Article

# Pushing the boundary of seasonal prediction with the lever of varying annual cycles

Jieru Ma [a,b,c], Jie Sun [b], Zhaohua Wu [b], Jianping Huang [a], Xiaofeng Xu [d], Yi Deng [e], Ming Cai [b,*]

[a] Collaborative Innovation Center for West Ecological Safety, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China
[b] Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee FL 32306, USA
[c] State Key Laboratory of Severe Weather and Institute of Tibetan Plateau Meteorology, Chinese Academy of Meteorological Sciences, Beijing 100081, China
[d] Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081, China
[e] School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta GA 30332, USA

A B S T R A C T

Predicting climate anomalies months in advance is of tremendous socioeconomic value. Facing both theoretical and practical constraints, this realm of "seasonal prediction" progressed slowly in recent decades. Here we devise an innovative scheme that pushes the boundary of seasonal prediction by recognizing and isolating distinct spatiotemporal footprints left by modes of climate variability that cause varying annual cycles in response to the solar forcing. The predictive power harnessed from these spatiotemporal footprints results in a prediction skill surpassing existing models for seasonal forecasts of eastern China rainfall, which is one of the most challenging seasonal prediction problems. By considering varying annual cycles explicitly, the new scheme is able to predict multi-provincial flood and/or drought occurrences seamlessly over an entire year. This novel scheme is generically applicable for improving seasonal forecasts over other monsoon regions and for critical climate variables such as surface temperature and Arctic sea-ice extent.

© 2022 Science China Press. Published by Elsevier B.V. and Science China Press. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Skillful predictions of precipitation variability months ahead bear fundamental importance in social and agricultural planning. Since 1958, the National Climate Center of the China Meteorological Administration has been holding an annual meeting in early spring, called "Huishang (会商)", to obtain consensus forecasts of precipitation anomalies of the following seasons over different regions of China [1,2]. Decision makers then use the forecasts to guide agricultural and industrial planning and to develop plans for flooding disaster mitigation. Due to the unique location of China and its orographic complexity, the precipitation in China is influenced by many major climate systems, such as the East Asian monsoon system, the South Asian monsoon system, the migration of the Western Pacific Subtropical High, and complicated mid-latitude circulation patterns and weather systems [3–9]. The most prominent features of precipitation in eastern China are the abrupt advancing and retreating of seasonal rain bands from South China to North China in the spatiotemporal domain and inhomogeneity of precipitation intensity [3,10]. However, this north-south

advancing and retreating is not regular year to year [11], making the seasonal predictions of regional precipitation in China an extremely challenging task [12,13]. Any improvement in seasonal forecasts of precipitation can bring significant socioeconomic benefits to many millions of people.

Despite their high skill in predicting Pacific SSTs (sea surface temperatures) for a lead time of up to 14 months [14,15], state-of-art coupled GCMs (general circulation models) still do not possess a useful skill in predicting precipitation for a lead time longer than 1 month [12,16,17]. As a result, seasonal prediction of precipitation anomalies relies heavily on statistical models. The recent advancement in ensemble forecasting has made it possible for the first time that the June mean rainfall in the middle and lower Yangtze River basin can be predicted by a dynamical seasonal forecasting system at a lead time of up to 4 months with significant skill [18]. Many modes of climate variability, such as El Niño-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and Indian Ocean Dipoles (IOD), are found to impact the precipitation variability in eastern China [6,7,19–21]. The indices representing these climate variability modes have been used to develop various statistical models for seasonal forecasts of seasonal/monthly means of precipitation anomalies over some key regions

* Corresponding author.
E-mail address: mcai@fsu.edu (M. Cai).

(e.g., Yangtze River valley and central-North China) during the monsoon rainy season [22–26].

Conventionally, predictands of statistical models are climate anomalies, which are defined as departures from the climatological mean annual cycle, which we refer to as the traditional annual cycle (TAC). TAC is the portion of a variable that exactly repeats itself at any temporal location of a year but is independent of the year (https://glossary.ametsoc.org/wiki/Annual_cycle). Because TAC does not bear additional predictable information besides repeating itself annually, seasonal forecasts for climate variables become tasks of predicting anomaly fields. This way of partition, to some extent, brings convenience at the first glance. However, this partition can work effectively only when the response of the climate system to yearly periodic solar forcing is linear and dominated by repetitive TAC with corresponding anomaly fields minor. From a physical perspective, a large portion of such defined anomaly fields of a specific climate variable (e.g., precipitation) is originated from modes of climate variability that interfere with its annual cycle in response to the solar forcing [22–26], and the remaining is due to internal linear/nonlinear interactions on different timescales, ranging from the synoptic, sub-seasonal to interannual scales. Thus, the traditionally defined anomaly fields may contain a significant component originating from spatiotemporal footprints left by modes of climate variability that act to modulate individual years' annual cycle in response to the solar forcing. The above argument implies that the traditional partition approach may potentially distort the interpretations of the physical origins of the anomalies, as elucidated in the work of Wu et al. [27].

The objective of this study is to explore a new avenue for improvement in seasonal forecasts by focusing on the part that comes from the annual response to solar forcing modulated by modes of climate variability, which we refer to as yearly varying annual cycles (VAC), consisting of yearly repeatable and non-repeatable spatiotemporal patterns. The repetitive pattern corresponds to TAC, but its amplitude varies yearly. Besides yearly varying amplitude, the non-repetitive patterns are drastically different from that of TAC. The difference between VAC and TAC is regarded as the interannual variability of varying annual cycles (IVAC). Physically, the IVAC of a climate variable corresponds to spatiotemporal footprints left by modes of climate variability that interfere with its annual cycle response to the solar forcing. In other words, IVAC anomalies express themselves through mechanisms that are shared with those that determine the annual cycle. Thus, IVAC is expected to be more predictable than the remaining part of the traditionally defined anomalies, because it is intrinsically related to individual years' annual cycle due to modulations of the annual cycle by modes of climate variability. Therefore, the skill of seasonal forecasts may be significantly improved by focusing on IVAC, provided that IVAC accounts for a significant portion of traditionally defined anomaly fields. The remaining part of the paper will attest to this expectation by demonstrating the utility of this VAC concept for improving the seasonal forecasts of precipitation in eastern China.

## 2. Diagnostics and predictions based on varying annual cycles

### 2.1. The varying annual cycle perspective

In the view of the traditionally defined annual cycle, a daily climate variable $P(s,t)$ after removing its long-time climatological mean, where $s$ denotes space domain and $t$ is time, is decomposed as

$$P(s,t) = P_{\mathrm{TAC}}(s, day) + P'(s,t), \tag{1}$$

where $P_{\mathrm{TAC}}$ and $P'$ represent, respectively, TAC and traditionally defined anomaly fields of $P(s,t)$ with respect to $P_{\mathrm{TAC}}$, and "day" is the (Julian) day of a year. VAC can be described in a domain that includes both spatial and Julian-day variations jointly, referred to as the "space-day" domain. It follows that $P(s,t)$ is decomposed alternatively as

$$P(s,t) = P_{\mathrm{VAC}}(s, day, year) + P^*(s,t), \tag{2}$$

where $P_{\mathrm{VAC}}$ and $P^*$ represent, respectively, VAC and anomaly (with respect to $P_{\mathrm{VAC}}$) fields, and "year" is the year. The relationships between $P_{\mathrm{TAC}}$ and $P_{\mathrm{VAC}}$, and between $P'$ and $P^*$, are

$$
\begin{aligned}
P_{\mathrm{VAC}}(s, day, year) &= P_{\mathrm{TAC}}(s, day) + P_{\mathrm{IVAC}}(s, day, year), \\
P'(s,t) &= P_{\mathrm{IVAC}}(s, day, year) + P^*(s,t),
\end{aligned}
\tag{3}
$$

where $P_{\mathrm{IVAC}}$ corresponds to the interannual variability of the annual cycle of $P(s,t)$.

As illustrated in Fig. 1, TAC is the annual pattern in the space-day domain that repeats itself exactly at any temporal location year after year ($P_{\mathrm{TAC}}$, Fig. 1b) whereas VAC consists of repeatable and non-repeatable annual patterns. As a result, the anomaly field with respect to TAC ($P'$, Fig. 1c) could be dominated by annual signals that reflect spatiotemporal footprints left by modes of climate variability that interfere with the annual cycle response to the solar forcing. VAC, on the other hand, would capture the spatially and temporally mutual and continuous evolution of a climate variable in a given year ($P_{\mathrm{VAC}}$, Fig. 1d) in response to the annually evolving solar forcing and modes of climate variability. The advantage of using the space-day domain is to naturally represent the repeatable "space-day" evolution that is directly related to the annual cycle and non-repeatable "space-day" variation resulting from the interference of climate variability modes with the annual cycle response to the solar forcing. The departure field from VAC ($P^*$, Fig. 1e) represents a collection of spatially and temporally "independent" events. It is expected that $P_{\mathrm{IVAC}}$ is more predictable than $P^*$ because it is intimately related to the annual cycle response to the solar forcing in the presence of modes of climate variability. The inclusion of $P^*$ in the traditionally defined anomaly ($P'$) would "contaminate" the varying annual cycle response to the solar forcing, resulting in less skillful forecasts.

