Article

Identifying source of predictability for vapor pressure deficit variability in the southwestern United States

Check for updates

Jiale Lou[1] ✉, Youngji Joh[2], Thomas L. Delworth[2] & Liwei Jia[2]

Atmospheric vapor pressure deficit (VPD) measures the difference between saturation vapor pressure and actual vapor pressure, and its variability is closely related to fire activity in the western United States (US). Here, we assess the forecast skill of monthly VPD variability using a state-of-the-art dynamical forecast system and statistical predictions, such as the persistence forecast and model-analog forecasts. In the model-analog framework, we select analog states resembling the observed initial conditions from the model space, and the subsequent evolution of those initial model-analogs yields forecast ensembles. Dynamical forecasts demonstrate skillful predictions of VPD variability in the western US, exceeding the persistence forecast skill, which indicates additional sources of VPD predictability within the climate system. To quantify the contribution of different climate variables to VPD prediction, we develop a weighted model-analog forecast and evaluate its skill in comparison to VPD-only and unweighted forecasts. Our findings suggest that sea surface temperature is a critical source of VPD predictability over the western US. The optimally weighted model-analog exhibits forecast skill for VPD variability comparable to that of the dynamical forecast system.

Seasonal-to-interannual climate predictions of fire-prone conditions provide valuable information for assessing the risk and potential severity of wildfires, as well as for implementing preparatory measures. Factors such as temperature, humidity, wind speed, and precipitation influence the likelihood and behavior of large-scale wildfire activity. Among various meteorological fire-prone indices[1–4], atmospheric vapor pressure deficit (VPD), defined as the difference between saturation and actual vapor pressure, has gained popularity due to its direct consideration of both heat and humidity and the close relationship between its variability and fire activity[5–13]. In essence, higher VPD values indicate drier atmospheric conditions, thereby increasing the likelihood of fire activity.

However, predicting VPD variability on seasonal time scales poses challenges. One of the challenges arises from model biases. Previous studies[14,15] have found that state-of-the-art climate models often misrepresent both the long-term trends and decadal variability of actual vapor pressure in the southwestern United States (SWUS). Addressing this issue requires thorough investigation to enhance the models' performance. In particular, Lou et al. [15] showed that the leading VPD mode (i.e., the first empirical orthogonal function (EOF) mode derived from the monthly VPD anomalies over the contiguous US), centered in the SWUS (loosely defined as the region in 24°-40°N, 90°-120°W), displays pronounced variations on inter-decadal time scales and is closely related to El Niño-Southern

Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO). Given that sea surface temperature (SST) variability related to ENSO and the PDO has been shown to be important precursors for the development of VPD variations in the SWUS[16–18], the natural follow-up questions here are to what extent is VPD variability in the SWUS predictable? How much contribution arises from SST in predicting VPD variations? To address these questions, we compare the state-of-the-art numerical forecasts generated by the Geophysical Fluid Dynamics Laboratory (GFDL) Seamless System for Prediction and Earth System Research (SPEAR) seasonal forecast system[19–21] with two benchmark forecasting tools (i.e., persistence and model-analog forecasts[22,23]) in this study.

Overall, a few studies have specifically examined the predictions and predictability of fire weather–related quantities. For instance, Di Giuseppe et al. [2] demonstrated that globally anomalous fire weather conditions can be confidently predicted only up to one month in advance, with skillful predictions extending to two months ahead in certain regions. However, in many cases beyond this timeframe, forecasts do not exhibit greater skill than climatology. Nevertheless, there exists an extended predictability window of up to 6–7 months ahead when anomalous fire weather results from large-scale phenomena, such as ENSO. Similarly, Chen et al. [24] suggested that ENSO accounts for about 1/3 of the total predictable global wildfire burned area. Some other studies[25,26] examined the relationship between ENSO and

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA. [2]Geophysical Fluid Dynamics Laboratory/NOAA, 201 Forrestal Road, Princeton, NJ, USA. ✉e-mail: jiale.lou@noaa.gov

fire-prone conditions. It remains an open question, as posted in refs. 27,28, whether and to what extent the forecast skill of VPD could be improved through the existence of windows of forecast opportunity related to atmospheric and oceanic variability.

Fully coupled operational climate models face challenges in decomposing each source of predictability that contributes to VPD forecasts. Here, we apply an alternative method – the model-analog technique[22,23] – to isolate sources of predictability related to VPD, SST, sea level pressure (SLP), and soil moisture (SM). The general idea of this approach is that two climate states will evolve in a similar fashion if they closely resemble each other at the initial states[29]. Initially, Lorenz[29] searched for observational (or 'naturally occurring'[29]) analogs within the observations but found that no pair of analogs was sufficiently similar to be useful for prediction. Later, to overcome the limitations of short observational records, studies[22,23,30,31] began utilizing large pre-existing climate model datasets as libraries for identifying initial model-analogs. Here, because the initial model-analog states are selected from the model space directly, it efficiently eliminates the initialization shock[32] that the traditional assimilation-initialized numerical models[22,30] commonly encounter. For similar reasons, the model-analog technique has also been employed to evaluate the performance of state-of-the-art numerical models[33].

Previous studies showed the usefulness of the model-analog approach to predict ENSO variability[22,23,31], the Indian ocean dipole[34], decadal variability in the Southern Ocean[35], sea level anomalies[36], and crop yields[37]. Specifically, refs. 22,34 used unweighted information from SST and sea surface height (SSH) to identify initial model-analogs for ENSO and IOD predictions, respectively, referring to this method as the equally weighted model-analog approach. 'Unweighted' here means all the variables contribute equally to the selection of the initial model-analogs. Similarly, while incorporating additional variables—such as sea surface and subsurface ocean temperatures and salinities—to skillfully predict decadal variability in the Southern Ocean, ref. 35 also applied an unweighted framework. It remains unclear how effective the traditional equally weighted model-analog technique is in predicting hydroclimatic variables such as VPD. In this study, we incorporate various predictors, such as SST from different ocean sectors, SLP, and SM, and design three sets of model-analog forecasting experiments to identify the sources of predictability that contribute to the seasonal forecast skill of VPD. These model-analog experiments include VPD-only, traditional equally weighted, and weighted model-analog approaches.

Results

Predictions and predictability of VPD variability

We began our analysis by examining the SPEAR seasonal forecast system[19]. To evaluate the forecast skill, we utilized an extensive set of SPEAR hindcasts[20,21] as described in the data and methods section. The hindcasts were then verified against the monthly ERA5 reanalysis datasets for the period 1992-2022.

Figure 1A presents an overview of the spatiotemporal characteristics of the leading VPD mode of variability, derived from EOF analysis of the monthly ERA5 reanalysis data for the period 1992–2022. The leading VPD mode explains 27.9% of the total variance and exhibits a center of action in the SWUS (loosely defined as the region in the black box in Fig. 1A). To ensure consistency in statistical and physical features when truncating the verification period to 1992–2022, we conducted EOF analysis on ERA5 reanalysis over the full period (1940–2022) and the pre-1992 period (1940–1991), respectively. Although the verification period is truncated, the spatial pattern and temporal variability of the leading EOF mode of the monthly VPD anomalies (Fig. 1 A) closely resemble that from the whole and pre-1992 periods (results not shown). Here, to maintain consistent physical interpretations and ensure comparability among the climate indices, we exclusively utilized area-averaged indices in our analysis. That is, the area-averaged monthly VPD anomalies in the southwestern US served as the predictand time series.

Figure 1B–D provide an overview of the SPEAR seasonal forecast system. Specifically, Fig. 1B presents the monthly initialized ensemble

forecasts for SWUS VPD, as well as ENSO and the PDO, represented by monthly area-averaged SST anomalies in the NINO3.4 region and the North Pacific, respectively (see data and methods section for details). In general, oceanic states contribute to predictability beyond the influence of initial conditions on seasonal to decadal timescales (e.g., Fig. 1 in ref. 38). Previous studies have shown that ENSO and PDO serve as important precursors for SWUS VPD evolution, as oceanic states influence large-scale teleconnection patterns, moisture transport, and land surface conditions. Given their established roles in modulating hydroclimate variability, comparing VPD predictions with SST-based predictions of ENSO and PDO helps elucidate the sources of predictability, a topic that will be explicitly discussed in later sections.

We can see that the ensemble forecast spread for the VPD quickly diverges, as indicated by the individual ensemble member trajectories (grey curves in Fig. 1B). Conversely, the ensemble forecast spreads for ENSO and the PDO generally align well with their respective verifications. This divergence in forecast accuracy is further reflected in the anomaly correlation (AC) skill (Fig. 1C), where the area-averaged SWUS VPD variability demonstrates the lowest forecast skill, while the NINO3.4 index and the PDO display generally higher forecast skill. Notably, when linear trends are removed from both the verification and forecasts, the AC skill of the PDO (red dashed curve in Fig. 1C) begins to decline by 0.11 at the seven-month forecast lead and continues to decrease until the end of the forecast period, with a total reduction in skill of 0.18 compared to forecasts that include trends. This suggests that the PDO trend significantly ($p < 0.05$) enhances its forecast skill, particularly at longer lead times. However, the presence of trend components does not significantly affect the forecast skill of ENSO or the area-averaged VPD variability for the verification period from 1992 to 2022 (not shown).

The seasonality of forecast skill for SWUS VPD, ENSO, and PDO is shown in Fig. 1D. Generally, forecast skill for these climate indices varies not only with the lead time but also with the season in which the forecast is initialized. This seasonality may stem from inherent seasonal predictability linked to each variable (e.g., ENSO spring predictability barrier[39–41]) or the models' capacity to simulate the climate system's dynamics at different times of the year. Notably, the dynamical forecasts can skillfully predict the SWUS VPD up to 5 months in advance by using December, January, and February 1st initialized hindcasts. Moreover, there appears to be a seasonal skill minimum where VPD forecasts show no skill when targeting summertime VPD and maximum skill when targeting springtime VPD, regardless of the initial months. While less skillful compared to ENSO and PDO forecasts, VPD forecast skill generally resembles their forecast skill (Fig. 1D). Figure 1D suggests that the seasonality of VPD forecast skill is closely related to that of ENSO and the PDO. A natural follow-up question arises: What role does SST play in contributing to the forecast skill of VPD, and to what extent? However, the challenge with breaking down sources of predictability in SPEAR seasonal forecasts lies in its coupled nature, making it difficult to isolate individual sources without running costly, traditionally assimilation-initialized atmospheric and oceanic hindcasts. Despite few attempts, such as pacemaker experiments[42,43] to isolate sources of predictability, numerical model-based equivalents are limited due to their computational cost. This is where the model-analog approach offers an advantage, which will be elaborated in the following sections.

