Using Explainable Machine Learning to Interpret the Effects of Policies on Air Pollution: COVID-19 Lockdown in London

Liang Ma*[1], Daniel J. Graham[1], Marc E.J. Stettler[1]

[1] Centre for Transport Studies, Department of Civil and Environmental Engineering,

Imperial College London

Short Paper submitted for presentation at the
11th Symposium of the European Association for Research in Transportation
Zurich, 2023

SHORT SUMMARY

Activity changes during the COVID-19 lockdown brought an unprecedented opportunity to understand the likely effectiveness of prospective air quality management policies on reducing air pollution. Using a regression discontinuity design for causal analysis, we show that the first UK national lockdown led to unprecedented decreases in road traffic yet incommensurate and heterogeneous responses in air pollution in London. At different locations, changes in air pollution attributable to the lockdown ranged from -50% to 0% for NO_2, 0% to +4% for O_3, -5% to +0% for PM_{10} and there was no response for $PM_{2.5}$. Using explainable machine learning, we show that the degree to which NO_2 pollution was reduced in an area was correlated with spatial features (including road freight traffic and proximity to a major airport and the city centre), and that existing inequalities in air pollution exposure were exacerbated: pollution reductions were greater in places with more affluent residents and better access to public transport services.

Keywords: Air pollution; Causal analysis; COVID-19; Explainable machine learning.

1. INTRODUCTION

Various interventions have been implemented in cities to improve air quality. However, the impact pathways from an intervention to air quality can incorporate various factors and complicated interactions among factors, which presents challenges to both isolating the intervention effect and quantifying the contribution of different factors to the net effect. Recently, the rapid development of interpretation methods for machine learning (ML) models to achieve explainable ML has provided an opportunity to gain insights into complicated relationships among high-dimensional variables (Molnar et al., 2022). Unlike black-box ML models, explainable ML seeks to make the outputs and processes of ML models more interpretable and understandable to humans and has been applied in various fields to support decision-making (Guidotti et al., 2018; Molnar et al., 2022).

Activity changes during the COVID-19 lockdown brought an unprecedented opportunity to understand the likely effectiveness of prospective emission control policies in improving air quality. Different methods have been applied to quantify the air quality impacts of lockdowns, including comparing air quality levels before and after the lockdown, using bottom-up simulations, predicting business-as-usual concentrations with ML models, and applying causal

* Proposed presenting author. Email: liang.ma13@imperial.ac.uk

Full paper submitted to Environmental Science & Technology on December 21, 2022 (under review)

inference methods (Dang & Trinh, 2021; Jephcote et al., 2021; Menut et al., 2020; Venter et al., 2020). Compared with other methods applied, causal inference methods generally have advantages in data requirement, model building, and the interpretation of effect estimates. However, due to the complexity of atmospheric processes, previous studies of this type had a typical limitation by using a parametric confounding control for weather conditions. Meanwhile, several studies found heterogeneous air quality effects of lockdowns (Fassò et al., 2021; Vadrevu et al., 2020). However, most of them focused on depicting the variation in effects across subsamples of cities or countries; few compared effects within a city or further evaluated the contribution of factors to the heterogeneity of effects.

In this paper, we provide an analysis of the changes in London's air quality attributable to the first UK COVID-19 lockdown with a causal inference approach. To overcome the typical limitation in confounding control, the meteorological normalisation technique is applied to non-parametrically control for weather conditions and seasonality effects. To further understand the spatial heterogeneity in the lockdown impacts, we additionally evaluate the contribution of different factors to the level of lockdown impacts by interpreting a predictive ML model with SHapley Additive exPlanations (SHAP) values. Various factors are considered, such as economy, demographics, transport demand, public transport supply, and geographical location. The identification of key factors affecting pollution reduction within a city can provide further guidance in improving air quality and help to shape our city with better and more equitable design, planning, and management.