### 2.2. Annual patterns of precipitation over eastern China

The climate variable considered in this study is $P(s,t)$, the daily precipitation field after removing its long-time climatological mean over eastern China (Fig. S1 online) zonally averaged from 105° to 123°E covering the period of 1980–2020 (see data section of Analysis procedures for details). Yearly maps of $P(s,t)$ over eastern China in the space-day domain that include $P_{\mathrm{TAC}}(s,day)$ are given in Fig. 2 whereas Fig. S2 (online) corresponds to their counterparts of traditionally defined precipitation anomalies $P'(s,t)$ written in the format of, $P'(s,day,year)$. The two most salient features of yearly evolving space-day maps that are well documented in the Refs. [3,10] are (1) more rain from late spring to summer (wet season) than the rest of the year (dry season), (2) the northward migration of seasonal transition of heavy rainfall centers has two sudden jumps: one around late May at about 25°N and the other around July at about 32°N. Fig. 2 also shows that there exist large year-to-year variations of space-day patterns of $P$. Examples include (1) year-to-year variation of the wet season starting dates (e.g., 1983 for an early start and 2011 for a late start), (2) more/less rainfalls in both south and north with (almost) the same starting and ending dates of the wet season as the climatological wet season (1994/2004), and (3) more/less rainfalls in the south of 32°N but less/more rainfalls in the north during the wet season (e.g., 1983/2003).
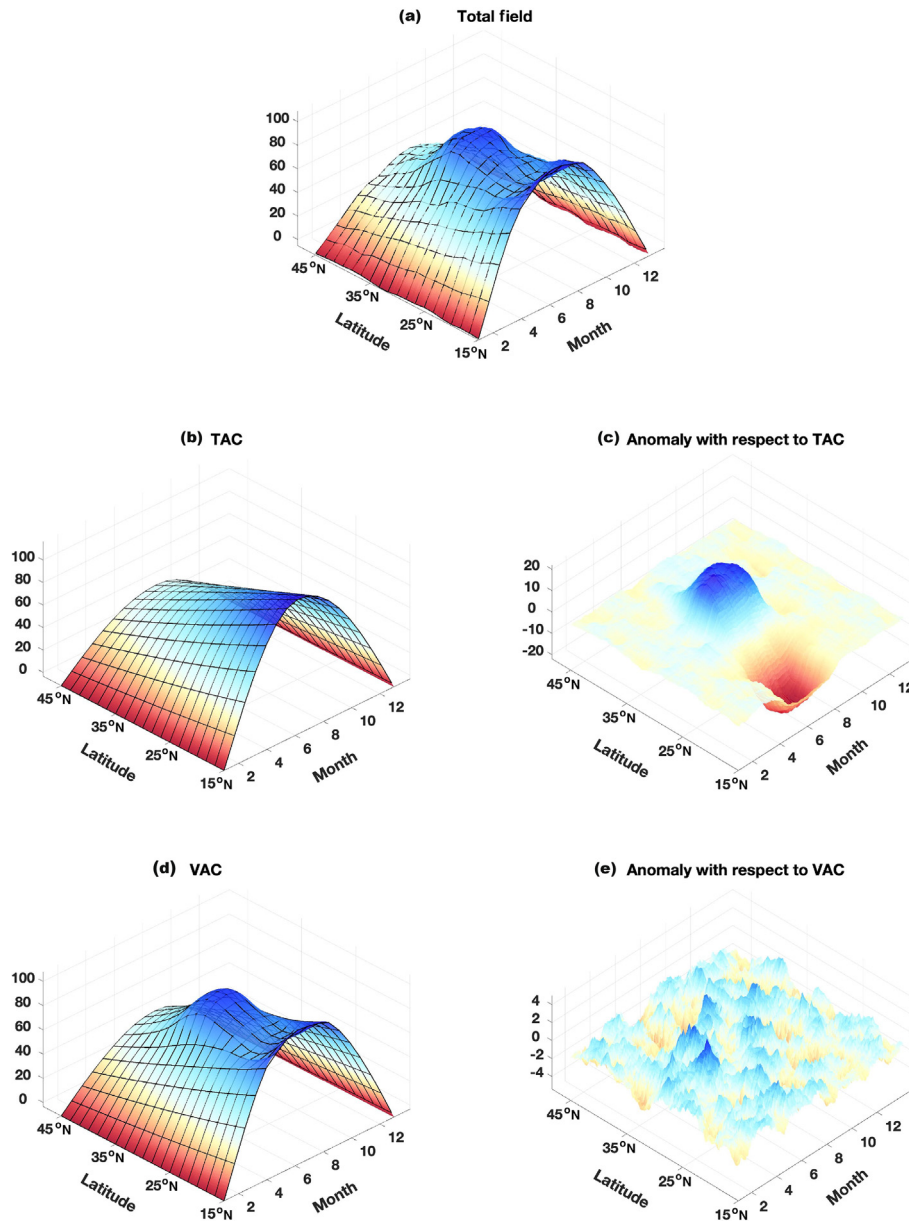
**Fig. 1.** Schematic diagram illustrating the conceptual difference between the traditional annual cycle (TAC) and varying annual cycle (VAC) perspectives. A meteorological field (say precipitation over eastern China) in a given year in the space-day domain (a) can be decomposed into TAC (b) and the anomaly with respect to TAC (c). Alternatively, it can be decomposed into VAC (d) and the anomaly with respect to VAC (e). The meteorological field shown here is generated synthetically for the purpose of illustration with an arbitrary scale, and is intended to be reminiscent of a flooding year in North China, such as 1990.

In this study, the traditionally defined annual cycle, $P_{TAC}(s, day)$, is obtained using data from 1980–2009, and the remaining 11 years (2010–2020) are used to evaluate the forecast skill of the statistical model developed using data from 1980 to 2009. To extract $P_{IVAC}$, we apply an empirical orthogonal function (EOF) method outlined in Box A of Fig. S3 (online) using the first 30 space-day maps shown in Fig. 2. Displayed in Fig. 3 are the first six EOF modes in the space-day domain (left column) and their principal components (PCs, right column). For each EOF mode, we obtain the space-day power spectra defined as the power spectra of the annual time series averaged over all latitudes. We regard those EOF patterns whose space-day power spectra of signals longer than the semi-annual scale account for at least 75% of their total variances as part of VAC (Fig. S4a online). It is seen from Fig. S4b (online) that EOF7 and higher have noticeably smaller contributions than those from EOF1-EOF6 to the total variance of $P'$. As a result, we only consider

the first 6 EOFs as the independent annual patterns (APs) of VAC, which are denoted as AP1, AP2, ···, AP6. Because these six APs account for nearly 94% of the total $P$ variance, $P_{VAC}$ reconstructed from these six APs (Fig. S5 online) is highly correlated with the total $P$ shown in Fig. 2 (Fig. S6 online). As expected, AP1 corresponds to $P_{TAC}$ (not shown here), as the correlation between them is very close to 1 (0.998) and the climatological mean of PC1 (175 mm/d, the right ordinate of Fig. 3b online) is very close to the amplitude of $P_{TAC}$ (about 175.5 mm/d). The departure of PC1 from the amplitude of $P_{TAC}$ corresponds to the part of $P'$ due to the year-to-year amplitude variation of $P_{TAC}$. Contributions to the total variance of $P'$ from the six APs are, respectively, 5.2%, 19.3%, 11.8%, 10.5%, 8.8%, and 7.7%.

