Deep learning to evaluate US NOx emissions using surface ozone predictions

Tai-Long He[1], Dylan B. A. Jones[1], Kazuyuki Miyazaki[2], Binxuan Huang[3], Yuyang Liu[1], Zhe Jiang[4], Edward Charles White[5], Helen M Worden[6], and John R. Worden[7]

[1]University of Toronto
[2]Jet Propulsion Laboratory
[3]Carnegie Mellon University
[4]University of Science and Technology of China
[5]Department of Physics
[6]National Center for Atmospheric Research (UCAR)
[7]JPL / Caltech

November 24, 2022

Abstract

Emissions of nitrogen oxides (NOx = NO + NO2) in the United States have declined significantly during the past three decades. However, satellite observations since 2009 indicate total column NO2 is no longer declining even as bottom-up inventories suggest continued decline in emissions. Multiple explanations have been proposed for this discrepancy including 1) the increasing relative importance of non-urban NOx to total column NO2, 2) differences between background and urban NOx lifetimes, and 3) that the actual NOx emissions are declining more slower after 2009. Here we use a deep learning model trained by NOx emissions and surface observations of ozone to assess consistency between the reported NOx trends between 2005-2014 and observations of surface ozone. We find that the 2005-2014 trend from older satellite-derived emission estimates produced at low spatial resolution best reproduce ozone in low NOx emission (background) regions, reflecting the blending of urban and background NOx in these low-resolution top-down analyses. The trend from higher resolution satellite-based estimates, which are more capable of capturing the urban emission signature, is in better agreement with ozone in high NOx emission regions, and is consistent with the trend based on surface observations of NO2. In contrast, the 2005-2014 trend from the US Environmental Protection Agency (EPA) National Emission Inventory (NEI) results in an underestimate of ozone. Our results confirm that the satellite-derived trends reflect anthropogenic and background influences and that the 2005-2014 trend in the NEI inventory is overestimating recent reductions in NOx emissions.

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

Key Points:

- Satellite-based emission estimates of NO_x are influenced by anthropogenic and background NOx, but their relative contributions are unclear
- Our deep learning model suggests that the EPA NEI emission inventory overestimates the 2005–2014 trend in US NO_x emissions
- Trends in high-resolution space-based emission estimates and in surface NO2 are consistent in high emission regions

Corresponding author: Tai-Long He, the@physics.utoronto.ca

Abstract

Emissions of nitrogen oxides (NO_x = NO + NO_2) in the United States have declined significantly during the past three decades. However, satellite observations since 2009 indicate total column NO_2 is no longer declining even as bottom-up inventories suggest continued decline in emissions. Multiple explanations have been proposed for this discrepancy including 1) the increasing relative importance of non-urban NO_x to total column NO_2, 2) differences between background and urban NO_x lifetimes, and 3) that the actual NO_x emissions are declining more slower after 2009. Here we use a deep learning model trained by NO_x emissions and surface observations of ozone to assess consistency between the reported NO_x trends between 2005-2014 and observations of surface ozone. We find that the 2005-2014 trend from older satellite-derived emission estimates produced at low spatial resolution best reproduce ozone in low NO_x emission (background) regions, reflecting the blending of urban and background NO_x in these low-resolution top-down analyses. The trend from higher resolution satellite-based estimates, which are more capable of capturing the urban emission signature, is in better agreement with ozone in high NO_x emission regions, and is consistent with the trend based on surface observations of NO_2. In contrast, the 2005-2014 trend from the US Environmental Protection Agency (EPA) National Emission Inventory (NEI) results in an underestimate of ozone. Our results confirm that the satellite-derived trends reflect anthropogenic and background influences and that the 2005-2014 trend in the NEI inventory is overestimating recent reductions in NO_x emissions.

1 Introduction

Air pollution is a major cause of mortality globally (Cohen et al., 2017). In this context, tropospheric ozone is a key pollutant that is produced photochemically by the oxidation of hydrocarbons in the presence of nitrogen oxides (NO_x = NO + NO_2). Air pollution regulations have resulted in dramatic reductions in emissions of NO_x. However, Jiang et al. (2018) suggested that NO_x emission estimates inferred from satellite observations (referred to as top-down estimates) indicate that there has been a slowdown in the reduction rate since 2009, compared to the bottom-up emission inventory reported by the US Environmental Protection Agency (EPA) National Emission Inventory (NEI). In contrast, it has been suggested that the slowdown in the reduction rate in the satellite-derived emission estimates does not indicate a discrepancy with the NEI inventory, but instead is due to the increasing relative influence of non-anthropogenic NO_x emissions on atmospheric NO_x as captured by the satellite measurements (Silvern et al., 2019). It has also been reported by J. Li and Wang (2019) that the satellite-derived trends are consistent with the trends in surface observations of NO_2 in high emission regions and that the discrepancy between the top-down and bottom-up trends are due to non-linearity in the relationship between NO_x emissions and the satellite observations of NO_2 in low emission "rural" regions. Here we use a data-driven deep learning (DL) model that predicts surface ozone abundances across the US, which allows us to assess the consistency of the inferred 2005–2014 trends in NO_x emissions with observed surface ozone.

Surface ozone in the United States is highly variable on both short and long time scales, reflecting the influence of meteorology, non-linearity in the ozone chemistry, and changes in the emissions of ozone precursor gases. Atmospheric models used to simulate the distribution of ozone typically do not reproduce the observed long-term trend in tropospheric ozone, partially due to large uncertainty in simulated ozone response to varying NO_x emissions (Miyazaki et al., 2020b). Furthermore, these models tend to overestimate summertime surface ozone abundances in the United States. For example, in an evaluation of 16 global models and one hemispheric model it was found that the models overestimated summertime daily maximum 8-h average (MDA8) surface ozone in the eastern United States by 10–20 ppb (Reidmiller et al., 2009; Travis et al., 2016).

Machine learning methods are now becoming more widely used for simulation of atmospheric composition (e.g., Keller & Evans, 2019; Seltzer et al., 2020). For example, Seltzer et al. (2020) used an artificial neural network to simulation surface ozone to assess the impact of ozone exposure on human health and crop yields. In this study, we apply a state-of-the-art DL model in predicting surface ozone in the continental US. The data-driven, U-shaped DL model employed here captures well both the long-term and short-term variability in summertime MDA8 ozone in the United States. Previous studies have used statistical methods to investigate the relationship between large-scale atmospheric circulation patterns and summertime surface ozone (Gardner & Dorling, 2000; Shen & Mickley, 2017). Recent achievements in DL over the past few years show that empirical models are able to learn both spatial and temporal patterns in the input data (Goodfellow et al., 2016). It has been suggested that DL approaches have the potential to improve our predictive ability and understanding in a wide range of challenges we have in Earth science (Reichstein et al., 2019). A key benefit of this DL approach is that it is independent of the chemical errors that are typically found in atmospheric chemical transport models used in air quality studies. It requires no a priori assumptions about the relationship between NO_x emissions and tropospheric ozone associated with changes in the lifetime of NO_x or in emissions of volatile organic compounds (VOCs). During training, the changing relationship between NO_x emissions and ozone is learned by the model, to the extent that these changes are reflected in the ozone observations. This benefit of DL is also a limitation, in that these approaches currently are incapable of providing direct mechanistic insights in the processes governing the learned relationships. Nevertheless, the model provides a useful tool to determine which putative trend in NO_x emissions is most consistent with ozone observations.