2. METHODS

The research framework of this paper is shown in Figure 1. The method for quantifying the causal air quality impacts mainly follows Ma et al. (2021b). The evaluation is applied at individual air quality monitoring sites within London for regulated air pollutants including NO_2, O_3, $PM_{2.5}$, and PM_{10}, and a non-regulated pollutant NO_x. Specifically, meteorological normalisation is first applied to control important confounders. A Gradient Boosting Decision Trees (GBDT) is used to consider complex relationships among model variables. A normalised concentration time series is derived by removing the variation in the observed concentrations that can be explained by input confounders. Change point detection (CPD) is then conducted to detect structural changes in the normalised concentration time series; the results are used to support the research period specification and test a key model assumption for the next step. A sharp regression discontinuity design (RDD) is then specified on the normalised concentrations where a site had at least one change point around the start of the lockdown T_0, i.e. showed a response. As a causal inference method, sharp RDD goes beyond comparing air quality before and after T_0. The inference is based on a trend function approximation and quantification of effect at the discontinuity of a trendline either side of T_0, which makes it less vulnerable to random noise and unrelated events (Ma et al., 2021b). With the estimated RDD coefficients, the lockdown effect τ is derived by stacking the impact from the current daily period and those from lagged periods. The interval estimates of τ is provided with a Monte Carlo simulation in Ma et al. (2021a). The τ estimates at different sites are aggregated with a bootstrapping in Ma et al. (2021a) for a city-wide mean.

To further interpret the spatial heterogeneity of impacts, 124 spatial features are evaluated based on their contribution in predicting the pollution reduction due to lockdown, using explainable ML (details in Figure 1). The evaluation is conducted at the Middle Layer Super Output Area (MSOA) level and focuses on the changes in annual mean concentrations of NO_2 in 2020 caused by the lockdown. Particularly, the annual mean concentration is estimated in two scenarios at individual sites: with and without lockdown. Mapping methods are then applied to estimate the pollution reductions attributable to the lockdown at MSOAs; the mapping model mainly follows Horálek et al. (2019), which combines a linear regression and ordinary kriging of residuals. A GBDT model is then built on the estimated pollution reduction with the features at MSOAs, separately for absolute and relative reductions. The key hyperparameters of GBDT are automatically tuned with Bayesian optimisation. As the widely used feature importance metrics (such as total gains) do not directly indicate how the level of the output variable is affected by features, we use the SHAP values (an approximation of Shapley values) to interpret our GBDT models. The Shapley value is a key solution in cooperative game theory and can also be classified as an additive feature attribution method for a local explanation of complex predictive models (Lundberg et al., 2020). An additive feature attribution method is characterised by assigning a contribution ϕ_i to each feature, with the sum of $\{\phi_i\}$ approximating the original model predict (Lundberg et al., 2020). Therefore, ϕ_i is in the same unit as the model's output and, consequently, can provide a better interpretation compared with using total gains. Moreover, Shapley values have advantages over the other methods of this class, as they can provide a single unique solution to assigning contributions with desirable properties (Lundberg et al., 2020).

