The hydroxyl radical (OH) is the most important oxidant on global and local scales in the troposphere. Urban OH controls the removal rate of primary pollutants and triggers the production of ozone. Intertannual trends of OH in urban areas are not well documented or understood due to the short lifetime and high spatial heterogeneity of OH. We utilize machine learning with observational inputs emphasizing satellite remote sensing observations to predict surface OH in 49 North American cities from 2005 to 2014. We observe changes in the summertime OH over one decade, with wide variation among different cities. In 2014, compared to the summertime OH in 2005, 3 cities show a significant increase of OH, whereas, in 27 cities, OH decreases in 2014. The year-to-year variation of OH is mapped to the decline of the NOx column. We conclude that these cities in this analysis are either in the NOx-limited regime or at the transition from a NOx suppressed regime to a NOx-limited regime. The result emphasizes that, in the future, controlling NOx emissions will be most effective in regulating the ozone pollution in these cities.

hydroxyl radical | ozone control | NOx emission

The hydroxyl radical (OH) is the dominant oxidant in the atmosphere. Reactions with OH initiate the chemistry of a variety of trace gases (1–3), including greenhouse gases such as methane (CH4) and primary pollutants including carbon monoxide (CO), nitrogen oxides (NOx = NO+NO2), and volatile organic compounds (VOCs). As a result, OH affects radiative forcing over a global scale and governs local pollution events over cities. In cities, the lifetime of the freshly emitted species is anticorrelated with the OH concentration (4). Laughner and Cohen (5) observe NOx lifetime directly from space and conclude that significant shifts in the NOx lifetime occurred between 2005 and 2014 among 30 North American cities, which, in turn, suggests an underlying variation in OH in these cities.

Moreover, OH defines the production of ozone (O3) (6, 7). Ground-level ozone is a major secondary pollutant that leads to negative impacts on human health and triggers the risk of death from respiratory causes (8). Ozone exceeds health-based standards most often during summertime. Ozone results from reactions involving two predominant ozone precursors: NOx and VOCs. In the presence of sunlight, OH oxidizes VOCs to form organic peroxy radicals (RO2), and then RO2 can react with NO to form NO2 which photolyzes to produce O3. Shown in Fig. 1, the ozone production rate P(O3) is a nonlinear function of both NOx and VOCs, and its dependence on NOx is similar to the dependence of OH on NOx. Two distinct regimes are observed. At high VOC reactivity and low NOx, both P(O3) and OH increase with enhancing NOx, and the corresponding regime is identified by the limiting reagent as a NOx-limited regime. In contrast, both P(O3) and OH show the opposite relationship with NOx in the NOx-saturated regime, which is characterized by high NOx compared to VOC reactivity. In terms of reducing ozone pollution, the effectiveness of the emissions control strategy depends on whether the photochemical regime of ozone formation is a NOx-saturated or NOx-limited regime. Therefore, OH is the ideal target molecule to examine the chemical regime; the relationship between OH and NOx at interannual time scales serves as an explicit proxy to monitor the shift of the chemical regimes.

Despite its pivotal role in oxidation chemistry, OH observations at urban or suburban areas are remarkably sparse. The longest record is reported in ref. 9 where continuous OH measurements were conducted at a rural site in southern Germany between 1999 and 2003. Other OH measurements, occurring either in situ or airborne, are limited to a shorter time scale ranging from days to months (10–17). The short lifetime (<1 s) of OH, owing to the high reactivity, means local in situ chemistry dominates the OH budget. As a result, the high spatial heterogeneity makes it impossible to describe OH chemistry over a large spatial scale solely from in situ OH observations.

In a recent paper, we developed a machine learning (ML) model to represent the OH chemistry emphasizing urban areas (18). The model aims to represent the OH chemistry simulated from a state-of-the-art chemical transport model relying on a
Fig. 1. A schematic of OH and P(O3) as a function of NO\textsubscript{x} at the photochemical steady state. We assume a NO\textsubscript{2}/NO ratio of 4, an alkyl nitrate branching ratio $\alpha$ of 0.04, and a HO\textsubscript{x} production rate of 0.3 ppt $\times$ s$^{-1}$.

