Supplementary information

Data-driven machine learning quantifies ozone transport in the Hangzhou Bay urban cluster

Summary:

Page11: Fig. S7. Calculation result of Backward Trajectory

Equation. (S1) The formula of Linear interpolation

$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) \quad \text{(S1)}$$

where (x0, y0) and (x1, y1) are the coordinates of the two points before and after the value to be filled, and (x, y) are the coordinates of the points that need to be supplemented.

Equations. (S2)-(S7) The formula of LSTM

$$f_t = \sigma\left(W_f \cdot [h_{t-1}; x_t] + b_f\right) \quad \text{(S2)}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}; x_t] + b_i) \quad \text{(S3)}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}; x_t] + b_C) \quad \text{(S4)}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad \text{(S5)}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}; x_t] + b_o) \quad \text{(S6)}$$

$$h_t = o_t * tanh\,(C_t) \quad \text{(S7)}$$

Where $W_f, W_i, W_C, W_o,$ and b_f, b_i, b_C, b_o are the parameters that the model needs to learn. σ is the sigmoid function, which acts as the activation function of the neural network to map variables between 0 and 1. C_t represents a vector of new candidate values that will be added to the learning process. o_t is the output value of the control unit, and ht is its value between -1 and 1.

Equation. (S8) The formula of MAE

$$MAE = \frac{1}{m}\sum_{i=1}^{m} |(y_i - \hat{y}_i)| \quad \text{(S8)}$$

where y_i is the value of the i^{th} forecast, $\hat{y}$ is the value of the i^{th} observation, and m is the total amount of data.

Equation. (S9) The formula of RMSE

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \quad \text{(S9)}$$

where y_i is the value of the i^{th} forecast, $\hat{y}$ is the value of the i^{th} observation, and m is the total amount of data. The lower the MAE and RMSE, the higher the prediction accuracy of the model.

Equation. (S10) The formula of PMC

$$PMC = \frac{Cov\ (X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \quad \text{(S10)}$$

where X, Y are two variables that require correlation coefficients, $\mathrm{Cov}\ (\mathrm{X}, \mathrm{Y})$ is the covariance of X and Y; $D(X)$ is the variance of X, and $D(Y)$ is the variance of Y.

