# From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model

Jiani Yang[a,1], Yifan Wen[b,1], Yuan Wang[a,c,2], Shaojun Zhang[b,2], Joseph P. Pinto[d], Elyse A. Pennington[e], Zhou Wang[f], Ye Wu[b], Stanley P. Sander[c], Jonathan H. Jiang[c], Jiming Hao[b], Yuk L. Yung[a,c], and John H. Seinfeld[e,2]

[a]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125; [b]School of Environment, Tsinghua University, Beijing 100084, China; [c]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109; [d]Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [e]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and [f]Department of Geography, University of Mainz, 55099 Mainz, Germany

Contributed by John H. Seinfeld, May 3, 2021 (sent for review February 12, 2021; reviewed by Russell R. Dickerson and Alma Hodzic)

**The large fluctuations in traffic during the COVID-19 pandemic provide an unparalleled opportunity to assess vehicle emission control efficacy. Here we develop a random-forest regression model, based on the large volume of real-time observational data during COVID-19, to predict surface-level $NO_2$, $O_3$, and fine particle concentration in the Los Angeles megacity. Our model exhibits high fidelity in reproducing pollutant concentrations in the Los Angeles Basin and identifies major factors controlling each species. During the strictest lockdown period, traffic reduction led to decreases in $NO_2$ and particulate matter with aerodynamic diameters <2.5 μm by –30.1% and –17.5%, respectively, but a 5.7% increase in $O_3$. Heavy-duty truck emissions contribute primarily to these variations. Future traffic-emission controls are estimated to impose similar effects as observed during the COVID-19 lockdown, but with smaller magnitude. Vehicular electrification will achieve further alleviation of $NO_2$ levels.**

COVID-19 | machine learning | air pollution | traffic emissions | vehicular electrification

In the urban environment, vehicular traffic is a principal source of air pollutants, including nitrogen oxides ($NO_x$ = NO + $NO_2$), carbon monoxide (CO), and carbonaceous particles. Secondary ozone ($O_3$) and particulate matter (PM) have adverse impacts on human health (1) by inducing dysfunction and deterioration of cardiovascular, respiratory, and immune systems (2). The COVID-19 pandemic led to unprecedented decreases in traffic-related emissions in megacities worldwide (3–5). Owing to the short chemical lifetime of $NO_x$ and the pandemic-induced emission changes, the well-defined and abrupt decrease in $NO_2$ has been captured by satellites as well as ground-based observations (6–8). However, changes in secondary pollutants like $O_3$ and a major portion of $PM_{2.5}$ (PM with aerodynamic diameters <2.5 μm) during the pandemic were diverse in different regions (7, 9), for which the major drivers remain unclear. Atmospheric chemical reactions serve as essential nonlinear links between emissions and atmospheric composition. Moreover, local meteorological factors, such as air temperature, humidity, radiation, and clouds, also strongly regulate photochemical formation of ozone and multiphase chemistry of secondary PM (6, 9–11). The response of secondary pollutants to COVID-19–induced emission changes remains poorly understood; existing studies provide limited insight into the consequent chemistry (7). Here, we disentangle the complex factors involving emissions, chemical reactions, pollutant transport, and meteorology to evaluate the effect of pandemic-induced or other dramatic emission changes on air quality.

Los Angeles (LA) has long been one of the most polluted cities in the United States (12). Surrounded by mountains on three sides and bounded by the Pacific Ocean, ideal conditions exist for pollutant buildup over the LA Basin and downwind areas (13, 14). Owing to the strict sulfur oxides ($SO_x$) emission control program established in 1978 and major improvements of motor vehicle engines, $SO_2$ and black carbon levels have significantly declined (15). However, organic aerosol concentrations, contributing to more than half of $PM_{2.5}$, have not declined as significantly as primary emissions (16, 17). The COVID-19–induced variability of air quality provides an opportunity to evaluate the efficacy of traffic mitigation strategies.

Diesel-powered heavy-duty vehicles and medium-duty vehicles, such as trucks and buses, comprise only a modest fraction of the total numbers of the on-road fleet in LA but disproportionately contribute to a large fraction of overall vehicle emissions (17, 18, 19). Even with installation of diesel particle filters and selective catalytic reduction (SCR) systems, unusually high emissions of $NO_x$ and lower SCR efficiency are still reported (20). In 2017, The California Air Resources Board (CARB) adopted a series of regulations including reduction of $NO_x$ emissions by 90% for new

## Significance

We capitalize on large variations of urban air quality during the COVID-19 pandemic and real-time observations of traffic, meteorology, and air pollution in Los Angeles to develop a machine-learning air pollution prediction model. Such a model can adequately account for the nonlinear relationships between emissions, atmospheric chemistry, and meteorological factors. Moreover, this model enables us to identify key drivers of air-quality variations and assess the effect of future traffic-emission controls on air quality. We unambiguously demonstrate that the full benefit from fleet electrification cannot be attained if focused only on mitigation of local vehicle emissions. To continue to improve air quality in Los Angeles, off-road emissions and those from volatile chemical products need to be more strictly regulated.

Author contributions: J.Y., Y. Wang, S.Z., and J.H.S. designed research; J.Y., Y. Wen, Y. Wang, and S.Z. performed research; J.Y., Y. Wen, Y. Wang, S.Z., J.P.P., E.A.P., Z.W., Y. Wu, S.P.S., J.H.J., J.H., Y.L.Y., and J.H.S. analyzed data; J.Y., Y. Wen, and S.Z. developed the RF model; and J.Y., Y. Wen, Y. Wang, S.Z., and J.H.S. wrote the paper.

Reviewers: R.R.D., University of Maryland, College Park; and A.H., National Center for Atmospheric Research.

The authors declare no competing interest.



[1]J.Y. and Y.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: yuan.wang@caltech.edu, zhsjun@tsinghua.edu.cn, or seinfeld@caltech.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102705118/-/DCSupplemental.

Published June 21, 2021.

heavy-duty diesel trucks (21), requiring truck manufacturers to transition from diesel trucks and vans to electric zero-emission trucks beginning in 2024, aiming for an all-zero-emission short-haul drayage fleet in ports and railyards by 2035 and zero-emission "last-mile" delivery trucks and vans by 2040 (22). An assessment of the air-quality-related benefit of the zero-emission delivery truck plan is lacking.

