VU Research Portal

On the reliability of composite analysis

Li, Lintao; Dolman, Albertus J.

Link to publication in VU Research Portal

On the reliability of composite analysis: an example of wet summers in North China

Lintao Li [a,b,*], Albertus J. Dolman [c]

[a] School of Geographical Science, Qinghai Normal University, Xining 810008, China
[b] Earth and Climate Cluster, Faculty of Science, VU University Amsterdam, Amsterdam, the Netherlands
[c] Royal NIOZ, the Netherlands Institute for Sea Research, Den Burg, 59, 1790 AB Texel, the Netherlands

ARTICLE INFO

ABSTRACT

The validity of composite analysis is seldom discussed despite the fact that it can yield conflicting results. Our study confirms its validity by adding a reliability analysis to the classical composite analysis. Based on the signals extracted from composite analysis, 10 of the 14 wet summers in North China (1951–2020) can be "predicted" retrospectively. This study suggested the cyclonic anomaly over Mongolia at 500- and 850-hPa is closely associated with wet summers in North China. Interestingly, we found the most profound effects come from the Southern Hemisphere, with high confidence levels and large magnitude of the composite anomalies. These composite results are further cross-validated. We show that, based on the signals extracted from composite analysis, previously unseen wet summers in North China can be predicted with the mean absolute percentage error (MAPE) around 6%.

1. Introduction

Composite analysis, sometimes also referred to as epoch analysis, is widely used to detect relations between climate phenomenon and potential causal factors. Composite analysis is considered as simple in the sense that only two basic steps are required to form a composite analysis (von Storch and Zwiers, 1999): form sets (Θ) of the events and calculate the composite mean values (e.g. Prein et al., 2017). In practice, the second step is often modified by calculating either the composite anomaly (e.g. Ullah et al., 2021) or the composite difference (e.g. Nan et al., 2021). Since the signals of physical mechanisms leading to atmospheric phenomena do exist within the historical data, the accumulation and averaging of successive events can amplify these signals while the stochastic background noise can be substantially reduced (Laken and Calogović, 2013). Composite analysis is thus a simple and powerful approach to detect how a low-amplitude signal finds its expression in other variables.

While composite approach appears simple and powerful in theory, in practice, inconsistent results can be produced by composite studies. For example, Gao et al. (2014) using composite analysis found that a stronger western North Pacific subtropical high (WNPSH) brings more rainfall to North China. In contrast, Hao et al. (2010) and Feng and Hu (2004) found that North China summer rainfall variation is not affected to any significant extent by the WNPSH. Zhao et al. (2010) even found a more westward (i.e. stronger) WNPSH is related to shorter rainy season and thus less summer rainfall over North China. Inconsistent results can be found in other applications of this approach (e.g. Laken and Calogović, 2013).

The key question is: to what extent can we still use composite analysis? In the present paper we will first apply a traditional composite analysis. Importantly, a post-processing approach is proposed to demonstrate the reliability of the composite results, using the example of wet summers in North China, a highly stressed area of ground water (Qiu et al., 2016). North China (35–40°N, 110–120°E, Fig. 1) is one of the most densely populated and most intensively irrigated regions in the world (Liu et al., 2015). The huge population, agricultural irrigation, and industrial development demand a much larger amount of water resources than that which can be supplied by the regional rainfall and river runoff. Similar to many other regions in the world (Pimentel et al., 2004), North China mines groundwater to meet the imbalance, but much heavier, at a speed far exceeding the recharge rate (Qiu et al., 2016; Huang and Pang, 2013).

Precipitation is the direct natural water supply, which recharges groundwater and relieves the environmental stress. In North China, a

---

wet summer (June, July, August, JJA) is very important for groundwater recharge since 69% of its annual precipitation is concentrated in summer (Qian and Zhou, 2014). In this paper, we are interested in climate conditions/patterns favourable to large summer precipitation in North China.

Numerous studies have documented the linkages between precipitation over North China and various climate factors during the past decades (e.g. Ouyang et al., 2014). Using long precipitation data series from Dai et al. (1997), Qian and Zhou (2014) associated multidecadal variability of North China aridity with the Pacific decadal oscillation (PDO) phase changes. Using composite analysis, they further found that under the PDO+ phase, a wave train similar to the Pacific-Japan pattern-like teleconnection suppressed the northward monsoon flow and resulted in a deficit of summer rainfall in North China. Using a similar precipitation dataset, however, Feng and Hu (2004) demonstrated that the Pacific-Japan teleconnection pattern has a rather weak effect on rainfall in North China. On interannual time scales, many studies associated the variation of summer rainfall in North China with the El Niño-Southern Oscillation (ENSO). They found that summer rainfall in North China is suppressed when El Niño is in the developing and mature phase while it is stimulated when an El Niño is fading (e.g. Zhang et al., 1999; Feng and Hu, 2004). A considerable amount of studies attributed the summer rainfall variation in North China to the interannual variation of the East Asian summer monsoon (EASM, e.g. Ding, 1994) while other studies found the Indian summer monsoon plays a role (e.g. Zhang et al., 1999; Feng and Hu, 2004). However the mechanisms responsible for the variation of both monsoon systems remain in debate (Zhao-S et al., 2015; Turner and Annamalai, 2012; Wu et al., 2012; Boos and Kuang, 2013; He et al., 2015). These factors of course do not operate in isolation, but interact in intricate connections across the globe (Zahn, 2003). These ocean-atmosphere teleconnections influence rainfall indirectly, and the signals of physical mechanisms leading to the rainfall variation are weak compared to their background noise.

Associations between atmospheric circulation and precipitation in North China have attracted much attention in previous studies. At the upper-level of the atmosphere, it is known that the East Asian westerly jet (EAJ) has a profound effect on precipitation in North China (Liang and Wang, 1998; Zhao-G et al., 2015; Qu and Huang, 2012). Ding et al. (2008) documented that the weakening of the tropical upper-level easterly jet (TEJ) contributed to the weakening of the Asian summer monsoon system and subsequently to the dry trend of the summer rainfall in North China. Yu et al. (2004) attributed the increased droughts in North China to the summer cooling at the upper troposphere over extra-tropical East Asia. At the middle-level of the atmosphere, the WNPSH is widely believed to be one of the key influencing factors of the EASM and subsequently the summer rainfall in North China (Li et al., 2016; Zhu et al., 2011; Gao et al., 2014). Meanwhile, the cyclonic anomaly over Mongolian area, which is induced by warmer air temperature anomaly, is emphasized by previous studies as important to wet summers in North China (Hao et al., 2010; Ding et al., 2008). Atmospheric circulation at the lower-level is generally believed to be the reflection and result of the aforementioned factors (Ding et al., 2008). It is also closely related to the water vapour supply for North China (Hao et al., 2010).