The sum of the products of the six APs and their PCs (after removing their climatological means) is $P_{IVAC}$ (Fig. S7 online), which is constructed according to
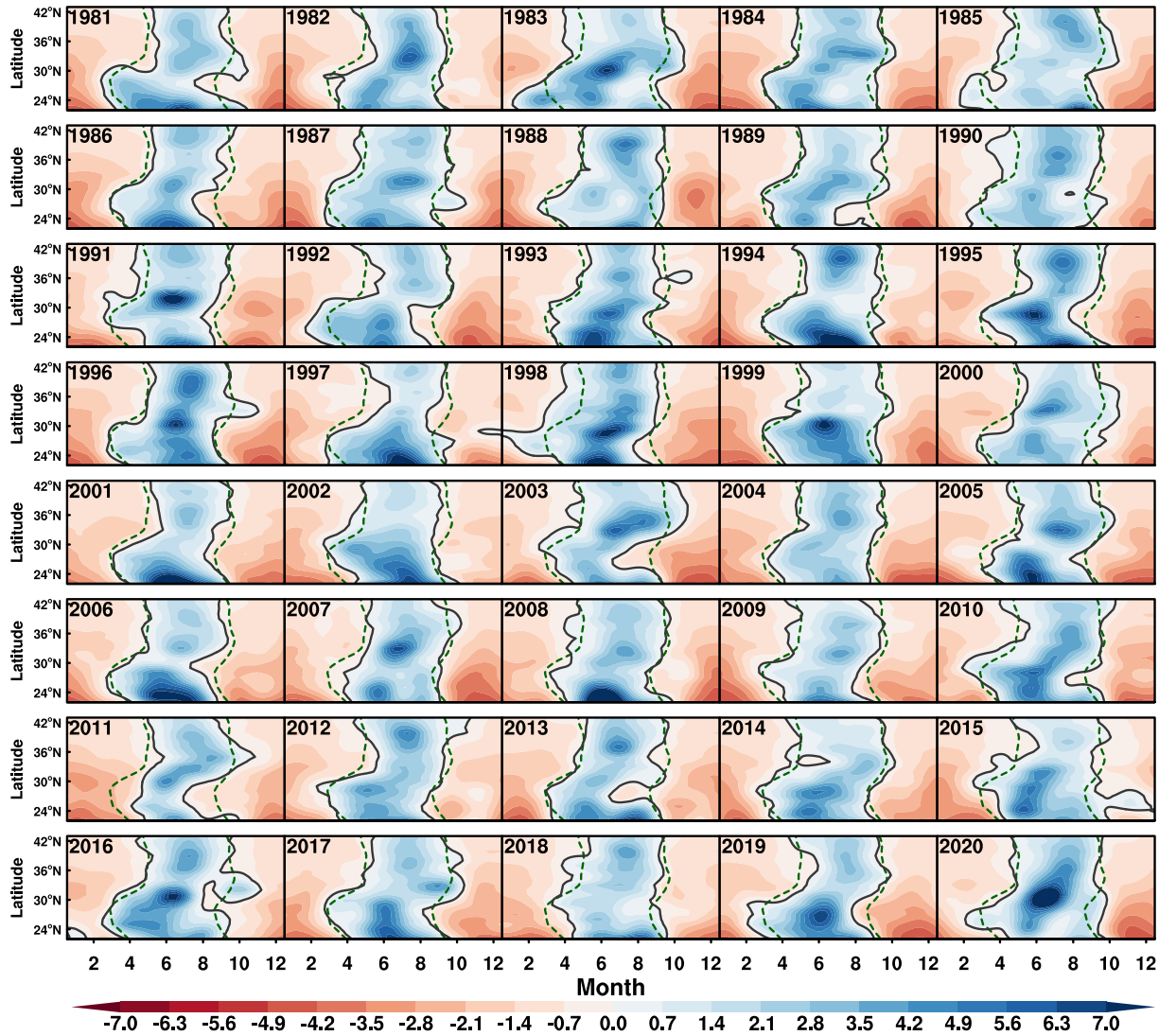
**Fig. 2.** Yearly maps of precipitation fields ($P$) over eastern China in the space-day domain. The ordinate is latitude and abscissa is month of a year with the year information indicated at the top left corner of each panel. Precipitation is in units of mm/d. The dashed green lines are the zero contour lines of the climatological (1980–2009) mean of the traditionally defined annual cycle ($P_{\mathrm{TAC}}$) of $P$, and the solid gray lines represent the zero contour lines of $P$. Note that the 1980 precipitation field is not included here for the sake of easy typesetting.

$$P_{\mathrm{IVAC}}(s, day, year) = \sum_{j=1}^{6} AP_j(s, day) PC_j^*(year), \qquad (4)$$

where

$$PC_j^*(year) = PC_j(year) - \overline{PC_j(year)}, \qquad (5)$$

where the overbar denotes the 30-year (1980–2009) mean values. Because $P_{\mathrm{IVAC}}$ accounts for nearly 63% of the total variance of $P'$, $P_{\mathrm{IVAC}}$ is the most important source of the traditionally defined precipitation anomalies. Fig. S8 (online) shows how $P_{\mathrm{IVAC}}$ contributes to $P'$ on a year-to-year basis, both in terms of all six APs collectively and each AP individually. There is a large year-to-year variation in the contribution to the $P'$ from $P_{\mathrm{IVAC}}$, ranging from 5% in 2017 to 87% in 1983. The amplitude of individual AP also exhibits pronounced year-to-year variations. For example, AP2 alone accounted for nearly 62% of the total variance of $P'$ in 1994 but was not present at all in 1991.

### 2.3. Successive orthogonalization of external factors

To identify the dominant modes of climate variability as potential predictors for $P_{\mathrm{IVAC}}$, we consolidate a total of 30 external factors, 20 of which have been considered in Refs. [25,26,28–47] for predicting spring/summer precipitation anomalies in some specific regions over eastern China and the remaining 10 are derived from these 20 factors. Their definitions, data sources, and references are provided in Tables S1 and S2 (online). The yearly time series of these external factors are derived from their 6-month mean values starting from September in the previous year and ending in February of the current year, covering the period of 1979/1980–2019/2020. For easy reference, we simply use the "previous year" as the year index for the yearly time series of these external factors (e.g., 1979 for the year 1979/1980). We use a successive orthogonalization method sketched in Box B of Fig. S3 (online) to remove any redundant information among these external factors, which makes the selected factors independent with one another. The
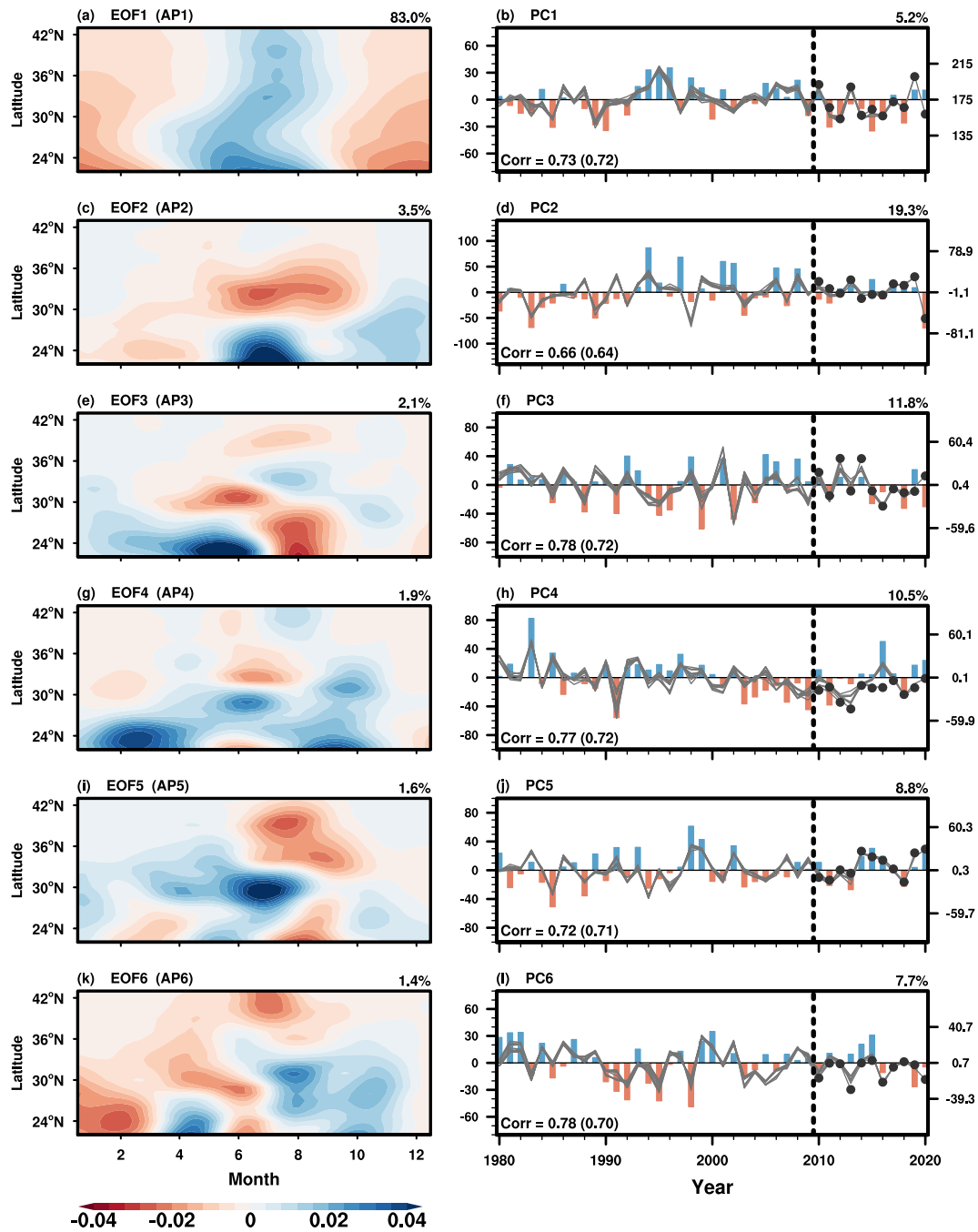
**Fig. 3.** Annual patterns (APs) of the precipitation fields (*P*) over eastern China. The panels on the left column are the 6 APs, corresponding to the first six empirical orthogonal functions (EOFs) derived from the first 30 years (i.e., *N* = 30) of *P* with ordinate being latitude and abscissa being month of a year. The values on the color bar are dimensionless and the values on the top right corner of each panel in the left column are the percentage variance of *P* (Fig. 1) explained by the individual APs. For the right columns, the bars plotted in the panels are yearly time series of the amplitudes of individual APs. The ordinate on the left is the anomalies (with their first 30-year mean values removed) whereas the ordinate on the right is the total values (i.e., including their first 30-year mean values), and both are in units of mm/d. The values on the top right corner of each panel are the percentage variance of traditionally defined precipitation anomaly (*P'*) explained by APs. The years on the left of the vertical dashed line correspond to the time period for the diagnosis and those on the right for predictions. The 11 gray curves correspond to the amplitudes of the APs obtained from the multi-regression models covering the periods of 1980–2010, 1980–2011, ⋯, and 1980–2020, respectively. Note that the 11 black dots in the end of the 11 gray curves represent forecasts for years of 2010, 2011, ⋯, 2020. The values on the bottom left corner of each panel are the correlation scores between the observed (bars) and predicted (gray curves) time series over the period of 1980–2009, and values inside the parentheses are the averaged correlation scores for the periods of 1980–2010, 1980–2011, ⋯, and 1980–2020.

