

Subseasonal Predictability of Diverse Persistent Wet-Cold Compound Extreme Events over China Based on Calibrating Dynamical and AI models

Wang Jiaxu

wangjiaxu23@mailsucas.ac.cn

China Academy of Meteorological Sciences <https://orcid.org/0009-0003-4233-2740>

Jie Wu

National Climate Center <https://orcid.org/0000-0002-7806-6718>

Hong-Li Ren

China Academy of Meteorological Sciences

Li Guo

National Climate Center

Lijuan Chen

National Climate Center

Chenguang Zhou

National Climate Center

Research Article

Keywords: Persistent wet–cold compound extreme events, Subseasonal predictability, S2S prediction systems, AI-based model, CDF calibration

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10683844/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Subseasonal Predictability of Diverse Persistent Wet-Cold
Compound Extreme Events over China Based on Calibrating
Dynamical and AI models**

Jiaxu Wang^{a,b}, Jie Wu^{c*}, Hong-Li Ren^a, Li Guo^c, Lijuan
Chen^c, Chenguang Zhou^c

^a Chinese Academy of Meteorological Sciences, Beijing 100081, China

^b College of Earth and Planetary Sciences, University of Chinese Academy of
Sciences, Beijing 100049, China

^c State Key Laboratory of Climate System Prediction and Risk Management/China
Meteorological Administration Key Laboratory for Climate Prediction Studies,
National Climate Centre, China Meteorological Administration, Beijing 100081,
China

*Corresponding author: Jie Wu; Email: wujie@cma.gov.cn

ORCID: Jiaxu Wang, <https://orcid.org/0009-0003-4233-2740>

Acknowledgments: The authors gratefully acknowledge the data providers
for making the observational and subseasonal prediction datasets used in
this study available.

Abstract Persistent wet–cold compound extreme events (PWCEs) exert substantial socioeconomic impacts over China, but their subseasonal predictability and physical sources remain poorly understood. Using an event-oriented framework, this study investigates the subseasonal prediction skill of three PWCE types—South China (SC), Western China (WC), and Central China (CC)—using dynamical and artificial intelligence (AI)-based subseasonal-to-seasonal (S2S) systems (CMA-CPSv3, ECMWF, NCEP, and CMA-FENGSHUN). A lead- and member-specific cumulative distribution function (CDF)-based calibration was applied to reduce systematic climatological and extreme-value biases in temperature and precipitation forecasts. After calibration, all models capture the aggregated PWCE spatial distributions and associated wet–cold anomaly patterns, with recognizable signals persisting to 15–20 days. Individual PWCEs are generally predictable at ~15-day lead times, with marked type-dependent differences: WC-type events show higher spatial consistency in continuous-field metrics, whereas SC- and CC-type events perform better in capturing event spatial extent. The CC type shows the highest overall predictability, associated with coherent mid- to high-latitude wave trains and suppressed Maritime Continent convection. PWCEs preceded by robust blocking highs and organized tropical convective anomalies are more predictable. ECMWF performs best at short lead times, while FENGSHUN performs better at medium-to-long lead times, especially for CC-type events. Dynamical models effectively translate large-scale circulation and convective signals into forecast skill, whereas FENGSHUN skill is largely decoupled from these physical sources. Temperature predictability mainly originates from mid- to high-latitude circulation regimes, while precipitation predictability relies more on tropical convection. These findings support improved subseasonal prediction and early warning of compound extremes.

Keywords Persistent wet–cold compound extreme events; Subseasonal predictability; S2S prediction systems; AI-based model; CDF calibration

1 Introduction

Persistent Wet–Cold Compound Extreme events (PWCEs) constitute a category of compound extremes arising from the concurrence of anomalous precipitation and low-temperature conditions. Characterized by pronounced cooling alongside persistent precipitation, these conditions collectively promote the development of hazardous weather phenomena, such as snowfall and freezing rain (Peng et al., 2021). Compared with traditional single extremes of cold temperature or heavy precipitation, PWCEs generally lead to more severe societal impacts and economic losses. Against the background of global warming, recent studies have suggested that the frequency of PWCEs may exhibit a decreasing trend, whereas their extreme intensity and associated destructive risks could further intensify (Zong et al., 2022; Wang et al., 2026). One of the most representative PWCEs occurred over southern China in 2008, affecting a vast population and causing substantial socioeconomic losses (Ding et al., 2008; Sun and Zhao, 2010). Thereafter, PWCEs have continued to occur frequently, including events in January 2018 (Sun et al., 2019), February 2022 (Wang and Li, 2024), December 2023 (Ye et al., 2025), and February 2024 (Zhao et al., 2025a; Ma et al., 2026). It is worth noting that the evolutionary characteristics and affected regions are different across these PWCEs. Some events propagate from west to east (e.g., 2018–01), whereas others develop through north–south cold air intrusions (e.g., 2023–12), while some are primarily concentrated over the eastern Tibetan Plateau. Recently, Wang et al. (2026) classified PWCEs into three distinct types based on their spatial characteristics: the Southern China type (SC), the Western China type (WC), and the Central China type (CC). These three types exhibit markedly different large-scale circulation anomalies over high latitudes, as well as distinct tropical signals and moisture transport pathways.

Given the significant societal and economic impacts of PWCEs, skillful sub-seasonal-to-seasonal (S2S) prediction is crucial as it offers a pivotal 2–4-week lead time for disaster prevention and mitigation efforts (Vitart et al., 2017). Such prediction has substantial practical value, as it bridges the gap between weather and seasonal forecasts (Brunet et al., 2010). Although the importance of S2S prediction has been increasingly recognized, it remains more challenging than weather and seasonal prediction (Mariotti et al., 2020). This difficulty arises because the lead time is long enough for the memory of atmospheric initial conditions to decay, yet not sufficiently long for slowly varying oceanic anomalies to exert a dominant influence (Robertson et al., 2020). Recent studies suggest that potential predictability may also arise from factors such as tropical MJO variability (Xiang et al., 2022), ENSO evolution (Zheng et al., 2019), land surface processes (Zhu et al., 2019), sea ice (Li et al., 2019), and tropical–extratropical teleconnections (Stan et al., 2022; Massonnet et al., 2026).

Initiated jointly by the WWRP/WCRP in 2013, the S2S prediction project has generated substantial hindcast and real-time forecast data from 12 international research and operational centers (Vitart et al., 2017, 2025). These datasets are widely employed for the early warning of high-impact weather and climate events, such as heatwaves, droughts, cold surges, and anomalous precipitation patterns (Domeisen et al., 2022; Domeisen et al., 2023; Sreeparvathy et al., 2025; Zheng et al., 2026). Evaluations of the direct forecast skill for temperature and precipitation based on S2S models indicate that the effective forecast lead time is generally limited to about 2–3 weeks, with prediction skill decreasing faster in extratropical regions than in the tropics (Vitart et al., 2017; Andrade et al., 2019; Vitart et al., 2025). However, recent verification of subseasonal extreme-event forecasts based on the ECMWF model indicates that prediction skill for cold surges extends to about 2–3 weeks, whereas skill for heavy precipitation events is even lower (Domeisen et al., 2022; Rivoire et al., 2023). Compared with single-

variable extremes, compound extreme events pose greater challenges for prediction systems, as they involve coordinated anomalies of multiple meteorological variables and their persistence. For instance, an evaluation of four PWCEs over southern China during the winter of 2008 by Zhang et al. (2020) showed that model prediction skill decreases markedly beyond about 10 days. Moreover, systematic research into subseasonal predictive capabilities for PWCEs is still relatively scarce, the accuracy of forecasts, the factors influencing predictability, and the main sources of error remain poorly understood. In addition, the frequent occurrence and diversity of PWCEs indicate that it remains a significant risk under current climate conditions, posing a major challenge for the prediction of such hazardous extreme events.

In recent years, AI-powered foundation models for weather and climate prediction have advanced rapidly. Models such as Pangu-Weather and GraphCast have shown that they can now match—or even surpass—the medium-range global forecasting skill of leading numerical weather prediction systems (Bi et al., 2023; Lam et al., 2023). More recently, models such as GenCast and FuXi-S2S have expanded AI’s application to ensemble and subseasonal prediction, demonstrating strong performance in global forecasting and in representing large-scale climate variability such as the Madden–Julian Oscillation (Price et al., 2024; Chen et al., 2024). In terms of extreme-event prediction, some studies suggested that AI models may improve forecasts of extreme weather and tropical cyclone tracks (Price et al., 2024; Bodnar et al., 2025). However, evaluations of high-impact events also highlight limitations of AI models, particularly in capturing peak intensities, sharp frontal structures, and regional-scale details (Charlton-Perez et al., 2024; Kochkov et al., 2024). Moreover, most previous evaluations of AI-based weather and climate models have focused on individual hazards, such as heatwaves, tropical cyclones, and extreme precipitation (Camps-Valls et al., 2025), compound extreme-event prediction, which relies on the joint behavior of multiple variables and processes (Zscheischler et al., 2020), has not been systematically evaluated yet. This raises an important question: how skillful are AI-based models in predicting compound extreme events, and how do their relative advantages and limitations compare with those of traditional dynamical subseasonal-to-seasonal (S2S) models?

This study will systematically evaluate the subseasonal prediction skill for PWCEs over China and explore the potential predictability sources using four operational dynamical and AI-based S2S prediction systems. Unlike previous assessments based mainly on individual variables, or case studies, it adopts an event-oriented framework to characterize variations in prediction skill across event types and to investigate the associated physical sources of predictability. Model performance is evaluated in terms of event characteristics, spatial patterns, and physical consistency. In addition, key circulation factors influencing the predictability of PWCEs are further examined, with particular attention to differences among the three PWCE types. The remainder of this paper is organized as follows. Section 2 introduces the data and methods. Section 3 examines the overall predictability and spatial structure of PWCE forecasts. Section 4 analyzes the associated circulation features and sources of predictability. Section 5 provides the conclusions and discussion.

2 Data and Methods

2.1 Data

This study employs three dynamical S2S prediction systems: the China Meteorological Administration Climate Prediction System version 3 (CMA-CPSv3), the European Centre for Medium-Range Weather Forecasts Integrated Forecasting System (ECMWF), and the National Centers for

Environmental Prediction Climate Forecast System version 2 (NCEP-CFSv2)—alongside the data-driven machine learning system FENGSHUN to evaluate the subseasonal predictability of PWCEs over eastern China. The corresponding initialization periods and forecast intervals are summarized in Table 1.

The CN05.1 daily gridded precipitation and 2-m temperature dataset is used to identify PWCEs and verify model prediction performance. It is constructed from long-term station observations across China using multiple interpolation methods (Wu and Gao, 2013), with a horizontal resolution of $0.25^\circ \times 0.25^\circ$ and high spatial and temporal consistency, enabling reliable representation of regional-scale climate variability.

Daily mean air temperature, geopotential height (GPH), sea level pressure, and wind fields used for circulation analysis are obtained from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis, with a horizontal resolution of $2.5^\circ \times 2.5^\circ$. The NCEP/NCAR reanalysis is selected primarily because of its long-term temporal consistency and wide availability, as well as its timely updating, which facilitates potential applications in near-real-time operational monitoring. To assess the robustness of the results, we cross-validate the main circulation features using the ERA5 reanalysis and find good consistency between the two datasets (figures not shown).