The prime motivation behind this work is to evaluate and show the reliability of composite analysis. To do so, a new post-processing method is developed in Section 4. This is illustrated using an example with practical importance, namely wet summers in North China. The classical

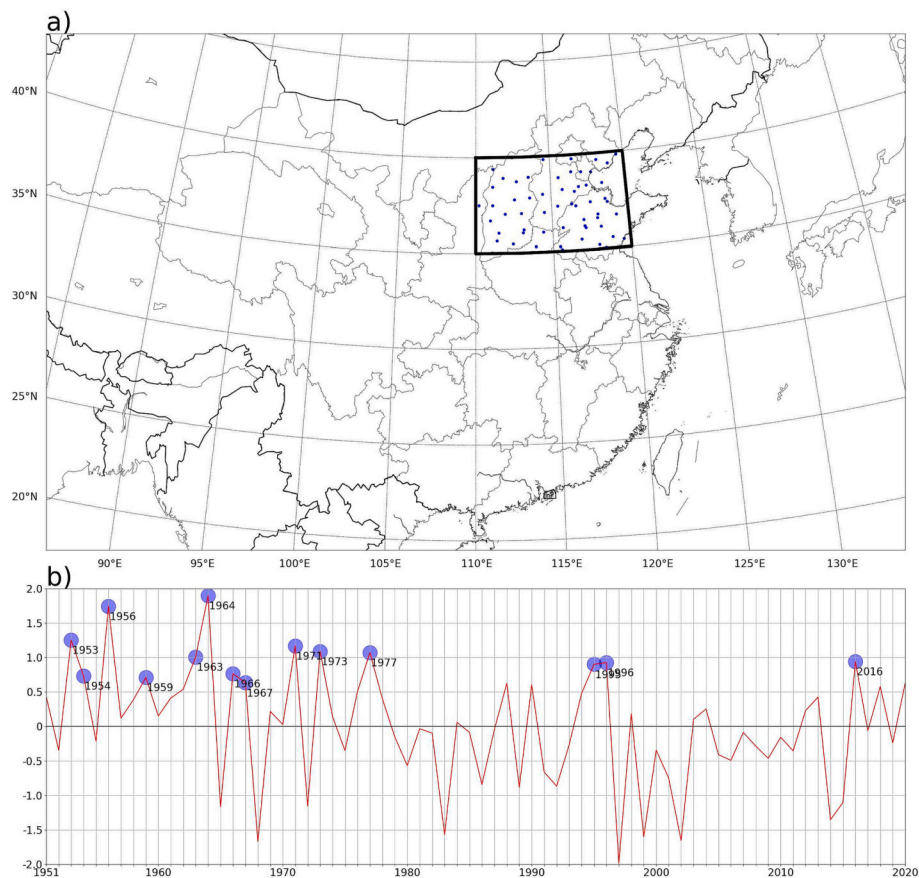**Fig. 1.** (a) The domain of North China in this study. The 58 national meteorological stations within North China are illustrated by the dots. (b) Time series of summer (June, July, and August, JJA) rainfall anomalies (mm/day) averaged from the 58 stations. The 14 wet summers are identified by purple dots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

composite analysis on the features favouring large summer precipitation in North China are identified and discussed in Section 3. Section 2 describes the data sources and methods used in the present study. A summary and some discussions are given in Section 5.

2. Data and methods

2.1. Data and target domain

The National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP/NCAR) reanalysis products (Kalnay et al., 1996), including the geopotential height (GPH), air temperature, and wind fields at pressure levels of 850, 500, and 200 hPa are used for the composite analysis and reliability analysis. The NCEP/NCAR data sets used in this study are provided by the Physical Sciences Laboratory (PSL) in National Oceanic and Atmospheric administration (NOAA, https://psl.noaa.gov/). All the reanalysis data sets are at a 2.5° latitude × 2.5° longitude resolution. The data for SST is from the NOAA Extended Reconstructed SST V5 (ERSST5, Huang et al., 2017), at a 2.0° latitude × 2.0° longitude resolution. Monthly accumulated precipitation data from 58 national meteorological stations within North China (Fig. 1a) are used for wet summer identification and for the calculation of reversibility in Section 4. The station data set is provided by the China Meteorological Administration (http://data.cma.cn/). All the datasets used in this study are from January 1951 to December 2020, which is mainly determined by the availability of the station precipitation data set.

We define North China as the area within 35–40°N and 110–120°E for the present study (Fig. 1a). The coordinates are "round" numbers, which fit the most stressed area of groundwater provided by Qiu et al. (2016). There is a consensus that a significant decreasing trend can be found for summer precipitation in North China from 1960s to the beginning of 2010s (e.g. Ding et al., 2008; Hao et al., 2010). While North China has noticeable wet summers in recent years, according to the 58 national station observations, most wet summers are still concentrated in 1951–1977 (Fig. 1b).

2.2. Composite analysis

It is necessary to define 'wet summer' for the present study. As discussed before, we are interested in large summer (JJA) precipitation in North China. Fig. 1b shows time series for precipitation anomalies in summer, based on data from the national meteorological stations shown in Fig. 1a. The 14 summers with largest precipitation amount (top 20%) are highlighted by purple dots and are hereafter referred to as wet summers.

The typical states (V_Θ) of wet summer can now be reconstructed by calculating the composite mean values of the 14 wet summers. The statistical significance is then determined by a two-tailed Student's *t*-test. By subtracting the mean climate state (V_{cli}) from the composite mean (V_{com}), we get the composite anomaly values (V_{ano}). The base period of 1951–2020 is used for calculating the mean climate state.

2.3. The notion of reversibility

As noted, the biggest advantage of composite analysis is that, by accumulating and averaging of a succession of selected events, the signals of the physical mechanisms driving a particular phenomenon can be amplified. Intuitively, if the signals are captured appropriately by the sets (Θ) of the selected events, most of the events (i.e. wet summers in North China) should be retrospectively predictable based on the captured signals. This is what we call "reversibility", similar to Boschat et al. (2016). In the statement below, the terminology "predict" actually means "predict retrospectively", unless otherwise noted.

Before introducing our new method to quantify the reliability, we give exact definitions of key terms used in this study. In the composite

anomaly figure for each interested variable, regions where the statistical significance exceeds a particular confidence level are either shaded (Fig. 2) or contoured (e.g. Fig. 3 a, c). These regions are named as key regions in the sense that the composite anomaly values within the key regions occur more often in wet summers than they do climatologically. The sign and magnitude of the regional mean composite anomaly value within a key region is named as key composite feature of the corresponding region or key feature in short. All the key features together in a particular composite anomaly figure for a certain variable build up the key composite features of the variable, or simply composite features. The composite features of all the interested variables define a composite scenario, which contains the accumulated signals of the physical mechanisms driving a particular phenomenon (e.g. the wet summers in North China).

In order to "predict" whether a phenomenon occurs, the term scenario similarity is defined. The steps of calculating scenario similarity proceed as follows. First, for each key region, we consider whether a score of zero or one should be assigned according to:

$$score = \begin{cases} 1, & \text{if } A > 0 \text{ and } B > \alpha A \\ 1, & \text{if } A < 0 \text{ and } B < \alpha A \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where A is the regional mean of the composite anomaly value, B is the mean anomaly value for a particular variable in a particular year within the same region as A, and α is the threshold value. It can be set to be between −1 and 1 (negative to positive correlation). In this study, we are only interested in α values between 0 and 1, with bins of 0.1 width.

Second, adding scores of all the identified key regions for all the interested variables derives the total score (s) for a particular year.

Third, the scenario similarity is defined as:

$$\frac{s}{N} \times 100\% \tag{2}$$

where s is the total score of a particular year, N is the number of all the key regions in the composite scenario. N is the largest score that the scenario of a particular summer can be assigned, in which case the scenario similarity equals 1. Therefore, N is effectively the potential full score.

The scenario similarity quantitatively describes the similarity between the climate signals captured by composite analysis and the signals of a particular summer, not only the similarity of magnitude, but also the similarity of spatial distribution. When the scenario similarity is high, it indicates that the signals of the physical mechanisms driving wet summers in North China are strong in a particular summer. Consequently, more summer rainfall in North China is expected in that summer.

The reversibility can be evaluated qualitatively by calculating the linear relationship between the summer precipitation anomalies and the scenario similarities (i.e. the signal strength of the physical mechanisms driving wet summer in North China). When the correlation coefficient is relatively large and the linear relationship is statistically significant, we can qualitatively say that the reversibility is existed and the corresponding composite analysis is therefore reliable.