Atmospheric chemical transport models have been widely used to examine the response of air pollutant concentrations to the changes of emissions and meteorological conditions. However, the challenge in preparing high-temporal-resolution emission profiles in a timely manner has limited a dynamic analysis of air-quality impacts resulting from the abrupt emission changes through the pandemic period. Recent studies have demonstrated the capability of predictive machine-learning (ML) models to capture the timing, magnitude, and major factors influencing real-time atmospheric responses to emission control measures (23–25). Compared with traditional chemical transport modeling, the ML technique has more flexibility in leveraging real-world data and possesses higher computational efficiency. Here, real-time data including traffic information from the California Department of Transportation (Caltrans), in situ surface-level pollutant concentrations and meteorology from the CARB, and population density and points of interest (physical location of compressed natural gas stations, power plants, landfills, etc.) at the city level are used within an ML framework to develop a model that can directly link atmospheric composition with societal factors. A supervised ML algorithm, the random-forest (RF) model, is employed to account for the nonlinear interactions between different input parameters without specifying any form of their relationships. We use this model to assess the sensitivity of $NO_2$, $O_3$, and $PM_{2.5}$ in the LA Basin to traffic emission changes at different stages of the COVID-19 lockdown by comparing predicted concentrations under different traffic emission scenarios. Moreover, by considering future climate changes and traffic emissions, we assess the possible benefits of future traffic evolution, including vehicular electrification, in 2035 and 2050.

## Results

**Identifying Key Factors Using RF Models.** Machine-learned geostatistical models are developed here to predict the concentrations of three major pollutants: $NO_2$, $O_3$, and $PM_{2.5}$ in the LA basin, using traffic information, meteorological conditions, and other socioeconomic factors as inputs (*SI Appendix*, Fig. S1). The models account for the nonlinear relationships among traffic emissions, atmospheric chemistry, and meteorological conditions. Additional model and data descriptions can be found in *SI Appendix*, *Extended Methods*. To evaluate the performance of the RF models, a commonly used fivefold cross-validation method is used (26, 27). As shown in Fig. 1, the models exhibit high fidelity in reproducing the observed $NO_2$ and $O_3$ concentrations, with coefficients of determination ($R^2$) of 0.88 and 0.86, respectively. The root-mean-square errors (RMSE) of the predicted $NO_2$ and $O_3$ concentrations are 3.45 and 4.32 ppb, respectively. The predicted $PM_{2.5}$ concentrations also show reasonable agreement with the ground-based observations, but with a smaller $R^2$ of 0.65. An underestimation of $PM_{2.5}$ starts to emerge when the $PM_{2.5}$ concentrations exceed 20 $\mu g/m^3$, corresponding to the 90th percentile in the $PM_{2.5}$ probability distribution function over LA. It is noted that RF models tend to have larger biases in predicting the extreme values due to fewer training data samples (28).

An important output of the RF model is a ranking of the relative importance of all input parameters. For $NO_2$, the three major governing factors are wind direction, nontruck vehicle miles traveled (VMT), and wind speed. The prominent rank of wind direction reflects the prevailing role of northwesterly and onshore winds in determining the spatially variable flow of pollutants received in the LA basin (29). The concentration of $NO_2$, which is a short-lived species, closely follows that of the traffic emission patterns. For example, $NO_2$ concentration is negatively correlated to wind speed due to the dilution effect and slightly increased as temperature decreases because of lower SCR efficiency at low temperatures (30). In contrast to $NO_2$, ozone variations are largely regulated by meteorological conditions. Moreover, the top five factors are all meteorology-related. Among them, the near-surface temperature ($T_{2m}$) exerts the largest influence through photochemical reactions forming ozone (31) and biogenic volatile organic compound (VOC) emission rates (32). Solar irradiance is a limiting factor that influences ozone-related photochemistry. For $PM_{2.5}$ prediction, ozone ranks as the most prominent, indicating the secondary source of aerosols in LA. Boundary layer height is the most relevant meteorological factor with $PM_{2.5}$ in the Gini importance ranking (see *SI Appendix*, *Extended Methods*), followed by relative humidity (RH), $T_{2m}$, and wind direction. Such a ranking of meteorological influence on $PM_{2.5}$ in LA is consistent with current understanding (9). Notably, a recent study on 8-y ground-based observations in Beijing, China showed the same importance ranking of meteorological factors (33). By using the points of interest (*SI Appendix*, Fig. S2) in the RF model, the influence of spatial contribution from crucial industrial locations (e.g., airport, wastewater treatment plants, power plants, and natural gas compressor stations) on air quality is identified. The model-predicted importance of airport-related emissions for $NO_2$ corroborates that air quality impacts of major airports need to be addressed for emission control (34). Volatile consumer and industrial chemical products are estimated to be a significant source of reactive VOCs and secondary organic aerosol formation in the LA Basin (35). To test the importance of different predictors on a time scale longer than hours, we rebuild the RF models using the daily means of the input data. The results from those models show generally similar ranking of predictors. The daily mean models retain 60%, 80%, and 80% of the top five most important predictors for $NO_2$, $O_3$, and $PM_{2.5}$, respectively, compared with the hourly models (*SI Appendix*, Fig. S3).