Table 1 Basic information of the S2S prediction models used in this study

Model	Institution	Resolution	Ensemble Size	Initial Year	Forecast Lead Time (days)
CMA-CPSv3	CMA	$0.45^\circ \times 0.45^\circ$	4	2006	60
ECMWF Cy-S5	ECMWF	$1.5^\circ \times 1.5^\circ$	11	2004	46
NCEP-CFSv2	NCEP	$1.5^\circ \times 1.5^\circ$	16	1999	44
FENGSHUN	CMA	$1.5^\circ \times 1.5^\circ$	100	2002	60

2.2 Methods

2.2.1 Identification of PWCE

The identification of PWCEs follows our previous study (Wang et al., 2026), in which both relative and absolute thresholds for temperature and precipitation are used to define PWCEs. The relative thresholds are calculated using a 15-day moving window, with the temperature threshold set as the 25th percentile of daily mean temperature and the precipitation threshold defined as the 50th percentile of non-zero precipitation days (daily precipitation > 0 mm), allowing each calendar day at each grid point to have a unique relative threshold. The absolute thresholds are set as below 4°C for temperature and above 1 mm/day for precipitation. These four thresholds are specified independently at each grid point and for each calendar day to identify grid-scale events. Spatiotemporally contiguous events are then extracted using the three-dimensional connected-component (CC3D) method proposed by Luo et al. (2022). This method accounts for connectivity in both the temporal (daily) and spatial (grid-based) dimensions. A WCE daily candidate is first identified when the number of contiguous grid points exceeds 200 on a given day. A PWCE event is defined when these daily candidates persist for at least three consecutive days with spatial overlap. A total of 252 PWCE events have been identified during 1961–2023 (see Wang et al., 2026 for further details). Following Zhang et al. (2011), a wet–cold intensity index (PT) is defined to quantify the overall intensity of PWCEs by summing the daily contributions of all grid points within each event, as described below.

$$PT_{ij} = \frac{P_{ij} - \overline{P_{ij}}}{P_s} - \frac{T_{ij} - \overline{T_{ij}}}{T_s}, \quad (1)$$

In this formulation, P_{ij} and T_{ij} are the daily precipitation and daily mean temperature at grid point j on day i , while $\overline{P_{ij}}$ and $\overline{T_{ij}}$ denote the corresponding climatological daily means. P_s and T_s are the climatological standard deviations of precipitation and temperature. Larger PT values indicate stronger wet–cold signals, characterized by higher precipitation and lower temperature.

We classify the observed PWCEs using the K-means clustering algorithm based on their initiation locations and trajectory centroid positions, which together capture both event onset and evolution pathways. Based primarily on the Calinski–Harabasz index (Calinski and Harabasz, 1974) and the SSE elbow criterion (Ketchen and Shook, 1996), and considering the robustness and physical interpretability of the results, the number of clusters is set to $k = 3$. PWCEs are then divided into three categories: Southern China (SC) type, Western China (WC) type and Central China (CC) type. Their respective spatial distributions and trajectories are shown in Online Resource 1 (Fig. S1).

2.2.2 Model bias calibration

To reduce the influence of systematic model biases on event identification and prediction skill evaluation, this study applies a cumulative distribution function (CDF)–based statistical bias calibration to the forecasts of temperature and precipitation from both dynamical and data-driven S2S prediction systems. The CDF-based bias calibration maps the empirical distribution of model forecasts onto the observed distribution, effectively correcting systematic biases in the mean, variance, and distributional shape, which has been widely adopted in S2S prediction studies (Gudmundsson et al., 2012; Wu et al., 2026). For temperature, we apply standard quantile mapping (QM) at each grid point and for different forecast leads with a 15-day running window by directly mapping the empirical CDF of model forecasts onto the observed CDF. Because daily precipitation exhibits strong intermittency and a high frequency of zero values, direct application of standard quantile mapping may produce undesired changes in wet- and dry-day frequencies (Gudmundsson et al., 2012; Maraun et al., 2017). Accordingly, we employ a two-step CDF-based calibration for precipitation. First, we constrain precipitation occurrence frequency using the observed cumulative probability of zero precipitation. Model precipitation with cumulative probabilities below this threshold is set to zero, ensuring consistency of wet-day and dry-day frequencies with observations. Second, quantile mapping is applied to wet-day samples only, mapping modeled wet-day precipitation onto the observed distribution to correct intensity biases while avoiding shifts in precipitation occurrence frequency (Maraun, 2013; Cannon et al., 2015; Ivanov and Kotlarski, 2017). Then the same PWCE identification criteria are applied to model forecast data since their mean and variance have been corrected.

2.2.3 Model verification methods

Model prediction skill can be evaluated using various metrics targeted for continuous and categorical variables, respectively. Here, we assess the prediction skill for PWCEs using the anomaly correlation coefficient (ACC; Li and Robertson, 2015) and the normalized root mean square error (NRMSE; Lee and Wang, 2014; Zhu et al., 2014). Before evaluation, model forecasts and observations are aligned in time and space and mapped onto the observation grid. Skill metrics are computed only at grid points with valid values in both datasets.

Two forms of ACC are used in this study. The non-centered ACC (ACCu) retains information on both spatial consistency and anomaly amplitude, while the centered ACC focuses on the similarity of spatial anomaly patterns after removing the spatial mean. The specific form of ACC used in each analysis

is indicated in the corresponding figure caption. The centered and non-centered ACC are respectively defined as:

$$ACC = \frac{\sum_{i=1}^N (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (F_i - \bar{F})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (2)$$

$$ACC_u = \frac{\sum_{i=1}^N F_i O_i}{\sqrt{\sum_{i=1}^N F_i^2} \sqrt{\sum_{i=1}^N O_i^2}} \quad (3)$$

In this formulation, F_i and O_i are the forecast and observed values at grid point i , respectively. $\bar{F}$ and $\bar{O}$ denote their spatial means over all valid grid points. The summation is taken over all valid grid points in the verification domain.

The NRMSE is used to quantify model forecast errors relative to the spatial variability of observed anomalies and is defined as the ratio of the root mean square error (RMSE) to the spatial standard deviation of the observed anomaly field (Lee and Wang, 2014), as given by:

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - \bar{O})^2}} \quad (4)$$

Here, $\bar{O}$ denotes the spatial mean of the observed field over the same set of valid grid points. Event-scale ACC and NRMSE are obtained by temporally averaging the daily values over each event.

To assess the ability of S2S models to discriminate PWCE occurrence, we employ binary verification metrics following Wilks (2019). Observations and forecasts are converted into daily binary fields at each grid point, with occurrence coded as 1 and non-occurrence as 0. At each grid point, a hit (H) occurs when both forecasts and observations indicate an event, a miss (M) when an event is observed but not forecast, a false alarm (F) when an event is forecast but not observed, and a correct negative (C) when neither indicates an event. The threat score (TS), which measures the overall accuracy of event forecasts, is defined as:

$$TS = \frac{H}{H+M+F} \quad (5)$$

The TS ranges from 0 to 1, with higher values indicating better forecast performance. The probability of detection (POD), which measures the fraction of observed events that are correctly forecast, is defined as:

$$POD = \frac{H}{H+M} \quad (5)$$

The POD ranges from 0 to 1, with higher values indicating better event detection.

3 Overall predictability and spatial structure

3.1 The reproduction of multi-events averaged characteristics

The frequency distributions of identified PWCEs from four S2S models after CDF-calibration for different lead times are presented in Fig.1. Note that for models with interval initialization (e.g., ECMWF), the lead time is extended by 1–3 days to locate the nearest initialization date, ensuring the continuity of the verification sample. In observations, PWCEs are mainly located from the Yellow River to south China, with pronounced maxima over the Yangtze River Basin and its southern regions. All four S2S models capture the main spatial features of the observed PWCE distribution, suggesting that the bias-corrected forecasts reliably represent PWCE-prone regions. With increasing forecast lead time, the intensity and spatial extent of the high-frequency centers show a slight weakening, especially for CMA

and ECMWF models. However, the geographical locations remain broadly consistent with the observations even for the lead time of 26 days, with the ACC still remaining around 0.9. In addition, the CDF-calibrated forecasts show a clear improvement in representing the event occurrence frequency compared with the original forecasts. In contrast, the uncorrected forecasts exhibit much lower ACC values, which decrease rapidly with increasing lead time. The spatial distribution also exhibits pronounced biases, with high-frequency centers predominantly located in western regions, particularly over the southern Qinghai–Tibet Plateau, which contradicts the observed PWCE distribution (Fig. S2).

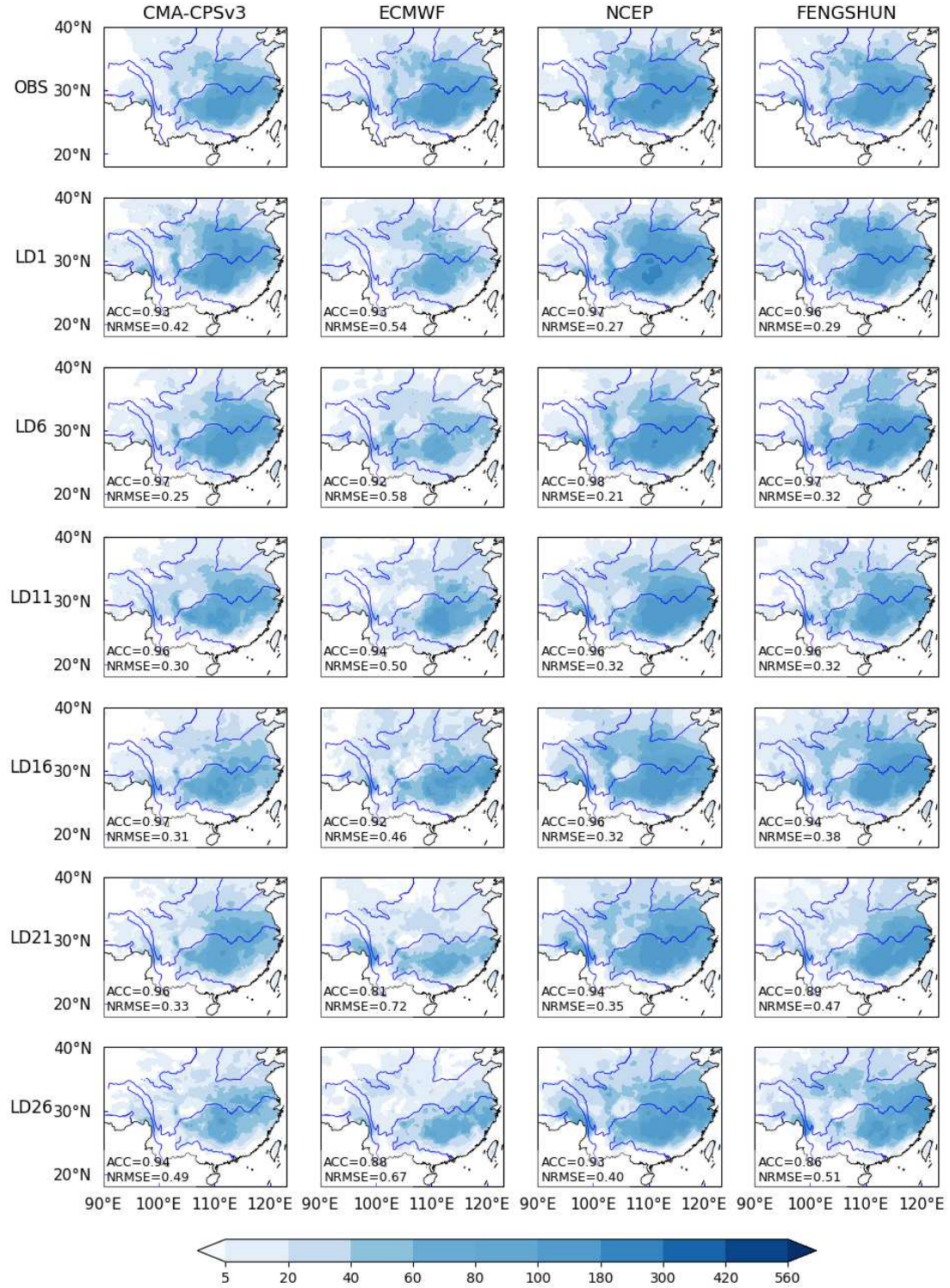


Fig. 1 Spatial distribution of the frequency of single-day, grid-point events occurring during PWCE periods, as identified from four S2S models (CMA-CPSv3, ECMWF, NCEP, and FENGSHUN) after CDF-based bias calibration. The first row shows the observations (OBS), and the remaining rows present the corresponding model results at different lead times (LD1, LD6, LD11, LD16, LD21, and LD26). Shading denotes the event occurrence frequency, with colors from light to dark indicating increasing frequency. The ACC and NRMSE for each panel are shown in the lower-left corner.

