ARTICLE

SCIENCE CHINA Earth Sciences
earth.scichina.com link.springer.com
https://doi.org/10.1007/s11430-025-1739-8

# Heatwave forecast skill in the mid-high latitudes of the Northern Hemisphere at subseasonal lead times: Regional and duration disparity

Haozhou LIU[1,2], Yimin LIU[1,2*], Qing BAO[3], Bian HE[1,2] & Wansuo DUAN[1,2]

[1] State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[2] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Laboratory of Atmospheric and Oceanic Dynamics (LAOD), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

* Corresponding author (email: lym@lasg.iap.ac.cn)

Received 25 April 2025; Revised 26 October 2025; Accepted 29 October 2025; Published online 11 December 2025

**Abstract** Accurate prediction of heatwaves is critical and remains a key objective of the Sub-seasonal to Seasonal (S2S) Prediction project. However, focusing on extreme cases and neglecting false alarms might overestimate regional forecast skill. Therefore, it is essential to evaluate the overall performance of current S2S models in predicting heatwaves. Using long-term datasets from current multiple subseasonal forecast systems, this study assessed the forecast skill for heatwaves in the mid-high latitudes of the Northern Hemisphere. Analysis revealed that heatwave frequency, intensity, and duration can be predicted up to 2 weeks in advance at the hemispheric scale. The S2S models generally predict the occurrence of heatwave 1–2 weeks in advance and warm anomalies 3–4 weeks ahead. Two notable patterns were identified: regional dependence and a duration effect. The former is exemplified by heatwaves in Europe being more predictable than those in East Asia and North America, which is attributed to higher inherent predictability and stronger model performance. The latter reflects the fact that long-duration heatwaves are more predictable than short-duration ones within the first lead 2 weeks, primarily owing to atmospheric predictability constraints. This study provides a more nuanced understanding of the current capabilities of S2S heatwave forecasts.

**Keywords** Heatwave, Subseasonal to seasonal (S2S), Forecast skill, Regional dependence, Duration effect

## 1. Introduction

Extreme heatwaves have profound impacts on societal productivity and activities (Somanathan et al., 2021), and pose substantial threats to human life, health, and safety (Howard and Krishna, 2022). Moreover, persistent heatwaves create conditions favorable for extreme events such as droughts and wildfires (Jyoteeshkumar reddy et al., 2021; Libonati et al., 2022). When co-occurring with other extreme events, such as heavy precipitation, the societal impacts can be amplified considerably (Zscheischler et al., 2018). Because the occurrence of heatwave events will become increasingly frequent in the future owing to climate change, it is expected that the associated risks will also be exacerbated (Meehl and Tebaldi, 2004; Domeisen et al., 2023). Therefore, the need to improve the ability to predict heatwaves has become urgent. Heatwaves often occur on subseasonal timescales. Sub-seasonal to seasonal (S2S) predictions are inherently complex owing to the influence of initial conditions and predictable signals from longer-term external forcing, a challenge often called the "predictability desert" (Vitart et al., 2012). Nevertheless, bridging this forecasting gap is essential in achieving seamless prediction systems (COPES, 2005). As part of this goal, the S2S Prediction project prioritizes improving the prediction accuracy of high-impact extreme events (Vitart and Robertson, 2018).

The onset of a heatwave is usually attributable to the occurrence of a local anticyclonic anomaly within the synoptic weather time scale. In such a situation, the combined effects of intensified clear-sky solar shortwave radiation and subsidence-induced adiabatic warming result in a notable positive surface air temperature anomaly (Barriopedro et al., 2023; Domeisen et al., 2023). Expanding to broader temporal and spatial scales, local circulation anomalies during heatwaves are often driven by Rossby wave activity and persistent atmospheric blocking systems (Kautz et al., 2022). When the timescale extends further to the subseasonal range or beyond, slow-evolving processes with memory, such as soil moisture and sea surface temperature, begin to exert influence (Jiang et al., 2023). Robertson and Vitart (2018) and Sciences et al. (2016) categorized the sources of subseasonal predictability signals. These processes encompass tropical and subtropical quasi-oscillatory variability, memory-driven factors such as land, ocean, sea ice, and stratosphere interactions, and external forcings like anthropogenic influences and volcanic activity. While these factors contribute to the development of heatwaves, they also provide stable and predictable signals that serve as valuable sources of predictability for subseasonal heatwave forecasts.

Pyrina and Domeisen (2023) examined the prediction skills of the European Centre for Medium-Range Weather Forecasts (ECMWF) S2S model in terms of the regional heatwave

**Citation:** Liu H, Liu Y, Bao Q, He B, Duan W. 2026. Heatwave forecast skill in the mid-high latitudes of the Northern Hemisphere at subseasonal lead times: Regional and duration disparity. Science China Earth Sciences, https://doi.org/10.1007/s11430-025-1739-8

characteristics in Europe, including their occurrence, duration, and intensity, highlighting a local predictability range of 1–3 weeks. However, research exploring the prediction skills of heatwave characteristics from the perspective of a broader spatial distribution remains lacking. Existing studies have illustrated that S2S models can predict the occurrence of heatwaves approximately 2 weeks in advance and forecast surface air temperature warm anomalies up to 4 weeks ahead or even longer (Vitart and Robertson, 2018; Domeisen et al., 2022). Notably, most of those studies focused on observed extreme heatwave events, possibly neglecting "normal" heatwave events and failing to account for false alarms. Since S2S models often exhibit higher forecast skill for extreme heatwaves (Wulff and Domeisen, 2019), this focus may inevitably lead to an overestimation of the capability of a model in forecasting heatwaves. Therefore, it is important to determine the extent to which current S2S models have demonstrated skill in heatwave prediction.

Current S2S models exhibit varying skill in forecasting heatwaves across different regions (Ford et al., 2018; Pyrina and Domeisen, 2023). As mentioned earlier, S2S models tend to show higher forecast skill for extreme heatwaves that last longer and are more intense. However, the underlying reasons for such forecast skill differences require further investigation. The objective of our study is to comprehensively evaluate the forecast skill of multiple current S2S models in predicting heatwaves using their long-term reforecast data, and to explore the root causes of the skill differences from the fresh perspective of predictability. The evaluation incorporates both deterministic forecast skill of the spatial distribution of heatwave characteristics and probabilistic accuracy of heatwave day predictions. Then, factors responsible for regional dependence and duration effect in forecast skill are analyzed.

## 2. Data and methods

### 2.1 Data

#### 2.1.1 Reforecast data

The S2S database currently contains 13 numerical models. However, the reforecast periods and frequencies vary across models (Vitart et al., 2017), making it a challenge to systematically evaluate the overall performance of current models in predicting heatwaves. To maximize the number of models included in the evaluation while avoiding bias owing to differing reforecast periods, we selected models with a reforecast span of at least 15 years within the period 2001–2020. Additionally, because the predictability of heatwaves typically extends up to one month (Domeisen et al., 2022), models with a low reforecast frequency might lead to inaccurate evaluations. Therefore, a reforecast frequency of at least four times per month was required for the models included in this study. Consequently, the following 10 models were selected for evaluation: CMA, CNR-ISAC, CNRM, CPTEC, ECCC, ECMWF, HMCR, IAP-CAS, KMA, and UKMO (Appendix Table S1). The variables used for the calculation include maximum 6-hourly temperature at 2-m height ($T2m$), daily instantaneous 500-hPa geopotential height ($Z_{500}$), and daily mean soil moisture in the top 100 cm ($SM_{100}$).

To ensure fairness and convenience in subsequent comparisons between different S2S models, we first reprocessed the reforecast data from each S2S model into daily reforecast data (Yang et al., 2018). The core idea of this method is to move a window to average the values near a missing lead time, thereby replacing the missing value. Xie et al. (2020) and Yang et al. (2018) have demonstrated that this reprocessing approach does not affect the calculated forecast skill. Specifically, for reforecast frequencies of 2 per week, 1 every 5 days, 1 every 7 days, and 4 per month, we used sliding window lengths of 5, 7, and 9 days, respectively. For computational simplicity and ease of comparison, we calculated the average of lead time intervals of 4–10, 11–17, 18–24, and 25–31 days as the forecast approximation for lead times of 1–4 weeks, respectively. This approach is particularly suitable for handling model climatology because of the lack of actual heatwave events available to serve as a reference regarding which day's forecast should be selected for evaluation.

#### 2.1.2 Observation and reanalysis datasets

Single reanalysis data might have a potential favor phenomenon on the individual model in a skill comparison (Wulff and Domeisen, 2019). Therefore, daily maximum air temperature at 2-m height ($TX$) from the ECMWF Reanalysis version 5 (ERA5) dataset (Hersbach et al., 2020), Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2) reanalysis (Gelaro et al., 2017), and Berkeley Earth Surface Temperature (BEST) observations (Rohde et al., 2013) were used. Additionally, hourly circulation and land data from ERA5 were used to investigate the causes of differences in forecast skill. These variables included geopotential height, air temperature, vertical pressure velocity, zonal wind, and meridional wind from 1000 to 10 hPa, together with soil moisture data from the top three layers (0–7, 7–28, and 28–100 cm) that were linearly weighted and summed to derive $SM_{100}$. The spatial resolutions of the datasets are as follows: ERA5 (0.25°×0.25°), MERRA2 (0.5°×0.625°), and BEST (1°×1°). All hourly data were first averaged to daily resolution for subsequent processing. This study focused on the mid-high-latitude summer (June-July-August, JJA) in the Northern Hemisphere from 2001 to 2020.

#### 2.1.3 Other data

Gridded population data from the Gridded Population of the World, version 4 (resolution: approximately 5 km×5 km) dataset (Center for International Earth Science Information Network-CIESIN-Columbia University, 2017) were used to identify social impact hotspots in the Northern Hemisphere. The global observation station distribution data were obtained from the Observing Systems Capability Analysis and Review Tool, which is a component of the World Meteorological Organization Integrated Global Observing System framework (https://oscar.wmo.int/surface/).

### 2.2 Definition

#### 2.2.1 Heatwave definition

Current S2S models tend to underestimate the intensity of heatwaves (Xie et al., 2020; Pyrina and Domeisen, 2023), which means that defining heatwaves using an absolute threshold method is unsuitable for these models. In this study, a heatwave at each grid point on land was defined as a period of at least 3 consecutive days during which the $TX$ exceeded the 90th percentile, calculated using a 15-day centered moving window based on the 2001–2020 climatology during JJA, which is similar to the approach adopted in previous studies (Mondal and

Mishra, 2021; Zhang et al., 2023). Brunner and Voigt (2024) recommended removing the mean seasonal cycle before calculating the extreme threshold to avoid time-, region-, and dataset-dependent biases, which can lead to underestimation of extreme frequencies. Therefore, the mean seasonal cycle was removed before calculating the extreme threshold. Furthermore, because the definition of marine heatwaves differs from that of land heatwaves (Hobday et al., 2016; Oliver et al., 2018), all heatwaves in this paper refer specifically to land heatwaves.

Heatwave characteristics include frequency, intensity, timing, duration, and spatial extent (Domeisen et al., 2023). This study focused on the frequency, intensity, and duration. Frequency refers to the total number of heatwaves occurring in a summer. Duration includes two metrics: longest duration, defined as the length of the longest heatwave in a summer, and average duration, representing the mean length of all heatwaves occurring in a given summer. Intensity also includes two metrics: maximum intensity, defined as the largest daily maximum temperature anomaly relative to the 90th percentile threshold recorded during heatwave days in a summer, and average intensity, calculated as the mean daily temperature anomaly relative to the same threshold across all heatwave days in a summer (Wang et al., 2024).

As illustrated in Fig. 1, the three datasets used in this study provide consistent depictions on the spatial distribution of heatwave characteristics across the Northern Hemisphere. Heatwaves are less frequent in low-latitude regions and more prevalent in high-latitude regions; however, their frequency in the observational data is generally higher than that in the reanalysis data (Fig. 1B). Moreover, heatwaves in low latitudes are typically characterized by shorter durations and weaker intensities, whereas those in mid- and high-latitudes exhibit longer durations and stronger intensities (Fig. 1D, 1F). This pattern aligns with the broader climatic characteristics of heatwaves in the Northern Hemisphere (Wang et al., 2024). Regions prone to prolonged heatwaves are concentrated primarily in eastern North America, eastern Europe, and central and western Siberia (Fig. 1C). Notably, the MERRA2 dataset displays abnormally long heatwave durations in low-latitude regions compared with those of the ERA5 and BEST datasets (Fig. 1D). Domeisen et al. (2023) highlighted that relative threshold and persistence-based heatwave definitions are less applicable in tropical regions. Thus, subsequent analysis focuses on mid- and high-latitude areas of the Northern Hemisphere (north of 23°N). For brevity, only the spatial distributions of the average duration and the average intensity are presented here; the distributions of longest duration and intensity, which follow similar patterns, are omitted.

#### 2.2.2 *Blocking intensity index*

The blocking intensity index (*BI*) for the two European regions is defined as the ratio of the area-weighted sum of $Z_{500}$ anomalies at grid points experiencing blocking to the total area of that region. The detailed steps of the specific definitions are as follows: (1) Determine whether a grid point is blocked based on its $Z_{500}$ northward and southward gradients, following the method of Scherrer et al. (2006), which is an extension of the method of Tibaldi and Molteni (1990). A grid point is considered blocked if the northward gradient (GHGN) is less than −10 m/(°latitude) and the southward gradient (GHGS) is greater than 0 m/(° latitude) for at least 5 consecutive days. During this period, the $b(t)_{\varphi_0,\lambda}$ of the grid point is assigned a value of 1; otherwise, it is assigned a value of 0. (2) Calculate the blocking intensity $B(t)$ per unit area within a region as a time-dependent function. (3) Normalize $B(t)$ to create a standardized *BI* for consistent comparison.

$$\begin{aligned}&\mathrm{GHGN}=\frac{Z_{500}(\varphi_N)-Z_{500}(\varphi_0)}{\varphi_N-\varphi_0}\\&\mathrm{GHGS}=\frac{Z_{500}(\varphi_0)-Z_{500}(\varphi_S)}{\varphi_0-\varphi_S}\\&\begin{cases}b(t)_{\varphi_0,\lambda}=1,\ \mathrm{GHGN}<-10\ \text{and}\ \mathrm{GHGS}>0\ \text{at least 5 days}\\ b(t)_{\varphi_0,\lambda}=0,\ \text{others}\end{cases}\end{aligned}\tag{1}$$

$$B(t)=\frac{\sum b(t)_{\varphi_0,\lambda}\cdot\cos\varphi_0\cdot Z'_{500\varphi_0,\lambda}(t)}{\sum\cos\varphi_0}\tag{2}$$

where $\varphi_0$ and $\lambda$ represent the latitude and longitude of any grid point between 34.5°N and 75°N, respectively with $|\varphi_{N/S}-\varphi_0|=15°$, and $Z'_{500\varphi_0,\lambda}$ denotes the $Z_{500}$ anomaly at the grid point ($\varphi_0$, $\lambda$) on a specific day relative to its climatological mean state.

### 2.3 Evaluation metrics

The root mean square error (RMSE) and the spatial anomaly correlation coefficient (ACC) were used to evaluate the deterministic forecast skill of the S2S model ensemble mean for heatwave characteristics in the mid- and high-latitudes of the Northern Hemisphere. The RMSE reflects the magnitude of deviation in amplitude, while the ACC measures the performance of models in predicting the spatial distribution of heatwave characteristics. Additionally, the temporal anomaly correlation coefficient (TCC) was employed to assess the forecasting skill of the S2S model for $Z_{500}$ and $\mathrm{SM}_{100}$ anomalies.

To evaluate the probabilistic forecast skill, the Brier Skill Score (BSS) and Continuous Ranked Probability Skill Score (CRPSS) were applied to measure the capability of models in predicting the occurrence of heatwaves. A higher BSS indicates better forecast skill, with a BSS of 0 representing no skill and a BSS of 1 corresponding to a perfect forecast (Wilks, 2019). In combination with Eq. (3), the BSS considers heatwave hits, false alarms, and missed alarms, providing a balanced measure of overall forecast performance:

$$\mathrm{BS}=\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2\tag{3}$$

$$\mathrm{BSS}=\frac{\mathrm{BS}_{\mathrm{cli}}-\mathrm{BS}}{\mathrm{BS}_{\mathrm{cli}}}\tag{4}$$

where $p_i$ and $o_i$ represent the forecast probability and observed frequency of the $i$-th sample, respectively. $N$ is the total number of samples (Brier, 1950), and $\mathrm{BS}_{\mathrm{cli}}$ refers to BS of the climatology.

As shown in Eq. (5), a CRPS value of zero indicates that the forecast probability perfectly matches the observation, corresponding to a perfect forecast (Hersbach, 2000). Higher CRPSS values denote higher forecast skill, while CRPSS $\leqslant 0$ implies that the forecast cannot be significantly distinguished from the historical climatology, and is considered to have no skill. For $\mathrm{CRPS}_{\mathrm{cli}}$, a climatological ensemble of 100 members was

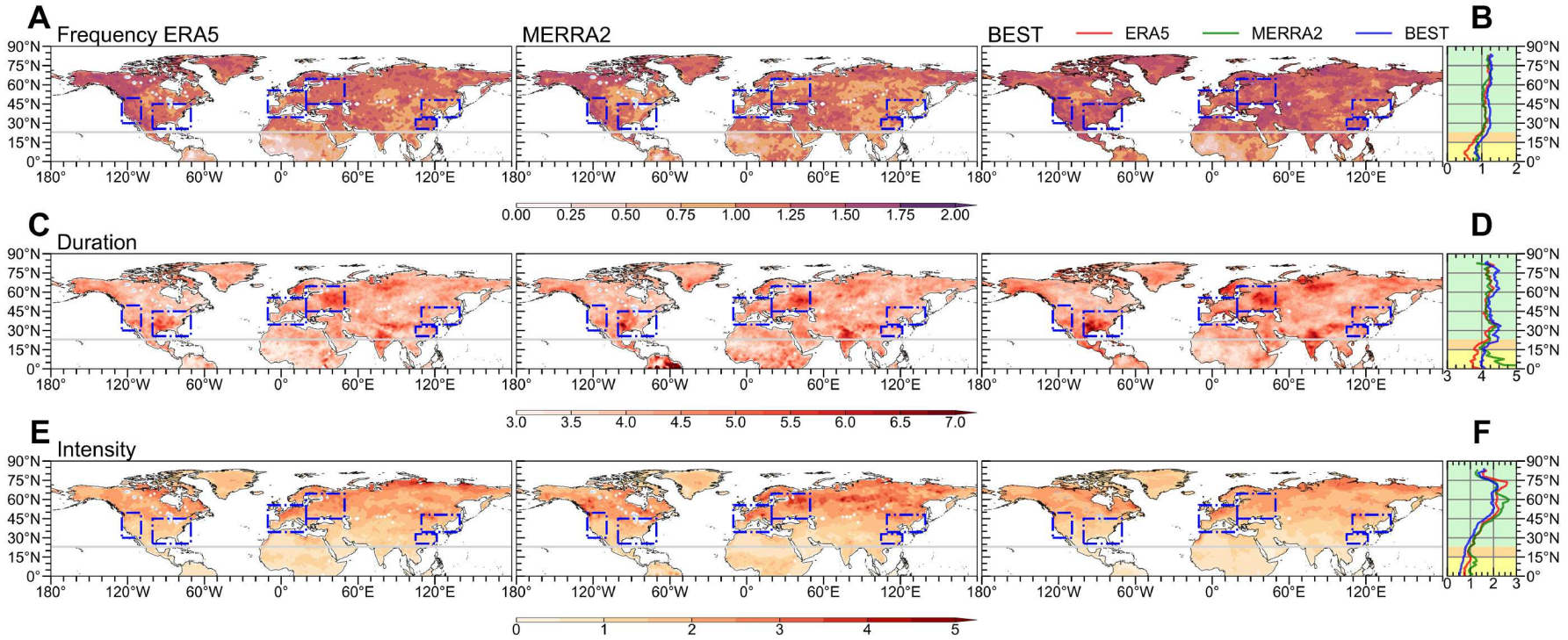


**Fig. 1** Spatial distribution of JJA land heatwave (A) frequency (unit: events/year), (C) average duration (unit: days/year), and (E) average intensity (unit: °C) from ERA5, MERRA2, and BEST (2001–2020). Panels (B), (D), and (F) show the corresponding latitudinal means for ERA5 (red), MERRA2 (green), and BEST (blue). Blue boxes highlight six social impact hotspots, detailed in Section 3.2.1. Gray line marks the latitude of 23°N.

constructed for each forecast target date by taking the ±2-day (5-day) window for each year from 2001 to 2020, based on observations or reanalysis data (Buizza and Leutbecher, 2015; Dai et al., 2021). The heatwave days used to compute CRPSS included not only those corresponding to observed events (i.e., model hits and missed alarms), but also false alarms, defined as cases where more than 50% of the ensemble members forecast a heatwave that was not observed. This inclusion aims to provide a more comprehensive assessment of the S2S models' probabilistic forecast skill for heatwaves.

$$\mathrm{CRPS} = \int_{-\infty}^{+\infty} \left(\mathrm{CDF}_{\mathrm{frc}}(x) - \mathrm{CDF}_{\mathrm{obs}}(x)\right)^2 \mathrm{d}x \tag{5}$$

$$\mathrm{CRPSS} = \frac{\mathrm{CRPS}_{\mathrm{cli}} - \mathrm{CRPS}}{\mathrm{CRPS}_{\mathrm{cli}}} \tag{6}$$

where $\mathrm{CDF}_{\mathrm{frc}}(x)$ and $\mathrm{CDF}_{\mathrm{obs}}(x)$ represent the cumulative distribution function of the ensemble forecast and observations, respectively. $\mathrm{CRPS}_{\mathrm{cli}}$ was computed in the same way as CRPS but using the climatology ensemble members.