**Role of Traffic Emissions during COVID-19.** During the COVID-19 pandemic traffic was abruptly reduced in late March and early April and then gradually recovered to the pre-COVID-19 level in LA (*SI Appendix*, Fig. S4). The time series of $NO_2$ generally followed the temporal variation of traffic in LA during the COVID-19 period. $O_3$ and $PM_{2.5}$ concentrations remained at a relatively low level in March and early April due to rainy and windy weather conditions. We conduct the RF model predictions with COVID-19 meteorology and pre–COVID-19 traffic information from on-road sensors (VMT, automobile type, etc.), so the differences between these model predictions and the observed pollution levels reflect the influence of the COVID-19–induced traffic emission reductions (Fig. 2*A*). During the strictest lockdown period (6 April to 12 April), traffic reduction led to decreases in the daily averaged $NO_2$ and $PM_{2.5}$ concentrations by 2.9 ppb and 1.1 $\mu g/m^3$, corresponding to fractional changes of –30.1% and –17.5%, respectively. These results are consistent with an independent chemical transport model assessment which reported that COVID-related emission reductions caused the population-weighted mean concentrations of $PM_{2.5}$ and $NO_2$ to decrease by 1.2 $\mu g/m^3$ (–14%) and 2.4 ppb (–22%) during 19 March to 20 April, respectively (36). In the later recovery period (8 May to 30 June), the all-traffic-induced fractional changes of $NO_2$ and $PM_{2.5}$ decrease to –19.0% and –6.0%, respectively. The traffic impacts on ozone differ from those of $NO_2$ and $PM_{2.5}$. A 2.1-ppb (5.7%) increase in maximum daily 8-h average (MDA8) $O_3$ by all traffic occurred during the strictest lockdown period. The ozone enhancement is mainly caused by the alleviation of the ozone titration by NO. There is also a contribution from the nonlinearity of ozone formation chemistry. The ozone production in LA is in the $NO_x$-saturated/VOC-limited regime under the traffic-as-usual

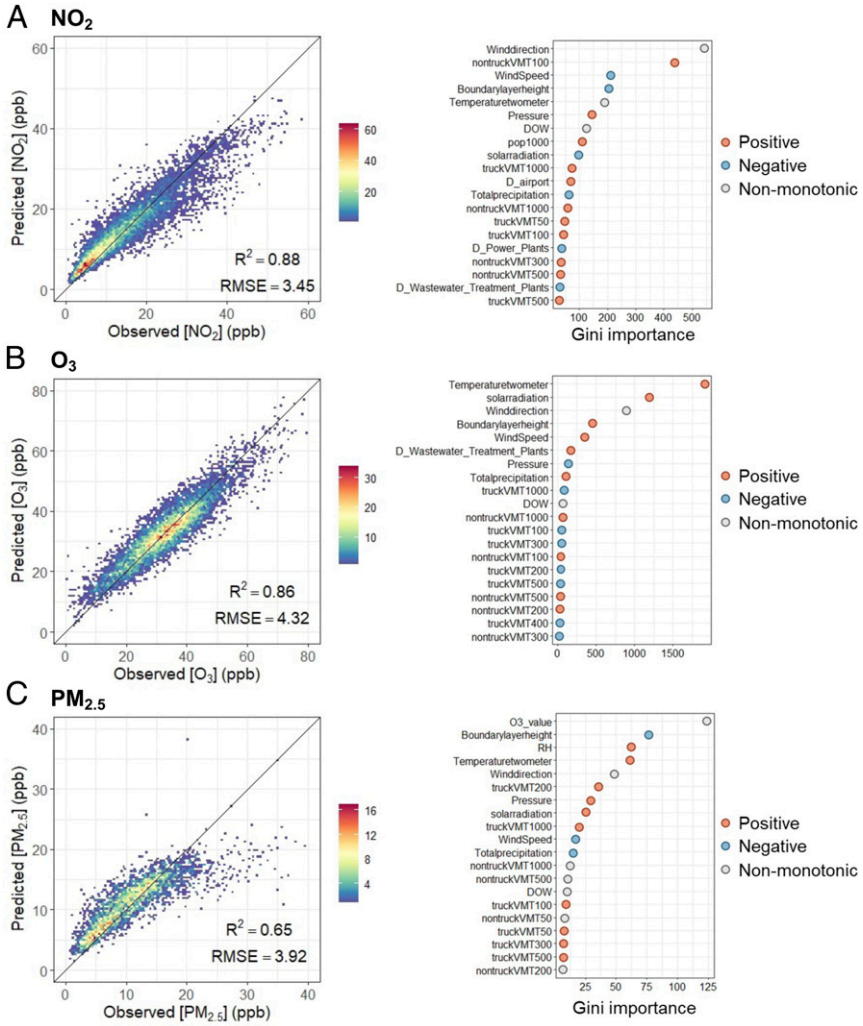


**Fig. 1.** Model performance and variable importance for three species: (*A*) $NO_2$, (*B*) $O_3$, and (*C*) $PM_{2.5}$ in LA. Cross-validated model $R^2$ and RMSE are calculated by using a fivefold cross-validation modeling performance for 24-h average concentrations. The color indicates the sample size for each dot. The variables are listed in order of importance from top to bottom. The horizontal axis represents the Gini index from the RF model. A larger value represents higher importance. The definitions of all predictors are provided in *SI Appendix*, Table S2.

scenario. Excessive $NO_x$ can serve as a sink for OH radicals, thus retarding the oxidation of VOC, sequestering ozone, or suppressing its production (7). Note that the LA Basin has a large portion of background ozone produced from local biogenic sources or transboundary transport (37). Hence, the fractional change of the COVID-19–induced ozone enhancement can be higher, if the anthropogenic ozone is used as a reference. We further differentiate the impacts from truck and nontruck vehicles by altering only the on-road truck activities according to the observations from different time periods (Fig. 2*B*). During the strictest lockdown period, truck emission reductions account for 61.1%, 81.6%, and 70.4% of all-traffic-induced changes in $NO_2$, MDA8 $O_3$, and $PM_{2.5}$, respectively. This result reinforces the fact that diesel trucks are a major source in the entire traffic sector.

To build a direct linkage between pollutant concentrations and traffic activity we also develop an emulator for each species based on our RF model results. The emulator can predict the relative changes of emissions as a function of the fractional changes in truck and nontruck VMT relative to the year 2019 level. $NO_2$ monotonically decreases along with the reduction in either truck or nontruck VMT (Fig. 3*A*). The reduction slope is steeper for trucks, indicating the larger emission factor of $NO_x$ for diesel engines. MDA8 $O_3$ generally increases with the reduction of truck traffic in a monotonic manner (Fig. 3*B*), while an overall decrease in MDA8 $O_3$ is found for the reduction of nontrucks. The distinctive impacts on ozone are likely explained by the fact that diesel trucks emit higher levels of $NO_x$ than nontrucks (38), but they share the similar nonmethane VOC emission factor (39, 40). Therefore, truck and nontruck emissions fall in $NO_x$-saturated and $NO_x$-limited regimes, respectively. This is also consistent with larger $NO_2$ susceptibility to reductions of truck than of nontruck emissions. The $PM_{2.5}$ linkage with traffic is more complicated, especially with regard to nontruck emissions. In contrast with the monotonic decrease of $PM_{2.5}$ in response to the reduction in truck VMT, the bended-curve (Fig. 3*C*) response of $PM_{2.5}$ is found along with the nontruck VMT reduction. Similar to MDA8 $O_3$, the overall magnitude of fluctuation of $PM_{2.5}$ is also smaller for nontruck (less than 0.1 $\mu g/m^3$) than that for truck. In general, regulation of trucks can be a more efficient way to lower $PM_{2.5}$ concentration than other vehicles.

