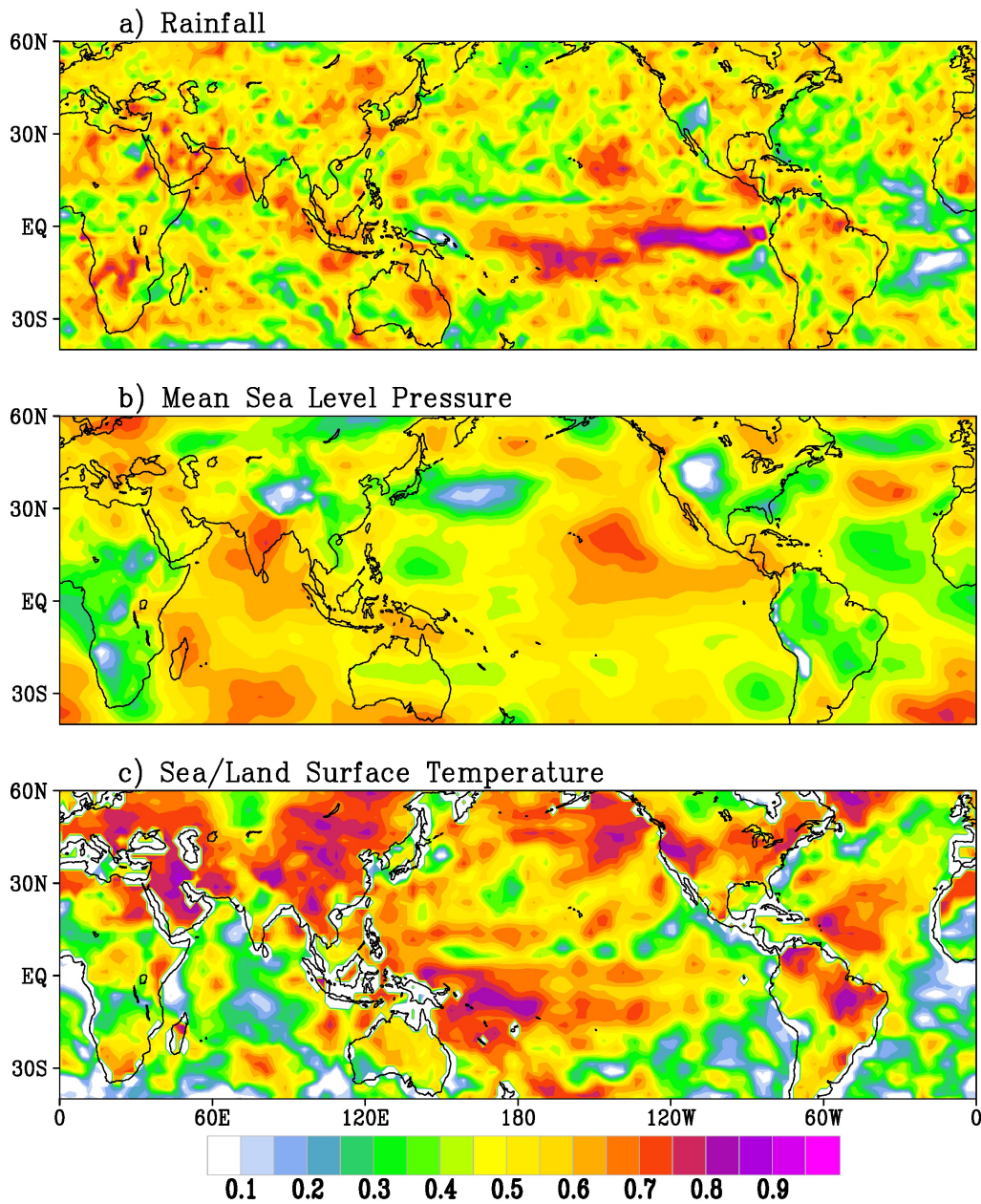


Figure 11. Maximum Correlation at each grid point using 100 thousand Bootstrapped re-sampling of 10 ensemble average from 52 ensembles and 41 years (1981-2021) simulation by IITM-CFS. Maximum correlation in JJAS-averaged a) Rainfall, b) Mean Sea Level Pressure, and c) surface temperature (SST over Ocean and 2m temperature over land).

these components. Furthermore, regions exhibiting paradoxical behaviour do not necessarily coincide with areas of significant correlation between the ‘signal’ and ‘noise’ (Figures 2).

