ARTICLE   **OPEN**

Seasonal predictability of the dominant surface ozone pattern over China linked to sea surface temperature

Yuan Chen[1], Dean Chen[2], Linru Nie[1], Wenqi Liu[1], Jingfang Fan[3], Xiaosong Chen [ID][3] and Yongwen Zhang [ID][1] ✉

Mitigation surface ozone pollution becomes increasingly pivotal in improving China's air quality. However, the impact of global sea surface temperature anomalies (SSTA) on the long-term predictability of China's surface ozone remains challenging. In this study, we employ eigen techniques to effectively characterize dominant surface ozone patterns over China, and establish cross-correlations between the dominant patterns and global SSTA time series. Our findings reveal that China's summer ozone pollution is strongly associated with crucial SSTA clusters linked to atmospheric circulations, i.e., the West Pacific Subtropical High and the Pacific-North American teleconnection pattern. For winter, ozone pollution is attributed to SSTA clusters related to the Southern Oscillation, the Madden-Julian Oscillation and others. We propose a multivariate regression model capable of predicting surface ozone patterns with a lead time of at least 3 months. Evaluation of our model using a testing dataset yields an R-value of around 0.5 between predicted and observed data, surpassing statistical significance threshold. This suggests the viability and potential applicability of our predictive model in surface ozone forecasting and mitigation strategies in China.

npj Climate and Atmospheric Science (2024)7:17 ; https://doi.org/10.1038/s41612-023-00560-7

INTRODUCTION

Ozone in the stratosphere forms the ozone layer, which is essential for absorbing the Sun's ultraviolet radiation and protecting life on Earth. However, Ozone in the lower levels of the troposphere is a kind of air pollutant that has serious damage to human life, such as corroding human lungs, and destroying crops and forest vegetation[1]. Numerous studies indicated that worldwide, several hundreds of thousands of premature deaths annually were associated with ground-level O_3 pollution[2]. Surface O_3, known as a secondary air pollutant, isn't directly emitted by vehicles or industries; rather, it is produced by the reaction of sunlight with nitrogen oxides and hydrocarbons in the air[3,4]. In recent years, it has become a significant pollutant in major urban areas of China[5–7]. Notably, a recent study indicated a substantial rise in O_3 pollution across Northern China. However, the trends of O_3 pollution in several regions of China do not exhibit significant increases compared to the influences of natural and climate variability[8]. As such, predicting surface O_3 concentrations is crucial for mitigating the damage caused by air pollution in China. Notably, long-term predictions are invaluable as they enable governments to strategically plan air pollution control measures months or even a year in advance.

Efforts to improve predictions of surface O_3 concentrations have involved investigating the primary factors influencing them[7]. Not only do emissions of surface O_3 precursors play a significant role[8,9], but large-scale ocean-atmosphere circulations also have substantial impacts on O_3 levels[10]. Connections between large-scale ocean-atmosphere circulations and local air pollution are critical for long-term air pollution forecasting[11]. Previous research has demonstrated that large-scale climatic patterns can be harnessed to predict surface-level O_3 concentrations in the United States up to a season in advance, attributed to the interactions between large-scale ocean-atmospheric circulations and the inherent long-term memory[10]. Additionally, it has been observed that the North Atlantic Oscillation-driven anomalous atmospheric circulations influence O_3 pollution in Europe by modulating the photochemical reactions involved[12,13].

Furthermore, certain meteorological conditions such as drought, elevated temperatures, and intense sunlight are conducive to the formation of surface O_3 pollution. Studies investigating the relationships between these meteorological conditions and broader climatic influences have unveiled the potential for seasonal prediction of surface O_3 concentrations[14]. Springtime warming in the Western Pacific Ocean, Western Indian Ocean, and Ross Sea is related to the interannual shifts in the frequency of simultaneous occurrences of heat waves and O_3 pollution in China during summer[15]. Arctic sea ice has also been identified as an indirect influence on surface O_3 levels in Northern China through its effect on meteorological patterns across Eurasia[16]. Rossby waves, known for triggering anticyclones in Northern China, create an environment characterized by stable air, high temperatures, and low humidity, which consequently traps particulate matter and O_3 pollutants near the surface[17,18]. Utilizing a global three-dimensional Goddard Earth Observing System Chemical transport model (GEOS-Chem), researchers have found that the East Asian summer monsoon plays a significant role in the interannual variability of summer surface O_3 in China[19]. Moreover, the El Niño/Southern Oscillation (ENSO), a paramount inter-annual climate phenomenon, has been linked to variations of the total ozone column amounts by affecting tropopause height[20–22]. Global climate change has also been implicated in the rising trends of surface O_3 concentrations in China[16,23].

However, extant literature primarily concentrates on O_3 pollution in specific regions of China (e.g. north China) affected by several teleconnection patterns or the long-term variability of O_3-related meteorological conditions under climate effects. These studies did not directly address the long-term prediction of O_3 pollution across China. It is essential to recognize that predictions

[1]Data Science Research Center, Faculty of Science, Kunming University of Science and Technology, 650500 Kunming, Yunnan, China. [2]Department of Physics, University of Helsinki, P.O. Box 48, 00014 Helsinki, Finland. [3]School of Systems Science/Institute of Nonequilibrium Systems, Beijing Normal University, 100875 Beijing, China. ✉email: zhangyongwen77@gmail.com

of O_3-related meteorological conditions and O_3 pollution itself are not completely equivalent. Moreover, there are O_3 fluctuation patterns due to the correlation of O_3 concentration between different regions within China[24]. Consequently, the objectives of this study encompass the identification of important fluctuation patterns in surface O_3 concentrations across China, and the examination of their associations with the global SSTA. SSTA has well-documented impacts on atmospheric circulation and further subsequent implications for ozone levels[25]. Furthermore, the long-lasting memory of SSTA, ranging from months to years, harbors potential seasonal prediction power[26,27].

To identify crucial patterns of O_3 fluctuations, eigenanalysis emerges as a potent and efficient technique for decomposing multiple fluctuation patterns into a set of independent principal modes. Common methods include classic principal component analysis (PCA) and empirical orthogonal function (EOF) analysis, recognized as efficient tools in climate studies and various investigations into O_3 pollution[10,24,28,29]. For instance, two dominant patterns of summer O_3 pollution in China have been discerned with the influence of the West Pacific Oceans[30]. Additionally, an advanced Eigen Microstate method, approached from a statistical physics standpoint[31–33], has been employed. Thus, here we focus on predicting the time series of the first dominant pattern, effectively addressing the intricate nature of ozone pollution across various regions in China, including the north and south. This involves extracting the system's eigen or principal modes and examining their temporal evolution.