**Air-Quality Benefit for Future On-Road Traffic Decarbonization.** The Paris Agreement aims to increase the percentage of zero emission vehicles to 25% by 2025, 80% by 2035, and 100% by 2050. Under the "Green New Deal" LA would build a clean and reliable power grid to empower the next generation of green transportation.

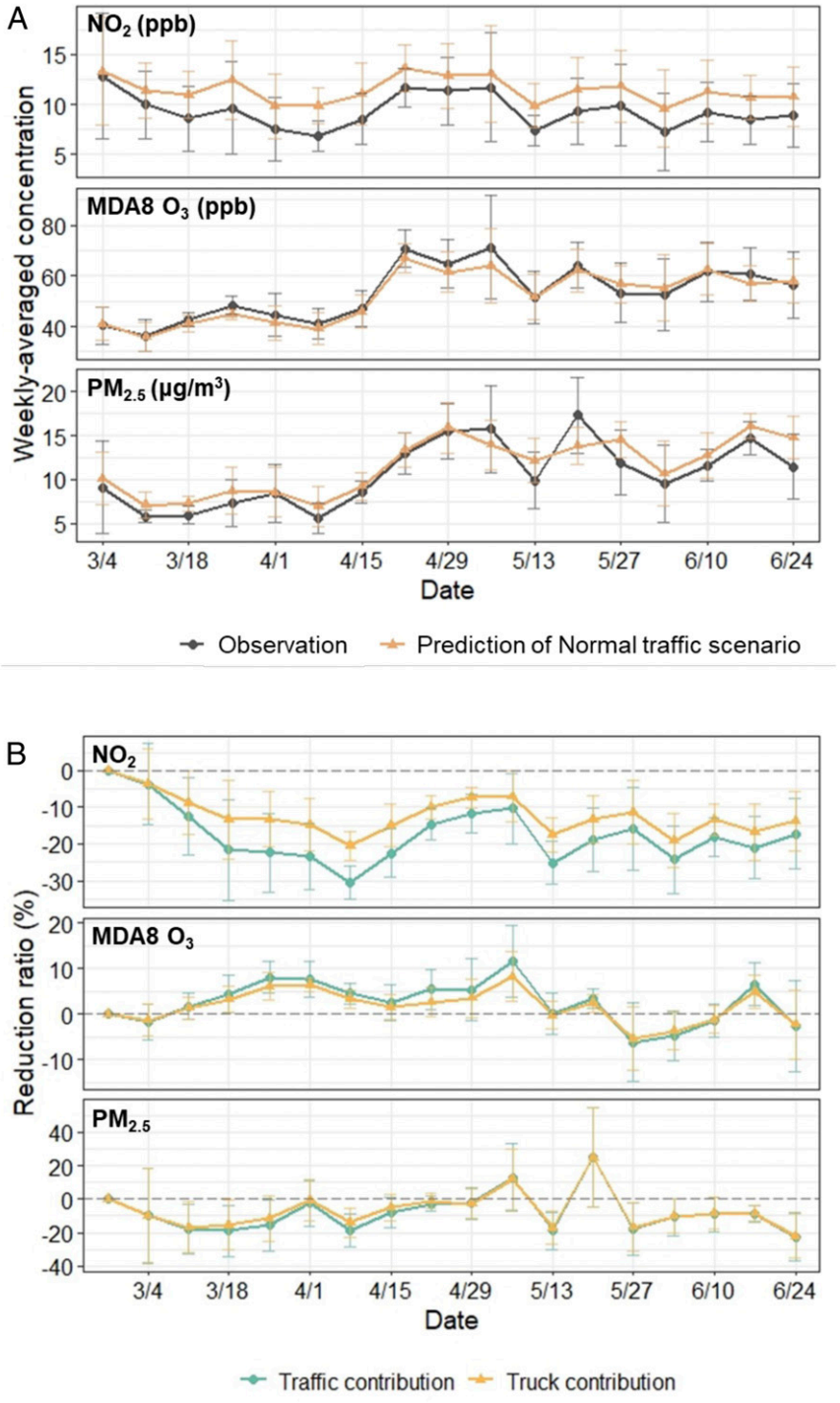


**Fig. 2.** Comparison of observations and predictions. (*A*) Comparison of observations and predictions of normal traffic scenario and (*B*) the impact of traffic reduction from total fleet and truck fleet on $NO_2$, $O_3$, and $PM_{2.5}$ concentrations during the lockdown period of the COVID-19 pandemic in LA. Each data point represents a weekly mean. The error bars are SDs from daily results in each week.

The baseline future traffic emission changes are provided by the 2017 version of the EMission FACtor (EMFAC), a model that estimates the official emission inventories of on-road mobile sources in California from 2000 to 2050 (41). Here we introduce three degrees of fleet electrification (also including other zero-emission vehicles like hydrogen fuel cell vehicles) based on the EMFAC emission inventories (*SI Appendix*, *Extended Methods* and Table S1). All the fractional changes in the truck or nontruck vehicles for the future scenarios are within their ranges in our RF training dataset, i.e., hourly observations during 2019 and 2020.

