Author’s response to reviewer comments

We are thankful for the constructive reviews. Below we list all referee remarks and suggestions (in *italics)* along with our responses. The revised version of the manuscript afterwards with all changes highlighted in red is provided at the end of this document.

Anonymous Referee #1

We thank the reviewer for his/her time and the thoughtful feedback. Below we list all referee remarks and suggestions (in *italics)* along with our responses.

Reviewer comment: *Keller et al. are investigating here the impact of COVID-19 restrictions on both NO_2 and O_3 surface concentrations. To estimate these changes taking into account the influence of the meteorology, they designed an interesting approach relying on a global simulation with the GEOS-CF model primarily bias-corrected using machine learning models. Compared to the recent studies covering this topic, the main strengths of this study are its spatial scale (since more than 5,000 stations in 46 countries are considered) and the fact that two important trace gases are included (NO_2 and O_3). The authors notably highlighted a strong variability of the NO_2 changes in general agreement with the level of mobility restrictions put in the different countries, while a lower response of O_3 is found. The paper is well written and relevant for our scientific community. It should thus be accepted after addressing one single major comment*
regarding the methodology and other minor suggestions.

Major comment :
- My major comment is about the machine learning methodology. Some information are missing or at least confusing, which explains why I classify it as a major comment but it might be only a minor one requiring only to provide more details in the text. My concerns are related to the way training and test datasets are obtained and how the machine learning models are tuned. Due to the substantial autocorrelation typically found in hourly air quality time series, using a random selection for splitting the datasets into training and validation data might lead to too optimistically good performances. For instance, the way I see it, if the model ingests a training data at a time t (and learn the corresponding model bias) and used for prediction at t+1, given that the model includes time features that allow to locate temporally this point, it will simply learn that around that time t, the model error is X, and then consider that the error at t+1 should also be close to X. In other words, the model might not learn properly the relationships between the model error and the features other than the temporal ones (most importantly, the meteorological parameters). In addition, it seems that no cross-validation (using for instance K-fold or time series cross-validation) is performed at any time (the word never appears in the manuscript), while this is important for tuning the models and ensuring robust estimates of their performance. Actually, it seems also that no tuning is performed during the preparation of the machine learning models. Also, there is often some confusion between the terms “training” (the phase in which you train your model), “validation” (the phase in which you tune your model and/or your select among different types of models) and “test” (the phase in which you evaluate the final performance of your final model already tuned). I guess what you mean here by “validation” is “test”? But since you are not mentioning how/if your models are tuned, it is quite confusing. Please clarify your methodology regarding these different points. On top of that, I agree with the comment of the editor that some discussion of the uncertainties of your approach should be included in the paper.

Author's response: We updated the machine learning methodology in the revised version of the manuscript and expanded its description. In the updated manuscript, all models have been trained using 8-fold cross validation, where the 8 batches represent quarterly chunks of model-observation pairs in order to minimize possible autocorrelation impacts. We also updated the notation of 'validation' and 'test' datasets.
These updates led to a slight deterioration of the model skill scores but have no discernible impact on the overall results and conclusions.

R: *Minor comments :*
- L73 : Which data availability at the hourly scale are you requiring for considering that a given day is valid? Please add this information
A: We only include days with at least 12 hours of valid data. We added this information to the manuscript.

R: *- Fig. 1 : Eventually, adding three panels zooming on North America, Europe and East Asia might be useful since the red points are completely hiding the blue and purple points in Europe and Asia (or if there are much less blue/purple points, you could plot them above the red points)*
A: We added figures with close-up maps of East Asia, Europe, and North America to the appendix.

R: *- L93 : You used OMI observations for scaling the anthropogenic emissions from 2010 to 2018. Why not scaling emissions up to 2019 included? Does the same procedure applied for 2019 highlights noticeable changes of NOx emissions between 2018 and 2019?*
A: We recognize that the initiali wording in this paragraph was misleading and we adjusted it in the updated version of the manuscript, with reference to the GEOS-CF description paper recently submitted for review (https://www.essoar.org/doi/10.1002/essoar.10505287.1).

R:*- L103 : Please indicate here that your XGBoost is making predictions at the hourly scale. Please also mention clearly that one XGBoost model is trained for each station, independently from the others.*
A: We added this information to the manuscript.

R: *- L110 : Which "mean" are you referring to here? The overall mean over the period 2018-2019? Or the seasonal monthly mean? Did you test applying the machine learning without removing outliers? Which performance is obtained ? Strongly stagnant conditions might lead to a peak of NO_2, and you want your model to learn this type of event, so at first sight, I don't understand why this step is needed (or even wanted). Please provide here a more complete justification of your methodological choices.*
A: The main motivation for this approach was to adjust for obviously erroneous observations, such as ozone or nitrogen dioxide concentrations of several thousand ppbv. Such values can occur in the OpenAQ database, whose values are reported in real-time and are not backfilled with quality-controlled data. To support this point, we performed two sensitivity simulations using more stringent thresholds of 3 or 4 standard deviations and did not find any change in our results.

R: *- L110 and Figure 2 : Please comment a bit more your results. Notably, I am wondering why your results are different at stations around #3000. Is this a specific region? Do you have any idea of the reason for that? Also, I am curious, why not simply normalising the RMSE by the average concentration? (rather than the range between 5th an 95th percentiles)*
A: We reordered the stations to reflect the four major regions considered in this study (China, Europe, USA, rest of the world). We chose the percentile window as the denominator for the NRMSE because it offers a better reflection of the concentration variability at the given site. The results using the RMSE normalized by the annual mean would look qualitatively very similar.

R: *- L116 : You mention 49 species and 31 modelled emissions : please provide the complete list of the species taken into account here (eventually in Supplementary Material or Appendix)*
A: The full list of input features is given in Table A2 in the Appendix.

R: *- Section 2.3 : Please indicate the features importance obtained with XGBoost, for both NO_2 and O_3. This is an information especially interesting in your study since you are using a lot of features, many of them probably not very useful for making the predictions (?)*
A: We added a new paragraph to the revised version of the manuscript, discussing the SHapely Additive exPlanations (SHAP) values for both the NO_2 and O_3 bias correctors in more detail. The SHAP values are similar to the ‘classic’ feature importance but better take into account the role of feature interactions. The distribution of all input feature importances is shown in the Figures A4 and A5 in the appendix.

R: *- L151 : Just to know, how these cities have been selected? Were they selected following an objective approach (for instance, all largest cities or cities with strongest data availability) or arbitrarily?*
A: We chose these 5 cities rather arbitrarily. Wuhan, Milan and New York represent early outbreak ‘hotspots’ that received a lot of media attention, and Taipei and Rio de Janeiro offer examples of different government responses to the pandemic (as also reflected in the data). We provide more detail on our motivation for showcasing these 5 cities in the revised version of the manuscript.

R: *- Fig. 5 : It would be useful to indicate the number of stations included in each country (for instance in the title of each panel).*
A: We provide the number of sites in the inset of each figure.

R: *- L182 : I would expect that the machine learning model (trained with data from late 2018 to end of 2019) would learn the reduction of NO_2 associated to the Chinese New Year, but the results for 2019 presented in Fig. 5 suggest that it is not the case. Any idea of the possible reason for that ? Is it simply because no training data are available in the first part of 2018 ? In any case, this could reduce the trust we have in the prediction done in 2020, at least for this specific country and this specific period of the year. Maybe including a new input variable representing if a day is holiday or not could help solving this issue.*

A: This is an excellent comment and the idea to add holidays as an additional input feature is intriguing (albeit somewhat cumbersome to implement on a global dataset!). We are actually quite happy to see that the model did not learn the NO_2 reduction associated with Chinese New Year, as such a behavior in our eyes would indicate a possible overfitting. Rather, we hope to capture the 'regular' model bias with the machine learning models and accept the fact that unusual events such as holidays cannot be captured. We feel this is the more conservative approach, especially since for China, we would have only one holiday to train the model on (year 2019).

R: *- The authors evaluated their machine learning models by checking the mean biases, errors and correlations over the entire period, which is fine for the analysis conducted in Sect. 3.1. The analysis of the diurnal variations of O3 and Ox is interesting but should come with an evaluation of the performance of the machine learning models at the diurnal scale : does the bias of the machine learning models show any diurnal variability ? I think it is important to show (eventually in the Supplement) and discuss a diurnal plot of the bias (similar to Fig. 8a) for both training and test datasets, just to ensure that the very small mean biases obtained over the entire period do not hide error compensation of stronger biases during specific times of the day.*
A: We added the hourly skill scores of the test data set in the appendix, and also note it in the discussion of the results. Note that the skill scores for the training and validation data show the same indifference to the time of the day.

R: *- L205 : I am not sure we can consider that the stronger seasonality of O3 compared to NO2 would bring more challenges since the seasonality is expected to be taken into account by the "month" feature. However, prediction business-as-usual O3 is possibly more challenging than for NO2 due to the more complex processes driving its concentration (e.g. secondary pollutant produced by more complex chemical reactions, involving more numerous precursors, potential strong influence of dry deposition, long-range transport)*
A: We agree with the reviewer and changed the wording in the updated version of the manuscript to reflect the fact that compared to NO_2, O_3 concentrations are much more influenced by large-scale processes and the local O_3 signal is thus expected to be much smaller.

R: *- L225 : The results shown for Ox are based on a subset of stations where both NO2 and O3 collocated measurements are available? If yes, is the same subset used for showing the results of O3 alone? Please clarify this point. In any case, it would be nice to have both O3 and Ox on a similar subset of stations to allow fair comparisons.*
A: The analysis is indeed based on the subset of stations where both NO_2 and O_3 observations are available. We clarified this in the manuscript.

R: - L262 : "natural background NO2"
A: We changed the wording as suggested.

R: *- L260 : Is this value of 80% obtained at global scale? How variable it is spatially (and more specifically, from one country to the other)?*

A: The 80% average sensitivity is the global mean value over the simulated sensitivity period (Dec-Jun). For the emission calculation, we updated the methodology and now use a variable NO_x/NO_2 ratio, depending on the inferred (percentage) NO_2 decrease. We acknowledge that this is still a simplification as the NO_x/NO_2 ratio is variable in both space and time. To account for this, we assign a rather large (absolute) uncertainty of 15% to the NO_x/NO_2 sensitivity ratio. We updated the manuscript, figures and tables accordingly.

***R:** - L263 : Why using EDGAR[2015] rather than HTAP[2018] ?*
A: We chose EDGAR over HTAP because it's baseline inventory is more up-to-date (2015 vs. 2010). We added this information to the manuscript.