Comparison with benchmark forecasts

To further evaluate the forecast skill of VPD variability, in this section, we compare the SPEAR seasonal forecast system with a couple of benchmark forecasts, including persistence forecasts and model-analog forecasts. Firstly, the forecast skill of VPD generated by the state-of-the-art climate model is better than persistence forecasts in most of the western US and southeastern Canada but poorer in the Midwest and eastern US at the first three-month forecast leads (Fig. 2). While the poorer forecast skill of SPEAR in those regions is of interest, they are not typically wildfire-prone areas and will warrant future investigation. A possible interpretation is that the model-simulated VPD variability damps too quickly in this region, resulting in
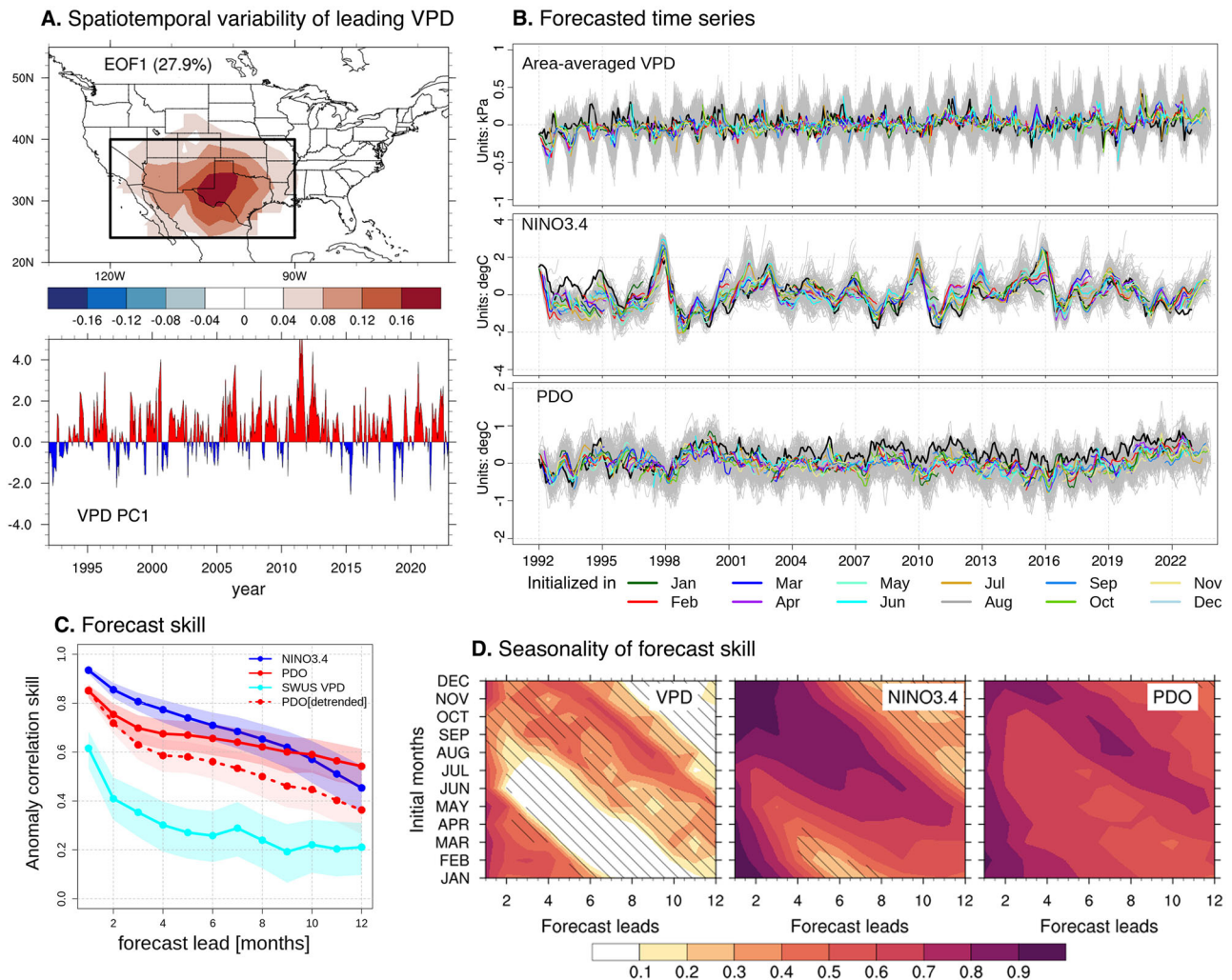
Fig. 1 | Spatiotemporal features of the leading vapor pressure deficit (VPD) mode and its predictability. A Spatial pattern of the leading VPD mode and corresponding PC time series. The black box indicates the southwest US area analyzed in this study. **B** Observed and predicted time series for the area-averaged VPD in the southwest US, NINO3.4, and the PDO. Black curves represent observations. Grey curves represent the forecast trajectories of individual forecast members with forecast leads up to 12 months, and colored curves represent the forecast trajectories of the ensemble mean forecasts with different colors denoting different initialization months. For instance, the dark green color represents the forecast trajectory initialized in January with forecast leads up to 12 months. **C** Ensemble-mean anomaly correlation (AC) skill for area-averaged VPD in the southwest US, NINO3.4, and the PDO, respectively, with forecast leads up to 12 months. The shading indicates the 5-95% confidence interval based on the bootstrapping significance test. **D** Ensemble-mean AC skill for area-averaged VPD in the southwest US, NINO3.4, and the PDO as a function of the initialization month (vertical axis) and forecast lead times up to 12 months (horizontal axis). AC skill is computed based on the entire 1992–2022 period using verification datasets from ERA5 reanalysis. Forecasts were generated from the SPEAR seasonal forecasting system with 15 ensemble members. Hatching indicates correlation coefficients not statistically significant at the 95% confidence level following ref. [63].

reduced persistence. Overall, the forecast skill of the VPD is generally high at the first forecast lead, with an overall skill greater than 0.5 in the SWUS (Fig. 2a). However, the forecast skill quickly declines with increasing forecast leads. After the initial three months' evolution, the VPD forecasts generally become unskillful (not shown), which is consistent with the forecast skill of the SWUS VPD time series in Fig. 1C.

Despite having lower forecast skill in predicting VPD variability in the SWUS, the persistence forecasts exhibit similar seasonality (Supplementary Fig. 1). Specifically, winter-initialized forecasts (from October to January) can skillfully predict SWUS VPD up to approximately six months in advance (Supplementary Fig. 1). However, the forecasts generally show no skill when targeting the summer and fall seasons. This implies that during winter, when ENSO activity peaks, ENSO provides a potential source of predictability that sustains SWUS VPD variability into the following spring.

Previous studies[23,31,35,44] showed that model-analog forecasts offer comparable, and in some cases, better forecast skill when predicting SST-based variability. This is primarily attributed to the direct selection of initial

states from the model space, which can effectively avoid initialization shock. To take account of uncertainties arising from model structures, here, we first use 12 CMIP6 models (see supplementary materials for details) as data libraries to search for the initial VPD model-analogs that closely resemble the observed VPD states, and the subsequent model evolutions form the basis of the forecast ensembles. To ensure a direct and fair comparison, we set the model-analog ensemble size to 15 to match SPEAR's.

Figure 3 illustrates the AC skill of the monthly VPD anomalies based on the model-analog technique. It is important to note that, as mentioned in Ding et al. [33], the lead-0 skill depicted in Fig. 3 does not measure forecast skill. Instead, it evaluates the alignment between the initial model-analog ensemble-mean states and observations, essentially assessing how accurately observed anomalies can be reconstructed within the model space (i.e., reconstruction skill). This method enables a comprehensive evaluation of the spatiotemporal structure of model variability compared to observations. To avoid confusion, the time series formed by the initial model-analogs and the corresponding AC skill at lead 0 are referred to as the reconstructed time
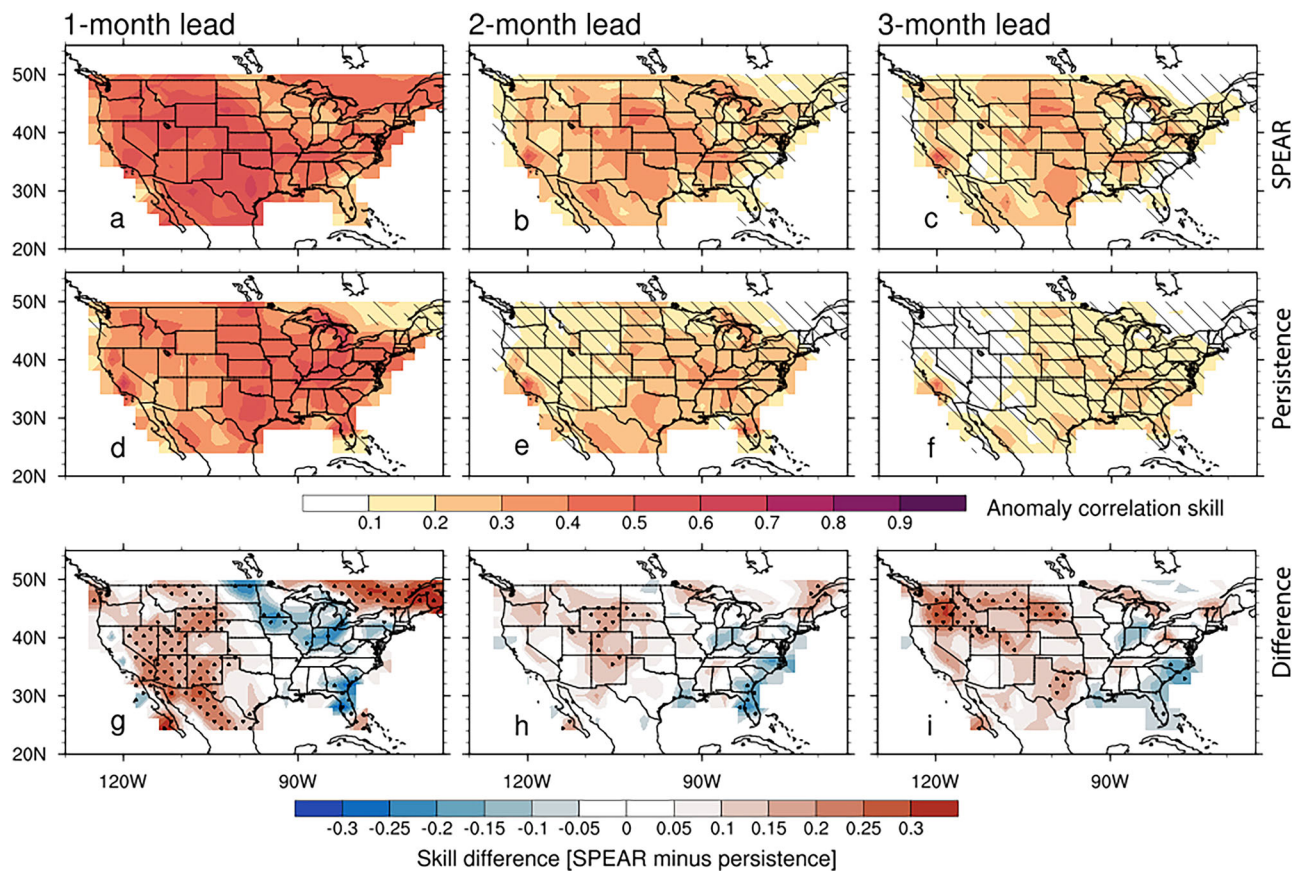
Fig. 2 | Anomaly correlation (AC) skill maps of monthly vapor pressure deficit (VPD) anomalies at the first three-month forecast leads. Forecasts are generated using (**a**–**c**) the SPEAR seasonal forecast system, and (**d**–**f**) persistence method. AC skill is computed based on the entire 1992–2022 period using verification datasets from ERA5 reanalysis. The SPEAR seasonal forecast system[19,21] consists of 15 ensemble members. Persistence forecasts assume the forecasted values are the same as its initial value. Hatching in panels **a**–**f** indicates correlation coefficients that are not statistically significant at the 95% confidence level, following ref. 63. Stippling in panels **g**–**i** indicates the forecast skill differences are statistically significant at the 95% confidence level following the Fisher Z-transformation test.

series and reconstruction skill, respectively. The interested reader could refer to Fig. 7 in Ding et al. [22] to understand how to ensure a fair comparison between operational numerical models and statistical models.

We can see that the reconstructed VPD time series closely resemble the verification data, with a temporal correlation of 0.98 for the period from 1941-2022 and 0.99 for the recent verification period of 1992-2022 (Fig. 3A), indicating the selected initial model-analogs can well capture the observed VPD features. Despite the high reconstruction skill (Fig. 3B) at month 0, the forecast skill quickly dropped to around 0.45 after the first month of evolution and continues to decline with the increased forecast leads. Overall, the VPD-only model-analog forecast skill varies across different CMIP6 models. Also, it is worth noting that the model-analog VPD forecast skill is lower than that generated by the SPEAR seasonal forecast system (black curve in Fig. 3C) and is comparable to the persistence forecasts (dashed curve in Fig. 3C). This is within expectation primarily because VPD variability is the only predictor in the initial model-analog selection (i.e., VPD-only experiment), which might lack longer-term memory on seasonal scales. In the next step, we will demonstrate that the model-analog technique becomes useful in identifying the potential sources of predictability and in addressing the question of the relative contribution of SST signal to VPD predictions.