2 Methods

2.1 A hybrid deep learning model to predict summertime surface ozone

A schematic of the model is given in Figure 1. The model has eight convolutional layers and three max pooling layers to extract the dominant features in the input data. Convolutional neural networks (CNNs) are the most fundamental model in DL and are able to efficiently capture spatial correlations in data. The weights in each CNN layer in the DL model perform convolutional calculations with the input and forward the output into subsequent layers. Max pooling layers are similar to convolutional layers, except that the convolution is replaced by a simple max transformation. Max pooling layers are used to further reduce data dimensionality and to extract dominant features. The optimization of the model is supervised by the "truth", which is the summertime MDA8 ozone measured by the AQS network in this study. The weights in the CNNs are optimized using the back-propagation algorithm (Rumelhart et al., 1986; LeCun et al., 1989), which employs the partial derivatives of cost function with respect to the truth.

We also embed the recurrent neural networks (RNNs) into our DL architecture. The RNNs were developed for sequential forecasting problems (Rumelhart et al., 1988), which showed strong skills in capturing dynamics hidden in data. The RNN model used in this study is the long-short term memory (LSTM) cell (Hochreiter & Schmidhuber, 1997), which is used to enhance the model's ability to capture the temporal variability in summertime ozone. In this study, the dynamics captured by the LSTM model includes both short-term daily variability and long-term trends in MDA8 ozone. We made the model deeper by stacking 3 LSTM cells in series to amplify its predictive skills.

After the input information gets compressed by the convolutional blocks and the LSTM cells, the latent vectors are projected to the output layer via a decoder that consists of a sequence of transposed convolutional layers and upsampling layers. Following Ronneberger et al. (2015), we added residual learning connections that forward the high-resolution features extracted by the encoder to the decoder for better localization of the

features learned by the DL model. These connections are helpful with faster convergence of the optimization, as they contain trainable weights that represents more direct relationship between input and output variables (H. Li et al., 2018).

The loss function to be optimized is defined as the mean squared error calculated in each grid box as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{1}$$

where y_i and $\hat{y}_i$ are the predicted and observed MDA8 ozone. The Pearson correlation coefficient between predictions and observations is used as an auxiliary metric of model performance. This performance evaluation is only computed in grid boxes where AQS measurements are available. This way the optimization of the model is not influenced by the imperfect observational coverage of the AQS data. The back-propagation algorithm is used to train this end-to-end architecture, with the ADAM optimization algorithm for a faster convergence (Kingma & Ba, 2014).

2.2 Summertime ozone predictors

Large-scale patterns in atmospheric circulation, sea surface temperatures (SSTs), and sea level pressure (SLP) influence summertime ozone variability in the United States on synoptic to interseasonal timescales (Shen & Mickley, 2017; Shen et al., 2015). To represent these large-scale processes, in addition to well-known proximate meteorological drivers of ozone variability, we have therefore selected the following MDA8 ozone predictors, focusing on the JJA period: anthropogenic emissions of NO_x, mean sea level pressure (MSLP), geopotential at 500 hPa level (Z), downward shortwave radiation (SSRD), sea surface temperature (SST), 2-meter temperature (T2M), and 2-meter dew point (D2M). The input NO_x emissions are separated into the following seven emissions sectors to better help with the training: agriculture (AGR), the power industry (ENE), the manufacturing industry (IND), residential and commercial (RCO), international shipping (SHP), surface transportation (TRA), and waste disposal (WST). The sector-based NO_x emissions provide geospatial information to the neural networks, which helps with the regression and localization of ozone levels.

2.3 Data

The meteorological data are from the ERA-Interim reanalysis (Dee et al., 2011) from the European Centre for Medium-Range Weather Forecasts (ECMWF), which have been regridded to a horizontal resolution of $1.5° \times 1.5°$. The NO_x emissions are from the CEDS inventory (Hoesly et al., 2018), and were regridded from their native resolution of 0.5° x 0.5° to $1.5° \times 1.5°$. We chose CEDS for the analysis as it was specifically developed to provide historical emissions for climate and atmospheric chemistry models. CEDS uses a sequential scaling approach in which emissions are first scaled to the Emission Database for Global Atmospheric Research (EDGAR) inventory, and then rescaled to the appropriate national inventory. In the United States, this rescaling is with respect to the NEI inventory. CEDS also smooths discontinuities in the NEI inventory, resulting in differences between CEDS and NEI. All the input data are cropped to a regional domain extending between 0°–72°N, and between 180°W–0° to encompass the North Pacific and the North Atlantic, where strong linkages were found between ocean forcing and summertime climate in the eastern United States (Shen & Mickley, 2017; Sutton & Hodson, 2005, 2007; Gill, 1980).

While CEDS emissions were used to develop the general DL framework, we used the EPA NEI inventory to evaluate the 2005-2014 bottom-up trends. Specifically, we used the NEI version 2014v1 downloaded from the EPA website (`https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data`). The top-down

estimates of NO_x emissions are from the Tropospheric Chemistry Reanalysis (TCR) data product (`https://tes.jpl.nasa.gov/tes/chemical-reanalysis/`). The TCR data products were generated with a Kalman-filter-based data assimilation system that assimilated satellite measurements of ozone, CO, NO_2, HNO_3, and SO_2 from the Ozone Monitoring Instrument (OMI), the SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY), the Global Ozone Monitoring Experiment (GOME-2), the Tropospheric Emission Spectrometer (TES), the Microwave Limb Sounder (MLS), and the Measurement Of Pollution In The Troposphere (MOPITT) satellite instrument. The TCR NO_x emissions used in Jiang et al. (2018) was an earlier version of the data product, described in (Miyazaki et al., 2015). The assimilated was conducted at a spatial resolution of $2.8°\times2.8°$. In this paper we also analyze the newer version of the TCR chemical reanalysis (TCR-2) (Miyazaki et al., 2020b), which utilized updated satellite observations at a higher model resolution of $1.1°\times1.1°$. The $1.5°\times1.5°$ resolution used for the deep learning model was chosen to be similar to the resolution of the TRC-2 product.

MDA8 ozone was estimated from ozone measurements from the EPA Air Quality System (AQS) (https://www.epa.gov/aqs). The MDA8 ozone were aggregated to $1.5°\times 1.5°$ grid boxes. The average value of MDA8 for each grid box will not be representative of any specific type of station data (e.g., road, industrial point source, etc), however, the aggregated data set provides a good measure of regional surface ozone and can be used to evaluate the impact of recent trends in NO_x emissions on ozone. By using this data set as the truth during training, the model is be able to capture the linkage between the NO_x emission trend and the trend in surface ozone on these scales.

3 Results

3.1 Predicting summertime ozone

The model is trained using AQS ozone observations from 1980 to 2009, and its performance evaluated using data from the subsequent five years. We do not train the model over the entire 1980–2014 period as our goal is to use the model in a predictive context to evaluate the putative NO_x emission trends after 2009. During the training and evaluation of the model we use the CEDS NO_x emissions. As shown in Figure 2, the predicted JJA MDA8 ozone concentrations between 2010–2014 are in good agreement with the AQS ozone observations. The mean error for the contiguous United States (CONUS), the northeastern United States, the southeastern United States, and the west coast is -0.09 ± 0.37 ppb, 0.28 ± 0.82 ppb, 0.12 ± 0.60 ppb, and 0.15 ± 0.54 ppb, respectively. We show in Figures S1 in the Supplementary Information that the errors are larger with a lower model resolution of $3°\times3°$, and we would expect improved performance at higher spatial resolution than the $1.5° \times 1.5°$ resolution. However, even at the course resolution these errors in ozone are significantly smaller than the 10–20 ppb by which conventional model simulations typically overestimate JJA MDA8 ozone in the eastern United States (Reidmiller et al., 2009). In sensitivity tests in which we exclude NO_x emissions as a predictor in the lower resolution version of the model, the predicted ozone abundances are significantly more biased across the CONUS (see Figure S2 and Table S1 in the Supplementary Information). The results indicate that during the training, the model is able to capture the changing relationship between NO_x emissions and surface ozone.