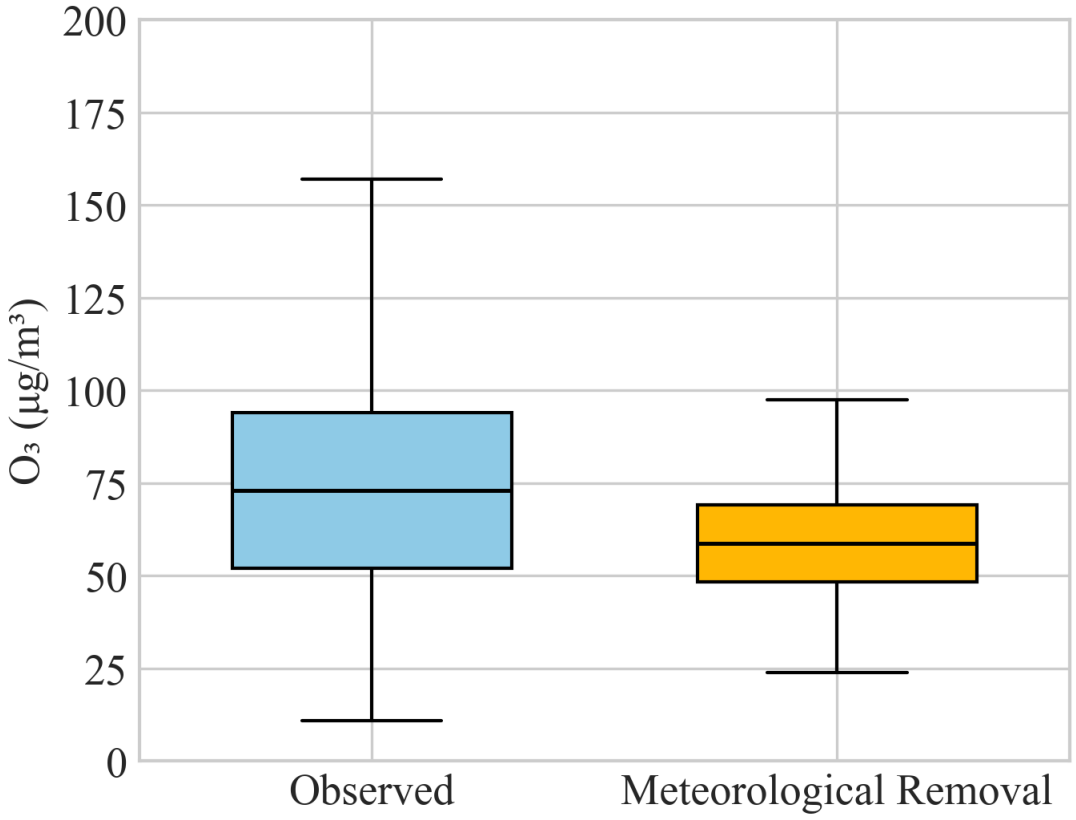

Fig. S1. Box plot comparing the distribution of hourly "Observed" O_3 concentrations with the "Meteorological Removal" O_3 predictions for June 2020. The plot illustrates the overall impact of the meteorological removal method, showing a systematic reduction in the median (central line), interquartile range (the box), and overall spread of the ozone distribution.

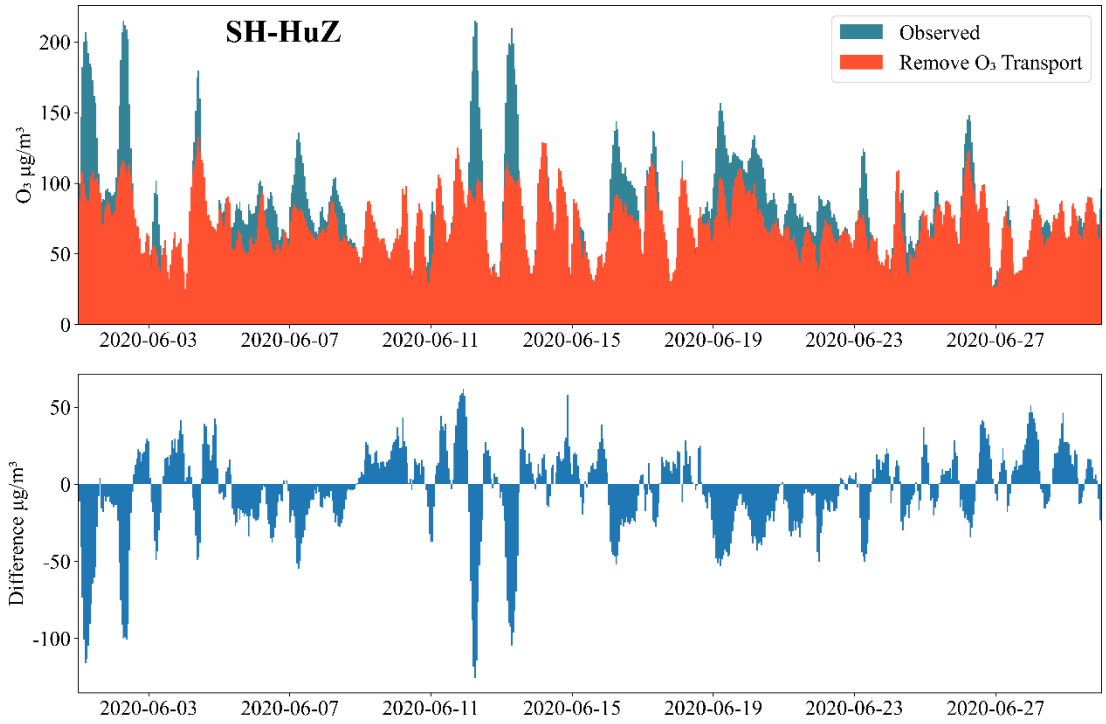

Fig. S2. Simulation results after Bi-LSTM Meteorological Removal (Huzhou to Shanghai). The blue-green area represents the observed hourly O_3 concentrations in Shanghai. The red line represents the simulated O_3 concentrations after applying the meteorological removal method to the source city's data. The dark blue bars represent the calculated contribution of transport.