The reversibility can further be evaluated quantitatively by assessing the accuracy of the "predicted" summer rainfall based on the scenario similarity. The ideal state is that a summer with the highest scenario similarity value has the largest amount of observed rainfall; the second scenario similarity value then corresponds to the second largest amount of rainfall, and so forth. In practice however, the ideal state does not exist. Therefore, we apply the root-mean-square (RMS) operation to the queue number errors between the predicted and the observed summer rainfall, with a smaller RMS value corresponding to higher prediction accuracy. We consider queue number here because return period or frequency is commonly used to understand (extreme) climate events (e. g. Read and Vogel, 2015; Mallakpour and Villarini, 2015). Based on the procedures we introduced here, other aspects (e.g. magnitude) of events
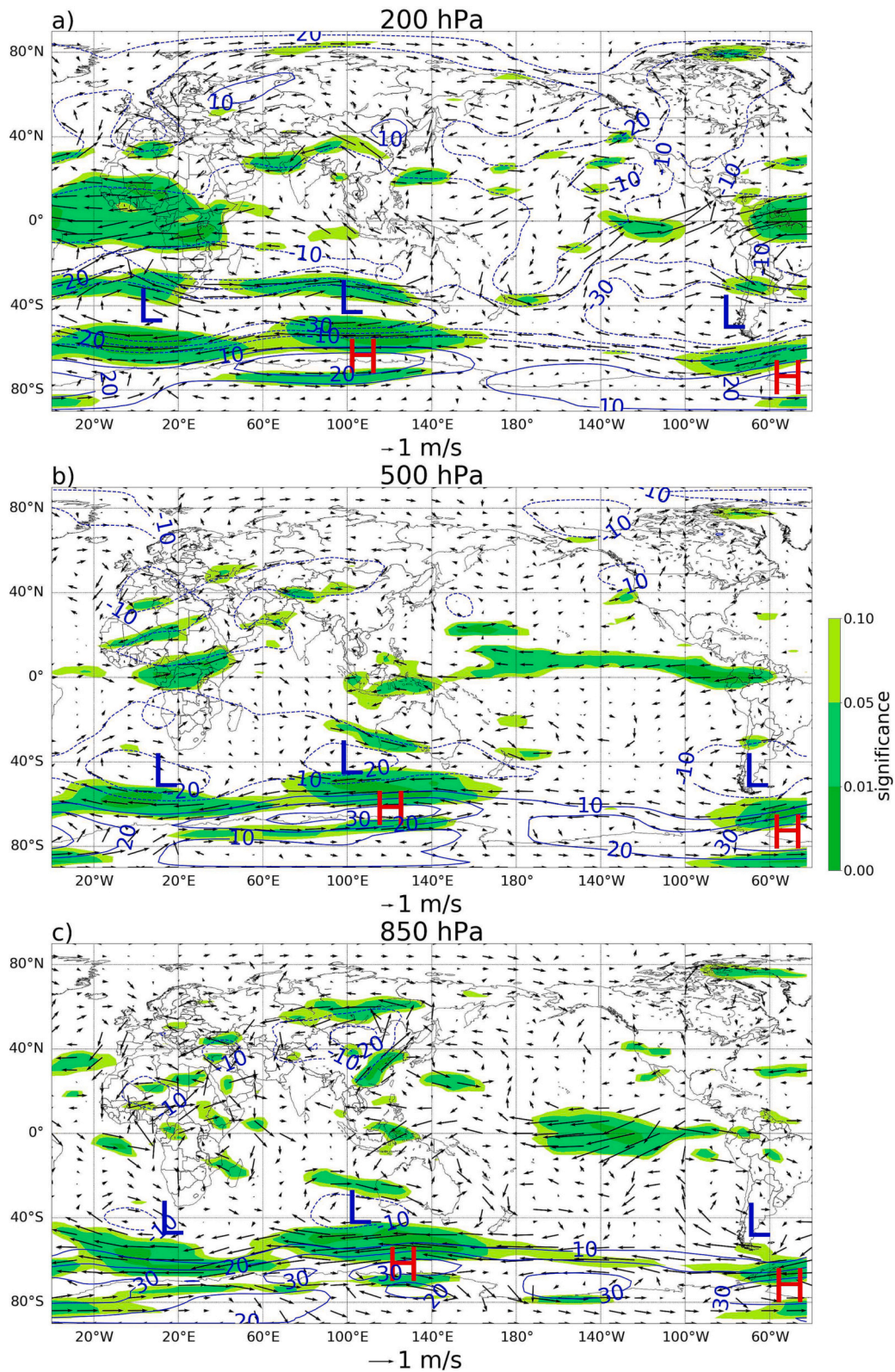
Fig. 2. Composite anomalies of geopotential height (GPH) and winds at (a) 200 hPa, (b) 500 hPa, (c) 850 hPa. The base period of 1951–2020 is used for calculating the mean climate state. Shadings denote grids where significance levels (*p*-value) of zonal wind anomalies are at 10, 5, and 1%. L and H denote the location of the cyclonic and anticyclonic anomalies, respectively.
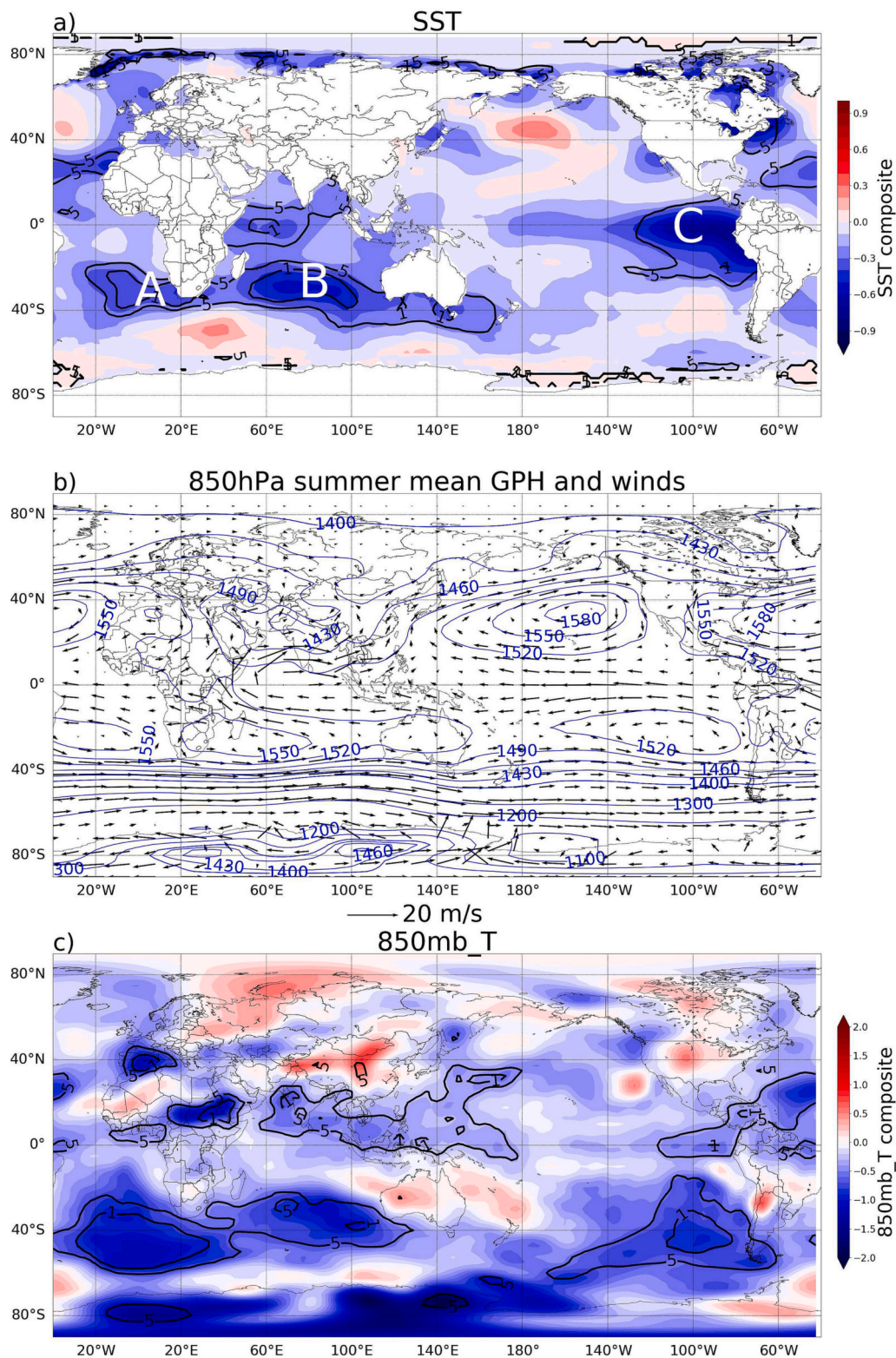
Fig. 3. (a) Composite anomalies of SST. (b) Long term summer mean GPH and winds at 850 hPa. (c) Composite anomalies of air temperature at 850 hPa. The base period of 1951–2020 is used for calculating the mean climate state. Black contours in (a) and (c) are significance levels of the temperature anomalies where the percent sign has been omitted. A, B, and C in (a) denote the noticeable key regions of the SST composite.