***R:** - Fig. 9a : Results shown in Fig. A7 figure are not exactly what I would expect and thus deserve more discussion, highlighting more clearly the potential uncertainties. For instance, if I understand correctly, NO_2 emissions (estimated using the OMI NO_2 tropospheric column taken here as a proxy of the NO_x emissions) would have decreased more strongly during February 2020 (before the lockdown) than in March-April (during the lockdown), which is likely not true. A potential issue I see here is that the authors are not taking into the influence of the meteorology on the NO_2 tropospheric columns.*
A: The main goal of the sensitivity simulation was to obtain NO_x/NO_2 sensitivity ratios for a wide variety of (realistic) emission changes. Rather than using a fixed NOx emission ratio (as e.g., done in Lamsal et al., 2011), we chose to use the OMI NO_2 tropospheric columns as a proxy for emission changes. This is an obvious oversimplification but serves the stated goal of the sensitivity simulation. We clarified this aspect in the updated version of the manuscript.

***R:** Also, the final number of 5% of reduction of global NOx emissions should also be discussed. Is it consistent with what we could expect during that COVID-19 period, namely a strong reduction of traffic emissions? (what is the contribution of traffic to global NOx emissions?). Therefore, please discuss in more detail this section.*
A: Traffic emissions are approximately 27% of total anthropogenic NO_x emissions. Using this information, we estimate that our derived NOx emission reductions correspond to 17-24% of global traffic emissions. We added a discussion on this to the manuscript.

***R:** - L305-317 : This paragraph corresponds to a new analysis and should thus be included in a dedicated section rather than in the conclusion. Also, a more detailed information should be provided regarding this work. The authors say "we assume a sustained reduction in global anthropogenic emissions of NOx, CO and VOCs". Which reductions were used for CO and VOC emissions ? Also, given that the estimated reduction of NOx emissions is highly variable in time (Fig. 9b), to what "sustained" corresponds here?*
A: We moved this analysis to its own paragraph (Section 3.4.) and added more detail on the methodology of this sensitivity experiment. The emission reduction used for the forecast simulation was fixed at -20%, i.e., assuming no variability in time. This is an obvious simplification but serves the stated purpose of the sensitivity experiment.

***R:** - Figures in Appendix : Please increase the resolution of these plots.*

A: We changed the layout to 4 panels per column to increase the resolution.

Anonymous Referee #2

We thank the reviewer for his/her time and thoughtful feedback. Below we list all referee remarks and suggestions (in *italics)* along with our responses.

Reviewer comment: *In studying the effect of COVID-19 restrictions on air pollution, the meteorological variability complicates a direct comparison with pre-lockdown periods. The authors are well aware of this, and tackle this problem by comparing ground observations against model simulations based on a business-as-usual emission inventory. Local modelling biases (due to representation error, wrong emissions, meteo, or chemistry) are corrected for by a machine learning approach, trained in a pre-lockdown period (2018- 2019). The paper is well written, and presents a sound and well-developed approach, hence I recommend its publications after addressing the following minor issues.*
I agree with the major comment of the previous reviewer to provide more details about the machine learning methodology and how the potential pitfall of autoregression of time series is dealt with.
Author's response: We overhauled the machine learning methodology in the revised version of the manuscript to better address the potential issue of auto-correlation, and overall expanded significantly on the description of the methodology and associated uncertainty estimation.

R: *For NO_2, The machine learning approach appears to be surprisingly powerful to adjust a rather coarse chemical transport model (25 x 25 km2 resolution) to the local situation, given the strong gradients found in cities. Figure 2: it would be interesting to make a distinction between (rural) background stations and street stations. Is the bias correction method sufficiently strong to solve the representation error of the latter category?*
A: We found no difference in skill scores between background sites and polluted sites, and added this information to the manuscript.

R: *Figure 4 shows underprediction of the uncorrected model for Milan and Taipei, overprediction for NYC, and alternating under- and overprediction for Wuhan. In my opinion, your analysis in 3.1 lacks some words about what we can learn from the modelling biases. Are representation errors dominating, or are we looking at e.g. wrong emission estimates?*
A: We added a (short) discussion about the possible reasons for the model bias to the revised version of the manuscript.

R: *Figure 5, just out of curiosity: is there a reason why so many observation sites in Romania measure significantly higher NO, than expected by the BCM?*
A: The large uncertainty range in Romania was caused by two sites with much higher NO_2 concentrations than the BCM. Because we used the overall 5/95% quantiles as uncertainty estimate, this resulted in the shown large uncertainty range. For the updated version, we completely overhauled the uncertainty calculation, which now in our view results in more realistic uncertainty estimates. For instance, our uncertainties are now based on the model-observation mismatches obtained on the test data, and the stated uncertainty estimates are higher for countries with only a few observations compared to countries with a dense network.

R: *Figure A1: Showing results for more Chinese mega-cities would be instructive, especially given the strong local observation network in China.*
A: We added three more Chinese cities to the analysis (Chongqing, Guangzhou, and Tianjin).

R: *Figures A1-A3: Sometimes strong NO2 reductions are already visible months before the official lockdown starts (e.g. Ljubljana, Vienna, Dublin, Boston, and Denver). Any explanation?*
A: Many countries issued 'soft' stay-at-home orders before the 'hard' lockdowns started, and in many locations the NO_2 observations start to reflect this change in human behavior ahead of the lockdowns. We discuss this now in more detail in the newly added Section 2.4 (Lockdown dates).

R: *Figures A1-A3: the blue and red lobes in the pre-COVID period can be used to estimate the error in your methodology and put the results (e.g. in Table A1-A3) in better perspective.*
A: As already mentioned above, we reworked the uncertainty estimates based on the model-observation mismatches on the test data. This is similar to the here suggested approach but a bit more restrictive as it is based on the test data only.

R: *Figures A1-A3: I am missing an indication of n, the number of observation sites used for each city.*
A: We added this information to the figures.

R: *Section 3.2: Personally, I find the results for O_3 less striking, although I directly admit that an O_3 analysis is more subtle and less straightforward than NO_2. Figure 8a shows the flattening of diurnal cycle, which is used to explain the marginal effect of the measures on average O_3 concentrations in Figure 6 and 7. I think it would be more interesting to see these figures for daily peak values of O_3, instead of daily mean values.*
A: We considered this but were worried about 'sensationalizing' our findings by focusing on the ozone peak values. While focusing on the afternoon (or daytime) ozone values is common, the goal of this study was to analyse the overall impact of COVID-19 lockdowns on ozone and we thus find it more appropriate to show the daily mean changes. The changes in afternoon ozone (as well as nighttime ozone) is discussed in detail in Section 3.2 and highlighted in Figure 8.

R: *Section 3.3, lines 247-252: I had to read this several times to understand, and I am still not sure if I do by now. First it is stated that NO_2 concentrations do not change 1:1 with changing NO_x emissions, but in the following sentence it is suggested that NO_2 columns from OMI are used to scale underlying NO_x emissions. Also, I can not deduce how the sensitivity study is set up exactly. Please rewrite.*
A: We updated the description of the sensitivity experiment in the revised version of the manuscript.

R: *Section 3.3: Your emission reduction results (e.g. Figure 9b) are potentially prone to sampling biases. According to Figure 5, the results for India are based on only 7 stations (!). Furthermore, as the ground-based monitoring stations are typically located in cities, the results* *reflect emission reductions within cities (such as traffic), but not necessarily emission reductions of other sectors such as industry or power plants. This should be addressed in a short discussion.*

A: Our emission estimates for countries such as India or Brazil are indeed susceptible to errors from a variety of sources, including sampling errors and the assumed NO_x/NO_2 ratio. We revisited the uncertainty calculation in the new version of the manuscript to better reflect these uncertainties, and also expanded the discussion in section 3.3 (in addition to adding a new section 2.3.4 dedicated to the calculation of uncertainty associated with the machine learning methodology).

R: *Conclusions: lines 305-313 describe an additional experiment about the effect of NOx emission reduction on surface ozone, which, according to my taste, should be shifted backward (e.g. in an additional section 3.4) before the conclusions start.*

A: We moved this analysis to a separate section 3.4 (Long-term impact of reduced NO_x emissions on surface O_3)

Referee #3

We thank Kirsten de Nooijer for the time taken to review this paper and for the thoughtful feedback. Below we list all referee remarks and suggestions (in *italics)* along with our responses.

Reviewer comment: *Major comment 1): Statistics need to be included. The aim of the paper is to quantify, uncertainties should be quantified as well. The reported numbers are easily disregarded without the proper statistics (e.g., p-values, t-tests or z-tests) and uncertainty ranges. This problem is present in figures 4 and 6 (but applies for all given emission changes in the manuscript). Here the difference between the BCM prediction and observations is shown, but without noting whether this difference is significant (or perhaps falls within the uncertainty range of the BCM prediction).*
Author's response: We updated the uncertainty estimation based on model-observation comparisons on the test dataset, and propagate the estimated uncertainties per location site to a city and country level. The numbers in the updated version of the manuscript include the estimated uncertainties. In addition, we highlight statistically significant concentration changes in the concentration tables provided in the Appendix (Tables A3-A8), using a (stringent) p-value of 0.001.

R: *Example 1, line 145: "For Wuhan, we find a reduction in NO_2 of 60% relative to the expected BCM value for February and March 2020, and similar decreases are found over Milan (60%) and New York (45%) starting in mid-March and lasting through April (Fig. 4; Tables A1-A3)." How certain are these numbers? Is it between 62% and 58%? Or between 70% and 50%? I urge the authors to please quantify the uncertainty of these numbers by providing uncertainty ranges or mentioning of significance. This could be implemented similar to Le Quéré et al. (2020), here reductions in emissions are provided by stating the range (representing ±1σ) instead of a single number.*
A: The estimated uncertainty ranges are provided in the updated version of the manuscript. The stated uncertainties take into account the number of observation sites, so that estimates that are based on fewer sites result in higher uncertainties (all else equal).