RESULTS

Spatial and temporal characteristics of the surface O_3 pattern in China

We first perform data preprocessing to derive the detrended surface O_3 and SSTA. In order to investigate the spatiotemporal characteristics of O_3 patterns over China, we employ eigen analysis on the detrended O_3 field for the summer and winter of the training period (June 2013 to February 2018); and obtain eigenvectors and principal modes for both seasons (refer to Methods). We find that the 1st eigenvalues account for 22.8% and 36.8% of the total variance in summer and winter, respectively. The spatial distributions of the corresponding eigenvector u_1 are predominantly positive across China in both seasons, with notable differences in regional concentration (see Fig. 1a and b). During the summer season, the most significant positive components, indicating elevated O_3 concentrations, are observed in the northern regions of China, specifically in Hebei and Shanxi, as illustrated in Supplementary Fig. 1a. This pattern is attributable to heightened precursor emissions[34] and the meteorological phenomenon of a high-pressure anomaly in north China. Consequently, this anomaly induces both high-temperature and low-humidity anomalies, fostering enhanced photochemical reactions, all of which contribute to O_3 production[30].

Conversely, during the winter season, Southern China emerges as the region with the highest O_3 concentration, as depicted in Supplementary Fig. 1b. This phenomenon can be attributed to the long-range transport of O_3 and O_3 precursors within the polluted air masses originating from the northern regions, alongside photochemical formation under dry and sunny weather conditions[7,35]. Additionally, the first principal component (PC1), which corresponds to the temporal evolution of the primary mode, is computed from Eq. (1) of Methods, and is illustrated in Fig. 1c and d. Notably, it displays trends that are akin to the average O_3 concentrations over China during the same period (see Supplementary Fig. 1c, d). In fact, PC1 captures the predominant fluctuation region for O_3 pollution in China, distinct from representing ozone fluctuations across all locations throughout the country.

The 2nd eigenvalues explain 11.3% and 11.7% of the total variance in summer and winter, respectively. The spatial distributions of the 2nd eigenvectors (Supplementary Fig. 2a and b) reveal a distinct boundary between northern and southern clusters, reflecting the influence of diverse weather systems on surface O_3 fluctuations. For example, Northern China is susceptible to Rossby waves, which can trigger heat or cold waves originating from Siberia[17,18], whereas Southern China is influenced by tropical circulations[36].

Global SSTA linked to the surface O_3 pattern in China

To elucidate the relationship between dominant surface O_3 patterns and global climate, we construct a network based on Methods. The cross-correlation between PC1 and the global SSTA time series is computed based on Eq. (2) (refer to Methods). To conduct significance tests, we generate synthetic correlation using two different null models: the White Noise Null Model (WNNM) and the Monthly Autocorrelation-preserving Null Model (MANM). In WNNM, the daily SSTA time series and PC1 is completely shuffled, resulting in a typical white noise time series. On the other hand, MANM preserves the order within a one-month time window, corresponding to the typical timescale of SSTA memory, and then shuffles the order of different time windows. This MANM test maintains the autocorrelation within a month but eliminates the true correlation between the two-time series. Supplementary Table 1 shows the reliability of the MANM test effectively identifying spurious correlations in comparison to the conventional T-test[10,27,37].

Figure 1e, f shows the Probability Density Functions (PDFs) of the absolute correlation $|CC^*_{v_1, s_i}|$ for the observed data (depicted in red), in contrast to WNNM (in green) and MANM (in blue). It is evident that the correlations for MANM are considerably larger than those of WNNM, which can be attributed to the influence of autocorrelation. However, the peaks of the PDFs are distinct between the observed and the null models for both summer and winter, as depicted in Fig. 1e and f. The observed data exhibits significantly larger correlations. To eliminate spurious correlations, we employ the 97.5% significance level of the MANM test as the threshold Δ (refer to Fig. 1e and f). Correlations below this threshold do not establish a link between PC1 and SSTA nodes.

Figure 2a and b illustrate the spatial distributions of significant correlations $CC^*_{v_1, s_i}$ exceeding the threshold across global SSTA for summer and winter, respectively. The regions with positive and negative correlations between PC1 and SSTA nodes are represented in red and blue, respectively. Certain regions exhibit robust correlations in both seasons (above 0.5 or below −0.5). The associated delay times are derived as outlined in the Methods section and are depicted in Fig. 2c and d. Notably, some SSTA regions have delay times exceeding 90 days, suggesting long-term memory behavior between these SSTA regions and surface O_3 in China, and the potential for predicting ozone levels more than one season in advance. To pinpoint critical SSTA regions, we identify the critical SSTA nodes where $|CC^*_{v_1, s_i}| > \Delta$ and $\tau^* > 90$. We select the four largest clusters formed by spatially contiguous critical SSTA nodes, as shown in Fig. 2e and f. These clusters are teleconnected to surface O_3 in China over distances exceeding 3000 km, with the exception of cluster CS3 in Fig. 2e, which is proximate to Eastern China.