abscissa of Fig. S9 (online) lists the 14 factors resulting from the successive orthogonalization based on the data from 1979–2008. The differences between the original yearly time series (vertical bars) and the orthogonalized counterparts (colored curves) of these 14 factors in Fig. S10 (online) correspond to the redundancies removed by the successive orthogonalization procedure.

### 2.4. Relating annual patterns to external factors

We next build six multi-regression models to relate amplitudes of the six APs to the 14 independent factors using factor and rainfall data in the 30-year period of 1979/1980–2008/2009 (referred to as the "diagnostic" period), following the procedures outlined in Box C of Fig. S3 (online) and Eqs. (9) and (10). It is seen from

Fig. S11 (online) that the order of the factors in terms of explained variances changes from one AP to another. It is of importance to note that the explained variance of each AP decreases rapidly as the number of the external factors such that the total variance explained by the 14 factors becomes saturated beyond 8–9 factors. Therefore, it suffices to consider the first 8 factors when building multi-regression models for reconstructing the yearly time series of APs. Because different APs involve a different combination of these 14 factors, we collectively have considered all of the 14 factors for the six APs. Shown in the grey curves of Fig. 3 are the regressed yearly time series of APs obtained from the multi-regression models that only include the first 8 factors given by Figs. S11a–f. Their correlations with the original year-to-year time series of these APs over the diagnostic period range from 0.66 to 0.78, indicating the regression models are skillful in capturing year-to-year amplitude variations of the six APs.

These 14 factors collectively capture about 51% variance of $P_{\text{IVAC}}$ (Fig. S12a online). The key features associated with the external factors that have the first 5 largest contributions to $P_{\text{IVAC}}$ are

(i) Multivariate ENSO index (MEI) (Fig. S12b online). The years immediately after the positive phase of MEI tend to have a very early start of rainfall seasons over South China and the Yangtze River valley with enhanced Meiyu rainfall. The reverse can be said of the negative phase of MEI. Previous studies [7,32,33] also indicate that in ENSO decaying years, there tends to be enhanced precipitation extending northeastward from southern China in January and February to the Yangtze River valley in March to May. The enhanced precipitation over the Yangtze River valley continues through the months of June, July, and August but less rainfall over southern China.

(ii) Tibetan Plateau snow depth (TPSD) (Fig. S12c online). The years immediately after a thicker snow depth over Tibetan Plateau tend to have a stronger rainy season over South China and lengthened Meiyu season over the Yangtze River valley, but drier summer over South China and North China Plain. The summer flooding over the Yangtze River and Yellow River basins has been attributed to anomalous moisture transport to eastern China after the above-normal TP snow cover in cold seasons [44,45]. Fig. S12c (online) further suggests that summer flooding over the Yangtze River and Yellow River basins is accompanied by a rapid transition from enhanced rainfall in spring to drier summer over South China.

(iii) Spatial mean sea level pressure over Mascarene (MSLP) (Fig. S12d online). The MSLP index describes a semipermanent subtropical ridge in Southern Hemisphere. According to Ref. [34], MSLP impacts the Asian summer monsoon through its modulation on the Somali jet. Our results indicate that years immediately after positive MSLP anomalies tend to have a drier spring over entire eastern China followed by a wetter summer over North China and South China, but the drought condition over the Yangtze River valley persists throughout the entire year.

(iv) Quasi Biennial Oscillation index (QBO) (Fig. S12e online). The years immediately after the westerly QBO phase in the tropical stratosphere tend to have to a shorter rain season over South China, a drier Meiyu season over the Yangtze River valley, a wetter rain season over the North China plain. The study [40] suggests that the QBO could be one of the important factors predicting the spring rainfall variability over East Asia, as the westerly QBO phase is accompanied by weakened Hadley circulation resulting in the equatorward retreat of the WPSH (western Pacific subtropical high) and the southward shift of East Asian jet. Such changes in

circulation are responsible for the southward displacement of the midlatitude rainband, resulting in decreased (increased) spring rainfall over the Yangtze River valley (South China). It well supports our results of the spring rainfall pattern in Fig. S12e (online). Our results also indicate that the wet season in South China ends earlier and a drier Meiyu season occurs in westerly QBO. In addition, the decreased rainfall in summer and early fall over South China is consistent with the finding of fewer tropical cyclones approaching South China in the westerly QBO phase [41].

(v) Spatial mean SST over North Pacific Ocean (NPSST) (Fig. S12f online). The years immediately after warm SST over North Pacific tend to have a shorter but stronger rain season over South China, a drier Meiyu season over the Yangtze River valley, but more rains in the wet season over North China. The warmer sea surface temperature (SST) of the tropical Northwest Pacific has been used as a predictor of the autumn drought of South China and the Yangtze River valley [48].

(vi) Antarctic Oscillation index (AAO) (Fig. S12g online). The years immediately after the positive phase of Antarctic oscillation tend to enhance summer rainfall over South China, but dry years north of the Yangtze River valley. As reported in previous studies [28,29], strong AAO events in spring and summer weaken the southwest monsoon in spring near South China, increase summer precipitation in the middle and lower reaches of the Yangtze River valley, and reduce rainfall over central-North China.

### 2.5. Forecast skill of $P_{\text{IVAC}}$

We next want to validate the six regression models in the context of predicting the temporal evolution of year-to-year variations of amplitudes of the six APs in the period of 2010–2020, referred to as the "prediction period". To forecast the amplitudes of APs in a target year, $(2009 + k)$ where $k = 1, 2, \cdots, 11$, we need to renormalize and re-orthogonalize the 14 external factors in the same order as they are considered in the diagnostic period so that they remain to be independent in the period from 1979 to the year before the target year (see Box D of Fig. S3 online and Analysis procedures for details). It is seen from Fig. S10 (online) that the 11 color lines of every factor, each representing the re-orthogonalized $(30 + k)$-year time series of the factor under the consideration, are all nearly indistinguishable from the 30-year time series constructed via the successive orthogonalization in the diagnostic period. This helps ensure that the re-orthogonalization does not change the inherent information content of the external factors that regulate amplitude variations of APs. The predicted amplitudes of APs in all the target years in the prediction period (11 black dots in Fig. 3) are obtained from the six regression models using the values of the 14 independent factors before the target year (Fig. S10 online) and the 11 sets of updated $6 \times 8$ regression coefficients using the data from the period of 1980 to the year before the target year. The regression models with the newly updated regression coefficients allow us to forecast the yearly time series of amplitudes of APs over the period from 1980 to the target year. Such time series not only yield the forecast for the amplitude of APs at the target year but also provide a new opportunity for validating the regression models over the diagnostic period and against the forecasts made in the early years. As indicated in Fig. 3, the forecasted yearly time series of APs made at the target year resembles closely to its counterparts not only over the diagnostic period (i.e., 1980–2009), but also over the prediction period (2010–2020). This suggests that the statistical relations between amplitude variations of APs and external factors are robust. As a result, their average correlation skill in predicting the original year-to-year time series of these AP amplitudes, ranging

from 0.64 to 0.72, is very close to the skill assessed in the diagnostic period.