The EMFAC model assumes that nontruck emissions will decrease by 54% in 2035 and 58% in 2050 as compared with 2019 (*SI Appendix*, Fig. S5). For truck emissions, CARB recently estimated that the low-$NO_x$ omnibus regulation would lead to 29% of $NO_x$ emission reduction in 2050 as compared with the original EMFAC results, which have been used as the baseline

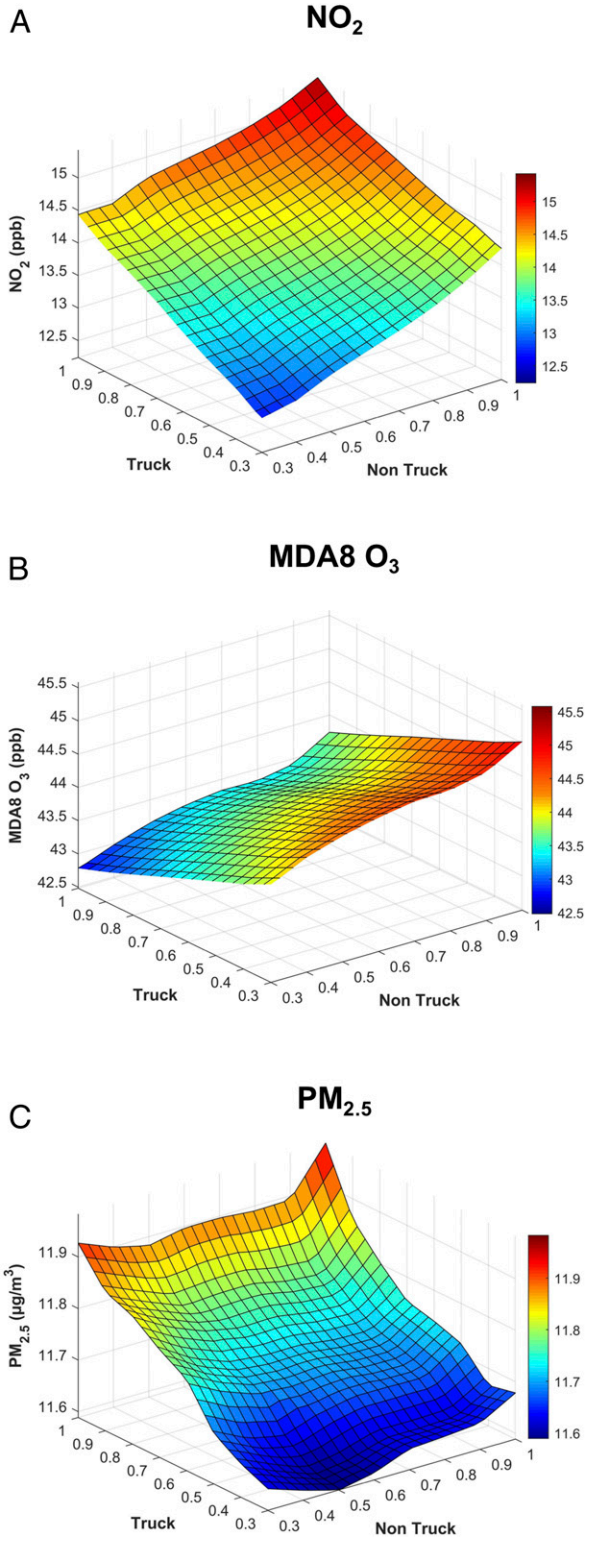


**Fig. 3.** Predicted annual-average concentrations. Distribution of (*A*) $NO_2$, (*B*) MDA8 $O_3$, and (*C*) $PM_{2.5}$ with different combinations of nontruck and truck activity fractional changes relative to the annual average level of 2019.

truck emissions without additional electrification (42). On the other hand, the EMFAC inventories assume greater truck activity caused by increases in intensity of consumer goods delivery in 2050 than in 2035. Therefore, compared to 2019, truck emissions would have comparable decreasing ratios in 2035 (by 55%) and 2050 (by 54%). The impacts of the future traffic emission reduction are pronounced: Compared to 2019, $NO_2$ would be reduced by 13.7% ± 1.3% in 2035 and 14.4% ± 0.9% in 2050. $PM_{2.5}$ would be reduced by 3.1% ± 0.4% in 2035 and 2.9% ± 0.4% in 2050 (Fig. 4 *A* and *F*). Similar to its behavior during the lockdown period of COVID-19, MDA8 $O_3$ is predicted to exhibit a reverse trend with a 0.6% ± 0.4% increase in 2035 and a 0.4% ± 0.3% increase in 2050. This result is also a combination of the $NO_x$-saturation regime and the ozone titration effect. Of note, the reduction ratios of $NO_2$ and $PM_{2.5}$ concentrations increase significantly from 2020 to 2035 due to the efficient reduction of traffic emissions, while the reductions slow down and even slightly rebound from 2035 to 2050 with the relatively limited emission reductions (*SI Appendix*, Fig. S6).

To further assess the impacts of fleet electrification on air quality we independently alter the electrification rates of total fleet mileage from the remaining parameters in EMFAC. Three scenarios are assessed here, representing moderate to aggressive electrification rates (*SI Appendix*, Table S1 and Fig. S7). Our first electrification scenario (E1) assumes moderate electrification rates, i.e., 10% and 5% for nontrucks and trucks in 2035. The 2035 electrification rate of the truck fleet is close to the Advanced-Clean-Trucks regulation benefit estimated by CARB, because more than 60% of class 8 trucks operating in California are registered as out-of-state vehicles that will not be mandatory to be electrified according to current federal plans (42). In 2035, as compared with 2019, E1 corresponds to emission reduction rates of 57% for both nontruck and truck. In 2050, the electrification rates of E1 are 20% for nontruck and 10% for truck, corresponding to emission reduction rates of 65% for nontruck and 59% for truck. As shown in Fig. 4 *B* and *G*, the RF model predicts that $NO_2$ will decrease by 14.7% ± 1.0% in 2035 and 15.8% ± 0.6% in 2050 under E1. Also, $PM_{2.5}$ will drop by 3.0% ± 0.3% in 2035 and 2.9% ± 0.4% in 2050. MDA8 $O_3$ is predicted to increase by 0.6% ± 0.4% in 2035 and 0.3% ± 0.3% in 2050. The other two future scenarios (E2 and E3) are more aggressive in electrifying vehicles than E1. Therefore, the magnitudes of the $NO_2$ reduction are enlarged in E2 and E3, and the reduction ratio achieves 19.1% ± 1.1% in 2050 under E3 with the most aggressive electrification ratios [i.e., 80% for nontrucks, which is close to the estimated electrification rate from California's Advanced Clean Cars program (43), and 40% for trucks]. The increasing ratio of MDA8 $O_3$ shrinks with higher electrification rates in both 2035 and 2050 (Fig. 4 *H* and *I*). Such a change in MDA8 $O_3$ reveals that LA would be evolving to less-$NO_x$-saturated conditions with further reduction of $NO_x$. However, $PM_{2.5}$ levels are less sensitive to progressive electrification. One possible reason is the unbalanced emission reduction in truck and nontruck fleets for future electrification. According to Fig. 3, $PM_{2.5}$ is more sensitive to trucks than to nontrucks. The relative higher emission contribution of trucks in the total fleet from E1 to E3 may explain the decrease in $PM_{2.5}$ reduction.