Variations in SSTA, exhibiting inertial memory over months to years, can induce anomalies in large-scale atmospheric circulation[38,39]. These atmospheric circulation changes, in turn, could have repercussions on air pollution and surface O_3 levels in China[40,41]. To delve into the underlying physical mechanisms associated with the identified SSTA clusters, we present the most relevant atmospheric circulation indexes corresponding to the identified SSTA clusters in Fig. 2e and f. Specifically, the summer clusters CS1, CS2, CS3, and CS4 are linked to the Walker
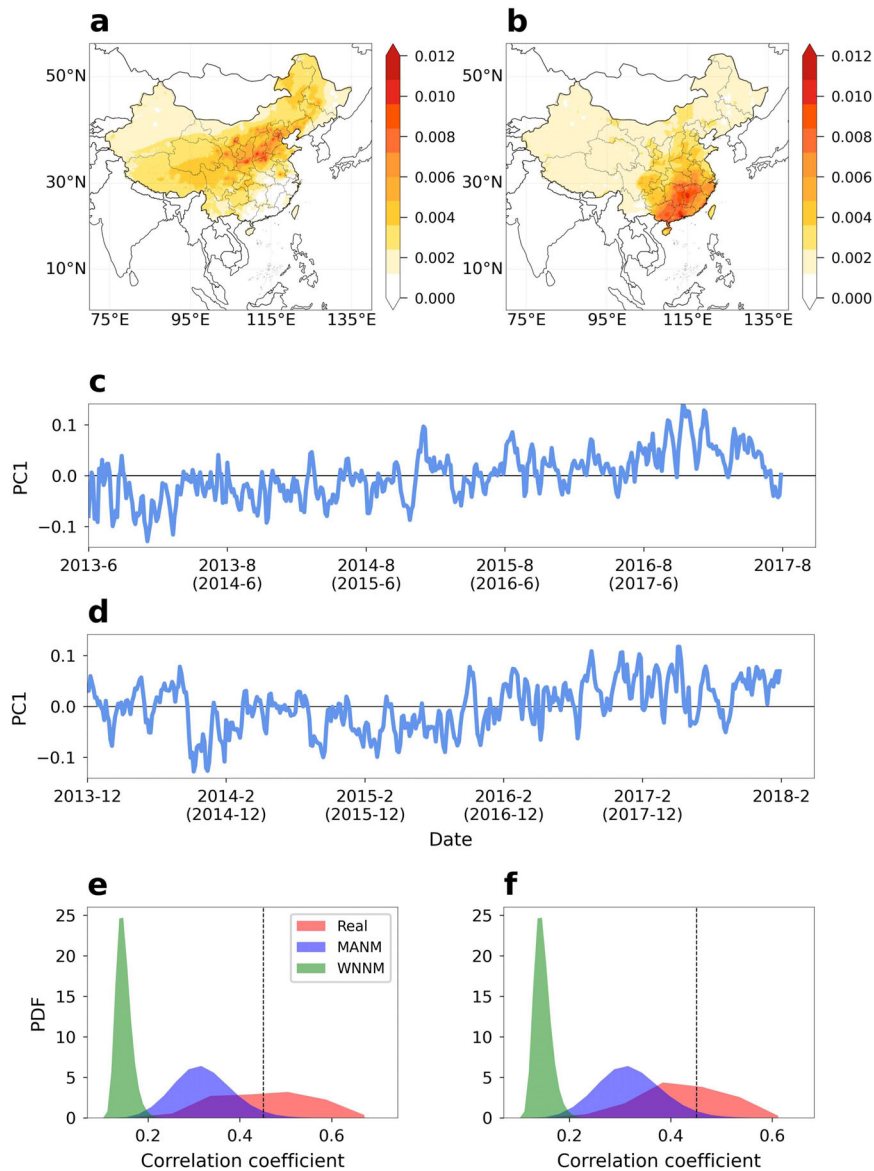
Fig. 1 Spatial and temporal characteristics of the surface O₃ pattern in China and significance tests. Spatial distribution of the 1st eigenvector of the O_3 principal mode in **a** summertime and **b** wintertime, respectively. Time series of PC1 in **c** summertime and **d** wintertime, respectively. PDFs of the absolute correlations $|CC^*_{v_1,s_i}|$ between PC1 and SSTA nodes for real data (red area) and the synthetic data for WNNM (green) and MANM (blue) for **e** summertime and **f** wintertime, respectively. Black dashed line represents 97.5th percentile of the synthetic correlations of MANM.

circulation (Niño 1+2)[42], the North Pacific High (NPH)[43], the West Pacific Subtropical High (WPSH)[44], and the Pacific-North American teleconnection pattern (PNA)[45], respectively (see Fig. 3a). The significant correlation coefficients above 0.5 between the atmospheric circulation indexes and the corresponding SSTA clusters are shown in Table 1. Furthermore, these atmospheric circulation indexes exhibit significant correlations with PC1 of surface O_3 pollution over months, as well as with the SSTA clusters (see Table 1).

Supplementary Figure 3 illustrates the time-delayed average atmospheric field anomalies following strong or weak atmospheric circulations. In Supplementary Figure 3a, the weak Walker circulation induces a high-pressure anomaly over the Western Pacific Ocean and Eastern Asia. Similarly, the weak NPH, the strong WPSH, and the strong PNA also lead to high-pressure anomalies in north China (see Supplementary Fig. 3c, e, g) associated with lower humidity, elevated temperatures, reduced

cloud cover, and intensified solar radiation. These conditions create favorable environments for photochemical reactions that generate surface O_3[16,20–23,46,47]. Conversely, the opposite atmospheric field anomalies are associated with low-pressure anomalies and low levels of surface O_3 pollution in Supplementary Fig. 3b, d, f, h.