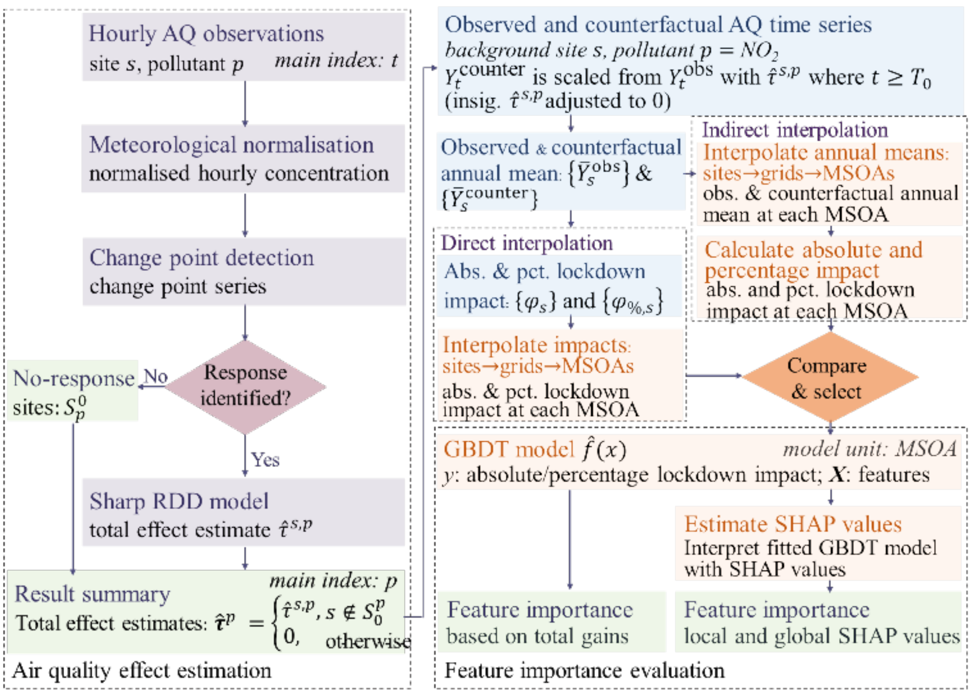

Figure 1. Graphical summary of the methodology for air quality effect estimation (left) and feature evaluation (right). The output/conclusion is coloured green. The left column (purple) goes through each concentration time series. The right column is either conducted on each monitoring site (blue) or focused on MSOAs (orange).

3. RESULTS AND DISCUSSION

COVID-19 lockdown effects on air quality

Road traffic in London dropped by up to 65% during the lockdown (Transport for London, 2020). However, our results show that the relative change in NO_2 caused by the lockdown ranged from -50% to 0% at different monitoring sites. Among them, 44% of the 45 sites close to roads (roadside sites) and 70% of the 33 sites measuring town-wide pollution levels (background sites) showed a small reduction (<10%) or a null response. Aggregating the effects across London, the lockdown reduced NO_2 concentrations by 12% at roadside sites and 7% at background sites on average. For other pollutants, the relative change attributable to the lockdown ranged from -62% to 0% for NO_x, 0% to 4% for O_3, -5% to +0% for PM_{10}, and there was no response for $PM_{2.5}$. Unlike NO_2 and NO_x, the regional contribution to PM and O_3 is substantial (Greater London Authority, 2020). Our results imply that reducing transport activities and restricting exhaust emissions are not sufficient to tackle air pollution, particularly for those at background locations and for pollutants that are largely affected by regional emission sources.

Our estimated air quality impacts of the lockdown are generally consistent with previous studies in that we find more marked reductions in NO_2 concentrations yet less significant changes in other regulated pollutants. However, our estimates are not as large as those of Jephcote et al. (2021), who reported an average reduction of 38% and 17% respectively in NO_2 and $PM_{2.5}$ concentrations and an average increase in O_3 concentrations of 8% across the UK during the first lockdown. Jephcote et al. (2021) estimated the impacts by comparing the air quality observations between the lockdown period and the same period in previous years. However, this approach may be biased by differing meteorological conditions and long-term trends in air quality. Particularly, air quality in the UK (both NO_2 and PM) has improved year on year in most major cities, including London (Department for Environment Food & Rural Affairs, 2020; Ma et al., 2021b). Therefore, comparing air pollution levels across different years is likely to overestimate the air quality impacts attributable to the lockdown.

Another study by Shi et al. (2021) reported abrupt but smaller than expected changes in air quality attributable to the lockdown in 11 cities globally. For London, they found that the lockdown changed background concentrations of NO_2, O_3, and $PM_{2.5}$ respectively by -8 $\pm$ 8%, -2 $\pm$ 8%, and +11 $\pm$ 17%. While our results for NO_2 and O_3 are similar, the results for $PM_{2.5}$ differ. Shi et al. (2021) estimated the lockdown impact by the relative change in air pollution before and after the lockdown in 2020 after subtracting the relative change over the same period in the average concentrations across the previous 4 years. Their estimated impacts for PM could be biased by PM episodes during their research period, between March and April in 2016-2020. PM episodes are a regular feature in springtime in western Europe; particularly, two episodes due to regional pollution transport were recorded in London in their specified post-lockdown period (Imperial College London, 2021). Although they applied a meteorological normalisation, this technique may not be effective to control for abrupt natural events or regional pollution transport (Shi et al., 2021). Consequently, the pollution increase estimated in Shi et al. (2021) for $PM_{2.5}$ may incorporate the influence of these recorded episodes. In contrast, our study focuses on the time around the start of the lockdown, using CPD for response identification and subsequently a sharp RDD for effect estimation, and are therefore less susceptible to the influence of pollution episodes.

