Supporting Information for "Deep learning to evaluate US NO_x emissions using surface ozone predictions"

Tai-Long He[1], Dylan B. A. Jones[1], Kazuyuki Miyazaki[2], Binxuan Huang[3], Yuyang Liu[4], Zhe Jiang[5], E. Charlie White[1], Helen M. Worden[6], John R. Worden[2]

[1]Department of Physics, University of Toronto

[2]Jet Propulsion Laboratory, California Institute of Technology

[3]School of Computer Science, Carnegie Mellon University

[4]Department of Computer Science, University of Toronto

[5]School of Earth and Space Sciences, University of Science and Technology of China

[6]Atmospheric Chemistry Observations and Modelling, National Center for Atmospheric Research

Contents of this file

1. Text S1 to S5
2. Figures S1 to S5
3. Table S1 to S2

Corresponding author: T.-L. He, Department of Physics, University of Toronto, 60 St George St, Toronto, ON, Canada, M5S 1A7. (the@physics.utoronto.ca)

Introduction This file contains supporting information for the main text of "Deep learning to evaluate US NO_x emissions using surface ozone predictions". Sections S1, S2 and S4 discuss the results from a variety of sensitivity tests, based on the experiments in the main text. Section S3 shows the regional domains used in the ozone-NO_x relationship analysis in the main text, and Sections S5 shows the spatial distribution of the high- and low-NO_x emission regions discussed in the main text.

Text S1: Sensitivity to model resolution

Figure S1 shows the predicted JJA MDA8 ozone at a model resolution of 3° x 3°. The errors are larger than for the 1.5° x 1.5° model, with a mean CONUS error of -0.27 ± 0.08 ppb. As shown in Table S1, the mean errors are -2.63 ± 0.18 ppb, 2.45 ± 0.16 ppb, and 0.95 ± 0.13 ppb for the Northeast, Southeast, and West Coast, respectively. The correlation between the predicted and observe ozone remain high across the United States ($R = 0.87$ for the CONUS).

Text S2: Sensitivity test of the impact of NO_x emissions on US MDA8 predictability

We conducted a sensitivity experiment with the 3° x 3° model in which we trained the model with only the meteorological predictors. The results of this experiment are shown in Figure S2 and Table S1. Using only the meteorological predictors the model captures well the ozone variability. For the CONUS, the model predicted MDA8 ozone with a correlation of $R = 0.81$ with only the meteorological predictors, compared to a value of $R = 0.87$ with the meteorological and NO_x emission predictors. However, without accounting for the NO_x emissions, the mean error in the predicted ozone is significantly larger, 4.50 ± 0.11 ppb compared to -0.27 ± 0.08 ppb (as indicated in Table S1).

Text S3: Regional definition for ozone-NO_x relationship analysis

Figure S3 shows the regional domains used for the ozone-NO_x relationship analysis in the main text. To capture the changing regional relationship between ozone abundances and NO_x emissions, we selected the southern California, southeastern US, and northeastern US domains shown, which are slightly more restricted geographically than the domains used in Figure 2 in the main text. We chose these more restricted domains to better isolate the regional ozone-NO_x relationships.

Text S4: Sensitivity to reduced training data

To evaluate the impact of reduced training data, we retrained the 3° x 3° model from 1980 to 2005, and tested it from 2005-2016. The time series of the predicted and observed MDA8 are plotted in Figure S4, and the error statistics for 2005-2009 and 2010-2016 are given in Table S2, respectively. Between 2005-2009, the MDA8 ozone predicted using the different NO_x trends all show good consistency over the US. However, after 2010, the bottom-up trends of NO_x resulted in an underestimation of MDA8 ozone relative to that from the top-down trends. The divergence is clearly visible in the time series of the monthly mean errors in Figure S4, with the EPA-based trend clearly producing the largest RMSE and negative bias after 2010. The results are consistent with those obtained with the higher resolution model. Even with the reduced training data, for the CONUS for 2010–2014 we obtain the smallest mean error with the TCR-2 trend and the largest error with the EPA trend.