In winter, clusters CW1, CW2, CW3, and CW4 are associated with the Southern Oscillation Indices (SOI)[48], the North Atlantic Oscillation (NAO)[49], the Amundsen Sea Low (ASL)[50], and the Madden-Julian Oscillation (MJO)[51], as depicted in Fig. 3b. The correlation coefficients between them are presented in Table 1. During winter, Southern China emerges as a dominant contributor to surface O_3 pollution, indicating mechanisms distinct from those in summer. SOI is negatively correlated with cluster CS1 and PC1 (see Table 1). Following the negative phase of SOI, East Asia experiences a strong north-to-south wind anomaly transporting O_3 pollution and O_3 precursors from northern regions (see
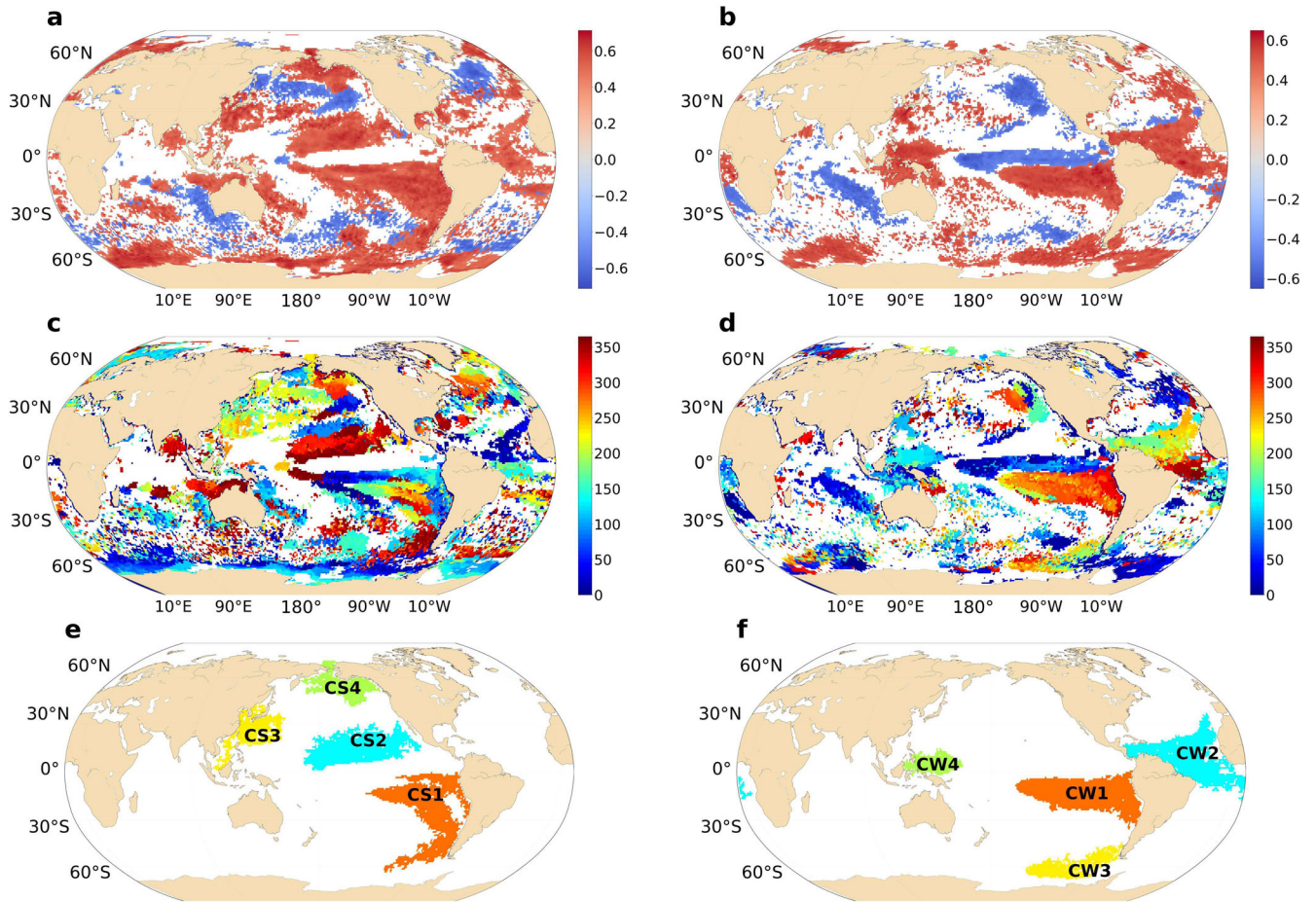
Fig. 2 Global SSTA linked to the surface O₃ pattern in China. Distributions of significant correlation coefficients over global SSTA nodes linked to PC1 of surface O₃ in China for the (**a**) summertime and **b** wintertime, respectively. Distributions of the corresponding delay times to the correlation coefficients for the (**c**) summertime and **d** wintertime. Four clusters of SSTA connected to the surface O₃ with the largest areas and long-term delay times more than 90 days for the (**e**) summertime and **f** wintertime, respectively.
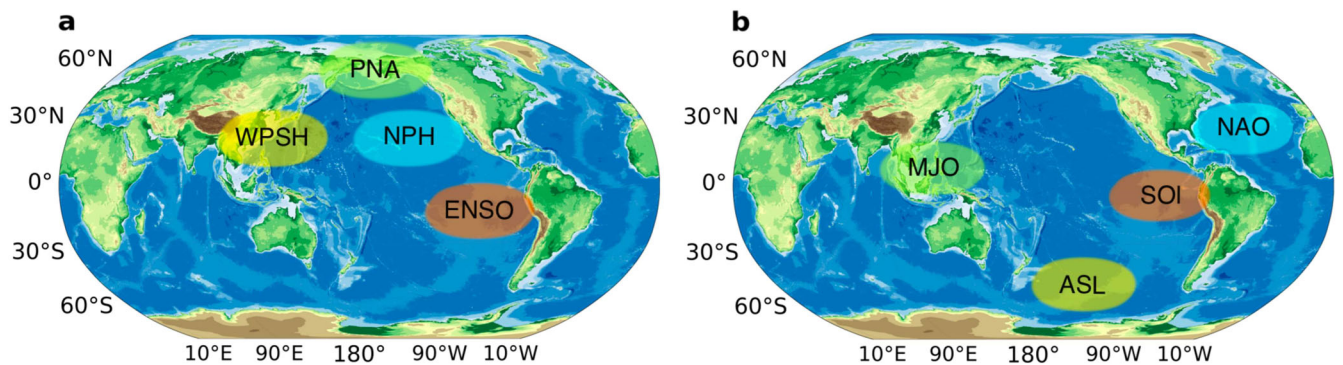
Fig. 3 Atmospheric circulations related to SSTA clusters. Sketch maps of four atmospheric circulations corresponding to four SSTA clusters, influencing PC1 of surface O₃ pollution in China during (**a**) summer and **b** winter.

Supplementary Fig. 4a). During the negative phase of NAO, ASL and MJO exhibit atmospheric fields with a similar wind pattern facilitating the transport of O₃ pollution (Supplementary Fig. 4c, e, g). Conversely, for the opposite phases, the atmospheric fields are depicted in Supplementary Fig. 4b, d, f, h. It is noteworthy that these atmospheric circulation indexes are interdependent; for instance, ASL can influence MJO and SOI, subsequently affecting O₃ pollution in China[52–55].

Seasonal prediction of the surface O₃ pattern in China

To assess the predictive power of these SSTA clusters for surface ozone in China, we employ a multiple linear regression model. The model, based on a combination of these clusters, predicts the time series of the first principal component (PC1) with a fifteen-day moving average for summertime and wintertime surface O₃ in China beyond 90 days. Using a training dataset from June 2013 to February 2018, the model is obtained by Eq. (4) (see Methods). The

Table 1. Relationships between the atmospheric circulation indexes and the corresponding SSTA clusters and PC1 of surface O_3 pollution in months.