Factors affecting lockdown effects

The first UK national lockdown is estimated to have caused spatially heterogeneous impacts on NO_2 concentrations in London. By using explainable ML, we find that the degree to which NO_2 was reduced in an area was mostly correlated with the proportion of heavy goods vehicles (HGVs) in road traffic before the lockdown, the distance to London Heathrow Airport (LHR), and the distance to the Central Activities Zone (city centre); see Figure 2. This finding is generally consistent with Yang et al. (2021), who found the NO_2 reduction in Los Angeles during the lockdown was primarily due to the changes in HGVs' activities. In addition to feature ranking, our results also show that a lower proportion of HGVs in road traffic and closer proximity to LHR are associated with greater pollution reductions in most MSOAs; however, different areas can have opposite effects from a short distance to the city centre.

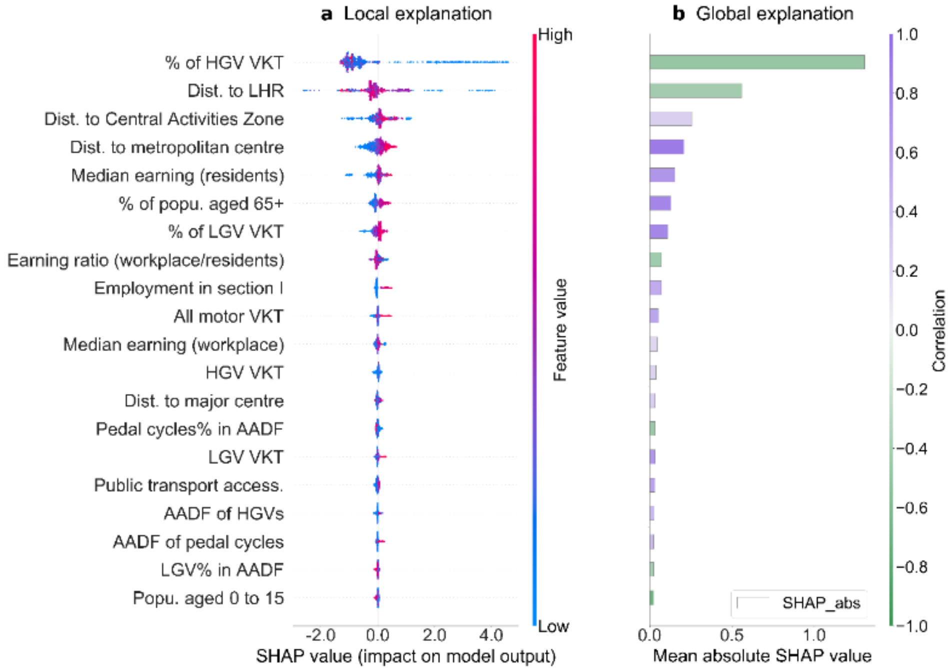

Figure 2. Estimated local SHAP values (left) and global feature importance (right) quantifying the contribution of each feature to the relative reduction in NO_2 due to lockdown. **a** Each point is a SHAP value specific to a feature and an MSOA; the colour of a point indicates the feature value for that MSOA; points are spread out in the y-axis to avoid overlap. **b** The global importance is calculated as the mean absolute SHAP value across MSOAs; the colour indicates the Pearson correlation between feature values and SHAP values across MSOAs; for example, a negative correlation (green) indicates that a higher feature value is generally associated with a smaller pollution reduction. Features are sorted by global feature importance; the top 20 features are plotted.