We next construct $P_{IVAC}$ from these individually regressed/forecasted AP amplitudes (Fig. S13 online). Fig. 4 indicates that the 30-year (41-year) mean of map correlations between Figs. S13 and S7 (online) is about 0.67 (0.61), which is nearly twice as high as the threshold of the "useful" skill (0.3) for climate predictions. Moreover, the correlation skill of the regressed/forecasted $P_{IVAC}$ against observed $P_{IVAC}$ is as high as 0.5 in 24 out of 30 years in the diagnostic period and 7 out of 11 years during the prediction period. Fig. S14 (online) shows that the correlation skill evaluated under the diagnostic mode is only slightly better than that evaluated under the prediction mode with the exception in 2016, in which the diagnostic skill is significantly higher than its prediction skill. This is suggestive that the high prediction skill seen in the prediction period (2010–2020) is not resultant of the over-fitting problem. Furthermore, the saturation of explained variance beyond 8–9 factors shown in Fig. S11 (online) is indicative that the potential issue of over-fitting is largely avoided by the removal of the redundant information among factors, as the inclusion of additional factors after re-orthogonalization helps little in increasing explained variance of a given VAC. It is worthwhile to point out that the correlation skill for predicting the full anomalies is equal to the product of the skill for the $P_{IVAC}$ field (Fig. 4) and the square root of the percentage variance (Fig. S8 online) explained by $P_{IVAC}$. On average, $P_{IVAC}$ explains about 63% of the full anomalies and the average correlation skill for $P_{IVAC}$ is 0.61. Therefore, the average correlation skill for predicting the full anomalies is equal to 0.48 (=0.61 × $\sqrt{0.63}$).

To further substantiate the cross validation result shown in Fig. 4, we have performed a series of 10,000 × 11 Monte-Carlo forecast experiments using these 14 factors as the predictors for $P_{IVAC}$. Specifically, we randomly select 30 years of data from 1980 to 2020 as the diagnostic period for obtaining regression coefficients of $P_{IVAC}$ against the 14 factors and predict $P_{IVAC}$ in the other 11 years. We repeat the random selections of the diagnostic period 10,000 times, which yields a total of 10,000 × 11 forecast skill evaluations and a total of 10,000 × 30 diagnostic skill evaluations. The results of Monte-Carlo forecast experiments (Fig. S15 online) support the conclusion that the good diagnostic skill (90% of diagnostic skill exceeds 0.3 correlation score) of our model does translate to good forecast skill (about 60% forecasts have a skill higher than 0.3), which is well above the forecasts made using white noise (about 25%). The ratio of diagnostic skill to prediction skill obtained in Monte-Carlo forecast experiments (60/90 or 0.67) is similar to the ratio of the average prediction skill in 2010–2020 (0.46) to

the diagnostic skill in 1980–2009 (0.67). Therefore, the Monte-Carlo forecast experiments, besides serving as a statistical cross validation, further indirectly suggest that the high skill is not due to over-fitting. In addition, we have evaluated the skill of "persistent forecasts", which is often used as a reference forecast ("climatology" is another reference forecast) for skill evaluation. Clearly, our forecasts outperform greatly persistent forecasts.

## 3. Conclusion and discussion

By virtue of being part of the world's strongest monsoons, the temporal evolution of precipitation in eastern China is strongly influenced by the annually evolving solar forcing. On the premise that climate variability related to yearly varying annual cycles is more predictable than the internal climate variability, we devise an innovative way to isolate the precipitation variability in eastern China associated with annual patterns that are not repeatable yearly from the remaining of the traditionally defined precipitation anomaly field. Specifically, we regard the spatial domain and day-to-day evolution in a year as the domain for spatiotemporal patterns of a meteorological variable (such as precipitation) that change yearly. To achieve this, we separate time $t$ explicitly into (Julian) days of a year and years and join the space domain and (Julian) days of the year as the base, referred to as the "space-day" domain, for spatiotemporal patterns. As a result, a yearly meteorological field (e.g., precipitation) in the space-day domain naturally contains the information of the (yearly) repeatable "space-day" evolution and (yearly) non-repeatable "space-day" variations. The former corresponds to the climatological annual cycle that is directly related to the solar forcing and the latter are part of the anomaly fields resulting from the interference of climate variability modes with the annual cycle response to the solar forcing.

We identify one repeatable and five non-repeatable APs in the precipitation field over eastern China. The IVAC associated with these six APs accounts for nearly 63% of the total variance of precipitation anomalies in eastern China, representing year-to-year variations of starting date, duration, and intensity of monsoon precipitation in eastern China. Therefore, year-to-year amplitude variations of repeatable and non-repeatable APs about their climatological mean values contribute substantially to year-to-year variations of monsoon precipitation in eastern China.

We envision that these APs encapsulate distinct spatiotemporal footprints of a set of external factors that regulate Asian monsoon systems and collectively describe atmosphere-ocean-land conditions for eastern China's precipitation response to the annually
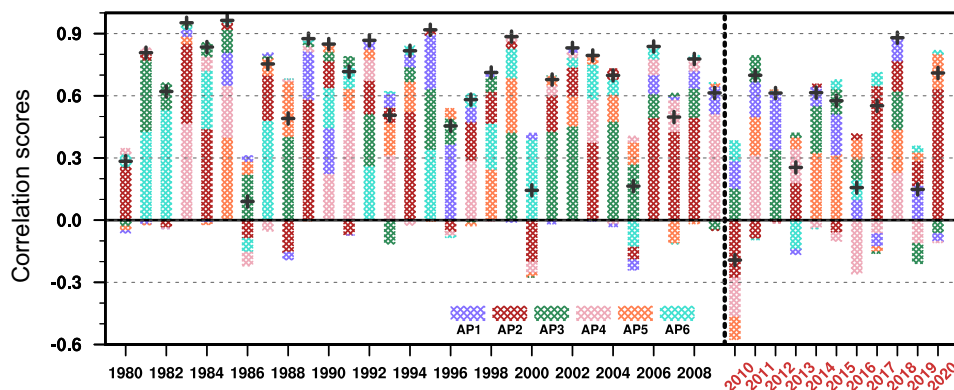


**Fig. 4.** Yearly time series of the correlation scores of the forecasts for the interannual variability of annual cycle ($P_{IVAC}$) against the observed $P_{IVAC}$. Colors are used to indicate the contributions to the correlation scores from individual APs and their sums equal to the values indicated by "+". The dashed vertical lines separate the diagnosis period (1980–2009) from the prediction period (2010–2020). The vertically stacked color bars are ordered in a bottom-up fashion following the decreasing absolute value of correlation coefficients.

evolving solar forcing in the following year. We apply a successive orthogonalization analysis to obtain 14 independent factors from a total of 30 external factors that have been used in the literature for predicting spring/summer precipitation anomalies in some specific regions over eastern China. Collectively, the prior year information of these 14 factors explains about 51% variance of the year-to-year variation of monsoon precipitation in eastern China associated with VAC. The six factors that have the largest contribution to VAC are (1) MEI (Multivariate ENSO index), (2) TPSD (Tibetan Plateau snow depth), (3) MSLP (spatial mean sea level pressure over Mascarene), (4) QBO (Quasi Biennial Oscillation index), (5) NPSST (spatial mean SST over North Pacific Ocean), and (6) AAO (Antarctic Oscillation index). These factors regulate the yearly response of precipitation in eastern China to the annually evolving solar forcing via advancing/delaying starting dates, lengthening/shortening of the monsoon rainfall season, and strengthening/weakening of its intensity in different latitude bands of eastern China.

It is of importance to note here that the ordering of the factors by the successive orthogonalization procedure adopted in this study is done without considering their ability to explain the APs explicitly, as these factors have been already proved in the literature to have useful skill in predicting spring/summer precipitation anomalies over eastern China. This choice allows us to treat these selected factors equally and independently in building the multi-regression models for predicting APs. We also note that an alternative choice is to construct the multi-regression models by combining regression and orthogonalization analyses jointly following the work of Ref. [49]. Because the two choices would select different factors from the 30 external factors as predictors, the resultant skill would be different (should all of the 30 factors are selected, the resultant skill would be identical, corresponding to the skill of the standard multiple regression analysis using all of the 30 factors).

To explore the prediction utility of the VAC perspective, we perform forecast experiments in the years of 2010–2020 using multivariate regressions models with the information of these 14 independent factors in the prior years for amplitudes of the six annual patterns identified from 1980–2009 (diagnostic period). The external factors each have distinct spatiotemporal footprints in eastern China's rainfall throughout the year as they represent atmosphere-ocean-land conditions in the prior year that regulate the next year's annual cycle response in rainfall to the annually evolving solar forcing. By independently forecasting the dominant spatiotemporal modes of rainfall in terms of the six APs, our approach reduces spatiotemporal noise contamination and enhances the utilization of the potential predictability offered by each external factor. It is shown that the correlation skill in predicting the original year-to-year time series of individual APs' amplitudes, ranging from 0.64 to 0.72, is very close to the skill assessed in the diagnostic period (1980–2009). The average correlation skill of the predicted $P_{\text{IVAC}}$ in the prediction period (2010–2020) is about 0.46, which is about a 30% reduction from the average correlation skill (0.67) evaluated in the diagnostic period. The statistical cross validation using Monte-Carlo forecast experiments confirms the small reduction (33%) from the diagnostic skill to the prediction skill. The 41-year average skill of 0.61 for the precipitation over an entire year in eastern China surpasses average forecast skills for just rainy seasons (e.g., spring and/or summer) demonstrated by existing dynamical (∼0.1), dynamic-statistics (∼0.38), and statistical (∼0.45) models [50–55]. Because the high correlation skill of our new approach is achieved for the continuously temporal evolution of spatial rainfall anomaly patterns over the entire year, instead of just over a certain period (e.g., the month of June, or July), such forecasts for rainfall anomalies over the entire year would help enable decision makers for more targeted mitigation planning on the flooding/drought throughout a year.