## 3. S2S model forecast skill

In this section, we assess the forecast skill of current S2S models in relation to heatwaves from two perspectives: deterministic and probabilistic forecast skill. The former evaluates the ability of the S2S model ensemble mean to reproduce the spatial distribution of observed heatwave characteristics, including frequency, duration, and intensity. The latter focuses on local regions, examining the probabilistic forecast skill of the models' ensemble member distribution for days during heatwaves.

### 3.1 Deterministic forecast skill at the hemispheric scale

Current S2S models exhibit similar deterministic forecast skill patterns for the spatial distribution of heatwave characteristics in the mid- and high-latitudes of the Northern Hemisphere (Fig. 2). From lead week 1 to lead week 2, forecast skill decreases sharply, while the rate of decline slows from lead week 2 to lead week 3. Beyond lead week 3, there is no notable change in forecast skill. To quantify the effective forecast time of the current S2S models for heatwave characteristics, a straightforward statistical test was designed. For a specific model at a given lead time (e.g., lead week 1/2/3/4), the forecast data were randomly replaced with data of any year from the model's historical climatology. If a historical year exists such that the calculated ACC was higher and the RMSE was lower than those of the original forecast, it was assumed that there is no forecast skill for the model in capturing the spatial distribution of heatwave characteristics at that lead time.

Fig. 2 shows that for most S2S models at lead weeks 3–4, the statistical threshold outperforms the original forecast, indicating that S2S models can reliably predict the spatial distribution of heatwave frequency, intensity, and duration in the mid- and high-latitudes of the Northern Hemisphere up to approximately lead week 2. Notably, the spatial distributions of the longest duration and maximum intensity are similar to the average duration and intensity, and different reanalysis datasets do not produce marked discrepancies in the assessment results (Appendix Fig. S1). Additionally, the ECMWF, ECCC, and CNR-ISAC models exhibit superior performance in predicting the spatial distribution of heatwave characteristics across the Northern Hemisphere.

### 3.2 Probabilistic forecast skill at the regional scale

#### 3.2.1 *Area cluster*

The systems influencing heatwaves vary across regions, often leading to distinct heatwave characteristics and corresponding forecast skill levels. Here, we focus on the region to the north of the Tropic of Cancer. As shown in Fig. 1A, 1C, 1E, the characteristics of heatwaves vary markedly across regions. In Western Europe (WEU) and Eastern Europe (EEU), heatwaves are typically defined by low frequency, long duration, and high intensity. In Northern East Asia (NEA) and Central China (CCH), heatwaves occur more frequently, but are generally shorter in duration and weaker in intensity. In North America, heatwave characteristics are mixed: those in Western North America (WNA) resemble those of East Asia with high frequency and shorter duration, while those in the Eastern United States (EUS) exhibit similarity to those in Europe with lower frequency and longer duration. Therefore, it is necessary to perform regional segmentation before examining forecast accuracy.

The prediction of extreme events on the S2S time scale is of considerable socioeconomic value and is a key focus of the S2S project (Vitart and Robertson, 2018). This study selected six densely populated regions within the mid- and high-latitudes of the Northern Hemisphere (Fig. 3) as high-impact hotspots, which have been widely studied in previous research, including WEU (Cai et al., 2024), EEU (Vitart and Robertson, 2018), CCH (Zhang et al., 2023), NEA (Hsu et al., 2020), and WNA and EUS (Domeisen et al., 2022). It is noteworthy that the heatwave characteristics within these six regions show general consistency (Fig. 1).

#### 3.2.2 *Probabilistic skill*

As the forecast lead time increases, the forecast skill for predicting the occurrence of days during heatwaves gradually decreases (Fig. 4A). Overall, the current 10 S2S models can predict the occurrence of heatwaves in WEU, EEU and NEA with a lead time of approximately 2 weeks, with both BSS and CRPSS remaining above zero, and at least 30% of the ensemble members predicting *TX* anomalies exceeding the 90th percentile (Fig. 4A, 4B). However, for CCH, WNA, and EUS, this predictive capability is generally limited to a lead time of 1 week. The skewness and kurtosis of the probability density function (PDF) for the ensemble forecasts gradually decrease as the lead time increases (Fig. 4B). For WEU, EEU, NEA, and CCH, warm anomalies can be predicted approximately 3 weeks in advance, with at least 40% of ensemble members forecasting *TX* anomalies reaching above-normal levels. By contrast, this lead time extends to around 4 weeks in North America (Fig. 4B).

This highlights that previous work on case studies might have overestimated the predictive skill of S2S models in relation to heatwaves. Not all heatwaves can be predicted 2 weeks before they occur; in CCH, WNA, and EUS, this lead time is overestimated by about one week in prior studies. Likewise, warm anomalies cannot all be forecasted at 4-week leads, with skill in Europe and East Asia also overestimated by roughly one week. This finding highlights evident regional variability: higher predictive skill for heatwaves in Europe, particularly WEU, and lower predictive skill in East Asia and North America, especially

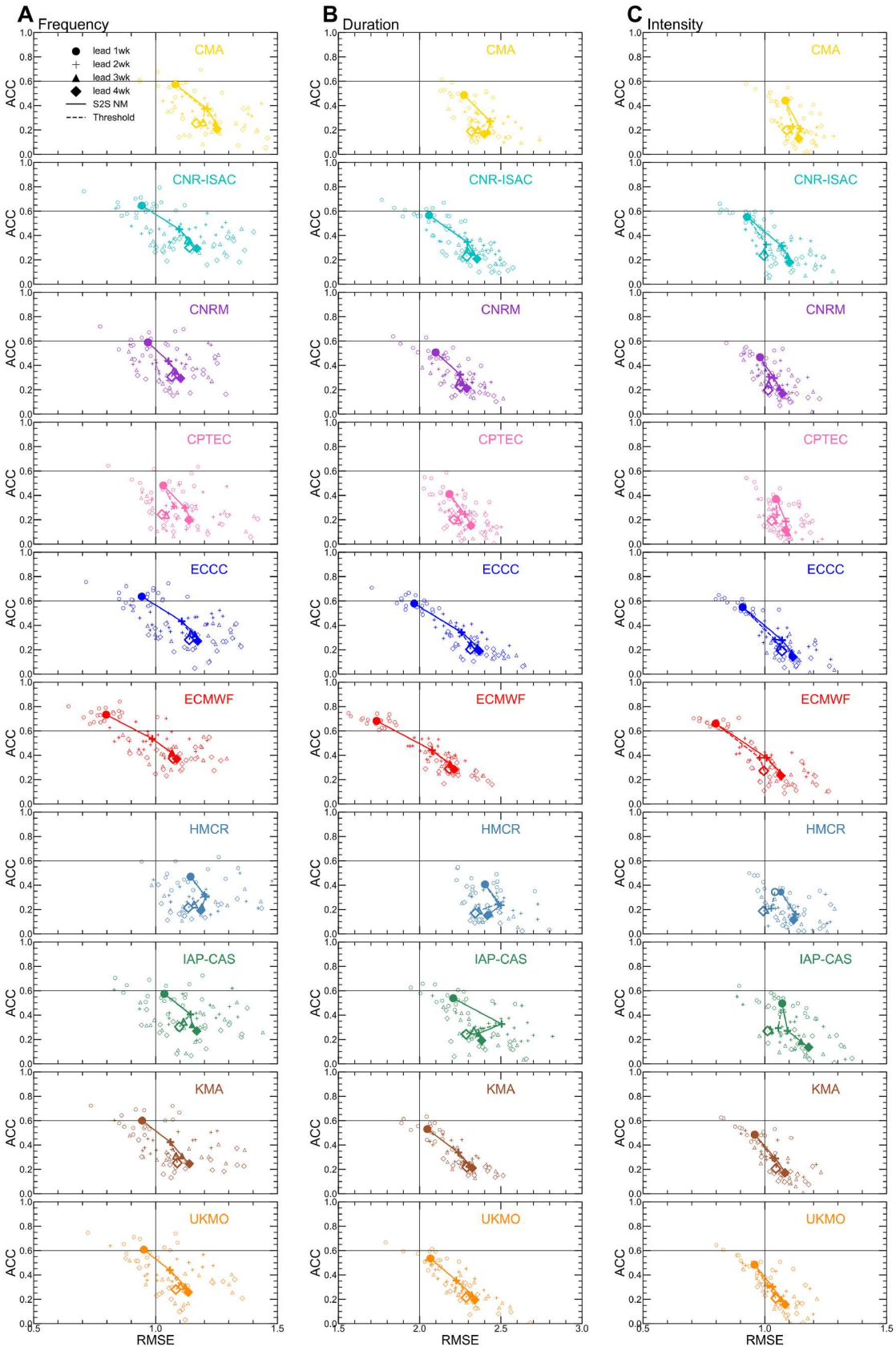


**Fig. 2** ACC and RMSE for heatwave (A) frequency, (B) duration, and (C) intensity across the mid- and high-latitudes of the Northern Hemisphere (north of 23°N) for 10 S2S models compared with the ERA5 reanalysis from lead weeks 1–4 (indicated by circles, plus signs, triangles, and diamonds, respectively). Small hollow markers show individual years, large solid markers denote the climatological mean, and large hollow markers indicate the statistical test thresholds. Solid/dashed lines connect large solid/hollow markers.

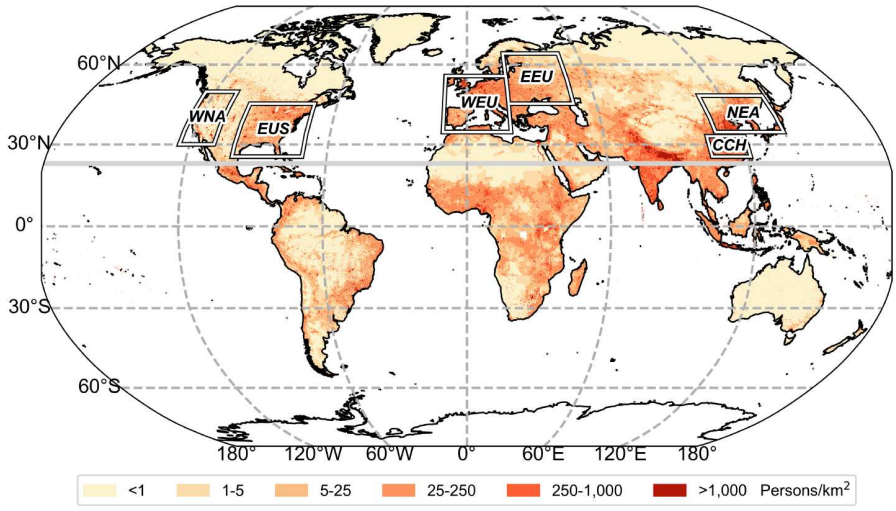


**Fig. 3** Global population density in 2020. Gray line marks the latitude of 23°N. Rectangular boxes represent six selected regions: Western Europe (WEU: 34.5°–55.5°N, 10.5°W–19.5°E), Eastern Europe (EEU: 45°–64.5°N, 19.5°–49.5°E), Northern East Asia (NEA: 34.5°–48°N, 109.5°–139.5°E), Central China (CCH: 25.5°–33°N, 105°–121.5°E), Western North America (WNA: 30°–49.5°N, 109.5°–124.5°W), and the Eastern United States (EUS: 25.5°–45°N, 70.5°–100.5°W).

CCH. To verify the robustness of these conclusions, we also conducted validation using reanalysis data, and the results showed no substantial differences (see Appendix Figs. S2 and S3).

Forecast skill tends to be higher for heatwaves with greater intensity and longer duration (Wulff and Domeisen, 2019). Taking the example of the ECMWF model, which is a model with excellent forecasting performance, most regions exhibit a clear pattern in which heatwaves of longer duration correspond to higher forecast skill at lead times of 1–4 weeks (Fig. 5B). An exception is observed in WNA, which is likely attributable to the limited sample size introducing potential randomness, i.e., only one long-duration heatwave has been recorded in this region. As such, this is not discussed further here. Building on these findings, we defined short-duration heatwaves as those lasting no more than 5 days and long-duration heatwaves as those lasting more than 5 days. With this classification, we were able to assess how well the current S2S models perform in forecasting short-duration and long-duration heatwaves.

As shown in Fig. 6A–6F, notably higher forecast skill is exhibited for long-duration heatwaves than for short-duration heatwaves. This difference is particularly evident at lead times of 1–2 weeks, where the PDF of the ensemble members for long-duration heatwaves demonstrates noticeably higher skewness and kurtosis. However, as the lead time extends to 3–4 weeks, the differences between the two types of heatwaves gradually diminish, with their PDFs overlapping substantially. Specifically, days during long-duration heatwaves can be predicted approximately 2 weeks in advance, while days during short-duration heatwaves can be predicted only approximately 1 week in advance, with at least 30% of the ensemble members forecasting *TX* anomalies above the 90th percentile. Similar conclusions are found using the reanalysis data, as shown in Appendix Figs. S4 and S5.

As discussed previously, the higher forecast skill for heatwaves in Europe could be partially attributed to the dominance of long-duration heatwaves in that region, for which the models inherently tend to exhibit higher forecast skill. However, these observations lead to a deeper question regarding the cause of the regional and duration-based differences in heatwave forecast skill.

## 4. Potential reasons for regional and duration-based variations in S2S forecast skill

Mu et al. (2017) highlighted that forecast skill is an evaluation of predictability, while predictability itself represents the theoretical upper limit of forecast skill. As discussed in Section 3, the effective lead time for forecasting heatwaves is within 1 month. Within this timeframe, the sources of predictability primarily stem from atmospheric and land surface processes (Mariotti et al., 2018). For subseasonal predictability, the two main land surface sources are soil moisture and snowpack (Dirmeyer et al., 2019). Given that this study focused on boreal summer, snowpack, which typically plays a dominant role in spring as a predictability source, might have already transformed into other memory signals, such as soil moisture. Therefore, soil moisture was considered the sole source of land surface predictability in this context. In this section, we explore the causes of forecast skill differences by considering the relationship between forecast skill, predictability limits, and the true value.

### 4.1 Causes of regional differences in forecast skill

As shown in Fig. 5A, it is evident that in WEU and EEU, longer-duration heatwaves are often associated with higher intensity. By contrast, this positive correlation between heatwave duration and intensity is not as pronounced in East Asia and North America, especially NEA and WNA. This observation prompts the following hypothesis: heatwaves in WEU and EEU might be influenced predominantly by a single dominant weather system, where a stronger regime often results in longer-lasting and more intense heatwaves. Additionally, owing to the singular nature of the controlling system, the predictability of such heatwaves tends to be higher. Conversely, heatwaves in East Asia and North America are more likely influenced by multiple competing systems with more complex physical processes. These systems

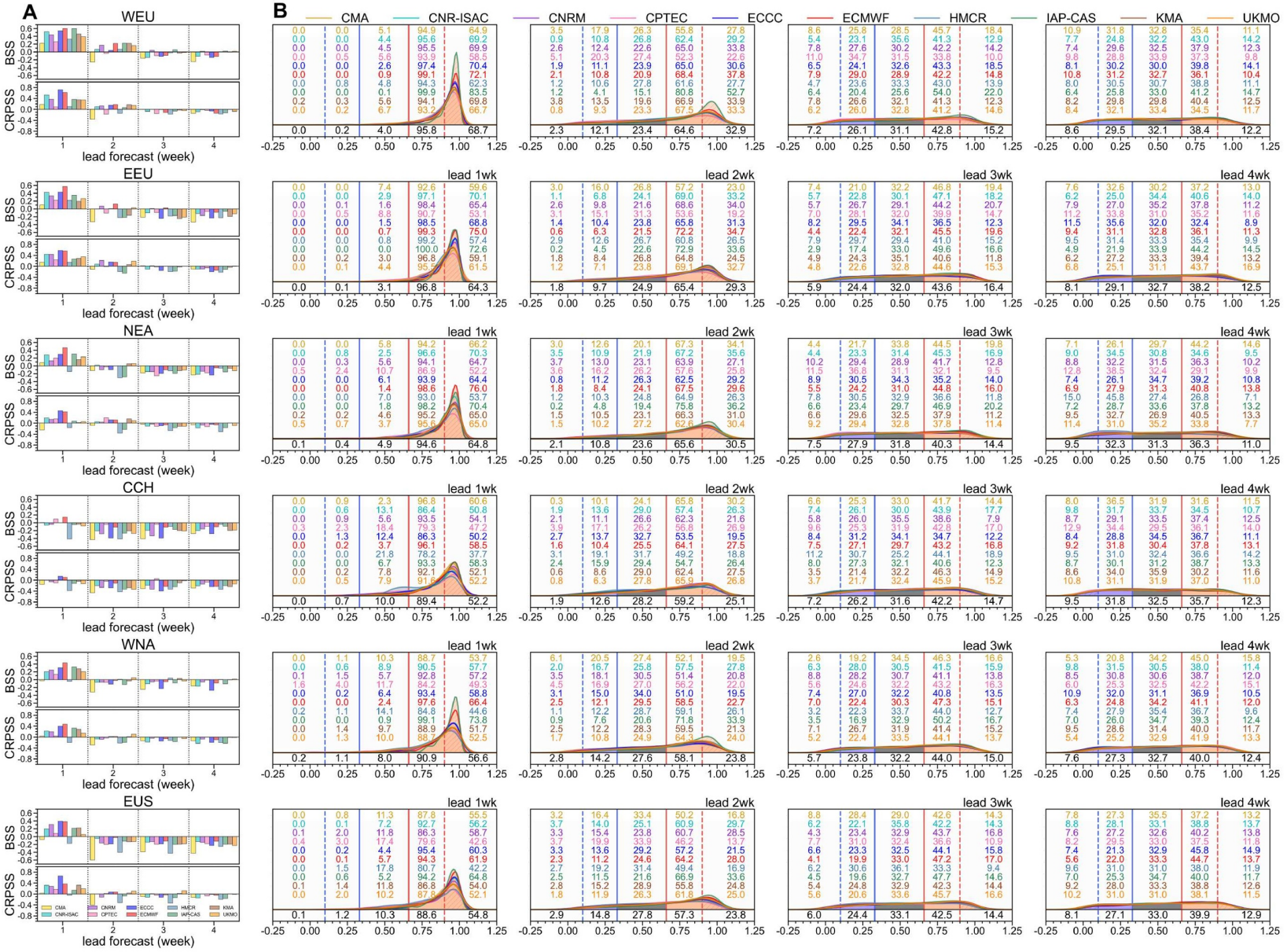


**Fig. 4** (A) BSS and CRPSS for forecasting the occurrence of heatwave days. (B) PDF of ensemble members at each *TX* percentile from 10 S2S models for observed heatwave days (BEST dataset) across six regions (top to bottom) and lead weeks 1–4 (left to right). In (B), the x-axis represents the percentiles of *TX* anomalies relative to the model climatology. PDF curves are divided by vertical lines: red dashed, red solid, blue dashed, and blue solid lines indicate the 90th, 66.7th, 10th, and 33.3rd percentiles, respectively. Shaded areas indicate below normal (<33.3%, blue), normal (33.3%–66.7%, gray), and above normal (>66.7%, red) conditions. Colored numbers (%) denote the percentage of ensemble members predicting each percentile range, while the black number below each subplot indicates the S2S multimodel mean.