R: *Example 2, line 228: "Compared to the BCM model, there has been an increase in the concentration of night time O3 (midnight-5.00 local time, Fig. 8a) by 1 part per billion by volume (ppbv = nmol mol-1) compared to the BCM, whereas Ox shows a decrease of 1 ppbv (Fig. 8b)." Is this reported 1 ppbv difference significant? I highly suggest you to report whether the modelled change is significantly different from the observations. The recent paper by Liu et al., (2020), also referenced by Keller et al., does report significance and thereby makes a more compelling case. Liu et al. derived uncertainty from 10000 Monte Carlo simulations from monthly statistics to estimate a 68% confidence interval. This procedure could be followed here as well. Another suggestion is to provide a paragraph on uncertainty estimation for the machine learning algorithm in the method section, similar to Petetin et al. (2020). Perhaps here the method of Hengl et al. (2017) could be useful. They describe a procedure for machine-learning uncertainty estimation with the use of the program R and the package 'xgboost'.*

A: We added a section on the uncertainty estimation to the methods (Section 2.3.4) and use these uncertainties to quantify the significance of our findings. Based on this, we conclude that the 1ppbv change is indeed statistically significant, and we now state so in the manuscript.

R: *- Major comment (2): It's unclear how numbers in the result section are constructed from the represented data, no calculation steps are mentioned in the method section. Most importantly, how is the reduction in global NOx emissions of 2.9 TgN calculated?*
A: We revisited the description of the emission calculation, offering much more detail on the methodology to hopefully make it easier to follow.

R: *Line 253 states the following: "This results in anthropogenic emission adjustment factors of 0.3 to 1.4 (Fig. A7)." Because of the lack of clarification on calculation steps or argumentation, it is unclear how the adjustment factors of 0.3 and 1.4 are determined. Is perhaps the approach of Mendoza & Russel (2001) used to derive adjustment factors for NO_x emissions? Please refer to the used methodology or provide the calculation steps. The in the manuscript referred figure A7 does not provide the calculations either (even though this seems to be suggested). Figure A7 only shows the monthly average perturbations applied to the 2018 anthropogenic base emissions, ranging from 0.5 to 1.5. As a consequence, the resulting quantification of reduction in emissions loses credibility.*
A: As already stated above, we updated the description of the emission calculation and also adjusted the uncertainty estimation, which now includes uncertainties for both the estimated NO_2 reductions and the assumed NO_2/NO_x ratio. The emission estimates reported in the revised version of the manuscript now include these uncertainty estimates.

R: *Lines 262-266: "Based on bottom-up emissions estimates for 2015 from the Emission Database for Global Atmospheric Research (EDGAR v5.0_AP, Crippa et al., 2018, 2020) and using a constant concentration/emissions ratio of 0.8 based on the best fit line obtained from the model sensitivity simulation (dashed purple line in Fig. 9a), we calculate that the total reduction in anthropogenic NO_x emissions due to COVID-19 containment measures during the first six months of 2020 amounted to 2.9 TgN (Fig. 9b and Table 2)."*
It is clear a calculation is performed, but not how. How is the quite important 2.9 TgN reduction in anthropogenic NOx emission due to COVID-19 containment constructed? The 2.9 TgN is not in the referred Table 2 nor in Figure 9b. I urge the authors to provide the taken calculation steps resulting in the (quite important) 2.9 TgN reduction in anthropogenic NO_x emission. This will improve the credibility of that given number.
A: The methodology to calculate the emissions is now described in much more detail, along with a discussion of the corresponding uncertainties.

R: *- Major comment (3): The manuscript mentions 'lockdown' situations but does not provide a definition of 'lockdown'. The restrictions vary per country (Ravindran & Shah, 2020) and the definition will have consequences on changes in NO_2 emissions. Some countries only enforced restrictions based on time, while keeping most forms of transport, schools and business open. Others have been reported to only had restrictions for part of the country. Please provide a definition of 'lockdown'.*

A: A clear definition of lockdowns is indeed complicated by the various responses, often even within regions of a country. We provide the list of used lockdown dates in the Appendix (Table A2). In general, we emphasize that the main purpose of the lockdown dates are to guide the reader in the interpretation of the figures, rather than using them at 'face value' for statistical analysis. The interpretation of lockdown dates is further complicated by the fact that many countries issued 'soft' lockdowns before the official lockdowns, which already altered human behavior and resulted in a decrease in NO_2 concentrations in advance of the official stay-at-home orders. We discuss this problem in the newly added Section 2.4 in the manuscript.

R: *Lines 156-162: For Taipei and Rio de Janeiro, the observations and the BCM show little difference (Fig. 4), consistent with the less stringent quarantine measures in these places. Other cities with only short term NO2 reductions of less than 25% include Atlanta (USA), Budapest (Hungary), and Melbourne (Australia), again correlating with the comparatively relaxed containment measures in these places (Fig. A1-A3). In contrast, Tokyo (Japan) and Stockholm (Sweden), which also implemented a less aggressive COVID-19 response, exhibit* NO_2 *reductions comparable to those of cities with official lockdowns (>20%), suggesting that economic and human activities were similarly subdued in those cities."*
This suggests that degrees of reduction in NO_2 *emissions are linked to severity in measures taken by local governments (e.g. lines 156-162), however, the severity of measures per country are not characterised. I suggest providing an overview of 'lockdowns' via a table including severity of measures and start and end dates. As an example, take a look at Ravidran & Shah (2020) where countries were classified on severity by introducing colour codes.*
A: We added the lockdown dates used in this study to the Appendix (Table A2) but refrain from adding a lockdown severity measure because we don't feel comfortable with such a number on a country scale. For instance, how should one evaluate the severity of the lockdown for the United States where some cities (e.g. New York) were under a complete lockdown while other places saw little (official) restrictions? Rather, we emphasize in the manuscript that the main reason for adding lockdown dates is to support the visualizations.

R: *Line 142: The start and end dates for these are from https://en.wikipedia.org/wiki/COVID19_pandemic_lockdowns or based on local knowledge."*
Because of Wikipedia's quickly changing contents, stating the start and end dates in a table will be an improvement on the derived results and will be more concrete than the stated 'local knowledge.
A: We added the list of lockdown dates to the Appendix and also provide the date at which the lockdown dates were accessed.

R: *Lines 21 22: Reductions in* NO_2 *correlate with timing and intensity of COVID-19 restrictions, ranging from 60% in severely affected cities (e. Wuhan, Milan) to little change (e. Rio de Janeiro, Taipei)."*
Also, the manuscript mentions correlations in timing and intensity of COVID-19 restrictions and reductions in NO_2 *(e.g. lines 21-22). A quantification of this correlation is however missing. Are these findings only based on eying the figures? Was a correlation test performed? I recommend adding quantification of the correlations.*

A: We didn't mean to use the word correlation in the literal sense here, and recognize that its use was misleading. We changed the wording accordingly as we don't think that a correlation analysis of the derived concentration changes to the lockdown dates is scientifically warranted.

-Minor comments:
R: *Table 1: The links for AEROS (Japan) and EPA Victoria (Melbourne, Australia) do not work.*
A: We couldn't find any issues with the links but updated them again in the updated version of the manuscript.

R: *119: Provide an argumentation on why all observations below or above 2 standard deviations from the mean are removed, contrary to Ma et al. (2020) where observations below or above 3 standard deviations were removed.*
A: We updated the discussion about the removal of outliers (and its motivation), and also conducted two sensitivity studies using a threshold of 3 and 4 standard deviations, respectively. These sensitivity runs did not show any change in our results.

R: *Figures 2 and 3: The presentation of the machine learning statistics could be simplified in form of a table. I fail to see how the representation of the machine learning statistics in a graph are useful to the reader (including the location#, since no information is supplied to deduct which location# is which location). I suggest replacing figures 2 and 3 by a table providing statistical performance, similar to Table 4 of Ivatt & Evans (2019).*
A: We updated the figure so that statistics are grouped by region (as suggested by another reviewer), and discuss the statistics in more detail in the newly added Section 2.3.3.

R: *Figures 4 and 6: Reductions in % are difficult to read in the figures, one must go back to the text for the actual numbers. Consider including the numbers in the figures, so they stand stronger by themselves. Both figures could be shortened on the x-as as well, starting at 2019. The (incomplete) data from 2018 does not contribute to the results. I would even consider replacing both figures 4 and 6 entirely by new figures that better meet the objective of quantifying the difference in reductions of* NO_2 *and* O_3 *concentrations (including notification of significance or uncertainty ranges, see major comment 1).*
A: The percentage reductions are provided in the figures in the appendix as well as the tables, and the uncertainties are stated in the tables. The main objective of Figures 4 and 6 is to introduce the overall concept of our methodology and to show comparisons of observations and model values before and after the bias-correction. The time range 2018-mid-2020 is shown to highlight the full extent of the analysis data and to highlight how the model-observation comparisons look like for the entire previous time period (where available). Most other figures in the manuscript focus on relative changes derived from the bias-corrected model (e.g., the figures in the Appendix or Figures 5 and 7), and we find it important to show the full time series of the baseline model (as well as the effect of the bias-correction) for both O_3 and NO_2 in at least one figure.

R: *191: Consider replacing the vague terms 'some countries' and 'most countries'. These results are stronger when presented in numbers, for example: '42 out of 46 countries...'*

A: We updated this to ‘29 out of 36 countries…’.

R: *208: Reconsider the phrasing of this result. Belgium, Italy, Luxembourg and Switzerland do not all four show pronounced peaks in early April, based on Figure 7.*
A: We changed the wording in the updated version of the manuscript.

R: *221-225: Consider including chemical equations of the mentioned processes to improve readability of this paragraph.*
A: Detailed explanation of ozone chemistry, including the chemical equations, are provided in the references. We added an additional reference to Seinfeld and Pandis (2016) and also provide another reference for the NO_x/NO_2 ratio (Shah et al., 2020) in the updated discussion of the emissions calculation.

R: *252-258: Move this text to methods, it seems out of place here in the result section.*
A: We expanded the description of the sensitivity simulation, so that this paragraph now hopefully seems less out of context. We prefer keeping it in this paragraph so that the entire section stands on its own.

R: *305-309: Move this text to methods as well, it seems out of place here in the conclusion section.*
A: This is now discussed in newly added section 3.4.

R: *Figure 10: Consider moving this figure to the result section instead of below the conclusion.*
A: This figure is now discussed in the newly added section 3.4.