The role of SST in predicting VPD anomalies
Lou et al. [15] showed that SST variability associated with tropical La Niña conditions appears to be an important precursor for the positive SWUS VPD anomalies. This SST signal can emerge a few months before the dry phase of SWUS VPD, providing a potential source of predictability on seasonal timescales. In this section, rather than relying solely on VPD, we

also incorporate SST information and other key variables closely related to VPD variability, such as SLP, SM in determining the initial model-analogs. This allows us to quantify the role of other climate variability in predicting VPD anomalies.

To ensure a direct comparison with the SPEAR seasonal forecast system, we utilize the SPEAR pre-industrial control (piControl) run (see more details in the data and methods section) as the data library for conducting the model-analog forecasts here. With the inclusion of the SST component, there is a general improvement in the VPD forecast skill compared to the VPD-only forecast (black curve in Fig. 4A), albeit sometimes marginally significant (Fig. 4A). This suggests that SST appears to be an important source of predictability when forecasting VPD anomalies. Interestingly, when incorporating the tropical Pacific SST, the seasonality of the VPD forecast skill (Fig. 4B) begins to resemble that of the SPEAR seasonal forecast system (i.e., Fig. 1D), with the maximum forecast skill observed during spring seasons. This is consistent with previous backward model-analog results, as demonstrated in Fig. 8 of Lou et al. [15], which shows that the SST signal persists longer when targeting the spring season. Moreover, the skill difference is readily apparent, as illustrated in Fig. 4C, where the inclusion of SST, especially SST in the tropical Pacific, significantly enhances the forecast skill. However, contrary to our expectation that SM could serve as an additional source of predictability through the reemergence mechanism[45], incorporating SM in the selection of initial model-analogs does not noticeably improve forecast skill (not shown).

Regarding the inclusion of SM, ref. 38 highlights that land states — particularly SM — can contribute to predictability in the window between weather forecasts and seasonal-to-interannual climate time scales, providing inter-seasonal memory and therefore enhancing seasonal forecasts.
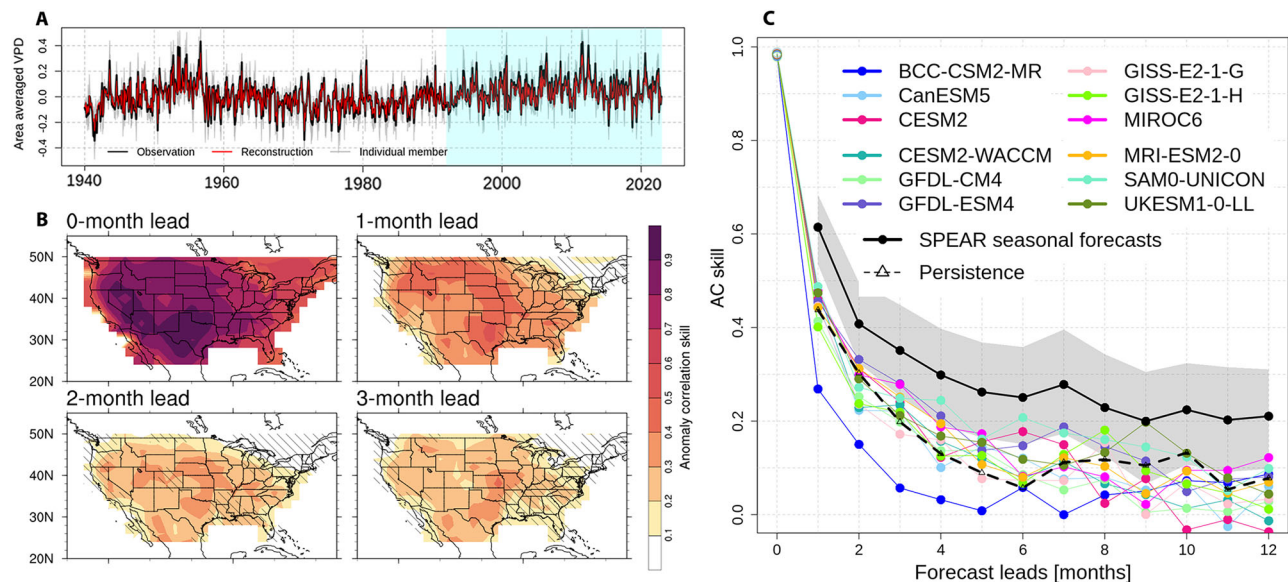
Fig. 3 | Anomaly correlation (AC) skill of monthly vapor pressure deficit (VPD). **A** Reconstructed VPD time series in the Southwest US (120-90°W, 24-40°N) at the initial month (lag=0) when model-analogs are selected. The blue shading indicates the verification period of 1922-2022. A dozen CMIP6 models constitute the model-analog library (see supplementary materials for details). The top 15 analogs from each model are selected as the ensemble members. **B** AC skill maps for the monthly VPD anomalies at forecast leads of 0-3 months, respectively. Hatching indicates correlation coefficients that are not statistically significant at the 95% confidence level, following ref. 63. The top 15 analogs from each model are selected, resulting in a total of 180 model-analog ensemble members (15 analogs × 12 libraries). The ensemble mean forecast is then obtained by computing the population mean of these 180 ensemble members. The AC skill maps in **B** represent the anomaly correlations between the model-analog ensemble-mean forecasts and ERA5 reanalysis over the period of 1992–2022. **C** AC forecast skill for the area-averaged VPD in the southwestern US. Each colored curve represents a forecast skill trajectory from one CMIP6 model, with one dot per forecast lead month. Colored open circles represent the reconstructed skill at forecast lead 0. For comparison, the AC skill generated by SPEAR seasonal forecasts is shown in black solid curve at leads of 1-12 months, and skill generated by the persistence forecasts is shown in black dashed curve with open triangles.

However, the lack of improvement in model-analog forecast skill when considering SM may be due to larger initial uncertainties in selecting SM model-analogs (Supplementary Fig. 2). In the western US, the average reconstruction skill for SM only reaches 0.69, which is much lower than that of other variables (for example, the VPD reconstruction skill shown in Fig. 3). This suggests that there may be an insufficient number of high-quality initial analogs that closely match the observed SM states. How different predictors might impact SM prediction skills is beyond the scope of this study and may be addressed in future work.

Why is the traditional model-analog forecast not as skillful as SPEAR?

So far, we have conducted VPD-only and traditional equally weighted model-analog experiments. When it comes to predicting SST variability such as ENSO, the traditional equally-weighted model-analog technique often yields comparable, and to some extent, even better forecast skill compared to state-of-the-art general circulation models[22,23,35]. However, despite the improvement in model-analog VPD forecasts with the inclusion of SST (i.e., Fig. 4), the overall forecast skill remains lower than that of the SPEAR seasonal forecast (Fig. 4B vs. Figure 1D). Potential explanations could include: 1) the library datasets from which the initial model-analogs are drawn have their own preferred states that may be biased in capturing the observed VPD characteristics adequately; 2) different forcing scenarios in the models, such as piControl and historical forcing, might impact the evolution of internal VPD variability and therefore the subsequent forecast skill; and 3) Atmospheric variability is inherently chaotic, exhibiting a higher degree of freedom and greater sensitivity to initial conditions compared to model-analog ENSO predictions.

To test these hypotheses, we started addressing the model biases. Here, we conducted a perfect-model (a model whose forecasts have no conditional or unconditional bias[22]) experiment (see data and methods section for details) to eliminate uncertainties arising from both the model structure and the observations. Since the training and verification datasets both come from the model-preferred states, the perfect-model forecast skill provides an estimate of the potential predictability of each variable.

We can see that the forecast skill improves when the tropical SST is included in the perfect model context (Fig. 5a), reaffirming the role of SST in predicting VPD variability on seasonal time scales. However, the overall forecast skill of the VPD anomalies is lower than that generated by the SPEAR hindcast (i.e., Fig. 1C). This does not necessarily imply that the model-analog forecast is not as good as the SPEAR seasonal forecast system, because the sample sizes vary dramatically from 31 years (SPEAR seasonal forecast system) to 2700 years (90 years*30 SPEAR-LE members) in the perfect-model experiment. We resampled the 2700-year verification period into different 31-year chunks, and the SPEAR seasonal forecast falls within the range of the bootstrapped skill generated by the perfect model (not shown), reflecting the influence of sampling sizes. The perfect-model experiment here clearly demonstrates that the predictability of VPD variability is fairly limited even in the model-preferred states. In contrast, ENSO-related forecast skill (Fig. 5b) remains skillful throughout the forecast leads with only SST being the predictor, consistent with previous perfect-model ENSO experiment[22]. In conclusion, even in the absence of observation and model uncertainties, the potential predictability of VPD variability diminishes rapidly as forecast leads increase, emphasizing the chaotic nature of atmospheric conditions.