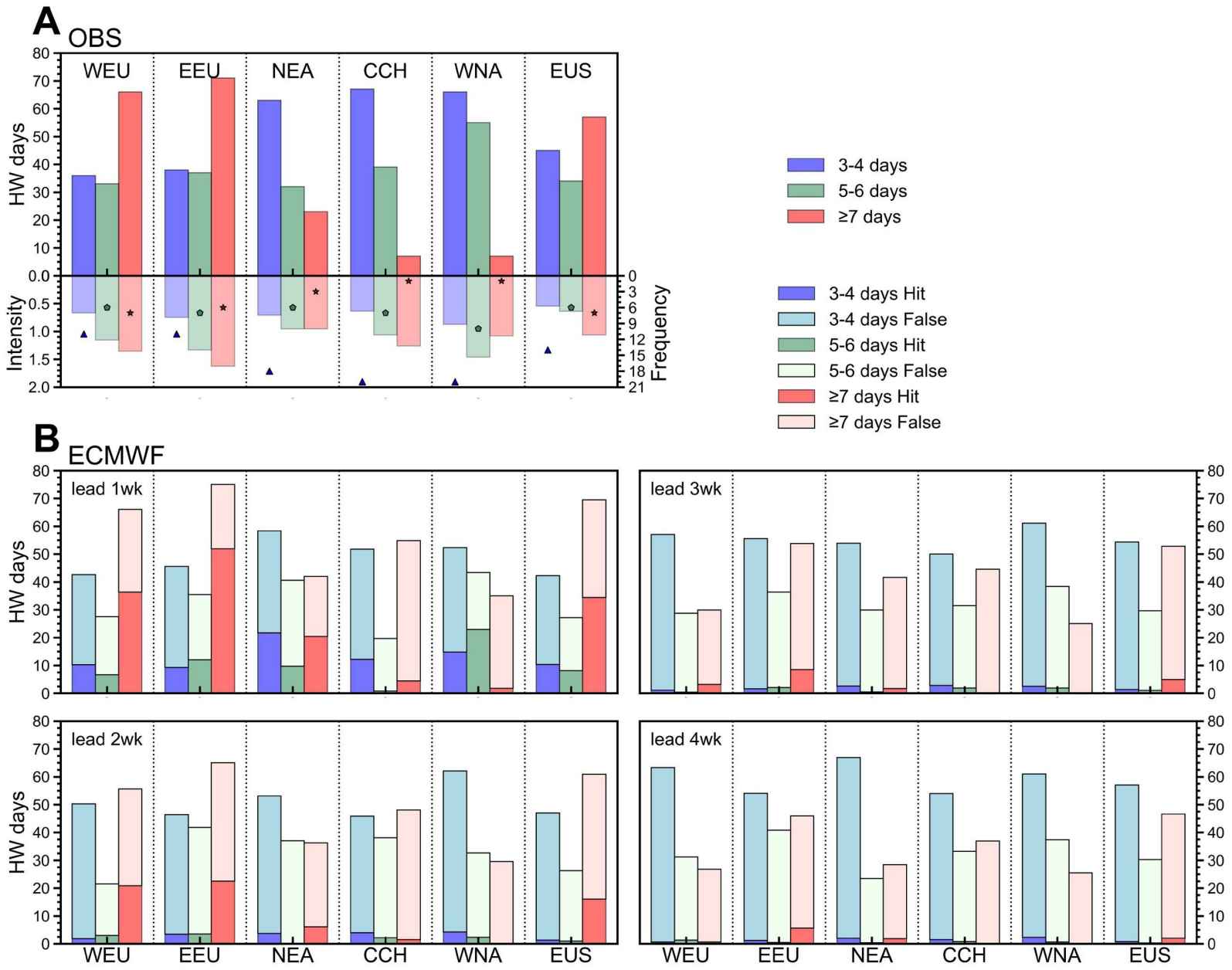


**Fig. 5** (A) Cumulative days, average intensity (unit: °C), and cumulative frequency (unit: events, indicated by markers) of heatwaves of 3–4 (blue), 5–6 (green), and ⩾7 days (red) observed in six regions from 2001 to 2020 (BEST observations). (B) Hit days (dark-colored bars) and false alarm days (light-colored bars) for heatwaves of 3–4, 5–6, and ⩾7 days predicted by the ECMWF S2S model from lead times of 1–4 weeks.

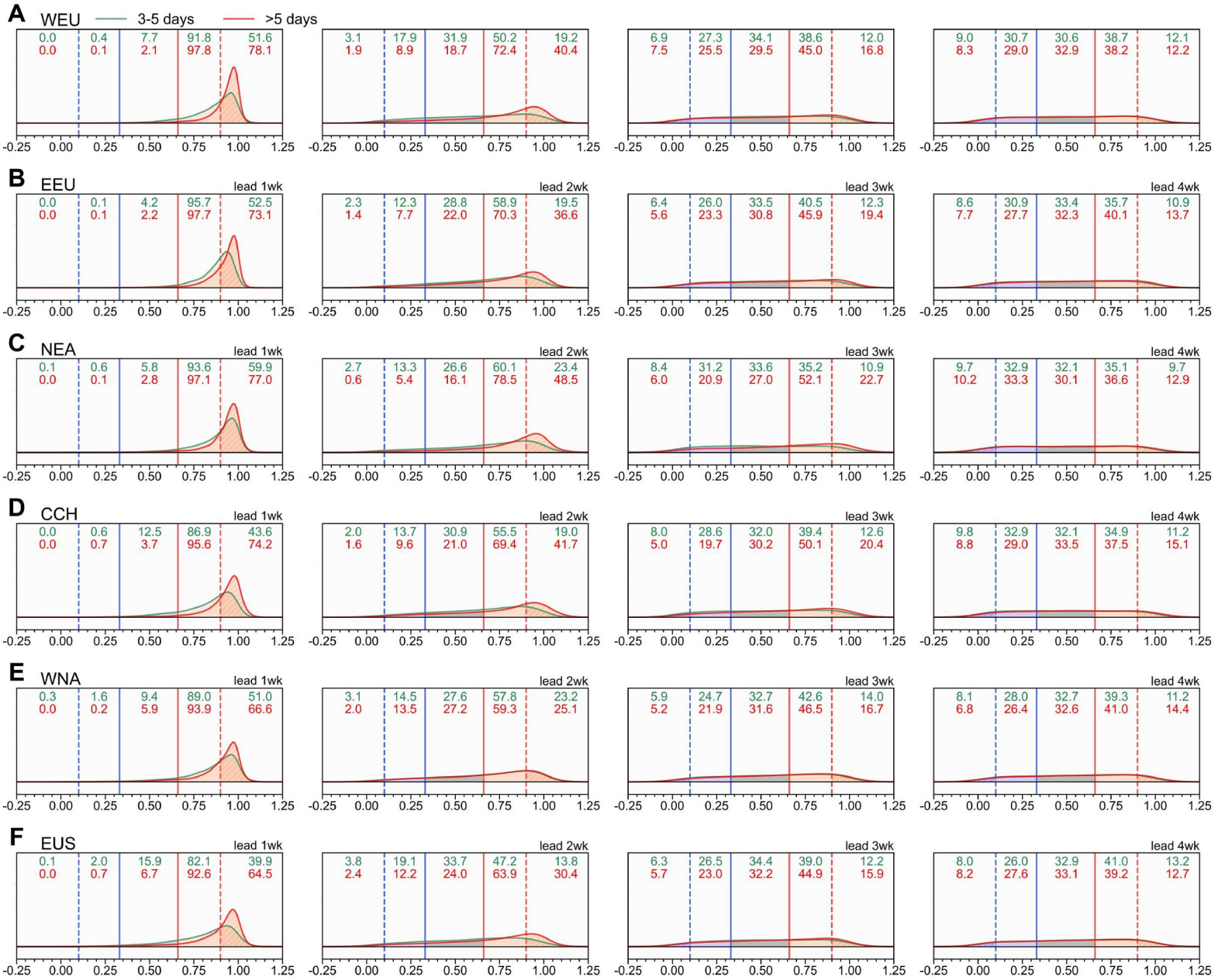


**Fig. 6** PDFs of ensemble members at each *TX* percentile from S2S models for (A–F) short-duration (green) and long-duration (red) heatwave days in six regions (top to bottom) based on BEST dataset, shown as the multimodel mean from lead weeks 1–4 (left to right). See Fig. 4 for detailed figure elements.

often "counterbalance" each other, making the relationship between heatwave duration and intensity weak, and generally reducing the forecast accuracy. For simplicity, we focus on the areas of WEU and EEU with higher forecast skill and NEA and CCH with lower forecast skill.

#### 4.1.1 *Predictable limit*

Regional circulation is often closely linked with surface air temperature (SAT). There is a statistically significant positive correlation between $Z_{500}$ and SAT across all days in the four regions analyzed (Fig. 7). During heatwave days, WEU, EEU, and NEA exhibit a similarly strong positive correlation between $Z_{500}$ and SAT. However, in CCH, this relationship is not statistically significant, suggesting that high pressure intensity does not strongly correspond to the intensity of heatwaves in CCH. Heatwave days in WEU and EEU consistently occur under high pressure anomalies, while in NEA and CCH, a small fraction of heatwave days occur under low pressure anomalies. This feature highlights the greater complexity of circulation patterns in East Asia compared with those in Europe, illustrating that the physical processes driving heatwaves in East Asia are intricate and chaotic. Consequently, this complexity results in lower predictability for heatwaves in the region.

To further validate the conclusions above, we examined the similarity of circulation anomalies during heatwaves across different regions. As shown in Appendix Fig. S6, the circulation similarity during heatwaves in Europe is notably higher than in East Asia. This holds true regardless of heatwave duration,

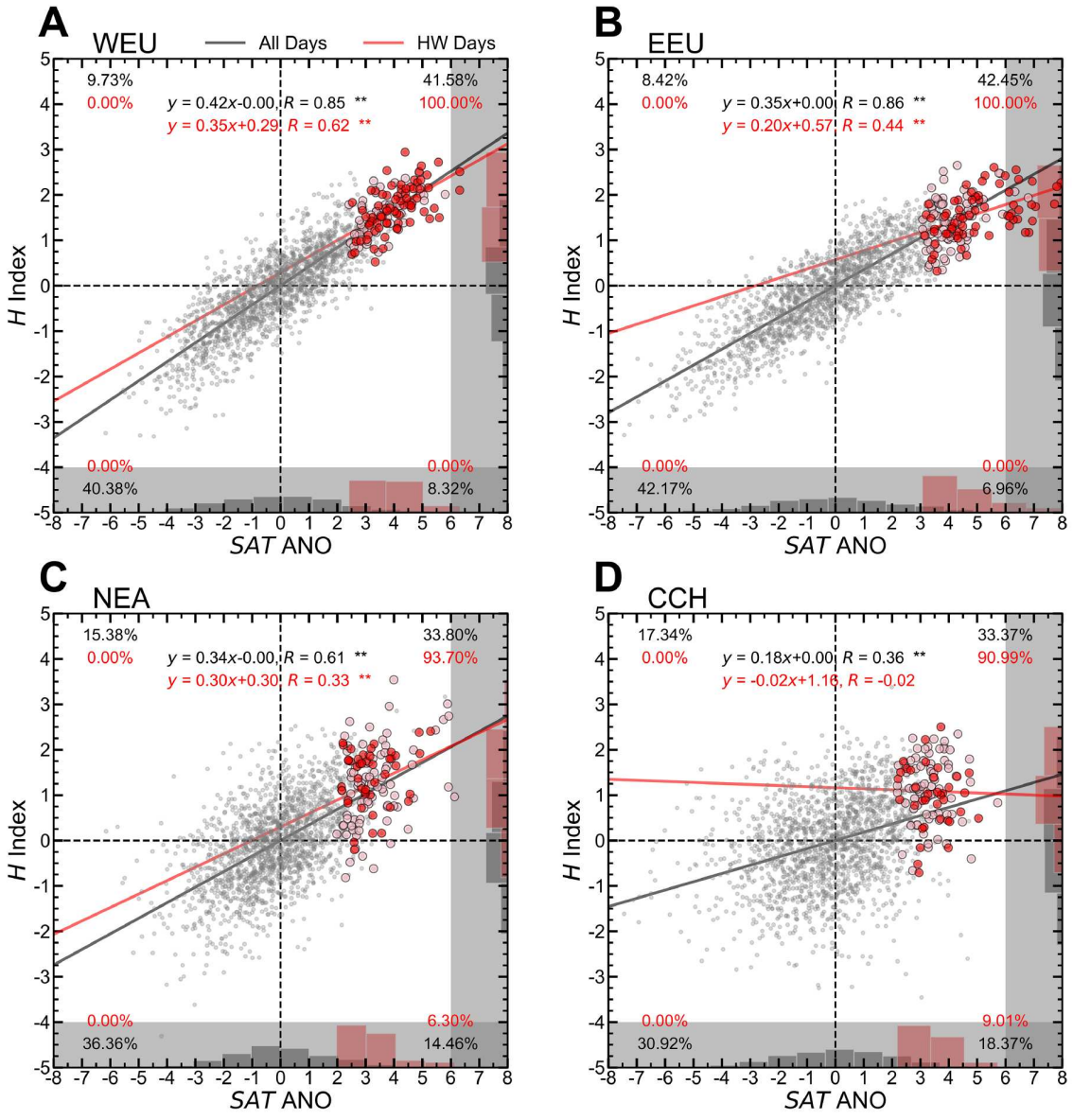


**Fig. 7** Scatterplots illustrating the relationship between standardized $Z_{500}$ anomalies and $T2$m anomalies for all days and heatwave days in (A–D) WEU, EEU, NEA, and CCH (2001–2020; ERA5). Gray, pink, and red dots represent all days, short-duration heatwaves, and long-duration heatwaves, respectively. Fitted curves are shown. Bar charts on the right and bottom show the PDFs of $Z_{500}$ anomalies and $T2$m anomalies. Linear fits, correlation coefficients, and $t$-test statistical significance levels ('*' for 95% and '**' for 99%) are indicated. Percentages of points in each quadrant relative to the total number of points are shown in the four corners of each subplot, with red for heatwave days and black for all days.

further indicating that the physical processes leading to heatwaves in East Asia are more chaotic than those in Europe.

The vertical structures of the tropospheric circulation during heatwave days under both positive and negative $Z_{500}$ anomalies were examined to further elaborate the spatial differences. It was found that in all four regions, during heatwaves associated with positive $Z_{500}$ anomalies, the circulation layer displays a quasi-barotropic structure (Appendix Fig. S7A–S7D), whereas during heatwaves under negative $Z_{500}$ anomalies, the circulation layer shows a baroclinic structure with low pressure anomalies in the lower troposphere and a ridge-trough pattern in the mid-troposphere (Appendix Fig. S7E–S7F). As illustrated in Fig. 8A–8D, the circulation patterns associated with heatwaves under positive $Z_{500}$ anomalies in WEU and EEU are linked to wave numbers 5 and 7 of the circumglobal Rossby wave over the mid-latitudes of the Northern Hemisphere (Kornhuber et al., 2020; Cai et al., 2024). By contrast, in NEA and CCH, the circulation patterns triggering heatwaves might involve multiple teleconnection systems, including eastward-propagating wave trains from mid-to-high latitudes and northward-propagating wave trains from tropical regions (Hsu et al., 2017; Nie et al., 2024; Xiao et al., 2024). This further underscores the complexity of the atmospheric circulation in East Asia. For heatwave days in NEA and CCH under negative $Z_{500}$ anomalies, these regions are often located behind an upper-level trough and ahead of a ridge, with subsidence in the mid-to-upper troposphere playing a significant role in the occurrence of heatwave days under such conditions (Fig. 8E, 8F).

#### 4.1.2 *Capability of S2S models in capturing regional predictable signals*

Soil moisture is another key source of $T2m$ predictability, with most heatwaves across the four regions occurring under dry soil anomalies (Appendix Fig. S8). Therefore, it is important to determine how well the S2S models capture these atmospheric and land-based predictable signals. As shown in Fig. 9A, the S2S models generally exhibit higher forecast skill for the atmospheric circulation in WEU and EEU compared with that in NEA and CCH. Similarly, for soil moisture predictions, all S2S models show better performance in Europe than in East Asia, except for the CMA model, which may be attributed to its lack of soil moisture assimilation (Fig. 9B). This indicates that the S2S models have stronger capability in reproducing the atmospheric and land processes directly associated with heatwave predictability in

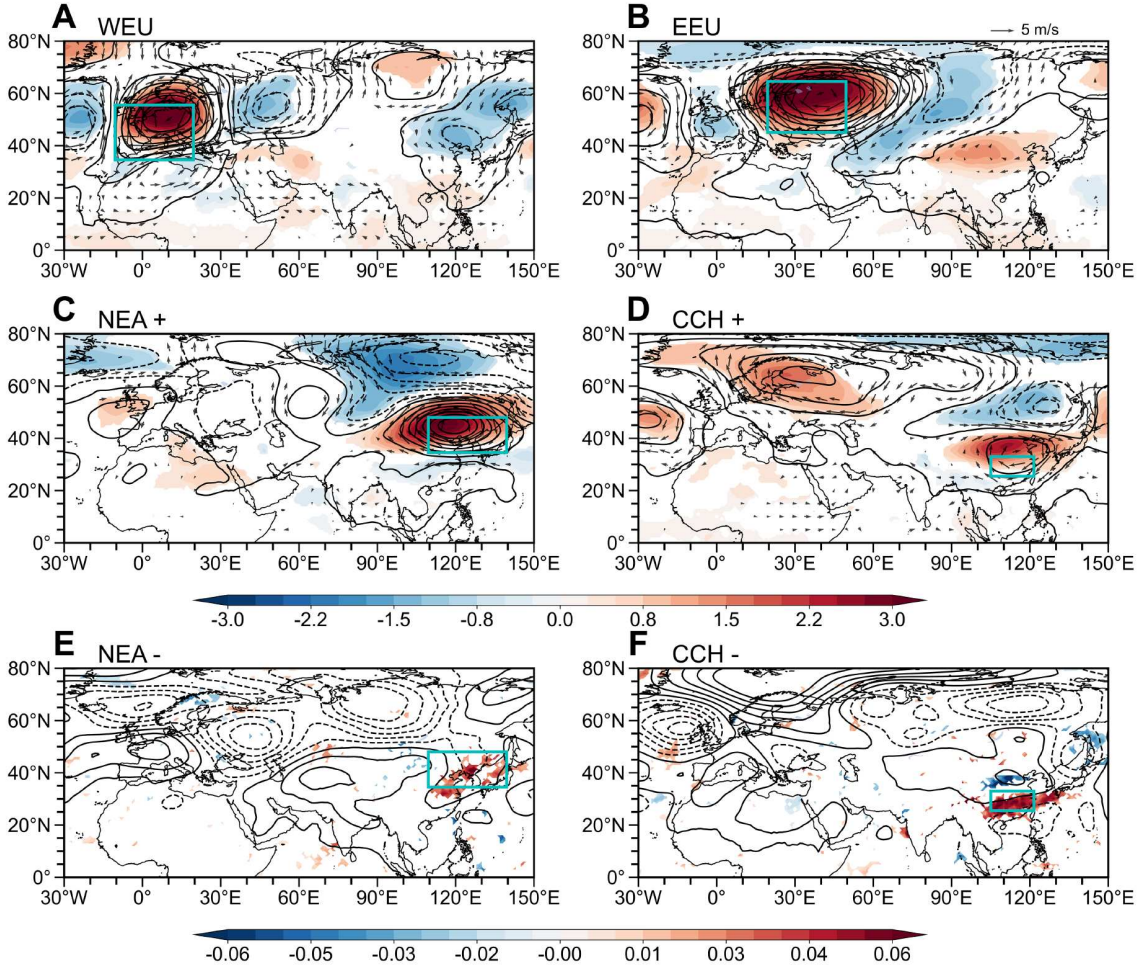


**Fig. 8** Composite anomalies for heatwave days under positive $Z_{500}$ anomalies from 2001 to 2020 in the ERA5 Reanalysis for (A–D) WEU, EEU, NEA, and CCH. Shading represents 500 hPa air temperature anomalies (unit: °C), contours represent geopotential height anomalies (unit: gpm), and vectors indicate wind anomalies (unit: m/s). Solid contours indicate positive anomalies, while dashed contours indicate negative anomalies. (E and F) Similar to (A–D); but for heatwave days under negative $Z_{500}$ anomalies in NEA and CCH regions, with shading representing vertical velocity anomalies (unit: Pa/s). Rectangular boxes outline the WEU, EEU, NEA, and CCH regions. Only shading and vectors exceeding the 95% confidence level are displayed.

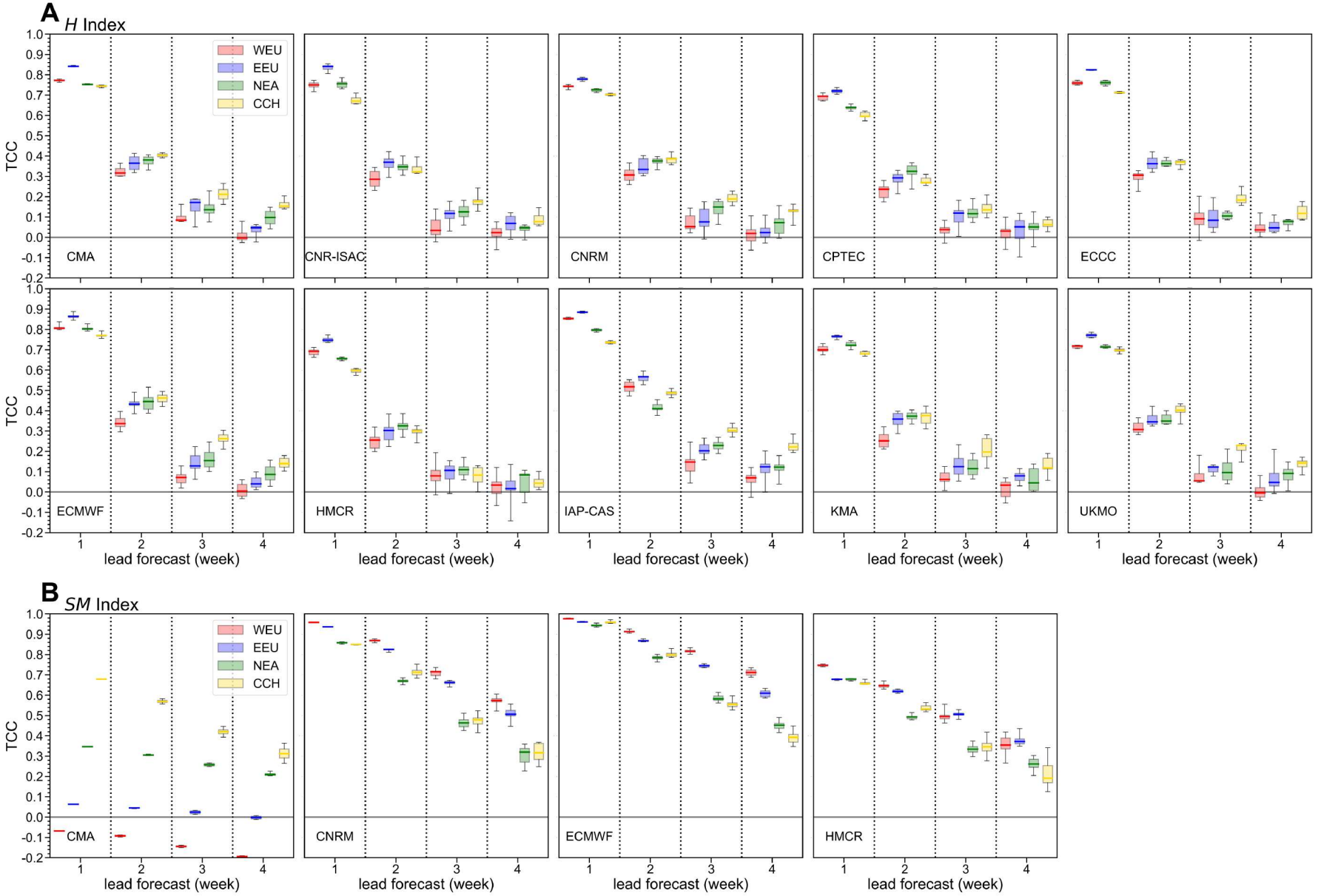


**Fig. 9** (A) Boxplots of TCC forecast skill for $Z_{500}$ anomalies during all days from lead times of 1–4 weeks for WEU (red), EEU (blue), NEA (green), and CCH (yellow). (B) TCC forecast skill for $SM_{100}$ in the four regions, showing results from the four S2S models that output $SM_{100}$. Upper and lower edges of the boxes represent the maximum and minimum values, respectively, with the median indicated by a solid line.