It should be pointed out that as with most statistical forecasts for monthly/seasonal mean anomalies, our forecasts tend to underestimate the amplitude of climate anomalies. We have confirmed that the correlation skill of our forecasts is negatively correlated with root mean square errors (RMSE). Therefore, forecasts with a large positive correlation score have a small RMSE. The mean value of RMSE of our forecasts is about 0.5 mm/d with the largest RMSE being about 0.9 mm/d, which is smaller than the mean RMSE of seasonal forecasts for precipitation made by the NOAA Climate Test Bed (https://www.cpc.ncep.noaa.gov/products/ctb/CTB/).

Fig. 5 highlights that the power of the VAC-based forecasting approach in terms of its capability of predicting occurrences of multi-provincial-scale hydrological extremes at least six months in advance. The top row of Fig. 5 corresponds to the year 1998, when summer flooding in the Yangtze River Valley killed more than 3000 people, destroyed over 15 million homes, and caused a total economic loss exceeding $36 Billion [56]. The hindcast initiated at the February of 1998, whose skill is about 0.71, unambiguously captures the March-July flood over most of eastern China and the subsequent dry months of the remaining year. The middle row is for the year 2011 when the most severe drought since 1961 occurred over eastern China leading to a total loss of nearly $48 Billion [57]. The correlation skill of the hindcast initiated at the February of 2011 is about 0.7, capturing the 2011 drought signature, including its overall spatiotemporal extent. The correlation skill of the hindcast initiated at the February of 2020 (bottom row) is about 0.71, capturing the massive flooding in 2020 in the Yangtze River valley and a dry monsoon season over North China Plain and South China.

This study demonstrates the utility of the VAC perspective for seasonal predictions using zonal mean anomalies, which can be naturally extended for spatial patterns of anomalies. In 4.7 of Analysis procedures, we outline a procedure that can be used to obtain the spatial patterns of anomalies directly from the VAC-based seasonal prediction of zonal mean anomalies. We have applied this procedure to obtain the spatial patterns of observed and forecasted $P_{\text{IVAC}}$ (Figs. S16–S18 online) from the results shown in Fig. 5. The results confirm that (1) the spatial patterns of the observed $P_{\text{IVAC}}$ also greatly capture the spatial patterns of the full rainfall anomalies in each of the three highlighted multi-provincial-scale hydrological extreme years over eastern China and (2) the good prediction skill for observed $P_{\text{IVAC}}$ can be transferred to the prediction skill for their spatial patterns, as summarized in Fig. S19 (online).

We here wish to add that the use of the September-February average adopted in this work for defining the predictors is mainly to demonstrate the utility of the VAC perspective for seasonal predictions of eastern China rainfall to be considered in the annual "Huishang (会商)" conducted in early spring of each year (i.e., March). The use of the information in January and February would allow us to retain the most updated information about the phase and amplitude of these climate modes for March's "Huishang". One could choose the July-December average for defining the predictors for "Huishang" in January, or the August-January average for "Huishang" in February, and so on.

In conclusion, climate variability manifested in VACs is a major source of predictability across interannual timescales, because these VACs express themselves through mechanisms that are shared with those that determine the annual cycle response to the solar forcing. The newly identified source of predictability helps push the boundary of statistical seasonal predictions to a new paradigm that includes the continuously spatiotemporal evolution of climate anomalies seamlessly over an entire year with a longer lead time. The two immediate benefits of the VAC perspective are that (1) it helps advance our understanding of modulations of annual cycle response to the solar forcing by modes of climate
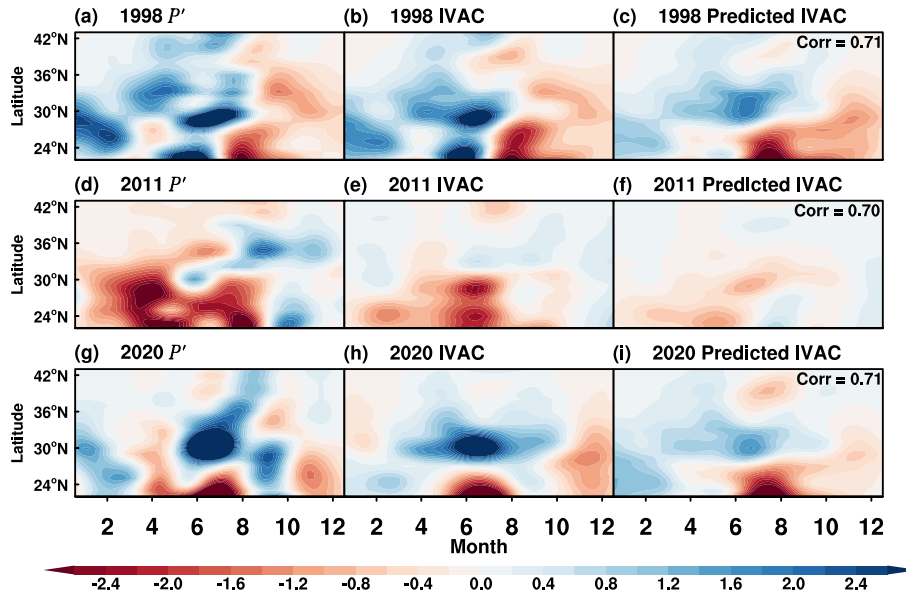
**Fig. 5.** Forecasts of 1998, 2011, and 2020 hydrological extremes over eastern China. (a)–(c), respectively, are the yearly maps of traditionally defined precipitation anomalies ($P'$), interannual variability of annual cycle ($P_{IVAC}$), and forecasted $P_{IVAC}$ in 1998 (a flooding year). (d)–(f) are same as (a)–(c), but for 2011 (a drought year). (g)–(i) are same as (a)–(c), but for 2020 (a flooding/drought year for Yangtze River valley/southern China). The ordinate is latitude and abscissa is month. The values on the top right corner of panels (c), (f), and (i) are the correlation scores between the observed and forecasted $P_{IVAC}$ in 1998, 2011, and 2020, respectively.

variability; and (2) it helps develop more skillful statistical models for seasonal forecasts where the dynamical models currently lack skill. The varying annual cycle perspective can be similarly adopted for the improvement of seasonal forecasts in other regions that are significantly impacted by major monsoon systems (e.g., India, Africa, and Australia) and for other climate variables besides precipitation, such as surface temperatures over lands and Arctic sea-ice extent.

# 4. Analysis procedures

## 4.1. Data

We use version 3 of the daily precipitation dataset compiled by the National Meteorological Information Center of the China Meteorological Administration, covering the period of 1951–2020 [58]. The daily precipitation dataset is derived from surface hourly rain gauge data over 2420 stations in China, most of which are located in eastern China [59]. Before the analysis, we first gridded the station precipitation data over the domain of (21° to 43°N, 105° to 123°E) at the resolution of (1° × 1° using an interpolation algorithm of Cessman style analysis. We then removed their 1980–2009 means, obtained their zonal mean values, and performed a 4-month running mean. The resultant daily fields are denoted as $P(s,t)$ covering the period of 1980–2020, where $s$ is latitude and $t$ is time. We note the main characteristics of the annual patterns are not sensitive to the temporal scale of the smoothing scheme (e.g., 1-month versus 4-month running mean) other than a shorter time running mean would retain a substantial faction of day-to-day variability at scales much shorter than semi-annual scale. The 4-month running mean is used for robustly capturing the signals at annual and semi-annual scales for the annual patterns. Supplementary Table S1 (online) lists the information of all external factors considered in this study, including their abbreviations, definitions, and references whereas Supplementary Table S2 (online) provides the information of data sources for deriving the external factors. The yearly time series of these external factors are derived from their 6-month mean values starting at September

of the previous year and ending at February of the current year, covering the period of 1979/1980–2019/2020.