470 Hence, the central issue underlying this problem is why the ANOVA method falls short in estimating climate predictability? What leads to the apparent inadequacy in separating a forecast into ‘signal’ and ‘noise’ components? Some of the answers can be found in the nature of the problem. Weather forecasting pertains to the predictability of the first kind, where the accuracy of initial conditions plays a primary role, and slowly evolving boundary conditions play a secondary role. Conversely, seasonal/decadal forecasting is categorized as predictability of the second kind, where slowly evolving boundary conditions
475 play a major role. Moreover, the physical processes which affect the sub-seasonal variability at various frequencies (e.g. lows and depressions, MJOs, etc) can also affect the seasonal mean. As the real world’s models are not perfect, these deficiencies can also add to the variability or ensemble spread in addition to that due to errors in the initial conditions.

Therefore, understanding the co-variability between sub-seasonal components and seasonal means within the context of predictability and prediction is of paramount importance. While prediction skill demonstrates systematic variations with the
480 covariance between sub-seasonal components and seasonal mean rainfall, the PPL remains quite invariant (Figure 3). It becomes evident that the model exhibits a broad and often skewed range of co-variability between sub-seasonal variance and seasonal mean in comparison to observations, with sub-seasonal components strongly linked to global predictors (see Figures 7, 9,). Additionally, ensemble spread (or noise component) is found to vary systematically with this co-variance and prediction skill (Figure 3, 7). Consequently, a robust methodology for estimating PPL must consider not only the errors in initial condi-
485 tions but also those arising from model physics, dynamics, and numerical methods employed. These findings underscore the critical significance of accurately simulating sub-seasonal components and their co-variability in seasonal prediction, shedding light on factors influencing prediction skill and identifying potential areas for model enhancement.

Building upon the work of Saha et al. (2019), we employ a simple method to estimate the maximum prediction skill by a model. This method serves to determine the upper bound on seasonal climate prediction skill attainable by a model. In
490 consideration of the constraint posed by a finite sample size (i.e., 52 ensembles for 41 years), and its potential impact on the accuracy of potential predictability estimates, we employ the Bootstrapping method. This statistical technique enables us to infer the maximum prediction skill of the entire ensemble population by re-sampling from the available ensembles, thereby mitigating the influence of sample size limitations on our estimates. Since estimates of PPL are model-dependent and subject to change with improvements in the model and initial conditions (e.g. Kumar et al., 2014), we propose that the
495 maximum prediction skill obtained through our method represents the model’s PPL at its current state of development and initial conditions. This approach effectively avoids paradoxical situations.

Here we use 41 years of model re-forecasts and corresponding observations. However, longer observational and re-forecast datasets would allow a more robust estimation of the predictability limit, particularly with respect to decadal and multi-decadal variability. Our diagnostic quantifies the range of prediction skill attainable by the model when evaluated against observations;
500 however, estimates of maximum skill should be interpreted with caution, as they may be affected by sampling uncertainty arising from the finite ensemble size. We emphasize that this approach is intended as a diagnostic sensitivity analysis rather than a directly implementable operational forecasting algorithm. In real-time prediction systems, the growth of initial condition errors,

together with model errors, limits the practical attainability of the predictability limit, and the apriori identification of “best” ensemble members is not feasible. Nevertheless, emerging AI–ML approaches, trained on longer model simulations, reanalyses, and observational datasets, offer a potentially promising pathway by learning complex subseasonal-to-seasonal relationships with large-scale predictors (e.g., Sharma et al., 2025), which may help mitigate error growth and enhance operational monsoon forecasting skill with longer lead time in the future.

Code and data availability. The code and scripts used in the preparation of this manuscript are publicly available in Yashas (2025) <https://doi.org/10.5281/zenodo.15369106>. The observational data sets used in the study are: 2m air temperature data over land from the Climatic Research Unit (CRU TS3.1) available in Harris et al. (2014). SST data is taken from EN4 reanalysis, which is available in Good et al. (2013). MSLP data is obtained from ERA5 reanalysis, available freely in Hersbach et al. (2020). Daily rainfall data from the Global Precipitation Climatology Project (GPCP Version 1.3) available in Huffman et al. (2001) for 1997-2021 and monthly rainfall data from GPCP version 2.3 for 1979-2021 available in Adler et al. (2020) . CFS re-forecast of monthly SST, MSLP, surface temperature and daily precipitation for June-to-September (1981-2021) with 52 ensemble members are publicly available in Saha (2024) <https://doi.org/10.5281/zenodo.13166897>. Freely available software GrADS is used for plot and data analysis (<http://cola.gmu.edu/grads/>).