Text S5: Spatial distribution of high- and low-NO_x emission regions

The $1.5° \times 1.5°$ grid cells shown as blue boxes in Figure S5 are the cells with high NO_x emissions that were determined following the approach of Li and Wang (2019). These cells are assumed to be representative of regions with anthropogenic emissions. All other grid cells in the CONUS domain are defined to be low-NO_x emission regions, and are assumed to be representative of "background" regions in the regional analysis discussed in the main text. Also shown in the figure is the definition of the regional domains for the CONUS, Northeastern US, Southeastern US, and West Coast used in the analysis.

References

Li, J., & Wang, Y. (2019). Inferring the anthropogenic nox emission trend over the united states during 2003–2017 from satellite observations: Was there a flattening of the emission tend after the great recession? *Atmospheric Chemistry and Physics Discussions*, *2019*, 1–35. Retrieved from https://www.atmos-chem-phys-discuss.net/acp-2019-472/ doi: 10.5194/acp-2019-472

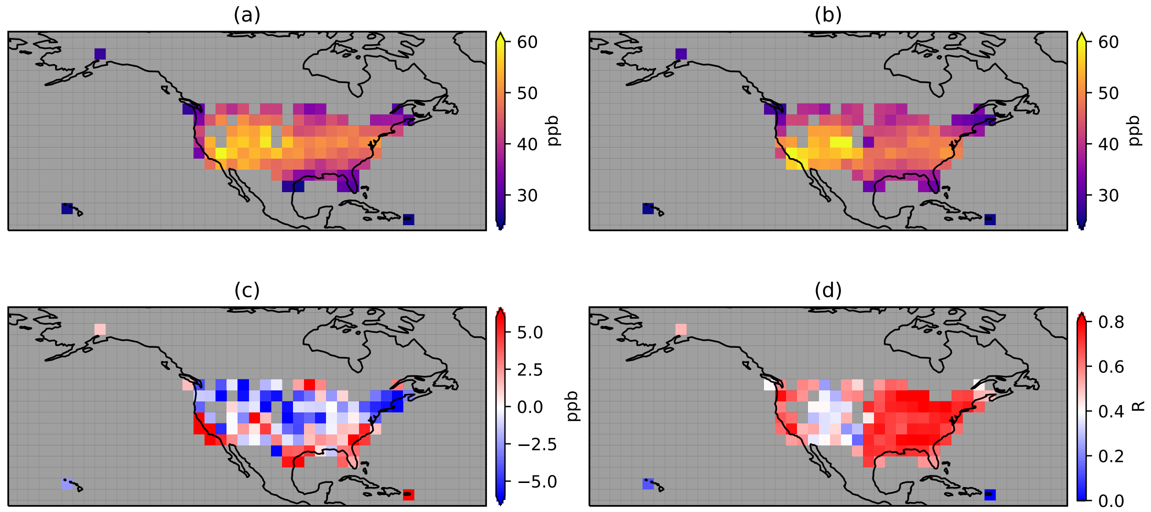

Figure S1. Observed (blue line) and predicted (orange line) daily (first column), 7-day averaged (second column), and monthly averaged (third column) JJA MDA8 ozone (in ppb) during the testing period (2010–2014) at a resolution of $3^\circ \times 3^\circ$.