## 4.2. Extraction of annual patterns of P

Following Eq. (2) and Box A of Fig. S3 (online), we first write $P(s, t)$ into $P(s,day,year)$. The annual patterns of $P$ are obtained from EOF analysis of $P(s,day,year)$ over the 30-year period of 1980–2009 by joining "*day*" with latitude "*s*" as the "space-day" domain for EOF patterns and "*year*" as the domain for their associated PCs, namely

$$P(s, day, year) = \sum_{j=1}^{30} EOF_j(s, day)PC_j(year),$$
$$\text{where} < \left(EOF_j(s, day)\right)^2 > \equiv 1, \tag{6}$$

where "< >" represents the average over the domain of $s$ and $day$ and $PC_j$ has the units of mm/d. As stated in the main text, the first six EOFs are referred to as annual patterns. The yearly time series of APs can be obtained from

$$PC_j(year) = < P(s, day, year)AP_j(s, day) >, \tag{7}$$

where $AP_j(s, day) = EOF_j(s, day)$, $j = 1, 2, 3, 4, 5,$ and $6$. The yearly time series of the fractional variance ($R_{AP}$) of traditionally defined $P'$ explained by $P_{IVAC}$ (Fig. S8 online) is obtained from

$$R_{AP}(year) = \frac{\sum_{j=1}^{6} \left(PC_j^*(year)\right)^2}{< \left[P'(s, day, year)\right]^2 >}, \tag{8}$$

where $PC_j^*$ is defined in Eq. (5).

## 4.3. Ordering and orthogonalization of external factors

As sketched in Box B of Fig. S3 (online), we first normalize each of the 30 external factors listed in Table S1 (online) so that their 1979–2008 means is zero and their variance is equal to one. For easy reference, we denote the original yearly time series of external factors as $\{F_m(year)\}$. To remove the redundant information among these 30 external factors, we follow the iterative proce-

dures outlined below to order and successively orthogonalize these external factors starting with $m = 1$:

(1) Construct a covariance matrix among these $(31-m)$ factors and find the external factor that explains the most amount of the variance at all other factors (including itself). This factor will be selected as the $m$th ordered factor.
(2) Re-construct the time series of the remaining $(30-m)$ factors by removing the portion that is explained by the $m$th ordered factor.
(3) Repeat (1) and (2) with the reconstructed time series with the value of $m$ increased by 1 till the variance of the remaining (re-constructed) factors all is less than 25%.

The completion of the iterative procedures (1)–(3) yields a set of $M^*$ ($M^* \leq 30$) independent yearly time series in the period of 1979–2008 with their variance no less than 25% of their original yearly time series. These independent factors will be used to construct multi-regression models for predicting AP amplitudes. According to Fig. S9 (online), the first 14 external factors (or $M^*$=14) can be regarded as the independent predictors for AP amplitudes as the remaining variance of each of the other 16 factors is no more than 25% after removing the part that can be explained by these 14 factors (see the last column of Fig. S9 online). For easy reference, we denote these independent predictors as $\widetilde{F}_m$.

### 4.4. Multi-regression model

As sketched in Box C of Fig. S3 (online), the multi-regression model for predicting interannual variability of a specific $AP_j$ is,

$$\widehat{PC}_j^*(year) = \sum_m \alpha_{m,j} \widetilde{F}_m(year - 1), \tag{9}$$

where

$$\alpha_{m,j} = \frac{\overline{PC_j^*(year)\widetilde{F}_m(year - 1)}}{\overline{\left(\widetilde{F}_m(year - 1)\right)^2}}. \tag{10}$$

In Eq. (9), the order of $m$ is arranged according to values of $(\alpha_{m,j})^2 \overline{\left(\widetilde{F}_m(year - 1)\right)^2} / \overline{\left(PC_j^*(year)\right)^2}$ from the largest to the smallest, as indicated in Fig. S11 (online). The percentage of $P_{IVAC}$ variance explained by external factors ($EV_m$) shown in Fig. S12a (online) is obtained from

$$EV_m = \frac{\sum_j^6 (\alpha_{m,j})^2 \overline{\left(\widetilde{F}_m(year - 1)\right)^2}}{\sum_{j=1}^6 \overline{\left(PC_j^*(year)\right)^2}}. \tag{11}$$

The space-day patterns of $P_{IVAC}$ associated with external factors ($PF_m(s, day)$, Fig. S12b–g (online) are obtained by

$$PF_m(s, day) = \sqrt{\overline{\left(\widetilde{F}_m(year - 1)\right)^2}} \sum_{j=1}^6 \alpha_{m,j} AP_j(s, day). \tag{12}$$

The yearly $P_{IVAC}$ patterns regressed against the external factors (Fig. S13 online) are obtained from Eq. (4) by replacing $PC_j^*(year)$ with $\widehat{PC}_j^*(Y_k)$.

### 4.5. Prediction of $P_{IVAC}$

We are now ready to predict $AP$ amplitudes at the year $Y_k$, where $Y_k$ denotes the year of $(2009 + k)$, $k = 1, 2, \cdots$ (Box D of Fig. S3 online). To do this, we first re-normalize and re-

orthogonalize the 14 independent predictors in the period from 1979 to the year $(Y_k-1)$ without changing their order during the re-orthogonalization for obtaining $\widetilde{F}_m(Y_k - 1)$. We next apply the multi-regression model for predicting $\widehat{PC}_j^*(Y_k)$ with the understanding of (1) that the overbar in Eq. (10) is applied to the period from 1980 to the year $(Y_k - 1)$ and (2) that the order of $m$ in (9) is the same as its original order for each $j$. The forecast for $P_{IVAC}$ at the year $Y_k$ is obtained from Eq. (4) by replacing $PC_j^*(year)$ with $\widehat{PC}_j^*(Y_k)$.

### 4.6. Prediction skill evaluation of $P_{IVAC}$

The yearly time series of the correlation scores of the forecasts for $P_{IVAC}$ (Fig. 4) is obtained from

$$S_1(year) = \frac{\sum_{j=1}^6 \widehat{PC}_j^*(year)PC_j^*(year)}{\sqrt{\sum_{j=1}^6 \left(\widehat{PC}_j^*(year)\right)^2} \sqrt{\sum_{j=1}^6 \left(PC_j^*(year)\right)^2}}. \tag{13}$$

### 4.7. Spatial patterns of observed and predicted $P_{IVAC}$

The spatial patterns associated with each AP can be obtained by regressing the yearly varying daily full (i.e., traditionally defined) anomalies on each grid point (or each station) against the yearly time series (i.e., principal component) of its corresponding AP. The resultant spatial patterns associated with each AP consist a series of 365 continuously evolving daily maps. The daily spatial patterns of the observed $P_{IVAC}$ in a given year can be obtained by summing up (over all of the six APs) the products of spatial patterns associated with each AP and its principal component value in that year. Similarly, the daily spatial patterns of the predicted $P_{IVAC}$ in a given year are obtained by summing up the products of spatial patterns associated with each AP and the predicted value of its principal component in that year.

### Conflict of interest

The authors declare that they have no conflict of interest.

### Author contributions

Ming Cai and Zhaohua Wu conceived the original idea for the study. Jieru Ma downloaded the data. Jieru Ma and Jie Sun performed the analysis and made graphic presentations. All authors discussed the results throughout the study and contributed to the writing.

## Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2022.12.026.

## Data availability

Data used in this study can be downloaded from the Uniform Resource Locator (URL) listed in Table S2 (online). The data generated from our calculations are available on request from the authors.

## References

[1] Ren HL, Wu YL, Bao Q, et al. The China multi-model ensemble prediction system and its application to flood-season prediction in 2018. J Meteorol Res 2019;33:540–52.

[2] Wang HJ, Fan K, Sun J, et al. A review of seasonal climate prediction research in China. Adv Atmos Sci 2015;32:149–68.

[3] Yeh T-C, Tao S, Li M. The abrupt change of circulation over the Northern Hemisphere during June and October in the atmosphere and the sea in motion. New York: Rockefeller University Press; 1959. p. 249–67.

[4] Lau KM, Li MT. The monsoon of East Asia and its global association—A survey. Bull Georg Acad Sci Am Meteorol Soc 1984;65:114–25.

[5] Tao SY, Chen LX. Review of recent research on the East Asia summer monsoon in China in monsoon meteorology. New York: Oxford University Press; 1987. p. 60–92.

[6] Wu GX, Duan AM, Liu YM, et al. Tibetan Plateau climate dynamics: recent research progress and outlook. Natl Sci Rev 2015;2:100–16.

[7] Zhang RH, Sumi A, Kimoto MA. Diagnostic study of the impact of El Niño on the precipitation in China. Adv Atmos Sci 1999;16:229–41.

[8] Wang B, Wu RG, Li T. Atmosphere-warm Ocean interaction and its impacts on AsianAustralian monsoon variation. J Clim 2003;16:1195–211.

[9] Ding Y, Wang Z, Sun Y. Inter-decadal variation of the summer precipitation in East China and its association with decreasing Asian summer monsoon. Part I: observed evidences. Int J Climatol 2010;28:1139–61.

[10] Chiang JCH, Swenson LM, Kong W. Role of seasonal transitions and the westerlies in the interannual variability of the East Asian summer monsoon precipitation. Geophys Res Lett 2017;44:3788–95.

[11] Zhang H, Griffiths M, Chiang J, et al. East Asian hydroclimate modulated by the position of the westerlies during Termination I. Science 2018;362:580–3.

[12] Wang B, Ding Q, Fu X, et al. Fundamental challenge in simulation and prediction of summer monsoon rainfall. Geophys Res Lett 2005;32:L15711.