The effectiveness of the CDF-based bias calibration in reproducing the spatial distribution of PWCEs is further evaluated using the ACC–NRMSE relationship between observations and the outputs of all four S2S models (Fig. 2). After calibration, all models shift clearly toward higher ACC and lower NRMSE values, indicating improved spatial consistency with observations and reduced magnitude errors. In contrast, the uncorrected results predominantly occupy the low-ACC and high-NRMSE region, with degradation worsening at longer lead times. As lead time increases, both corrected and uncorrected forecasts show declining ACC and rising NRMSE, reflecting the inherent decay of S2S prediction skill. Across all lead times, the CDF-corrected results consistently outperform their uncorrected counterparts. Specifically, after CDF calibration, the mean ACC across six lead times increases from 0.314 to 0.935, and the mean NRMSE decreases from 2.960 to 0.406, corresponding to an absolute ACC gain of 0.621 and an 86.3% reduction in NRMSE. These results demonstrate that CDF calibration substantially enhances the representation of PWCEs.

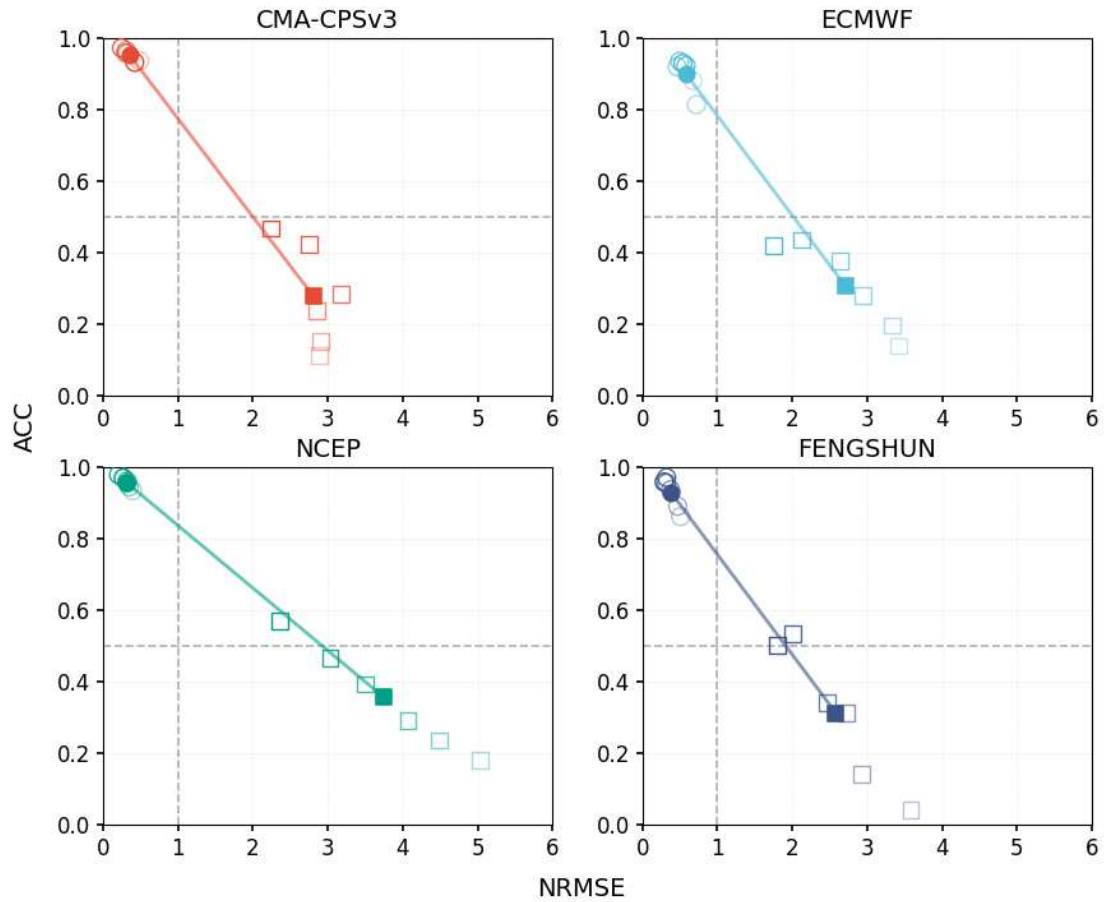


Fig. 2 ACC–NRMSE relationship for PWCE event-day frequency distributions from observations

and four S2S models before and after CDF calibration. Squares and circles denote uncorrected and corrected results, respectively, at leads of 1, 6, 11, 16, 21, and 26 days. Lines connect their midpoints to illustrate the overall impact of CDF calibration. Transparency decreases with lead time increasing.

From the multi-events averaged perspective, we evaluate how well the S2S models reproduce the observed wet-cold intensity structures of different PWCE types. Figs. 3–5 show the spatial distributions of predicted PT fields for the three PWCE types at different forecast lead times based on observation-identified PWCEs. In general, at short lead times (LD1 and LD6), across all the PWCE types, the four models generally reproduce the broad locations of the observed PT centers with the ACCu remaining around 0.9 and NRMSE below 1.0, although the amplitudes are slightly weaker. After LD11, the NRMSE approaches or exceeds 1.0 across all models and event types and increases progressively thereafter, indicating growing forecast errors. A more pronounced degradation in prediction skill emerges after LD16, as the ACCu and NRMSE show diverse characteristics across different models and PWCE types.

Specifically, for SC-type events, although the NRMSE increases substantially after LD11, the observed spatial pattern remains reasonably well captured by most models until LD16, with the ACCu remaining above 0.9. However, by LD21, CMA-CPSv3 and ECMWF exhibit spurious negative PT values over the northern and southern parts of the event domain, respectively, resulting in a dramatic drop in ACCu to negative values. In contrast, NCEP and FENGSHUN still retain the wet-cold spatial pattern, albeit with reduced amplitude.

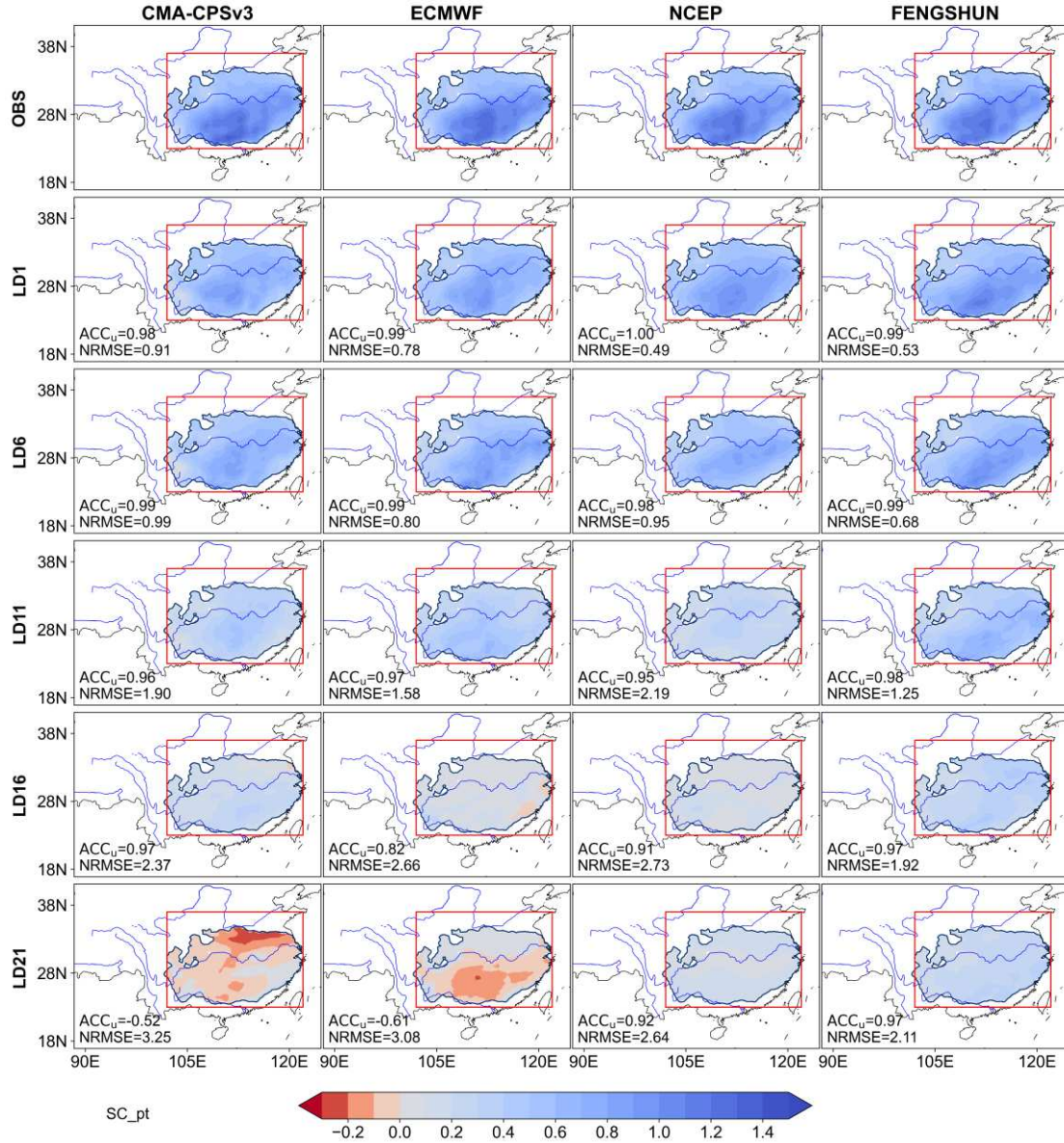


Fig. 3 Spatial distributions of PT values during SC-type PWCEs from four S2S models at different lead times. The first row shows the corresponding observations, and the remaining rows show model results. Shading indicates PT values, the red rectangle marks the main impact region, and the black contour outlines the primary occurrence region. ACC_u and NRMSE calculated from the composite PT fields are shown in the lower-left corner of each panel.

For WC-type events, except for CMA-CPSv3, the ECMWF, NCEP, and FENGSHUN models effectively maintain the western PT structure even at longer lead times (LD16 and LD21), with the ACC_u remaining above 0.85. In particular, FENGSHUN exhibits the strongest performance, maintaining an ACC_u above 0.9 and an NRMSE below 1.0 through LD11, which demonstrates its robust capability to capture the wet-cold anomaly pattern over the eastern Tibetan Plateau.