and prediction measurements beside the RMS operation (e.g. Kutner et al., 2005; Hyndman and Koehler, 2006) can be easily adapted.

2.4. Cross-validation and sensitivity analysis

While reversibility analysis shows the reliability of composite analysis, one may want to simply know whether signals extracted by composite analysis can predict previously unseen climate events. Therefore,

our reversibility analysis is further cross validated. Since the data sets of 14 wet summers is severely limited, leave-one-out cross-validation is used (Cawley and Talbot, 2003; Arlot and Celisse, 2010). Specifically, we hold-out one wet summer and calculate composite scenario of the other 13 wet summers. Then the previously unseen wet summer (i.e. the hold-out summer) precipitation is predicted based on its scenario similarity. This process iterates 14 times, each iteration a different wet summer is held-out for validation while the remaining 13 summers are used for calculating composite scenario. At last, the accuracy of the prediction is measured by the well-known mean absolute percentage error (MAPE) method, which is widely used to measure the forecast accuracy. Let O_t and F_t denote the observed and forecasted values at data point t, respectively. According to Kim and Kim (2016), the the absolute percentage error (APE) is defined as:

$$APE = \sum_{t=1}^{N} \left| \frac{O_t - F_t}{O_t} \right| \tag{3}$$

and the the MAPE is defined as:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{O_t - F_t}{O_t} \right| \tag{4}$$

To test the sensitivity of the composite analysis and the reliability analysis, we repeated all analyses using other gridded datasets including ERA5 (http://climate.copernicus.eu/products/climate-reanalysis) and HadSST4 (https://www.metoffice.gov.uk/hadobs/hadsst4/). The results generated from ERA5 and HadSST4 agree well with these from NCEP/NCAR and ERSST5. Therefore, most results presented in this study are derived from NCEP/NCAR and ERSST5, unless otherwise noted.

3. Composite results

In order to examine large scale atmospheric conditions associated with wet summers in North China, global composite analyses were performed for the 14 wet summers shown in Fig. 1b. The composite results are shown in Figs. 2 and 3. Composite anomalies which are significant at the 90, 95, and 99% confidence levels are shown in shadings (Fig. 2) or contours (Fig. 3a, c). These areas are named as key regions in the sense that they contain important signals of physical mechanisms leading to a wet summer in North China. The corresponding anomaly patterns within key regions are named as key features, which are closely associated with wet summers in North China in the sense that they occur significantly more often during wet summers than they do climatologically. Figs. 2 and 3 are displayed in a global view because the identified key regions are distributed worldwide. As such, valuable signals will not be missed and the reliability of the composite analysis can be achieved properly.

At 200 hPa (Fig. 2a), no significant signal of the EAJ depicted in previous studies near 40 N can be detected. The TEJ is strengthened, while the statistical significant area of the intensified TEJ is mainly over the tropical Atlantic. Interestingly, the most noticeable key features are located in the Southern Hemisphere: intensified upper-level westerlies near 30°S and weakened westerlies near 60°S. This suggests that the Southern Hemisphere plays a significant role in the dynamical processes associated with wet summers in North China, which is in line with Li et al. (2012). The weakened westerlies near 60°S may associated with a negative phase of the Southern Annular Mode (SAM, http://www.bom.gov.au/climate/enso/history/ln-2010-12/SAM-what.shtml). For the mechanisms how SAM influences the precipitation in the North Hemisphere, the reader is referred to Dou et al. (2020) and its references. Fig. 2a shows that the anomalies of the westerly winds in the Southern Hemisphere are generally induced by three cyclonic anomalies and two anticyclonic anomalies (Ls and Hs in Fig. 2a). The cyclonic anomalies are located over the southeast Atlantic, the south Indian, and the

southeast Pacific Oceans near 40–45°S, and the anticyclonic anomalies are near (100°E, 70°S) and (50°W, 75°S). According to Peixoto and Oort (1992), those upper-level anomalies are strongly depended on the forcing at the surface. Indeed, comparing the air temperature anomaly maps at 200, 500 (not shown here), and 850 hPa (Fig. 3c), we find that significant cool anomalies can be found in regions with cyclonic anomalies at 200 hPa (Fig. 2a). Corresponding to the anticyclonic anomalies, warm anomalies can be found at 500 hPa near (100°E, 70°S) and 850 hPa near (50°W, 75°S). While the cool anomalies generally reach the 1% significance level, neither of the warm anomalies are significant, even for the 10% significance level.

At 500 hPa (Fig. 2b), the three cyclonic anomalies and the two anticyclonic anomalies at 200 hPa, as well as the associated westerly anomalies are noticeable as well. The tropical easterly jet is intensified significantly over the Pacific. Similar to the distribution of key regions at 200 hPa, noticeable signals are mainly found in the Southern Hemisphere and the tropical oceans. In the Northern Hemisphere, a noticeable climatic signal is the wave train between 40 and 60°N. The cyclonic anomaly over Mongolia is a profound feature of this wave train due to its high confidence level and its short distance to North China. According to Ding et al. (2008), this cyclonic anomaly is associated with more frequent invasion of cold air from high latitude to North China. The cold air interacts with relatively warm and moist air over North China and thus brings more summer rainfall there.

At 850 hPa (Fig. 2c), the pattern of circulation anomaly is similar to 500 hPa, including the weakened westerlies in the Southern Hemisphere and the strengthened easterlies over the tropical Pacific. In the Northern Hemisphere, the large cyclonic anomaly over Mongolia and the associated wind anomalies are the nearest and thus the most direct processes associated with wet summer in North China. On the long term summer mean 850 hPa GPH map, a trough lays over Mongolia (Fig. 3b). A cyclonic anomaly indicates that, during wet northern summer, the trough is either deeper or more persistent than its climatological mean condition, or both. This deeper and more persistent low level trough favours wet summer in North China. On one hand, its southwest flank transports more cold air from northwest down to North China, which is verified by the stronger zonal winds to its south-southwest. On the other hand, on its southeast flank, more warm moist air from south can be transported to North China. This is verified by the stronger zonal and meridional winds on its southeast flank (Fig. 2c). When cold air from northwest meets the warm moist air from south, rainfall is more likely to occur (Li and Dolman, 2016).