Europe, resulting in higher forecast skill for that region. Additionally, the distribution of observation stations is denser in Europe than in East Asia, which helps reduce the initial uncertainty of the observational field and supports higher forecast skill in that region (Appendix Fig. S9).

## 4.2 Causes of forecast skill differences in relation to heatwave duration

To investigate the causes behind the forecast skill differences related to heatwave duration, we focused on Europe, where the differences are more pronounced, and the region offers a larger sample of long-duration heatwaves for analysis. Long-duration heatwaves in Europe are often associated with prolonged and stronger blocking systems, exhibiting correlations with strong statistical significance (Fig. 10A, 10B). This indicates that longer-lasting heatwaves are linked to more consistent and less chaotic background circulation patterns. Although the relationship between soil moisture and *T*2m is statistically more significant in the WEU region, suggesting stronger land-atmosphere coupling, a higher proportion of days during long-duration heatwaves occur under anomalously dry soil moisture conditions compared with those in short-duration heatwaves (Fig. 10C, 10D). This implies that during long-duration heatwaves, the land surface provides a more predictable signal that contributes to enhanced predictability.

We also analyzed the performance of S2S models in forecasting circulation and soil moisture anomalies during both types of heatwaves. As shown in Appendix Fig. S10, at lead week 1, the models do not exhibit notable superiority in predicting circulation and soil moisture anomalies during long-duration heatwaves compared with those of short-duration heatwaves. This finding suggests that the models' ability to simulate key physical processes during heatwaves does not vary substantially with

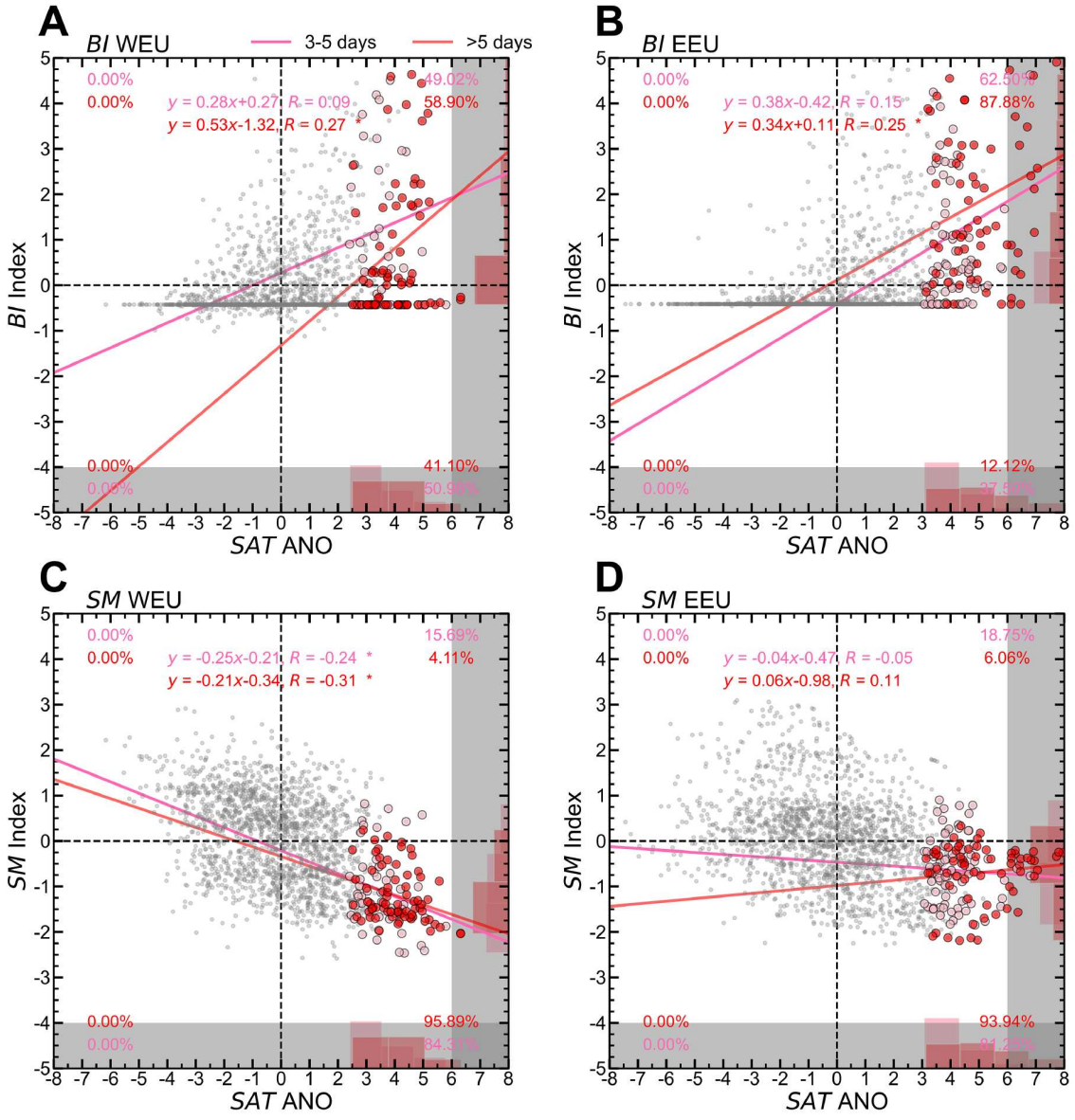


**Fig. 10** Similar to Fig. 7, but showing scatterplots of (A, B) the relationship between *BI* anomalies and *T*2m anomalies, and (C, D) the relationship between $SM_{100}$ anomalies and *T*2m anomalies. Pink dots represent short-duration heatwaves, red dots represent long-duration heatwaves.

heatwave duration in the target regions. However, at lead week 2, the models demonstrate higher forecast skill for circulation anomalies during long-duration heatwaves. This might be attributable to the forecast period approaching the atmospheric predictability limit (Lorenz, 1969), as previously discussed, with longer-lasting and more intense blocking systems providing additional sources of predictable signals. In summary, the differences in local heatwave forecast skill caused by heatwave duration are likely constrained primarily by the limits of atmospheric predictability.

## 5. Summary and discussion

Previous studies mostly focused on extreme events or failed to account for false alarms, which might have overestimated the forecast skill of S2S models for heatwaves. Current S2S models show forecast skill for the spatial distribution of heatwave characteristics (frequency, duration, and intensity) in the Northern Hemisphere mid-to-high latitudes (north of 23°N) at a lead time of approximately 2 weeks. At the regional scale, the S2S models can forecast heatwave occurrences approximately 2 weeks in advance for WEU, EEU, and NEA regions, but only around 1 week ahead for CCH, WNA, and EUS, with the latter overestimated by roughly one week in previous case studies (Domeisen et al., 2022). The potential effective forecast time for heatwaves is longer: warm anomalies can be predicted up to approximately 4 weeks in advance in North America and about 3 weeks ahead in Europe and East Asia, with the latter also overestimated by nearly one week in earlier assessments (Domeisen et al., 2022).

The S2S models show regional differences in forecast skill, with higher skill for heatwaves in Europe compared with those in East Asia and North America. The lower chaos of the circulation in Europe contributes to stronger heatwave predictability. Additionally, the models' greater ability to replicate key physical processes related to heatwaves reduces the uncertainty, and the denser distribution of observation stations in Europe further decreases the uncertainties of the initial fields. Both factors together enhance the ability of S2S models to capture predictable signals in Europe. This "stronger predictability combined with stronger modeling capability" ("stronger plus stronger") ensures superior forecast performance in the region (Fig. S11). By contrast, East Asia exhibits the opposite characteristics, resulting in lower forecast skill.

The S2S models show differences in forecast skill related to heatwave duration, with long-duration heatwaves typically corresponding to higher forecast skill. Specifically, the S2S models can predict long-duration heatwaves up to 2 weeks in advance, but only 1 week in advance for short-duration heatwaves. As the forecast lead time extends further, the difference in skill between the two types of heatwaves becomes less pronounced. In Europe, long-duration heatwaves often occur under more persistent and well-defined strong blocking circulation patterns, which provide a more predictable atmospheric background. Additionally, the memory effect of soil moisture often supplies extra predictable signals. This suggests that the difference in forecast skill related to heatwave duration is mainly influenced by the "predictable limit". While S2S models typically underestimate heatwave intensity and duration, the extent to which this model bias contributes to skill differences between long- and short-duration events—relative to the predictable limit —remains unclear.

This study focused on the regions where such differences in heatwave forecast skill are more pronounced—Europe and East Asia—leaving open whether similar conclusions hold for North America. When discussing the causes of forecast skill differences related to heatwave duration, this study emphasized local drivers such as blocking and soil moisture, but larger-scale factors (e.g., planetary waves, sea surface temperature, sea ice, snow cover, and stratospheric processes) may also modulate heatwaves through interactions with atmosphere. It is noteworthy that this study does not further explore how duration variations influence regional forecast skill differences. For instance, WEU, EEU, and EUS all experience more long-duration events, yet Europe's forecast skill exceeds that of EUS (Fig. 5A). Investigating this could help to understand regional heatwave characteristics and their driving mechanisms. Additionally, the higher forecast skill for longer-duration heatwaves warrants further exploration, especially in WNA, where more observational samples are needed. Furthermore, the ability to anticipate warm anomalies earlier in North America raises the question of whether stronger oceanic influence at longer lead times plays a role, which requires further examination.

Previous studies analyzing the mechanisms of heatwaves focused primarily on anomalous high pressure systems, where reduced cloud formation allows more solar radiation to reach the surface, increasing sensible heating. Additionally, anomalous subsidence-driven adiabatic warming provides conditions favorable for extreme heat (Barriopedro et al., 2023). However, in this study, it was observed that high temperatures in East Asia can occasionally occur under anomalous low pressure systems. The underlying physical processes behind such cases remain to be investigated. Liu et al. (2024) suggested that reduced cloud cover and increased sensible heat flux in the Yangtze River Basin during the early stages of a heatwave promote the onset of high temperatures, whereas increased cloud cover and reduced sensible heating during later stages contribute to the gradual decay and end of the heatwave. This finding indicates that statistical analyses based on the entire heatwave period might be somewhat insufficient. It is recommended that the dynamic processes governing heatwave onset, development, and cessation be investigated, focusing on more detailed mechanisms to identify potential windows of opportunity for subseasonal heatwave predictions.

The complex circulation over East Asia is shaped by multiple factors: the Tibetan Plateau heating influences the South Asian High (Liu et al., 2020), the Western Pacific Warm Pool convection affects the Western Pacific Subtropical High (Huang and Sun, 1994), mid-high latitude Rossby wave trains induce teleconnections (Li et al., 2019), and tropical convection triggers northward-propagating intraseasonal oscillation wave trains (Hsu et al., 2017). Additionally, land-atmosphere feedback, such as soil moisture, further modulates the local circulation (Jiang et al., 2023). By contrast, Europe, situated in the mid-high latitudes, exhibits a much smaller Burger number, with advection-dominated dynamics that generally follow a quasi-barotropic structure (Hoskins, 1987; Wu et al., 2002). Consequently, circulation patterns tend to be simpler and less chaotic. Understanding heatwaves under barotropic and baroclinic regimes is crucial for improving S2S model forecasts. Developing tailored ensemble initialization strategies for these regimes might enhance model performance, which is something that we aim to

explore in the future. The Subseasonal Experiment project uses multimodel ensembles to assess systematic model uncertainties (Pegion et al., 2019). Extending this approach to multimodel ensemble forecasts within the S2S project could further enhance heatwave predictions, making it a valuable avenue for future research.

The explanation of the causes behind heatwave forecast skill differences in this study remains relatively brief. Questions remain about the relative contributions of the predictable limit, model simulation uncertainties, and uncertainties in initial observation data. We believe that further predictability studies are necessary to deepen our understanding of subseasonal heatwave forecasting. Encouragingly, during this research, it was found that S2S reforecast data could reproduce the predictability differences shown in Fig. 7 and Appendix Fig. S8. This suggests that current S2S models provide a solid foundation for quantifying the predictable limits and the ability of models to capture predictable signals. Such research will enhance our understanding of subseasonal heatwave forecasts, and provide theoretical support for reducing initial condition and model uncertainties. It will also inform the selection of ensemble forecasting methods and the optimization of observational networks, thereby improving subseasonal heatwave forecast quality under limited resources. Improved prediction of extreme events will ultimately strengthen decision-making strategies and enhance society's resilience to risk and disasters.


## Acknowledgement

The authors thank three anonymous reviewers for their valuable comments that helped to improve the manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 42288101). We also thank for the technical support of the National Large Scientific and Technological Infrastructure "Earth System Numerical Simulation Facility" (https://cstr.cn/31134.02.EL).


## Supporting information

The supporting information is available online at http://earth.scichina.com and http://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

## Compliance and ethics

The authors declare no conflict of interest.

## References


Barriopedro D, García-Herrera R, Ordóñez C, et al. 2023. Heat waves: Physical understanding and scientific challenges. Rev Geophys, 61: e2022RG000780

Brier G W. 1950. Verification of forecasts expressed in terms of probability. Mon Weather Rev, 78: 1–3

Brunner L, Voigt A. 2024. Pitfalls in diagnosing temperature extremes. Nat Commun, 15: 2087

Buizza R, Leutbecher M. 2015. The forecast skill horizon. Q J R Meteorol Soc, 141: 3366–3382

Cai F, Liu C, Gerten D, et al. 2024. Sketching the spatial disparities in heatwave trends by changing atmospheric teleconnections in the Northern Hemisphere. Nat Commun, 15: 8012

Center for International Earth Science Information Network-CIESIN-Columbia University. 2017. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10, doi: 10.7927/H4DZ068D

COPES W. 2005. The World Climate Research Programme Strategic Framework 2005-2015: Coordinated Observation and Prediction of the Earth System (COPES). WMO. 65

Dai G, Mu M, Li C, et al. 2021. Evaluation of the forecast performance for extreme cold events in east Asia with subseasonal-to-seasonal data sets from ECMWF. J Geophys Res-Atmos, 126: 2020JD033860

Dirmeyer P A, Gentine P, Ek M B, et al. 2019. Land surface processes relevant to sub-seasonal to seasonal (S2S) prediction. In: Robertson A W, Vitart F, eds. Sub-Seasonal to Seasonal Prediction. Amsterdam: Elsevier. 165–181

Domeisen D I V, White C J, Afargan-Gerstman H, et al. 2022. Advances in the subseasonal prediction of extreme events. Bull Amer Meteorol Soc, 103: E1473–E1501

Domeisen D I V, Eltahir E A B, Fischer E M, et al. 2023. Prediction and projection of heatwaves. Nat Rev Earth Environ, 4: 36–50

Ford T W, Dirmeyer P A, Benson D O. 2018. Evaluation of heat wave forecasts seamlessly across subseasonal timescales. NPJ Clim Atmos Sci, 1: 20

Gelaro R, McCarty W, Suárez M J, et al. 2017. The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim, 30: 5419–5454

Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast, 15: 559–570

Hersbach H, Bell B, Berrisford P, et al. 2020. The ERA5 global reanalysis. Q J R Meteorol Soc, 146: 1999–2049

Hobday A J, Alexander L V, Perkins S E, et al. 2016. A hierarchical approach to defining marine heatwaves. Prog Oceanogr, 141: 227–238

Hoskins B J. 1987. Diagnosis of forced and free variability in the atmosphere. In: Cattle H, eds. Atmospheric and Oceanic Variability. Bracknell, UK: James Glaisher House. 57–73

Howard S, Krishna G. 2022. How hot weather kills: The rising public health dangers of extreme heat. BMJ, 378: o1741

Hsu P C, Lee J Y, Ha K J, et al. 2017. Influences of boreal summer intraseasonal oscillation on heat waves in monsoon Asia. J Clim, 30: 7191–7211

Hsu P C, Qian Y, Liu Y, et al. 2020. Role of abnormally enhanced MJO over the Western Pacific in the formation and subseasonal predictability of the record-breaking Northeast Asian heatwave in the summer of 2018. J Clim, 33: 3333–3349

Huang R, Sun F. 1994. Impacts of the thermal state and the convective activities in the tropical western warm pool on the summer climate anomalies in East Asia (in Chinese). Chin J Atmos Sci, 18: 141–151

Jiang J, Liu Y, Mao J, et al. 2023. Extreme heatwave over Eastern China in summer 2022: The role of three oceans and local soil moisture feedback. Environ Res Lett, 18: 044025

Jyoteeshkumar reddy P, Sharples J J, Lewis S C, et al. 2021. Modulating influence of drought on the synergy between heatwaves and dead fine fuel moisture content of bushfire fuels in the Southeast Australian region. Weather Clim Extrem, 31: 100300

Kautz L A, Martius O, Pfahl S, et al. 2022. Atmospheric blocking and weather extremes over the Euro-Atlantic sector—A review. Weather Clim Dynam, 3: 305–336

Kornhuber K, Coumou D, Vogel E, et al. 2020. Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. Nat Clim Chang, 10: 48–53

Li J, Zheng F, Sun C, et al. 2019. Pathways of influence of the Northern Hemisphere mid-high latitudes on East Asian climate: A review. Adv Atmos Sci, 36: 902–921

Libonati R, Geirinhas J L, Silva P S, et al. 2022. Assessing the role of compound drought and heatwave events on unprecedented 2020 wildfires in the Pantanal. Environ Res Lett, 17: 015005

Liu W, Shi N, Wang H, et al. 2024. Thermodynamic characteristics of extreme heat waves over the middle and lower reaches of the Yangtze River Basin. Clim Dyn, 62: 3877–3889

Liu Y, Lu M, Yang H, et al. 2020. Land-atmosphere-ocean coupling associated with the Tibetan Plateau and its climate impacts. Natl Sci Rev, 7: 534–552

Lorenz E N. 1969. The predictability of a flow which possesses many scales of motion. Tellus A-Dynamic Meteorol Oceanogr, 21: 289

Mariotti A, Ruti P M, Rixen M. 2018. Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. NPJ Clim Atmos Sci, 1: 4

Meehl G A, Tebaldi C. 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305: 994–997

Mondal S, Mishra A K. 2021. Complex networks reveal heatwave patterns and propagations over the USA. Geophys Res Lett, 48: e2020GL090411

Mu M, Duan W S, Tang Y M. 2017. The predictability of atmospheric and oceanic motions: Retrospect and prospects. Sci China Earth Sci, 60: 2001–2012

Nie Y, Ren H L, Zuo J, et al. 2024. Eurasian mid-latitude jet stream bridges an Atlantic to Asia summer teleconnection in heat extremes. Environ Res Lett, 19: 044003

Oliver E C J, Donat M G, Burrows M T, et al. 2018. Longer and more frequent marine heatwaves over the past century. Nat Commun, 9: 1324

Pegion K, Kirtman B P, Becker E, et al. 2019. The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. Bull Amer Meteorol Soc, 100: 2043–2060

Pyrina M, Domeisen D I V. 2023. Subseasonal predictability of onset, duration, and intensity of European heat extremes. Q J R Meteorol Soc, 149: 84–101