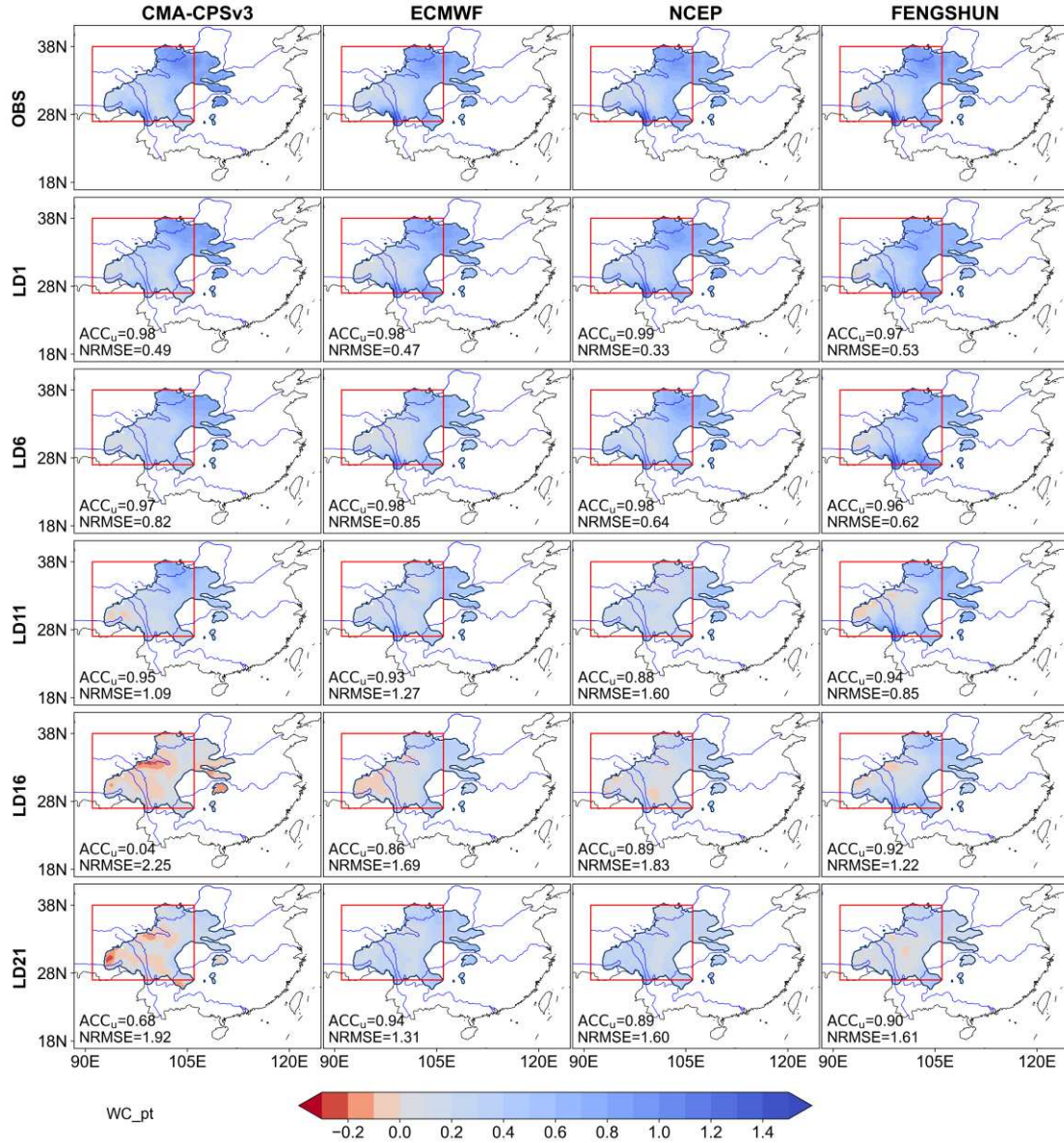


Fig. 4 Same as Fig. 3, but for WC-type PWCEs.

For CC-type events, the broad positive PT center weakens after LD11, with degradation becoming more pronounced after LD16, particularly in CMA-CPSv3 and NCEP. Meanwhile, ECMWF and FENGSHUN retain moderate to high spatial consistency, with an ACC_u remaining above 0.7 and positive PT values broadly distributed across the event domain. Overall, PT forecasts exhibit larger amplitude errors after LD11, as indicated by the increasing NRMSE, while the loss of spatial skill beyond LD16 varies across all models. Among them, FENGSHUN exhibits the slowest skill decay, most realistically capturing the wet-cold anomalies at longer lead times.

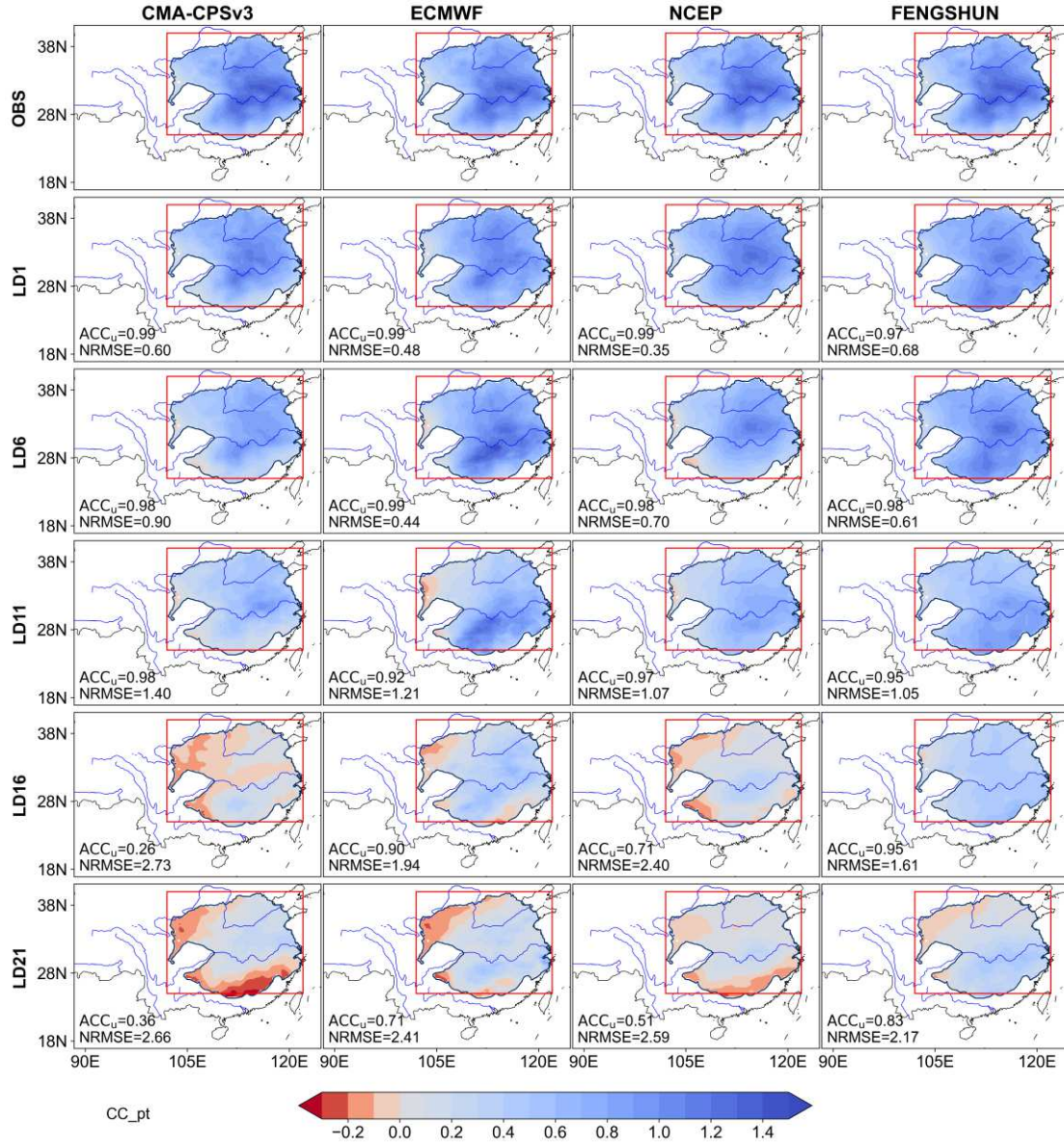


Fig. 5 Same as Fig. 3, but for CC-type PWCEs.

To diagnose the evolution of raw model biases in the two PT components, Figure 6 illustrates the joint distributions of the standardized precipitation (P) and negative temperature (-T) anomalies during PWCE periods across the S2S models. P and -T are generally positively correlated. As lead time increases, the model means in all four models shift away from the observed mean toward weaker PT amplitudes, but the bias evolution differs among models. ECMWF and NCEP remain relatively close to the observed mean at short lead times and then exhibit a gradual weakening of both components. CMA-CPSv3 shows a modest precipitation bias at short lead times, followed by a joint reduction in the precipitation and cold-temperature components. FENGSHUN shows the largest precipitation underestimation even at short lead times, while its cold-temperature component weakens progressively with increasing lead time.

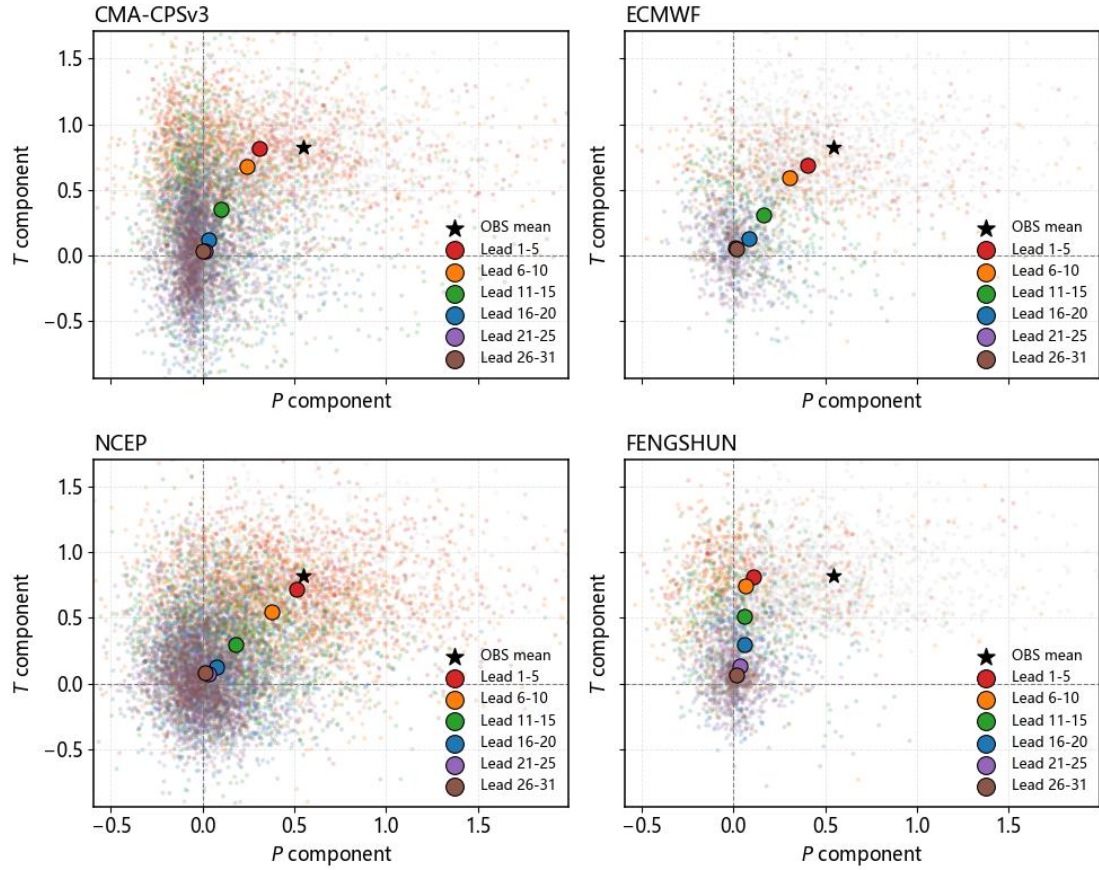


Fig. 6 Comparison of the precipitation (P, x-axis) and temperature (T, y-axis) components of daily PT values during PWCE periods from four S2S models (CMA-CPSv3, ECMWF, NCEP, and FENGSHUN) at different forecast lead times before CDF calibration. Circles denote the mean values from the model results, while stars indicate the observed mean values.