The second hypothesis stresses that different external forcing scenarios might impact the evolution of VPD variability and therefore the subsequent forecast skill. To test this, we conduct the same model-analog hindcast experiment, but with historical simulations from the SPEAR-LE as the library (the detailed experiment design can be found in the 'data and methods' section). Firstly, it is evident that the inclusion of SST improves VPD prediction regardless of which library datasets are used (Fig. 6A). However, the influence of different forcing scenarios (i.e., piControl vs. historical) on forecast skill is not coherent. That is, some lead times and initial months show improvement, while others exhibit a decrease in skill (Fig. 6B). There is no overall skill improvement when comparing different
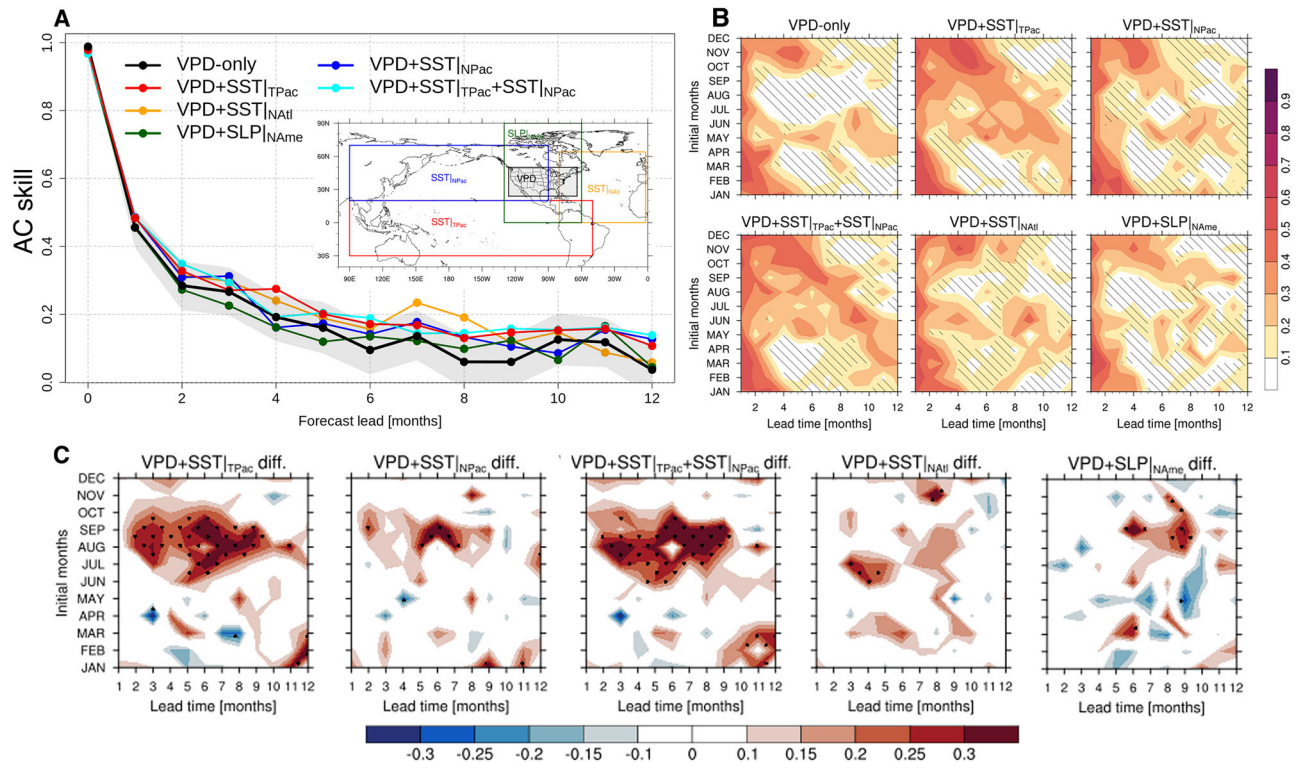
Fig. 4 | Anomaly correlation (AC) skill of monthly vapor pressure deficit (VPD) using a model-analog technique. A AC skill of the monthly area-averaged VPD anomalies in the Southwest US (120-90°W, 24-40°N). Initial model-analogs are selected based on VPD only, tropical Pacific SST, North Pacific SST, North Atlantic SST, and North American SLP, respectively. The grey shading indicates the 5-95% confidence interval of the VPD-only forecast skill. The sub-panel indicates the regions used to select the initial model-analogs. The top 15 analogs from the SPEAR piControl simulation (3000 years) are selected as the ensemble members. **B** Ensemble-mean AC skill for area-averaged VPD in the southwest US as a function of initialization month (vertical axis) and forecast lead times up to 12 months (horizontal axis). The model-analog forecasts were generated by incorporating different predictors related to VPD, tropical Pacific SST, North Pacific SST, North Atlantic SST, and North American SLP, respectively. Hatching indicates correlation coefficients that are not statistically significant at the 95% confidence level, following ref. 63. **C** The AC skill difference is calculated by subtracting each model-analog experiment from the baseline VPD-only experiment. Stipples indicate the skill difference is statistically significant at the 95% confidence level.
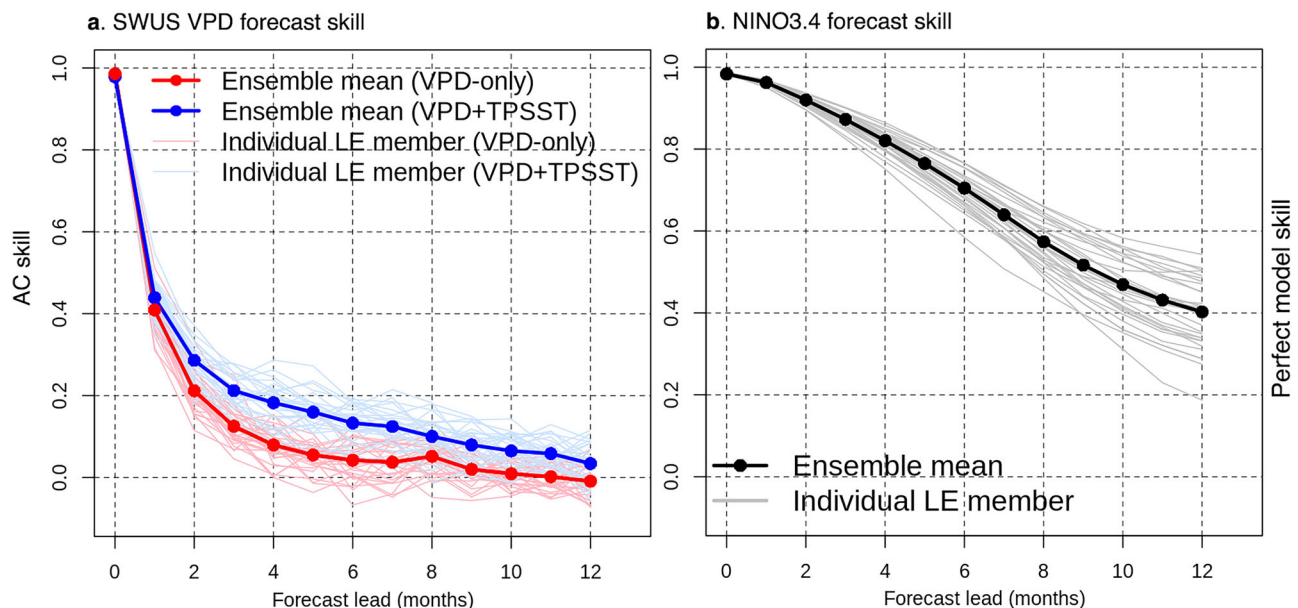

Fig. 5 | Perfect-model anomaly correlation (AC) forecast skill. a AC skill of the SWUS VPD time series generated by the perfect-model experiment (see the text for details). Red curves represent the forecast skill using VPD-only, while blue curves represent the inclusion of tropical SST in addition to VPD. **b** AC skill of the NINO3.4 time series generated by perfect-model experiment. Thin curves represent individual SPEAR-LE members used as the verification, while thick lines represent the ensemble mean.
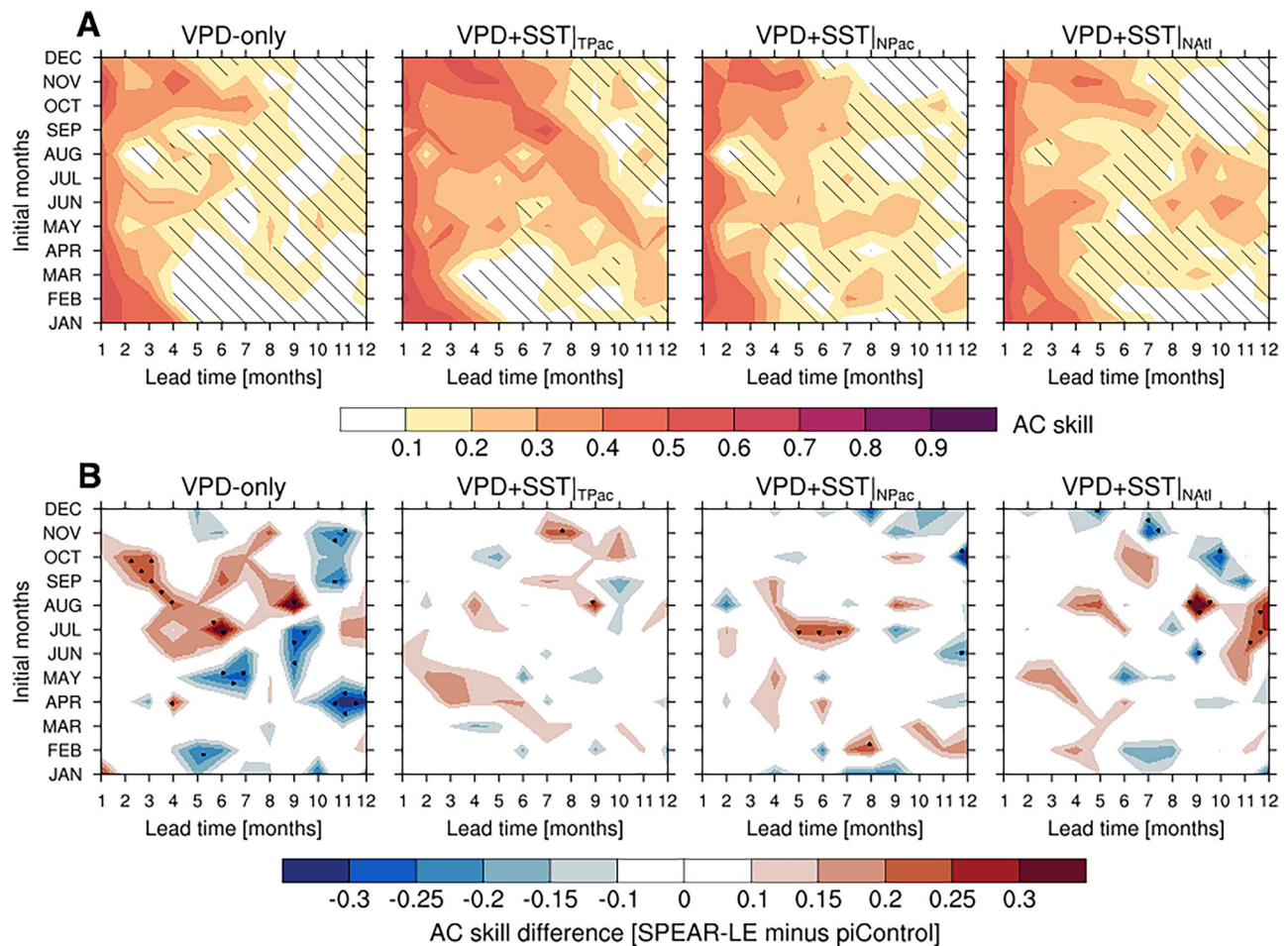
Fig. 6 | Anomaly correlation skill of the area-averaged VPD in the southwest US. **A** same as Fig. 4B, but using historical runs taken from SPEAR large ensembles (LE) as the data library to search for the initial model-analogs. **B** The skill difference between the SPEAR-LE and SPEAR piControl. The stipple indicates the difference is statistically significant ($p < 0.05$) following Fisher Z-transformation test.

forcing scenarios. This suggests that the time-evolving external forcing might not dramatically alter the overall evolution of VPD variability and, consequently, the forecast skill.

Lastly, atmospheric variability is inherently chaotic, potentially exhibiting greater sensitivity to initial conditions relative to its oceanic counterpart. To investigate this hypothesis, we estimated the signal-to-noise ratio (SNR) and examined the respective signal and noise components. First, the theoretical relationship between the AC skill and the SNR[46–49] can be simply expressed as $AC = s/\sqrt{1 + s^2}$ with s denoting SNR. The curve in Fig. 7a displays how the expected forecast skill increases nonlinearly as the SNR increases if the ensemble size is infinite. The expected AC skill approaches unity for a larger signal relative to the spread.

We further decompose the SNR into the corresponding noise and signal components by showing the violin plots in Fig. 7b, c. Figure 7b compares the ensemble spread for two variables (SWUS VPD and NINO3.4) across two sets of experiments: Model-analog and SPEAR seasonal forecast. Here, the model-analog hindcast is taken from VPD and the tropical Pacific SST experiment as an example (i.e., red curve in Fig. 4A). Note that the results are not sensitive to which model-analog experiment we examined (not shown).

Here, the forecast spread and signal in Fig. 7b, c are shown at a one-month forecast lead, which we use as a proxy for the initial uncertainties or, generally, uncertainties that are sensitive to the initial conditions. This approximation is necessary because the SPEAR seasonal forecast system is initialized on the first day of each month. The so-called month-0.5[22] or month-0[20] forecast represents the mean of the

first forecast month, effectively centered in the middle of the calendar month (see Fig. 7 in Ding et al.[22] for reference). In contrast, the equivalent model-analog forecast is initialized using the monthly mean observations centered on the previous month (i.e., month 0). This alignment ensures that the 1-month-lead model-analog forecast and the SPEAR month-0.5 forecast are verified at the same time. For consistency, we refer to both as the month-1 forecast, with subsequent lead times following the same naming convention.