**Regional Climate Change on Air Quality.** The effect of climate change on meteorological conditions is a key factor in modulating urban pollution. The responses of different pollutants to four key meteorological variables are probed here via idealized perturbation experiments using the RF models (*SI Appendix*, Fig. S8). The model shows that $PM_{2.5}$ is enhanced by RH via the promotion of heterogeneous chemistry to form secondary aerosols in aerosol water (9). Increasing photochemistry via solar radiation tends to increase $PM_{2.5}$ and $O_3$ at the expense of $NO_2$. Both $NO_2$ and $PM_{2.5}$ concentrations are elevated by a lower boundary layer height. $O_3$ shows the opposite responses due to concurrent $O_3$ titration and lower boundary layer height at nighttime. Higher surface temperature fosters ozone production and further promotes secondary aerosol formation. Our RF model is capable of capturing the monotonic increasing relationship between MDA8

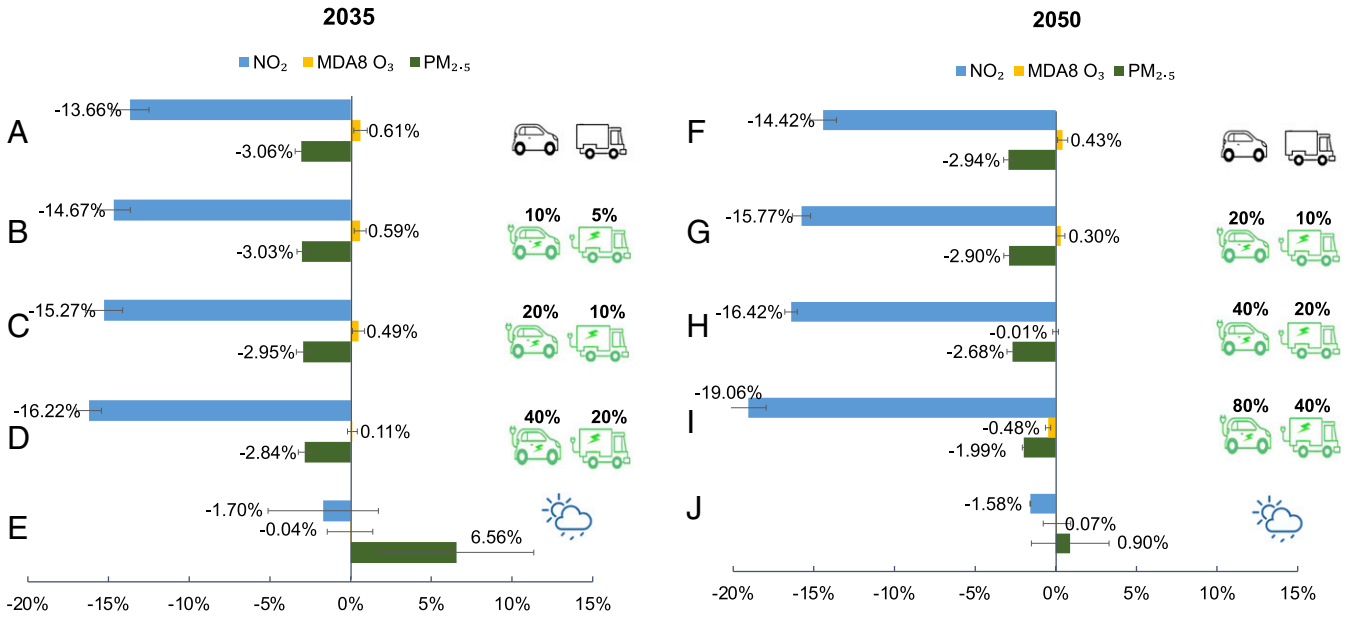


**Fig. 4.** Reduction ratios of $NO_2$, MDA8 $O_3$, and $PM_{2.5}$ concentrations under different traffic scenarios in 2035 and 2050 relative to 2019. *A* and *F* represent baseline traffic emission scenarios from EMFAC; *B–D* and *G–I* represent three electrification scenarios; *E* and *J* future climate change scenarios in 2035 and 2050, respectively. The error bars represent uncertainty of model predictions calculated by the Monte Carlo method. Random sampling was repeated for 100 times considering uncertainty of each variable in prediction of each scenario.

O3 concentration and ambient temperature. The d(O3)/d(T) in 2019 is ~1.5 ppb/°C, which is generally consistent with a previous study (44) which reported a post-2002 climate penalty factor (CPF) of 2 ppb/°C for the eastern United States, and the CPF is expected to decrease as $NO_X$ emissions have been greatly reduced since then.

Additional model predictions were performed to assess the impacts of future regional climate change on air quality in LA. Future meteorological variables near 2035 and 2050 are projected from the multimodel ensemble simulations of the Climate Model Intercomparison Project Phase 6 (CMIP6; see *SI Appendix, Extended Methods* and Fig. S9), while the same traffic level as 2019 will be adopted. Our RF models predict that the annual mean concentrations of $NO_2$ and MDA8 $O_3$ will decrease while $PM_{2.5}$ will increase at different rates around 2035 and 2050 (Fig. 4 *E* and *J*). The rates of change for the three species in 2050 are not in proportion with the changes in 2035, reflecting the highly nonlinear relationship of climate change and with air pollution over a few decades. Future climate changes are estimated to exert a higher influence on $O_3$ and $PM_{2.5}$ concentrations than traffic amount and type in 2035, demonstrating the pronounced impacts of meteorology on these two species. This result is also consistent with the ranking of variable importance in the RF models (Fig. 1). The uncertainty of future climate change is estimated by the spread among different CMIP6 models and different ensemble members.