Author contributions. YS and SKS conceptualized the idea and carried out model simulation and data analysis, and wrote the manuscript. SP, MK and UV contributed to the discussion, plotting Figures and manuscript writing.

Competing interests. There are no competing interests.

Acknowledgements. We thank MoES, the Government of India and the Director IITM for all the support in carrying out this work.

520 References

- Adler, R., Wang, J.-J., Sapiano, M., Huffman, G., Chiu, L., Xie, P. P., Ferraro, R., Schneider, U., Becker, A., Bolvin, D., Nelkin, E., Gu, G., , and NOAA CDR Program: Hybrid gridded demographic data for the world, 1950-2020, <https://doi.org/10.5281/zenodo.3768003>, 2020.
- Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, *Nature*, 525, 47–55, 2015.
- Borah, P. J., Venugopal, V., Sukhatme, J., Muddebihal, P., and Goswami, B. N.: Indian monsoon derailed by a North Atlantic wavetrain, 525 *Science*, 370, 1335–1338, 2020.
- Bröcker, J., Charlton-Perez, A. J., and Weisheimer, A.: A statistical perspective on the signal-to-noise paradox, *Quarterly Journal of the Royal Meteorological Society*, 149, 911–923, 2023.
- Charney, J. G. and Shukla, J.: Predictability of Monsoons, in: *Monsoon Dynamics*, edited by Lighthill, J. and Pearce, R. P., pp. 99–108, Cambridge University Press, Cambridge, 1981.
- 530 Cottrell, F. M., Screen, J. A., and Scaife, A. A.: Signal-to-noise errors in free-running atmospheric simulations and their dependence on model resolution, *Atmos. Sci. Lett.*, 25(6), <https://doi.org/https://doi.org/10.1002/asl.1212>, 2024.
- Duchon, C. E.: Lanczos filtering on one and two dimensions, *J. Appl. Meteorol.*, 18, 1016–1022, 1979.
- Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, *Geophys. Res. Lett.*, 41, 5620–5628, 2014.
- 535 Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarplay, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, 1089(D22), 8851, [doi:10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296), 2003.
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *J. Geophys. Res.: Oceans*, 118, 6704–6716, <https://doi.org/10.1002/2013JC009067>, 2013.
- 540 Griffies, S. M., Harrison, M. J., Pacanowski, R. C., and Rosati, A.: A Technical guide to MOM4, GFDL Ocean Group Technical Report 5, GFDL, 337 pp, 2004.
- Harris, I. P. D. J., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations-The CRU TS3.10 57 dataset, *Int. J. Climatol.*, 34, 623–642, <https://doi.org/10.1002/joc.3711>, 2014.
- Hazra, A., Chaudhari, H. S., Saha, S. K., Pokhrel, S., and Goswami, B. N.: Progress towards achieving the challenge of Indian summer 545 monsoon climate simulation in a coupled ocean-atmosphere model, *J. Adv. Model. Earth Syst.*, 9, 2268–2290, 2017.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, 550 J.-N.: The ERA5 global reanalysis, *Q. J. R. Meteorol. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hu, Z.-Z., Kumar, A., and Zhu, J.: Dominant modes of ensemble mean signal and noise in seasonal forecasts of SST, *Climate Dynamics*, 56, 1251–1264, 2021.
- Huffman, G. J., Adler, R. F., Morrissey, M. M., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multi-satellite observations, *J. Hydrometeorol.*, 2, 36 – 50, <https://rda.ucar.edu/datasets/dsd728003/>, 2001.
- 555 Jain, S., Scaife, A. A., and Mitra, A. K.: Skill of Indian summer monsoon rainfall prediction in multiple seasonal prediction systems, *Clim. Dyn.*, 52, 5291 – 5301, 2019.