Overall, the model is able to capture both the short-term and long-term dynamics of MDA8 ozone well (Figure 3). The predicted MDA8 ozone over the United States have ubiquitously high correlations with the observations. However, low temporal correlations are found in the Intermountain West ($R \approx 0.40$), where there are fewer AQS observations. Also, this region is strongly influenced by free troposphere background ozone abundances rather than local or regional precursor emissions (Zhang et al., 2014). Including wind fields and wildfire emissions as additional predictors may improve the pre-

dictability of MDA8 ozone in the Intermountain West, as wildfires and transport from the western U.S. could have an impact on ozone in this region. The year-to-year variability of surface ozone is also shown to be related to stratospheric intrusions in spring (Zhang et al., 2014; Lin et al., 2015) and the emissions of NO_x from lightning in summer (Zhang et al., 2014).

The US NO_x emissions have dramatically decreased since the 1990s due to air pollution regulations, and there has been regional changes in the ozone-NO_x relationship associated with these emission reductions, reflecting the non-linearity in the ozone chemistry (He et al., 2020). As a result of this non-linearity, the same ozone concentration can be produced at low and high levels of NO_x emissions, reflecting the fact that the chemistry is a source of ozone at low NO_x emissions and a sink for ozone at high NO_x emissions. This non-linearity is also influenced by VOC emissions. At low NO_x emission, where ozone increases with increasing NO_x, the chemistry is described as being NO_x-sensitive. In contrast, at high NO_x emissions, where ozone decreases with increasing NO_x emissions, but increases with increasing VOC emissions, the chemistry is considered to be VOC-sensitive. Here we show that the DL model not only predicts well the short-term ozone variability, but also captures the regionally-dependent chemical relationship between ozone and NO_x emissions. Since the model was trained with data from 1980 to 2009 and data after 2010 were not in the training samples, we chose the following three periods to examine the ozone sensitivity to NO_x emissions the over the past 3 decades: 1986–1990, 2001–2005 and 2010–2014. Because of the 1.5° resolution, we cannot explicitly examine changes at urban scales, so instead we analyzed the chemical relationship between ozone and NO_x emissions for the northeastern US, the southeastern US, and southern California, as shown in Figure S3 in the Supplementary Information.

The relationships between summertime ozone and NO_x emissions for the 3 selected time periods are shown in Figure 4. For the northeastern US, between 1986–1990 the extreme values of surface ozone observations between 1986–1990 exhibit a slight negative slope, whereas the slope of median values is almost flat. The results suggest that the ozone photochemical regime was transitional between the VOC-sensitive and NO_x-sensitive regimes during this time. This transitional photochemical regime was observed for 2001–2005 and 2010–2014 periods as well. For the southeastern US and southern California, the ozone-NO_x relationship in 1986–1990 has a turning point around 10×10^{-11} kgN m^{-2} s^{-1}, suggesting a NO_x-sensitive regime in regions of low NO_x emissions (less than 10×10^{-11} kgN m^{-2} s^{-1}) and a transition regime (between VOC- and NO_x-sensitive conditions) in locations with higher NO_x emissions. By 2010–2014, the the southeastern US and southern California become more NO_x-sensitive, particular the southeastern US.

Comparison of Figures 4a–4c with Figures 4d–4f shows that the ozone-NO_x relationships for all three regions over the three time periods are correctly predicted by the DL model. Since the model is trained with data from 1980 to 2009, the agreement between the observed and predicted ozone relationships for the 1980s and early 2000s is somewhat expected. However, the regional consistency between the modeled and predicted ozone relationships for 2010–2014 shown in Figure 4 suggests that through training the model is able to learn the changing, regionally-dependent chemical relationship between surface ozone and NO_x emissions in the US at the $1.5°\times1.5°$ spatial scales to which we aggregated the data.

3.2 Trend of anthropogenic NO_x emissions over the United States after 2010

The trend in the annual mean NO_x emissions from the NEI bottom-up inventory as well as from top-down emission estimates from Jiang et al. (2018) and TCR-2 (Miyazaki et al., 2020a) are shown in Figure 5. As can be seen, there is good agreement in the NO_x

emission trend in the different inventories between 2005, when the top-down inventories became available, and 2010. However, after 2010 the top-down inventories suggest a significant slowdown in the rate of reduction of NO_x emissions in the United States (Jiang et al., 2018). Included in Figure 5 is the trend in surface NO_2 from observations from the AQS network. The AQS NO_2 trend suggests a smaller reduction in NO_x emissions than the NEI inventory between 2005-2010, but not as pronounced as the slowdown observed in the top-down inventories.

Evaluating these emission trends using conventional atmospheric chemical transport models is challenging due to the fact that those models are impacted by deficiencies in the employed chemical mechanisms and dynamical parameterizations. The DL model captures the relationship between MDA8 and its predictors based on the input in situ and meteorological data only, and is able to mitigate the impact of a majority of the sources of error in conventional atmospheric models.

To evaluate the trends in the NO_x emissions, we use the trained DL model to predict MDA8 ozone from 2010 to 2014 using the CEDS NO_x emissions scaled by the different annual trends shown in Figure 5. The CEDS inventory is scaled as follows:

$$E_i^m = E_i^{CEDS} \cdot \beta^m \tag{2}$$

where E_i^{CEDS} is the CEDS emissions for month i, β^m is the annual scaling factor that captures the trend shown in Figure 5 for a given inventory m, and E_i^m is the resulting scaled NO_x emissions used in the model prediction of MDA8 ozone. For each NO_x trend, we run an ensemble of 22 ozone predictions.

The error statistics for the predicted MDA8 ozone are shown in Figure 6. The observed AQS NO_2 trend results in a mean error of -0.20 ± 0.38 ppb across the CONUS, which is statistically indistinguishable from the standard results (-0.09 ± 0.37 ppb) obtained with the CEDS inventory (shown in Figure 3) and that based on the TCR-2 trend (-0.12 ± 0.38 ppb). In contrast, our results indicate that the NEI and Jiang et al. trends are statistically inconsistent, with the NEI trend resulting in a larger negative bias of -0.87 ± 0.39 ppb and the Jiang et al. trend producing a positive bias of 0.15 ± 0.39 ppb. Averaged across the United States, the satellite-based TCR-2 trend produces the smallest mean errors in predicted ozone. In a sensitivity test in which we trained the model using data from 1980-2005 and predicted MDA8 ozone for 2005-2016, the four trends all produced consistent ozone predictions between 2005–2010, but diverged after 2010, with the NEI trend producing the largest negative bias in predicted ozone (See Figure S4 in the Supplementary Information).