Furthermore, our results reveal that the existing health inequalities within the city were exacerbated during the lockdown: areas with lower-income residents, which have also been exposed to high levels of air pollution before the pandemic (Barnes et al., 2019), typically experienced smaller NO_2 pollution reductions during the lockdown. Particularly, areas with median incomes below the 10^{th} percentile experienced a mean decrease in pollution reduction of 0.6 percentage points (pp), while those with median incomes above the 90^{th} percentile experienced a mean increase in pollution reduction of 0.3 pp (Figure 3). Moreover, we find places with higher public transport accessibility levels were generally associated with a larger pollution reduction (Figure 3). Therefore, our results highlight the link between existing social inequalities and the effects of air pollution control policies.

While close proximity to the city centre is estimated to enhance pollution reduction in several cases (Figure 2), the areas that are closer to metropolitan town centres were commonly less affected compared with the areas further away (Figure 3). As people tended to make local trips during the lockdown, our result implies the potential of alleviating air pollution in the city centre by revitalising local town centres to reduce the need for travel. Particularly, a shift towards remote/hybrid working may continue in the UK after the pandemic (Transport for London, 2020), which is likely to further enhance the role of town centres. However, policymakers should be mindful of the potential deterioration of air quality around town centres following the development.

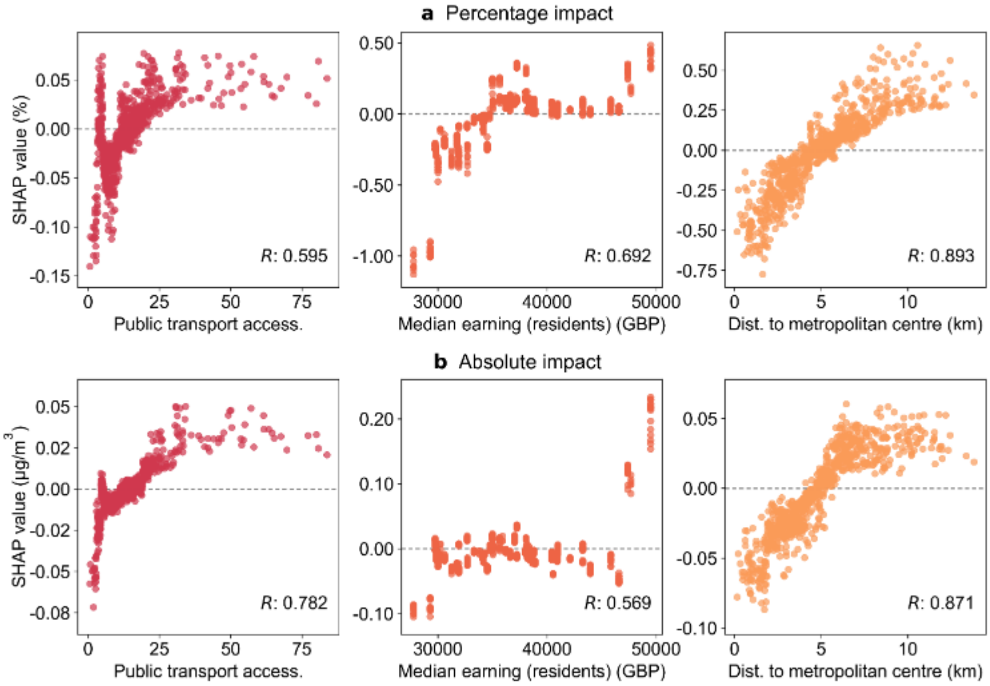

Figure 3. Relationship between public transport accessibility level, median gross annual pay for residents, the distance to the nearest metropolitan town centre, and the SHAP value for **a** the relative impact (pp) and **b** the absolute impact (µg/m^3). Each point shows the feature value (x-axis) and corresponding SHAP value (y-axis) at a particular MSOA. The sign of the SHAP value indicates how the feature value contributes to the predict of the lockdown impact at that MSOA (positive: increasing pollution reduction; negative: decreasing pollution reduction). The Pearson correlation (R) between the feature value and the corresponding SHAP value across MSOAs is labelled on each plot.

4. CONCLUSIONS

Our study shows that the unprecedented decrease in transport activities following the COVID-19 lockdown led to an incommensurate reduction in air pollution. Improving air quality in cities requires a multi-faceted set of policies to control emissions across sectors and full consideration of inequalities. Within the transport sector, a sustained effort is necessary to consistently guide the city for health and equality, such as by reducing emissions for freight transport and airport-related activities, facilitating and encouraging active travel and public transport, and integrating transport planning with land use planning.