3.2 The predictability of individual PWCE

Figure 7 shows the variation of prediction skill for PWCE PT fields with forecast lead time for the four S2S models after CDF calibration. For all events (Fig. 7a), ACCu values are relatively high at short lead times but decrease substantially after about two weeks. Using 0.5 as a threshold, the average useful prediction limit reaches approximately 11 days for the ECMWF model and 8 days for the other models. However, based on the 95% statistical significance thresholds calculated using the effective degrees of freedom—which range from approximately 0.2 to 0.3 and vary across PWCE types—the average skillful prediction limit extends to 20 days for FENGSHUN and 15–17 days for the other dynamical models.

Differences in prediction skill are evident among PWCE types. WC-type events retain positive ACCu values for a longer period and show the slowest decay, with the useful prediction limit of FENGSHUN extending to 15 days, suggesting relatively higher predictability in the continuous PT-field metric (Noted that the event number is less for WC-type). By contrast, SC- and CC-type events exhibit a more rapid reduction in ACCu after the medium lead range, although the ECMWF model maintains a useful prediction limit of about 12 days for these types. The four models also show distinct lead-time-dependent behaviors. ECMWF generally performs better at short lead times, especially within the first 0–10 days, whereas FENGSHUN shows a slower skill decay and retains relatively higher ACCu values

at medium-to-long lead times. These results indicate that PWCE prediction skill is both event-type dependent and model dependent.

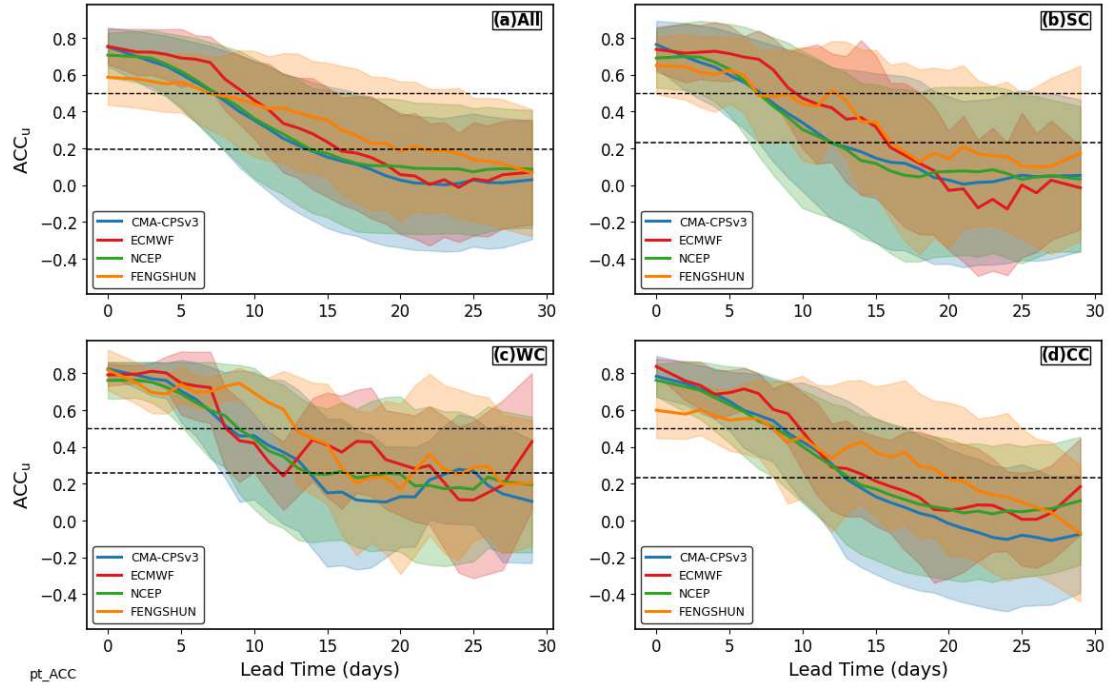


Fig. 7 Lead-time dependence of PWCE prediction skill for the four S2S models (CMA-CPSv3, ECMWF, NCEP, and FENGSHUN) after CDF calibration. Shown are event-based ACCus of PT fields as a function of forecast lead time, defined relative to the observed event onset, for all events and for the three PWCE types. The ACCus are calculated within type-dependent spatial masks: 90° – 122° E, 23° – 40° N for all events; 102° – 122° E, 23° – 37° N for SC; 91° – 106° E, 27° – 38° N for WC; and 102° – 122° E, 25° – 40° N for CC.

Figure 8 illustrates the variation of spatial binary matching skill, as measured by TS, with forecast lead time for the four S2S models based on observation-identified PWCE events. TS quantifies the spatial overlap of grid points between the model-identified and observed event extents. A similar improvement is evident in POD (Fig. S3), with calibration increasing the fraction of observed PWCE grid points correctly detected by the models. For all events (Fig. 8a), TS generally decreases with forecast lead time across all models, indicating that the ability to accurately capture the spatial extent of observed PWCEs weakens as lead time increases. The CDF-corrected results exhibit higher TS at nearly all lead times, demonstrating that bias calibration effectively improves the identification of event spatial extent. Overall, the FENGSHUN model generally exhibits the highest TS, except at very short lead times (0–2 days). Using 0.3 as a threshold, FENGSHUN maintains skillful predictions up to a lead time of about 15 days, with particularly strong performance for CC- and SC-type events. Except that, ECMWF performs better than the other dynamical models, maintaining an average skillful range of about 5–10 days. Notably, the initial forecast skills are relatively low for all models (below 0.5), indicating that it remains challenging to directly identify PWCE events even at very short lead times. Together with the ACCu results in Fig. 7, these results suggest that WC-type events show relatively better continuous-field spatial consistency, whereas CC-type events have clearer advantages in event-area matching.

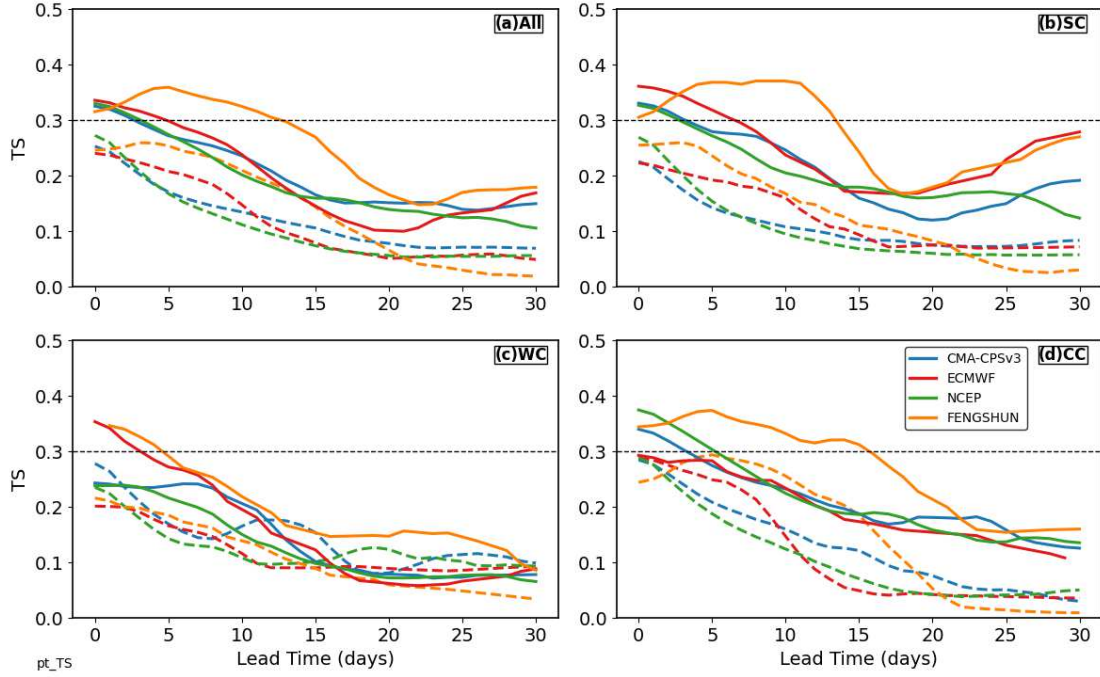


Fig. 8 Lead-time dependence of TS for PWCEs from four S2S models before and after CDF calibration, shown for all events and the three PWCE types. Solid and dashed lines indicate corrected and uncorrected results, respectively.

4 Circulation and physical sources of predictability

We further assess the models' capability to reproduce mid- to high-latitude circulation anomalies associated with diverse PWCEs. Taking FENGSHUN and ECMWF as representatives (Fig. 9; see Fig. S4 for CMA-CPSv3 and NCEP), both models effectively capture the dominant circulation signatures at short-to-medium lead times (0–10 days)—specifically, the meridional “north-positive south-negative” 500-hPa GPH dipole and the enhanced Siberian High at the surface—with ACCu skills exceeding 0.4. Beyond 15 days, simulation fidelity degrades, and skill exhibits notable model- and event-type dependencies. Specifically, FENGSHUN excels in capturing the northwest-southeast wave train of CC-type events ($\text{ACCu} > 0.5$ at 15-day lead time), consistent with its superior PWCE prediction skill (Figs. 7d, 8d). Meanwhile, ECMWF better maintains the meridional dipole structure over western China for WC-type events up to 20 days, supporting its higher prediction skill for this subtype (Fig. 7c). For SC-type, the circulation simulation performance of FENGSHUN is slightly higher than that of ECMWF models, but the anomalous magnitudes are underestimated in both models.

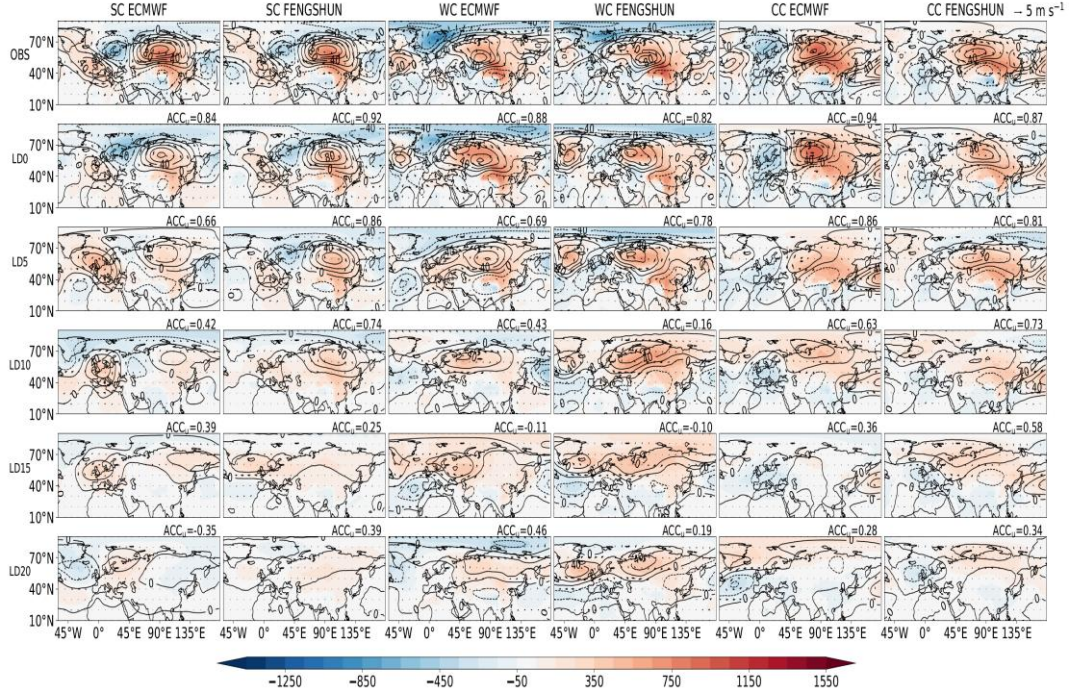


Fig. 9 Composite high-latitude circulation anomalies on PWCE types from the FENGSHUN and ECMWF models at different forecast lead times. Contours show 500-hPa GPH (Z500), and shading indicates sea level pressure (MSL) anomalies. Stippling denotes regions where MSL anomalies pass the significance test. The ACCu of each model result relative to the corresponding observations in the first row is shown in the upper-right corner of each panel.