To investigate whether the satellite observations of NO_2 are more representative of non-anthropogenic NO_x in rural regions after 2010 (Silvern et al., 2019; J. Li & Wang, 2019), we segregated the predictions into high-NO_x and low-NO_x regions according to whether the average NO_x emission in a given grid box is greater than or less than 1×10^{11} molec cm^{-2} s^{-1}, respectively, following J. Li and Wang (2019). We assume that these high-NO_x regions are strongly influenced by anthropogenic emissions, whereas the low-NO_x regions are more representative of background NO_x conditions (see Figure S5 in the Supplementary Information for the spatial distribution of these high-NO_x and low-NO_x emission regions). As shown in Figure 6, the NO_x emissions scaled by the observed AQS NO_2 trend produce ozone predictions with the smallest error (0.03 ± 0.53 ppb) in regions with high NO_x emissions. In these high NO_x regions the NEI trend results in an ozone bias of -0.87 ± 0.37 ppb. For the low-NO_x regions, the best performance is obtained with the Jiang et al. trend, with a bias of 0.06 ± 0.37 ppb.

Our results agree with Silvern et al. (2019), suggesting that the satellite-based NO_x trend in Jiang et al. (2018) is more representative of background NO_x conditions. They also confirm the finding of Jiang et al. (2018) that the 2005–2015 NEI trend is overestimating reductions in NO_x emissions. We find that the more recent satellite-based trend

from TCR-2 is relatively consistent with the AQS NO_2 trend (in both high- and low-NO_x regions). As noted in Section 2.3, the TRC-2 satellite-based NO_x emission product is an update of that used in Jiang et al. (2018), and a key difference between the two products that could explain the consistency of the TCR-2 and AQS NO_2 trends is that the TCR-2 product was derived at higher spatial resolution (1.1° x 1.1° compared to 2.8° x 2.8°), which offers a better means of discriminating between anthropogenic and background NO_x. This updated emission product, TCR-2, is in agreement with J. Li and Wang (2019) who found that the trend in satellite observations of NO_2 is consistent with the AQS NO_2 data in urban regions, but reflects a more gradual decrease in NO_2 in rural regions.

4 Conclusions

We have developed a state-of-the-art DL model to predict summertime daily MDA8 ozone in the U.S. The model uses 13 predictors, including large-scale meteorological variables and sector-specific anthropogenic emissions of NO_x. The model was trained with observed summertime MDA8 ozone data from 1980 to 2009 and tested with data from 2010 to 2014. We found that the model captured well the daily variability in MDA8 ozone across the United States, predicting ozone with a correlation of $R = 0.88$ and a mean error of -0.09 ± 37 ppb. Regionally, the model has high predictability of ozone in the eastern U.S. and on the west coast ($R > 0.85$), but low predictability in the Intermountain West ($R \approx 0.4$).

We used the model to evaluate trends in NO_x emissions between 2005–2014 inferred from top-down and bottom-up inventories, in the context of the model predictions of surface ozone. Our analysis suggested that care is needed in interpreting top-down satellite-based emission estimates as the satellite observations are affected by a combination of anthropogenic NO_x emissions and rural NO_x conditions. The trend in the satellite-based NO_x emission estimates in Jiang et al. (2018) is more indicative of the trend in background NO_x. However, our results confirm that the recent higher resolution TCR-2 satellite-based emission inventory is consistent with the surface NO_2 trend in regions of high anthropogenic NO_x emissions. The results also confirm that the NEI inventory is overestimating the reduction in NO_x emissions after 2010, which Jiang et al. (2018) attributed to the growing relative contribution of less-stringently regulated emissions from diesel and off-road vehicles not accounted for in the NEI inventory. Our analysis demonstrates the potential utility of DL for air quality studies. The DL architecture employed here is generic and flexible. It can be utilized to realize other high-dimensional predictions, given the spatial and temporal dynamics in the data.

Acknowledgments
This work was supported by the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2019-06804) and the Canadian Space Agency (grant 16SUASEMIS). Computations were performed on the Graham supercomputer of Compute Ontario and Compute Canada. Part of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). We thank B. C. McDonald for helpful discussions about the NO_x emission inventories.

The satellite-derived NO_x emissions from Jiang et al. (2018) can be obtained from the sources cited in Jiang et al. (2018). The NEI data (version 2014v1) can be downloaded from the EPA website (`https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data`). The TCR-2 NO_x emission data are available for download from `https://tes.jpl.nasa.gov/chemical-reanalysis/`. The EPA AQS surface NO_2 data are available for download from `https://www.epa.gov/aqs`. The CEDS anthropogenic emission data were originally downloaded from `https://esgf-node`

.llnl.gov/search/input4mips/, and were processed by the Support Team for the GEOS-Chem model. The processed CEDS data used here are available for download from http://wiki.seas.harvard.edu/geos-chem/index.php/CEDS_anthropogenic_emissions.

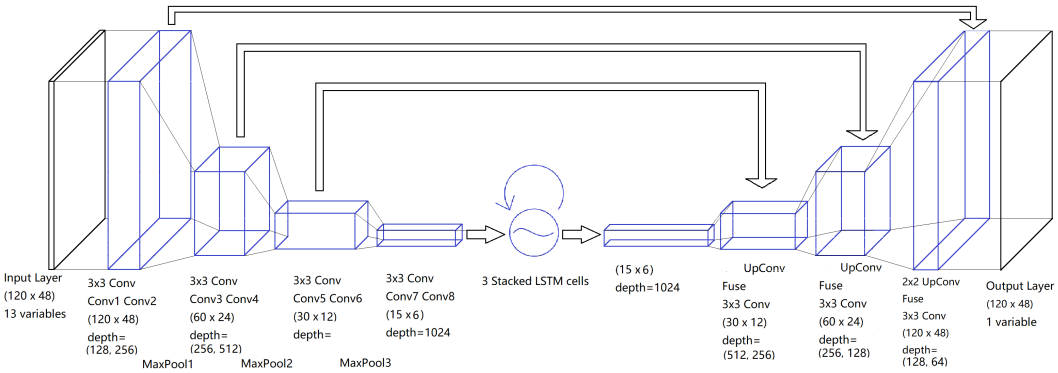

Figure 1. Deep learning model to predict JJA MDA8. The model consists of an input layer with 13 channels for the ozone predictors, eight convolution and three max pooling layers to extract the dominant features in the data, and three stacked LSTM cells to capture the dynamics in the data. Compressed data are then passed to transposed convolution layers for projection to the output layer. The three arrows at the top indicate the residual learning connections that forward the high-resolution features extracted by the encoder to the decoder for better localization of the features.