In summary, we leverage the unprecedented large variations of road traffic spanning a few months in LA during the COVID-19 pandemic to probe the impacts of future decarbonization policies. An ML model is developed for LA to predict $NO_2$, $O_3$, and $PM_{2.5}$ concentrations based on real-time traffic data and meteorological measurements. Capitalizing on the high fidelity and computing efficiency of this predictive RF model, we demonstrate the significant contribution of traffic, especially from heavy-duty trucks, to pollutant variations in the first few months of the COVID-19 pandemic. Future decarbonization policies are estimated to impose effects on air quality similar to COVID-19, but with smaller magnitude. Large-scale fleet electrification will achieve further alleviation of $NO_2$ levels and is likely to transition LA to a less-$NO_x$-saturated regime of $O_3$ formation. However, the benefit from fleet electrification on $PM_{2.5}$ may be not attained if it is focused only on mitigation of on-road emissions. Moreover, emission standards of out-of-state vehicles should be aligned with those of the local fleet under federal efforts, and off-road emissions and those of volatile chemical products need to be more strictly regulated.

## Methods

Hourly data over 1.5 y (January 2019 to June 2020) serve as input to the RF models. Key input parameters include processed traffic activity (truck/non-truck VMT), meteorology (wind speed/direction, near-surface temperature, boundary layer height, precipitation, solar radiation, pressure, and RH), temporal information (weekday/weekend and holiday), population density, distance to nearby points of interest, etc. Complete information can be found in *SI Appendix*. A wide range of temporal variability is explicitly considered, from diurnal, daily, weekly, and seasonal timescales. The hourly temporal resolution of the training data is sufficiently high to capture the lifetimes of the three targeted species. The predictive capability is separately developed at 11 sites for $PM_{2.5}$, 18 sites for $O_3$, and 22 sites for $NO_2$, covering the populous urban areas in the LA Basin (*SI Appendix*, Fig. S2). Additional data description and experiment designs can be found in *SI Appendix*. Note that our training dataset include both $NO_x$-limited and $NO_x$-insensitive regimes over certain locations and time periods. Therefore, our site-specific RF models do reproduce the signs of $NO_X$-insensitive or $NO_X$-limited regimes outside urban areas (*SI Appendix*, Fig. S12). A recent study used two-decade air-quality data from the same sites to identify the long-term photochemical regimes over the LA area and reported a subset of air quality sites in periurban areas or downwind areas represent $NO_x$-insensitive or even $NO_x$-limited regimes during the afternoon period (31).

**Data Availability.** All study data are included in the article and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** Y. Wang, S.P.S., J.H.J., and Y.L.Y. acknowledge support by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. E.A.P. and J.H.S. acknowledge support by the Samsung Corporation (award SAMS.2019GRO). S.Z. acknowledges support by the National Key Research and Development Program of China (grant 2017YFC0212100), the National Natural Science Foundation of China (grant 41977180), and Ford Motor Company. J.Y. acknowledges Leo Gallagher and Daniel Kitowski at the California Department of Transportation, Thomas E. Morrell at the Caltech Library, Jin Tao for helpful information on data inputs, and Yu Zhou at Tsinghua University for useful discussions.