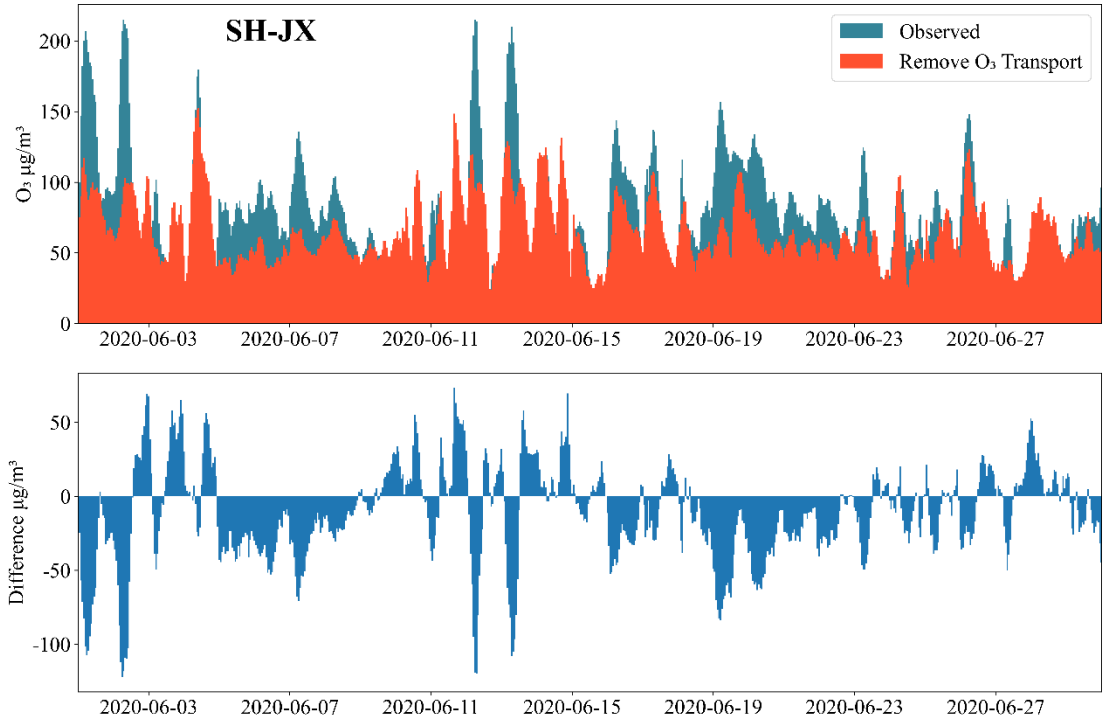

Fig. S3. Simulation results after Bi-LSTM Meteorological Removal (Jiaxing to Shanghai). The blue-green area represents the observed hourly O_3 concentrations in Shanghai. The red line represents the simulated O_3 concentrations after applying the meteorological removal method to the source city's data. The dark blue bars represent the calculated contribution of transport.