Before examining Fig. 7 in detail, we first highlight the overall skill comparison between the model-analog forecasts and the SPEAR seasonal forecast at a one-month lead. The model-analog forecast skill for VPD at this lead time is generally lower than that of the SPEAR seasonal system regardless of the model-analog experiment (i.e., VPD-only, equally weighted, and optimally weighted experiments) used. In contrast, the SST-based NINO3.4 model-analog forecast skill is comparable to that of the numerical system at this forecast lead (not shown). We can see that the VPD forecasts show a much higher spread than in the ENSO predictions (Fig. 7b), particularly notable in the broader and higher distribution, reflecting the chaotic nature of the atmosphere. Despite exhibiting a very similar spread distribution in the VPD forecasts (Fig. 7b), the SPEAR model-analog shows a smaller signal component compared to the SPEAR seasonal forecasts (indicated by the thick middle lines in Fig. 7c). This results in a lower SNR and consequently reduced AC skill in the model-analog forecast. In contrast, the ENSO forecast signal is fairly comparable in both the model-analog and SPEAR seasonal forecasts (Fig. 7c). Despite having a larger spread in the model-analog experiment, the ENSO forecast skill turns out to be
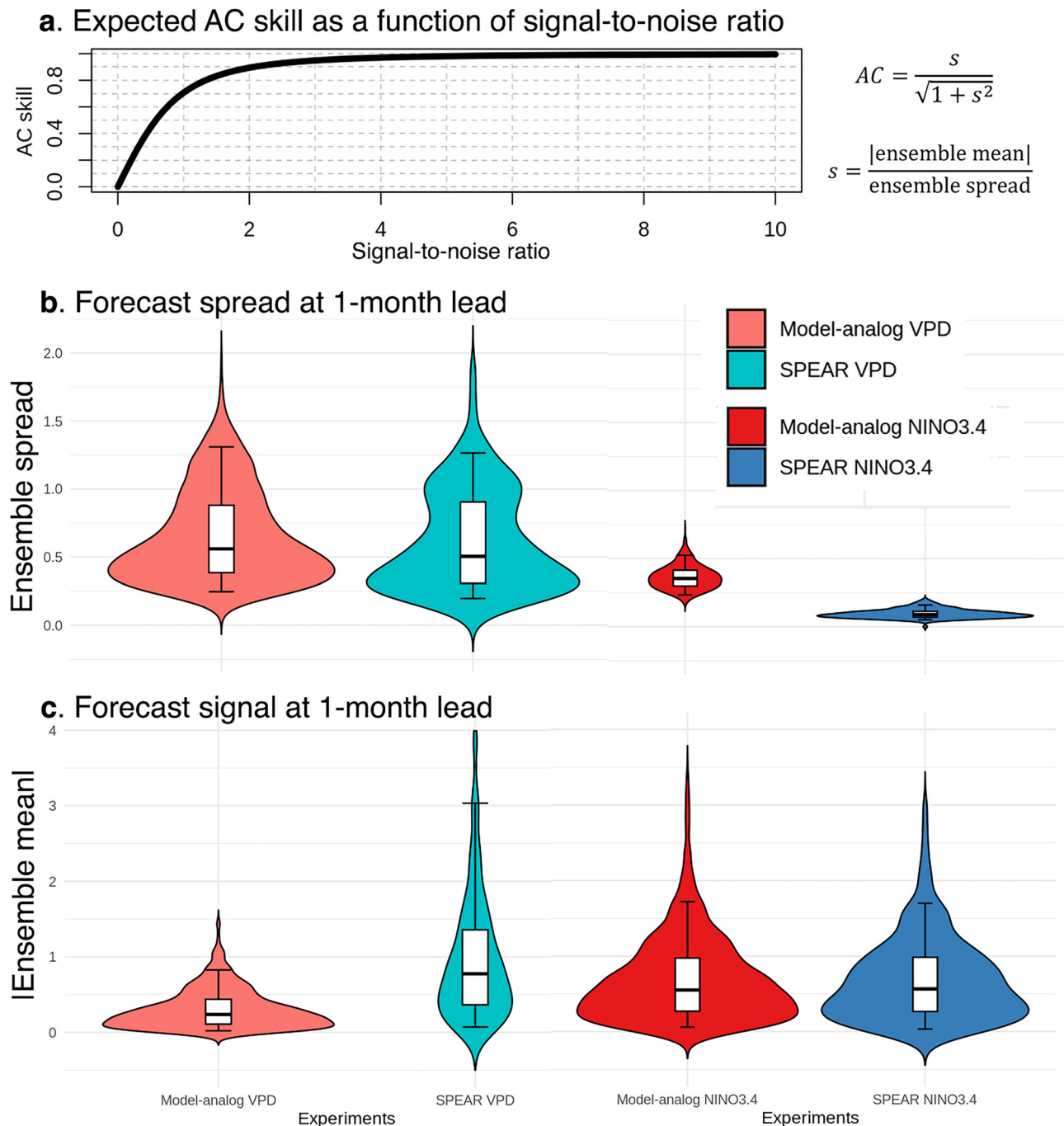
Fig. 7 | Estimation of anomaly correlation (AC) forecast skill and signal-to-noise ratio (SNR). a Theoretical relationship between AC skill and SNR when ensemble size is infinite. **b** and (**c**) violin plots for the distribution of forecast noise (ensemble spread; panel (**b**) and signal (ensemble mean; panel (**c**) at the one-month forecast lead for the SWUS VPD and NINO3.4 time series based on SPEAR model-analog experiment and SPEAR seasonal forecast. The noise and signal are normalized based on the standard deviation of the corresponding verifications. The central line in the middle of the violin plot represents the median of the data. The box plots embedded within the violin plots represent the range from the first quartile (Q1) to the third quartile (Q3) of the data. This range, known as the interquartile range (IQR), contains the middle 50% of the data points, with the middle thick line indicating the median and the whiskers showing the 5-95% intervals. The shape of the violin plots shows how the data is distributed. The width of the violin plots in panels (**b**) and (**c**) indicates the density of the predictand. Wider sections represent higher density, while narrower sections indicate lower density.

comparable (not shown) due to a much stronger signal outperforming the respective noise.

Figure 7 illustrates that statistical methods, including model-analog approaches, rely on pre-existing analogs and empirical relationships rather than explicitly solving the governing physical equations. As a result, they are more sensitive to noise in the initial conditions and less effective at capturing dynamically consistent responses, leading to a weaker signal (Fig. 7c). In

contrast, numerical models integrate forward from well-defined initial conditions, allowing them to resolve the growth of predictable signals, particularly in the first month when initial conditions exert a strong influence. Even when selecting the best-matched analogs, statistical methods such as the model-analog technique inherently smooth over a range of potential initial conditions, further diluting the signal—particularly when predicting complex atmospheric variables. Interestingly, these constraints
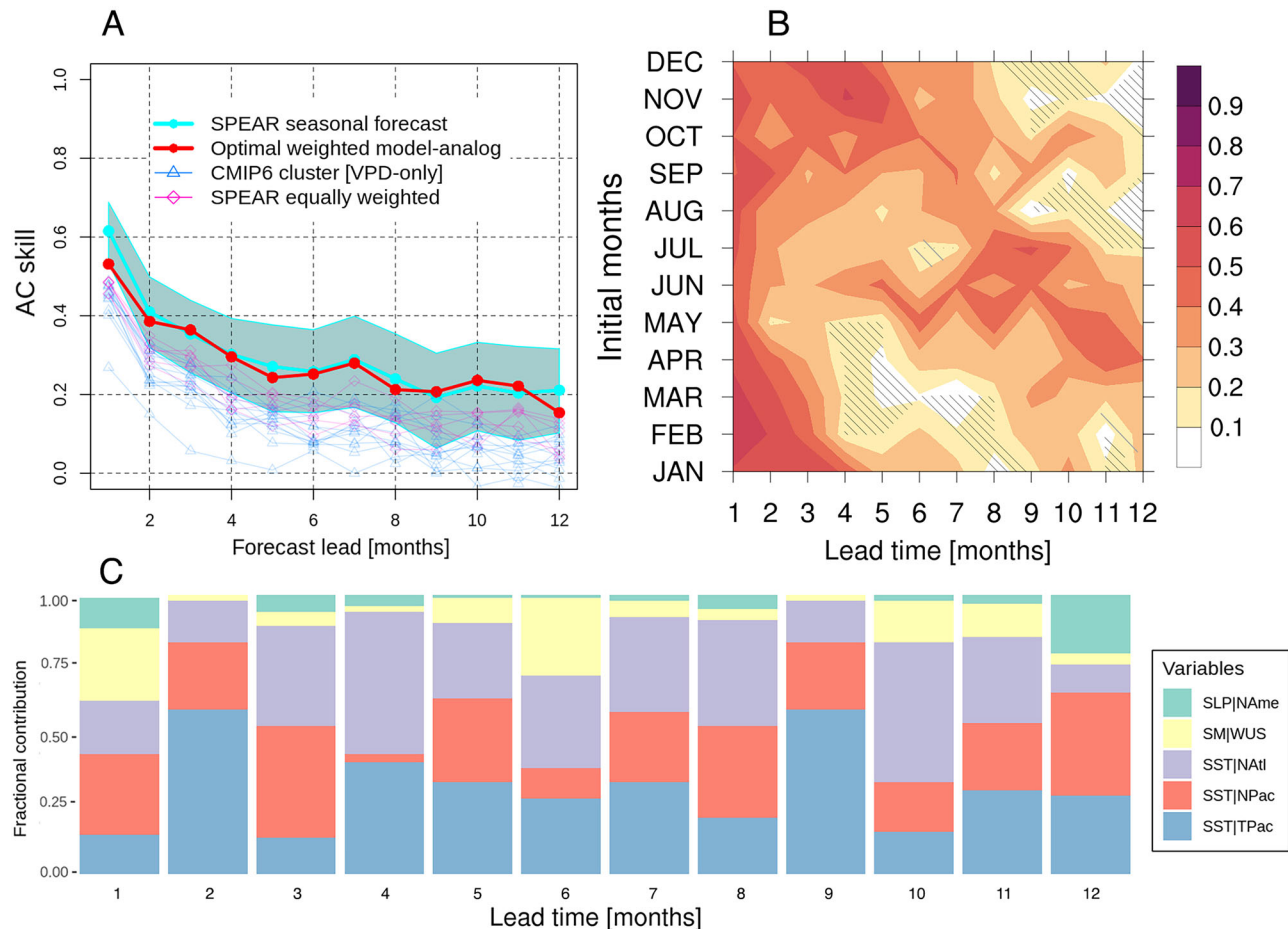
Fig. 8 | Anomaly correlation (AC) skill of the weighted model-analog hindcasts.
A AC skill of the VPD anomalies in the southwest US generated by the SPEAR seasonal forecast system (cyan), and weighted model-analog (red). For comparison, the equally weighted model-analog skill is shown in purple (same as in Fig. 4A), and VPD-only model-analog skill is in light blue (same as in Fig. 3C). The cyan shading denotes 5-95% confidence interval of SPEAR seasonal forecast system. **B** AC skill for area-averaged VPD in the southwest US as a function of initialization month (vertical axis) and forecast lead times up to 12 months (horizontal axis). The forecast skill here is generated using weighted model-analog experiment. Hatching indicates the AC skill is not significant at 95% confidence level. **C** Fractional weights showing the relative contribution arising from each predictor in forecasting the area-averaged VPD variability in the southwest US (the detailed description can be found in the text).

have a limited impact on ENSO predictions, primarily because ENSO exhibits low-dimensional variability[22], where deterministic processes play a dominant role over stochastic influences. As a result, both the numerical and statistical models can effectively capture its evolution, leading to relatively robust and comparable predictions despite the inherent uncertainties.

It is worth noting that, at short lead times (i.e., 1-month lead), numerical models benefit from precise initial conditions, allowing them to resolve predictable signals effectively. However, as lead time increases, predictability is increasingly governed by low-frequency climate modes (e.g., ENSO, PDO) and boundary forcings (e.g., SST anomalies). Analog-based methods can capture these slow-evolving features just as well as, if not better than, numerical models, especially when the models struggle with long-term drift and biases. This aspect will be further discussed in the following section.

How to improve the model-analog forecast?
Recently, instead of solely relying on root mean square distance (RMSD) to determine the initial model-analog, some machine learning-based techniques have been employed to optimize the selection process, by incorporating, for example, linear inverse model[50], convolutional neural network[51], and interpretable neural network[52] into the initial model-analog determination. Generally, instead of using traditional equally weighted model-analogs, the idea of these methods is to find a spatially weighted mask that quantifies how important each grid point is for determining whether two

climate states (or predictors and predictands) will evolve similarly. To achieve this, we randomly perturbed the weight parameters, as shown in Eq. (4), 1000 times to reflect the fractional contribution arising from each variable, associated with SST, SLP, and SM. Subsequently, we trained the parameters on a shifting 30-year training period. After determining the optimal weight parameters at each window, we averaged the weight mask and applied a weighted model-analog to the separate verification period. Notably, we also implement two additional perturbation strategies (see data and methods for details): increasing the perturbations to 5,000 iterations and training the parameters over the full 1940–1991 period. These modifications do not notably impact the results.