		Niño 1+2	NPH	WPSH	PNA
	Atmospheric Circulation Index	CS1	CS2	CS3	CS4
	Corresponding SSTA cluster				
Summer	Lagged correlation between the circulation index and PC1	0.74 ± 0.18	-0.65 ± 0.21	0.79 ± 0.16	0.71 ± 0.19
	Lagged correlation between the circulation index and SSTA	0.72 ± 0.18	-0.59 ± 0.22	0.79 ± 0.16	0.75 ± 0.18
		SOI	NAO	ASL	MJO
	Atmospheric Circulation Index	CW1	CW2	CW3	CW4
	Corresponding SSTA cluster				
Winter	Lagged correlation between the circulation index and PC1	-0.63 ± 0.21	-0.42 ± 0.25	-0.53 ± 0.23	-0.58 ± 0.22
	Lagged correlation between the circulation index and SSTA	-0.66 ± 0.2	-0.46 ± 0.24	-0.64 ± 0.21	-0.54 ± 0.23

R-value R_0 between predicted and observed data are 0.89 and 0.81 for summer and winter, respectively (Fig. 4a, b). For the testing dataset from June 2018 to February 2023, the R-value R_1 is lower than those of the training dataset but remains around 0.5 for both seasons (Fig. 4a, b). To evaluate the significance of the model's predictive performance, we conduct the MANM test by shuffling the time series of SSTA nodes within each cluster. The significance level of R_1 for the prediction model exceeds the 97.5th percentile compared to the null model (Fig. 4c, d). Supplementary Fig. 5 presents the absolute correlation coefficients between each SSTA cluster and PC1 for each year within the testing periods, which reveal variations in predictive power across different years and for different SSTA clusters.

DISCUSSION

In this study, we explored the realm of long-term surface ozone (O_3) pollution prediction in China. Initially, we employed eigen techniques to extract dominant principal modes characterizing China's summer and winter O_3 pollution patterns. Notably, the spatial distribution corresponding to the first mode exhibits consistent patterns, with critical regions suffering high levels of O_3 pollution dominating this mode, a distinction apparent between summer and winter. Conversely, the spatial distribution associated with the second eigenvalue reveals a distinct boundary between northern and southern clusters, driven by differing meteorological conditions in Northern and Southern China.

Furthermore, we calculated cross-correlations with time delays between PC1 and the SSTA time series. This analysis highlights four crucial SSTA clusters that significantly influence PC1 of O_3 pollution. Our findings reveal that summer O_3 pollution is linked to the Walker circulation, the North Pacific High, the West Pacific Subtropical High, and the Pacific-North American teleconnection pattern corresponding to the four SSTA clusters. These atmospheric circulations with anomalies can create favorable environments for photochemical reactions that generate surface O_3 pollution. Winter O_3 pollution is associated with the Southern Oscillation, the North Atlantic Oscillation, the Amundsen Sea Low, and the Madden-Julian Oscillation.

To enhance predictive capabilities, we proposed a statistical model to forecast the first principal component of O_3 pollution in China for both summer and winter seasons. This model is based on PC1 and its association with the states of the identified SSTA clusters, with a lead time of at least 3 months. With the training dataset, our model demonstrated high prediction accuracy, achieving R-values of 0.89 and 0.81 for summer and winter, respectively. In the testing dataset, the R-values remains close to 0.5 for both seasons. The performance of our prediction model indicates its proficiency in capturing general trends in the time series, despite its limitations in predicting short-term fluctuations.

The findings from our study can assist communities in anticipating climate conditions affecting ozone pollution one season in advance, allowing for the implementation of emission control measures to minimize the adverse impacts of deteriorating air quality due to unfavorable climate conditions.

METHODS

Data

The study utilizes the maximum daily 8 h average (MDA8) surface O_3 concentration dataset (version 2) over China from June 2013 to February 2023. The dataset, with a grid resolution of $0.1° \times 0.1°$, is sourced from the Tracking Air Pollution in China (TAP). This dataset merges ground measurements, satellite retrievals, chemical transport model, land-use information with machine learning models and meteorology fields, ensuring its reliability[56,57]. For each grid j, a time series $\{x_j(t)\}$ of O_3 concentration is available. This study focuses on summertime (June-July-August) and wintertime (December-January-February). The global daily average Sea Surface Temperature (SST) data from the ERA5 Reanalysis[58], with a resolution of $0.25° \times 0.25°$ from 1991 to 2023.

Data detrending

The O_3 concentration dataset is divided into two periods: (1) June 2013 to February 2018, serving as the training period for model fitting; (2) June 2018 to February 2023, serving as the testing period to assess the model's reliability. The mean seasonal cycle (calculated based on the training period) for each calendar day is subtracted from the O_3 data for both periods. For SST, the data from 1991 to 2023 is divided into three parts (1991–2000, 2001–2010 and 2011–2023) to obtain the SSTA. The first two parts are based on the entire decade data to remove the seasonal cycle, while the last part is based on the data of the first 7 years (excluding the year corresponding to the test set).

Principal modes of O_3 fluctuation

The detrended O_3 field over China, comprising N grids and a time length of M, can be represented as a matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t, \ldots, \mathbf{y}_M)^T$, where $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{tN})$ represents the state of system at time t[28]. The covariance matrix $\mathbf{C} = \mathbf{Y}^T\mathbf{Y}/M$ is constructed, and its eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N \geq 0$ and eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$ are obtained by solving the eigen equation $\mathbf{Cu} = \lambda\mathbf{u}$. All eigenvectors are normalized and orthogonal to each other. The projection of the fluctuation field $\mathbf{Y}$ onto the n-th eigenvector $\mathbf{Y}$ onto the n-th eigenvector $\mathbf{u}_n = (\mathbf{u}_{n1}, \mathbf{u}_{n2}, \ldots, \mathbf{u}_{nN})^T$ results in the n-th principal component:

$$\mathbf{v}_n = \mathbf{Y}\mathbf{u}_n, \tag{1}$$

whose elements $v_{tn}, t = 1, 2, \ldots, M$ (it also can denote as a time series $\{v_n(t)\}$). The contributions of principal components to the system's evolution are sorted in decreasing order, with the first principal component being the most crucial.
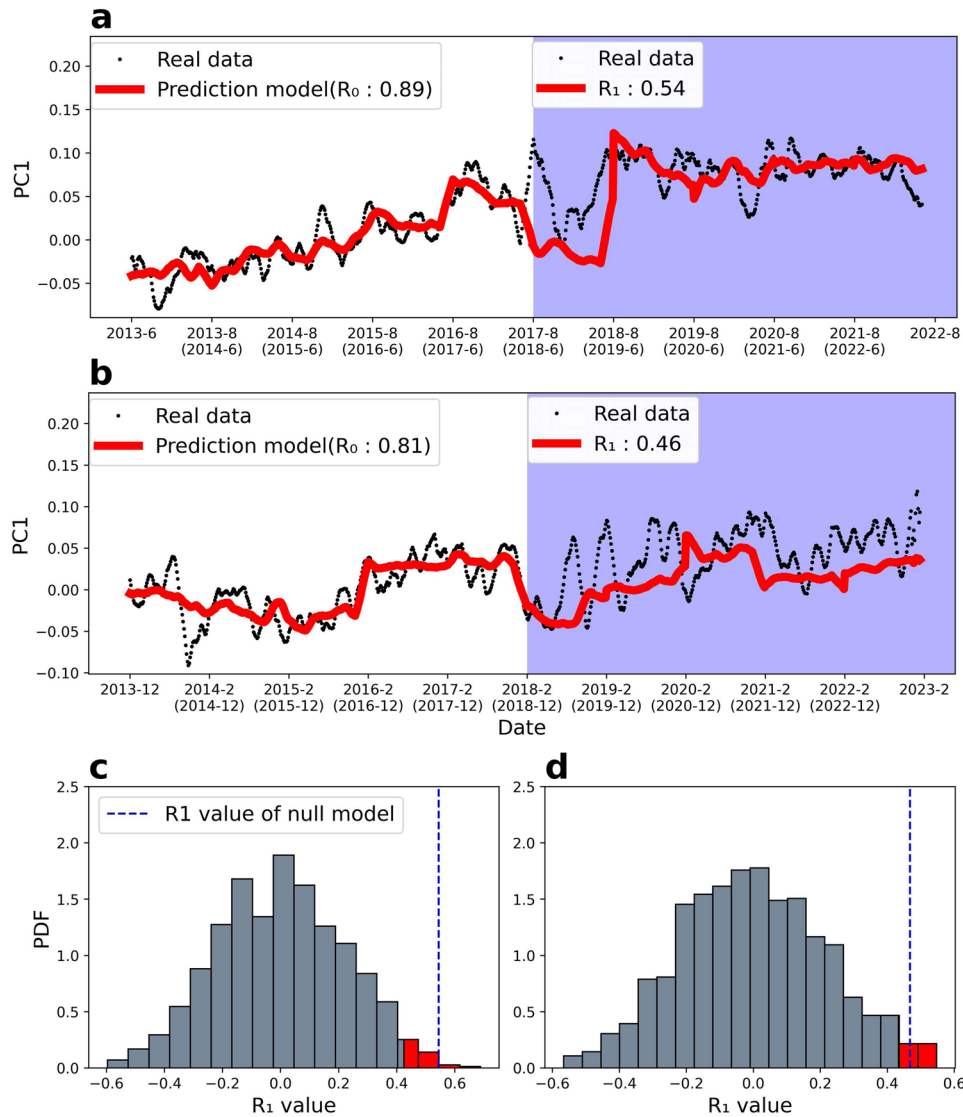
Fig. 4 Seasonal prediction of the surface O_3 pattern in China. Prediction of time series of PC1 (with the fifteen-days moving average) for the surface O_3 in China based on the multiple linear regression modeling for **a** summertime and **b** wintertime, respectively. The blue shadow area shows the prediction for the testing dataset. Black dotted line represents the real PC1 time series with the 15-day moving average. Significance test of R-value for the prediction of the testing dataset for **c** summertime and **d** wintertime, respectively. The red area is above the 97.5th percentile.

Network constructing

In our network, the nodes are defined as the principal components of O_3 fluctuation (with a primary focus on the first mode) and SSTA grids[59]. The links, on the other hand, represent the connections between them. For a given SSTA grid i, the SSTA time series is denoted as $\{s_i(t)\}$. To quantify the relations between the PC1 of O_3 fluctuation v_1 and the SSTA time series, we calculate the cross-correlation function[60–63]:

$$CC_{v_1,s_i}(\tau) = \frac{\langle (v_1(t) - \overline{v_1}) \cdot (s_i(t-\tau) - \overline{s_i}) \rangle}{\sqrt{\langle [v_1(t) - \overline{v_1}]^2 \rangle} \cdot \sqrt{\langle [s_i(t-\tau) - \overline{s_i}]^2 \rangle}}, \quad (2)$$

where $\overline{v_1}$ and $\overline{s_i}$ represent the averages over the periods; $0 \leq \tau \leq 365$ is the delay time. When considering either summertime or wintertime, the variable t spans intermittent periods of summer and winter across various years. Specifically, for summertime, the range is $365(yy-1) + 152 \leq t \leq 365(yy-1) + 243$, for wintertime, it is $365(yy-1) + 335 \leq t \leq 365yy + 59$, where yy represents each

respective year index from the initial value 1. We determine the maximum absolute value of the cross-correlation as CC_{v_1,s_i}^*, and record the corresponding delay time as τ^*. When $\tau^* > 90$, the SSTA grid can provide the potential prediction power more than one season in advance. Thus, a network link is defined when $|CC_{v_1,s_i}^*| > \Delta$ and $\tau^* > 90$; otherwise, there is no link between v_1 and s_i. Δ is the threshold to select the significant correlation, which we set as the 97.5% significant level of the MANM test (see Fig. 1e and f).

Multiple linear regression model

In this study, we construct a multivariate linear model by combining correlated clusters identified in previous steps. For SSTA cluster C_j, the time series of the weighted averages of SSTA is denoted as:

$$WS_j(t) = \frac{\sum_{i \in C_j} s_i(t) \cdot |CC_{v_1,s_i}^*|}{\sum_{i \in C_j} |CC_{v_1,s_i}^*|}. \quad (3)$$

The multiple linear regression model is then constructed as follows:

$$v'_1(t) = \sum_{j=1}^{N_c} a_j \cdot WS_j(t - \tau_j^*) + b, \qquad (4)$$

where $v'_1(t)$ represents the predicted PC1, N_c is the number of selected clusters, and τ_j^* is the delay time exceeding 90 days between PC1 and WS_j, identified based on Eq. (2) for the training set. The parameters a_j ($j = 1, \ldots, N_c$) and b are fitted using five-year time window data and remain fixed for predicting PC1 in the following year. As the prediction extends to subsequent years, the parameters are updated based on the past five years.