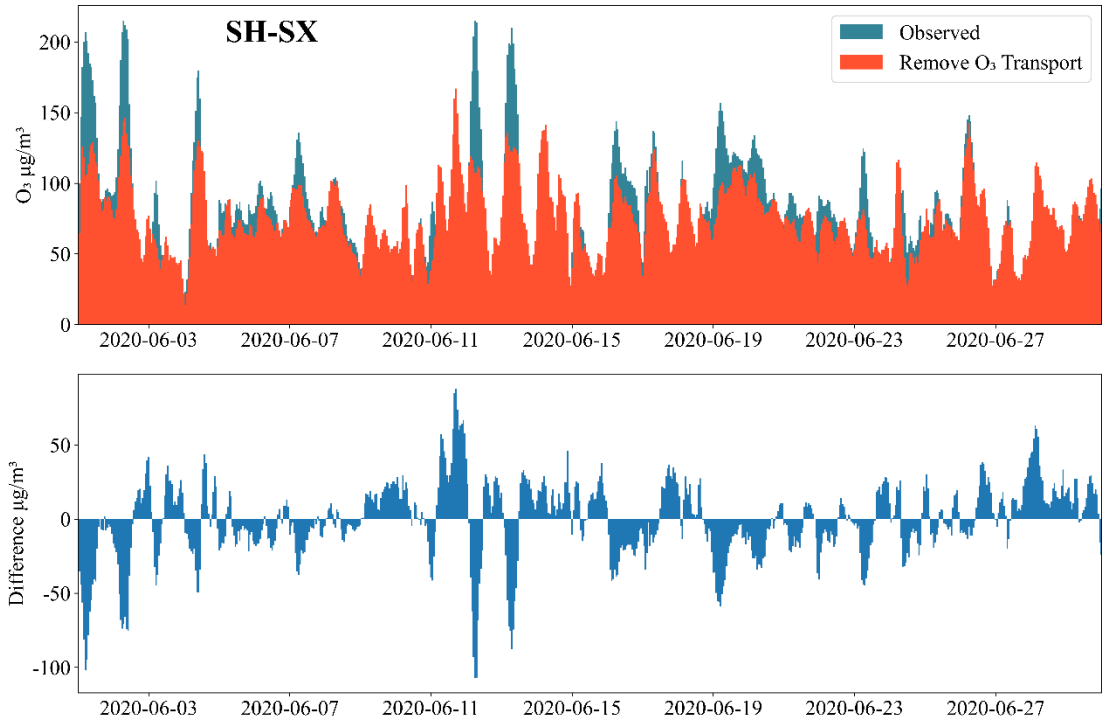

Fig. S4. Simulation results after Bi-LSTM Meteorological Removal (Shaoxing to Shanghai). The blue-green area represents the observed hourly O_3 concentrations in Shanghai. The red line represents the simulated O_3 concentrations after applying the meteorological removal method to the source city's data. The dark blue bars represent the calculated contribution of transport.

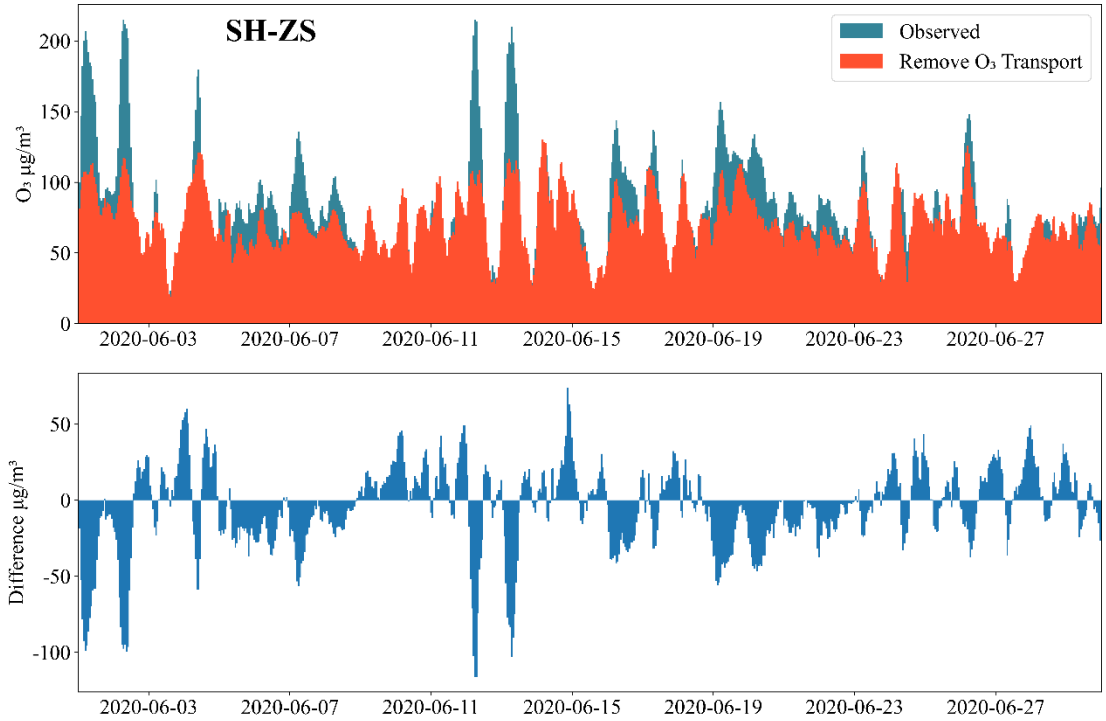

Fig. S5. Simulation results after Bi-LSTM Meteorological Removal (Zhoushan to Shanghai). The blue-green area represents the observed hourly O_3 concentrations in Shanghai. The red line represents the simulated O_3 concentrations after applying the meteorological removal method to the source city's data. The dark blue bars represent the calculated contribution of transport.