Tropical convective anomalies and the corresponding low-level circulation are crucial for PWCE moisture transport and serve as primary predictability sources. Fig. 10 evaluates the capability of the ECMWF dynamical model and the FENGSHUN AI-based model to capture these features (see Fig. S5 for CMA-CPSv3 and NCEP). Observations show that SC-type events feature enhanced convection over the Maritime Continent (MC) and South China Sea (SCS), driving moisture-laden southwesterlies toward southern China. Conversely, WC- and CC-type events exhibit widespread positive OLR anomalies over the MC and western Pacific (WP). For CC-type events, this suppressed convection stimulates an anticyclonic anomaly to its northwest, reinforcing the WNPAC and driving southwesterlies toward central China. Both models capture these anomalous tropical patterns at lead times up to 10 days. Beyond this window, predictability diverges among PWCE types. CC-type events retain predictability at medium lead times (15–20 days), particularly for FENGSHUN ($ACC_u > 0.5$), which successfully reproduces the distinct convective-suppression northwest of the MC, reflecting the more coherent convection–circulation coupling inherent to CC-type events. In contrast, the tropical signals for SC- and WC-type events (enhanced convection over the SCS and suppressed convection over the eastern MC) weaken after 10 days. Furthermore, an inter-model contrast emerges: FENGSHUN better maintains the spatial structures of OLR and circulation anomalies for WC- and CC-type events, whereas ECMWF performs relatively better for SC-type events, consistent with their respective skills in predicting PWCE PT distributions (Fig. 7).

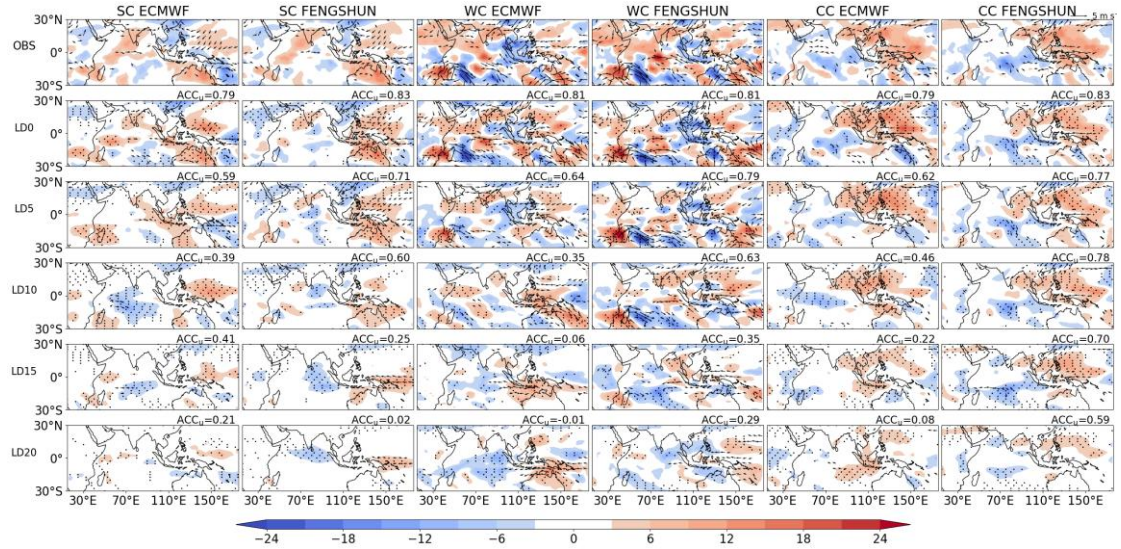


Fig. 10 Composite low-latitude circulation anomalies on the first day of the three PWCE types from the FENGSHUN and ECMWF models at different forecast lead times. Shading indicates outgoing longwave radiation (OLR) anomalies, and stippling denotes regions where OLR anomalies pass the significance test. Vectors show low-level winds, with only wind speeds exceeding 1 m s^{-1} displayed. ACCu relative to observations is shown in the upper-right corner.

To assess the model's performance in capturing key circulation evolution features and identify potential predictability windows across PWCE types, we classify events into high- and low-skill groups based on the PT-index ACCu at a 20-day lead time (LD20) from FENGSHUN, the best-performing model. For its twice-weekly initializations, the nearest available initial field to LD20 is used to ensure sample completeness. Given the sample sizes (SC: 28, WC: 8, CC: 24), the 50th percentile serves as the threshold, yielding 14/14, 4/4, and 12/12 high-/low-skill events, respectively. Composites of observations and LD20 forecasts are then constructed to diagnose the evolution of mid-latitude circulation anomalies (Figs. 11, Fig. S1).

For SC-type events, high-skill composites at LD20 feature a prominent Ural blocking high. This anomaly extends eastward, reinforcing the Siberian High and steering cold air intrusions southward into eastern China. FENGSHUN captures this evolution reasonably well, albeit with a westward-shifted positive GPH anomaly and underestimated circulation intensity. Conversely, low-skill events lack a coherent Ural blocking; instead, they exhibit weak, alternating GPH anomalies over the North Atlantic. In forecasts, these weak anomalies decay rapidly, failing to form the typical SC-type dipole structure. This divergence is quantitatively reflected in the Z500 ACCu, which remains >0.5 at LD20 for high-skill events but drops to -0.38 for low-skill events.

Similar predictability dichotomies characterize WC- and CC-type events. For WC-type, high-skill observations show a North Atlantic dipole-like GPH anomaly that triggers the downstream positive anomaly inducing the PWCE. For CC-type, high-skill events originate with a positive GPH anomaly over Northern Europe that propagates eastward along a great-circle path as a planetary-scale wave train, accompanied by a significant western Pacific subtropical high. FENGSHUN generally reproduces these upstream precursors and downstream wave train evolutions, despite inevitable intensity underestimation. In contrast, low-skill events exhibit chaotic and displaced upstream precursors in observations, preventing realistic simulation. For example, in low-skill CC-type forecasts beyond LD10, the GPH

anomaly weakens and fails to propagate equatorward; consequently, the wave train is deflected poleward, failing to establish the coherent meridional dipole pattern over the continents.

Overall, high-skill events (Fig. 11) are consistently preceded by robust and well-organized blocking highs. The associated positive GPH anomalies exhibit coherent eastward and southward development, effectively guiding the cold air intrusions. FENGSHUN largely captures these observed spatiotemporal evolutions; although the simulated intensities are somewhat underestimated, the model successfully reproduces the key circulation modes up to 20 days in advance. Based on the Z500 spatial patterns and ACCu metrics, CC-type events demonstrate the highest overall circulation predictability among the three categories, closely linked to their highly systematic wave train structures.

In contrast, low-skill events are generally associated with weaker, less coherent, and spatially displaced upstream circulation anomalies (Fig. S6). These signals decay rapidly in the FENGSHUN forecasts, with increasing amplitude and phase errors beyond LD10, preventing the development of the observed downstream circulation patterns.

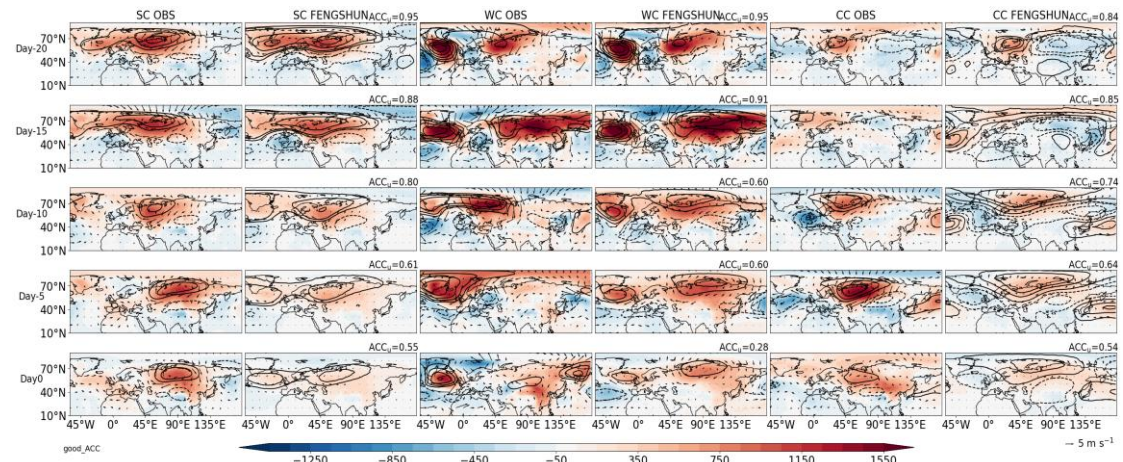


Fig. 11 Lead-lag composites of mid- to high-latitude circulation anomalies for high-skill PWCEs from the FENGSHUN model. For each event type, the left and right panels show the observed and corresponding FENGSHUN forecast composites for the same events, respectively. Shading, vectors, and contours indicate sea level pressure anomalies, 850-hPa wind anomalies, and 500-hPa geopotential height anomalies. The upper-right value in each forecast panel indicates the Z500 ACCu.

The evolution characteristics of tropical convective and low-level circulation anomalies associated with high-skill and low-skill PWCE events are further presented in Figures 12 and Fig. S7 respectively to identify the tropical sources of predictability. High-skill events are consistently characterized by structured, large-scale, eastward-propagating convective anomalies. Specifically, for high-skill WC-type events, the primary observed precursor is strong convective suppression over the eastern Indian Ocean and Maritime Continent (MC), which slowly propagates eastward from Day -20 to event onset. For high-skill CC-type events, vigorous convection over the South China Sea and MC, as well as suppressed convection over the Indian Ocean, serve as crucial precursors. These early-stage anomalies propagate eastward, dynamically forcing the establishment of the western North Pacific anticyclone (WNPAC). FENGSHUN successfully captures these preceding convective signals and their subsequent large-scale evolution in high-skill events (particularly for WC- and CC-types), maintaining an ACCu of 0.47 at Day 0.

Conversely, low-skill events exhibit much weaker and spatially dispersed convective precursors.

For example, the initial suppression signals over the MC for WC-type and the Indian Ocean for CC-type are significantly diminished. Consequently, FENGSHUN struggles to simulate the subsequent realistic convection and circulation patterns, as reflected by substantially lower ACCu values. It is worth noting that although the high-/low-skill classifications are defined based on FENGSHUN, these distinct convective characteristics are generally consistent across the other dynamical models (figures omitted).

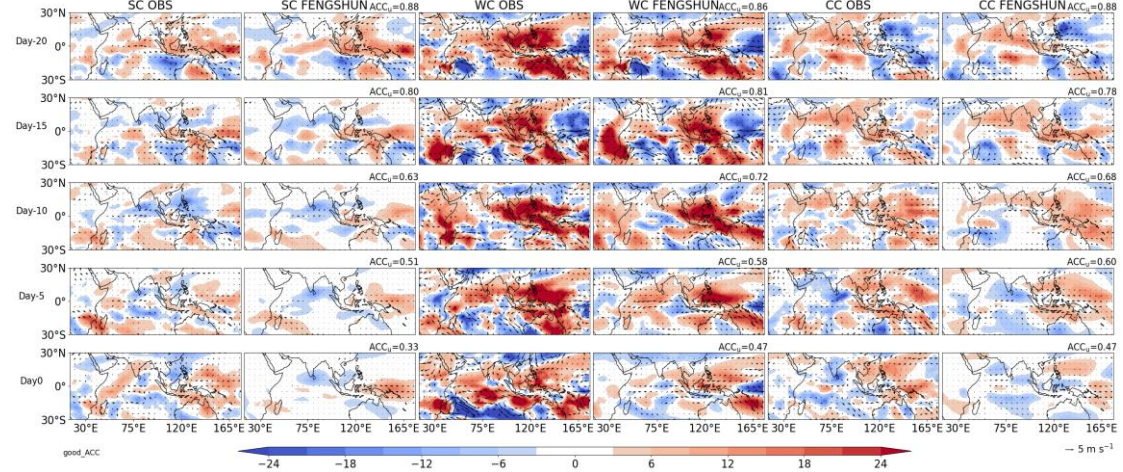


Fig. 12 Lead-lag composites of low-latitude circulation anomalies for high-skill PWCEs from the FENGSHUN model. For each event type, the left and right panels show the observed and corresponding FENGSHUN forecast composites for the same set of high-skill events, respectively. Shading denotes outgoing longwave radiation anomalies, and vectors denote 850-hPa wind anomalies. The upper-right value in each forecast panel indicates the OLR ACCu.