DATA AVAILABILITY

The meteorological data are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF), accessible at https://cds.climate.copernicus.eu/#!/home. The China Surface O_3 Concentration Dataset (Version 2) can be retrieved from http://tapdata.org.cn/?page_id=129. Derived data supporting the results of this research are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The source codes for the analysis of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Liu, H. et al. Ground-level ozone pollution and its health impacts in China. *Atmos. Environ.* **173**, 223–230 (2018).
2. Zhang, J., Wei, Y. & Fang, Z. Ozone pollution: a major health hazard worldwide. *Front. Immunol.* **10**, 2518 (2019).
3. Lippmann, M. Health effects of tropospheric ozone. *Environ. Sci. Technol.* **25**, 1954–1962 (1991).
4. Silva, R. A. et al. Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change. *Environ. Res. Lett.* **8**, 034005 (2013).
5. Verstraeten, W. W. et al. Rapid increases in tropospheric ozone production and export from China. *Nat. Geosci.* **8**, 690–695 (2015).
6. Wang, Y. et al. Sensitivity of surface ozone over China to 2000–2050 global changes of climate and emissions. *Atmos. Environ.* **75**, 374–382 (2013).
7. Wang, T. et al. Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* **575**, 1582–1596 (2017).
8. Yu, P., Zhang, Y., Meng, J. & Liu, W. Statistical significance of PM$_{2.5}$ and O$_3$ trends in China under long-term memory effects. *Sci. Total Environ.* **892**, 164598 (2023).
9. Bloomer, B. J., Stehr, J. W., Piety, C. A., Salawitch, R. J. & Dickerson, R. R. Observed relationships of ozone air pollution with temperature and emissions. *Geophys. Res. Lett.* **36**, L09803 (2009).
10. Shen, L. & Mickley, L. J. Seasonal prediction of US summertime ozone using statistical analysis of large scale climate patterns. *Proc. Natl. Acad. Sci.* **114**, 2491–2496 (2017).
11. Yin, Z., Wang, H., Liao, H., Fan, K. & Zhou, B. Seasonal to interannual prediction of air pollution in China: review and insight. *Atmos. Oceanic Sci. Lett.* **15**, 100131 (2022).
12. Christoudias, T., Pozzer, A. & Lelieveld, J. Influence of the North Atlantic Oscillation on air pollution transport. *Atmos. Chem. Phys.* **12**, 869–877 (2012).
13. Pausata, F. S. R., Pozzoli, L. & North Atlantic Oscillation and tropospheric ozone variability in Europe: model analysis and measurements intercomparison. *Atmos. Chem. Phys.* **12**, 6357–6376 (2012).
14. Yin, Z., Li, Y. & Cao, B. Seasonal prediction of surface O_3-related meteorological conditions in summer in north China. *Atmos. Res.* **246**, 105110 (2020).
15. Gao, M. et al. Large-scale climate patterns offer preseasonal hints on the co-occurrence of heat wave and O_3 pollution in China. *Proc. Natl. Acad. Sci.* **120**, e2218274120 (2023).
16. Yin, Z. et al. Links of climate variability in Arctic sea ice, Eurasian teleconnection pattern and summer surface ozone pollution in north China. *Atmos. Chem. Phys.* **19**, 3857–3871 (2019).
17. Zhang, Y., Fan, J., Chen, X., Ashkenazy, Y. & Havlin, S. Significant impact of Rossby waves on air pollution detected by network analysis. *Geophys. Res. Lett.* **46**, 12476–12485 (2019).
18. Dong, Y. et al. The impact of synoptic patterns on summertime ozone pollution in the north China plain. *Sci. Total Environ.* **735**, 139559 (2020).
19. Zhang, L., Liao, H. & Li, J. Impacts of Asian summer monsoon on seasonal and interannual variations of aerosols over Eastern China. *J. Geophys. Res. Atmos.* **115**, D00K05 (2010).
20. Shaman, J. & Tziperman, E. Summertime ENSO-North African-Asian jet teleconnection and implications for the Indian monsoons. *Geophys. Res. Lett.* **34**, L11702 (2007).
21. Zhang, J., Tian, W., Wang, Z., Xie, F. & Wang, F. The influence of ENSO on northern midlatitude ozone during the winter to spring transition. *J. Clim.* **28**, 4774–4793 (2015).
22. Zhang, J. et al. Influence of the El Niño Southern Oscillation on the total ozone column and clear-sky ultraviolet radiation over China. *Atmos. Environ.* **120**, 205–216 (2015).
23. Jiang, Z. et al. Impact of Western Pacific Subtropical High on ozone pollution over Eastern China. *Atmos. Chem. Phys.* **21**, 2601–2613 (2021).
24. Yin, Z., Cao, B. & Wang, H. Dominant patterns of summer ozone pollution in Eastern China and associated atmospheric circulations. *Atmos. Chem. Phys.* **19**, 13933–13943 (2019).
25. Deser, C. & Phillips, A. S. Atmospheric circulation trends, 1950–2000: The relative roles of sea surface temperature forcing and direct atmospheric radiative forcing. *J. Clim.* **22**, 396–413 (2009).
26. Monetti, R. A., Havlin, S. & Bunde, A. Long-term persistence in the sea surface temperature fluctuations. *Physica A* **320**, 581–589 (2003).
27. Gao, M. et al. Seasonal prediction of Indian wintertime aerosol pollution using the ocean memory effect. *Sci. Adv.* **5**, eaav4157 (2019).
28. Hannachi, A., Jolliffe, I. T. & Stephenson, D. B. Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.* **27**, 1119–1152 (2007).
29. Zhang, Y., Fan, J., Li, X., Liu, W. & Chen, X. Evolution mechanism of principal modes in climate dynamics. *New J. Phys.* **22**, 093077 (2020).
30. Zhang, Y. et al. Evolution of summer surface ozone pollution patterns in China during 2015–2020. *Atmos. Res.* **291**, 106836 (2023).
31. Hu, G., Liu, T., Liu, M., Chen, W. & Chen, X. Condensation of eigen microstate in statistical ensemble and phase transition. *Sci. China Phys. Mech.* **62**, 1–8 (2019).
32. Sun, Y. et al. Eigen microstates and their evolutions in complex systems. *Commun. Theor. Phys.* **73**, 065603 (2021).
33. Chen, X. et al. Eigen microstates and their evolution of global ozone at different geopotential heights. *Chaos* **31**, 071102 (2021).
34. Li, M. et al. Persistent growth of anthropogenic non-methane volatile organic compound (NMVOC) emissions in China during 1990–2017: drivers, speciation and ozone formation potential. *Atmos. Chem. Phys.* **19**, 8897–8913 (2019).
35. Zheng, J., Zhong, L., Wang, T., Louie, P. K. K. & Li, Z. Ground-level ozone in the Pearl River Delta region: Analysis of data from a recently established regional air quality monitoring network. *Atmos. Environ.* **44**, 814–823 (2010).
36. Lianshou, C., Huibang, L., Yihong, D. & Hui, Y. An overview of tropical cyclone and tropical meteorology research progress. *Adv. Atmos. Sci.* **21**, 505–514 (2004).
37. Student. The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
38. Mo, K. C. & Thiaw, W. M. Ensemble canonical correlation prediction of precipitation over the sahel. *Geophys. Res. Lett.* **29**, 11–1 (2002).
39. Ndiaye, O., Ward, M. N. & Thiaw, W. M. Predictability of seasonal Sahel rainfall using GCMS and lead-time improvements through the use of a coupled model. *J. Clim.* **24**, 1931–1949 (2011).
40. Zhai, S. et al. Fine particulate matter (PM$_{2.5}$) trends in China, 2013–2018: Separating contributions from anthropogenic emissions and meteorology. *Atmos. Chem. Phys.* **19**, 11031–11041 (2019).
41. Yin, Z. & Ma, X. Meteorological conditions contributed to changes in dominant patterns of summer ozone pollution in Eastern China. *Environ. Res. Lett.* **15**, 124062 (2020).
42. Trenberth, K. E. & Stepaniak, D. P. Indices of El Niño evolution. *J. Clim.* **14**, 1697–1701 (2001).
43. Schroeder, I. D. et al. The North Pacific High and wintertime pre-conditioning of California current productivity. *Geophys. Res. Lett.* **40**, 541–546 (2013).
44. Liu, Y., Li, W., Ai, W. & Li, Q. Reconstruction and application of the monthly Western Pacific Subtropical High indices. *J. Appl. Meteor. Sci.* **23**, 414–423 (2012).
45. Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.* **115**, 1083–1126 (1987).