Global Impact of COVID-19 Restrictions on the Surface Concentrations of Nitrogen Dioxide and Ozone

Christoph A. Keller[1,2], Mat. J. Evans[3,4], K. Emma Knowland[1,2], Christa A. Hasenkopf[5], Sruti Modekurty[5], Robert A. Lucchesi[1,6], Tomohiro Oda[1,2], Bruno B. Franca[7], Felipe C. Mandarino[7], M. Valeria Díaz Suárez[8], Robert G. Ryan[9], Luke H. Fakes[3,4], Steven Pawson[1]

[1]NASA Global Modeling and Assimilation Office, Goddard Space Flight Center, Greenbelt, MD, USA

[2]Universities Space Research Association, Columbia, MD, USA

[3]Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York, YO10 5DD, UK

[4]National Centre for Atmospheric Science, University of York, York, YO10 5DD, UK

[5]OpenAQ, Washington, DC, USA

[6]Science Systems and Applications, Inc., Lanham, MD, USA

[7]Municipal Government of Rio de Janeiro, Rio de Janeiro, Brazil

[8]Secretaria de Ambiente, Quito, Ecuador

[9]School of Earth Sciences, The University of Melbourne, Australia

Correspondence to: Christoph A. Keller (christoph.a.keller@nasa.gov)

Abstract. Social-distancing to combat the COVID-19 pandemic has led to widespread reductions in air pollutant emissions. Quantifying these changes requires a business-as-usual counterfactual that accounts for the synoptic and seasonal variability of air pollutants. We use a machine learning algorithm driven by information from the NASA GEOS-CF model to assess changes in nitrogen dioxide (NO_2) and ozone (O_3) at 5,756 observation sites in 46 countries from January through June 2020. Reductions in NO_2 coincide with timing and intensity of COVID-19 restrictions, ranging from 60% in severely affected cities (e.g., Wuhan, Milan) to little change (e.g., Rio de Janeiro, Taipei). On average, NO_2 concentrations were 18 (13-23)% lower than business as usual from February 2020 onward. China experienced the earliest and steepest decline, but concentrations since April have mostly recovered and remained within 5% to the business-as-usual estimate. NO_2 reductions in Europe and the US have been more gradual with a halting recovery starting in late March. We estimate that the global NO_x ($NO+NO_2$) emission reduction during the first 6 months of 2020 amounted to 3.1 (2.6-3.6) TgN, equivalent to 5.5 (4.7-6.4)% of the annual anthropogenic total. The response of surface O_3 is complicated by competing influences of non-linear atmospheric chemistry. While surface O_3 increased by up to 50% in some locations, we find the overall net impact on daily average O_3 between February - June 2020 to be small. However, our analysis indicates a flattening of the O_3 diurnal cycle with an increase in night time ozone due to reduced titration and a decrease in daytime ozone, reflecting a reduction in photochemical production.

The O_3 response is dependent on season, time scale, and environment, with declines in surface O_3 forecasted if NO_x emission reductions continue.

1 Introduction

The stay-at-home orders imposed in many countries during the Northern Hemisphere spring of 2020 to slow the spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, hereafter COVID-19), led to a sharp decline in human activities across the globe (Le Quéré et al., 2020). The associated decrease in industrial production, energy consumption, and transportation resulted in a reduction in the emissions of air pollutants, notably nitrogen dioxide (NO_2) (Liu et al., 2020a; Dantas et al., 2020; Petetin et al., 2020; Tobias et al., 2020; Le et al., 2020). Nitrogen oxides (NO_x=NO+NO_2) have a short atmospheric lifetime and are predominantly emitted during the combustion of fossil fuel for industry, transport and domestic activities (Streets et al., 2013, Duncan et al., 2016). Atmospheric NO_2 concentrations thus readily respond to local changes in NO_x emissions (Lamsal et al., 2011). While this may provide both air quality and climate benefits, a quantitative assessment of the magnitude of these impacts is complicated by the natural variability of air pollution due to variations in synoptic conditions (weather), seasonal effects, and long-term emission trends as well as the non-linear responses between emissions and concentrations. Thus, simply comparing the concentration of pollutants during the COVID-19 period to those immediately before or to the same period in previous years is not sufficient to indicate causality. An emerging approach to address this problem is to develop machine-learning based 'weather-normalization' algorithms to establish the relationship between local meteorology and air pollutant surface concentrations (Grange et al., 2018; Grange and Carslaw, 2019; Petetin et al., 2020). By removing the meteorological influence, these studies have tried to better quantify emission changes as a result of a perturbation.

Here we adapt this weather-normalization approach to not only include meteorological information but also compositional information in the form of the concentrations and emissions of chemical constituents. Using a collection of surface observations of NO_2 and ozone (O_3) from across the world from 2018 to present (Section 2.1), we develop a 'bias-correction' methodology for the NASA global atmospheric composition model GEOS-CF (Section 2.2) which corrects the model output at each observational site based on the observations for 2018 and 2019 (Section 2.3). These biases reflect errors in emission estimates, sub-gridscale local influences (representational error), or meteorology and chemistry. Since the GEOS-CF model makes no adjustments to the anthropogenic emissions in 2020, and no 2020 observations are included in the training of the bias corrector, the bias-corrected model (hereafter BCM) predictions for 2020 represent a business-as-usual scenario at each observation site that can be compared against the actual observations. This allows the impact of COVID-19 containment measures on air quality to be explored, taking into account meteorology and the long-range transport of pollutants. We first apply this to the concentration of NO_2 (Section 3.1), and then O_3 (Section 3.2) and explore the differences between the counterfactual prediction and the observed concentrations. In Section 3.3 we explore how the observed changes in the NO_2 concentrations relate to emission of NO_x, and in Section 3.4 we speculate what the COVID-19 restrictions might mean for the second half of 2020.

2 Methods

2.1 Observations

Our analysis builds on the recent development of unprecedented public access to air pollution model output and air quality observations in near real-time. We compile an air quality dataset of hourly surface observations for a total of 5,756 sites (4,778 for NO_2 and 4,463 for O_3) in 46 countries for the time period January 1, 2018 to July 1, 2020, as summarized in Fig. 1 and Table 1. More detailed maps of the spatial distribution of observation sites over China, Europe, and North America are given in Fig A1-3. The vast majority of the observations were obtained from the OpenAQ platform and the air quality data portal of the European Environment Agency (EEA). Both platforms provide harmonized air quality observations in near real-time, greatly facilitating the analysis of otherwise disparate data sources. For the EEA observations, we use the validated data (E1a) for years 2018-2019 and revert to the real-time data (E2a) for 2020. For Japan, we obtained hourly surface observations for a total of 225 sites in Hokkaido, Osaka, and Tokyo from the Atmospheric Environmental Regional Observation System (AEROS) (MOE, 2020). To improve data coverage in under-sampled regions, we further included observations from the cities of Rio de Janeiro (Brazil), Quito (Ecuador), and Melbourne (Australia). All cities offer continuous, hourly observations of NO_2 and O_3 over the full analysis period, thus offering an excellent snapshot of air quality at these locations. We include all sites with at least 365 days of observations between Jan 1, 2018 and December 31, 2019, and an overall data coverage of 75% or more since the first day of availability. Only days with at least 12 hours of valid data are included in the analysis. The final NO_2 and O_3 dataset comprise 8.9×10^7 and 8.2×10^7 hourly observations, respectively.

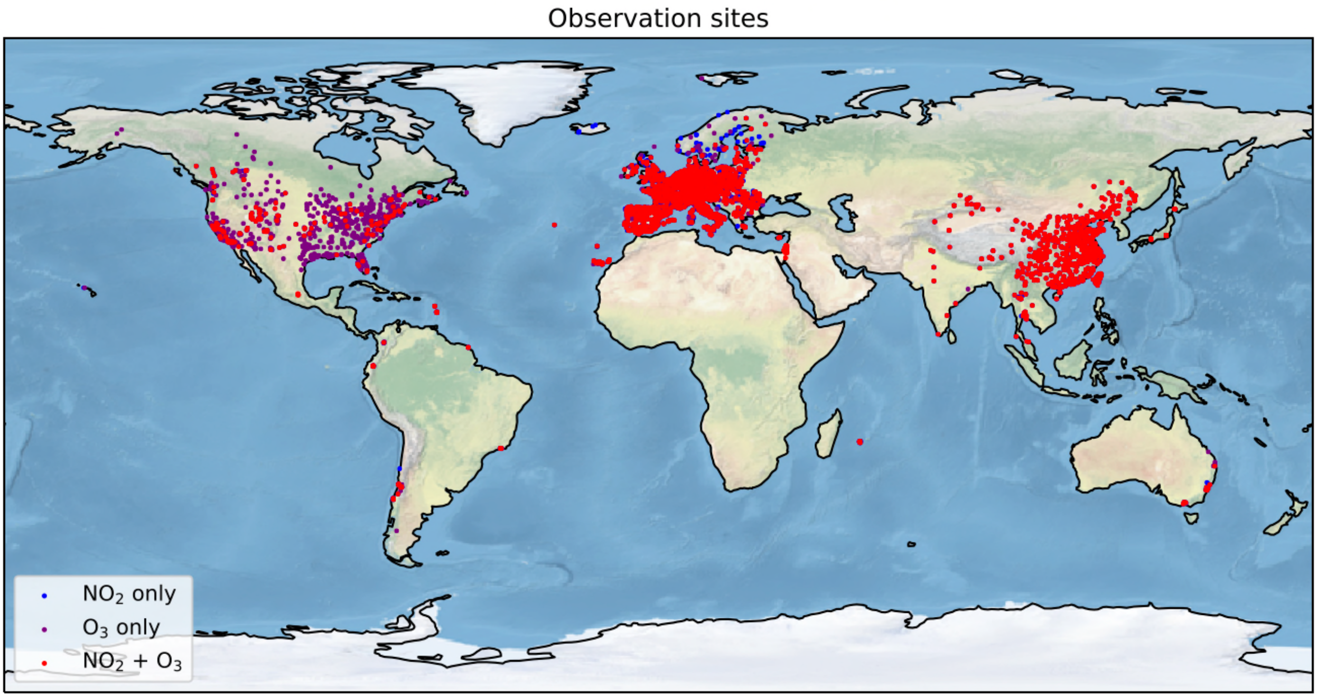


Figure 1: Location of the 5,756 observation sites included in the analysis. Red points indicate sites with both NO_2 and O_3 observations (3,485 in total), purple points show locations with O_3 observations only (978 sites) and blue points show locations with NO_2 observations only (1,293 sites). See Appendix for detailed maps for North America, Europe, and China.

Table 1: Observational data sources used in the analysis. Time period covers Jan 1, 2018 - July 1, 2020.