The importance of the intensified trough to the rainfall in North China has long been documented (e.g. Dai et al., 2005). As for the water vapour transport, it is generally believed to be either due to the WNPSH or due to the Tibetan low, which drives the Eastern Asian summer monsoon (e.g. Li et al., 2016). According to our composite study, however, the anomaly of warm moist southerly flow is mainly due to the intensified trough over Mongolia, since only weak and non-significant intensification can be found from its two counterparts (Fig. 2b, c).

Since the circulation anomalies are frequently documented as being forced by the SST (e.g. Peixoto and Oort, 1992; Boschat et al., 2016), the composite anomalies of SST were calculated and shown in Fig. 3a. Despite the SST pattern having remarkable interannual variation, two regions show up as significantly associated with wet summer in North China. The most noticeable feature is the robust cold chain from 20°W to 150°E near 30°S. Within this cold chain, four spots are characterized by a high confidence level, which is above 99%. The two largest spots are located to the west and east of South Africa, and are marked as spot A and B in Fig. 3a, respectively. The second region (spot C) is located in the eastern tropical Pacific, with somewhat lower confidence level of 95%, but noticeable anomaly magnitude (approximately -1 °C near the cool center).

The cold chain from 20°W to 150°E near 30°S shows up as the most noticeable SST anomaly feature not only for the large scope and magnitude of the anomaly but also for its high confidence level. This

cold chain, together with warmer SST to its south, sets up a very similar temperature pattern of the low-level air in the Southern Hemisphere (Fig. 3c). Cold air near 40°S, together with warmer air to its south, reduces the temperature contrast between 40°S and 60°S. Consequently, the pressure gradient is lower than the climate mean and the prevailing westerly winds are weaker in wet summer, agreeing with the circulation pattern discussed above.

Another noticeable feature revealed in Fig. 3a is the dramatic cooling of the eastern to central tropical Pacific during wet summer in North China. Despite the fact that the confidence level of the cooling anomaly is not as high as the cold chain near South Africa, the magnitude of the anomaly is much larger, approximately −1 °C near the cooling center (Fig. 3a). This cooling anomaly may imply that the summer rainfall in North China can be associated with ENSO and prefers the La Niña episode. However, this depends on the definition of ENSO. On the one hand, when the Oceanic Niño Index was calculated based on the Niño 3.4 index defined by Trenberth (1997), we found that the wet years of 1953 and 1963 are identified as El Niños, 1954, 1964, 1971 and 1973 are La Niñas, and all other seven wet years are identified as neutral. In this case, no significant association can be found between ENSO and wet summers in North China based on data sets uses in the present study. On the other hand, when we define El Niño as the positive summer SST anomaly in Niño 3.4 Region, La Niña as the negative anomaly, and without neutral year, we found that except 1953, 1963, 1966, and 1977, all the other nine wet years are La Niñas. In this case, wet summer in North China is associated with ENSO.

4. Reliability of composite analysis

To show the usefulness of composite analysis, an algorithm is needed to evaluate the quality of composite result objectively.

4.1. Evaluate the composite result qualitatively

To evaluate the composite result objectively, we calculate the scenario similarities for key regions defined by significance levels (p-value) of 0.1, 0.05, and 0.01, threshold values (α in eq. 1) between 0 and 1, with bins of 0.1 width. In the present study, the scenario similarity is calculated based on 13 composite features, corresponding to 13 variables including the geopotential height, air temperature, zonal and meridional winds at pressure levels of 850, 500, and 200 hPa, and SST. The two figures in the Appendix give an example showing how key region, key feature, and composite feature defined. The linear relationships between the summer precipitation anomalies and the scenario similarities are calculated and plotted in Fig. 4.

Fig. 4 is a scatter diagram illustrating statistical relations between summer precipitation anomalies in North China and the corresponding total scores (s), under three confidence levels and ten threshold values (α). The values of s and N are displayed instead of the scenario similarities in expression (2) because showing the N values is easier for the discussions below, also because the scenario similarity is easy to be derived from N by expression (2). Fig. 4 clearly suggests that positive relationships exist between the scenario similarities and summer rainfall anomalies. All the linear relationships are statistically significant above the confidence level of 99.9% and all the correlation coefficients are larger than 0.5 (not shown here). Based on the linear relationship between the scenario similarity and summer rainfall anomaly, wet summer in North China can now be predicted by the scenario similarity (recall its definition in Section 2).

Since wet summers in North China can now be predicted, it is safe to say that when the climate condition is similar to the composite scenario shown in Section 3, a wet summer in North China is more likely to happen (the probability is 10 out of 14, see below). Therefore, the reversibility of the composite results discussed in Section 3 has now been shown to exist and the composite results are reliable.
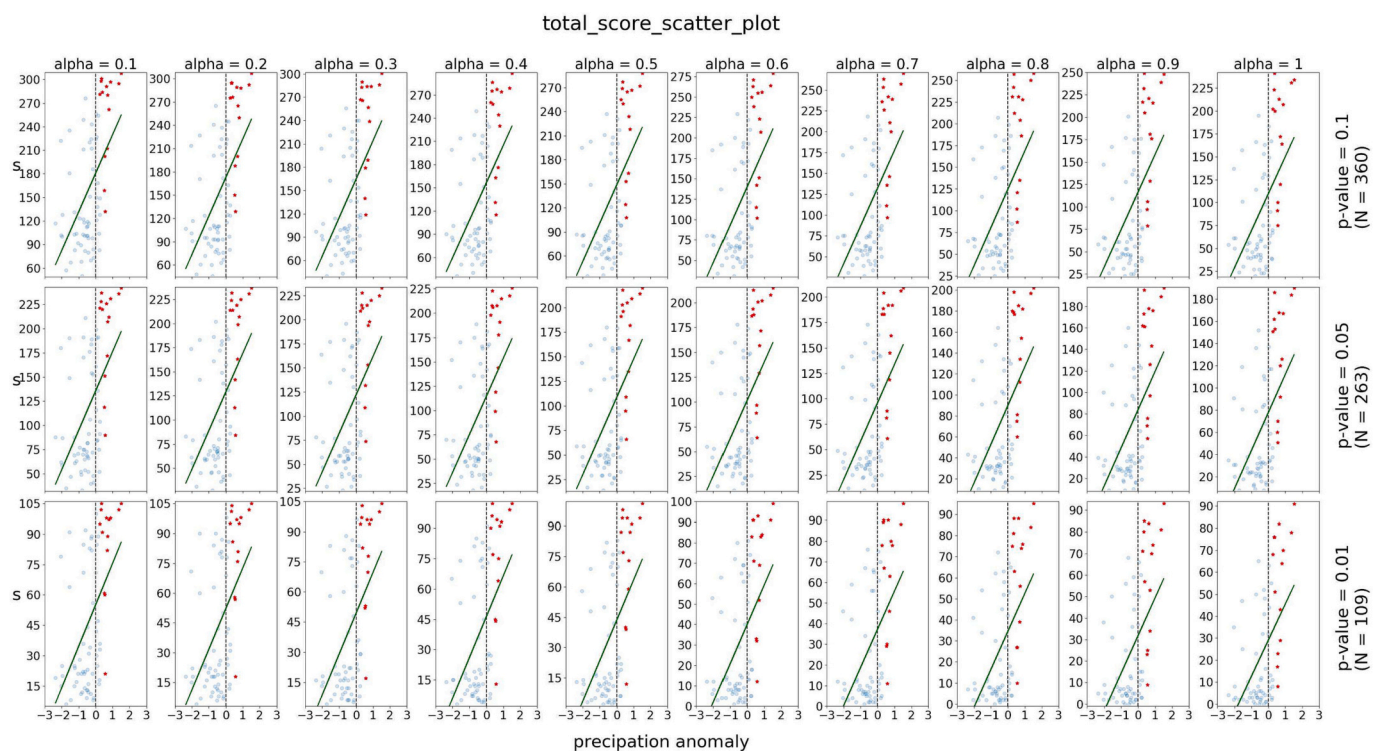
Fig. 4. Scatter diagram illustrating statistical relations between summer precipitation anomalies in North China and the corresponding total scores(s in expression 2), under particular confidence levels (p-values) and a variety of threshold values(α). The 14 red dots are the selected wet summers shown in Fig. 1b. The values of s and N are displayed instead of the scenario similarities, which can be derived by expression (2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Evaluate the composite result quantitatively