- Kang, I.-S. and Shukla, J.: Dynamic Seasonal Prediction and Predictability of the Monsoon, in: *The Asian Monsoon*, edited by Wang, B., pp. 585–612, Praxis, Springer, Berlin, Heidelberg, https://doi.org/10.1007/3-540-37722-0_15, 2006.
- Klavans, J., Cane, M., and Clement, A. e. a.: NAO predictability from external forcing in the late 20th century, *npj Clim. Atmos. Sci.*, 4,22, <https://doi.org/https://doi.org/10.1038/s41612-021-00177-8>, 2021.
- 560 Kumar, A., Peng, P., and Chen, M.: Is There a Relationship between Potential and Actual Skill?, *Mon. Weather Rev.*, 142, 2220–2227, 2014.
- Kumar, K. K., Hoerling, M., and Rajagopalan, B.: Advancing dynamical prediction of Indian monsoon rainfall, *Geophys. Res. Lett.*, 32, L08 704, doi:10.1029/2004GL021 979, 2005.
- Lorenz, E. N.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130 – 141, 1963.
- 565 Lorenz, E. N.: Three approaches to atmospheric predictability, *Bul. Amer. Meteor. Soc.*, 50, 345 – 351, 1969.
- Mayer, B., Düsterhus, A., and Baehr, J.: When Does the Lorenz 1963 Model Exhibit the Signal-To-Noise Paradox?, *Geophysical Research Letters*, 48, e2020GL089 283, 2021.
- Mooley, D. and Parthasarathy, B.: Indian summer monsoon and the east equatorial pacific sea surface temperature, *Atmosphere-Ocean*, 22, 23–35, <https://doi.org/10.1080/07055900.1984.9649182>, 1984.
- 570 Peixoto, J. P. and Oort, A. H.: *Physics of Climate*, American Institute of Physics, New York, U.S.A., 520 pp., 1992.
- Pokhrel, S., Saha, S. K., Dhakate, A., Rahman, H., Chaudhari, H. S., Salunke, K., Hazra, A., Sujith, K., and Sikka, D. R.: Seasonal prediction of Indian summer monsoon rainfall in NCEP CFSv2: forecast and predictability error, *Clim. Dyn.*, 46, 2305–2326, DOI 10.1007/s00 382–015–2703–1, 2016.
- Rajeevan, M., Unnikrishnan, C. K., and Preethi, B.: Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability, *Clim. Dyn.*, 38, 2257–2274, 2012.
- 575 Rowell, D. P.: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulation, *J. Clim.*, 11, 109–120, 1998.
- Rowell, D. P., Folland, C. K., Maskell, K., and Ward, M. N.: Variability of summer rainfall over tropical North Africa (1906–92): Observations and modeling, *Q. J. R. Meteorol. Soc.*, 121, 669–704, 1995.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Dool, H. V. D., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C. Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, *Bul. Amer. Meteor. Soc.*, 91, 1015 – 1057, 2010.
- 580 Saha, S., Moorthi, S., Wu, X., Wang, J., Pan, H.-L., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mensez, M. P., Dool, H. V. D., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, *J. Clim.*, 27, 2185 – 2208, 2014a.
- Saha, S. K.: Why does the signal-to-noise paradox exist in seasonal climate predictability?, <https://doi.org/10.5281/zenodo.13166897>, 2024.
- Saha, S. K., Pokhrel, S., Chaudhari, H. S., Dhakate, A., Shewale, S., Sabeerali, C. T., Salunke, K., Hazra, A., Mahaptra, S., and Rao, A. S.: Improved simulation of Indian summer monsoon in latest NCEP climate forecast system free run, *Int. J. Climatol.*, 35, 1628–1641, 2014b.
- 590 Saha, S. K., Pokhrel, S., Salunke, K., Dhakate, A., Chaudhari, H. S., Rahaman, H., Sujith, K., Hazra, A., and Sikka, D. R.: Potential Predictability of Indian Summer Monsoon Rainfall in NCEP CFSv2, *J. Adv. Model. Eart. Syst.*, p. DOI: 10.1002/2015MS000542, 2016a.
- Saha, S. K., Sujith, K., Pokhrel, S., Chaudhari, H. S., and Hazra, A.: Predictability of global Monsoon Rainfall in NCEP CFSv2, *Clim. Dyn.*, 47, 1693–1715, 2016b.