Name	**Countries**	**Sites**	**Source**
OpenAQ	Australia, China, India, Hong Kong, Taiwan, Thailand, Canada, Chile, Colombia, United States	2410	https://openaq.org/
EEA	Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom,	3101	https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm
AEROS	Japan	225	http://soramame.taiki.go.jp/Index.php
EPA Victoria	Australia (Melbourne)	4	http://sciwebsvc.epa.vic.gov.au/aqapi/Help
Secretaria de Ambiente, Quito	Ecuador (Quito)	8	http://www.quitoambiente.gob.ec/ambiente/index.php/datos-horarios-historicos

Municipal Government of Rio de Janeiro	Brazil (Rio de Janeiro)	8	http://www.data.rio/datasets/dados-hor%C3%A1rios-do-monitoramento-da-qualidade-do-ar-monitorar

2.2 Model

Meteorological and atmospheric chemistry information at each of the air quality observation sites is obtained from the NASA Goddard Earth Observing System Composition Forecast (GEOS-CF) model (Keller et al., 2020). GEOS-CF integrates the GEOS-Chem atmospheric chemistry model (v12-01) into the GEOS Earth System Model (Long et al., 2015; Hu et al., 2018) and provides global hourly analyses of atmospheric composition at 25x25 km^2 spatial resolution, available in near real-time at https://gmao.gsfc.nasa.gov/weather_prediction/GEOS-CF/data_access/ (Knowland et al., 2020). Anthropogenic emissions are prescribed using monthly Hemispheric Transport of Air Pollution (HTAP) bottom-up emissions (Janssens-Maenhout et al., 2015), with imposed weekly and diurnal scale factors as described in Keller et al. (2020). The same anthropogenic base emissions are used for years 2018-2020. Therefore, GEOS-CF does not account for any anthropogenic emission changes since 2018, notably any anthropogenic emission reductions related to COVID-19 restrictions. However, it does capture the variability in natural emissions such as wildfires (based on the Quick Fire Emissions Dataset, QFED) (Darmenov and Da Silva, 2015), or lightning and biogenic emissions (Keller et al., 2014). While the meteorology and stratospheric ozone in GEOS-CF are fully constrained by pre-computed analysis fields produced by other GEOS systems (Lucchesi, 2015; Wargan et al., 2015), no trace-gas observations are directly assimilated into the current version of GEOS-CF. It thus provides a “business as usual” estimate of NO_2 and O_3 that can be used as a baseline for input into the meteorological normalization process.

2.3 Machine learning bias correction

2.3.1 Overall strategy

We use the XGBoost machine learning algorithm (https://xgboost.readthedocs.io/en/latest/#) (Chen and Guestrin, 2016; Frery et al., 2017) to develop a machine learning model to predict the time-varying bias at each observation site at an hourly scale. XGBoost uses the Gradient Boosting framework to build an ensemble of decision trees, trained iteratively on the residual errors to stage-wise improve the model predictions (Friedman, 2001). Based on the 2018-2019 observation-model differences, the machine learning model is trained to predict the systematic (recurring) model bias between hourly observations and the co-located model predictions. These biases can be due to errors in the model, such as emission estimates, sub-gridscale local influences (representational error), or meteorology and chemistry. Since model biases are often site-specific, we train a separate machine learning model for each site.
For each location, we split the 2-year training dataset into 8 quarterly segments (Jan-Mar, Apr-Jun, etc.) and train the model 8 times, each time omitting one of the segments (8-fold cross validation). The default XGBoost model parameters are used, with a learning rate of 0.3, minimum loss reduction of 0, maximum tree depth of 6, and L1 and L2 regularization terms of 0 and 1, respectively. Once trained, the

final model prediction at each location consists of the average prediction of the eight models. To test the final (multi-)model prediction, we omit the center week of each training segment from the 8-fold cross validation and use it for testing only. This approach aims to reduce the auto-correlation signal that can lead to overly optimistic machine-learning results (Kleinert et al., 2020) while still including data from all four seasons in the testing.

The observations used in this analysis are not always quality-controlled, which can cause issues if erroneous observations are included in the training, such as unrealistically high O_3 concentrations of several thousand ppbv. As an ad-hoc solution to this problem, we remove all observations below or above 2 standard deviations from the annual mean from the analysis. Sensitivity tests using more stringent thresholds of 3 or even 4 standard deviations resulted in no significant change in our results.

2.3.2 Evaluation of model predictors

The input variables fed into the XGBoost algorithm are provided in Table A1. The input features encompass 9 meteorological parameters (as simulated by the GEOS-CF model: surface north- and eastward wind components, surface temperature and skin temperature, surface relative humidity, total cloud coverage, total precipitation, surface pressure, and planetary boundary layer height), modelled surface concentrations of 51 chemical species (O_3, NO_x, carbon monoxide, VOCs, and aerosols), and 21 modelled emissions at the given location. In addition, we provide as input features the hour-of-day, day of week, and month of the year; these allow the machine learning model to identify systematic observation-model mismatches related to the diurnal, weekly and seasonal cycle of the pollutants. In addition, for sites with observations available for the full two years, we provide the calendar days since Jan 1, 2018 as an additional input feature to also correct for inter-annual trends in air pollution, e.g., due to a steady decrease in emissions not captured by the model. This follows a similar technique to Ivatt and Evans (2020) and Petetin et al. (2020).

Gradient boosted tree models consist of a tree-like decision structure, which can be analysed to understand how the model uses the input features to make a prediction. Particularly useful in this context is the SHapely Additive exPlanations (SHAP) approach, which is based on game-theoretic Shapely values and represents a measure of each feature's responsibility for a change in the model prediction (Lundberg et al., 2018). SHAP values are computed separately for each individual model prediction, offering detailed insight into the importance of each input feature to this prediction while also considering the role of feature interactions (Lundberg et al., 2020). In addition, combining the local SHAP values offers a representation of the global structure of the machine learning model.
Figure A4 shows the distribution of the SHAP values for all NO_2 predictors separated by polluted sites (left panel) and non-polluted sites (right panel), with polluted sites defined as locations with an annual average NO_2 concentration of more than 15 ppbv. Generally, the model-predicted (unbiased) NO_2 concentration is the most important predictor for the model bias, followed by the hour of the day, the day since Jan 1st 2018 ('Trendday'), and a suite of meteorological variables including wind speed (u10m, v10m), planetary boundary hight (zpbl), and specific humidity (q10m). All of these factors are expected to highly impact NO_2 concentrations and it is thus not surprising that the model biases are most sensitive to them. While there is considerable spread in the feature importance across the

individual sites, there is little overall difference in the feature ranking between polluted vs. non-polluted sites.

Figure A5 shows the SHAP value distribution for all O_3 predictors, again separated into polluted and non-polluted sites (using the same definition as for the NO_2 sites). Unlike for NO_2, the bias-correction models for polluted sites exhibit different feature sensitivities than the non-polluted sites. At polluted locations, the availability of reactive nitrogen (NO_2, NO_y, PAN) is the dominant factor for explaining the model O_3 bias, reflecting the tight chemical coupling between NO_x and O_3 (Seinfeld and Pandis, 2016). This is followed by the month of the year, total precipitation (tprec) and O_3 concentration, again variables expected to be correlated to O_3. At non-polluted sites, the uncorrected O_3 concentration is on average the most relevant input feature for the bias correctors, followed by the month of the year and the odd oxygen concentration (ox). The non-polluted sites are generally more sensitive to wind speed, reflecting the fact that O_3 production and loss at these locations is less dominated by local processes compared to the polluted sites.

2.3.3 Machine learning model skill scores

Figures 2 and 3 summarize the machine learning model statistics for NO_2 and O_3, respectively. The normalized mean bias (NMB), normalized root mean square error (NRMSE), and Pearson correlation coefficient (R) at each site are shown for both the training (blue) and the test (red) dataset. We define NMB as mean bias normalized by average concentration at the given site, and the NRMSE as the root mean square error normalized by the range of the 95-percentile concentration and 5-percentile concentration. Rather than using the mean as the denominator for the NRMSE, we choose the percentile window as a better reference point for the concentration variability at a given site. Using the mean as the denominator for the NRMSE would lead to very similar qualitative results.

For both NO_2 and O_3, the bias-corrected model predictions show no bias when evaluated against the training data, NRMSE's of less than 0.3, and correlation coefficients between 0.6-1.0 (NO_2) and 0.75-1.0 (O_3). Compared to the training data, the skill scores on the test data show a higher variability, with an average NMB of -0.048 for NO_2 and -0.034 for O_3, a NRMSE of 0.25 (NO_2) and 0.18 (O_3), and a correlation of 0.64 (NO_2) and 0.84 (O_3). We find no significant difference in skill scores between background vs. polluted sites or different countries.

A number of factors likely contribute to the poorer statistical results at some of the sites. Importantly, some sites might be prone to overfitting if the training data includes events that are not easily generalizable, such as unusual emission activity (e.g., biomass burning, fireworks, closure of nearby point source, etc.) or weather patterns not frequently observed. Also, the availability of test data at some locations is weak (less than 50%), which can contribute to a poorer skill score.

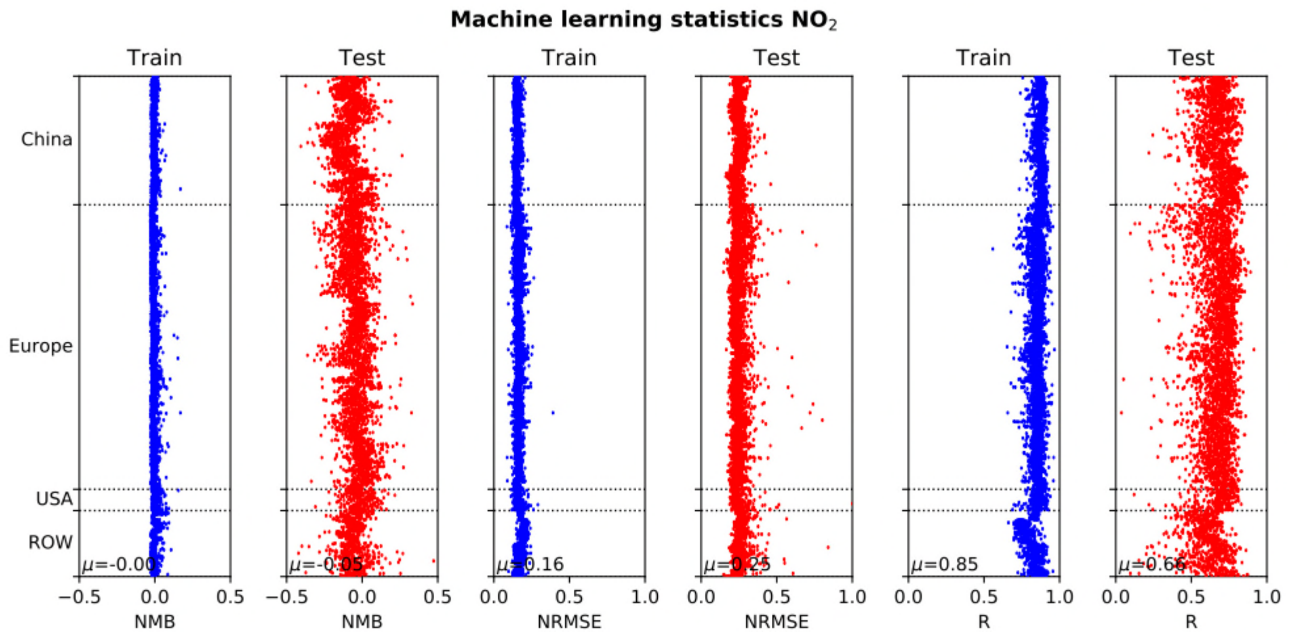


Figure 2: Machine learning statistics between hourly observations and the corresponding bias-corrected model predictions for each observation location. Shown are the normalized mean bias (NMB), normalized root mean square error (NRMSE) and Pearson correlation coefficient (R) for the training data (blue) and the test data (red). Data sorted by region: China, Europe, United States (USA), and rest of the world (ROW). The mean values across all locations are shown in the figure inset.

