Supporting Information for

Chinese Blue Days: A novel index and spatio-temporal variations

Su Wang[1,2], Gang Huang[1,2,3,4], Jintai Lin[5], Kaiming Hu[1,6], Lin Wang[7],
Hainan Gong[1,6]

*1 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and
Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of
Sciences, Beijing 100029, China*

2 University of Chinese Academy of Sciences, Beijing 100049, China

*3 Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National
Laboratory for Marine Science and Technology，Qingdao 266237，China*

4 Joint Center for Global Change Studies (JCGCS), Beijing 100875, China

*5 Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric
and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China*

*6 Center for Monsoon System Research, Institute of Atmospheric Physics, Chinese
Academy of Sciences, Beijing, China*

*7 Key Laboratory of Regional Climate-Environment for Temperate East Asia,
Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029,
China*

1. Quantity controls

To improve the accuracy and continuity of data we used, a series of strict quality control (HUSAR et al 2000, Lin et al 2014, Lin et al 2014, Guo et al 2017) were implemented.

Selection criteria for meteorological data included the following: (1) a month was considered to be sufficient if there were 6 or fewer missing days within the month; (2) a year was considered to be complete if all months were sufficient according to (1); (3) a station was eliminated when the variation in longitude or latitude extended beyond 50 km, or the variation in height above sea level extended beyond 100 m; (4) to get continuous records, the average of adjacent stations within a 25 km radius was used to

replace missing data.

Visibility data are selected as follows: (1) only visibility data at 14:00, Beijing time are selected; (2) a station should be removed if the median value passes 11 km but the maximum does not exceed 12 km; (3) a station should be removed if there are fewer than 5 distinct values; (4) a station should be removed if the missing data exceed 1% of the total data; (5) to get continuous records, the average of adjacent stations within a25 km radius was used to replace missing data.

1. Daily weather phenomena records used

Lacking weather phenomena records from the National Meteorological Information Center, we choose data recorded by weather websites, their addresses are as following:

(1) http://www.tianqihoubao.com/lishi/index.htm,

(2) http://tianqi.2345.cn/wea_history/71141.htm,

(3) https://www.baidutianqi.com/history/,

(4) http://www.tianqishi.com/lishi/.

After comparing data from 4 different websites, we find consistent records results, proving the reliability to some extent.

2. Methodology

The nonparametric Mann-Kendall (MK) test has been widely used to assess abrupt change. Previous studies find the existence of a positive serial correlation will increase the possibility of rejecting the null hypothesis while a negative serial correlation will reject (Yue et al 2002, Yue et al 2002, Novotny et al 2007, Villarini et al 2011, Ouarda et al 2019). Therefore, to limit the influence of serial correlation on the MK test, a popularly revised Mann-Kendall test has been used (Yue et al 2004, Zhang et al 2006, Ouarda et al 2019), the main steps are shown as the following:

(1) First, the serial correlation analysis method in Haan (2002) is used, the equation is:

$$\rho_k = \frac{Cov(x_t, x_{t+k})}{Var(x_t)} = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(x_t - \bar{x}_t)(x_{t+k} - \bar{x}_t)}{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x}_t)^2} \qquad (1)$$

where X_t (t=1, 2, ...) is the tested time series, X_{t+k} is the same time series with a time lag of k, and $\bar{X}_t$ is the mean of the time series. If ρ_k falls between the 95% confidence level lines, it means the tested series are independent at the 95% confidence level.

The average annual CBDI time series were tested and the result is shown below in Fig. S1. We notice that almost all series are statistically independent, except data at lag2, where we can see an obvious positive correlation.
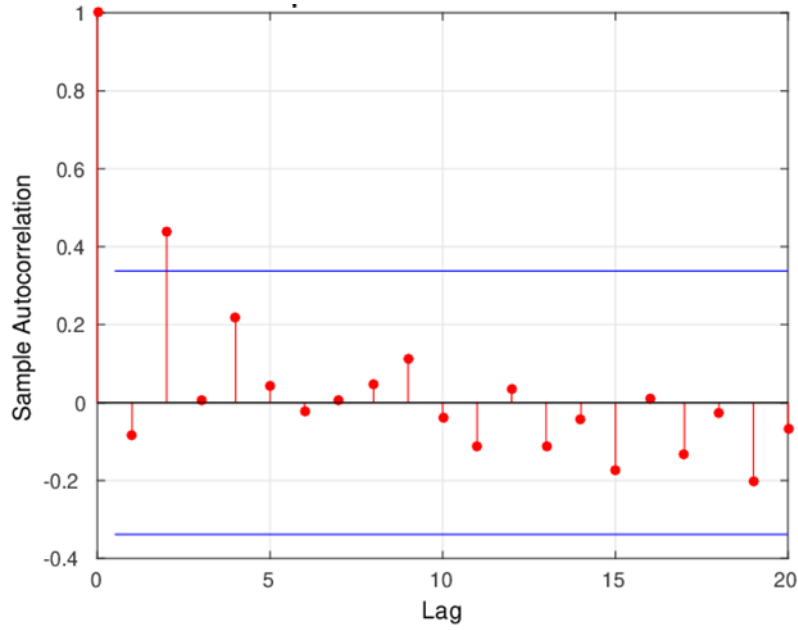

Figure 1. The first-order autocorrelations of average annual CBDI during 1980-2014. The 95% confidence bounds are plotted in blue, a number outside the blue lines indicates a significant trend.