Building on the identified roles of mid- to high-latitude circulation and tropical convection as predictability sources, Figure 13 quantifies their linkages to forecast skill. The results reveal a robust linear relationship between high-latitude Z500 and near-surface air temperature, particularly for CC- and SC-type events, explaining their relatively high predictability. Conversely, the connection between low-latitude convective activity (OLR) and precipitation is relatively weaker, reflecting the multiscale and nonlinear nature of precipitation processes. Among the models, ECMWF exhibits the highest correlation levels, followed by CMA-CPSv3 and NCEP, implying that ECMWF more effectively translates large-scale circulation and convective signals into skillful forecasts of key PWCE variables. Notably, FENGSHUN's forecast skill shows weak correlation with these physical predictability sources. This decoupling likely reflects its fundamental nature as a data-driven AI model, which operates independently of the physical cause-effect chains governing traditional dynamical models.

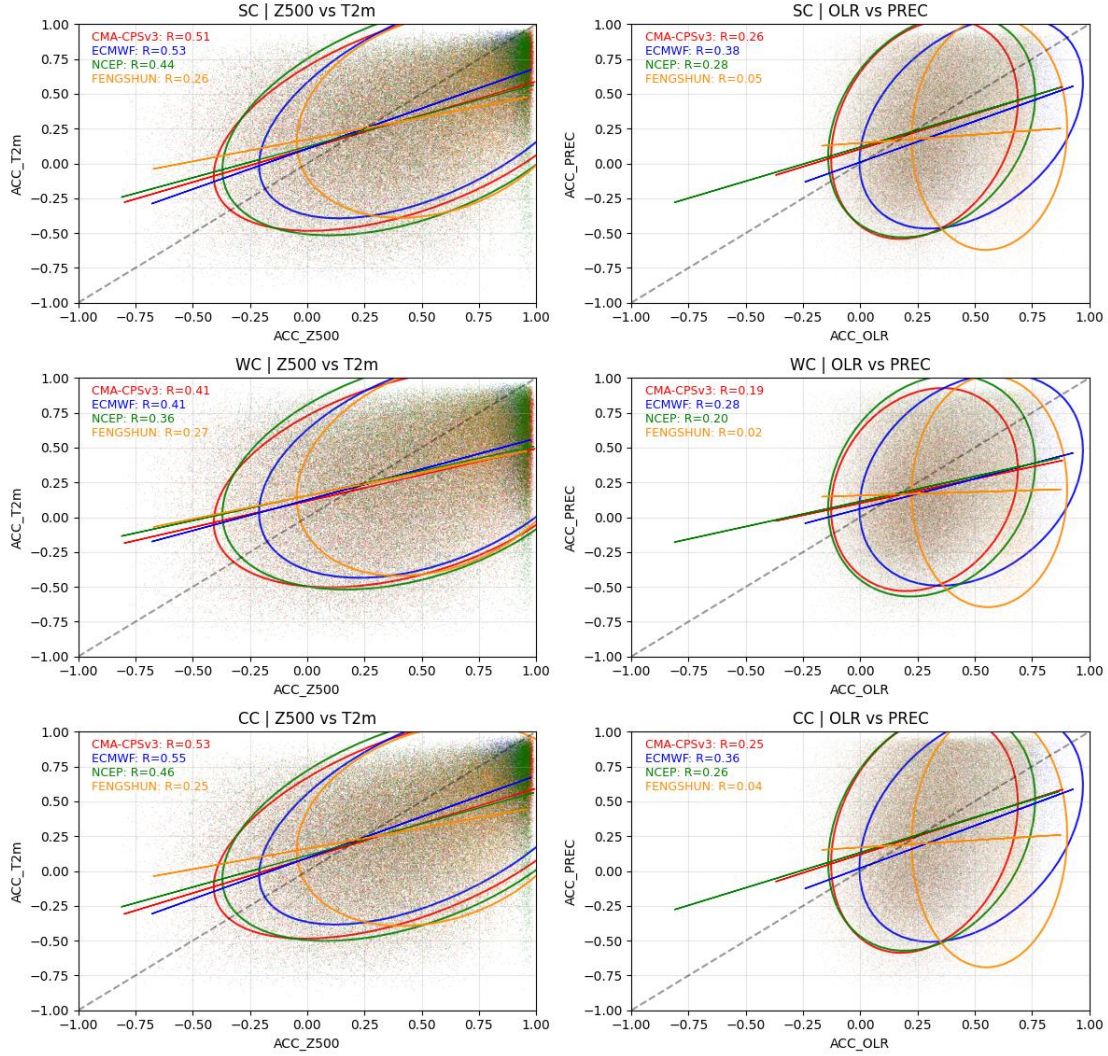


Fig. 13 Scatterplots of ACCu between model and observed fields for Z500–T2m (left column) and OLR–PREC (right column) over the primary occurrence regions of the three PWCE types (SC, WC, and CC) before bias calibration. Each row corresponds to one event type, from top to bottom: SC, WC, and CC. Dots represent all available samples during the extended winter season, namely November–March (NDJFM), for lead times of 0–21 days ($\text{Lead} \leq 21$), with different colors indicating different S2S models. Solid lines denote linear regression fits for each model, and ellipses illustrate the two-dimensional sample distributions. The value shown in each panel is the correlation coefficient (R) between the two ACCu series for the corresponding model, quantifying the coherence between the two prediction skills.

5 Conclusions and Discussions

Based on the three types of persistent wet–cold compound events (PWCEs; SC, WC, and CC) over eastern China identified by Wang et al. (2026), this study systematically has evaluated the subseasonal predictability of PWCEs and explored the possible predictability sources using four dynamical and AI-based S2S prediction systems: CMA-CPSv3, ECMWF, NCEP, and FENGSHUN. By adopting an event-oriented evaluation framework, this study has examined prediction skills in terms of both the event-aggregate physical fields and the spatial matching of individual event occurrences, thereby providing a comprehensive assessment of PWCE predictability and its potential physical origins.

Results indicate that CDF-based calibration substantially enhances the fidelity of S2S forecasts in representing PWCE occurrence frequency, spatial distributions, and wet–cold anomaly (PT) structures. Post-calibration, both dynamical and AI-based models effectively capture the aggregated spatial patterns of PWCE frequency and PT values, with recognizable signals persisting up to 15–20 lead days. Notably, the FENGSHUN model exhibits the slowest skill decay and the most robust representation of wet–cold anomaly structures at extended lead times.

Regarding the predictability of individual PWCE events, skillful forecasts are generally achievable at lead times of 10–15 days, though with pronounced dependencies on event type and model. Specifically, models exhibit higher prediction skill for the continuous wet–cold intensity (PT) fields of WC-type events, whereas they demonstrate stronger capability in capturing the event-area matching metric of SC- and CC-type events. Furthermore, model performance exhibits distinct lead-time-dependent complementarities: the ECMWF model shows superior skill during the short range (within 5 days), while the FENGSHUN model maintains a clear advantage at intermediate to extended lead times (5–20 days).

S2S models effectively capture the key mid- to high-latitude circulation and tropical convection anomalies associated with PWCEs at lead times up to 10–15 days. Among the three categories, the coherent mid- to high-latitude wave train and the suppressed convection over the Maritime Continent characterizing CC-type events are most reliably simulated, consistent with their highest overall predictability. High-skill events are consistently preceded by robust and well-organized precursors, specifically Ural blocking for SC-type, North Atlantic dipole-like anomalies triggering downstream development for WC-type, and a systematic planetary-scale wave train coupled with organized convective anomalies over the Maritime Continent for CC-type. These findings underscore that persistent and spatially coherent circulation and convective precursors constitute primary sources of PWCE predictability. Notably, the linkage between large-scale physical signals and forecast skill is robust for dynamical models, whereas FENGSHUN’s skill is largely decoupled from these physical predictability sources, reflecting its data-driven nature. Ultimately, temperature predictability primarily originates from mid- to high-latitude circulation regimes, while precipitation predictability relies more heavily on tropical convection patterns.

Although this study systematically evaluates the subseasonal predictability of PWCEs and their associated circulation–convection features, several issues merit further study. First, the CDF-based calibration effectively corrects the marginal distributions of temperature and precipitation; however, it does not explicitly constrain the inter-variable covariance critical for compound extremes. Because PWCE predictability fundamentally depends on the coupled behavior of wet and cold anomalies, future bias correction strategies should prioritize preserving this physical dependency as well as multi-model ensemble forecast information. Furthermore, the calibrated skill may be somewhat inflated by the lack of strict out-of-sample verification. Second, the inherent differences in hindcast periods, ensemble sizes, and initialization configurations among the S2S models preclude a strict ranking of model performance. Instead, the observed complementary strengths—ECMWF’s advantage at short lead times driven by superior initialization, versus FENGSHUN’s sustained skill at medium-to-long ranges due to its data-driven architecture—highlight that different models may capture distinct aspects of PWCE predictability. Third, our composite analysis reveals that high-skill events are consistently preceded by robust precursors, whereas low-skill events suffer from inherently disorganized circulation backgrounds. This strong flow-regime dependence suggests that future efforts should shift from composite diagnostics to process-oriented evaluations, specifically examining how models simulate the dynamical coupling between mid- to high-latitude wave trains and tropical convection. Notably, the decoupling of

FENGSHUN’s forecast skill from these physical predictors poses a critical challenge: unraveling the “black-box” mechanisms by which AI models implicitly represent these teleconnections is essential for building trustworthy AI-based forecasting systems. Finally, given the inherently low TS values even at short lead times, deterministic event-area matching remains highly challenging; thus, future evaluations must transition toward probabilistic frameworks to better quantify event risk. Ultimately, developing hybrid dynamical–AI approaches—leveraging the physical consistency of dynamical models to constrain data-driven architectures—may provide a transformative pathway for advancing subseasonal PWCE prediction.