Robertson A, Vitart F. 2018. Sub-Seasonal to Seasonal Prediction: The Gap

Between Weather and Climate Forecasting. Amsterdam: Elsevier. 569
Rohde R, Muller R, Jacobsen R, et al. 2013. Berkeley Earth temperature averaging process. Geoinfor Geostat-An Overview, 1: 1–13
Scherrer S C, Croci-Maspoli M, Schwierz C, et al. 2006. Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. Int J Climatol, 26: 233–249
Sciences N A, Earth D, Studies L, et al. 2016. Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Washington, DC: National Academies Press. 351
Somanathan E, Somanathan R, Sudarshan A, et al. 2021. The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing. J Political Economy, 129: 1797–1827
Tibaldi S, Molteni F. 1990. On the operational predictability of blocking. Tellus A-Dynamic Meteorol Oceanogr, 42: 343
Vitart F, Ardilouze C, Bonet A, et al. 2017. The subseasonal to seasonal (S2S) prediction project database. Bull Amer Meteorol Soc, 98: 163–173
Vitart F, Robertson A W. 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. NPJ Clim Atmos Sci, 1: 3
Vitart F, Robertson A W, Anderson D L. 2012. Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. B World Meteorol Organ, 61: 23
Wang Y, Zhou W, Wang C. 2024. Physical mechanism of the rapid increase in intense and long-lived extreme heatwaves in the Northern Hemisphere since 1980. Sci China Earth Sci, 67: 2103–2121
Wilks D S. 2019. Statistical Methods in the Atmospheric Sciences. Amsterdam: Elsevier. 818
Wu G, Chou J, Liu Y, et al. 2002. Dynamics of the Formation and Variation of Subtropical Anticyclones (in Chinese). Beijing: Science Press. 314
Wulff C O, Domeisen D I V. 2019. Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers. Geophys Res Lett, 46: 11520–11529
Xiao H, Xu P, Wang L. 2024. The Unprecedented 2023 North China heatwaves and their S2S predictability. Geophys Res Lett, 51: e2023GL107642
Xie J, Yu J, Chen H, et al. 2020. Sources of subseasonal prediction skill for heatwaves over the Yangtze River Basin revealed from three S2S models. Adv Atmos Sci, 37: 1435–1450
Yang J, Zhu T, Gao M, et al. 2018. Late-July barrier for subseasonal forecast of summer daily maximum temperature over Yangtze River Basin. Geophys Res Lett, 45: 12610–12615
Zhang T, Deng Y, Chen J, et al. 2023. An energetics tale of the 2022 megaheatwave over central-eastern China. NPJ Clim Atmos Sci, 6: 162
Zscheischler J, Westra S, van den Hurk B J J M, et al. 2018. Future climate risk from compound events. Nat Clim Change, 8: 469–477

(Editorial handling: Dehai LUO)

《中国科学》杂志社
SCIENCE CHINA PRESS

# 北半球中高纬热浪的次季节预报技巧：区域性及热浪持续时间差异性

刘昊宙[1,2]，刘屹岷[1,2*]，包庆[3]，何编[1,2]，段晚锁[1,2]

1. 中国科学院大气物理研究所，地球系统数值模拟与应用全国重点实验室，北京 100029
2. 中国科学院大学地球与行星科学学院，北京 100049
3. 中国科学院大气物理研究所，大气和海洋动力学研究室，北京 100029
* 通讯作者，E-mail: lym@lasg.iap.ac.cn

2025-04-25收稿，2025-10-26修回，2025-12-11接受，2025-12-11网络版发表
国家自然科学基金项目(42288101)资助

**摘要** 准确预测热浪至关重要，其作为次季节至季节(S2S)预测计划的核心目标之一也备受关注。先前研究多聚焦于极端个例且未考虑空报，可能高估了模式对热浪的预报技巧，因此有必要全面评估当前S2S模式对热浪的预报能力。本研究基于当前多个S2S模式长期历史回算数据，系统评估了北半球中高纬热浪的次季节预报技巧。结果表明：半球空间尺度上的热浪特征(频数、强度和持续时间)可被S2S模式提前大约2周预测。模式可提前1–2周预测热浪发生，提前3–4周预测暖异常。热浪预报技巧表征出区域性及持续时间的差异性，体现为欧洲地区热浪预报技巧更高，东亚和北美地区预报技巧较低，这主要归因于欧洲区域热浪可预报性更高同时模式对该地区的模拟能力更强。同时长期型热浪较短期型热浪在提前2周预报时间范围内预报技巧更高，这主要是受大气内禀可预报上限的制约。本研究为了解当前S2S模式对热浪的次季节预报能力提供了全面的认识。

**关键词** 热浪，次季节至季节(S2S)，预报技巧，区域性，持续时间差异性

## 1 引言

极端热浪对社会的生产生活影响深远(Somanathan等，2021)，严重威胁人们的生命健康安全(Howard和Krishna，2022)。此外，长期持续的热浪还将为干旱野火等极端事件提供有利的背景条件(Libonati等，2022；Jyoteeshkumar reddy等，2021)。当热浪与极端降水等其他极端事件复合发生时，其社会危害将被进一步放大(Zscheischler等，2018)。而在未来，伴随气候变化下热浪事件更加频发，其潜在风险还将继续加剧(Meehl和Tebaldi，2004；Domeisen等，2023)。因此提升对热浪的预测水平迫在眉睫。热浪通常发生在次季节时间尺度内，次季节至季节(S2S)预测由于其同时受初始条件和外强迫可预测信号源的影响，往往更加复杂，被誉为“可预报荒漠”(Vitart等，2012)。尽管如此，弥合这一缺陷对全面实现无缝预测至关重要(COPES，2005)。而作为这一目标的一部分，S2S预测计划更将提升对极端高影响事件的预测准确性放在重中之重(Vitart和Robertson，2018)。

热浪的发生通常是由于在天气尺度范围内局地出现反气旋环流异常、晴空太阳短波辐射增加和下沉绝热增温共同作用使得地表气温出现暖异常(Barriopedro等，2023；Domeisen等，2023)。在更大时空尺度上，热浪期间的局地环流异常往往是由Rossby波活动和长期持续的大气阻塞系统引起(Kautz等，2022)。当时间尺度延伸至次季节尺度及以上，土壤湿度和海表温度等具有记忆效应的慢变过程开始发挥作用(Jiang等，2023)。Robertson和Vitart(2018)及Sciences等(2016)对次季节可预测信

中文引用格式：刘昊宙, 刘屹岷, 包庆, 何编, 段晚锁. 2026. 北半球中高纬热浪的次季节预报技巧：区域性及热浪持续时间差异性. 中国科学: 地球科学, doi: 10.1360/N072025-0166
英文引用格式：Liu H, Liu Y, Bao Q, He B, Duan W. 2026. Heatwave forecast skill in the mid-high latitudes of the Northern Hemisphere at subseasonal lead times: Regional and duration disparity. Science China Earth Sciences, https://doi.org/10.1007/s11430-025-1739-8

号的来源进行了分类。这些过程涵盖了热带与副热带准周期性振荡变率、有记忆性的慢变过程(如陆地、海洋、海冰及平流层影响)，以及外部强迫(包括人为影响和火山活动等)。这些要素一方面作为驱动因子影响热浪的发生，而另一方面从次季节预测角度来说，它们也为进行热浪的次季节预测提供了稳定的可预测信号源。

Pyrina和Domeisen(2023)探究了欧洲中期天气预报中心(ECMWF)S2S模式对欧洲区域性热浪特征(发生、持续时间和强度)的预报技巧，指出热浪特征的局地可预报时效约为1–3周，然而从更大空间尺度探讨热浪特征空间分布预报技巧的研究仍相对匮乏。已有研究表明S2S模式大约可以提前2周预测热浪发生，提前4周乃至更长时间预测出暖异常(Vitart和Robertson，2018；Domeisen等，2022)。需要说明的是，此类研究多聚焦于观测中发生的极端热浪个例，而忽视了“常规”热浪事件并且未能考虑空报。而S2S模式常对极端热浪事件展示出更高的预报技巧(Wulff和Domeisen，2019)，因此先前研究将难免高估模式对热浪的预报技巧。所以明确当前S2S模式对热浪的真实预报水平是必要的。

当前S2S模式对热浪的预报技巧呈现出区域差异性(Ford等，2018；Pyrina和Domeisen，2023)。同时如前所述，S2S模式对持续时间更长强度更强的热浪事件也展示出更高的预报技巧。而造成这些预报技巧差异性背后的成因仍有待深入探究。本研究旨在基于多S2S模式长期历史回算数据，系统性评估当前模式对热浪的预报技巧，并尝试从可预报性角度探究造成预报技巧差异性的潜在成因。其中热浪的预报技巧评估包含两个方面：从热浪特征空间分布角度进行的确定性预报技巧评估和从热浪日预测准确性角度进行的概率预报技巧评估。之后本文还进一步分析了造成热浪预报技巧区域差异性及持续时间差异性的潜在原因。

## 2 数据和方法

### 2.1 数据

#### 2.1.1 回算数据

当前S2S数据库共包含13家数值模式，然而各模式回算数据的历史时间覆盖范围与预报频率不尽相同(Vitart等，2017)，这给系统性评估当前S2S模式对热浪的整体预报水平带来了一定困难。为尽可能选择更多模式评估同时又避免由于时间覆盖范围不同而引起模式间对比的不公平，本研究选择了在2001–2020年间回算年数不少于15年的S2S模式进行后续的评估工作。同时鉴于模式对热浪的预报时效通常在一个月以内(Domeisen等，2022)，过低的回算频率可能会造成评估不准确，因此本研究要求被选定模式至少每月有四次回算。最终共选取了10家S2S模式：CMA、CNR-ISAC、CNRM、CPTEC、ECCC、ECMWF、HMCR、IAP-CAS、KMA 和 UKMO(网络版附表S1)。使用的变量包括逐6小时最高2米气温($T2m$)、逐日瞬时500 hPa位势高度($Z_{500}$)及逐日平均浅层100 cm土壤湿度($SM_{100}$)。

为确保后续各模式间比较的公平性和便捷性，本研究先统一将回算频率各不相同的S2S模式数据处理为逐日回算数据(Yang等，2018)。此方法的核心思想是对某一回算缺失日，取用其临近范围内滑动窗口值的平均以替换缺失值，Xie等(2020)和Yang等(2018)已表明此种处理方式并不会影响后续算得的预报技巧。具体而言，对回算频率为每周2次、每5天1次、每7天1次和每月4次的S2S模式，分别取用5、7和9天的滑动窗口进行处理。为简化计算并便于比较，本研究将4–10天、11–17天、18–24天和25–31天的时效平均值分别作为模式提前1–4周的预报近似。这种方法更适用于在缺乏实际热浪事件作为参照情况下，对模式气候态进行处理。

#### 2.1.2 观测和再分析数据

在比较模式间预报技巧时，单一再分析资料可能会对个别模式产生偏好(Wulff和Domeisen，2019)。因此我们采用了多源观测和再分析数据，其中包括来自ERA5(ECMWF Reanalysis version 5；Hersbach等，2020)和MERRA2(Modern-Era Retrospective Analysis for Research and Applications；Gelaro等，2017)再分析数据集，以及来自BEST(Berkeley Earth Surface Temperature；Rohde等，2013)观测数据集中的日最高2米气温($TX$)。此外，在探究热浪预报技巧差异性成因时还用到了ERA5的逐小时环流及陆面数据，这些变量主要包括：1000–10 hPa的位势高度、气温、垂直速度、纬向风速和经向风速，及陆面顶部三层(0–7、7–28和28–100 cm)的土壤湿度数据，土壤湿度均通过线性加权求和后得到$SM_{100}$。各数据集的空间分辨率为：ERA5(0.25°×0.25°)、MERRA2(0.5°×0.625°)和BEST(1°×1°)。后续计算中所有逐小时数据均被提前处理为日平均，同时本研究聚焦于2001–2020年北半球中高纬度地区的夏季(6–8月，JJA)。

#### 2.1.3 其他数据

本研究采用了空间分辨率约5 km×5 km的全球网格人口数据集(Gridded Population of the World，version 4；Center for International Earth Science Information Network-

CIESIN-Columbia University，2017)以识别北半球高社会影响的热点区域。全球观测站点分布数据集来自世界气象组织全球观测系统框架下的观测系统能力分析与审查工具(Observing Systems Capability Analysis and Review Tool，https://oscar.wmo.int/surface/)。

## 2.2 定义

### 2.2.1 热浪定义

当前S2S模式的预测往往倾向于低估热浪事件的强度(Xie等，2020；Pyrina和Domeisen，2023)，这说明对所有模式进行统一的绝对阈值热浪定义是不合适的。因此在本研究中，定义每个陆地格点至少连续3天$TX$超过该模式90%相对阈值为一次热浪事件，相对阈值的计算采用和先前研究(Mondal和Mishra，2021；Zhang等，2023)一致的方式，取气候态2001–2020年JJA某日前后7天累计15天的滑动窗口来计算。Brunner和Voigt(2024)表明在计算极端阈值前应先去除季节循环以避免由时间、区域和数据依赖性偏差引起的低估极端事件频数的问题，因此本研究在计算相对阈值前已事先去除季节循环。此外，鉴于海洋的热浪定义方式与陆地热浪不同(Hobday等，2016；Oliver等，2018)，本文后续所述热浪均为陆地热浪。

热浪的特征包含频数、强度、发生时间、持续时间和空间范围(Domeisen等，2023)，本研究主要关注其中的频数、强度和持续时间特征。频数为一年夏季中热浪发生的累计次数。持续时间包含最长持续时间及平均持续时间，分别定义为一年夏季中发生的最长一次热浪的持续时间，及一年夏季中发生的所有热浪的平均持续时间。强度也包含最大强度和平均强度，前者定义为一年夏季中所有热浪日日最高气温较90%相对阈值温度异常的最大值，后者定义为一年夏季中所有热浪日日最高气温较90%相对阈值温度异常的平均值(王雨晴等，2024)。

如图1所示，本研究中所采用的三套再分析和观测数据集对北半球热浪特征的空间分布刻画基本一致。热浪频数的空间分布呈现低纬少、高纬多的特点，观测中的热浪频数整体较再分析数据中偏多(图1B)。此外，低纬度地区热浪持续时间短、强度弱，高纬度地区热浪持续时间长、强度强(图1D，1F)，这与北半球热浪特征的更长期气候态分布相一致(王雨晴等，2024)。容易发生长期热浪的区域主要集中在北美东部、欧洲东部和西伯利亚地区中西部(图1C)。值得注意的是，MERRA2在低纬度区域呈现的热浪持续时间较ERA5和BEST异常偏长(图1D)，同时Domeisen等(2023)也指出基于相对阈值和持续时间而定义热浪的指标可能在热带地区不适用，因此本文后续的研究区域主要聚焦于北半球中高纬度地区(23°N以北)。在此仅展示了热浪平均持续时间及平均强度的空间分布，最长持续时间及最大强度的空间分布与前者类似而未展示。

### 2.2.2 阻塞强度指数

后文中针对欧洲两区域定义的阻塞强度指数($BI$)为某区域内发生阻塞的格点强度面积加权求和与区域总面积之比。具体定义如下：(1) 基于格点的$Z_{500}$南北梯度判定其是否发生阻塞，主要参考了Scherrer等(2006)中的阻塞定义方式，该方案是对更早期Tibaldi和Molteni(1990)中阻塞定义的拓展。如果一个格点的$Z_{500}$至少连续5天向北梯度(GHGN)小于−10 m/(°纬度)且向南梯度(GHGS)大于0 m/(°纬度)，则被视作发生了阻塞，记此期间该格点的为1，否则被记作0。(2) 计算区域内单位面积的阻塞强度$B(t)$，其为随时间变化的函数。(3) 将$B(t)$标准化得到$BI$便于后续比较。

$$\begin{aligned}&\mathrm{GHGN}=\frac{Z_{500}(\varphi_N)-Z_{500}(\varphi_0)}{\varphi_N-\varphi_0}\\&\mathrm{GHGS}=\frac{Z_{500}(\varphi_0)-Z_{500}(\varphi_S)}{\varphi_0-\varphi_S}\\&\begin{cases}b(t)_{\varphi_0,\lambda}=1,\mathrm{GHGN}<-10\text{及}\mathrm{GHGS}>0\text{至少5天}\\b(t)_{\varphi_0,\lambda}=0,\text{其他}\end{cases}\end{aligned}\tag{1}$$

$$B(t)=\frac{\sum b(t)_{\varphi_0,\lambda}\cdot\cos\varphi_0\cdot Z'_{500\varphi_0,\lambda}(t)}{\sum\cos\varphi_0}\tag{2}$$

式中，$\varphi_0$和$\lambda$为34.5°N至75°N范围内任一格点的纬度和经度，且$|\varphi_{N/S}-\varphi_0|=15°$。$Z'_{500\varphi_0,\lambda}$为某日$(\varphi_0,\ \lambda)$格点上$Z_{500}$较气候平均态的异常值。

## 2.3 评估指标

本研究采用均方根误差(RMSE)和空间距平相关系数(ACC)来评估S2S模式的集合平均对北半球中高纬度地区热浪特征的确定性预报技巧。其中RMSE可以反映偏差大小，ACC可用于衡量模式对空间分布的预报性能。此外，还用时间距平相关系数(TCC)评估S2S模式对$Z_{500}$和$\mathrm{SM}_{100}$异常的预报技巧。

此外，本研究还使用了Brier技巧评分(BSS)和连续分级概率技巧评分(CRPSS)来评估模式对热浪发生预报的概率预报技巧。BSS值越高则预报技巧越高，BSS为0时代表无预报技巧，为1时则对应着完美预报(Wilks，2019)。如公式(3)所示，BSS综合考虑了模式对热浪的命中、空报和漏报，因此更能客观体现模式对热浪的整

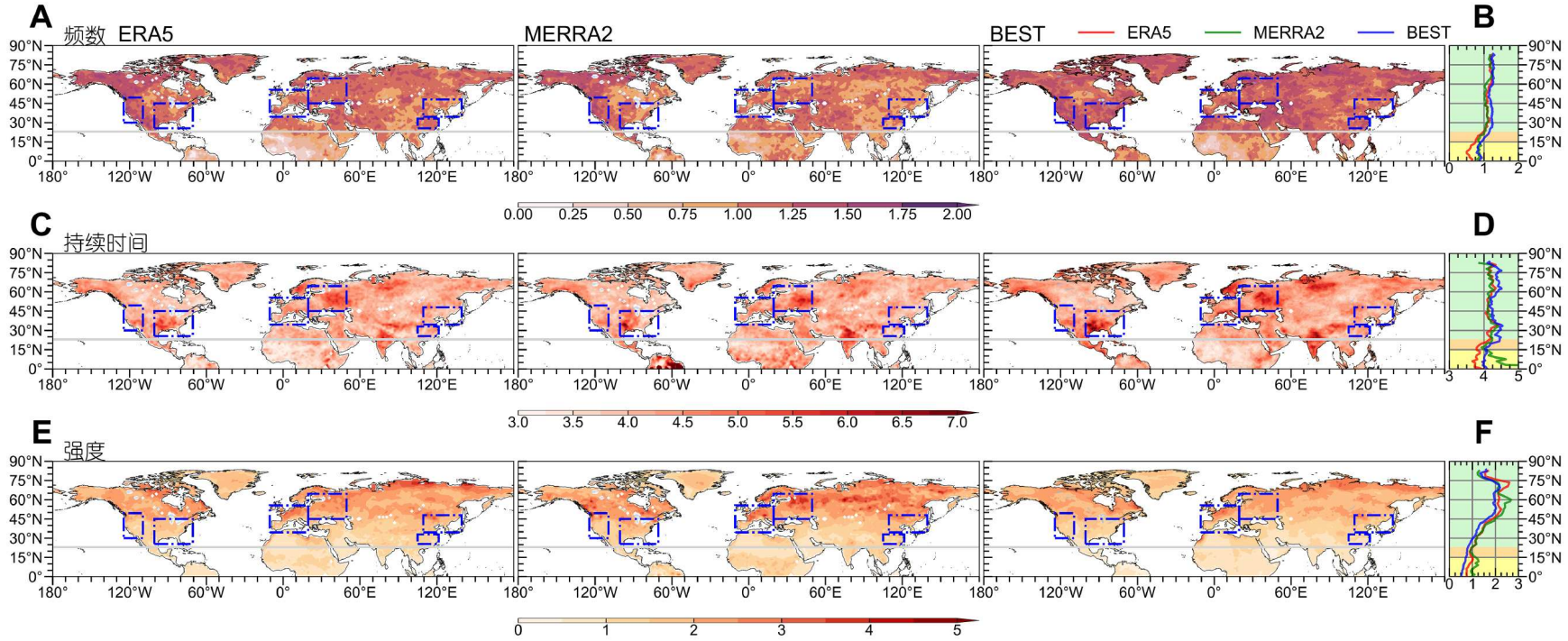


**图 1** ERA5、MERRA2及BEST数据集2001–2020年JJA陆地热浪特征的空间分布(A)频数(单位：次/年)、(C)平均持续时间(单位：天/年)和(E)平均强度(单位：℃)。(B、D和F)为对应热浪特征的纬向平均折线图(ERA5、MERRA2和BEST分别以红、绿和蓝色线标示)。蓝色框对应3.2.1节中六个社会影响热点区。灰色线表示23°N线

体预报水平。

$$\mathrm{BS}=\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2 \tag{3}$$

$$\mathrm{BSS}=\frac{\mathrm{BS_{cli}}-\mathrm{BS}}{\mathrm{BS_{cli}}} \tag{4}$$

式中，$p_i$和$o_i$分别表示第$i$个样本的预报概率及观测概率，$N$为总样本个数(Brier，1950)，$\mathrm{BS_{cli}}$为BS的参考气候态。

如公式(5)所示，当CRPS为0时说明预报累积概率密度函数与观测完全一致，对应完美预报(Hersbach，2000)。而CRPSS值越大时预报技巧越高，当CRPSS≤0时代表预报结果与历史气候态相近，即不再认为其有预报技巧。在计算$\mathrm{CRPS_{cli}}$时，针对每个预报终止日期所构建的100个历史气候态集合成员，是基于观测或再分析数据中2001–2020年该日前后2天累计5天窗口提取而成(Buizza和Leutbecher，2015；Dai等，2021)。此外，用于计算CRPSS的热浪日不仅包含观测中的热浪事件(即模式中命中和漏报的热浪日)，还包括模式中空报的热浪日，此处被定义为有超过50%的模式集合成员预报出热浪，但观测中却并未发生热浪的日子。这种处理方式可以为当前S2S模式对热浪的预报提供更全面客观的概率预报技巧评估。

$$\mathrm{CRPS}=\int_{-\infty}^{+\infty}\left(\mathrm{CDF_{frc}}(x)-\mathrm{CDF_{obs}}(x)\right)^2\mathrm{d}x \tag{5}$$

$$\mathrm{CRPSS}=\frac{\mathrm{CRPS_{cli}}-\mathrm{CRPS}}{\mathrm{CRPS_{cli}}} \tag{6}$$