The optimal model-analog AC skill at a one-month forecast lead is 0.07 lower than that of the SPEAR seasonal forecast (0.54 vs. 0.61), though this difference is not statistically significant at the 95% confidence level. Potential reasons for this discrepancy were discussed in the previous section. Nevertheless, it is noteworthy that the overall forecast skill remains comparable, as shown by the red and cyan curves in Fig. 8A. Meanwhile, when comparing the VPD-only and equally weighted model-analog experiments as shown in Figs. 3 and 4, weighted model-analogs and SPEAR seasonal forecast system generally have better forecast skill. Although there is a slightly higher forecast skill across various initial months (Fig. 8B), it remains consistent to see that the minimum skill occurs when targeting the late summer and early fall seasons, particularly when the forecasts are

initialized in the first half of the year. This forecast skill reemergence has also been reported by Lou et al. [23] in ENSO predictions. Specifically, their forecast initialized in late summer will experience skill that first slowly declines, then declines more rapidly at lead times of approximately 6–9 months. Subsequently, the skill increases with further lead time, plateauing at about 15 months before decreasing again for longer lead times. Despite having shorter periods of skillful forecasts, the minimum VPD forecast skill aligns with the seasonality of ENSO predictions and reflects temporarily lower SNR during the boreal summer.

Figure 8C displays the averaged weight parameters throughout the moving 30-year training period that used to determine the optimal AC skill. The composition of contributing variables varies across different lead times, indicating dynamic interactions among these factors and their influence on forecast skill. Remarkably, no single predictor completely dominates for any forecast lead, reflecting the complexity of the predicted system. Additionally, sampling errors may also contribute to the observed variability in weight distributions, as the limited training sample size can introduce fluctuations in parameter estimation. These uncertainties further emphasize the challenges in identifying a consistently dominant predictor across different forecast leads. However, it is noteworthy that, compared to SLP and SM, SST contributes a large portion to VPD forecast skill, reaffirming the crucial role of SST in seasonal VPD predictability. Although this study does not explicitly discuss the role of Atlantic SST, it shows a consistent impact on VPD predictions (Fig. 8C), in agreement with previous literature[17,53,54]. The implication of the weighted model-analog here is that if we can utilize more sophisticated tools, such as machine learning, to better identify and select the initial model-analogs, and to better train the cost functions, it is likely that the VPD forecast skill can be further improved accordingly.

Discussion and conclusion

Predicting fire activity on seasonal time scales is challenging due to the complexity of the processes involved, limitations in observational data, and the compounding effects of multiple concurrent factors such as ignitions, fire weather conditions, and vegetation[55]. While predicting individual fire events remains in its infancy, forecasting fire-prone climatic conditions is feasible assuming climatic processes act as top-down controls on the regional patterns of seasonal changes in fire behavior[56]. Previous studies[6,12,57] have identified robust relationships between atmospheric VPD and fire activity, indicating that higher VPD corresponds well with larger burned areas in the western US. This study aims to investigate the seasonal predictions and predictability of monthly VPD anomalies in the western US. It seeks to quantify the seasonal forecast skill of monthly VPD variability, and to assess the contributions of SST and other variables to predicting VPD variations in the southwest US.

Building upon previous work (Lou et al. [15]), which used a backward model-analog technique to emphasize the critical role of La Niña as a precursor to positive VPD conditions in the SWUS, this study applies a forward model-analog technique. Here, instead of tracing backward in time to identify potential precursors, we move forward in time to quantify how quickly the predictable signal dissipates in the spread of forecast noise given a set of initial conditions that closely resemble the observed states.

Through a series of model-analog experiments (Figs. 3–8), we demonstrated that the weighted model-analog approach is comparable to the state-of-the-art numerical model (Fig. 8). Both weighted model-analog and SPEAR seasonal forecast outperform the persistence forecasts and equally weighted model-analog, which in turn outperforms the VPD-only model-analog. Despite some skill differences, it is consistent across all model-analog experiments that SST in the tropical Pacific is a critical source of predictability for forecasting VPD variability in the SWUS on seasonal time scales. Meanwhile, the intrinsic predictability arising from atmospheric VPD variability is limited in both the real world and perfect-model experiments (Fig. 5). The implication of the weighted model-analog showing better forecast skill is that employing more sophisticated model-analog determination methods, which include better calibration and selection of predictors and initial model-analogs, can further enhance

forecast skillfulness. For example, one could consider incorporating sub-surface oceanic processes related to variables such as ocean heat content or bottom temperature to refine the cost functions (i.e., predictors). Alternatively, employing machine learning frameworks[52] might improve the assignment of weighting parameters.

Regarding the seasonality of forecast skill, both the numerical model and empirical model suggest that VPD variability in the late summer and early fall is the least predictable (Figs. 1 and 8). However, regardless of the initial months, the forecast skill of VPD is likely to recover in the subsequent winter and spring seasons, highlighting the inherent challenges in predicting VPD anomalies during warmer seasons. This is partially due to the seasonal cycle of ENSO and particularly due to the intrinsic predictable signal being smaller than its noise in those seasons.

Lastly, we demonstrate that even the weighted model-analog approach exhibits slightly lower forecast skill for VPD variability compared to its numerical counterpart, particularly at the 1-month forecast lead (Fig. 8). Upon examining the SNR, we showed that despite having a similar noise distribution (Fig. 7), the model-analog method underestimates the predictable signal. This underestimation of SNR subsequently leads to lower forecast skill in predicting atmospheric VPD variability. In contrast, the wider forecast spread in ENSO model-analogs does not appear to notably affect their forecast skill. This disparity arises because the predictable ENSO signal predominates in determining SNRs, with noise playing a secondary role in ENSO model-analog predictions. The slightly lower VPD forecast skill at the 1-month forecast lead can be improved by, for example, employing more advanced machine learning tools to better identify the initial model-analogs or by incorporating additional predictors that can provide a more comprehensive description of VPD dynamics.

Finally, while VPD is a critical indicator of wildfire risk and exhibits seasonal forecast skill, its predictability does not necessarily translate directly into fire predictability due to several key limitations. First, vegetation availability plays a crucial role in fire occurrence. Although high VPD indicates atmospheric dryness, the presence of sufficient dry fuel is necessary for wildfires to develop. For instance, a wet winter may limit fuel buildup, reducing fire risk even when VPD is high. Second, wildfires require ignition sources, whether from human activity (e.g., campfires, power lines) or natural causes (e.g., dry lightning). Since VPD does not account for ignition factors, fire predictions remain inherently uncertain. Finally, wind and other meteorological conditions remarkably influence fire behavior. Even under high VPD conditions, strong winds can drive rapid fire spread, while calm conditions may suppress fire growth. Given the complex interplay of these factors, future studies will further explore the role of wind and other fire-prone conditions in improving fire predictability.

Data and Methods
Observed and simulated Data

We utilize monthly data of 2-meter dewpoint temperature and 2-meter air temperature sourced from the ECMWF Reanalysis v5 (ERA5) reanalysis[58] to calculate VPD. Furthermore, we investigate oceanic and atmospheric conditions from the ERA5 reanalysis, incorporating monthly SST, SLP, and vertically averaged SM down to 100 centimeters from the land surface. All variables undergo remapping to a regular 2° x 2° grid and are area-weighted prior to analysis. Given our study's focus on comprehending fire weather conditions in the US, we apply a mask to exclude the VPD values over the ocean during VPD examination, created using SST data sourced from HadISST[59], which is similarly remapped to a 2°x2° grid. The NINO3.4 index is defined as the monthly area-averaged SST anomalies in 5°S-5°N, 170°-120°W. The PDO index is defined in the same manner but in the North Pacific domain (25°N–45°N, 140°E–145°W; i.e., Region 1 in Henley et al. [60]).

To conduct the forward model-analog hindcasts, we utilize SPEAR pre-industrial control (piControl) experiment and the historical large ensembles (Delworth et al. [19]). SPEAR is a newly developed coupled general circulation model (CGCM) created by the GFDL. We use both the 3000-year piControl run and 30 ensemble members of historical simulations that cover 1921–2010 as the model library to search for the model-analogs (the

Article

initial states that closely resemble the observed target states). This CGCM incorporates all radiative forcings and land cover from both anthropogenic and natural sources, including greenhouse gases, aerosols, ozone, solar irradiance, and volcanic aerosols. In addition to the monthly SST, SLP, and SM at different depths ranging down to 100 centimeters, monthly outputs of 2-meter relative humidity, and 2-meter temperature are used to compute VPD. All model outputs are remapped to a 2°x2° grid and latitudinally weighted. More details regarding the SPEAR-LE simulations can be found in Delworth et al. [19].

The model-analog hindcast is also performed using a dozen Coupled Model Intercomparison Project Phase 6 (CMIP6) models, identical to those utilized in Lou et al. (2024). These CMIP6 datasets encompass monthly 2-meter air temperature and relative humidity derived from the piControl experiments conducted across 12 CMIP6 climate models (see Supplementary Table 1 for detailed specifications). The pre-industrial CMIP6 forcings encompass repeating seasonal cycles, including CO2 and other well-mixed greenhouse gases, solar irradiance, ozone, aerosols, and land use[61,62]. Linear detrending is applied to eliminate potential climate drift, and all datasets are remapped onto a regular 2° longitude by 2° latitude grid prior to analysis.

The significance of the correlation coefficients r was determined using the method outlined by Davis[63]. This method adjusts for the effective number of degrees of freedom resulting from serial correlation. A basic t-statistic was then employed to evaluate the statistical significance of the correlations and compute the critical r values between the two time series. The bootstrapping test is employed to obtain the 5-95% confidence intervals of the forecast skill, using 1000 replications.

SPEAR seasonal forecast system

We use the SPEAR seasonal forecast system[19,21] in this study. SPEAR represents the latest generation of GFDL modeling system designed for seasonal to multidecadal prediction and projection. This coupled ocean-atmosphere-land-sea ice model operates at medium resolution, with an atmospheric and land resolution of 50 km and 33 atmospheric vertical levels. Detailed specifications are outlined in Delworth et al. [19] and Lu et al. [21].

In assessing the prediction skill of the SPEAR system, an extensive series of reforecasts (also known as hindcasts) was conducted. Over the period from January 1992 to December 2022, a 15-member ensemble of reforecasts was generated for each month. Each reforecast spanned a duration of 12 months and was initialized using reanalysis data from the first day of each month. In this study, we examine the forecast skill of VPD variability, ENSO and the PDO. The anomalies are computed by removing the lead-time-dependent climatology.

Estimation of VPD variability

VPD quantifies the amount of water vapor the air can hold at a specific temperature before saturation occurs. A higher VPD indicates a drier atmosphere. It can be calculated using monthly 2-meter temperature (T) and monthly 2-meter dew-point temperature (T_d) with the following equation[64]:

$$\text{VPD} = e_s(T) - e_a = c_1 * \exp\left(\frac{c_2 * T}{c_3 + T}\right) - c_1 * \exp\left(\frac{c_2 * T_d}{c_3 + T_d}\right) \quad (1)$$

where $c_1 = 0.611$ kPa, $c2 = 17.5$, and $c_3 = 240.978$ °C. The units of T and T_d are in °C, and the resulting VPD is in kPa.

In the model simulations, we compute dewpoint temperature T_d using the relative humidity (RH) and temperature (T) at surface level:

$$T_d = \frac{a_1 * (\ln\left(\frac{RH}{100}\right) + \frac{a_2 * T}{a_1 + T})}{a_2 - (\ln\left(\frac{RH}{100}\right) + \frac{a_2 * T}{a_1 + T})} \quad (2)$$

where, $a_1 = 243.04$, $a_2 = 17.625$, respectively. After obtaining T_d, the simulated VPD is then computed by applying Eq. (1).