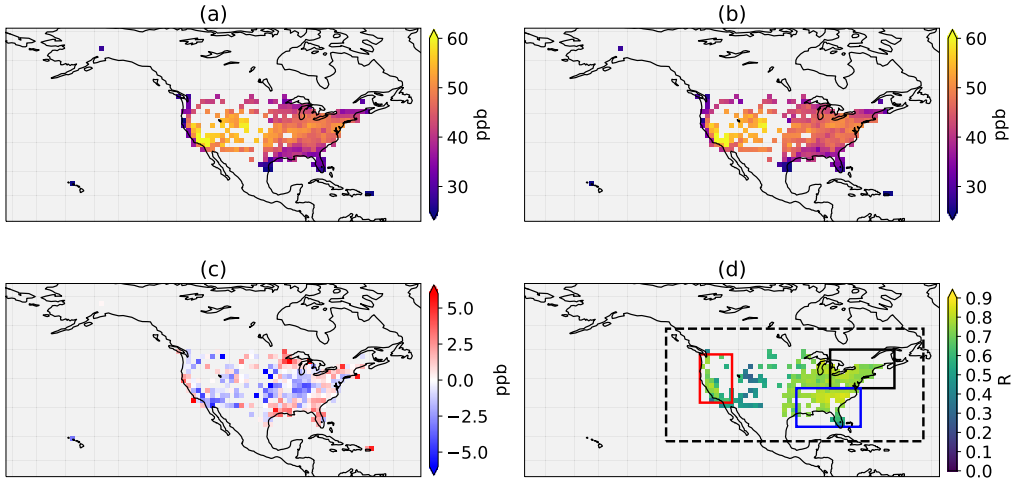

Figure 2. Observed (top left) and predicted (top right) mean JJA MDA8 ozone during 2010–2014. Also shown (bottom left) is the absolute error (in ppb) for the predicted minus observed MDA8 ozone. The errors are calculated where the AQS observations are located. Correlation (R) between the observed and predicted MDA8 ozone in each grid box is shown in bottom right. Also shown in bottom right are the definitions of the CONUS, Northeastern US, Southeastern US and the West coast domains in blacked dashed box, black solid box, blue box, and red box, respectively.

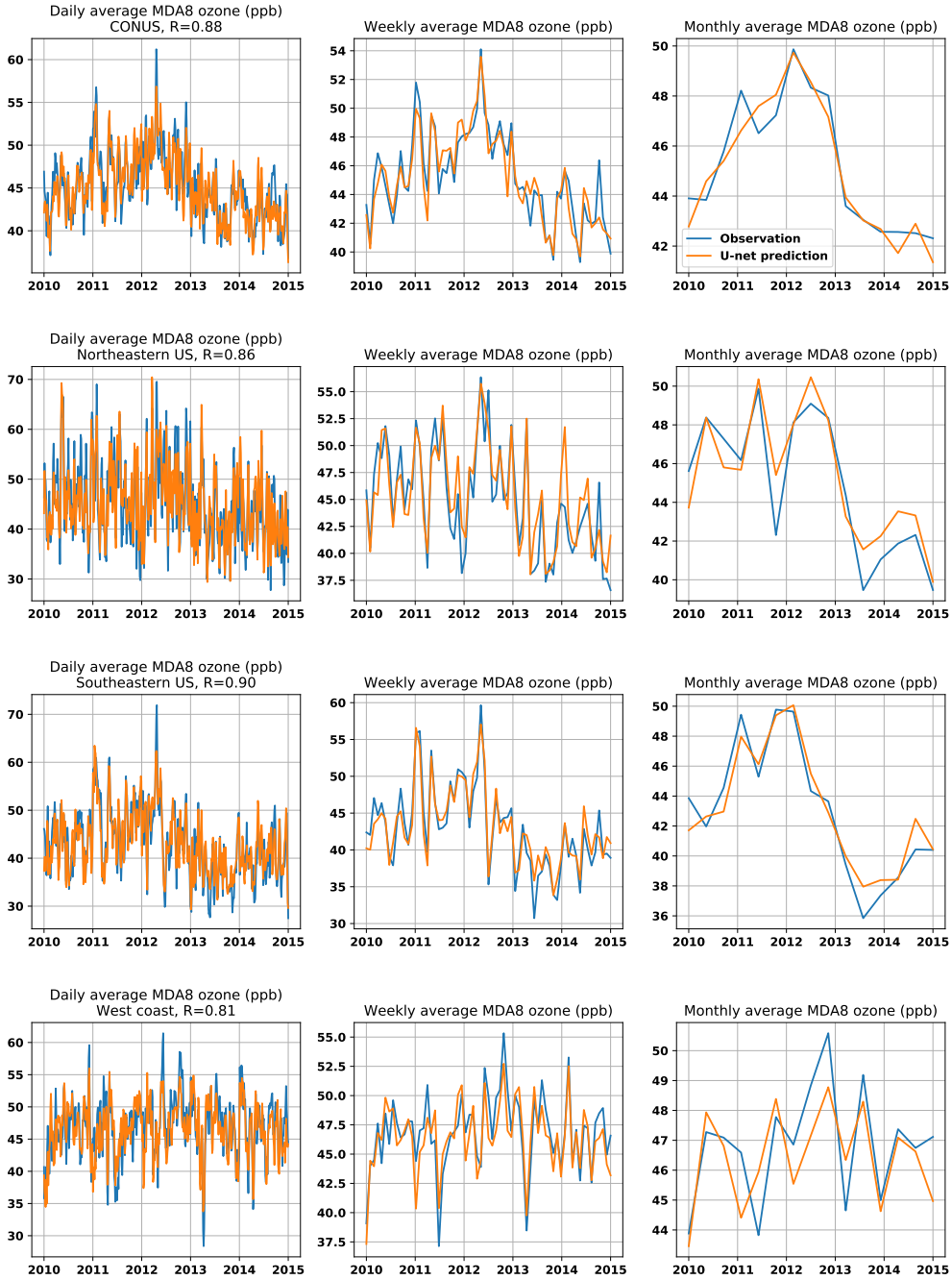

Figure 3. Observed (blue line) and predicted (orange line) daily (first column), 7-day averaged (second column), and monthly averaged (third column) JJA MDA8 ozone (in ppb) during the testing period (2010–2014). Shown are the time series for the CONUS (first row), the northeast (second row), the southeast (third row), and the west coast (last row). The regional definitions are shown in Figure 2.

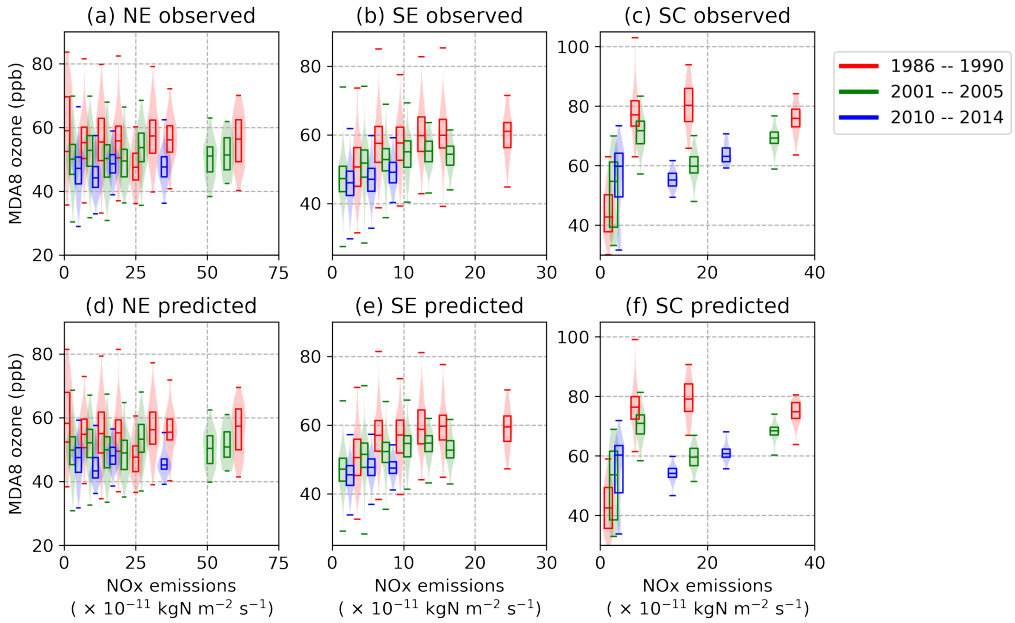

Figure 4. The changing relationship between MDA8 ozone and NO_x emissions for the three periods: 1986–1990 (red), 2001–2005 (green) and 2010–2014 (blue). Panels (a–c) show the relationship between the AQS ozone observations and NO_x emissions, whereas panels (d–f) show the relationship between the DL-predicted ozone and NO_x emissions. NO_x emissions and MDA8 ozone levels are smoothed by 4-day averaging windows. Note that the first two periods 1986–1990 and 2001–2005 are within the training data set, but 2010–2014 is not used in the training process. The three columns from left to right are results for the northeastern US (a and d), the southeastern US (b and e), and the southern California (c and f) regions.