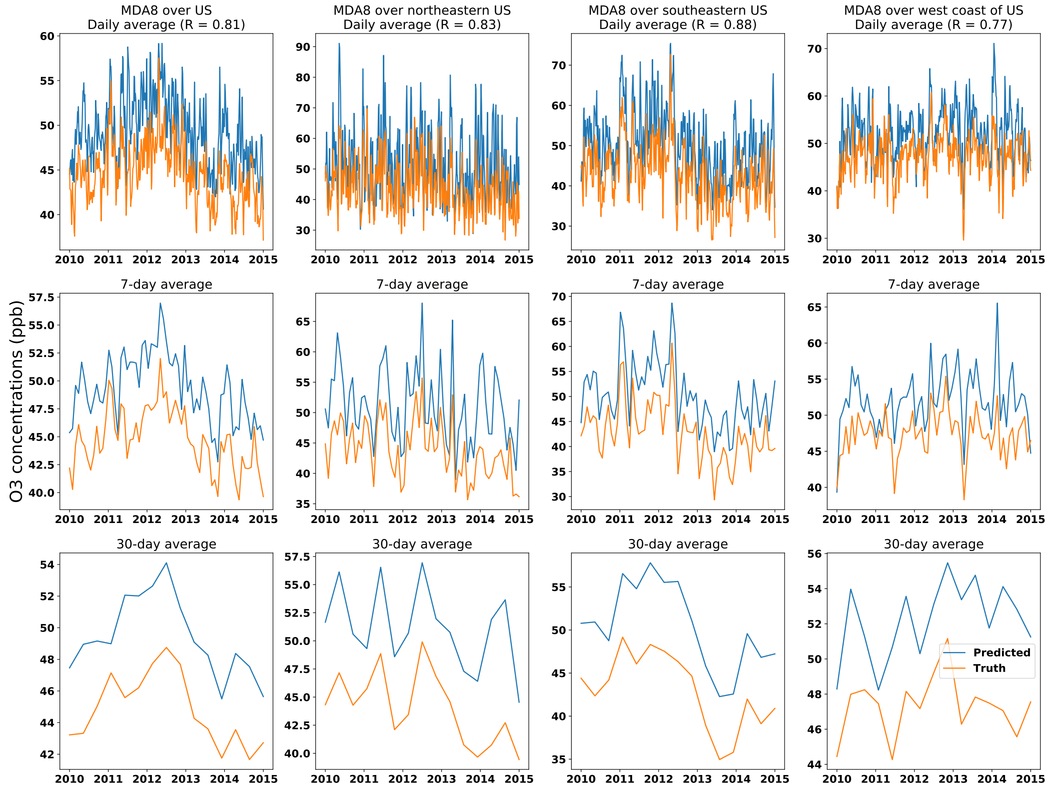

Figure S2. Observed (blue line) and predicted (orange line) daily (first row), 7-day averaged (second row), and monthly averaged (third row) JJA MDA8 ozone (in ppb) during 2010–2014 with only meteorological predictors. Shown are the time series for the CONUS (first column), the northeast (second column), the southeast (third column), and the west coast (last column).

Table S1. Regional error statistics for the model evaluation in the period of 2010–2014 for the model configured with the meteorological and NO_x emissions predictors and for the experiment using only the meteorological predictors. Shown are the mean errors, the standard error on the mean (SEM), and the R.

Predictors	Meteorological and NO_x		Meteorological	
Region	Mean Error ± SEM (ppb)	R	Mean Error ± SEM (ppb)	R
US	-0.27 ± 0.08	0.87	4.50 ± 0.11	0.81
Northeastern US	-2.63 ± 0.18	0.86	3.78 ± 0.25	0.82
Southeastern US	2.45 ± 0.16	0.88	10.88 ± 0.21	0.86
West coast	0.95 ± 0.13	0.81	5.37 ± 0.16	0.79

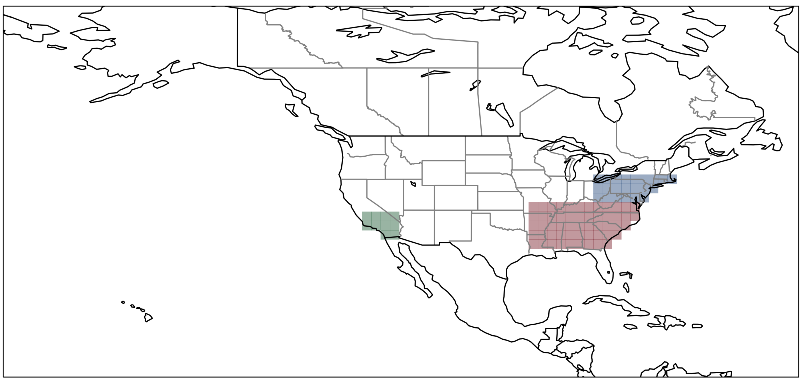

Figure S3. Domains for the Northeast, Southeast, and the southern California regions, which are indicated by the boxes shaded in blue, red and green, respectively. The domains are used for the ozone-NO_x relationship analysis in Figure 4 in the main text.