Model-analog technique

As documented in Ding et al. [22], the root-mean-square distance (RMSD) is used as a metric to determine the model-analogs. Specifically, the initial model-analogs at each time t are selected by minimizing the distance between the target state $x(t)$ and each library state $y(t')$. The target state is defined as the state at the initialization time, while the library consists of all pre-existing model states obtained from a general circulation model. The RMSD can be established using,

$$d(t, t') = \sum_{i=1}^{I} \frac{1}{J_i} \sqrt{\sum_{j=1}^{J_i} \left[\frac{x_j^i(t)}{\sigma_X^i} - \frac{y_j^i(t')}{\sigma_Y^i}\right]^2} \quad (3)$$

Here, $d(t,t')$ represents the RMSD calculated when comparing the observed state at time t with the simulated state at time t', The index i corresponds to a variable, with I representing the total number of variables, while j refers to a spatial degree of freedom (such as numbers of gridpoints) within the training region, with J_i denoting the total number of spatial locations for the i-th variable. Note that each variable (i.e., $x_j^i(t)$ and $y_j^i(t')$) here is normalized by its own domain-averaged standard deviation (i.e., σ_X^i and σ_Y^i) to equally weigh the variables. In this context, equal weighing means all the standardized variables contribute equally to the determination of the initial model-analogs. To account for seasonality, library states are restricted to the same calendar month as the target state.

The domain-averaged RMSD values are then sorted in ascending order, and the K best simulated states with lowest RMSD are chosen as the ensemble of initial states, indicated by the set $\{y(t'_1), y(t'_2), \ldots, y(t'_k), \ldots, y(t'_K)\}$ with k the model-analog index and t'_k the time stamp of this analog in the library. The subsequent evolution of this ensemble within the corresponding simulation, $\{y(t'_1 + \tau), y(t'_2 + \tau), \ldots, y(t'_k + \tau), \ldots, y(t'_K + \tau)\}$, forms the model-analog forecast ensemble for $x(t + \tau)$ at lead time τ months. For data libraries on the order of several hundred years in length, an ensemble size of $K \sim 10$–20 was found to give the optimal results[22]. Here, we simply use $K = 15$ to be consistent with the SPEAR seasonal forecast system.

Equally weighted model-analog forecasting approaches as shown in Eq. (3) have proven successful in predicting ENSO variability[22,23]. Specifically, SST and sea surface height (SSH) anomalies can yield ENSO forecast skill comparable to cutting-edge numerical models (e.g., Fig. 1 in Lou et al. [23]). The success in predicting ENSO is primarily due to the low dimensionality of ENSO variability, as SST and SSH are sufficient to capture most of its evolution. This raises the research questions addressed in this study: Can a similar approach be applied to predict much more chaotic atmospheric variability? And would the traditional equally weighted method be adequate for forecasting such a complex system?

Building on the traditional equally weighted model-analog approaches, we further add a weighting parameter w_i in this study to reflect the relative importance of different contributing factors as follows,

$$d(t, t') = \underbrace{\frac{1}{J} \sqrt{\sum_{j=1}^{J} \left[\frac{x_j(t)}{\sigma_X} - \frac{y_j(t')}{\sigma_Y}\right]^2}}_{\text{term1}} + \underbrace{\sum_{i=1}^{I} w_i \bullet \frac{1}{J_i} \sqrt{\sum_{j=1}^{J_i} \left[\frac{x_j^i(t)}{\sigma_X^i} - \frac{y_j^i(t')}{\sigma_Y^i}\right]^2}}_{\text{term2}}$$

$$(4)$$

The total RMSD on the left is determined by two components. Term 1 reflects the RMSD of the targeted predictand, which in this study is the monthly VPD anomalies. The second term represents the summation of RMSD from all other contributing factors, weighted by the parameter w_i. In this study, the contributing factors include SST in tropical Pacific [30°S-20°N, 90°E-50°W], North Pacific [20-70°N, 90°E-90°W], and North Atlantic [0°-64°N,80°W-0°], SLP in North America [0°-90°N, 60°-130°W], and vertically averaged SM in the western US [24°-50°N, 100°-126°W]. These hotspots are visualized in Fig. 4 for reference. The index I represents the total number of contributing factors ($I = 5$ in this case).

When $w_i = 0$, the RMSD is determined solely by term1 in Eq. (4), which is referred to as VPD-only model-analog experiments in this study. When w_i is a binary index (either 0 or 1), the approach simplifies to the traditional equally weighted model-analog forecasting as shown in Eq. (3). In the weighted model-analog experiment, the weighting parameter w_i is randomly perturbed with the constraint that their sum equals 1 (i.e., $\sum_{i=1}^{I} w_i = 1$), thereby reflecting the relative importance of these contributing factors in determining the initial model-analogs. Please note that weighting is positive and applied to the spatially normalized RMSD in this context. Readers are kindly reminded not to confuse this with other potential weighting approaches, such as weighting on k model-analog forecast ensembles or different library models, as these were not used in this study. Instead, the model-analog ensemble mean forecasts are derived directly from the population mean of K model-analog forecast ensembles.

To understand the role of external forcing on VPD forecast skill, we use both the piControl and historical SPEAR-LE simulations as the data library. For the historical SPEAR-LE simulations, we restrict the library to the 30-year period preceding the forecast time, ensuring that no future information —unavailable at the time of forecasting—is included. Meanwhile, we also conduct perfect-model forecast experiments based on the SPEAR-LE simulations. In this experiment, each ensemble member of the SPEAR-LE serves as the verification dataset in turn, while the remaining 29 SPEAR ensemble members constitute the data library to search for the initial model-analogs. Because both the training and verification datasets originate from the model's preferred states, the perfect-model forecast skill offers an estimation of the potential predictability of each variable[44,65]. We kindly remind the reader not to confuse the K model-analog ensemble members ($K = 15$ in this study, consistent with the SPEAR seasonal forecast system) with the ensemble members of the historical SPEAR large ensemble simulations, which consist of 30 members.

We apply cross-validation in generating weighted model-analog forecasts to ensure a fair assessment of forecast skill when comparing with the numerical model. Specifically, the observational dataset is divided into training and verification periods: the verification period spans 1992–2022, corresponding to the SPEAR seasonal forecast system, while the training period covers 1940–1991. Additionally, we implement a moving 30-year window within the training period to determine the optimal weighting parameters w_i — combinations that maximize forecast skill. The weighting parameters averaged over these 30-year windows (e.g., 1940–1969, 1941–1970, …, 1962–1991) are then used as the final weighting factors and validated against the separate verification period (1992–2022). The use of averaged weighting parameters over moving windows aims to enhance robustness while mitigating sampling errors.

We developed three bootstrapping strategies to perturb the weighting parameters. First, we perturbed the weighting parameters 1,000 and 5,000 times, respectively, and applied the resulting weighting masks to the full training period from 1940 to 1991 to determine the optimal weighting configuration. Alternatively, we perturbed the weighting parameters 1,000 times but applied the same set of weighting masks separately to each individual 30-year training window to identify the optimal weighting combination for each specific period. The averaged weighting parameters were then used. All three bootstrapping methods produced similar results. To reduce sampling errors arising from the limited training data, we adopted the latter approach.

Finally, we provide a detailed description of the procedure used to construct the weighted model-analog forecasts. This step-by-step breakdown is intended to clarify the workflow and demonstrate that no information from the verification period was used during the development of empirical model as stated in ref. 66.

Data preparation phase:
1. ERA5 anomaly calculation: Monthly ERA5 anomalies are calculated using a 30-year moving climatology, where each climatology is defined as the most recent 30-year period relative to the forecast month. For example, a forecast initialized in July 1993 uses 1964–1993 as its reference climatology. This moving window approach prevents the use

of future information that could bias the forecast evaluation, as noted in Risbey et al.[67] and Lou et al.[23].
2. Separating verification and training datasets: The ERA5 dataset is partitioned into a training period (1940–1991) and a verification period (1992–2022), with the latter overlapping the period of the SPEAR seasonal forecast system.

Training phase:
3. RMSD calculation in training period: The RMSD is computed for each variable using data from the training period, following Eq. (3).
4. Generation of random weighting parameters: A set of 1,000 random weighting parameters is generated, constrained such that the weights sum to 1. These weights are applied to the RMSD values from Step 3, following Eq. (4).
5. Model-analog forecasts (training period): The weighted RMSD values obtained in Step 4 (or Eq. 4) are sorted in ascending order. For each of the 1,000 perturbations (i.e., each set of weights), this yields a set of distance values $d(t,t')$. From each set, the top K analogs with the smallest RMSD values are selected as follows, $d(t,k) = \min_{t'} d(t,t')$, where $k = 1, 2, …K$. For each perturbation, ensemble model-analog forecasts are generated by extracting the subsequent evolution of the selected K analog members from the training period (1940–1991).
6. AC skill computation (training period): The K-member ensemble-mean AC forecast skill is computed for each perturbation. To ensure robustness, the training period is divided into 30-year moving windows (e.g., 1940–1969, 1941–1970, …, 1962–1991), and AC skill is computed separately within each window. This captures temporal variations in forecast skill (Lou et al.[23]).
7. Selection of optimal weighting configuration: For each 30-year window and forecast lead, the 1,000 AC scores are ranked in descending order. The weighting configuration yielding the highest AC skill is selected as optimal for that window.
8. Final weighting mask: The optimal weighting masks from each 30-year window are averaged to obtain a final weighting mask, representing the mean of 23 windows from 1940–1991. The final weighting mask is shown in Fig. 8 C in the text.

Verification phase:
9. RMSD calculation in verification period: The RMSD is calculated for the verification period (1992–2022) following the same procedure as in Step 3.
10. Application of final weighting mask: The final weighting mask from Step 8 is applied to the RMSD values from the verification period, following Eq. (4).
11. Selection of initial model-analogs: The weighted RMSD values are ranked in ascending order, and the top 15 model-analogs with the smallest RMSD are selected, identifying the most optimal initial analog states within the data library.
12. Model-analog forecast generation: The subsequent evolution of these selected analogs is used to generate the model-analog forecasts.
13. Ensemble-mean forecasts: The ensemble mean of the 15-member model-analog forecasts is computed and compared against ERA5 verification data for 1992–2022.
14. Skill assessment: The AC forecast skill is evaluated, and the results are visualized to assess forecast performance.

By following these steps, we ensure that no information from the verification period (1992–2022) is used at any stage of the empirical model development, including during the generation or selection of weighting parameters. This separation upholds the fairness of the forecast skill assessment[66,67].

Data availability
Monthly ERA5 datasets are available for download at https://cds.climate. copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-

means?tab=form. CMIP6 datasets can be downloaded from https://esgf-node.llnl.gov/search/cmip6/. The monthly model outputs of SPEAR-LE used in this study are accessible at https://www.gfdl.noaa.gov/spear_large_ensembles/.

Code Availability
The R scripts used to calculate the VPD and conduct model-analog forecasts are available through the Zenodo open data repository: https://doi.org/10.5281/zenodo.15149221.