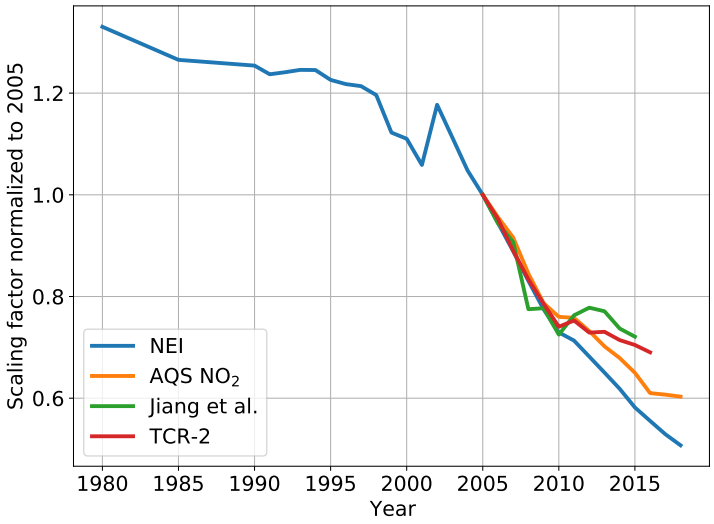

Figure 5. Relative change (normalized to 2005) in annual mean anthropogenic NO_x emissions for the United States from the bottom-up NEI inventory (blue line) and from the top-down inventories from TCR-2 (red line) and Jiang et al. (2018) (green line). Also shown is the trend in AQS NO_2 measurements (orange line).

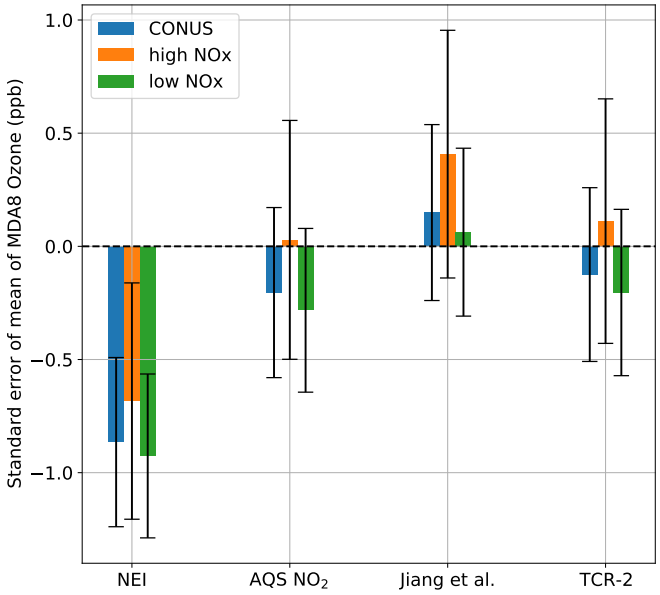

Figure 6. Mean error statistics for predicted MDA8 ozone for 2011–2014 for the CONUS (blue bars) and high-NO_x (orange bars) and low-NO_x (green bars) emission regions, based on NO_x emissions scaled by the NEI, AQS NO_2, TRC-2, and Jiang et al. (2018) trends. Error bars indicate the standard error on the mean.

References

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, *389*(10082), 1907 - 1918. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140673617305056 doi: https://doi.org/10.1016/S0140-6736(17)30505-6

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553-597. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828 doi: 10.1002/qj.828

Gardner, M., & Dorling, S. (2000). Meteorologically adjusted trends in uk daily maximum surface ozone concentrations. *Atmospheric Environment*, *34*(2), 171 - 176. Retrieved from http://www.sciencedirect.com/science/article/pii/S1352231099003155 doi: https://doi.org/10.1016/S1352-2310(99)00315-5

Gill, A. E. (1980). Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, *106*(449), 447-462. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49710644905 doi: 10.1002/qj.49710644905

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (http://www.deeplearningbook.org)

He, H., Liang, X.-Z., Sun, C., Tao, Z., & Tong, D. Q. (2020). The long-term trend and production sensitivity change in the us ozone pollution from observations and model simulations. *Atmospheric Chemistry and Physics*, *20*(5), 3191–3208. Retrieved from https://acp.copernicus.org/articles/20/3191/2020/ doi: 10.5194/acp-20-3191-2020

Hochreiter, S., & Schmidhuber, J. (1997, November). Long short-term memory. *Neural Comput.*, *9*(8), 1735–1780. Retrieved from http://dx.doi.org/10.1162/neco.1997.9.8.1735 doi: 10.1162/neco.1997.9.8.1735

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., ... Zhang, Q. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (ceds). *Geoscientific Model Development*, *11*(1), 369–408. Retrieved from https://www.geosci-model-dev.net/11/369/2018/ doi: 10.5194/gmd-11-369-2018

Jiang, Z., McDonald, B. C., Worden, H., Worden, J. R., Miyazaki, K., Qu, Z., ... Boersma, K. F. (2018). Unexpected slowdown of us pollutant emission reduction in the past decade. *Proceedings of the National Academy of Sciences*, *115*(20), 5099–5104. Retrieved from https://www.pnas.org/content/115/20/5099 doi: 10.1073/pnas.1801191115

Keller, C. A., & Evans, M. J. (2019). Application of random forest regression to the calculation of gas-phase chemistry within the geos-chem chemistry model v10. *Geoscientific Model Development*, *12*(3), 1209–1225. Retrieved from https://gmd.copernicus.org/articles/12/1209/2019/ doi: 10.5194/gmd-12-1209-2019

Kingma, D. P., & Ba, J. (2014, Dec). Adam: A Method for Stochastic Optimization. *arXiv e-prints*, arXiv:1412.6980.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989, Dec). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541-551. doi: 10.1162/neco.1989.1.4.541

Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). *Visualizing the loss*

landscape of neural nets.