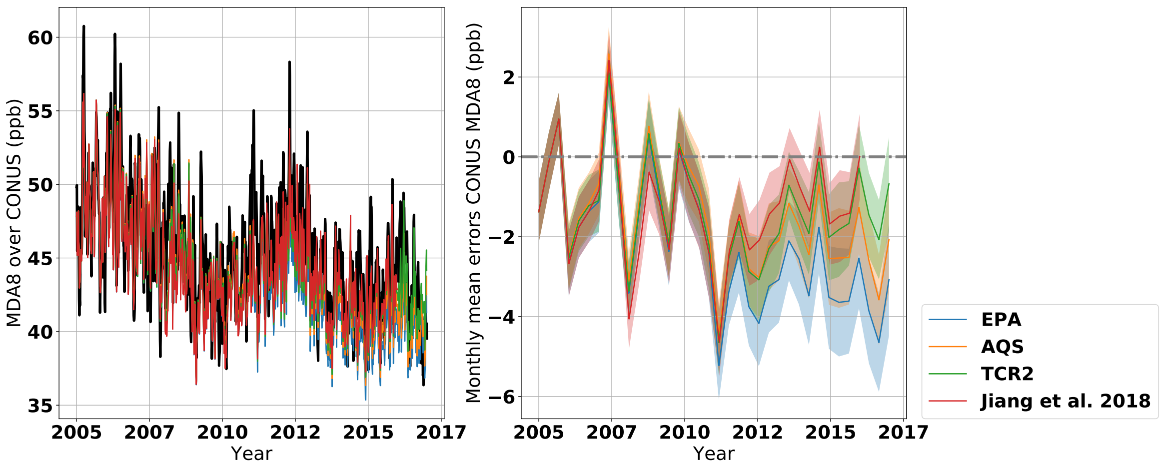

Figure S4. Observed and predicted daily mean (left) and monthly mean errors (right) of MDA8 ozone between 2005–2016 (2005–2015 for Jiang et al.). Shown are the AQS ozone observations (black line) and the model predictions based on the NO_x emissions scaled by the EPA (blue line), AQS NO_2 (orange line), TCR-2 (green line), and the Jiang et al. (red line) trends.

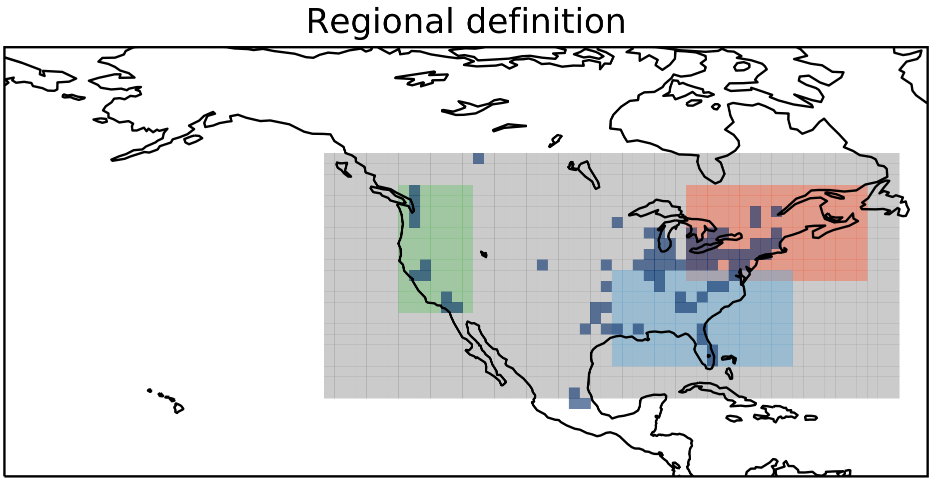

Figure S5. The spatial distribution of high- and low-NO_x emission regions. High-NO_x regions are indicated as dark blue grid cells. Also shown are the domains for the Northeast, Southeast, and West Coast regions, which are indicated by the boxes shaded in red, blue and green, respectively. The CONUS domain is shaded in grey.

Table S2. MDA8 ozone error statistics for the CONUS for 2005–2009 and 2010–2016 (2010–2015 for Jiang et al.).

	2005–2009		2010–2016 (2010–2015 for Jiang et al.)	
NO_x trend	Mean Error ± SEM (ppb)	R	Mean Error ± SEM (ppb)	R
EPA	-0.94 ± 0.11	0.83	-2.63 ± 0.08	0.81
AQS	-0.72 ± 0.11	0.83	-1.73 ± 0.08	0.81
TCR-2	-0.86 ± 0.11	0.84	-1.33 ± 0.08	0.79
Jiang et al.	-1.04 ± 0.11	0.82	-1.24 ± 0.08	0.81