## **Supplementary Information**

## Impact of Western Pacific Subtropical High on Ozone Pollution over East China

Zhongjing Jiang<sup>1</sup>, Jing Li<sup>1†</sup>, Xiao Lu<sup>2</sup>, Cheng Gong<sup>3</sup>, Lin Zhang<sup>1</sup>, Hong Liao<sup>4</sup>

1 Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China

2 School of Engineering and Applied Sciences, Harvard University, Cambridge, USA

3 Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

4 Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China

†Corresponding authors: jing-li@pku.edu.cn

## 1. Data quality control procedure

As raw ozone data released by the China National Environmental Monitoring Centre (<u>www.cnemc.cn</u>) may suffer from various problems like instrumental errors and systematic bias, a series of quality control was applied to remove outliers and invalid measurements. We describe the data quality control protocol in the following five consecutive steps.

(1) Setting up lower and upper boundaries:

We set lower and upper limit values for valid range of ozone concentrations of  $[0,1000] \ \mu g \ m^{-3}$ . Although ozone concentration can reach several hundred in some extreme cases, data larger than 1000  $\ \mu g \ m^{-3}$  was found to always show a discontinuity with adjacent data. We thus considered it erroneous and marked them as NaN.

(2) Removing a string of data with the same value:

If in 5 consecutive hours, at least 4 are not NaN but having the same value, these data were eliminated.

(3) Removing abrupt jumps in data:

As the product of a stationary chemical reaction, it is unreasonable to find huge spikes in the ozone time series. Here we used two tests to remove these unrealistic changes. Bilateral test (at least 7 consecutive data in 7 hours) is used to remove cases like Fig.S1 a and unilateral test (at least 4 consecutive data in 4 hours) is used to remove cases when the right side or left side is a NaN value, examples of which are shown in Fig.S1 b&c.

- (4) Continuity check:
  - In 12 consecutive hours, if there is at most 1 valid data, we set this series of data as NaN.
  - In 24 consecutive hours, if there is at most 3 valid data, we set this series of data as NaN.
  - 3) In 48 consecutive hours, if there is at most 5 valid data, we set this series of data as NaN.
- (5) Checking the number of valid data:

We abandoned the sites which has more than half missing values in a month.

These quality control procedures removed approximately 2.35% of the raw data from 2014 to 2018. For each individual site, 84% to 100% (average 98% for all sites) of the hourly measurements were retained. After completing these data quality control procedures, although the amount of available data is reduced, the analysis results should be more reliable since the apparently unreasonable values are excluded. In addition, as the invalid measurement appears more frequently at night and ozone concentration always reach its peak in the afternoon, daily maximum 8 h average (MDA8) ozone metrics used in this study is less affected by the quality-control related data reduction.



Figure S1. Three the examples of invalid data.

## 2. Model evaluation



**Figure S2.** Observed (symbols) and simulated (filled contours) seasonal mean surface ozone MDA8 concentration (ppbv) over China for summer (JJA) from 2014 to 2018 (a-e).