Two scenarios represent a high VOC reactivity (VOC\textsubscript{r} = 10 s$^{-1}$) condition and low VOC reactivity condition (VOC\textsubscript{r} = 1 s$^{-1}$), respectively.

small set of observed constraints that are available with records over one decade. The ML model with observational inputs from satellite remote sensing yields estimates of surface OH across 49 North American cities for the time period of 2005–2014.

**Variation of Urban OH over One Decade**

The 49 selected cities exhibit a full range of variation in OH chemistry as characterized by the HCHO columns and NO\textsubscript{2} columns. Both HCHO columns and NO\textsubscript{2} columns are acquired from satellite-based observations and serve as indicators for VOCs (19–21) and NO\textsubscript{x} (22, 23), respectively. We use the ratio of the HCHO column and the NO\textsubscript{2} column (satellite HCHO/NO\textsubscript{2}) to reflect the relative availability of NO\textsubscript{x} and total organic reactivity to hydroxyl radicals (24).

Fig. 2 shows the ratio of average HCHO and NO\textsubscript{2} columns during April to September between 2005 and 2014 over selected cities. The satellite HCHO/NO\textsubscript{2} varies by a factor of 3, reflecting a range in NO\textsubscript{x} and VOC chemistry. High satellite HCHO/NO\textsubscript{2} is indicative of an abundance of VOC relative to NO\textsubscript{x} and presents in southeast US cities. It is consistent with VOC emissions mostly from biogenic sources, for instance, isoprene from vegetation (25).

| A | B |
|---|---|
| ![Map of satellite-based summertime average of HCHO/NO\textsubscript{2} between 2005 and 2014 over 49 North American cities.](https://doi.org/10.1073/pnas.2117399119) | ![Frequency distribution of satellite HCHO/NO\textsubscript{2}.](https://doi.org/10.1073/pnas.2117399119) |

Fig. 2. The satellite HCHO/NO\textsubscript{2} over selected 49 cities. (A) The map of satellite-based summertime average of HCHO/NO\textsubscript{2} between 2005 and 2014 over 49 North American cities. (B) The frequency distribution of satellite HCHO/NO\textsubscript{2}.
cities along the West Coast and the Northwest United States exhibit a relatively low satellite HCHO/NO\textsubscript{2} ratio. Those cities feature large anthropogenic NO\textsubscript{x} emissions, whereas the VOC emissions are unreactive compared to the Southeast United States (26).

It also worth emphasizing the variation of the NO\textsubscript{2} and HCHO columns over one decade. Anthropogenic NO\textsubscript{x} emissions exhibited a significant reduction over the United States due to stringent air pollution regulations during the 2005–2014 period. This is reflected in the decline in the satellite-observed NO\textsubscript{2} column. The selected cities include examples with the largest NO\textsubscript{2} column as well as the most pronounced reduction (SI Appendix, Fig. S1). The observed NO\textsubscript{2} column decreases by 33%, on average, over this time period. The largest decline of NO\textsubscript{2} column by up to 55% is observed in Los Angeles, followed by other populous cities including Chicago, Boston, and New York. Austin observes the smallest decrease, 11%. We also note the smaller decrease of the NO\textsubscript{2} column over urban regions in the central United States, which is thought to be partially due to the offset of increasing soil NO\textsubscript{2} emission (27). Compared to the NO\textsubscript{2} column, HCHO columns show no consistent interannual variation (SI Appendix, Fig. S2). Among the selected cities, the relative change of HCHO column between 2005 and 2014 ranges from −15 to +10%. The causes of 10-y variation in the HCHO column are complex. While the anthropogenic VOC emissions from vehicles and industry declined (28), volatile chemical product emissions may be growing (29). Also, the changes in anthropogenic VOC emissions are masked by the large HCHO background driven by biogenic sources (30).