References

1. Quilcaille, Y., Batibeniz, F., Ribeiro, A. F. S., Padrón, R. S. & Seneviratne, S. I. Fire weather index data under historical and shared socioeconomic pathway projections in the 6th phase of the Coupled Model Intercomparison Project from 1850 to 2100. *Earth Syst. Sci. Data* **15**, 2153–2177 (2023).
2. Di Giuseppe, F. et al. Global seasonal prediction of fire danger. *Sci. Data* **11**, 128 (2024).
3. Jain, P., Castellanos-Acuna, D., Coogan, S. C. P., Abatzoglou, J. T. & Flannigan, M. D. Observed increases in extreme fire weather driven by atmospheric humidity and temperature. *Nat. Clim. Chang.* **12**, 63–70 (2022).
4. Sukkar, A., Dogru, A. O., Alganci, U. & Seker, D. Z. Conceptual design of a nationwide spatial decision support system for forest fire prevention and fighting. *Appl Geomat.* **16**, 347–363 (2024).
5. Abatzoglou, J. T., Williams, A. P., Boschetti, L., Zubkova, M. & Kolden, C. A. Global patterns of interannual climate–fire relationships. *Glob. Change Biol.* **24**, 5164–5175 (2018).
6. Abatzoglou, J. T. & Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl Acad. Sci.* **113**, 11770–11775 (2016).
7. Barbero, R., Abatzoglou, J. T., Larkin, S., Kolden, C. A. & Stocks, B. Climate change presents increased potential for very large fires in the contiguous United States. *Int. J. Wildland Fire* **24**, 892–899 (2015).
8. Barkhordarian, A., Saatchi, S. S., Behrangi, A., Loikith, P. C. & Mechoso, C. R. A Recent Systematic Increase in Vapor Pressure Deficit over Tropical South America. *Sci. Rep.* **9**, 15331 (2019).
9. Burke, M. et al. The changing risk and burden of wildfire in the United States. *Proc. Natl Acad. Sci.* **118**, e2011048118 (2021).
10. Ficklin, D. L. & Novick, K. A. Historic and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. *J. Geophys. Res.: Atmos.* **122**, 2061–2079 (2017).
11. Gamelin, B. L. et al. Projected U.S. drought extremes through the twenty-first century with vapor pressure deficit. *Sci. Rep.* **12**, 8615 (2022).
12. Seager, R. et al. Climatology, Variability, and Trends in the U.S. Vapor Pressure Deficit, an Important Fire-Related Meteorological Quantity. *J. Appl. Meteorol. Climatol.* **54**, 1121–1141 (2015).
13. Yuan, W. et al. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* **5**, eaax1396 (2019).
14. Simpson, I. R. et al. Observed humidity trends in dry regions contradict climate models. *Proc. Natl Acad. Sci.* **121**, e2302480120 (2024).
15. Lou, J., Joh, Y. & Delworth, T. The role of long-term trends and internal variability in altering fire weather conditions in the western United States. Preprint at https://doi.org/10.21203/rs.3.rs-3956903/v1 (2024).
16. Jiang, P., Yu, Z. & Acharya, K. Drought in the Western United States: Its Connections with Large-Scale Oceanic Oscillations. *Atmosphere* **10**, 82 (2019).
17. Mo, K. C., Schemm, J.-K. E. & Yoo, S.-H. Influence of ENSO and the Atlantic Multidecadal Oscillation on Drought over the United States. *J. Clim.* **22**, 5962–5982 (2009).
18. Schoennagel, T., Veblen, T. T., Romme, W. H., Sibold, J. S. & Cook, E. R. ENSO and PDO Variability Affect Drought-Induced Fire Occurrence in Rocky Mountain Subalpine Forests. *Ecol. Appl.* **15**, 2000–2014 (2005).
19. Delworth, T. L. et al. SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. *J. Adv. Model. Earth Syst.* **12**, e2019MS001895 (2020).
20. Jia, L. et al. Skillful Seasonal Prediction of North American Summertime Heat Extremes. *J. Clim.* **35**, 4331–4345 (2022).
21. Lu, F. et al. GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions. *J. Adv. Model Earth Syst.* **12**, e2020MS002149 (2020).
22. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Skillful Climate Forecasts of the Tropical Indo-Pacific Ocean Using Model-Analogs. *J. Clim.* **31**, 5437–5459 (2018).
23. Lou, J., Newman, M. & Hoell, A. Multi-decadal variation of ENSO forecast skill since the late 1800s. *npj Clim. Atmos. Sci.* **6**, 1–14 (2023).
24. Chen, Y., Morton, D. C., Andela, N., Giglio, L. & Randerson, J. T. How much global burned area can be forecast on seasonal time scales using sea surface temperatures? *Environ. Res. Lett.* **11**, 045001 (2016).
25. Shabbar, A., Skinner, W. & Flannigan, M. D. Prediction of Seasonal Forest Fire Severity in Canada from Large-Scale Climate Patterns. *J. Appl. Meteorol. Climatol.* **50**, 785–799 (2011).
26. Switanek, M. B., Troch, P. A. & Castro, C. L. Improving Seasonal Predictions of Climate Variability and Water Availability at the Catchment Scale. *J. Hydrometeorol.* **10**, 1521–1533 (2009).
27. Bedia, J. et al. Seasonal predictions of Fire Weather Index: Paving the way for their operational applicability in Mediterranean Europe. *Clim. Serv.* **9**, 101–110 (2018).
28. Frías, M. D., Herrera, S., Cofiño, A. S. & Gutiérrez, J. M. Assessing the Skill of Precipitation and Temperature Seasonal Forecasts in Spain: Windows of Opportunity Related to ENSO Events. *J. Clim.* **23**, 209–220 (2010).
29. Lorenz, E. N. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *J. Atmos. Sci.* **26**, 636–646 (1969).
30. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Diagnosing Secular Variations in Retrospective ENSO Seasonal Forecast Skill Using CMIP5 Model-Analogs. *Geophys. Res. Lett.* **46**, 1721–1730 (2019).
31. Lenssen, N. et al. Strong El Niño Events Lead to Robust Multi-Year ENSO Predictability. *Geophys. Res. Lett.* **51**, e2023GL106988 (2024).
32. Mulholland, D. P., Laloyaux, P., Haines, K. & Balmaseda, M. A. Origin and Impact of Initialization Shocks in Coupled Atmosphere–Ocean Forecasts*. *Monthly Weather Rev.* **143**, 4631–4644 (2015).
33. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Relating CMIP5 Model Biases to Seasonal Forecast Skill in the Tropical Pacific. *Geophys. Res. Lett.* **47**, e2019GL086765 (2020).
34. Wu, Y. & Tang, Y. Diagnosing seasonal forecast skill of the Indian Ocean Dipole mode using model-analogs. *Journal of Atmospheric and Oceanic Technology* **1**, 1037–1047 (2023).
35. Zhang, L. et al. Skillful decadal prediction skill over the Southern Ocean based on GFDL SPEAR Model-Analogs. *Environ. Res. Commun.* **5**, 021002 (2023).
36. Gu, Q. et al. Exploring multiyear-to-decadal North Atlantic sea level predictability and prediction using machine learning. *npj Clim. Atmos. Sci.* **7**, 1–15 (2024).
37. Anderson, W. et al. Preseason maize and wheat yield forecasts for early warning of crop failure. *Nat. Commun.* **15**, 7262 (2024).
38. Ek, M. B. Land Surface Hydrological Models. in *Handbook of Hydrometeorological Ensemble Forecasting* (eds. Duan, Q. et al.) 1–42 (Springer, Berlin, Heidelberg, 2018).
39. Chen, H.-C., Tseng, Y.-H., Hu, Z.-Z. & Ding, R. Enhancing the ENSO Predictability beyond the Spring Barrier. *Sci. Rep.* **10**, 984 (2020).
40. McPhaden, M. J. Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.* **30**, 1480 (2003).

41. Wang-Chun Lai, A., Herzog, M. & Graf, H.-F. ENSO Forecasts near the Spring Predictability Barrier and Possible Reasons for the Recently Reduced Predictability. *J. Clim.* **31**, 815–838 (2018).

42. Deser, C., Guo, R. & Lehner, F. The relative contributions of tropical Pacific sea surface temperatures and atmospheric internal variability to the recent global warming hiatus. *Geophys. Res. Lett.* **44**, 7945–7954 (2017).

43. Molteni, F., Kucharski, F. & Farneti, R. Multi-decadal pacemaker simulations with an intermediate-complexity climate model. *Weather Clim. Dyn.* **5**, 293–322 (2024).

44. Ding, H. & Alexander, M. A. Multi-Year Predictability of Global Sea Surface Temperature Using Model-Analogs. *Geophys. Res. Lett.* **50**, e2023GL104097 (2023).

45. Kumar, S., Newman, M., Wang, Y. & Livneh, B. Potential Reemergence of Seasonal Soil Moisture Anomalies in North America. *J. Clim.* **32**, 2707–2734 (2019).

46. Bröcker, J., Charlton–Perez, A. J. & Weisheimer, A. A statistical perspective on the signal-to-noise paradox. *Q. J. R. Meteorol. Soc.* **149**, 911–923 (2023).

47. Kumar, A. Finite Samples and Uncertainty Estimates for Skill Measures for Seasonal Prediction. *Monthly Weather Rev.* **137**, 2622–2631 (2009).

48. Kumar, A. & Hoerling, M. P. Analysis of a Conceptual Model of Seasonal Climate Variability and Implications for Seasonal Prediction. *AMS.* **81**, 255–264 (2000).

49. Sardeshmukh, P. D., Compo, G. P. & Penland, C. Changes of Probability Associated with El Niño. *J. Clim.* **13**, 4268–4286 (2000).

50. Shin, J., Park, S., Shin, S., Newman, M. & Alexander, M. A. Enhancing ENSO Prediction Skill by Combining Model-Analog and Linear Inverse Models (MA-LIM). *Geophys. Res. Lett.* **47**, e2019GL085914 (2020).

51. Toride, K. et al. Using Deep Learning to Identify Initial Error Sensitivity for Interpretable ENSO Forecasts. Preprint at https://doi.org/10.48550/arXiv.2404.15419 (2024).

52. Rader, J. K. & Barnes, E. A. *Optimizing Seasonal-to-Decadal Analog Forecasts with a Learned Spatially-Weighted Mask*. https://essopenarchive.org/users/558849/articles/649473-optimizing-seasonal-to-decadal-analog-forecasts-with-a-learned-spatially-weighted-mask?commit=4fe3bf52a0e86deb1e84ad52e586d45790e54dfb (2023) https://doi.org/10.22541/essoar.168748463.32520571/v1.

53. McCabe, G. J., Palecki, M. A. & Betancourt, J. L. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl Acad. Sci.* **101**, 4136–4141 (2004).

54. Kitzberger, T., Brown, P. M., Heyerdahl, E. K., Swetnam, T. W. & Veblen, T. T. Contingent Pacific–Atlantic Ocean influence on multicentury wildfire synchrony over western North America. *Proc. Natl Acad. Sci.* **104**, 543–548 (2007).

55. Moritz, M. A. et al. Climate change and disruptions to global fire activity. *Ecosphere* **3**, art49 (2012).

56. Turco, M. et al. Skilful forecasting of global fire activity using seasonal climate predictions. *Nat. Commun.* **9**, 2718 (2018).

57. Williams, A. P. et al. Correlations between components of the water balance and burned area reveal new insights for predicting forest fire area in the southwest United States. *Int. J. Wildland Fire* **24**, 14–26 (2014).

58. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

59. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.: Atmos.* **108**, D002670 (2003).

60. Henley, B. J. et al. A Tripole Index for the Interdecadal Pacific Oscillation. *Clim. Dyn.* **45**, 3077–3090 (2015).

61. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

62. Meinshausen, M. et al. Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.* **10**, 2057–2116 (2017).

63. Davis, R. E. Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.* **6**, 249–266 (1976).

64. Bolton, D. The Computation of Equivalent Potential Temperature. *AMS.* **108**, 1046–1053 (1980).

65. Amaya, D. J. et al. Linking Projected Changes in Seasonal Climate Predictability and ENSO Amplitude https://doi.org/10.1175/JCLI-D-23-0648.1 (2025).

66. Hastie, T., Tibshirani, R. & Friedman, J. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer, New York). https://doi.org/10.1007/978-0-387-84858-7 (2009).

67. Risbey, J. S. et al. Standard assessments of climate forecast skill can be misleading. *Nat. Commun.* **12**, 4346 (2021).

Acknowledgements

Author contributions

J.L. conducted the analyses and led the writing, with contributions from all the authors. Y.J., T.L.D, and L.J. helped shape the scientific ideas, and actively contributed to the discussions, writing, and manuscript polishing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-025-01028-6.

Correspondence and requests for materials should be addressed to Jiale Lou.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.