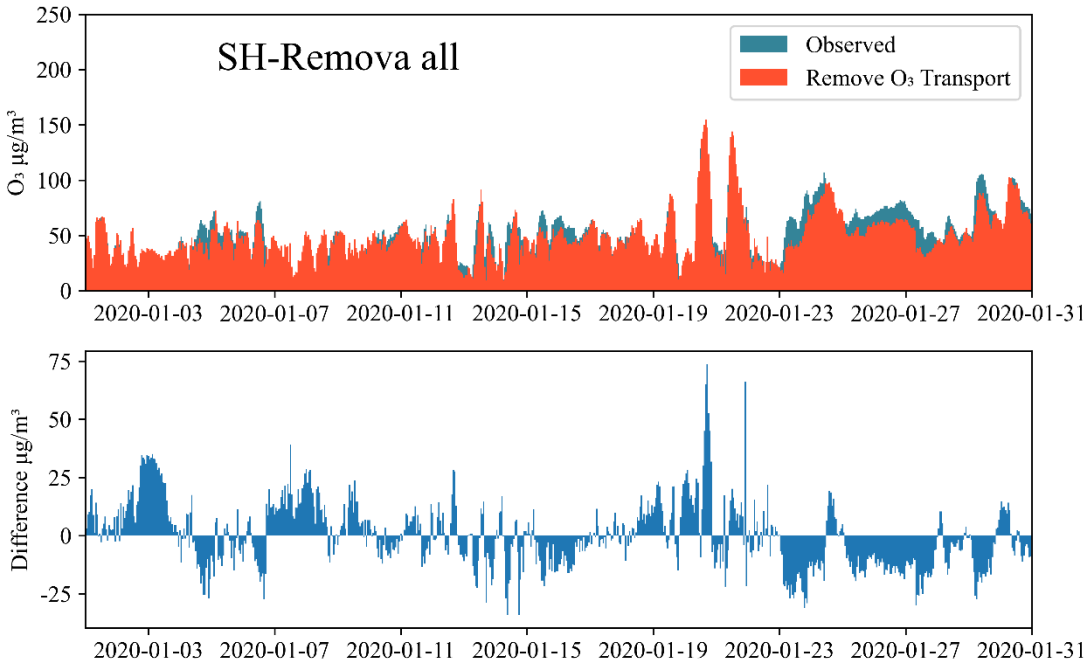

Fig. S6. Quantification of total O_3 transport from the entire Hangzhou Bay urban agglomeration to Shanghai in January 2020.

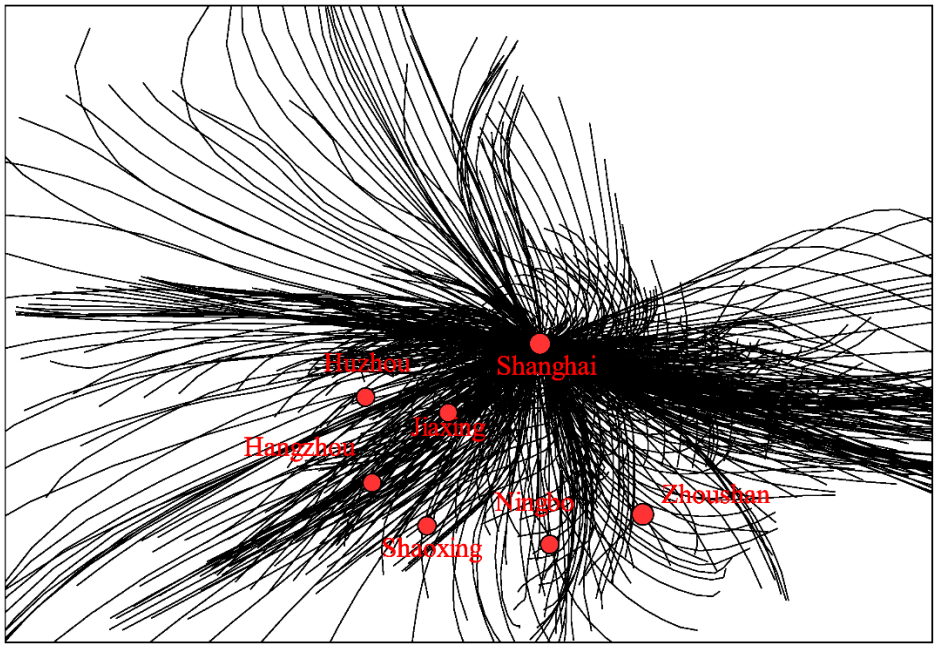

Fig. S7. Calculation result of Backward Trajectory