Until now, we have shown that the scenario similarity is positively related to rainfall anomaly in North China. The composite analysis formed in Section 3 is thus reliable, not only because that the key features identified by the composite analysis occur significantly more often during wet summers in North China than they do climatologically, but also because that, based on the signals accumulated from the indentified key features, wet summers can generally be "predicted".

However, in many cases, it may require a quantitative evaluation of the predictability/reversibility of signals extracted from composite analysis. For example, Fig. 4 gives 30 examples of linear models predicting wet summers in North China based on the scenario similarity. Despite the fact that these linear models are similar on visual inspection, and despite all the linear relationships are statistically significant, differences of predictability do exist between these linear models. The differences are more obvious for the distribution of the red dots, which correspond to the selected wet summers shown in Fig. 1b. In this case, one may want to know, which model has better predictability of wet summers?

The accuracy of the prediction is used to calculate the accuracy of reversibility. We apply a root-mean-square (RMS) operation to the queue number difference between the predicted and the observed summer rainfall. Table 1 is an example illustrating how this process is carried out for the condition when *p*-value is 0.05 and α is 1, which corresponds to the scatter plot in the second row and the last column in Fig. 4. It can be seen from Table 1 that 10 of the 14 wet summers are predicted as the top 20%. Specifically, the wettest summer observed in 1964 is predicted as the wettest as well.

The results of the RMS operation for all the conditions displayed in Fig. 4 are shown in Table 2. It appears that a higher confidence level used in composite analysis does not necessarily leads to a higher accuracy of reversibility. The reason lies in the key region identification. A higher confidence level normally leads to fewer and smaller key regions (see key regions shown in Fig. 2 as examples), and thus smaller potential full score (*N*). This is further verified by Fig. 4. On its right side, the potential full scores are 360, 263, and 109 for confidence levels of 90, 95, and 99%, respectively. Since a higher accuracy of reversibility can be reached by the confidence level of 95%, it is reasonable to believe that signals from the additional regions, which are identified by confidence level of 95% besides the key regions identified by 99%, do have influence on the occurrence of wet summers in North China.

Similarly, it can be seen from Table 2 that a higher threshold in calculating the scenario similarity does not necessarily leads to a higher accuracy of reversibility either. A higher threshold increases the similarities in magnitude. However, it makes more key regions get the score of zero. Signals from these regions are thus excluded in predicting wet summer in North China. It is not clear whether signals of the excluded regions are more powerful in predicting wet summers than the improvement of magnitude similarity by increasing the threshold (Table 2). This explains why the ensemble-mean scenario similaritiy is used in Section 4.3 for the cross-validation.

Note that scores indicating scenario similarity shown in Fig. 4 are not suitable for predicting neutral or dry years. For these purposes, corresponding composite analysis based on neutral and dry events are required respectively. It also explains the reason why only 14 wet years are used for the RMS calculation in Table 1. In general, the events used in the RMS operation should be the same as the events used in forming the composite analysis.

4.3. Cross-validation

To examine the reliability of composite analysis explectly, the reversibility is further cross-validated. First, we hold-out one wet summer (1953), and calculate the ensemble-mean scenario similarities for the remaining 13 wet summers. Fig. 5 shows the linear relationship between the regional mean summer rainfall and the ensemble-mean scenario similarity of the remaining 13 wet summers (blue triangles). The black line is the fitted linear regression model. Second, we calculate the ensemble-mean scenario similarity of the hold-out year (1953). Third, the observed regional mean rainfall in the summer of 1953 is 5.3 mm per day and the predicted value is 5. Therefore, the absolute percentage error (APE) of the prediction is 0.3 and the absolute percentage error is 5.8% (rounded numbers).

This process iterates 14 times. Each time, a different wet summer is heldout for validation while the remaining 13 summers are used for calculating composite scenario. Table 3 lists the observed and predicted rainfall for each wet summer and the corresponding APE value. Overall, the previously unseen wet summers can be predicted with the mean absolute percentage error around 6%, further showing the power of composite analysis.

5. Summary and discussions

Composite analyses have been frequently performed to form hypotheses of various climate relations. However, conflicting conclusions can be obtained (Boschat et al., 2016; Laken and Calogović, 2013). According to Laken and Calogović (2013) several issues may affect composite results, including (1) signal-to-noise ratios related to spatio-temporal restrictions; (2) interference from variability in data at time-scales greater than those concerning hypothesis testing; (3) biases imposed by the use of improper normalization procedures. While the present study was not designed to remove these issues, we developed a method to quantitatively evaluate the reliability of the composite analysis which allows us to detect unsuitable settings.

Such a complete composite study (i.e. the traditional composite analysis plus reliability analysis) is illustrated using wet summers in North China as an example. Traditional composite analyses on the global atmospheric circulation and SST features have been implemented to detect signals driving wet summers in North China. The analyses suggested that the cyclonic anomaly over Mongolia at 500- and 850-hPa is closely associated with wet summer in North China. Similar result has been achieved by previous composite studies (e.g. Ding et al., 2008). The importance of the intensified trough to the rainfall in North China has long been documented. For example, Ding et al. (2008) demonstrated that it is associated with more frequent invasion of cold air from high latitude to North China. The cold air interacts with relatively warm and moist air over North China and thus brings more summer rainfall there.

Table 1

An example illustrating how to conduct the root-mean-square (RMS) operation. The values correspond to the condition when *p*-value is 0.05 and α is 0.2 in Fig. 4. Specifically, the queue number of predicted rainfall is derived from *s/N*, which corresponds to the scenario similarity in Eq. (2).

Year	Queue number of observed rainfall	Queue number of predicted rainfall	difference	square
1964	1	1	0	0
1956	2	3	1	1
1953	3	5	2	4
1971	4	12	8	64
1973	5	13	8	64
1977	6	19	13	169
1963	7	4	−3	9
2016	8	30	22	484
1996	9	26	17	289
1995	10	28	18	324
1966	11	7	−4	16
1954	12	2	−10	100
1959	13	6	−7	49
1967	14	9	−5	25
sum				1598
mean				114.14
square root (RMS)				10.7

Table 2

The result of the root-mean-square operation for key regions defined by significance levels (*p*-value) of 0.1, 0.05, and 0.01 and threshold values (α in Eq. (1)) between 0 and 1, with bins of 0.1 width.