- 595 Saha, S. K., Hazra, A., Pokhrel, S., Chaudhari, H. S., Sujith, K., Rai, A., Rahaman, H., and Goswami, B. N.: Unraveling the Mystery of Indian Summer Monsoon Prediction: Improved Estimate of Predictability Limit, *J. Geophys. Res.*, 124, 1962–1974, 2019.
- Saha, S. K., Hazra, A., Pokhrel, S., Chaudhari, H. S., Sujith, K., Rai, A., Rahaman, H., and Goswami, B. N.: Reply to Comment by E. T. Swenson, D. Das, and J. Shukla on “Unraveling the Mystery of Indian Summer Monsoon Prediction: Improved Estimate of Predictability Limit”, *J. Geophys. Res.*, 125, e2020JD033 242. <https://doi.org/10.1029/2020JD033 242>, 2020.
- 600 Saha, S. K., Konwar, M., Pokhrel, S., Hazra, A., Chaudhari, H. S., and Rai, A.: Interplay between subseasonal rainfall and global predictors in modulating interannual to multidecadal predictability of the ISMR, *Geophys. Res. Lett.*, 48, e2020GL091 458. <https://doi.org/10.1029/2020GL091 458>, 2021.
- Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., Eade, R., Fereday, D., Folland, C., Gordon, M., et al.: Skillful long-range prediction of European and North American winters, *Geophysical Research Letters*, 41, 2514–2519, 2014.
- 605 Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, *npj Clim. Atmos. Sci.*, 1,28, doi:10.1038/s41 612–018–0038–4, 2018.
- Scheffe, H.: *The Analysis of Variance*, John Wiley and Sons, New York, 1959.
- Schneider, T. and Griffies, S. M.: A Conceptual Framework for Predictability Studies, *J. Clim.*, 12, 3133–3155, 1999.
- Sévellec, F. and Drijfhout, S.: The signal-to-noise paradox for interannual surface atmospheric temperature predictions, *Geophysical Research Letters*, 46, 9031–9041, 2019.
- 610 Sharma, D., Das, S., Chakraborty, D., Mitra, A., and Goswami, B. N.: Improving Indian summer monsoon rainfall prediction using deep learning up to two years in advance, *Q. J. R. Meteorol. Soc.*, n/a, e70 023, <https://doi.org/https://doi.org/10.1002/qj.70023>, 2025.
- Shukla, J.: Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting, *Science*, 282, 728 – 731, 1998.
- Shukla, J.: Monsoon Mysteries, *Science*, 318, 204–205, 2007.
- 615 Smith, D., Eade, R., Scaife, A., Caron, L.-P., Danabasoglu, G., DelSole, T., Delworth, T., Doblas-Reyes, F., Dunstone, N., Hermanson, L., et al.: Robust skill of decadal climate predictions, *npj Climate and Atmospheric Science*, 2, 13, 2019.
- Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L., Caron, L.-P., Counillon, F., et al.: North Atlantic climate far more predictable than models imply, *Nature*, 583, 796–800, 2020.
- Strommen, K. and Palmer, T. N.: Signal and noise in regime systems: A hypothesis on the predictability of the North Atlantic Oscillation, *Quarterly Journal of the Royal Meteorological Society*, 145, 147–163, 2019.
- 620 Walker, G. T.: Correlation in seasonal variations of weather, IX. A further study of world weather, *Memoirs of the Indian Meteorological Department*, 24, 275 – 332, 1924.
- Weisheimer, A., Decremmer, D., MacLeod, D., O’Reilly, C., Stockdale, T., Johnson, S., and Palmer, T.: How confident are predictability estimates of the winter North Atlantic Oscillation?, *Quarterly Journal of the Royal Meteorological Society*, 145, <https://doi.org/10.1002/qj.3446>, 2018.
- 625 Westra, S. and Sharma, A.: An Upper Limit to Seasonal Rainfall Predictability?, *J. Clim.*, 23, 3332–3351, 2010.
- Winton, M.: A reformulated three-layer sea ice model, *J. Atmos. Ocean. Tech.*, 17, 525 – 531, 2000.
- Wu, X., Simmonds, I., and Budd, W. F.: Modeling of Antarctic sea ice in a general circulation model, *J. Clim.*, 10, 593 – 609, 1997.
- Yang, D., Tang, Y., Zhang, Y., and Yang, X.: Information-based potential predictability of the Asian summer monsoon in a coupled model, *J. Geophys. Res.*, 117, D03119, doi:10.1029/2011JD016 775, 2012.
- 630 Yashas, S.: Codes used in the preparation of the article "Why does the signal-to-noise paradox exist in seasonal climate predictability?", <https://doi.org/10.5281/zenodo.15369106>, 2025.

Yoon, J.-H. and Chen, T.-C.: Water vapor budget of the Indian monsoon depression, *Tellus A*, 57, 770–782, 2005.

Zhang, W. and Kirtman, B.: Understanding the signal-to-noise paradox with a simple Markov model, *Geophysical Research Letters*, 46, 635 13 308–13 317, 2019.

Zhang, W., Kirtman, B., Siqueira, L., Clement, A., and Xia, J.: Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddy global coupled model, *Climate dynamics*, 56, 2895–2913, 2021.