1. K. Zhang, S. Batterman, Near-road air pollutant concentrations of CO and PM2.5: A comparison of MOBILE6.2/CALINE4 and generalized additive models. *Atmos. Environ.* **44**, 1740–1748 (2010).
2. D. Liang *et al*., Urban air pollution may enhance COVID-19 case-fatality and mortality rates in the United States. *Innovation (N Y)* **1**, 100047 (2020).
3. F. Pomponi, J. Hart, J. H. Arehart, B. D'Amico, Buildings as a global carbon sink? A reality check on feasibility limits. *One Earth* **3**, 157–161 (2020).
4. Z. Liu *et al*., Near-real-time monitoring of global $CO_2$ emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* **11**, 5172 (2020).
5. C. Le Quéré *et al*., Temporary reduction in daily global $CO_2$ emissions during the COVID-19 forced confinement. *Nat. Clim. Chang.* **10**, 647–653 (2020).
6. M. Bauwens *et al*., Impact of coronavirus outbreak on $NO_2$ pollution assessed using TROPOMI and OMI observations. *Geophys. Res. Lett.* **47**, e2020GL087978 (2020).
7. J. H. Kroll *et al*., The complex chemical effects of COVID-19 shutdowns on air quality. *Nat. Chem.* **12**, 777–779 (2020).
8. F. Liu *et al*., Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. *Sci. Adv.* **6**, eabc2992 (2020).
9. T. Le *et al*., Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. *Science* **369**, 702–706 (2020).
10. Y. Zhao *et al*., Substantial changes in nitrogen dioxide and ozone after excluding meteorological impacts during the COVID-19 outbreak in mainland China. *Environ. Sci. Technol. Lett.* **7**, 402–408 (2020).
11. J. Wu *et al*., Aerosol-photolysis interaction reduces particulate matter during wintertime haze events. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9755–9761 (2020).
12. R. Gottlieb, M. Vallianatos, R. Freer, P. Dreier, *The Next Los Angeles: The Struggle for a Livable City* (University of California Press, 2006).
13. K. W. Wong *et al*., Mapping $CH_4$ : $CO_2$ ratios in Los Angeles with CLARS-FTS from mount Wilson, California. *Atmos. Chem. Phys.* **15**, 241–252 (2015).
14. D. Wunch, P. O. Wennberg, G. C. Toon, G. Keppel-Aleks, Y. G. Yavin, Emissions of greenhouse gases from a North American megacity. *Geophys. Res. Lett.* **36**, L15810 (2009).
15. S. Sonwani, P. Saxena, Identifying the sources of primary air pollutants and their impact on environmental health: A review. *Int. J. Eng. Tech. Res.* **6**, 20 (2016).
16. C. S. Christoforou, L. G. Salmon, M. P. Hannigan, P. A. Solomon, G. R. Cass, Trends in fine particle concentration and chemical composition in southern California. *J. Air Waste Manag. Assoc.* **50**, 43–53 (2000).
17. A. C. Lloyd, T. A. Cackette, Diesel engines: Environmental impact and control. *J. Air Waste Manag. Assoc.* **51**, 809–847 (2001).
18. M. J. Haugen, G. A. Bishop, A. Thiruvengadam, D. K. Carder, Evaluation of heavy- and medium-duty on-road vehicle emissions in California's south coast air basin. *Environ. Sci. Technol.* **52**, 13298–13305 (2018).
19. M. Kelp *et al*., Sensitivity analysis of area-wide, mobile source emission factors to high-emitter vehicles in Los Angeles. *Atmos. Environ.* **223**, 117212 (2020).
20. C. V. Preble, R. A. Harley, T. W. Kirchstetter, Control technology-driven changes to in-use heavy-duty diesel truck emissions of nitrogenous species and related environmental impacts. *Environ. Sci. Technol.* **53**, 14568–14576 (2019).
21. California Air Resources Board, Heavy-duty omnibus regulation (2021). https://ww2.arb.ca.gov/rulemaking/2020/hdomnibuslownox. Accessed 4 January 2021.
22. California Air Resources Board, California takes bold step to reduce truck pollution (2020). https://ww2.arb.ca.gov/news/california-takes-bold-step-reduce-truck-pollution. Accessed 27 October 2020.
23. Y. Wang *et al*., Four-month changes in air quality during and after the COVID-19 lockdown in Six megacities in China. *Environ. Sci. Technol. Lett.* **7**, 802–808 (2020).
24. S. K. Grange *et al*., COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban areas. *Atmos. Chem. Phys. Discuss.* **21**, 4159–4185 (2021).
25. M. Lovrić *et al*., Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. *Environ. Pollut.* **274**, 115900 (2021).
26. Y. Bengio, Y. Grandvalet, No unbiased estimator of the variance of K-fold cross-validation. *J. Mach. Learn. Res.* **5**, 1089–1105 (2004).
27. S. Arlot, M. Lerasle, Why V=5 is enough in V-fold cross-validation. *HAL-Inria* (2014). https://hal.inria.fr/hal-00743931v2. Accessed 31 March 2021.
28. M. C. Robinson, R. C. Glen, A. A. Lee, Validating the validation: Reanalyzing a large-scale comparison of deep learning and machine learning models for bioactivity prediction. *J. Comput. Aided Mol. Des.* **34**, 717–730 (2020).
29. C. K. Wong *et al*., Monthly trends of methane emissions in Los Angeles from 2011 to 2015 inferred by CLARS-FTS observations. *Atmos. Chem. Phys.* **16**, 13121–13130 (2016).
30. D. L. Hall *et al*., Using near-road observations of CO, $NO_y$, and $CO_2$ to investigate emissions from vehicles: Evidence for an impact of ambient temperature and specific humidity. *Atmos. Environ.* **232**, 117558 (2020).
31. H. A. Parker, S. Hasheminassab, J. D. Crounse, C. M. Roehl, P. O. Wennberg, Impacts of traffic reductions associated with COVID-19 on southern California air quality. *Geophys. Res. Lett.* **47**, e2020GL090164 (2020).
32. S. E. Pusede *et al*., On the temperature dependence of organic reactivity, nitrogen oxides, ozone production, and the impact of emission controls in San Joaquin Valley, California. *Atmos. Chem. Phys.* **14**, 3373–3395 (2014).
33. T. Su, Z. Li, Y. Zheng, Q. Luan, J. Guo, Abnormally shallow boundary layer associated with severe air pollution during the COVID-19 lockdown in China. *Geophys. Res. Lett.* **47**, e2020GL090041 (2020).
34. N. Hudda, T. Gould, K. Hartin, T. V. Larson, S. A. Fruin, Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. *Environ. Sci. Technol.* **48**, 6628–6635 (2014).
35. B. C. McDonald *et al*., Volatile chemical products emerging as largest petrochemical source of urban organic emissions. *Science* **359**, 760–764 (2018).
36. Z. Jiang *et al*., Modeling the impact of COVID-19 on air quality in southern California: Implications for future control policies. *Atmos. Chem. Phys. Discuss.*, 10.5194/acp-2020-1197 (2020).
37. D. D. Parrish, L. M. Young, M. H. Newman, K. C. Aikin, T. B. Ryerson, Ozone design values in Southern California's air basins: Temporal evolution and U.S. background contribution. *J. Geophys. Res. Atmos.* **122**, 11166–11182 (2017).
38. R. O'Driscoll, M. E. J. Stettler, N. Molden, T. Oxley, H. M. ApSimon, Real world $CO_2$ and $NO_x$ emissions from 149 Euro 5 and 6 diesel, gasoline and hybrid passenger cars. *Sci. Total Environ.* **621**, 282–290 (2018).
39. B. C. McDonald, D. R. Gentner, A. H. Goldstein, R. A. Harley, Long-term trends in motor vehicle emissions in U.S. urban areas. *Environ. Sci. Technol.* **47**, 10022–10031 (2013).
40. A. A. May *et al*., Gas- and particle-phase primary emissions from in-use, on-road gasoline and diesel vehicles. *Atmos. Environ.* **88**, 247–260 (2014).
41. California Air Resources Board, EMFAC2017 volume III technical documentation V1.0.2 July 20, 2018. https://ww3.arb.ca.gov/msei/downloads/emfac2017-volume-iii-technical-documentation.pdf. Accessed 21 December 2020.
42. California Air Resources Board, Appendix D—Emissions inventory methods and results for the proposed amendments (2020). https://ww3.arb.ca.gov/regact/2020/hdomnibuslownox/appd.pdf. Accessed 4 January 2021.
43. California Air Resources Board, Vision scenario planning (2021). https://ww2.arb.ca.gov/resources/documents/vision-scenario-planning. Accessed 4 January 2021.
44. B. J. Bloomer, J. W. Stehr, C. A. Piety, R. J. Salawitch, R. R. Dickerson, Observed relationships of ozone air pollution with temperature and emissions. *Geophys. Res. Lett.* **36**, L09803 (2009).