(2) The effective sample size (n^*) has been proposed to modify the variance $V(S)$ to $V^*(S)$(Bayley et al 1946)

$$n^* = \frac{n}{1+2\cdot\sum_{k=1}^{n-k}(1-\frac{k}{n})\cdot\rho_k} \qquad (2)$$

$$V^*(S) = V(S)\cdot\frac{n}{n^*} \qquad (3)$$

where V(S) is the original MK variance and n is the actual size of the sample.

(3) Then, the $V^*(S)$ has been put into the traditional MK test. The final result indicates that no significant trend exists in the time series (not shown), therefore the result we used was sufficient to effectively detect the effect of the serial

correlation on the MK test.

3. Summary Statistics for Annual and Seasonally Averaged CBID

Table S1. Summary Statistics for Annual and Seasonally Averaged CBID

	annual	spring	summer	autumn	winter
mean	161.1	39.3	29	45.5	46.1
Trend(days/decade)	1.6*	1.4	0.3	-0.3	0.2
R^2	0.12	0.38	0.036	0.02	0.005
SD	4.7	2.4	1.8	1.9	2.38

SD, standard deviation; the trends are for 1980– 2014

R^2, R-square coefficient; *, pass the 95% significance of the level

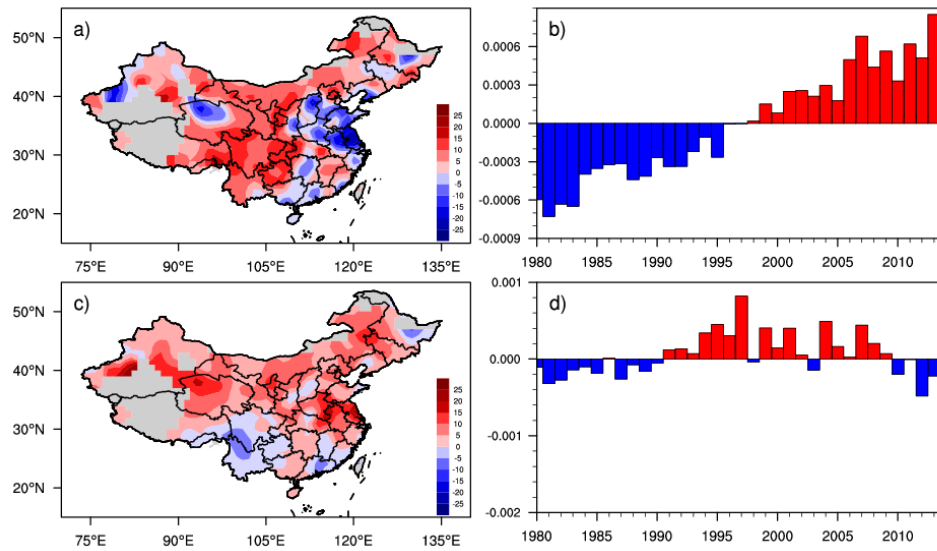## 4. EOF pattern of the CBDI during 1980-2014

Figure S2. The spatial pattern of the EOF1 (a), EOF2 (c) mode of the CBDI, and the time series of the first principal (PC1) (c) and PC2 (d) component over China from 1980 to 2014.

Reference

Bayley, G. V.，J. M. Hammersley 1946 THE EFFECTIVE NUMBER OF INDEPENDENT OBSERVATIONS IN AN AUTOCORRELATED TIME SERIES Journal of the Royal Statistical Society Series B-Statistical Methodology **82** 184-197.

Guo, J., T. Su, Z. Li, et al. 2017 Declining frequency of summertime local-scale precipitation over eastern China from 1970 to 2010 and its potential link to aerosols Geophysical Research Letters **4411** 5700-5708.

HUSAR, B. Rudolf, HUSAR, et al. 2000 Distribution of continental surface aerosol extinction based on visual range data Atmospheric Environment **3429** 5067-5078.

Lin, J., A. V. Donkelaar, J. Xin, et al. 2014 Clear-sky aerosol optical depth over East China estimated from visibility measurements and chemical transport modeling Atmospheric Environment **951** 258-267.

Lin, J. T., A. van Donkelaar, J. Y. Xin, et al. 2014 Clear-sky aerosol optical depth over East China estimated from visibility measurements and chemical transport modeling Atmospheric Environment **95** 258-267.

Novotny, E. V.，H. G. Stefan 2007 Stream flow in Minnesota Indicator of climate change Journal of Hydrology **3343(4)** 319-333.

Ouarda, T. B. M. J., L. A. Yousef，C. Charron 2019 Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change International Journal of Climatology **394** 2306-2323.

Villarini, G., J. A. Smith, M. L. Baeck, et al 2011 On the frequency of heavy rainfall for the Midwest of the United States. Journal of Hydrology. **4001(2)** 103-120.

Yue, S., P. Pilon G. Cavadias 2002 Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series Journal of Hydrology **2591(4)** 254-271.

Yue, S., P. Pilon, B. Phinney, et al. 2002 The influence of autocorrelation on the ability to detect trend in hydrological series Hydrological Processes **169** 1807-1829.

Yue, S.， C. Y. Wang 2004 The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series Water Resources Management **183** 201-218.

Zhang, Q., C. Y. Xu, S. Becker, et al. 2006 Sediment and runoff changes in the Yangtze River basin during past 50 years. Journal of Hydrology **3313(4)** 511-523.