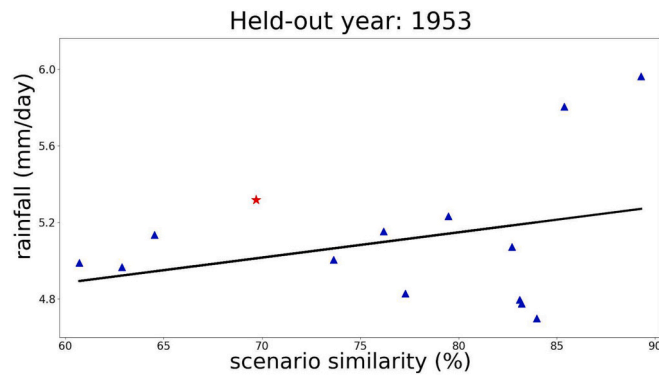|	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0 0.1	11.7	11.7	11.6	11.6	11.5	11.6	11.4	11.7	11.8	11.6
0.05	11.6	11.6	11.9	12	11.2	11.1	11.2	10.9	11	10.7
0.01	14.5	15	13.8	14.6	14.1	13.2	13.1	12.8	12.3	12.3

Fig. 5. Scatter diagram illustrating statistical relations between scenario similarity and summer precipitation. The black line is the fitted linear regression model, based on values of the remaining 12 wet summers (blue triangles). The red star shows the relation between the observed precipitation and scenario similarity for the hold-out year of 1953. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

A table showing the cross-validation results. The prediction accuracy is evaluated by the mean absolute percentage error (MAPE).

Hold-out year	Observed precip	Predicted precip	APE (%)
1953	5.3	5.0	5.8
1954	4.8	5.1	7.0
1956	5.8	5.1	12.9
1959	4.8	5.1	6.6
1963	5.1	5.1	0.0
1964	6.0	5.1	14.8
1966	4.8	5.0	3.0
1967	4.7	5.1	8.4
1971	5.2	5.0	4.7
1973	5.2	4.9	4.5
1977	5.1	4.9	5.2
1995	5.0	5.0	0.0
1996	5.0	4.9	2.2
2016	5.0	5.2	3.2
MAPE			**6.01**

Besides the importance on rainfall formation mechanism, we found the intensified trough plays a role in water vapour transport as well. Interestingly, we found the most profound effects come from the Southern Hemisphere, not only for the high confidence levels shown in Figs. 2 and 3, but also for the large magnitude of the composite anomalies, and for the large size of the key regions we defined in Section 4. We did however not set out to provide a novel explanation for these from a physical perspective.

Rather we aimed to address the question: to what extent can we believe the result of composite analysis? To answer this question, we proposed a post-processing method to the traditional composite analysis technique. The reliability of composite analysis was evaluated by a simple and straightforward notion. Similar to Boschat et al. (2016), the phrase 'reversibility' was used: when the signals driving wet summer in North China have been extracted appropriately, in turn, wet summers should be retrospectively predicted based on these signals. In the present study, we show that the composite analysis is quite reliable when it has been formed appropriately. Based on the signals from composite analysis, 10 of 14 wet summers could be predicted under various threshold values (α) and significance levels (*p*-value). Cross-validation further shows that, based on the signals extracted from composite analysis, previously unseen wet summers can be predicted with mean absolute percentage error (MAPE) around 6%.

The advantage of the reliability analysis lies in the following aspects. First, rather than using several boxes to extract signals subjectively, here we extract signals based on key regions objectively. The definition of key regions is of great importance. On the one hand, it guarantees that only the correct signals above a particular confidence level are extracted. Therefore the unrelated background noises will be removed in our reliability analysis. On the other hand, significant signals are extracted globally. Therefore valuable information will not be missed. Second, αA divides the magnitude of composite anomalies into ten degrees and thus the scenario similarity does not rely solely on any particular composite value. In this way, signals extracted based on various threshold values (α) can be compared and analyzed as discussed in Section 4.

Signals carried by different key regions are of different importance. Intuitively, we may think signals from larger key regions are more important than smaller key regions, signals from key regions with larger magnitude of composite anomalies are more important than smaller magnitudes, and signals from key regions with higher confidence levels are more important than lower confidence levels. For the last case, we shown that a higher confidence level used in composite analysis does not necessarily leads to a higher reversibility because signals from sufficient key regions matters.

One way to compare the relative importance of signals carried by different key regions is calculating the reversibility based on signals from each key region. Then all the reversibility can be quantified according to the method we proposed in Section 4. In theory, the most important signal comes from the key region corresponding to the least RMS value, the second most important signal corresponds to the second least RMS, and so forth. In practice, however, it is more difficult. The signals are generally so weak that no single signal can effectively predict wet summers in North China. Therefore, signals from more than one key region should be used to compare the overall relative importance. In this case, the number of the reversibility analysis we introduced in Section 4 will have to be repeated for ($2^N - N - 1$) times. Considering the large N values showing in Fig. 4, this will have severe computational limitations. More efficient algorithms may need to be developed to further release the power of composite analysis, and help to understand the nature of atmospheric relations better.

The present study evaluated the validity and reliability of the composite analysis by investigating the climatic factors of wet summers in North China. To our knowledge, this is the first study showing the reliability of the composite results. Therefore, it is safe to limit our conclusions to the specific case of wet summers in North China at this moment. Future studies can apply our methodology to other regions for other climate phenomena using different datasets. In this way, hopefully, the reliability of the composite analysis can be widely accepted.

CRediT authorship contribution statement

Lintao Li: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing,

Visualization. **Albertus J. Dolman:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2023.106881.

References

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79. https://doi.org/10.1214/09-SS054.

Boos, W., Kuang, Z., 2013. Sensitivity of the south Asian monsoon to elevated and non-elevated heating. Sci. Rep. 3, 1192. https://doi.org/10.1038/srep01192.

Boschat, G., Simmonds, I., Purich, A., Cowan, T., Pezza, A.B., 2016. On the use of composite analyses to form physical hypotheses: An example from heat wave-SST associations. Sci. Rep. 6, 29599. https://doi.org/10.1038/srep29599.

Cawley, G.C., Talbot, N.L., 2003. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. Pattern Recogn. 36, 2585–2592. https://doi.org/10.1016/S0031-3203(03)00136-5.

Dai, A., Fung, I.Y., Genio, A.D.D., 1997. Surface observed global land precipitation variations during 1900-88. J. Clim. 10, 2943–2962. https://doi.org/10.1175/1520-0442(1997)010<2943:SOGLPV>2.0.CO;2.

Dai, X.G., Fu, C.B., Wang, P., 2005. Interdecadal change of atmospheric stationary waves and North China drought. Chin. Phys. 14, 850.

Ding, Y., 1994. Monsoons over China. Kluwer Academic Publishers.

Ding, Y., Wang, Z., Sun, Y., 2008. 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. 28, 1139–1161. https://doi.org/10.1002/joc.1615.

Dou, J., Wu, Z., Li, J., 2020. The strengthened relationship between the yangtze river valley summer rainfall and the southern hemisphere annular mode in recent decades. Clim. Dyn. 54, 1607–1624.

Feng, S., Hu, Q., 2004. Variations in the teleconnection of ENSO and summer rainfall in northern China: a role of the Indian summer monsoon. J. Clim. 17, 4871–4881. https://doi.org/10.1175/JCLI-3245.1.

Gao, H., Jiang, W., Li, W., 2014. Changed relationships between the east Asian summer monsoon circulations and the summer rainfall in eastern China. Journal of Meteorological Research 28, 1075–1084. https://doi.org/10.1007/s13351-014-4327-5.

Hao, L., Min, J., Ding, Y., Wang, J., 2010. Relationship between reduction of summer precipitation in North China and atmospheric circulation anomalies. Journal of Water Resource and Protection 2, 569–576. https://doi.org/10.4236/jwarp.2010.26065.