式中，$\mathrm{CDF_{frc}}(x)$和$\mathrm{CDF_{obs}}(x)$分别为集合预报和观测所对应的累积概率密度函数。$\mathrm{CRPS_{cli}}$的计算同CRPS一致，但集合成员为上述的历史气候态集合成员。

## 3 模式的预报技巧

在本节中，我们从确定性和概率预报技巧两方面评估了当前S2S模式对热浪的预报水平。前者考察S2S模式集合平均对观测中热浪特征(频数、强度和持续时间)空间分布的再现能力，后者则是着眼局地考察模式集合成员的概率分布对热浪日的预报水平。

### 3.1 半球尺度的确定性预报技巧

当前多家S2S模式对北半球中高纬热浪特征空间分布的确定性预报技巧呈现一致性(图2)。从预报提前1到2周，预报技巧快速下降，提前2到3周预报技巧下降速度减缓，在超过提前3周之后，预报技巧变化不显著。为量化S2S模式对热浪特征的预报时效，此处设计了一种简单的统计检验方式，具体如下：针对某个模式的特定预报时效(提前1/2/3/4周)，将预报数据与该模式历史气候态中的任意一年随机调换，如果存在某一历史年份使计算所得ACC大于且RMSE小于原预报数据，则认为在该预报时效，模式可能已丧失对热浪特征空间分布的预报能力。

图2中显示对当前大多数S2S模式来说，基于历史气候年计算所得统计阈值在提前3–4周预报结果甚至优于原始预报数据，表明S2S模式可在提前2周左右预报北半球中高纬度地区热浪频数、强度和持续时间的空间分布特征。特别说明，最长持续时间和最大强度结果与平均持续时间和平均强度类似，且不同再分析数据集作为参考对评估结果无显著影响(网络版附图S1)。此外，ECMWF、ECCC和CNR-ISAC模式在北半球热浪特征空间分布预测结果上表现优异。

### 3.2 区域尺度的概率预报技巧

#### 3.2.1 区域划分

不同地区影响热浪发生的关键系统不同，使其往往具有不同的热浪特征，并对应着不同的预报技巧水平。在此本研究聚焦于北回归线以北地区，如图1A、1C、1E所示，不同区域间热浪特征呈现出显著差异：西欧(WEU)和东欧(EEU)地区的热浪呈现出频数少、强度强且持续时间长的特征，而在东北亚(NEA)和华中(CCH)地区，热浪发生更频繁，持续时间较短且强度也较弱。在北美地区，西北美(WNA)热浪呈现出近似东亚的频数多和持续时间短的特征，而美国东部(EUS)则呈现出近似欧洲热浪的频数少，持续时间长的特征(图5A)。因此，有必要在评估热浪预报准确性前对其区域进行划分。

极端事件的次季节准确预测具有极高的社会价值并且被视作S2S预测计划的关键一环(Vitart和Robertson，2018)。本研究选择了北半球中高纬六个人口密集区作为高社会影响的热点区域(图3)，这些区域已被先前研究广泛关注，其主要包括：WEU(Cai等，2024)、EEU(Vitart和Robertson，2018)、CCH(Zhang等，2023)、NEA(Hsu等，2020)、WNA和EUS(Domeisen等，2022)。值得一提的是在这六个区域中热浪特征分布也基本呈现出一致性(图1)，说明了分区的科学性。

#### 3.2.2 概率预报技巧

伴随预报时间提前，S2S模式对热浪日发生的预报技巧逐渐下降(图4A)。整体而言，本研究中的10个S2S模式对WEU、EEU和NEA地区的热浪大约可提前2周预测其发生，这可以通过该区域提前两周的BSS和CRPSS显著大于0，且至少30%集合成员预测出*TX*超过90%分位数体现。而对CCH、WNA和EUS地区的热浪，S2S模式

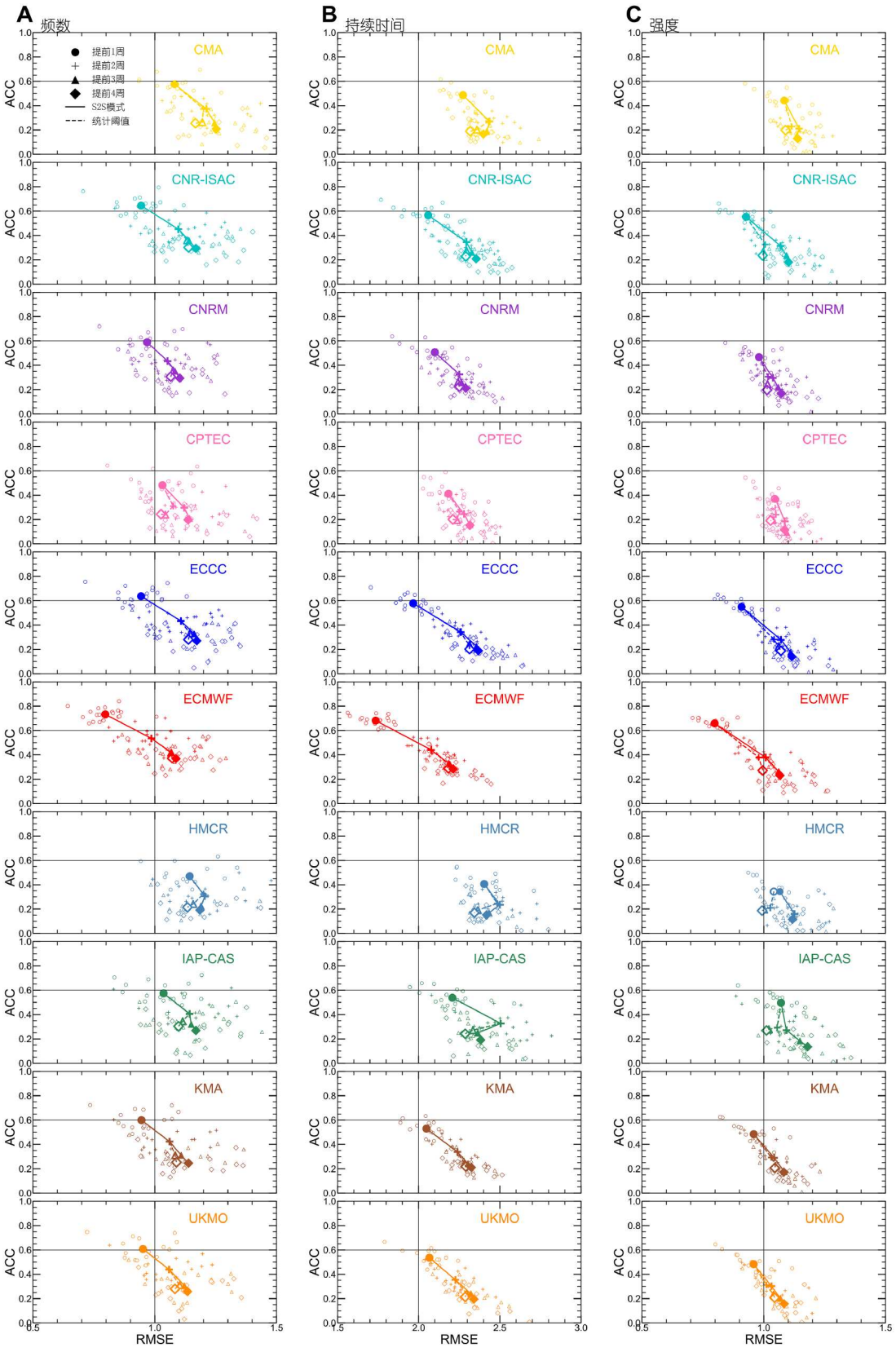


**图 2**　10个S2S模式对北半球中高纬(23°N以北)热浪特征空间分布的ACC和RMSE预报技巧(A)频数、(B)持续时间和(C)强度。参考数据集为ERA5，提前1–4周预报分别以圆形、加号、三角和菱形表示。小型空心符号代表单一年份，大型实心符号代表所有年份的气候态平均，大型空心符号表示统计检验阈值。大型实心/空心符号分别以实/虚线相连接

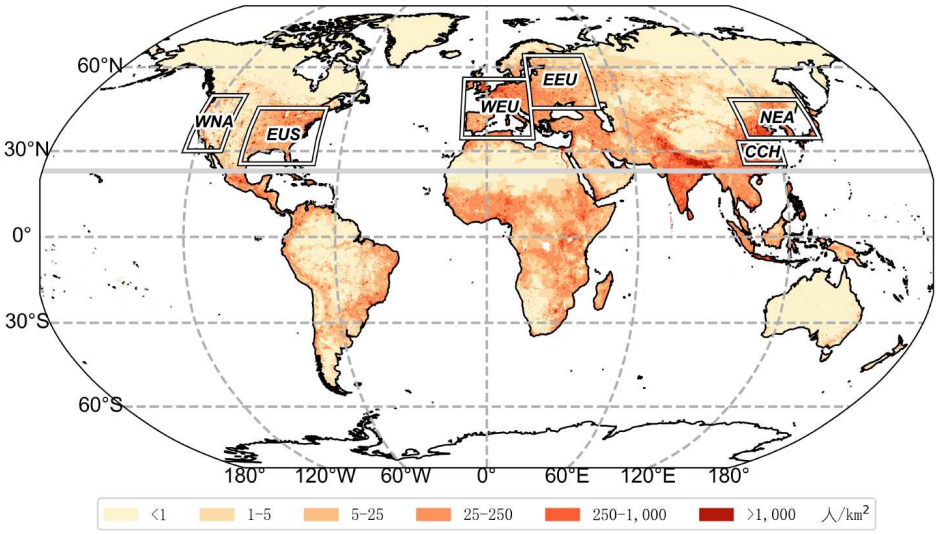


**图3** 2020年全球人口密度分布图。灰线为23°N纬线，矩形框代表六个选中区域：西欧(WEU： 34.5°–55.5°N，10.5°W–19.5°E)、东欧(EEU：45°–64.5°N，19.5°–49.5°E)、东北亚(NEA：34.5°–48°N，109.5°–139.5°E)、华中(CCH：25.5°–33°N，105°–121.5°E)、西北美(WNA：30°–49.5°N，109.5°–124.5°W)及美国东部(EUS：25.5°–45°N，70.5°–100.5°W)

仅可约提前1周预测其发生(图4A，4B)。集合预报概率密度分布函数(PDF)的偏度和峰度也伴随预报时间提前而逐渐减小(图4B)。对WEU、EEU、NEA和CCH地区而言，S2S模式大约能在热浪发生前3周预报出暖异常，可以通过该时段内至少40%的集合成员预报出*TX*达到“正异常”水平来体现。而在北美地区，这一预报时效则可被延伸至提前4周左右(图4B)。

这表明先前广泛基于热浪个例的评估工作可能在一定程度上高估了S2S模式对热浪的预报技巧。并非所有热浪均能在其发生前2周被提前预测，在CCH、WNA和EUS地区，这一预报时效被高估了一周左右；同时，也不是所有热浪均可以在提前4周被模式预报出暖异常，这一预报时效在欧洲和东亚地区同样被高估了一周左右。该结论呈现出显著的区域差异性，整体表现为：欧洲地区热浪预报技巧更高，特别是WEU区域，而东亚和北美地区预报技巧更低，特别是CCH区域。为进一步验证上述结论的稳健性，本研究还采用了其他再分析数据集作参考进行评估，所得结论与先前并无太大差异(附图S2，S3)。

强度越强持续时间越长的热浪往往对应着更高的预报技巧(Wulff和Domeisen，2019)。以预报性能优异的ECMWF为例，在提前1–4周，绝大部分地区均呈现出持续时间越长的热浪预报技巧也越高(图5B)。而WNA区域除外，这可能与该地区仅经历过一次长持续时间的热浪有关，参考样本少具有偶然性，在此不多做讨论。基于上述发现，我们定义了短期型和长期型两种热浪：前者被定义为持续时间不超过5天的热浪，后者被定义为持续时间大于5天的热浪。在此分类基础上，本研究进一步评估了当前S2S模式对长短期型热浪的预报技巧。

如图6A–6F所示，长期型热浪较短期型热浪预报技巧更高。这一差异在提前1–2周尤为显著，表现为该时段内模式集合成员对长期型热浪预报的PDF分布有更大的偏度和峰度。而当预报时间提前至3–4周，两种类型热浪间的差异逐渐减少，其PDF分布近乎重合。具体而言，长期型热浪大约可被提前2周预测发生，而短期型热浪仅可约提前1周被预测发生，可以通过至少30%集合成员预测出*TX*异常超过90%分位数体现。使用再分析数据所得结论与上述结果类似，详见附图S4和S5。

如前所述，欧洲地区热浪的预报技巧更高，可能是因为该区域热浪发生以长期型为主(图5A)，也因此其往往对应更高的预报技巧。但这些现象又引出了一个新的问题，造成热浪预报技巧区域差异性和持续时间差异性的原因是什么?

## 4 热浪预报技巧区域性及持续时间差异性的潜在成因

穆穆等(2017)指出，预报技巧应被理解为对可预报性的一种估计，而可预报性可以被视作是预报技巧的上限。由第3节描述可知S2S模式对热浪的有效预报时间在1个月以内，在此时间范围内，可预测信号的主要来源是大气及陆面过程(Mariotti等，2018)。而陆面的两个主要可预测信号源为土壤湿度和积雪(Dirmeyer等，2019)，

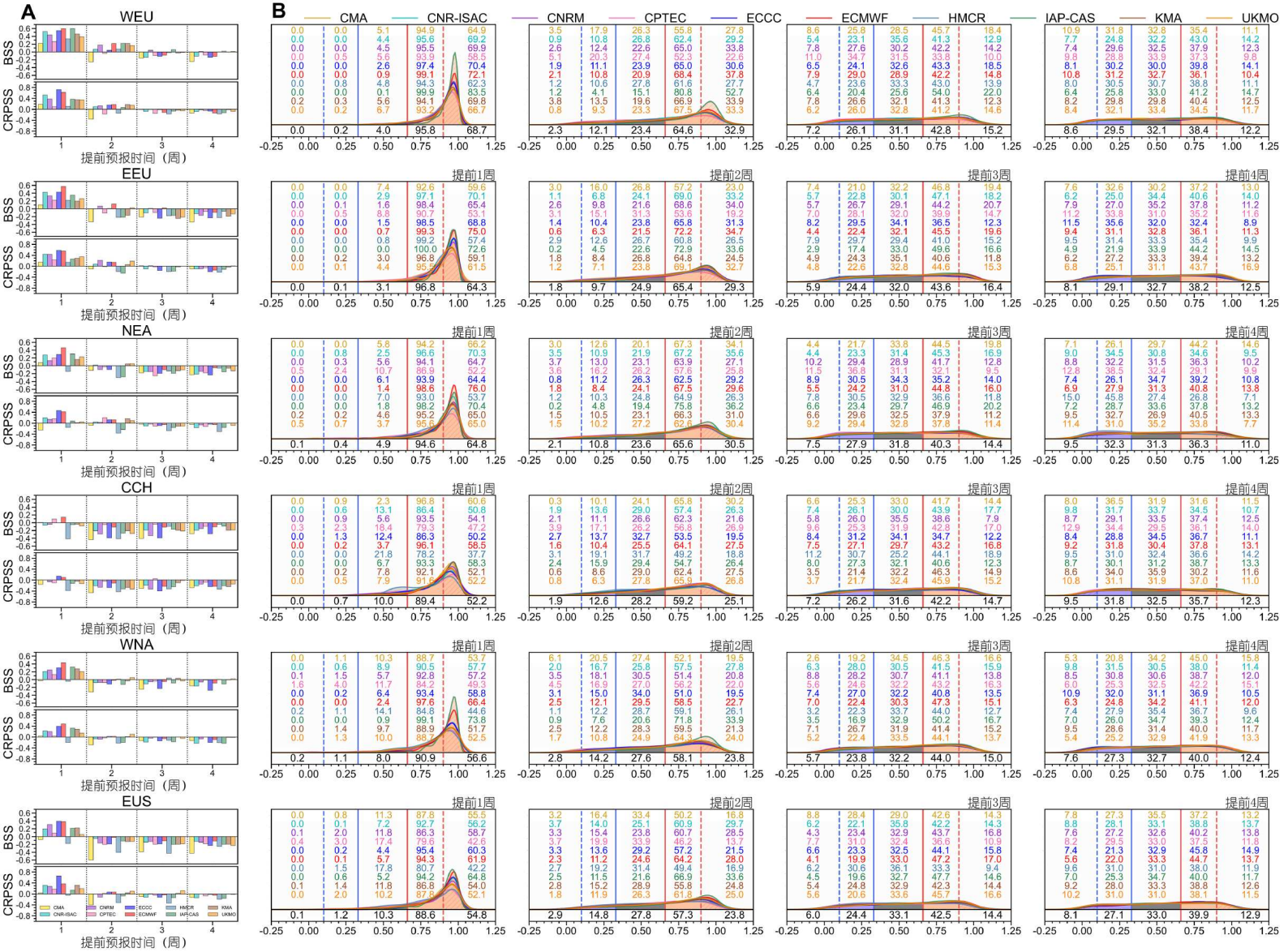


**图 4**　S2S模式对北半球中高纬六区域热浪的提前1–4周概率预报技巧（A）预报热浪日发生的BSS和CRPSS（B）10个S2S模式对BEST观测数据集中实际发生热浪日，预测的各*TX*分位数对应的集合成员PDF分布，从上至下为6个区域，从左至右为提前1–4周。图（B）中x轴表示模式预报出的*TX*异常相较于模式气候态中的百分位数。PDF曲线由红色虚竖线（90%）、红色实竖线（66.7%）、蓝色虚竖线（10%）和蓝色实竖线（33.3%）间隔，填色区域对应“负异常”（<33.3%，蓝色）、“正常”（33.3%–66.7%，灰色）和“正异常”（>66.7%，红色）。不同颜色数字（%）代表不同S2S模式在该分位数区间的集合成员数占比，子图下方的黑色数字表示多模式平均的集合成员数占比

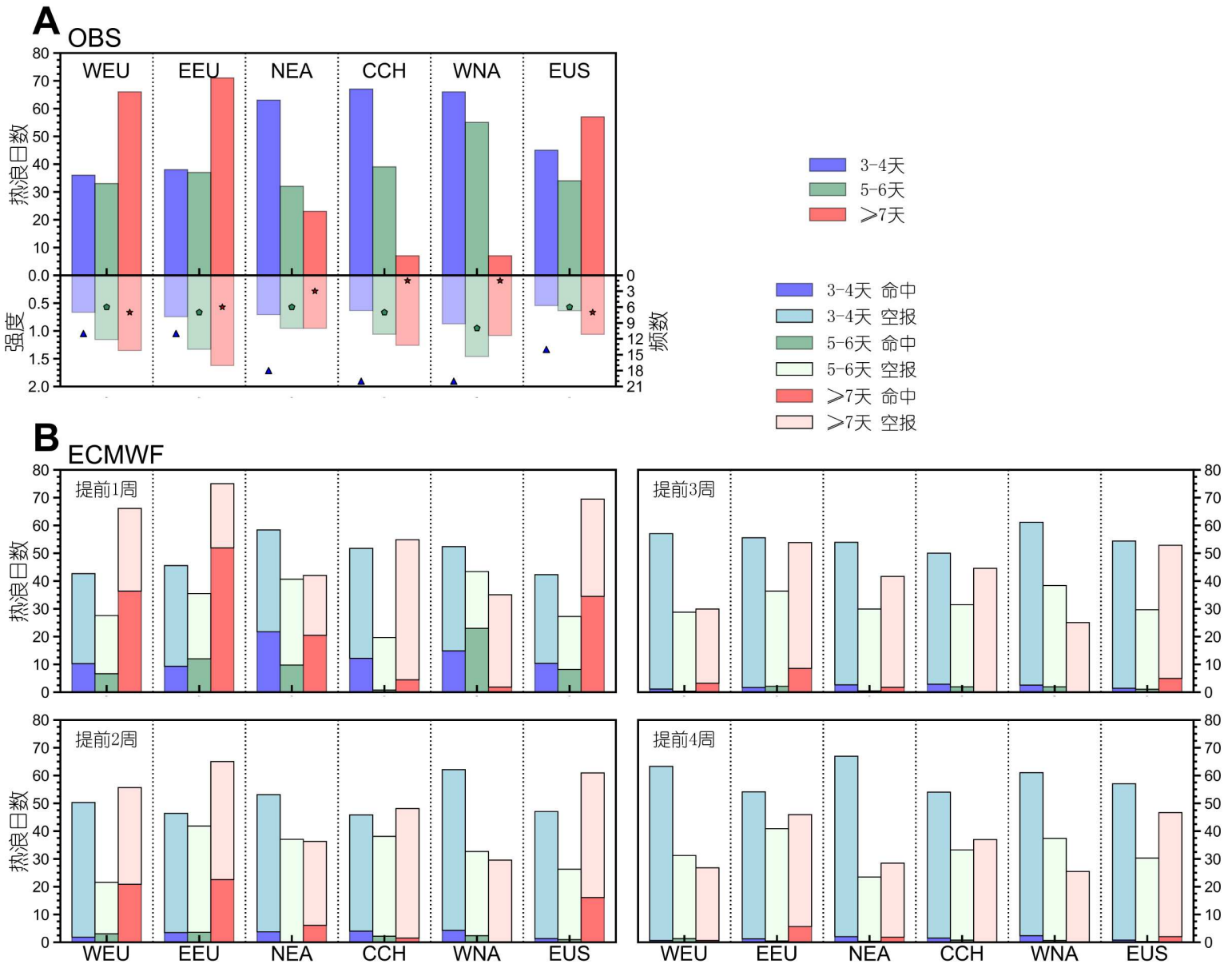


**图 5** 不同持续时间长度的热浪观测特征及在ECMWF中的预测性能(A)2001–2020年BEST数据中六个区域观测的持续时间长度为3–4天(蓝)、5–6天(绿)和≥7天(红)热浪的累计天数、平均强度(单位：℃)和累计频数(单位：次，以符号标记)。(B)ECMWF模式提前1–4周预测的三种持续时间长度热浪命中日数(深色柱状)和空报日数(浅色柱状)