The interannual variation in both the NO\textsubscript{2} column and the HCHO column leads to variation in OH. Compared to summertime in 2005, the relative change of urban OH in 2014 ranges from −17 to +11% over the 49 cities (Fig. 3). We find that 30 cities show statistically significant changes in surface OH between 2005 and 2014. In four cities, including Los Angeles, New Orleans, New York, and Toronto, the increase of summertime averaged OH is larger than 4%. Thirty out of 49 cities witness a decline of annual OH in 2014 larger than −4%. Uncertainties are

Fig. 3. The OH variation over selected 49 cities. (A) The map of the relative difference (Rel. diff) of summertime OH between 2005 and 2014 over 49 North American cities, and (B) the frequency distribution.
described in Materials and Methods. The full OH trends between 2005 and 2014 for each city are shown in SI Appendix, Fig. S3.

The Relationship between OH and NO₂ Column
Mapping the predicted OH to the concurrent NO₂ column depicts the relationship between OH and NO₂, and it reveals which chemical regime the urban environment falls in. For each city, we select the areas covering both city center and surrounding areas. If we assume that VOC reactivity is similar in all grids for a given urban area for a given year, the relationship between OH concentration and NO₂ column provides information about the dominant chemical regime. Assuming minimal variation in the VOC reactivity is supported by the small annual variation in HCHO shown in SI Appendix, Fig. S2 and similar weekend/weekday VOC reactivity observed in, for example, refs. 31–33.

Recall from Fig. 1 that the dependence of OH on NOₓ is nonlinear, and the slope of OH vs. NOₓ varies with its chemical regime. In the NOₓ-saturated regime, OH is inversely correlated with NOₓ, yielding a negative slope. The turnover point between the NOₓ-saturated regime and NOₓ-limited regime leads to a slope fluctuating around the zero point. Conversely, in the NOₓ-limited regime, the slope is positive. The farther away from the turnover point, the larger the slope.

Fig. 4 shows how two cities, Los Angeles and Denver, illustrate different relationships between OH and NO₂ column and how it has shifted over the years. In 2005, Los Angeles was characterized by high NOₓ (34). The average NO₂ column was \(9.2 \times 10^{15}\) molecules per cm², and the largest NO₂ column was \(2.4 \times 10^{16}\) molecules per cm². The OH from the ML model also shows a wide spread within the city, where OH concentration varies by a factor of 2. We note a significant inverse relationship; a higher NO₂ column corresponds to a lower OH concentration, yielding a slope of \(-0.1\) (Fig. 4A). The observed slope matches the theoretical pattern of a NOₓ-saturated regime as shown in Fig. 1. In contrast, Los Angeles in 2014 presented a much lower NO₂ column. The average NO₂ column is 60% lower than in 2005. Compared to the remarkable decline in the NO₂ column, the OH concentration shows a slight enhancement, 10% higher, on average, than those in 2005. As a consequence, the negative correlation between OH and the NO₂ column in 2005 vanishes.

\[\text{Fig. 4. The density plot between OH and the NO}_2\text{ column in 2005 and 2014 over (A) Los Angeles and (B) Denver. For each city, both OH and the NO}_2\text{ column are normalized to their annual averages in 2005. The lines and corresponding slopes denote the linear regression results between normalized OH and the normalized (Norm.) NO}_2\text{ column.}\]
In 2014, OH presented a strong nonlinear dependence on the NO$_2$ column. Although a linear regression fails to describe the full pattern between OH and the NO$_2$ column, the slope from linear fitting shows a large fluctuation near the zero point, indicative of the turnover point between the NO$_2$-saturated regime and the NO$_2$-limited regime. Expanding the relationship analysis of OH and NO$_2$ column to each year between 2005 and 2014 shows a continuous increase of the slope, indicating a consistent transition from a NO$_2$-saturated regime (SI Appendix, Fig. S4).