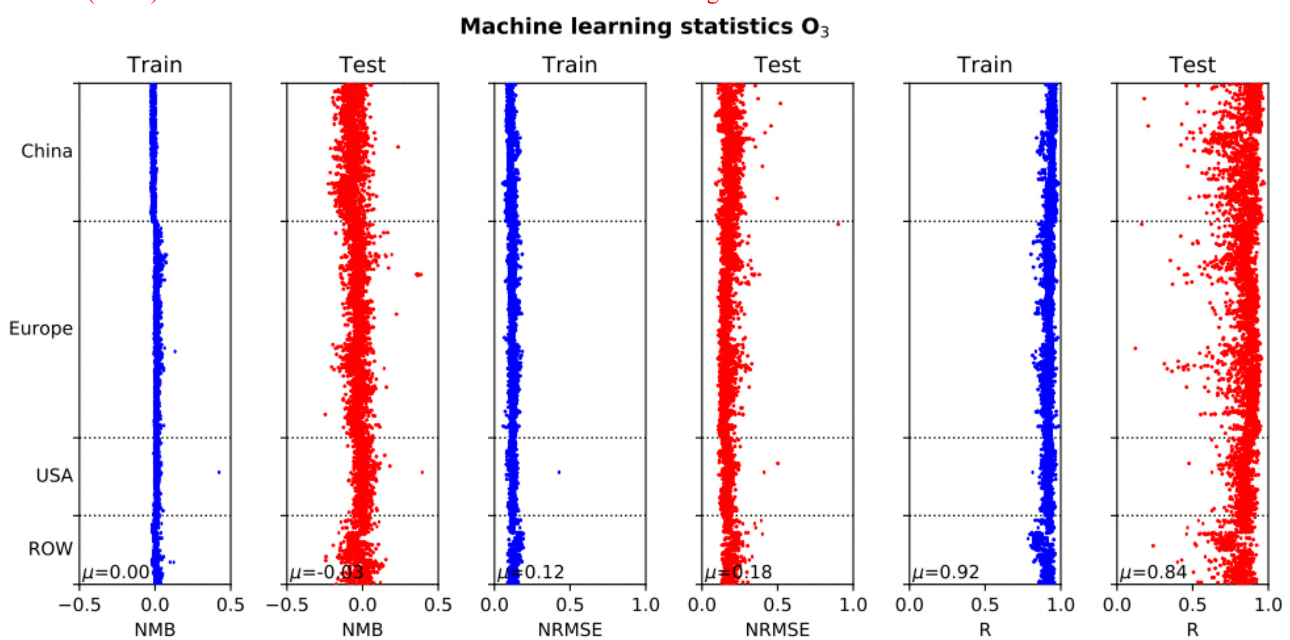


Figure 3: As Figure 2 but for O_3.

2.3.4 Uncertainty estimation

To quantify the uncertainty of an individual model predictions at any given site, we use the standard deviation of the model-observation differences on the test data. For sites with 100% test data coverage, this represents the standard deviation from a sample of 1344 hourly model-observation pairs. The thus obtained individual NO_2 prediction uncertainties range between 4.4 – 25 ppbv (average = 8.7 ppbv) at polluted sites and 0.1 – 15 ppbv at clean sites (average of 5.0 ppbv). On a relative basis, this corresponds to an average uncertainty of 45% at polluted sites and 65% at clean sites. For O_3, we obtain an average individual prediction uncertainty of 14 ppbv (4.9 – 32 ppbv) at polluted sites and 9.1 ppbv (3.0 – 43 ppbv) at clean sites, corresponding to an average relative uncertainty of an individual prediction of 29% and 33% at polluted and clean sites, respectively.
The results presented in this paper are averages aggregated over multiple hours and locations, and the reported uncertainties are adjusted accordingly by calculating the mean uncertainty $\bar{\sigma}$ from the above-described hourly uncertainties σ_i:

$$\bar{\sigma}^2 = \sum_{i=1}^{N} \left(\frac{\sigma_i}{N}\right)^2$$

2.4 Lockdown dates

To support interpretation and guide visualizations, we include approximate national lockdown dates in all figures. The start and end dates for these are from https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdowns (as of July 1, 2020) or based on local knowledge, with the full list of start and end dates given in Table A2. It should be noted that in many countries, lockdown policy varied regionally and many locations enacted ‘soft’ stay-at-home orders before the official lockdowns. Human behavior is therefore expected to have changed considerably in many locations before the official lockdowns went into force.

3 Results

3.1 Nitrogen dioxide

Figure 4 shows the weekly mean observations of NO_2 concentration, the GEOS-CF estimate and the BCM prediction based on the machine-learning predictor trained on 2018-2019 for the five cities of Wuhan (China), Taipei (Taiwan), Milan (Italy), New York (USA) and Rio de Janeiro (Brazil) from January 2018 through June 2020. We choose these five cities for illustration as they represent a diverse level of socio-economic development and due to the cities’ variable responses to the COVID-19 pandemic. These five cities are also illustrative of the varying quality of the uncorrected GEOS-CF predictions compared to the observations. For example, as shown by the dashed grey lines vs. the solid black lines in Fig. 4, the uncorrected model predictions are in good agreement with observations in Rio de Janeiro but underestimate the observed NO_2 concentrations in Taipei and Milan while overestimating concentrations over New York. These differences are a combination of the observation-model scale mismatch (25x25 km^2 vs. point observation) and model errors, such as the simulated spatiotemporal distribution of NO_x emissions or the modelling of the local boundary layer. The model-observation mismatch is particularly pronounced for Wuhan, where the model does not capture the observed seasonal cycle, pointing to errors in the imposed seasonal cycle of NO_x emissions in the model.

In contrast to the uncorrected model predictions, the BCM closely follows the observations for years 2018 and 2019 (dashed black lines in Fig. 4). The grey region in Fig. 4 shows the start and end of the implementation of COVID-19 containment measures. Once containment is implemented, observed concentrations start to diverge from the BCM prediction for Wuhan, Milan and New York (Fig. 4). For Wuhan, we find a reduction in NO_2 of 54 (48-59)% relative to the expected BCM value for February and March 2020, and average decreases of 30-40% are found over Milan (24-43%) and New York (20-34%) starting in mid-March and lasting through April (Fig. 4; Tables A3-A5). For cities where restrictions have been mainly removed (Wuhan, Milan) concentrations rise back towards the BCM value, although in neither city are the concentrations fully restored to what might be expected based on the business-as-usual GEOS-CF simulation.

Looking more broadly at cities around the globe, 53 of the 64 specifically analysed cities feature NO_2 reductions of between 20-50% (Fig. A6-A8 and Tables A3-A5). Most locations issued social distancing recommendations prior to the legal lockdowns and observed NO_2 declines often precede the official lockdown date by 7-14 days (e.g., Brussels, London, Boston, Phoenix, and Washington, DC).

For Taipei and Rio de Janeiro, the observations and the BCM show little difference (Fig. 4), consistent with the less stringent quarantine measures in these places. Other cities with only short-term NO_2 reductions of less than 25% include Atlanta (USA), Prague (Czech Republic), and Melbourne (Australia), again fitting with the comparatively relaxed containment measures in these places (Fig. A6-A8). In contrast, Tokyo (Japan) and Stockholm (Sweden), which also implemented a less aggressive COVID-19 response, exhibit NO_2 reductions comparable to those of cities with official lockdowns (>20%), suggesting that economic and human activities were similarly subdued in those cities.

Substantial differences exist between cities in South America, with Rio de Janeiro and Santiago de Chile showing little change thus far in 2020, whereas Quito (Ecuador) and Medellin (Colombia) experienced a greater than 50% reduction in NO_2 after the initiation of strict restrictions measures in mid-March (Fig. A8 and Table A5). Concentrations in Medellin rebounded sharply in April and May, while concentrations in Quito remained 55 (52-58)% below business as usual throughout May and only started to return back to normal in June.