[13] Wu ZW, Wang B, Li JP, et al. An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO. J Geophys Res 2009;114: D18120.

[14] Chen D, Cane MA, Kaplan A, et al. Predictability of El Niño over the past 148 years. Nature 2004;428:733–6.

[15] Kanamitsu M, Kumar A, Juang HH, et al. NCEP dynamical seasonal forecast system 2000. Bull Amer Meteorol Soc 2002;83:1019–37.

[16] Saha S, Moorthi S, Wu XR, et al. The NCEP climate forecast system version 2. J Clim 2014;27:2185–208.

[17] Ding Y, Liang P, Liu YJ, et al. Multiscale variability of Meiyu and its prediction: a new review. J Geophys Res 2020;125. e2019JD031496.

[18] Martin GM, Dunstone NJ, Scaife AA, et al. Predicting June mean rainfall in the middle/lower Yangtze River Basin. Adv Atmos Sci 2020;37:22–41.

[19] Lau KM. The East Asian summer monsoon rainfall variability and climate teleconnection. J Meteorol Soc Jpn 1992;70:211–42.

[20] Chang CP, Zhang YS, Li T. Interannual and interdecadal variation of the East Asian summer monsoon rainfall and tropical SSTs. Part 1: roles of the subtropical ridge. J Clim 2000;13:4310–25.

[21] Ma ZG. The interdecadal trend and shift of dry/wet over the central part of North China and their relationship to the Pacific Decadal Oscillation (PDO). Chin Sci Bull 2007;52:2130–9.

[22] Ding YH, Shi XL, Liu YM, et al. Multi-year simulations and experimental seasonal predictions for rainy seasons in China by using a nested regional climate model (RegCM_NCC). Part I: sensitivity study. Adv Atmos Sci 2006;23:323–41.

[23] Wang HJ, Fan K. A new scheme for improving the seasonal prediction of summer precipitation anomalies. Weather Forecast 2009;24:548–54.

[24] Wu B, Zhou TJ, Li T. Seasonally evolving dominant interannual variability modes of East Asian climate. J Clim 2009;22:2992–3005.

[25] Wang B, Lee JY, Xiang BQ. Asian summer monsoon rainfall predictability: a predictable mode analysis. Clim Dyn 2015;44:61–74.

[26] Fan K, Wang HJ, Choi YJ. A physically-based statistical forecast model for the middle-lower reaches of the Yangtze River valleys summer rainfall. Chin Sci Bull 2008;53:602–9.

[27] Wu ZH, Schneider EK, Kirtman BP, et al. The modulated annual cycle: an alternative reference frame for climate anomalies. Clim Dyn 2008;31:823–41.

[28] Wang HJ, Fan K. Central-North China precipitation as reconstructed from the Qing dynasty: signal of the Antarctic Atmospheric Oscillation. Geophys Res Lett 2005;32:L24705.

[29] Nan S, Li JP. The relationship between the summer precipitation in the Yangtze River valley and the boreal spring Southern Hemisphere annular mode. Geophys Res Lett 2003;30:2266.

[30] He S, Gao Y, Li F, et al. Impact of Arctic Oscillation on the East Asian climate: a review. Earth-Sci Rev 2017;164:48–62.

[31] Guan ZY, Yamagata T. The unusual summer of 1994 in East Asia: IOD teleconnections. Geophys Res Lett 2003;30:1544.

[32] Wu RG, Hu ZZ, Kirtman BP. Evolution of ENSO-related rainfall anomalies in East Asia. J Clim 2003;16:3742–58.

[33] Wang B, Liu J, Kim HJ, et al. Northern Hemisphere summer monsoon intensified by mega-El Niño/southern oscillation and Atlantic multidecadal oscillation. Proc Natl Acad Sci USA 2013;110:5347–52.

[34] Xue F, Wang HJ, He JH. Interannual variability of Mascarene high and Australian high and their influences on summer rainfall over East Asia. Chin Sci Bull 2003;48:492–7.

[35] Wang B, Li J, Cane MA, et al. Toward predicting changes in the land monsoon rainfall a decade in advance. J Clim 2018;31:2699–714.

[36] Sung MK, Kwon WT, Baek HJ, et al. A possible impact of the North Atlantic Oscillation on the East Asian summer monsoon precipitation. Geophys Res Lett 2006;33:L21713.

[37] Huang J, Higuchi K, Shabbar A. The relationship between the North Atlantic Oscillation and El Niño-Southern Oscillation. Geophys Res Lett 1998;25:2707–10.

[38] Yim SY, Wang B, Xing W. Prediction of early summer rainfall over South China by a physical-empirical model. Clim Dyn 2014;43:1883–91.

[39] Zhu Y, Wang H, Ma J, et al. Contribution of the phase transition of Pacific Decadal Oscillation to the late 1990s' shift in East China summer rainfall. J Geophys Res 2015;120:8817–27.

[40] Seo J, Choi W, Youn D, et al. Relationship between the stratospheric quasi-biennial oscillation and the spring rainfall in the western North Pacific. Geophys Res Lett 2013;40:5949–53.

[41] Ho CH, Kim HS, Jeong JH, et al. Influence of stratospheric quasi-biennial oscillation on tropical cyclone tracks in the western North Pacific. Geophys Res Lett 2009;36:L06702.

[42] Nnamchi HC, Li JP. Influence of the South Atlantic Ocean dipole on West African summer precipitation. J Clim 2011;24:1184–97.

[43] Xu H, Xu M, Xie SP, et al. Deep atmospheric response to the spring Kuroshio over the East China Sea. J Clim 2011;24:4959–72.

[44] Wu RG, Kirtman BP. Observed relationship of spring and summer East Asian rainfall with winter and spring Eurasian snow. J Clim 2007;20:1285–304.

[45] Wang ZQ, Duan AM, Wu GX. Time-lagged impact of spring sensible heat over the Tibetan Plateau on the summer rainfall anomaly in East China: case studies using the WRF model. Clim Dyn 2014;42:2885–98.

[46] Zhu Y, Wang H, Zhou W, et al. Recent changes in the summer precipitation pattern in East China and the background circulation. Clim Dyn 2011;36:1463–73.

[47] Li XY, Lu RY. Breakdown of the summertime meridional teleconnection pattern over the western North Pacific and East Asia since the early 2000s. J Clim 2020;33:8487–505.

[48] Wang L, Chen W, Zhou W, et al. Teleconnected influence of tropical Northwest Pacific sea surface temperature on interannual variability of autumn precipitation in Southwest China. Clim Dyn 2015;45:2527–39.

[49] Hu X, Taylor PC, Cai M, et al. Inter-model warming projection spread: inherited traits from control climate diversity. Sci Rep 2017;7:4300.

[50] Webster PJ, Hoyos C. Prediction of monsoon rainfall and river discharge on 15–30 day time scales. Bull Amer Meteorol Soc 2004;85:1745–65.

[51] Liu BQ, Zhu CW. Potential skill map of predictors applied to the seasonal forecast of summer rainfall in China. J Appl Meteorol Sci 2020;31:570–82 (in Chinese).

[52] Huang Y, Wang H, Zhang P. A skillful method for precipitation prediction over eastern China. Atmos Oceanic Sci Lett 2022;15:100133.

[53] MacLachlan C, Arribas A, Peterson KA, et al. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. Q J R Meteorol Soc 2015;141:1072–84.

[54] Xing W, Wang B, Yim SY. Long-lead seasonal prediction of China summer rainfall using an EOF-PLS regression-based methodology. J Clim 2016;29:1783–96.

[55] Lang XM, Wang HJ. Improving extraseasonal summer rainfall prediction by merging information from GCMs and observations. Weather Forecast 2010;25:1263–74.

[56] Lu RY. Anomalies in the tropics associated with the heavy rainfall in East Asia during the summer of 1998. Adv Atmos Sci 2000;17:205–20.

[57] Lu E, Cai WY, Jiang ZH, et al. The day-to-day monitoring of the 2011 severe drought in China. Clim Dyn 2014;43:1–9.

[58] Zhang Q, Zhao Y, Fan S. Development of hourly precipitation datasets for national meteorological stations in China. Torrential Rain Disasters 2016;35:182–6 (in Chinese).

[59] Day JA, Fung I, Liu WH. Changing character of rainfall in eastern China, 1951–2007. Proc Natl Acad Sci USA 2018;115:2016–21.

Jieru Ma received her Ph.D. degree in the College of Atmospheric Sciences, Lanzhou University in 2022. She participated in a Joint Training Ph.D. Project to pursue study in the Department of Earth, Ocean and Atmospheric Science, Florida State University as a research scholar from 2019 to 2022. Her research focuses on climate variability and prediction, especially climate dynamics and physics in East Asia and the Tibetan Plateau.

Ming Cai is a professor in the Department of Earth, Ocean and Atmospheric Science, Florida State University. His research focuses on atmospheric general circulation, large-scale atmospheric dynamics, dynamics and physics of climate variability and climate change, and their predictions.