46. Zhou, B. & Xia, D. Interdecadal change of the connection between winter North Pacific Oscillation and summer precipitation in the Huaihe river valley. *Sci. China Earth Sci.* **55**, 2049–2057 (2012).

47. Yang, X. Y., Wang, G. & Keenlyside, N. The Arctic sea ice extent change connected to Pacific decadal variability. *Cryosphere* **14**, 693–708 (2020).

48. Trenberth, K. E. Signal versus noise in the Southern Oscillation. *Mon. Wea. Rev.* **112**, 326–332 (1984).

49. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).

50. Hosking, J. S., Orr, A., Bracegirdle, T. J. & Turner, J. Future circulation changes off West Antarctica: Sensitivity of the Amundsen Sea Low to projected anthropogenic forcing. *Geophys. Res. Lett.* **43**, 367–376 (2016).

51. Kiladis, G. N. et al. A comparison of OLR and circulation-based indices for tracking the MJO. *Mon. Wea. Rev.* **142**, 1697–1715 (2014).

52. Liu, Y., Hu, Z. Z. & Wu, R. Cooperative effects of tropical Pacific and Atlantic SST forcing in southern China winter precipitation variability. *Clim. Dyn.* **55**, 2903–2919 (2020).

53. Zhou, Q., Chen, W. & Zhou, W. Solar cycle modulation of the ENSO impact on the winter climate of East Asia. *J. Geophys. Res. Atmos.* **118**, 5111–5119 (2013).

54. Jung, M. I., Son, S. W., Kim, H. & Chen, D. Tropical modulation of East Asia air pollution. *Nat. Commun.* **13**, 5580 (2022).

55. Zhang, Z. et al. Possible influence of the Antarctic Oscillation on haze pollution in north China. *J. Geophys. Res. Atmos.* **124**, 1307–1321 (2019).

56. Xiao, Q. et al. Tracking PM$_{2.5}$ and O$_3$ pollution and the related health burden in China 2013–2020. *Environ. Sci. Technol.* **56**, 6922–6932 (2021).

57. Xue, T. et al. Estimating spatiotemporal variation in ambient ozone exposure during 2013–2017 using a data-fusion model. *Environ. Sci. Technol.* **54**, 14877–14888 (2020).

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

59. Tsonis, A. A., Swanson, K. L. & Roebber, P. J. What do networks have to do with climate? *Bull. Am. Meteorol. Soc.* **87**, 585–596 (2006).

60. Yamasaki, K., Gozolchiani, A. & Havlin, S. Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.* **100**, 228501 (2008).

61. Wang, Y. et al. Dominant imprint of Rossby waves in the climate network. *Phys. Rev. Lett.* **111**, 138501 (2013).

62. Wang, Y., Gozolchiani, A., Ashkenazy, Y. & Havlin, S. Oceanic El-Niño wave dynamics and climate networks. *New J. Phys.* **18**, 033021 (2016).

63. Berezin, Y., Gozolchiani, A., Guez, O. & Havlin, S. Stability of climate networks with time. *Sci. Rep.* **2**, 666 (2012).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Y.C.: Investigation, Visualization, Analysis, Writing-Origional draft, Reviewing, Editing. D.C.: Methodology, Writing, Reviewing, Editing. L.N.: Methodology, Writing, Reviewing, Editing. W.L.: Methodology, Writing, Reviewing, Editing. J.F.: Methodology, Writing, Reviewing, Editing. X.C.: Methodology, Writing, Reviewing, Editing. Y.Z.: Investigation, Conceptualization, Analysis, Methodology, Writing, Reviewing, Editing, Supervision.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00560-7.

Correspondence and requests for materials should be addressed to Yongwen Zhang.

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

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