He, B., Wu, G., Liu, Y., Bao, Q., 2015. Astronomical and hydrological perspective of mountain impacts on the Asian summer monsoon. Sci. Rep. 5, 17586. https://doi.org/10.1038/srep17586.

Huang, T., Pang, Z., 2013. Groundwater recharge and dynamics in northern China: Implications for sustainable utilization of groundwater. Procedia Earth and Planetary Science 7, 369–372. https://doi.org/10.1016/j.proeps.2013.03.182.

Huang, B., Thorne, P., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J., Menne, M., Smith, T.M., Vose, R., Zhang, H.M., 2017. Extended reconstructed sea surface temperature, version 5 (ersstv5): upgrades, validations, and intercomparisons. J. Clim. 30, 8179–8205.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am.

Meteorol. Soc. 77, 437–471. https://doi.org/10.1175/1520-0477(1996)077¡0437:TNYRP¿2.0.CO;2.

Kim, S., Kim, H., 2016. A new metric of absolute percentage error for intermittent demand forecasts. Int. J. Forecast. 32, 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003.

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied Linear Statistical Models. McGraw-Hill Irwin.

Laken, B.A., Calogović, J., 2013. Composite analysis with Monte Carlo method: An example with cosmic rays and clouds. J. Space Weather Space Clim. 3, A29. https://doi.org/10.1051/swsc/2013051.

Li, L., Dolman, A.J., 2016. A synoptic overview and moisture trajectory analysis of the "7.21" heavy rainfall event in Beijing. Journal of Meteorological Research 30, 103–116. https://doi.org/10.1007/s13351-016-5052-z.

Li, Y., Li, J., Feng, J., 2012. A teleconnection between the reduction of rainfall in southwest western Australia and North China. J. Clim. 25, 8444–8461. https://doi.org/10.1175/JCLI-D-11-00613.1.

Li, L., Dolman, A.J., Xu, Z., 2016. Atmospheric moisture sources, paths, and the quantitative importance to the eastern Asian monsoon region. J. Hydrometeorol. 17, 637–649. https://doi.org/10.1175/JHM-D-15-0082.1.

Liang, X.Z., Wang, W.C., 1998. Associations between China monsoon rainfall and tropospheric jets. Q. J. R. Meteorol. Soc. 124, 2597–2623. https://doi.org/10.1002/qj.49712455204.

Liu, Y., Pan, Z., Zhuang, Q., Miralles, D., Teuling, A., Zhang, T., An, P., Dong, Z., Zhang, J., He, D., Wang, L., Pan, X., Bai, W., Niyogi, D., 2015. Agriculture intensifies soil moisture decline in northern China. Sci. Rep. 5, 11261. https://doi.org/10.1038/srep11261.

Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the Central United States. Nat. Clim. Chang. 5, 250–254. https://doi.org/10.1038/nclimate2516.

Nan, S., Zhao, P., Chen, J., Liu, G., 2021. Links between the thermal condition of the tibetan plateau in summer and atmospheric circulation and climate anomalies over the eurasian continent. Atmos. Res. 247, 105212.

Ouyang, R., Liu, W., Fu, G., Liu, C., Hu, L., Wang, H., 2014. Linkages between ENSO/PDO signals and precipitation, streamflow in China during the last 100 years. Hydrol. Earth Syst. Sci. 18, 3651–3661. https://doi.org/10.5194/hess-18-3651-2014.

Peixoto, J.P., Oort, A.H., 1992. Physics of Climate. American Institute of Physics.

Pimentel, D., Berger, B., Filiberto, D., Newton, M., Wolfe, B., Karabinakis, E., Clark, S., Poon, E., Abbett, E., Nandagopal, S., 2004. Water resources: Agricultural and environmental issues. BioScience 54, 909–918. https://doi.org/10.1641/0006-3568(2004)054[0909,WRAAEI]2.0.CO;2.

Prein, A.F., Liu, C., Ikeda, K., Trier, S.B., Rasmussen, R.M., Holland, G.J., Clark, M.P., 2017. Increased rainfall volume from future convective storms in the us. Nat. Clim. Chang. 7, 880–884.

Qian, C., Zhou, T., 2014. Multidecadal variability of North China aridity and its relationship to PDO during 1900-2010. J. Clim. 27, 1210–1222. https://doi.org/10.1175/JCLI-D-13-00235.1.

Qiu, J., Gao, Q., Wang, S., Su, Z., 2016. Comparison of temporal trends from multiple soil moisture data sets and precipitation: the implication of irrigation on regional soil moisture trend. Int. J. Appl. Earth Obs. Geoinf. 48, 17–27. https://doi.org/10.1016/j.jag.2015.11.012.

Qu, X., Huang, G., 2012. Impacts of tropical Indian Ocean SST on the meridional displacement of east Asian jet in boreal summer. Int. J. Climatol. 32, 2073–2080. https://doi.org/10.1002/joc.2378.

Read, L.K., Vogel, R.M., 2015. Reliability, return periods, and risk under nonstationarity. Water Resour. Res. 51, 6381–6398. https://doi.org/10.1002/2015WR017089.

von Storch, H., Zwiers, F.W., 1999. Statistical Analysis in Climate Research. Cambridge University Press.

Trenberth, K.E., 1997. The definition of El Niño. Bull. Am. Meteorol. Soc. 78, 2771–2777. https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

Turner, A.G., Annamalai, H., 2012. Climate change and the south Asian summer monsoon. Nat. Clim. Chang. 2, 587–595. https://doi.org/10.1038/nclimate1495.

Ullah, W., Wang, G., Lou, D., Ullah, S., Bhatti, A.S., Ullah, S., Karim, A., Hagan, D.F.T., Ali, G., 2021. Large-scale atmospheric circulation patterns associated with extreme monsoon precipitation in Pakistan during 1981–2018. Atmos. Res. 253, 105489.

Wu, G., Liu, Y., He, B., Bao, Q., Duan, A., Jin, F.F., 2012. Thermal controls on the Asian summer monsoon. Sci. Rep. 2, 404. https://doi.org/10.1038/srep00404.

Yu, R., Wang, B., Zhou, T., 2004. Tropospheric cooling and summer monsoon weakening trend over East Asia. Geophys. Res. Lett. 31, L22212. https://doi.org/10.1029/2004GL021270.

Zahn, R., 2003. Monsoon linkages. Nature. 421, 324–325. https://doi.org/10.1038/421324a.

Zhang, R., Sumi, A., Kimoto, M., 1999. A diagnostic study of the impact of El Niño on the precipitation in China. Adv. Atmos. Sci. 16, 229–241. https://doi.org/10.1007/BF02973084.

Zhao, P., Yang, S., Yu, R., 2010. Long-term changes in rainfall over eastern China and large-scale atmospheric circulation associated with recent global warming. J. Clim. 23, 1544–1562. https://doi.org/10.1175/2009JCLI2660.1.

Zhao-G, G., Huang, G., Wu, R., Tao, W., Gong, H., Qu, X., Hu, K., 2015. A new upper-level circulation index for the east Asian summer monsoon variability. J. Clim. 28, 9977–9996. https://doi.org/10.1175/JCLI-D-15-0272.1.

Zhao-S, S., Li, J., Yu, R., Chen, H., 2015. Recent reversal of the upper-tropospheric temperature trend and its role in intensifying the east Asian summer monsoon. Sci. Rep. 5, 11847. https://doi.org/10.1038/srep11847.

Zhu, Y., Wang, H., Zhou, W., Ma, J., 2011. Recent changes in the summer precipitation pattern in East China and the background circulation. Clim. Dyn. 36, 1463–1473. https://doi.org/10.1007/s00382-010-0852-9.