References

- Bi K, Xie L, Zhang H, et al (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619:533–538. <https://doi.org/10.1038/s41586-023-06185-3>
- Bodnar C, Bruinsma WP, Lucic A, et al (2025) A foundation model for the Earth system. *Nature* 641:1180–1187. <https://doi.org/10.1038/s41586-025-09005-y>
- Brunet G, Shapiro M, Hoskins B, et al (2010) Collaboration of the Weather and Climate Communities to Advance Subseasonal-to-Seasonal Prediction. *Bull Amer Meteor Soc* 91:1397–1406. <https://doi.org/10.1175/2010BAMS3013.1>
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics* 3:1–27. <https://doi.org/10.1080/03610927408827101>
- Camps-Valls G, Fernández-Torres M-Á, Cohrs K-H, et al (2025) Artificial intelligence for modeling and understanding extreme weather and climate events. *Nat Commun* 16:1919. <https://doi.org/10.1038/s41467-025-56573-8>
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate* 28:6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Charlton-Perez AJ, Dacre HF, Driscoll S, et al (2024) Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán. *npj Clim Atmos Sci* 7:93. <https://doi.org/10.1038/s41612-024-00638-w>
- Chen L, Zhong X, Li H, et al (2024) A machine learning model that outperforms conventional global subseasonal forecast models. *Nat Commun* 15:6425. <https://doi.org/10.1038/s41467-024-50714-1>
- Davini P, Cagnazzo C, Gualdi S, Navarra A (2012) Bidimensional Diagnostics, Variability, and Trends of Northern Hemisphere Blocking. *Journal of Climate* 25:6496–6509. <https://doi.org/10.1175/JCLI-D-12-00032.1>
- De Andrade FM, Coelho CAS, Cavalcanti IFA (2019) Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Clim Dyn* 52:5451–5475. <https://doi.org/10.1007/s00382-018-4457-z>
- Ding Y, Wang Z, Song Y, Zhang J (2008) Causes of the unprecedented freezing disaster in January 2008 and its possible association with the global warming. *Acta Meteorol Sin* 66:808–825. <https://doi.org/10.11676/qxxb2008.074> (in Chinese)
- Domeisen DIV, Butler AH, Charlton-Perez AJ, et al (2020) The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 1. Predictability of the Stratosphere. *JGR Atmospheres* 125:e2019JD030920. <https://doi.org/10.1029/2019JD030920>
- Domeisen DIV, Eltahir EAB, Fischer EM, et al (2022a) Prediction and projection of heatwaves. *Nat Rev Earth Environ* 4:36–50. <https://doi.org/10.1038/s43017-022-00371-z>
- Domeisen DIV, White CJ, Afargan-Gerstman H, et al (2022b) Advances in the Subseasonal Prediction of Extreme Events: Relevant Case Studies across the Globe. *Bulletin of the American Meteorological Society* 103:E1473–E1501. <https://doi.org/10.1175/BAMS-D-20-0221.1>
- Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol Earth Syst Sci* 16:3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>
- Hoskins B (2013) The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quart J Royal Meteor Soc* 139:573–584.

<https://doi.org/10.1002/qj.1991>
 Ivanov MA, Kotlarski S (2017) Assessing distribution-based climate model bias correction methods over an alpine domain: added value and limitations. *Intl Journal of Climatology* 37:2633–2653. <https://doi.org/10.1002/joc.4870>
 Ketchen Jr. DJ, Shook CL (1996) THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE. *Strat Mgmt J* 17:441–458. [https://doi.org/10.1002/\(SICI\)1097-0266\(199606\)17:6%253C441::AID-SMJ819%253E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6%253C441::AID-SMJ819%253E3.0.CO;2-G)
 Kochkov D, Yuval J, Langmore I, et al (2024) Neural general circulation models for weather and climate. *Nature* 632:1060–1066. <https://doi.org/10.1038/s41586-024-07744-y>
 Lam R, Sanchez-Gonzalez A, Willson M, et al (2023) Learning skillful medium-range global weather forecasting. *Science* 382:1416–1421. <https://doi.org/10.1126/science.adi2336>
 Lee J-Y, Wang B (2014) Future change of global monsoon in the CMIP5. *Clim Dyn* 42:101–119. <https://doi.org/10.1007/s00382-012-1564-0>
 Li F, Orsolini YJ, Keenlyside N, et al (2019) Impact of Snow Initialization in Subseasonal-to-Seasonal Winter Forecasts With the Norwegian Climate Prediction Model. *JGR Atmospheres* 124:10033–10048. <https://doi.org/10.1029/2019JD030903>
 Li S, Robertson AW (2015) Evaluation of Submonthly Precipitation Forecast Skill from Global Ensemble Prediction Systems. *Monthly Weather Review* 143:2871–2889. <https://doi.org/10.1175/MWR-D-14-00277.1>
 Luo M, Lau N, Liu Z, et al (2022) An Observational Investigation of Spatiotemporally Contiguous Heatwaves in China From a 3D Perspective. *Geophysical Research Letters* 49:e2022GL097714. <https://doi.org/10.1029/2022GL097714>
 Maraun D (2013) Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *Journal of Climate* 26:2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>
 Maraun D, Shepherd TG, Widmann M, et al (2017) Towards process-informed bias correction of climate change simulations. *Nature Clim Change* 7:764–773. <https://doi.org/10.1038/nclimate3418>
 Mariotti A, Baggett C, Barnes EA, et al (2020) Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bulletin of the American Meteorological Society* 101:E608–E625. <https://doi.org/10.1175/BAMS-D-18-0326.1>
 Masato G, Hoskins BJ, Woollings T (2013) Winter and Summer Northern Hemisphere Blocking in CMIP5 Models. *Journal of Climate* 26:7044–7059. <https://doi.org/10.1175/JCLI-D-12-00466.1>
 Massonnet F, Goessling H, Yuan X, et al (2026) The role of sea ice in subseasonal to seasonal predictability. In: *Sub-seasonal to Seasonal Prediction*. Elsevier, pp 321–354. <https://doi.org/10.1016/B978-0-443-31538-1.00005-1>
 Peng J-B, Bueh C, Xie Z-W (2021) Extensive Cold-Precipitation-Freezing Events in Southern China and Their Circulation Characteristics. *Adv Atmos Sci* 38:81–97. <https://doi.org/10.1007/s00376-020-0117-4>
 Price I, Sanchez-Gonzalez A, Alet F, et al (2025) Probabilistic weather forecasting with machine learning. *Nature* 637:84–90. <https://doi.org/10.1038/s41586-024-08252-9>
 Rivoire P, Martius O, Naveau P, Tuel A (2023) Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over Europe. *Nat Hazards Earth Syst Sci* 23:2857–2871. <https://doi.org/10.5194/nhess-23-2857-2023>
 Robertson AW, Vitart F, Camargo SJ (2020) Subseasonal to Seasonal Prediction of Weather to Climate

with Application to Tropical Cyclones. *JGR Atmospheres* 125:e2018JD029375.
<https://doi.org/10.1029/2018JD029375>

Sreeparvathy V, Debdut S, Mishra A (2025) A Review of Advances in Flash Drought Research: Challenges and Future Directions. *Earth's Future* 13:e2025EF006603.
<https://doi.org/10.1029/2025EF006603>

Stan C, Zheng C, Chang EK-M, et al (2022) Advances in the Prediction of MJO Teleconnections in the S2S Forecast Systems. *Bulletin of the American Meteorological Society* 103:E1426–E1447.
<https://doi.org/10.1175/BAMS-D-21-0130.1>

Sun B, Wang H, Zhou B (2019) Climatic Condition and Synoptic Regimes of Two Intense Snowfall Events in Eastern China and Implications for Climate Variability. *JGR Atmospheres* 124:926–941.
<https://doi.org/10.1029/2018JD029921>

Sun J, Zhao S (2010) The Impacts of Multiscale Weather Systems on Freezing Rain and Snowstorms over Southern China. *Weather and Forecasting* 25:388–407.
<https://doi.org/10.1175/2009WAF2222253.1>

Themeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change* 112:449–468. <https://doi.org/10.1007/s10584-011-0224-4>

Vitart F, Ardilouze C, Bonet A, et al (2017) The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society* 98:163–173.
<https://doi.org/10.1175/BAMS-D-16-0017.1>

Vitart F, Robertson AW (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Clim Atmos Sci* 1:3. <https://doi.org/10.1038/s41612-018-0013-0>

Vitart F, Robertson AW, Brookshaw A, et al (2025) The WWRP/WCRP S2S Project and Its Achievements. *Bulletin of the American Meteorological Society* 106:E791–E808.
<https://doi.org/10.1175/BAMS-D-24-0047.1>

Wang J, Wu J, Ren H-L, Chen L (2026) Diversity and Formation of Persistent Wet–Cold Compound Extreme Events in China. *J Meteorol Res* 40:341–360. <https://doi.org/10.1007/s13351-026-5209-3>

Wang M, Li X (2024) The Joint Influences of the South Asian Jet Wave Train and Eurasian Teleconnection on the Wet–Cold Extreme over South China in February 2022. *Journal of Climate* 37:6235–6252. <https://doi.org/10.1175/JCLI-D-23-0457.1>

Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences*, 4th ed., Academic Press.

Wu J, Jin F (2021) Improving the MJO Forecast of S2S Operation Models by Correcting Their Biases in Linear Dynamics. *Geophysical Research Letters* 48:e2020GL091930.
<https://doi.org/10.1029/2020GL091930>

Xiang B, Harris L, Delworth TL, et al (2022) S2S Prediction in GFDL SPEAR: MJO Diversity and Teleconnections. *Bulletin of the American Meteorological Society* 103:E463–E484.
<https://doi.org/10.1175/BAMS-D-21-0124.1>

Ye Y, Qian C, Dai A, et al (2025) Attribution of a record-breaking cold event in the historically warmest year of 2023 and assessing future risks. *npj Clim Atmos Sci* 8:14. <https://doi.org/10.1038/s41612-024-00886-w>

Zhang H, Qin J, Li Y (2011) Climatic background of cold and wet winter in southern China: part I observational analysis. *Clim Dyn* 37:2335–2354. <https://doi.org/10.1007/s00382-011-1022-4>

Zhang K, Li J, Zhu Z, Li T (2021) Implications from Subseasonal Prediction Skills of the Prolonged Heavy Snow Event over Southern China in Early 2008. *Adv Atmos Sci* 38:1873–1888.

<https://doi.org/10.1007/s00376-021-0402-x>
 Zheng C, Chang EK, Kim H, et al (2019) Subseasonal to Seasonal Prediction of Wintertime Northern Hemisphere Extratropical Cyclone Activity by S2S and NMME Models. *JGR Atmospheres* 124:12057–12077. <https://doi.org/10.1029/2019JD031252>
 Zheng T, Zheng Z, Fan P, et al (2026) The characteristics of heavy rainfall during the rainy season over North China and prediction skill of NCEP S2S. *Atmospheric Research* 330:108626. <https://doi.org/10.1016/j.atmosres.2025.108626>
 Zhu H, Chen H, Zhou Y, Dong X (2019) Evaluation of the subseasonal forecast skill of surface soil moisture in the S2S database. *Atmospheric and Oceanic Science Letters* 12:467–474. <https://doi.org/10.1080/16742834.2019.1663123>
 Zhu H, Wheeler MC, Sobel AH, Hudson D (2014) Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. *Monthly Weather Review* 142:1556–1569. <https://doi.org/10.1175/MWR-D-13-00222.1>
 Zong HF, Bueh CL, Peng JB, Lin D (2022) Combined disaster events of extensive and persistent low temperatures, rain/snow, and freezing in Southern China: objective identification and key features. *Chin J Atmos Sci* 46:1055–1070. <https://doi.org/10.3878/j.issn.1006-9895.2108.21052> (in Chinese)
 Zscheischler J, Martius O, Westra S, et al (2020) A typology of compound weather and climate events. *Nat Rev Earth Environ* 1:333–347. <https://doi.org/10.1038/s43017-020-0060-z>

Statements and Declarations**Funding**

Supported by the Joint Research Project for Meteorological Capacity Improvement (23NLTSZ003), National Key R&D Program of China (2024YFC3013100), Major science and technology project of the Xizang Autonomous Region (XZ202402ZD0006), and China Meteorological Administration (CMA) Youth Innovation Team (CMA2024QN06).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Jiaxu Wang: Methodology, Formal analysis, Investigation, Visualization, Writing—original draft. Jie Wu: Conceptualization, Methodology, Supervision, Writing – review and editing, Funding. Hong-Li Ren: Supervision, Writing – review and editing. Li Guo: Methodology, Writing – review and editing. Lijuan Chen: Writing – review and editing. Chenguang Zhou: Writing – review and editing.

Data Availability

The observational and subseasonal prediction datasets used in this study are publicly available from their respective data providers. The CN05.1 daily precipitation and temperature datasets were obtained from the China Meteorological Administration, and the S2S hindcast datasets were obtained from the corresponding prediction centers and data platforms described in the Data and Methods section. The processed datasets generated during the current study are available from the corresponding author upon reasonable request.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.docx](#)