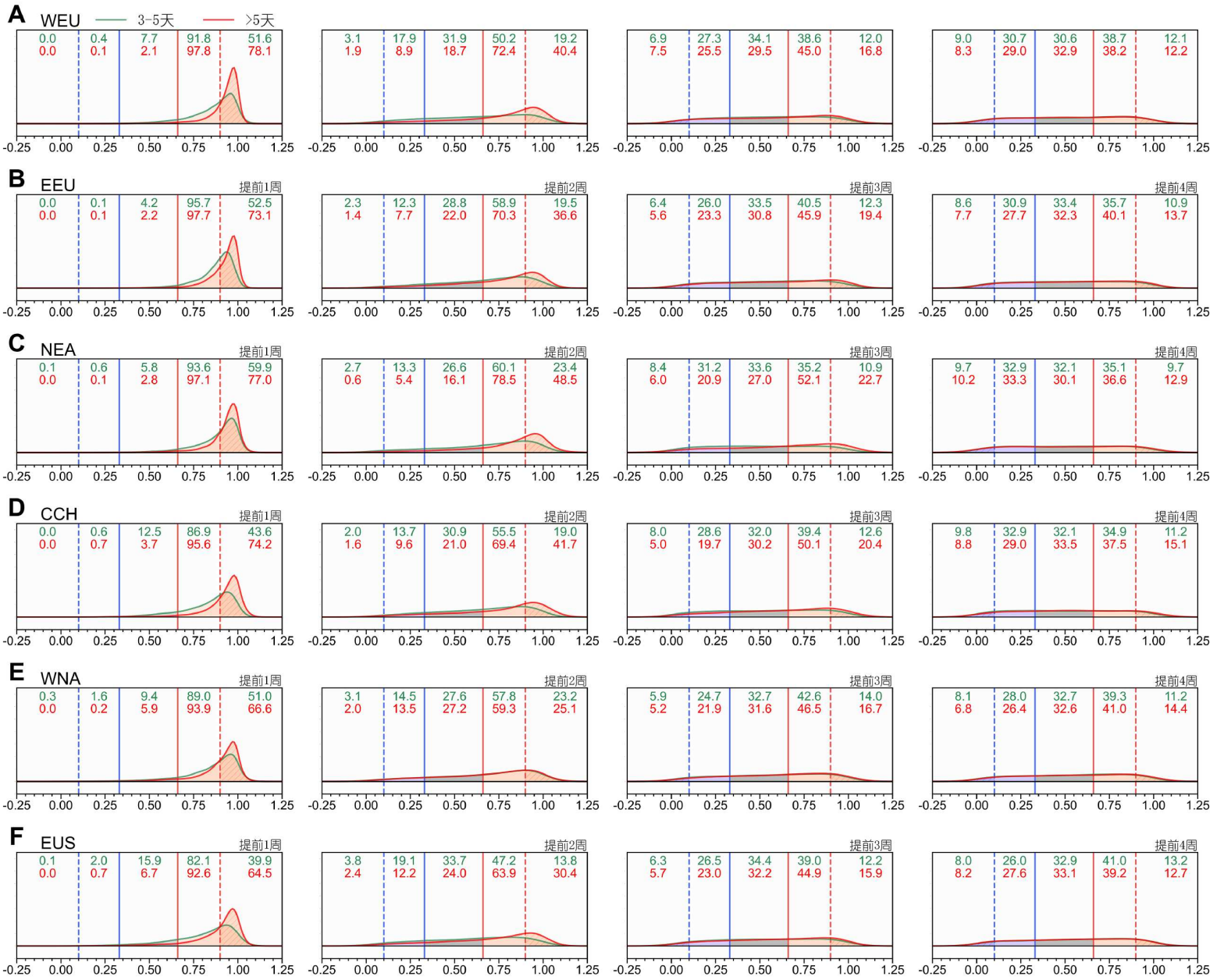


**图 6** S2S模式在长短期型热浪预报中各*TX*分位数对应集合成员PDF分布(A–F)从上至下分别为WEU、EEU、NEA、CCH、WNA和EUS六个区域。图中仅展示提前1–4周(从左至右)多模式平均结果，长短期型热浪分别以红色和绿色表示，参考数据集为BEST。其余图片要素设置参见图4

考虑到本研究聚焦于北半球夏季，而积雪主要在春季起作用，到夏季时源自积雪的可预测信号可能已被转换为其他形式的记忆信号存储起来(如土壤湿度)。因此本节中陆面可预测信号来源仅考虑土壤湿度。在本节中，我们将尝试从预报技巧、可预报性和真值关系的角度去说明造成热浪预报技巧差异性的成因。

### 4.1 预报技巧区域差异性成因

如图5A所示，WEU和EEU区域热浪的持续时间越长，其强度往往也更强，而在东亚和北美地区持续时间与热浪强度间的这一正相关关系却并不显著，特别是NEA和WNA区域。由此引出一项猜测：WEU和EEU区域的热浪往往受单一天气系统控制，天气系统越强，热浪持续时间越长，强度也越强。而正因为影响系统单一，其预测准确率也往往更高。而东亚和北美地区的热浪则相反，其影响系统较多且物理过程复杂，不同系统间相互牵制，热浪难以被单一系统长期控制。因而热浪持续时间与强度间的关系并不显著，预测准确性也较低。为简单起见，在此我们选择预报技巧较高的欧洲(WEU与EEU)和预报技巧较低的东亚(NEA与CCH)来说明热浪预报技巧区域差异性的背后成因。

#### 4.1.1 可预报性上限

区域环流与其地表气温(SAT)间往往联系密切。四个区域所有日数的$Z_{500}$和SAT间均存在显著的正相关关系(图7)。单就热浪日而言，WEU、EEU和NEA区域$Z_{500}$及SAT间同样呈现出显著的正相关关系，而在CCH区域这一关系却并不显著，说明该地区高压异常强度和热浪强度间并无明确对应关系。进一步可以发现WEU和EEU区域的所有热浪日均出现在高压异常环流背景下，而在NEA和CCH地区，却仍有少部分热浪日出现在低压异常的环流背景下(图7C，7D)。侧面昭示了东亚地区环流背景场较欧洲地区更复杂，驱动东亚热浪出现的物理过程更具混沌性，因此对应该地区更低的热浪可预报性。为进一步验证上述结论，我们还考察了不同区域热浪期间环流异常的相似性。如附图S6所示，欧洲地区热浪期间的环流相似性明显大于东亚地区，而且无论热浪持续时间长短，这一结论均成立。同样表明东亚地区形成热浪的物理过程较欧洲地区更具混沌性。

进一步从$Z_{500}$正负异常背景下出现热浪日的对流层环流垂直结构特征来分析其区域差异性。可以发现四个区域在$Z_{500}$正异常下出现的热浪日，其整层环流呈现准正压结构(附图S7A–S7D)，而在$Z_{500}$负异常下出现的热浪日，其环流结构则呈现出斜压结构特征，主要表现为对流层低层为低压异常，而对流层中层位于脊前槽后的特征(附图S7E，S7F)。如图8A–8D所示，WEU和EEU在$Z_{500}$正异常下的热浪环流型与北半球中纬度地区5波型和7波型的环球Rossby波列有关(Kornhuber等，2020；Cai等，2024)。而在NEA和CCH区域，触发热浪的环流型涉及多个遥相关系统，既包括北半球中高纬度地区的东传波列也包含热带地区向北传的波列(Hsu等，2017；Nie等，2024；Xiao等，2024)，进一步说明了东亚地区环流的复杂性。而就NEA和CCH出现在$Z_{500}$负异常下的热浪日而言，该地区上空呈现出脊前槽后的环流形势，对流层中上层的下沉运动在此类背景下热浪日的出现中发挥了关键作用(图8E，8F)。

#### 4.1.2 S2S模式捕捉区域可预测信号的能力

土壤湿度也是$T$2m次季节可预测信号的关键来源之一，四个区域的绝大部分热浪均出现在土壤干异常背景下(附图S8)。那当前S2S模式对大气和陆面可预测信号来源的预报技巧如何? 如图9A，各S2S模式对环流的预报基本均呈现出在欧洲(WEU和EEU)较东亚(NEA和CCH)的预报技巧更高。同时几乎所有模式对欧洲土壤湿度的预报技巧也明显高于东亚，除CMA模式外，这可能与其未进行土壤湿度同化有关(图9B)。这表明S2S模式对直接影响欧洲热浪的大气和陆面过程具有更强的模拟能力，因而S2S模式在该地区热浪预测中具有更高的预报技巧。此外，欧洲观测站点分布较东亚更密集，这有助于降低该地区观测场的初始不确定性，为保障更高的热浪预报技巧提供了有力支撑(附图S9)。

### 4.2 预报技巧持续时间差异性成因

在分析热浪持续时间引起热浪预报技巧差异性的成因时，本研究聚焦于此差异性更显著的欧洲地区进行考察，且该区域也拥有更多的长期型热浪样本量以供研究。欧洲的长期型热浪往往发生在长期维持的强阻塞环流背景下，其相关关系异常显著(图10A，10B)。这说明持续时间更长的热浪常常对应更简单稳定的环流背景场。此外，虽然WEU区域土壤湿度与$T$2m的相关关系更显著说明该地区陆-气耦合作用更强，但不论在WEU还是EEU，与短期型热浪相比，总会有更多占比的长期型热浪发生在土壤湿度干异常背景下(图10C，10D)。这表明在长期型热浪期间，陆面过程也为该类热浪预测提供了更多可预测信号。

本研究也评估了S2S模式对两类持续时间型热浪期间的环流及土壤湿度异常的预报技巧。如附图S10所示，在提前1周时，模式对长期型热浪期间环流及土壤湿度异

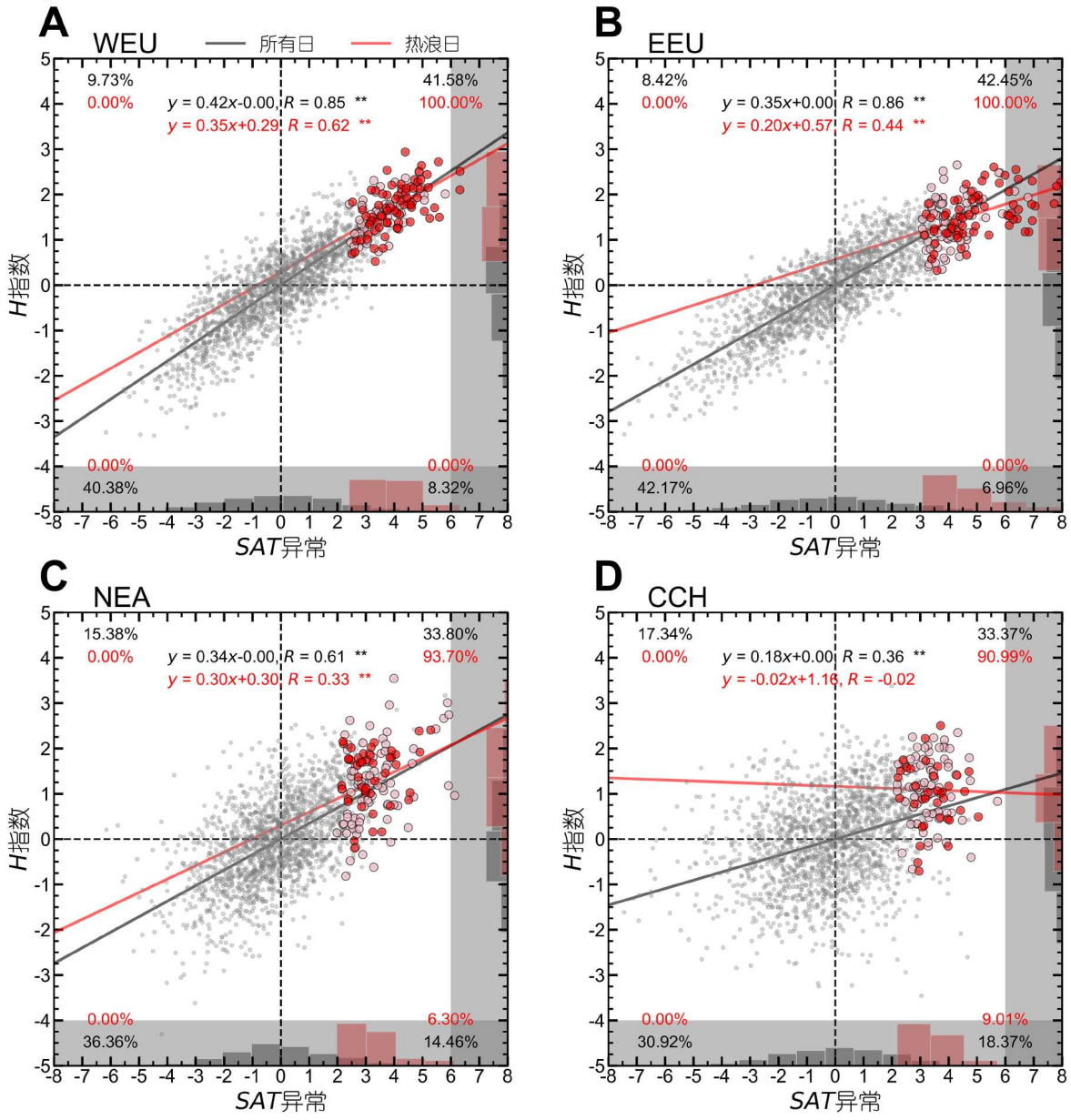


图 7　标准化$Z_{500}$异常和$T2$m异常关系散点图(A–D)分别为WEU、EEU、NEA和CCH区域(2001–2020；ERA5)。灰色、粉色和红色打点分别代表所有日、短期型和长期型热浪日，线性拟合曲线也在图中一并展示。各子图右侧和下方的条形图分别对应$Z_{500}$异常和$T2$m异常的PDF分布。图中还标记了线性拟合、相关关系及$t$检验的统计显著性水平(“*”表示95%和“**”表示99%)。每个象限的底角标注了该象限内格点数占总格点数的百分比，其中红色代表热浪日，黑色代表所有日

常的预报并未呈现出显著优越性。这表明S2S模式对欧洲热浪期间关键物理过程的模拟能力几乎不随热浪持续时间的变化而变化。而在提前2周时，模式对长期型热浪期间环流异常的预报技巧更高，这可能归因于预报时段已临近大气的可预报性上限(Lorenz，1969)，如前所述，长期维持的阻塞系统和土壤湿度干异常背景能够提供更多的可预测信号。综上，热浪持续时间引起的预报技巧差异性可能主要受大气可预报性上限的制约。

## 5　讨论和总结

先前研究大多基于极端个例且未考虑空报影响，在一定程度上高估了S2S模式对热浪的预报技巧。当前S2S模式对北半球中高纬(23°N以北)热浪特征(频数、持续时间和强度)空间分布的预报技巧在提前2周左右。就区域尺度而言，S2S模式在WEU、EEU和NEA地区大约可提前2周预报热浪发生，而在CCH、WNA和EUS地区仅可约提前1周预报热浪发生，对后者的预报时效在先前研

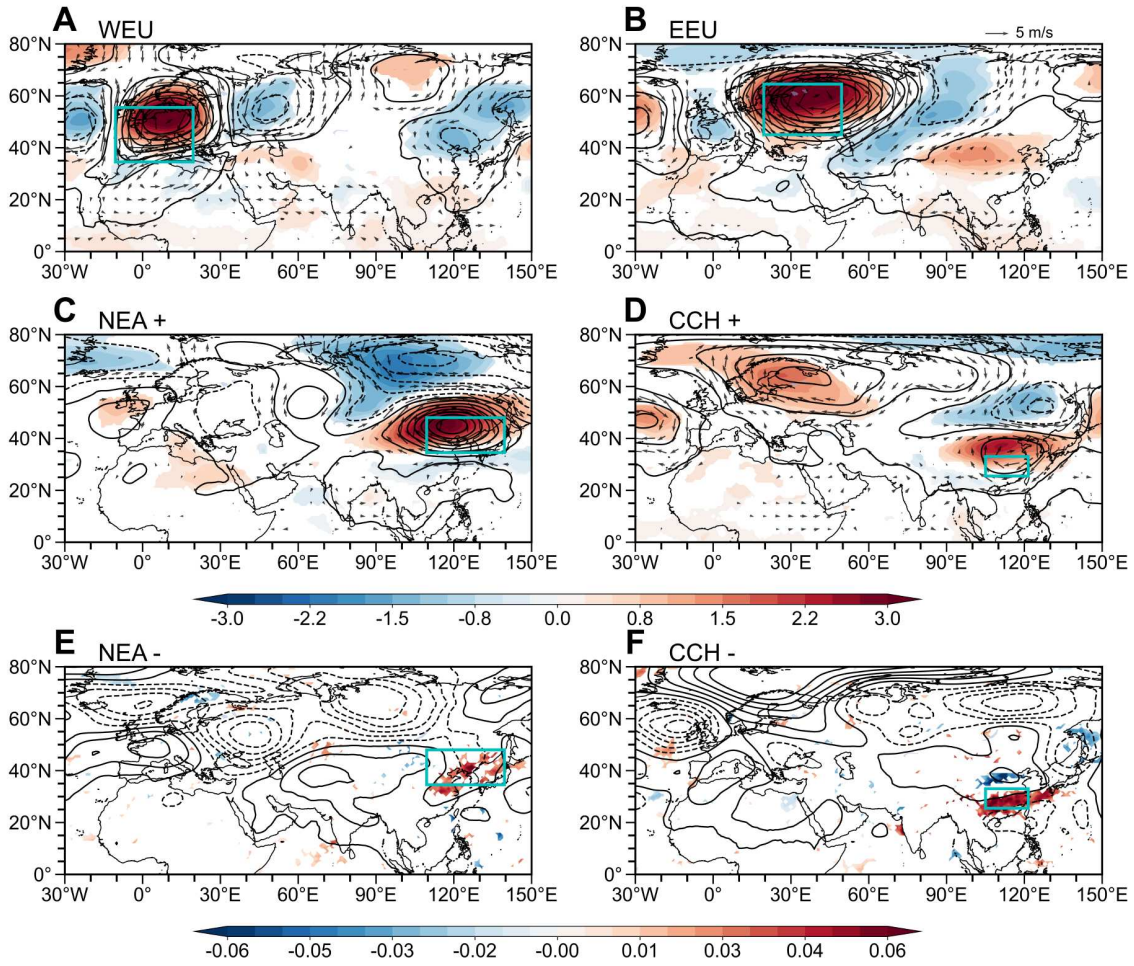


**图 8** 2001–2020年ERA5再分析数据集正负$Z_{500}$异常下热浪日的合成分析环流场(A–D)分别为WEU、EEU、NEA和CCH在$Z_{500}$正异常下热浪日的情况。填色为500 hPa气温异常(单位：℃)，等值线为位势高度异常(单位：gpm)，矢量箭头为风异常(单位：m/s)。图中实线对应正异常，虚线代表负异常。(E和F)类似于(A–D)，但为NEA和CCH区域在$Z_{500}$负异常下热浪日的情况，填色代表垂直速度异常(单位：Pa/s)。矩形框标记了WEU、EEU、NEA和CCH区域。图中仅展示统计显著性水平超过95%的填色和矢量箭头

究中被高估1周左右(Domeisen等，2022)。S2S模式对热浪的潜在预报时效更久：在北美地区约可在热浪发生前4周预报暖异常，而在欧洲和东亚地区，这一提前时间仅为3周左右，后者的预报时效同样被先前研究高估约1周(Domeisen等，2022)。

S2S模式对热浪的预报技巧呈现区域差异性，主要表现为欧洲较东亚和北美地区热浪预报技巧更高。究其原因，欧洲地区环流混沌性更弱，热浪可预报性更强。同时模式对影响热浪的关键物理过程(大气和陆面)模拟能力更强，减少了模式不确定性；观测站点分布更多降低了初始分析场的不确定性，二者共同作用使得S2S模式在欧洲地区捕获可预测信号的能力更强。“更强可预报性叠加更强模拟能力”(“强强联合”)保障了欧洲热浪预报技巧更高(附图S11)，而东亚地区则与之相反。

S2S模式对热浪的预报技巧呈现持续时间差异性，长期型热浪往往对应更高的预报技巧。具体表现在，S2S模式能提前2周预报长期型热浪出现，但仅能提前1周预报短期型热浪出现。而当预报时间线继续拉长，二者间的差异不再明显。以此差异性显著的欧洲地区为例，长期型热浪往往出现在持续稳定的强阻塞背景下，环流可预报性更高，同时具有记忆性的土壤湿度常常也会提供更多的可预测信号。因而热浪预报技巧的持续时间差异性主要是受大气内禀可预报上限制约引起。但在S2S模式预报中总倾向于低估热浪的强度和持续时间，这种模式偏差是否也会影响长短期型热浪预报技巧的差异性，在此并未讨论。如有影响，与可预报性上限相比，其各自的相对贡献又占多少，这些问题有待进一步探讨。