In contrast, the OH–NO$_2$ column relationship in Denver indicates that NO$_2$-limited chemistry dominated the Denver plume between 2005 and 2014 (SI Appendix, Fig. S5), and Fig. 4B shows the OH–NO$_2$ column relationship in 2005 and 2014. In 2005, the NO$_2$ column was moderate compared to Los Angeles; the average NO$_2$ column was 2.8 × 10$^{13}$ molecules per cm$^2$, and the largest NO$_2$ column was 7.2 × 10$^{15}$ molecules per cm$^2$. A positive correlation between OH and the NO$_2$ column is observed, with a slope of 0.30. The slope of the correlation suggests Denver was already NO$_2$ limited in 2005. In 2014, on average, there was a 16% decrease in the NO$_2$ column. However, the peak NO$_2$ column decreased by 36%, reflecting the effective emission control near the city center. The positive correlation between OH and NO$_2$ column continues in 2014, with its fitted slope increasing from 0.30 to 0.36. The enhancement in slope is consistent with, in 2014, moving farther away from the turnover point, along with the decline in the NO$_2$ column.

We extend the investigation of the relationship between OH and NO$_2$ column to 49 cities. In both 2005 and 2014, we conduct the linear regression between normalized OH and normalized NO$_2$ column; these slopes are summarized in Fig. 5, and the slopes for all years between 2005 and 2014 are shown in SI Appendix, Fig. S6. The year 2005 sees a wide scatter of the slope between normalized OH and normalized NO$_2$ column. Five cities, including New Orleans, Los Angeles, Tampa, and Boston, present a negative relationship between OH and NO$_2$ column. Sixteen cities yield a slope fluctuating around zero and a correlation coefficient ($r^2$) lower than 0.6. The rest of the cities observe strong positive correlations ($r^2 > 0.6$), with the slope larger than 0.2.

In 2014, the slopes in 45 out of 49 cities are more positive than they were in 2005. Cities such as Los Angeles have a negative slope closer to zero, indicating a transition from the NO$_2$-saturated regime to the turnover point of two regimes. Most cities whose slope is positive in 2005 increase further, demonstrating a shift within the NO$_2$-limited regime.

### Implication for Controlling Ozone Pollution

In 2021, 50 areas and 196 counties are designated as ozone nonattainment areas where 8-h ozone levels fail to meet the National Ambient Air Quality Standards, affecting over 120 million people (35). Efforts devoted to reducing high ozone have focused on reduction of O$_3$ precursors emitted by anthropogenic sources. NO$_x$ and VOC emission reductions have occurred (26, 36). However, the effectiveness of emission control on lowering ozone level encompasses a large variation and is associated with the chemical regime of the urban environment. In NO$_2$-saturated regime, NO$_x$ emission reductions are detrimental to the mitigation of ozone pollution in the short term. In the NO$_2$-limited regime, there are immediate benefits. Based on our conclusion of the chemical regime shifts, most North American cities were NO$_2$ limited in 2014. If further controls on NO$_2$ emission are prioritized for regulating ozone, the regulations will likely be effective.

### Conclusion

We leverage ML combined with satellite observations, to estimate surface OH in 49 cities. We found changes in the summertime OH between 2005 and 2014, ranging from a decrease of 17% to an enhancement of 11%. We observe a shift of chemical regimes, with a NO$_2$-limited regime now in effect in most cities. Thus, continued reduction of NO$_x$ emissions will effectively control ozone.

### Materials and Methods

The analysis code is available at https://doi.org/10.5281/zenodo.5296044 (37), and intermediate datasets are available at https://doi.org/10.6078/D1FM75 (38).

**Observational and Model Records Used.** All observational inputs are publicly available. We use observations from the Ozone Monitoring Instrument (OMI). The NO$_2$ retrievals are from Version 3.08 of the Berkeley High Resolution (BEHR) OMI NO$_2$ product (https://behr.cchem.berkeley.edu/home/), and HCHO retrievals are from the Quality Assurance for Essential Climate Variables project.
**ML Model.** We utilize a gradient boosted tree model to represent the OH chemistry over urban areas during summertime. It is a supervised ML model and is optimized for predicting the surface OH by learning the training data.