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

Lin, M., Fiore, A. M., Horowitz, L. W., Langford, A. O., Oltmans, S. J., Tarasick, D., & Rieder, H. E. (2015, May 12). Climate variability modulates western us ozone air quality in spring via deep stratospheric intrusions. *Nature Communications*, *6*, 7105 EP -. Retrieved from https://doi.org/10.1038/ncomms8105 (Article)

Miyazaki, K., Bowman, K., Sekiya, T., Eskes, H., Boersma, F., Worden, H., ... Ogochi, K. (2020a). An updated tropospheric chemistry reanalysis and emission estimates, tcr-2, for 2005–2018. *Earth System Science Data Discussions*, *2020*, 1–64. Retrieved from https://essd.copernicus.org/preprints/essd-2020-30/ doi: 10.5194/essd-2020-30

Miyazaki, K., Bowman, K., Yumimoto, K., Walker, T., & Sudo, K. (2020b). Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis. *Atmospheric Chemistry and Physics*, *20*(2), 931–967. Retrieved from https://acp.copernicus.org/articles/20/931/2020/ doi: 10.5194/acp-20-931-2020

Miyazaki, K., Eskes, H. J., & Sudo, K. (2015). A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of omi, mls, tes, and mopitt satellite data. *Atmospheric Chemistry and Physics*, *15*(14), 8315–8348. Retrieved from https://acp.copernicus.org/articles/15/8315/2015/ doi: 10.5194/acp-15-8315-2015

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*(7743), 195-204. Retrieved from https://doi.org/10.1038/s41586-019-0912-1 doi: 10.1038/s41586-019-0912-1

Reidmiller, D. R., Fiore, A. M., Jaffe, D. A., Bergmann, D., Cuvelier, C., Dentener, F. J., ... Zuber, A. (2009). The influence of foreign vs. north american emissions on surface ozone in the us. *Atmospheric Chemistry and Physics*, *9*(14), 5027–5042. Retrieved from https://www.atmos-chem-phys.net/9/5027/2009/ doi: 10.5194/acp-9-5027-2009

Ronneberger, O., Fischer, P., & Brox, T. (2015, May). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, arXiv:1505.04597.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536. Retrieved from https://doi.org/10.1038/323533a0 doi: 10.1038/323533a0

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Neurocomputing: Foundations of research. In J. A. Anderson & E. Rosenfeld (Eds.), (pp. 673–695). Cambridge, MA, USA: MIT Press. Retrieved from http://dl.acm.org/citation.cfm?id=65669.104449

Seltzer, K. M., Shindell, D. T., Kasibhatla, P., & Malley, C. S. (2020). Magnitude, trends, and impacts of ambient long-term ozone exposure in the united states from 2000 to 2015. *Atmospheric Chemistry and Physics*, *20*(3), 1757–1775. Retrieved from https://acp.copernicus.org/articles/20/1757/2020/ doi: 10.5194/acp-20-1757-2020

Shen, L., & Mickley, L. J. (2017). Seasonal prediction of us summertime ozone using statistical analysis of large scale climate patterns. *Proceedings of the National Academy of Sciences*, *114*(10), 2491–2496. Retrieved from https://www.pnas.org/content/114/10/2491 doi: 10.1073/pnas.1610708114

Shen, L., Mickley, L. J., & Tai, A. P. K. (2015). Influence of synoptic pat-

terns on surface ozone variability over the eastern united states from 1980 to 2012. *Atmospheric Chemistry and Physics*, *15*(19), 10925–10938. Retrieved from https://www.atmos-chem-phys.net/15/10925/2015/ doi: 10.5194/acp-15-10925-2015

Silvern, R. F., Jacob, D. J., Mickley, L. J., Sulprizio, M. P., Travis, K. R., Marais, E. A., ... Lamsal, L. N. (2019). Using satellite observations of tropospheric no2 columns to infer long-term trends in us nox emissions: the importance of accounting for the free tropospheric no2 background. *Atmospheric Chemistry and Physics*, *19*(13), 8863–8878. Retrieved from https://www.atmos-chem-phys.net/19/8863/2019/ doi: 10.5194/acp-19-8863-2019

Sutton, R. T., & Hodson, D. L. R. (2005). Atlantic ocean forcing of north american and european summer climate. *Science*, *309*(5731), 115–118. Retrieved from https://science.sciencemag.org/content/309/5731/115 doi: 10.1126/science.1109496

Sutton, R. T., & Hodson, D. L. R. (2007). Climate response to basin-scale warming and cooling of the north atlantic ocean. *Journal of Climate*, *20*(5), 891-907. Retrieved from https://doi.org/10.1175/JCLI4038.1 doi: 10.1175/JCLI4038.1

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., ... Zhou, X. (2016). Why do models overestimate surface ozone in the southeast united states? *Atmospheric Chemistry and Physics*, *16*(21), 13561–13577. Retrieved from https://www.atmos-chem-phys.net/16/13561/2016/ doi: 10.5194/acp-16-13561-2016

Zhang, L., Jacob, D. J., Yue, X., Downey, N. V., Wood, D. A., & Blewitt, D. (2014). Sources contributing to background surface ozone in the us intermountain west. *Atmospheric Chemistry and Physics*, *14*(11), 5295–5309. Retrieved from https://www.atmos-chem-phys.net/14/5295/2014/ doi: 10.5194/acp-14-5295-2014

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

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

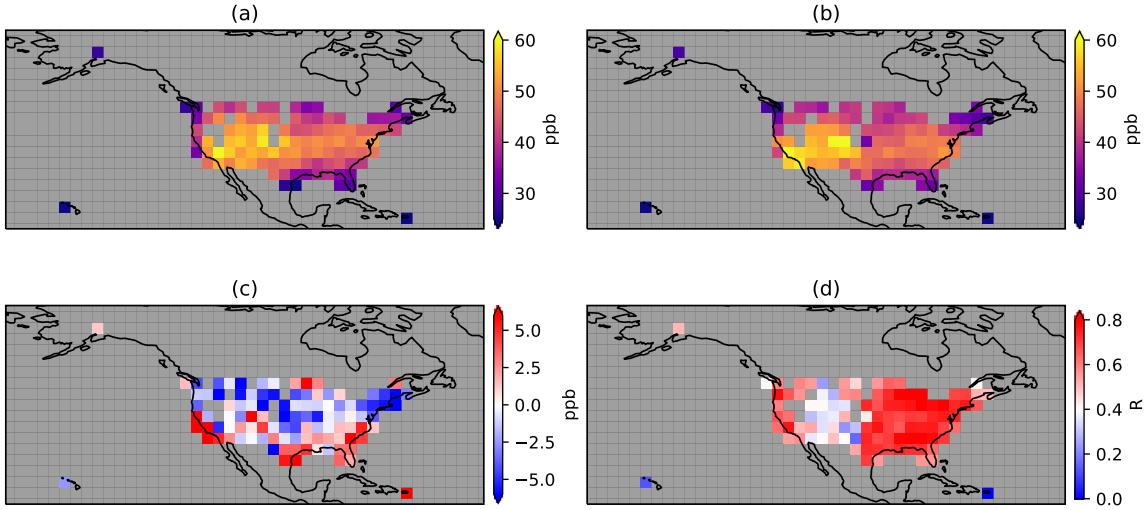

Figure S1. Observed (blue line) and predicted (orange line) daily (first column), 7-day averaged (second column), and monthly averaged (third column) JJA MDA8 ozone (in ppb) during the testing period (2010–2014) at a resolution of $3° \times 3°$.