ACKNOWLEDGMENTS

L M was funded by the Dixon and Skempton Scholarships from the Department of Civil and Environmental Engineering, Imperial College London.

REFERENCES

Barnes, J. H., Chatterton, T. J., & Longhurst, J. W. S. (2019). Emissions vs exposure: Increasing injustice from road traffic-related air pollution in the United Kingdom. *Transportation Research Part D: Transport and Environment*, *73*, 56–66. https://doi.org/10.1016/j.trd.2019.05.012

Dang, H. A. H., & Trinh, T. A. (2021). Does the COVID-19 lockdown improve global air quality? New cross-national evidence on its unintended consequences. *Journal of Environmental Economics and Management*, *105*, 102401. https://doi.org/10.1016/j.jeem.2020.102401

Fassò, A., Maranzano, P., & Otto, P. (2021). Spatiotemporal variable selection and air quality impact assessment of COVID-19 lockdown. *Spatial Statistics*, 100549. https://doi.org/10.1016/j.spasta.2021.100549

Greater London Authority. (2020). *Air Quality In London 2016-2020*. https://www.london.gov.uk/sites/default/files/air_quality_in_london_2016-2020_october2020final.pdf

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv*, *51*(93). https://doi.org/10.1145/3236009

Horálek, J., Schreiberová, M., Schneider, P., Kurfürst, P., Schovánková, J., & Ďoubalová, J. (2019). *European air quality maps for 2016* (Issue 9).

Imperial College London. (2021). *LAQN Pollution Episodes*. https://www.londonair.org.uk/london/asp/publicepisodes.asp

Jephcote, C., Hansell, A. L., Adams, K., & Gulliver, J. (2021). Changes in air quality during COVID-19 'lockdown' in the United Kingdom. *Environmental Pollution*, *272*, 116011. https://doi.org/10.1016/j.envpol.2020.116011

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Ma, L., Graham, D. J., & Stettler, M. E. J. (2021a). Air quality impacts of new public transport provision: A causal analysis of the Jubilee Line Extension in London. *Atmospheric Environment*, *245*. https://doi.org/10.1016/j.atmosenv.2020.118025

Ma, L., Graham, D. J., & Stettler, M. E. J. (2021b). Has the ultra low emission zone in London improved air quality? *Environmental Research Letters*, *16*(12), 124001. https://doi.org/10.1088/1748-9326/ac30c1

Menut, L., Bessagnet, B., Siour, G., Mailler, S., Pennel, R., & Cholakian, A. (2020). Impact of lockdown measures to combat Covid-19 on air quality over western Europe. *Science of the Total Environment*, *741*, 140426. https://doi.org/10.1016/j.scitotenv.2020.140426

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2022). General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. In *Lecture Notes in Computer Science* (Vol. 13200, pp.

39–68). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-04083-2_4

Shi, Z., Song, C., Liu, B., Lu, G., Xu, J., Van Vu, T., Elliott, R. J. R., Li, W., Bloss, W. J., & Harrison, R. M. (2021). Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Science Advances*, *7*(3), 6696–6709. https://doi.org/10.1126/sciadv.abd6696

Transport for London. (2020). *Travel in London Report 13*. http://content.tfl.gov.uk/travel-in-london-report-13.pdf

Vadrevu, K. P., Eaturu, A., Biswas, S., Lasko, K., Sahu, S., Garg, J. K., & Justice, C. (2020). Spatial and temporal variations of air pollution over 41 cities of India during the COVID-19 lockdown period. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-72271-5

Venter, Z. S., Aunan, K., Chowdhury, S., & Lelieveld, J. (2020). COVID-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 18984–18990. https://doi.org/10.1073/pnas.2006853117

Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J. P., Pennington, E. A., Wang, Z., Wu, Y., Sander, S. P., Jiang, J. H., Hao, J., Yung, Y. L., & Seinfeld, J. H. (2021). From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(26), 2102705118. https://doi.org/10.1073/pnas.2102705118