We prepare training data solely from a chemical transport model, WRF-Chem. We use six parameters as predictors, including four surface variables (\( \text{JO}^0\text{D}, \text{H}_2\text{O}, \) temperature, and pressure) and two column variables (\( \text{NO}_2 \) column and \( \text{HCHO} \) column). The calculation of two column variables incorporates the averaging kernel from the satellite products to emulate the quantities observed from space, referring to section 7.4 of Lauhner et al. (39). We select the WRF-Chem grids covering 49 North American cities, and each grid has a spatial scale of 12 km × 12 km. To represent each city, we select grid cells in a circular area with the radius varying from 0.5° to 1° around the city center based on the city size and the surrounding interference. The selection of 49 cities is consistent with Russell et al. (36) and Lauhner and Cohen (5). We constrain the time to ~13:30 local hour from April through September between 2005 and 2014 to present summertime conditions where the ozone pollution due to active photochemistry is most of a concern. For each city, we only filter out the days with lightning occurrence and with cloud fraction larger than 0.2.

Zhu et al. (18) evaluate the accuracy of this ML model and prove that the ML model using these six parameters is capable of capturing 76% of the OH variability and yields an RMSE of 2.1 × 10⁶ molecules per cm². Therefore, the ML model serves as an efficient alternative to OH simulation using the computationally expensive chemical transport model.

**OH Predictions.** We combine model simulations and satellite observations to construct the observation-based inputs, and use the ML model to predict OH. Among the six features, we take \( \text{JO}^0\text{D}, \text{H}_2\text{O}, \) temperature, and pressure from the WRF-Chem outputs. Since meteorological parameters are constrained by the North American Regional Reanalysis every 3 h, we expect that these parameters are in good agreement with the observations. We use satellite retrievals described above to obtain both \( \text{NO}_2 \) column and \( \text{HCHO} \) column features. Based on the availability of OMI satellite observations and BEHR retrieval products that are optimized to remove bias on the spatial resolution of cities, we confine our study to 8 y between 2005 and 2014. OH predictions for years 2006 and 2010 are absent due to lack of WRF-Chem or BEHR retrievals.

Zhu et al. (18) calculate the uncertainty of the OH predictions. Besides those inherent in the ML model, the uncertainties from observed input parameters, especially the \( \text{NO}_2 \) column and the \( \text{HCHO} \) column, propagate through the ML model configuration and lead to nonnegligible influence on model performance. Considering the uncertainties of the retrievals products for both the \( \text{NO}_2 \) column and the \( \text{HCHO} \) column, we estimate the RMSE of OH predictions as 3.2 × 10⁶ molecules per cm².

In order to compare our model to previous studies, we present the summertime average OH (Eq. 1).

\[
\text{OH}_{\text{sum}} = \frac{\sum_{\text{day}, \text{city}} \sum_{\text{grid}, \text{city}} \text{OH}_{\text{day}, \text{grid}, \text{city}}}{\sum_{\text{day}, \text{city}} \sum_{\text{grid}, \text{city}}} .
\]

The number of OH predictions used for calculation varies, as it depends on the number of grids for each city region as well as the number of valid days. The number of individual grid OH predictions ranges between 1,000 and 16,000. The minimum number corresponds to summertime OH in 2014 over Orlando, FL, which is consistent with a relatively small city size and fewer valid days in the analysis after filtering the days with lightning occurrence. We treat each OH prediction as an independent variable and use the lower bound of the number of observations; the RMSE of summertime average OH is 1.0 × 10⁶ molecules per cm². It is a conservative estimate to compensate for the fact that the OH predictions are not perfectly orthogonal.

We then, for each city, assess the relative difference of summertime average OH between 2005 and 2014 (rel. OH hereinafter). The summertime average OH across all cities ranges from 8.7 × 10⁶ molecules per cm² to 1.33 × 10⁷ molecules per cm² in 2005, and ranges from 8.2 × 10⁶ molecules per cm² to 1.25 × 10⁷ molecules per cm² in 2014. Considering the estimate of RMSE of 1.0 × 10⁶ molecules per cm², we calculate that the upper bound of the 5D of rel. OH is 2%. Therefore, we take 4% as a threshold in the analysis; a rel. OH larger than 4% is considered statistically significant with 95% confidence.