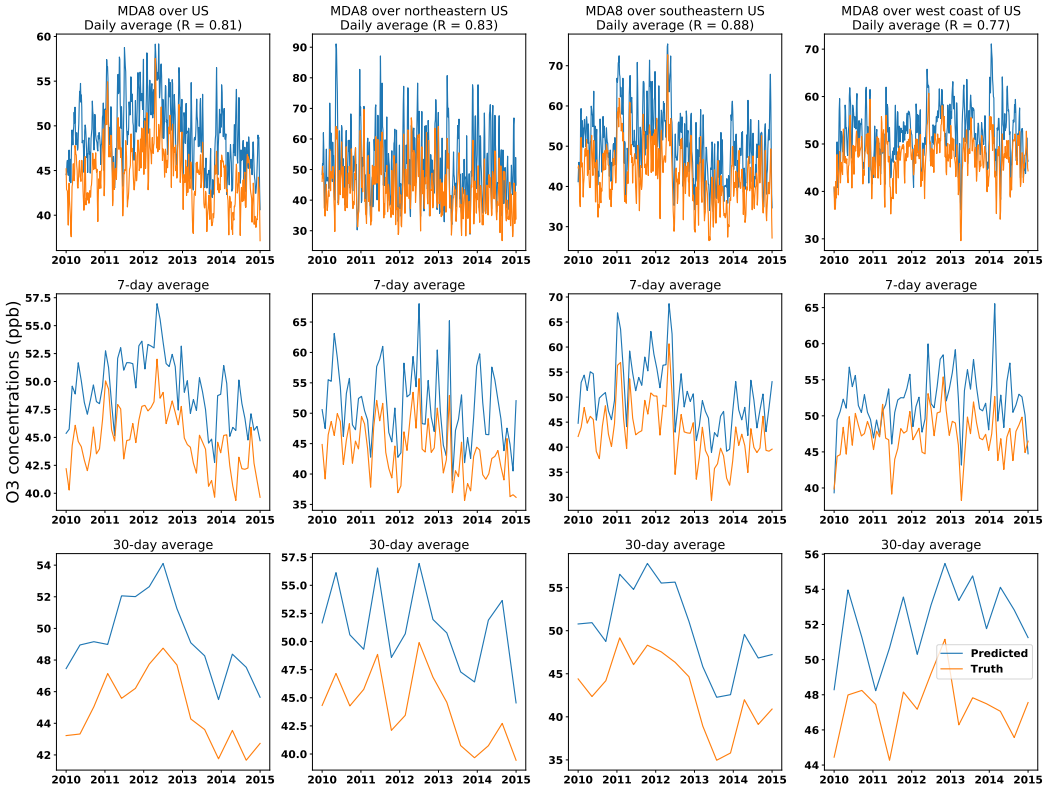

Figure S2. Observed (blue line) and predicted (orange line) daily (first row), 7-day averaged (second row), and monthly averaged (third row) JJA MDA8 ozone (in ppb) during 2010–2014 with only meteorological predictors. Shown are the time series for the CONUS (first column), the northeast (second column),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

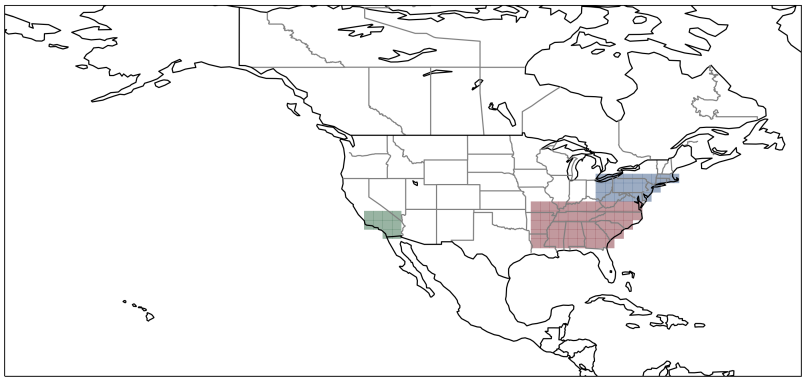

Figure S3. Domains for the Northeast, Southeast, and the southern California regions, which are indicated by the boxes shaded in blue, red and green, respectively. The domains are used for the ozone-NO_x relationship analysis in Figure 4 in the main text.

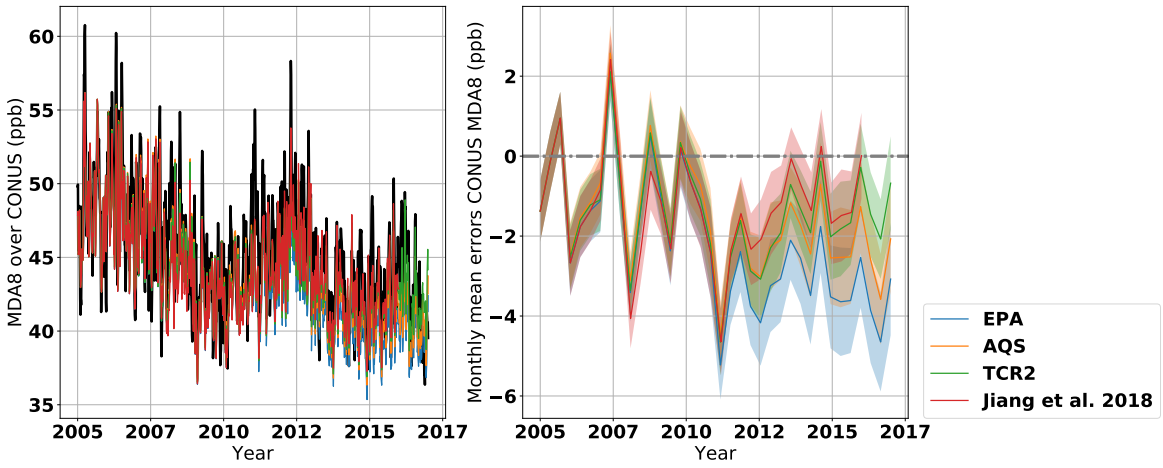

Figure S4. Observed and predicted daily mean (left) and monthly mean errors (right) of MDA8 ozone between 2005–2016 (2005–2015 for Jiang et al.). Shown are the AQS ozone observations (black line) and the model predictions based on the NO_x emissions scaled by the EPA (blue line), AQS NO_2 (orange line),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

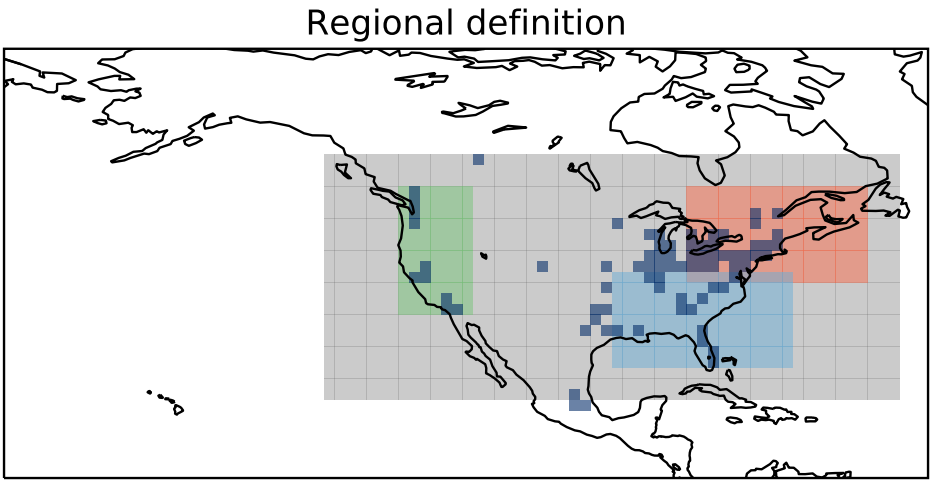

Figure S5. The spatial distribution of high- and low-NO_x emission regions. High-NO_x regions are indicated as dark blue grid cells. Also shown are the domains for the Northeast, Southeast, and West Coast regions, which are indicated by the boxes shaded in red, blue and green, respectively. The CONUS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