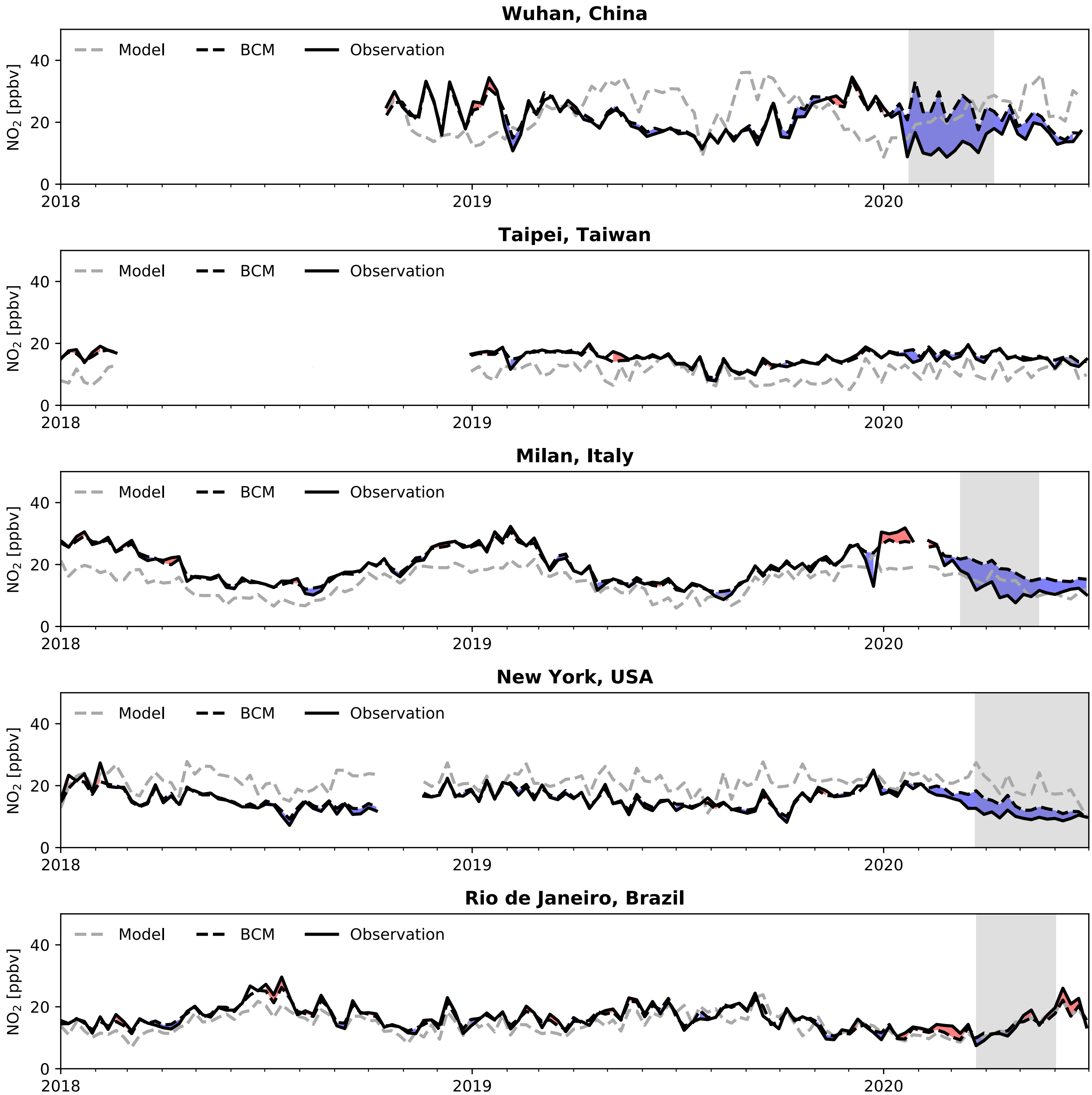


Figure 4. Comparison of NO_2 surface concentrations (ppbv = nmol mol-1) for Wuhan, Taipei, Milan, New York, and Rio de Janeiro for January 2018 through June 2020. Observed values are shown in solid black, the original GEOS-CF model simulation is shown in dashed grey, and the BCM predictions are in dashed black. The area between observations and BCM predictions is shaded blue (red) if observations are lower (higher) than BCM predictions. Grey areas represent the period of lockdown. Shown are the 7-day average mean values for the 9, 18, 19, 14 and 2 observational sites in Wuhan, Taipei, Milan, New York, and Rio de Janeiro, respectively. Observations for China are only available starting in mid-September 2018.

To evaluate the large-scale impact of COVID-19 restrictions on air quality, we aggregate the individual observation-model comparisons by country. We note that our estimates for some countries (e.g., Brazil, Colombia) are based on a single city and likely not representative of the whole country. On a country level, we find the sharpest and earliest drop in NO_2 over China, where observed concentrations fell, on average, 55 (51-59)% below their expected value in early February when restrictions were implemented (Fig. 5). Concentrations remained at this level until late February, at which point they started to increase until restrictions were significantly relaxed in early April. Our analysis suggests that Chinese NO_2 concentrations have recovered to within 5 (1-9)% of the business as usual since then. For 2019 (dashed line in Fig. 5) the BCM shows a reduction in NO_2 concentrations around Chinese New Year (5th February 2019), and it is likely that some reduction around the equivalent 2020 period (25th January 2020) would have occurred anyway. However, the 2020 reductions are significantly larger and more prolonged than in 2019. Similar to China, India shows large reductions in NO_2 concentration (58 (49-67)%) coinciding with the implementation of restrictions in mid-March (Fig. 5); however, NO_2 concentrations have not yet recovered by the end of June, reflecting the prolonged duration of lockdown measures. Other areas of Asia, such as Hong Kong and Taipei, implemented smaller restrictions than China or India and they show significantly smaller decreases (less than 20%).

For Europe and the United States, we find widespread NO_2 reductions averaging 22 (19-25)% in March and 33 (30-36)% in April (Fig. 5). In some countries, recovery is evident as lockdown restrictions are removed or lessened (e.g., Greece, Romania) but in 29 out of 36 countries, concentrations remain 20% or more below the business-as-usual scenario throughout May and June.

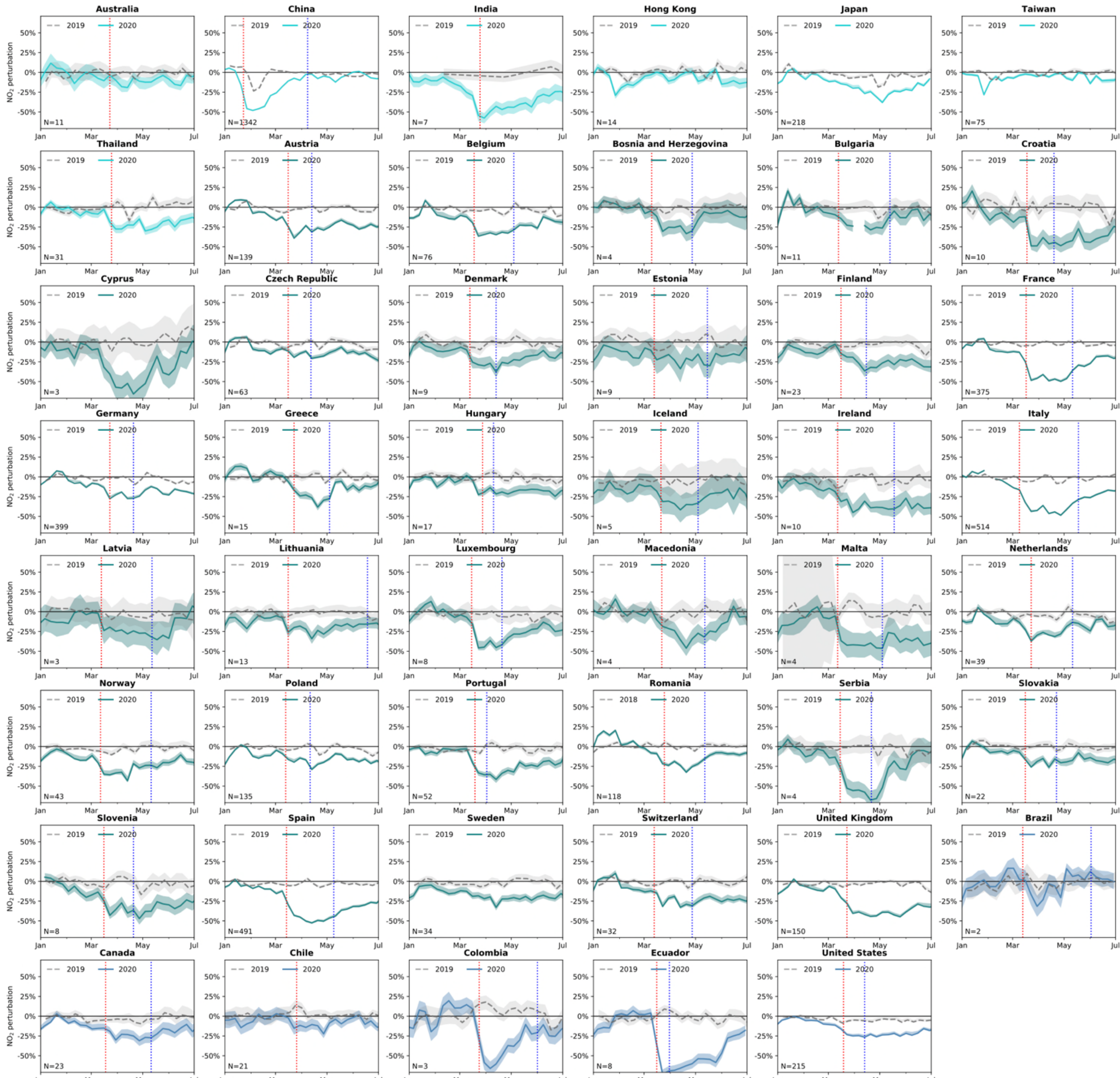


Figure 5: Seven-day average fractional difference between observed NO_2 and the BCM predictions for 46 countries between January 1 through June 30, aggregated from all sites across each country (number of sites in the bottom left of each panel). The thick line indicates the mean across all sites for the first half of 2020, with the shaded area representing the uncertainty estimate. Differing colours indicate differing regions (cyan: Asia & Australia; green: Europe; blue: Americas). The grey dashed line indicates the equivalent average for the same six month period in 2019. The red dashed vertical line indicates COVID-19 restriction dates, and the blue line indicates the beginning of easing measures.

3.2 Ozone

We follow the same methods for developing a business-as-usual counterfactual for O_3 as we did for NO_2 in section 3.1. Any change in local O_3 concentration arising from COVID-19 restrictions is set against a large seasonal increase in (background) concentrations in the Northern Hemisphere springtime (Fig. 6). Due to the longer atmospheric lifetime of O_3 compared to NO_2, the local O_3 signal is expected to be comparatively small. This makes attributing changes in O_3 concentration more challenging than for NO_2. Our analysis shows an O_3 increase of up to 50% for some periods in cities with large NO_2 reductions (e.g., Wuhan, Milan, Quito; Fig. 3 and Fig. A9-A11), but there is much less convincing evidence for a systematic O_3 response across cities or on a regional level (Fig. 7). For example, our analysis shows little O_3 difference in Beijing and Madrid during lockdown despite NO_2 declines comparable to Wuhan or Milan (Fig. A9-A11). O_3 enhancements of up to 20% are found over Europe (e.g., Belgium, Luxembourg, Serbia), with a peak in early April, approximately two weeks after lockdown started (Fig. 7).