**Relationship between OH and \( \text{NO}_2 \) Column.** To reconcile the large city-wise variation in both the OH and the \( \text{NO}_2 \) column, both OH and the \( \text{NO}_2 \) column are normalized to their 2005 averages for a given city. We apply linear regressions to fit the normalized OH to the normalized \( \text{NO}_2 \) column and collect the slopes from the linear fitting. The shifts in slopes between OH and the \( \text{NO}_2 \) column are discussed in the Relationship between \OH and \text{NO}_2 Column.

**Data Availability.** Zip data have been deposited in “Supporting data for Estimate of OH trends over one Decade in North American cities” (https://doi.org/10.6078/D1FM75). The analysis code is available at Zenodo, https://doi.org/10.5281/zenodo.5296044.

**ACKNOWLEDGMENTS.** We acknowledge support from NASA Grant 80NSSC19K0945. We acknowledge high-performance computing support from the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley and Chief Information Officer. We acknowledge use of the WRF-Chem preprocessor and the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer. We acknowledge use of the WRF-Chem model configuration with the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley and Chief Information Officer. We acknowledge use of the WRF-Chem library and the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley and Chief Information Officer.
25. A. Guenther et al., Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). Atmos. Chem. Phys. 6, 3181–3210 (2006).

26. S. Kharol et al., Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America. Atmos. Environ. 118, 236–245 (2015).

27. G. Vinken, K. Boersma, J. Maasakkers, M. Adon, R. Martin, Worldwide biogenic soil NO2 emissions inferred from OMI NO2 observations. Atmos. Chem. Phys. 14, 10363–10381 (2014).

28. US Environmental Protection Agency. Air pollutant emissions trends data 2016. https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data. Accessed 23 March 2022.

29. B. C. McDonald et al., Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Science 359, 760–764 (2018).

30. L. Zhu et al., Long-term (2005–2014) trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing emissions of volatile organic compounds. Geophys. Res. Lett. 44, 7079–7086 (2017).

31. S. Berle, U. Platt, M. Weng, T. Wagner, Weekly cycle of NOx by GOME measurements: A signature of anthropogenic sources. Atmos. Chem. Phys. 3, 2225–2232 (2003).

32. B. Kaynak, Y. Hu, R. Martin, C. Sioris, A. Russell, Comparison of weekly cycle of NOx satellite retrievals and NOx emission inventories for the continental United States. J. Geophys. Res. Atmospheres 114, D05302 (2009).

33. A. R. Russell, L. C. Valin, E. J. Bucsela, M. O. Wenig, R. C. Cohen, Space-based constraints on spatial and temporal patterns of NOx emissions in California, 2005–2008. Environ. Sci. Technol. 44, 3608–3615 (2010).

34. R. Gottlieb, R. Freer, M. Vallianatos, P. Dreier, The Next Los Angeles: The Struggle for a Livable City (University of California Press, 2006).

35. US Environmental Protection Agency. 8-hour ozone non-attainment areas (2015 standard). https://www3.epa.gov/airquality/greenbook/jbtc.html. Accessed 15 August 2021.

36. A. R. Russell, L. C. Valin, R. C. Cohen, Trends in OMI NO2 observations over the United States: Effects of emission control technology and the economic recession. Atmos. Chem. Phys. 12, 12197–12209 (2012).

37. Q. Zhu, qdzhu/ml-pred-urban-oh: First release. Zenodo. https://zenodo.org/record/5296044#.YjtxRDMKqU. Deposited 27 August 2021.

38. Q. Zhu, J. L., Laughner, R. C. Cohen, Supporting data for “Direct observation of changing NOx lifetime in North American cities.” Dryad. https://datadryad.org/stash/share/o60CWC0bly2nWz26b713sMwa1PiMh7mNh743N76Dwr. Accessed 23 March 2022.

39. J. L. Laughner, Q. Zhu, R. C. Cohen, The Berkeley high resolution tropospheric NO2 product. Earth Syst. Sci. Data 10, 2069–2095 (2018).