本研究在分析热浪预报技巧区域差异性成因时，仅对差异性更明显的欧洲和东亚地区进行了说明，北美地区是否有类似结论有待进一步验证。在讨论热浪预报技

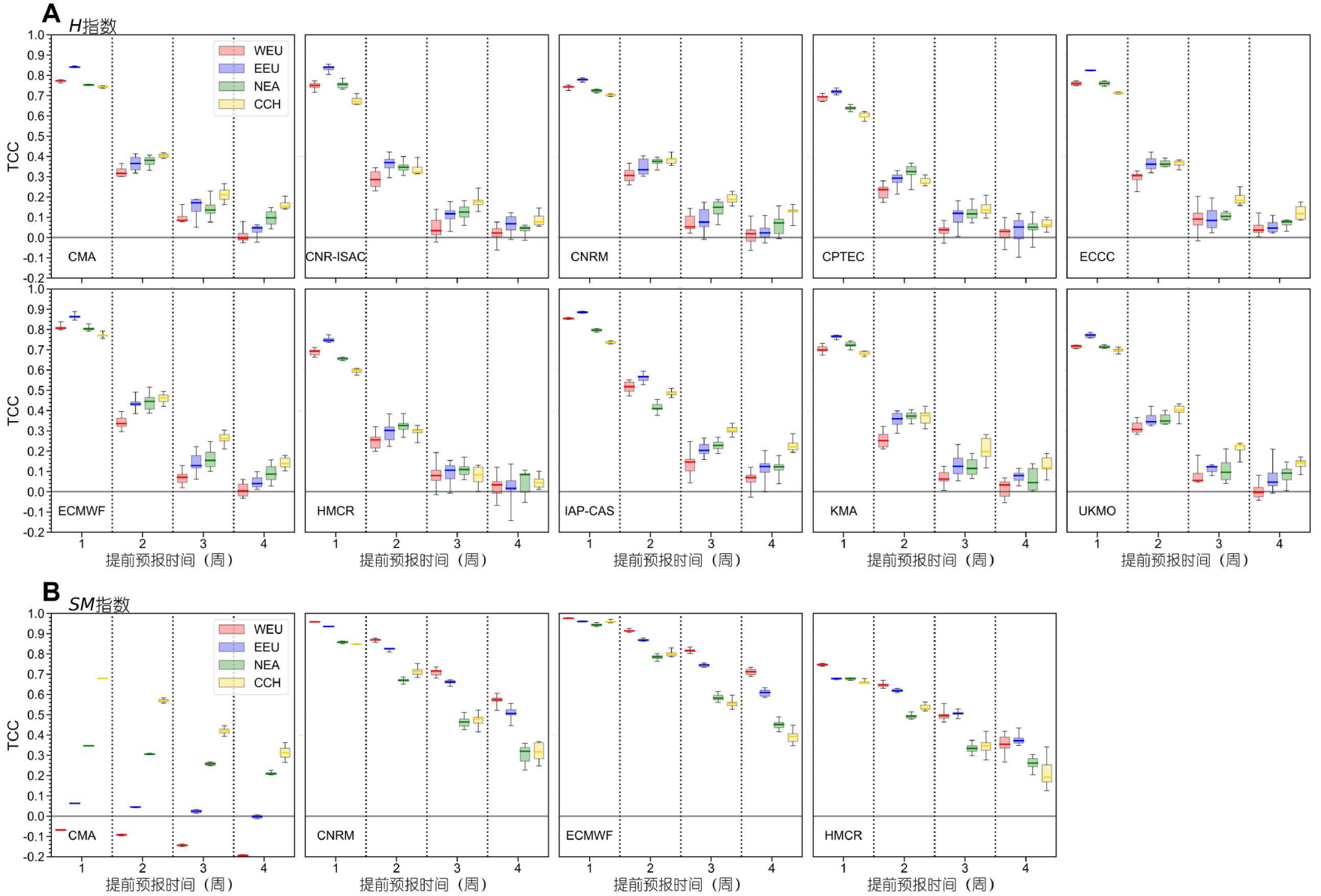


**图 9** S2S模式对所有日$Z_{500}$异常和$SM_{100}$异常提前1–4周的TCC预报技巧箱线图（A）$Z_{500}$异常和（B）$SM_{100}$异常（此处仅展示输出$SM_{100}$的四个S2S模式）。WEU、EEU、NEA和CCH分别以红、蓝、绿和黄色表示。箱线图的上界和下界对应最大最小值，中位数以实线表示

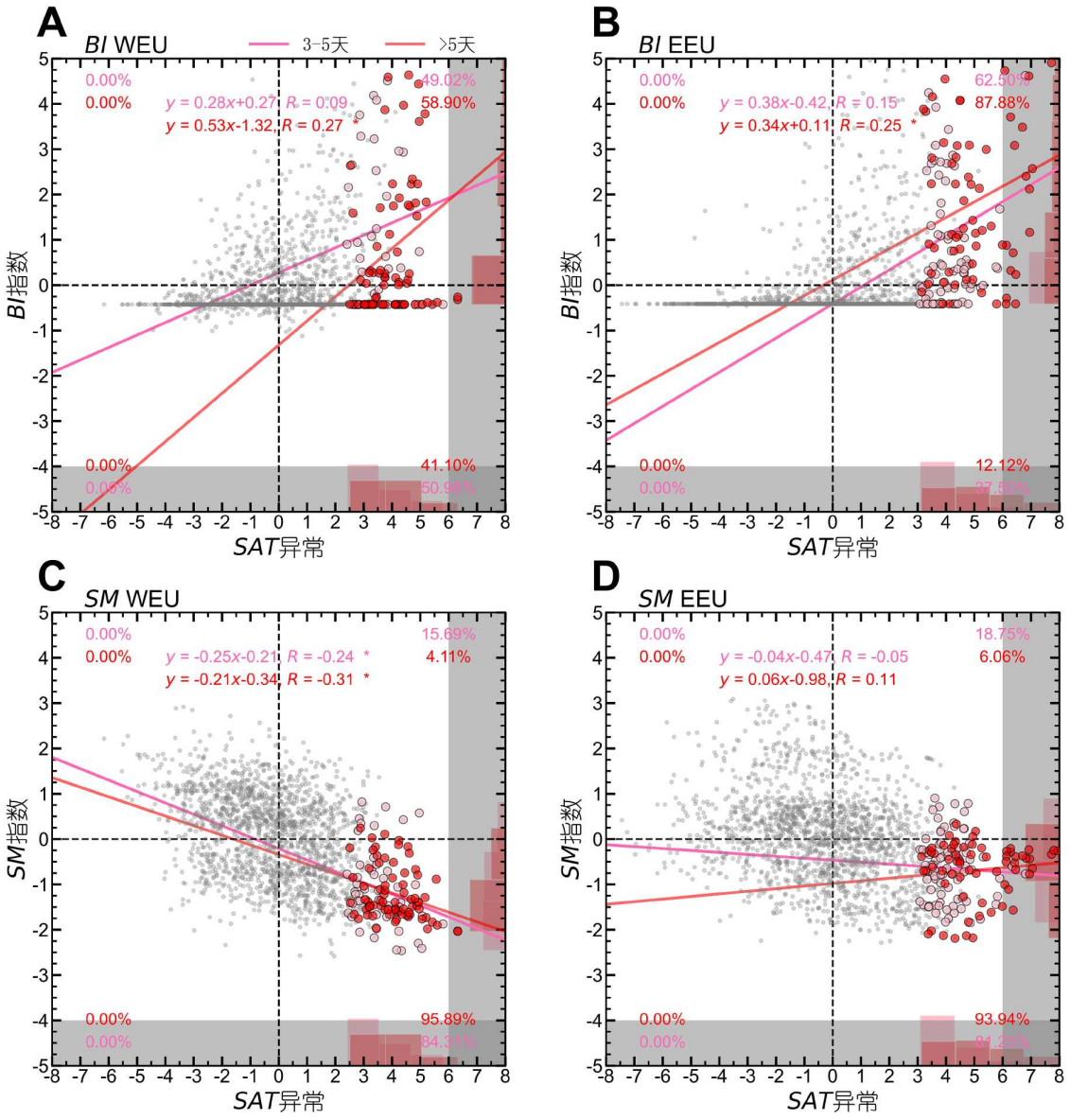


**图 10** *BI*异常和标准化$SM_{100}$异常与*T*2m异常关系散点图(A, B)*BI*异常与*T*2m异常间关系，(C, D)$SM_{100}$异常与*T*2m异常间关系。图片要素设置参考图7，粉色打点为短期型热浪，红色打点为长期型热浪

巧持续时间差异性成因时，本研究着重强调了局地驱动因子(阻塞系统及土壤湿度)的作用，而更大尺度的行星波、海表面温度、海冰、积雪和平流层过程可能也将通过与大气相互作用进而影响热浪的发生发展，此处并未考虑。同时，本研究也未进一步探究热浪持续时间差异性是否会影响到热浪预报技巧的区域差异性，例如：WEU、EEU和EUS均经历了较多的长期型热浪，但欧洲的预报技巧却要显著高于EUS(图5A)，探究此问题将有助于进一步理解各区域热浪特征的成因机制。此外，热浪持续时间越长预报技巧越高这一结论是否足够稳健，在WNA地区仍有待进一步讨论，需要在积累更多观测样本后再进行分析。更进一步的，北美地区似乎可以提前热浪发生更久预报暖异常，其原因是否在于该地区在更长预报时间上受海洋影响，有待进一步考究。

先前研究分析热浪机制成因时大多聚焦于异常的高压系统，成云减少，更多太阳短波辐射抵达地面，感热增加，同时异常的下沉绝热增温也为极端高温的出现提供了有利条件(Barriopedro等，2023)。但在此研究中发现东亚地区即使在少数低压异常环流背景下也可能出现高温情况，其背后的详细物理过程是什么有待深入研究。Liu等(2024)指出，长江流域热浪发生后的前几天成云减少感热增加利于高温出现，但在热浪后期成云增加感热

减少使得热浪逐渐衰减趋于结束。由此说明基于全热浪时段的统计分析可能是不够的，我们建议从动态过程角度研究热浪发生发展及结束的物理机制，着眼微观以期为热浪的次季节预测提供更多机遇窗口。

东亚地区环流形势复杂多变的成因机制多种多样：西侧青藏高原热源能影响南亚高压的位置和强度(Liu等，2020)，东侧西太平洋暖池对流会影响西太平洋副热带高压的位置(黄荣辉和孙凤英(1994))，北侧有中高纬度地区东传的Rossby波列遥响应(Li等，2019)及南侧热带对流会触发向北传播的季节内振荡波列(Hsu等，2017)。此外，东亚大气环流还受到土壤湿度等局地陆-气反馈影响(Jiang等，2023)。相较之下，位于中高纬度地区的欧洲Burgur数很小，以平流运动为主，一般为准正压结构控制(Hoskins，1987；吴国雄等，2002)，环流形势简单。因此，理解两类处于正斜压环流状态下热浪的成因机制对提升S2S模式的预报技巧具有非凡意义。进一步地，针对二者特征而产生相应的集合预报初始扰动方案，可能将有助于提升S2S模式对热浪的预报技巧，这也是本文作者们未来将致力探索的方向。次季节试验采用多模式集合预报来考量模式系统性偏差的不确定性影响(Pegion等，2019)，参考该试验，若对S2S预测计划中多模式同样进行集合预报，是否能改进热浪预测，这也是一个有价值的问题待进一步探索。

本文对热浪预报技巧差异性成因的解释相对简略，关于可预报性上限、模式不确定性和初始场不确定性在其中的相对贡献大小并未深入探究，但我们认为有必要开展相关可预报性研究以加深对次季节预测的认识。令人鼓舞的是在本研究过程中发现，S2S模式回算数据同样可以再现图7和附图S8中的可预报性差异现象，这说明当前S2S模式有能力对热浪的可预报性上限及模式捕捉可预测信号的能力进行量化研究。开展此类研究将增强人们对热浪次季节预测的认识，对如何减少初值及模式不确定性提供理论支撑。还将为如何选择集合预报方案和如何优化观测站网布局提供参考，以期在有限资源条件下提供更高水平的热浪次季节预测。而对极端事件更高质量的预测终将强化决策策略，并提升社会应对极端风险与灾害的韧性和能力。

**致谢** 感谢三位匿名审稿人提出的宝贵意见，这些意见有助于改进本文。感谢国家重大科技基础设施“地球系统数值模拟装置”(https://cstr.cn/31134.02.EL)提供的技术支持。

**补充材料** 本文的补充材料见网络版(http://earthcn.scichina.com)。补充材料为作者提供的原始数据，作者对其学术质量和内容负责。

## 参考文献

黄荣辉, 孙凤英. 1994. 热带西太平洋暖池的热状态及其上空的对流活动对东亚夏季气候异常的影响. 大气科学, 18: 141–151
穆穆, 段晚锁, 唐佑民. 2017. 大气海洋运动的可预报性思考与展望. 中国科学: 地球科学, 47: 1166–1178
王雨晴, 周文, 王春在. 2024. 北半球高强度-长持续时间型极端热浪事件1980年以来频率快速增加的机制分析. 中国科学: 地球科学, 54: 2133–2151
吴国雄, 丑纪范, 刘屹岷, 等. 2002. 副热带高压形成和变异的动力学问题. 自然科学进展. 314
Barriopedro D, García-Herrera R, Ordóñez C, et al. 2023. Heat waves: Physical understanding and scientific challenges. Rev Geophys, 61: e2022RG000780
Brier G W. 1950. Verification of forecasts expressed in terms of probability. Mon Weather Rev, 78: 1–3
Brunner L, Voigt A. 2024. Pitfalls in diagnosing temperature extremes. Nat Commun, 15: 2087
Buizza R, Leutbecher M. 2015. The forecast skill horizon. Q J R Meteorol Soc, 141: 3366–3382
Cai F, Liu C, Gerten D, et al. 2024. Sketching the spatial disparities in heatwave trends by changing atmospheric teleconnections in the Northern Hemisphere. Nat Commun, 15: 8012
Center for International Earth Science Information Network-CIESIN-Columbia University. 2017. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10, doi: 10.7927/H4DZ068D
COPES W. 2005. The World Climate Research Programme Strategic Framework 2005-2015: Coordinated Observation and Prediction of the Earth System (COPES). WMO. 65
Dai G, Mu M, Li C, et al. 2021. Evaluation of the forecast performance for extreme cold events in east Asia with subseasonal-to-seasonal data sets from ECMWF. J Geophys Res-Atmos, 126: 2020JD033860
Dirmeyer P A, Gentine P, Ek M B, et al. 2019. Land surface processes relevant to sub-seasonal to seasonal (S2S) prediction. In: Robertson A W, Vitart F, eds. Sub-Seasonal to Seasonal Prediction. Amsterdam: Elsevier. 165–181
Domeisen D I V, White C J, Afargan-Gerstman H, et al. 2022. Advances in the subseasonal prediction of extreme events. Bull Amer Meteorol Soc, 103: E1473–E1501
Domeisen D I V, Eltahir E A B, Fischer E M, et al. 2023. Prediction and projection of heatwaves. Nat Rev Earth Environ, 4: 36–50
Ford T W, Dirmeyer P A, Benson D O. 2018. Evaluation of heat wave forecasts seamlessly across subseasonal timescales. NPJ Clim Atmos Sci, 1: 20
Gelaro R, McCarty W, Suárez M J, et al. 2017. The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim, 30: 5419–5454
Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast, 15: 559–570
Hersbach H, Bell B, Berrisford P, et al. 2020. The ERA5 global reanalysis. Q J R Meteorol Soc, 146: 1999–2049
Hobday A J, Alexander L V, Perkins S E, et al. 2016. A hierarchical approach to defining marine heatwaves. Prog Oceanogr, 141: 227–238
Hoskins B J. 1987. Diagnosis of forced and free variability in the atmosphere. In: Cattle H, eds. Atmospheric and Oceanic Variability. Bracknell, UK: James Glaisher House. 57–73
Howard S, Krishna G. 2022. How hot weather kills: The rising public health dangers of extreme heat. BMJ, 378: o1741
Hsu P C, Lee J Y, Ha K J, et al. 2017. Influences of boreal summer intraseasonal oscillation on heat waves in monsoon Asia. J Clim, 30: 7191–7211
Hsu P C, Qian Y, Liu Y, et al. 2020. Role of abnormally enhanced MJO over

the Western Pacific in the formation and subseasonal predictability of the record-breaking Northeast Asian heatwave in the summer of 2018. J Clim, 33: 3333–3349
Jiang J, Liu Y, Mao J, et al. 2023. Extreme heatwave over Eastern China in summer 2022: The role of three oceans and local soil moisture feedback. Environ Res Lett, 18: 044025
Jyoteeshkumar reddy P, Sharples J J, Lewis S C, et al. 2021. Modulating influence of drought on the synergy between heatwaves and dead fine fuel moisture content of bushfire fuels in the Southeast Australian region. Weather Clim Extrem, 31: 100300
Kautz L A, Martius O, Pfahl S, et al. 2022. Atmospheric blocking and weather extremes over the Euro-Atlantic sector—A review. Weather Clim Dynam, 3: 305–336
Kornhuber K, Coumou D, Vogel E, et al. 2020. Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. Nat Clim Chang, 10: 48–53
Li J, Zheng F, Sun C, et al. 2019. Pathways of influence of the Northern Hemisphere mid-high latitudes on East Asian climate: A review. Adv Atmos Sci, 36: 902–921
Libonati R, Geirinhas J L, Silva P S, et al. 2022. Assessing the role of compound drought and heatwave events on unprecedented 2020 wildfires in the Pantanal. Environ Res Lett, 17: 015005
Liu W, Shi N, Wang H, et al. 2024. Thermodynamic characteristics of extreme heat waves over the middle and lower reaches of the Yangtze River Basin. Clim Dyn, 62: 3877–3889
Liu Y, Lu M, Yang H, et al. 2020. Land-atmosphere-ocean coupling associated with the Tibetan Plateau and its climate impacts. Natl Sci Rev, 7: 534–552
Lorenz E N. 1969. The predictability of a flow which possesses many scales of motion. Tellus A-Dynamic Meteorol Oceanogr, 21: 289
Mariotti A, Ruti P M, Rixen M. 2018. Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. NPJ Clim Atmos Sci, 1: 4
Meehl G A, Tebaldi C. 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305: 994–997
Mondal S, Mishra A K. 2021. Complex networks reveal heatwave patterns and propagations over the USA. Geophys Res Lett, 48: e2020GL090411
Nie Y, Ren H L, Zuo J, et al. 2024. Eurasian mid-latitude jet stream bridges an Atlantic to Asia summer teleconnection in heat extremes. Environ Res Lett, 19: 044003
Oliver E C J, Donat M G, Burrows M T, et al. 2018. Longer and more frequent marine heatwaves over the past century. Nat Commun, 9: 1324
Pegion K, Kirtman B P, Becker E, et al. 2019. The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. Bull Amer Meteorol Soc, 100: 2043–2060
Pyrina M, Domeisen D I V. 2023. Subseasonal predictability of onset, duration, and intensity of European heat extremes. Q J R Meteorol Soc, 149: 84–101
Robertson A, Vitart F. 2018. Sub-Seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting. Amsterdam: Elsevier. 569
Rohde R, Muller R, Jacobsen R, et al. 2013. Berkeley Earth temperature averaging process. Geoinfor Geostat-An Overview, 1: 1–13
Scherrer S C, Croci-Maspoli M, Schwierz C, et al. 2006. Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. Int J Climatol, 26: 233–249
Sciences N A, Earth D, Studies L, et al. 2016. Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Washington, DC: National Academies Press. 351
Somanathan E, Somanathan R, Sudarshan A, et al. 2021. The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing. J Political Economy, 129: 1797–1827
Tibaldi S, Molteni F. 1990. On the operational predictability of blocking. Tellus A-Dynamic Meteorol Oceanogr, 42: 343
Vitart F, Ardilouze C, Bonet A, et al. 2017. The subseasonal to seasonal (S2S) prediction project database. Bull Amer Meteorol Soc, 98: 163–173
Vitart F, Robertson A W. 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. NPJ Clim Atmos Sci, 1: 3
Vitart F, Robertson A W, Anderson D L. 2012. Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. B World Meteorol Organ, 61: 23
Wilks D S. 2019. Statistical Methods in the Atmospheric Sciences. Amsterdam: Elsevier. 818
Wulff C O, Domeisen D I V. 2019. Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers. Geophys Res Lett, 46: 11520–11529
Xiao H, Xu P, Wang L. 2024. The Unprecedented 2023 North China heatwaves and their S2S predictability. Geophys Res Lett, 51: e2023GL107642
Xie J, Yu J, Chen H, et al. 2020. Sources of subseasonal prediction skill for heatwaves over the Yangtze River Basin revealed from three S2S models. Adv Atmos Sci, 37: 1435–1450
Yang J, Zhu T, Gao M, et al. 2018. Late-July barrier for subseasonal forecast of summer daily maximum temperature over Yangtze River Basin. Geophys Res Lett, 45: 12610–12615
Zhang T, Deng Y, Chen J, et al. 2023. An energetics tale of the 2022 mega-heatwave over central-eastern China. NPJ Clim Atmos Sci, 6: 162
Zscheischler J, Westra S, van den Hurk B J J M, et al. 2018. Future climate risk from compound events. Nat Clim Change, 8: 469–477

（责任编委：罗德海）