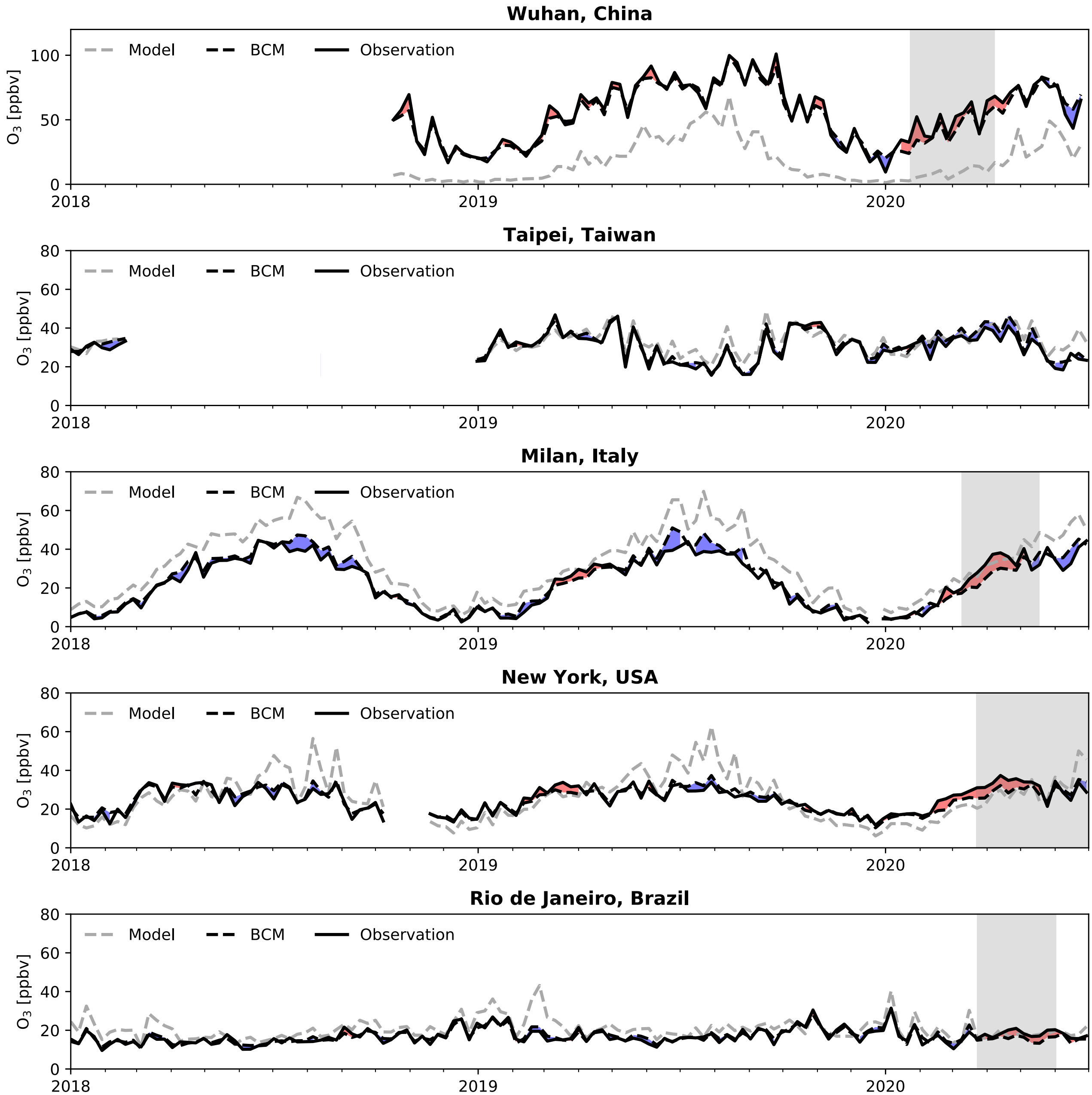


Figure 6: Comparison of O_3 surface concentrations for Wuhan, Taipei, Milan, New York, and Rio de Janeiro for January 2018 through June 2020. Observed values are shown in solid black, the original GEOS-CF model simulation is shown in dashed grey, and the BCM predictions are in dashed black. The area between observations and BCM predictions is shaded blue (red) if observations are lower (higher) than BCM predictions. The grey areas represent the period of lockdown. Shown are the 7-day average mean values for the 9, 18, 19, 14 and 4 observational sites in Wuhan, Taipei, Milan, New York, and Rio de Janeiro, respectively. Observations for China are only available starting in mid-September 2018.

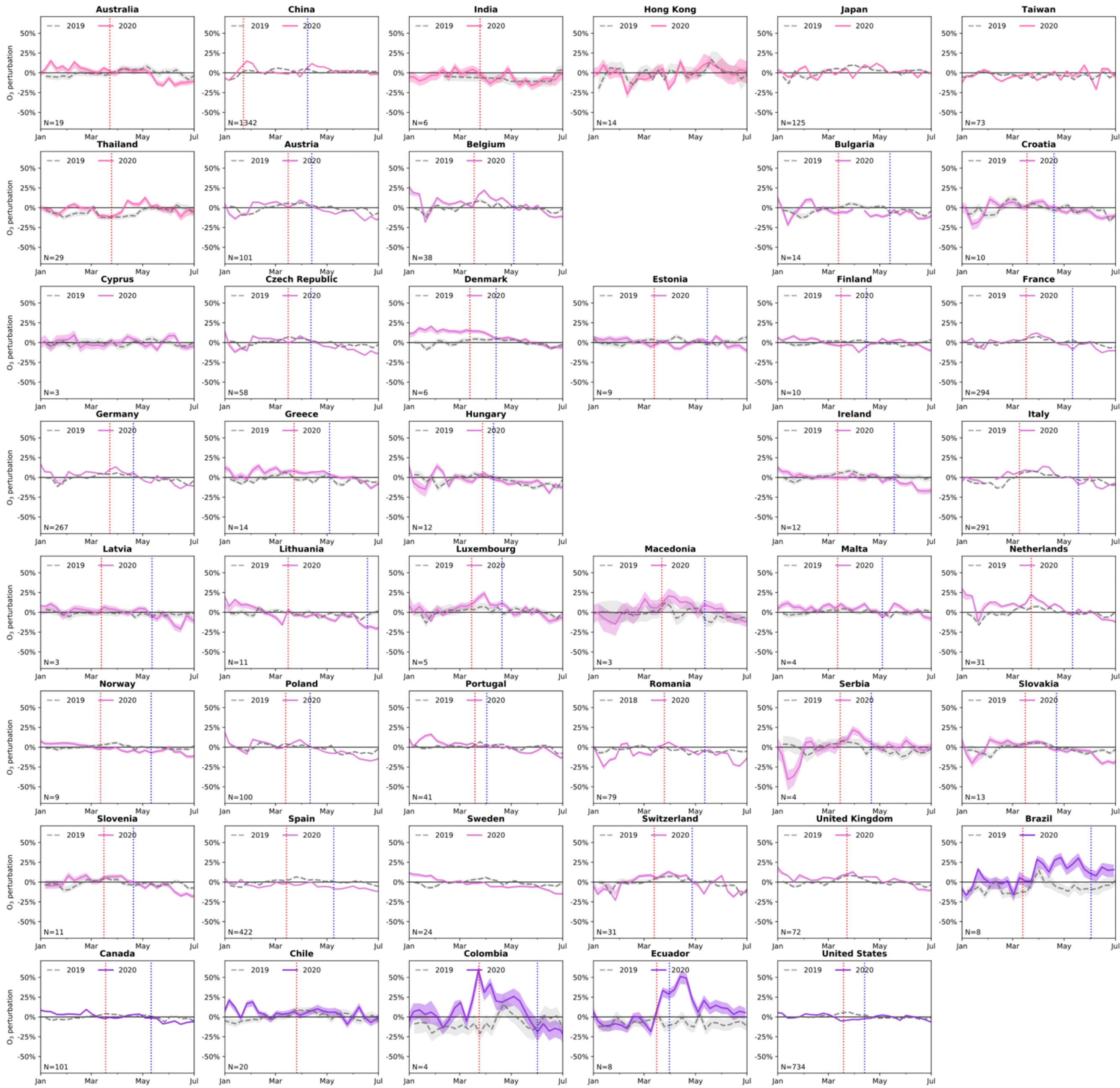


Figure 7: Similar to Figure 5 but for O_3 and without Bosnia and Herzegovina and Iceland. Differing colors indicate differing regions (pink: Asia & Australia; light purple: Europe; dark purple: Americas).

The analysis of O_3 is complicated by its nonlinear chemical response to NO_x emissions. In the presence of sunlight, O_3 is produced chemically from the oxidation of volatile organic compounds in the presence of NO_x (Seinfeld and Pandis, 2016). Therefore, a decline in NO_x emissions could decrease O_3 production and thus suppress O_3 concentrations. On the other hand, the process of NO_x titration, in which freshly emitted NO rapidly reacts with O_3 to form NO_2, acts as a sink for O_3 (Seinfeld and Pandis, 2016). Odd oxygen (O_x=NO_2+O_3) is conserved when O_3 reacts with NO and thus offers a tool for separating these competing processes. Figure 8 presents the global mean diurnal cycle for O_3 and O_x for the 5-month period since February 1, 2020 for both the observations and the BCM model, based on the individual hourly predictions at each observation site aggregated by local hour. The analysis of O_3 and O_x is based on the same set of observation sites where both NO_2 and O_3 observation are available (see Fig. 1). Compared to the BCM model, there has been an increase in the concentration of night time O_3 (midnight-5.00 local time, Fig. 8a) by 1 part per billion by volume (ppbv = nmol mol^{-1}) compared to the BCM, whereas O_x shows a decrease of 1 ppbv (Fig. 8b). While these changes are small in magnitude, they represent a multi-month aggregate over 3,485 observation sites that are statistically significant at the 1% confidence interval. It should be noted that the biases of the machine learning models show little diurnal variability (Fig. A12-13), suggesting that this result is not caused by poor model performance during specific times of the day.

Our results indicate that during the night, reduced NO emissions led to a reduction in O_3 titration, allowing O_3 concentrations to increase. During the afternoon, we find that O_3 concentrations are lower by 1 ppbv (Fig. 8a), while observed O_x concentrations are lower than the baseline model by almost 2 ppbv at 14:00 local time (Fig. 8b). We attribute the lower O_x to reduced net O_x production due to the lower NO_x concentration, but as titration is also reduced, daytime O_3 concentrations are little changed. Overall changes to mean O_3 concentrations are small, but there is a flattening of the diurnal cycle. As shown in the lower panels in Fig. 8, both factors - enhanced night time O_3 and reduced daytime O_x - are more pronounced at locations where pre-existing NO_2 concentrations are high (> 15 ppbv). This suggests that the observed O_3 deviations from the BCM are indeed coupled to NO_x reductions due to COVID-19 restrictions, given that those are most pronounced at polluted sites.

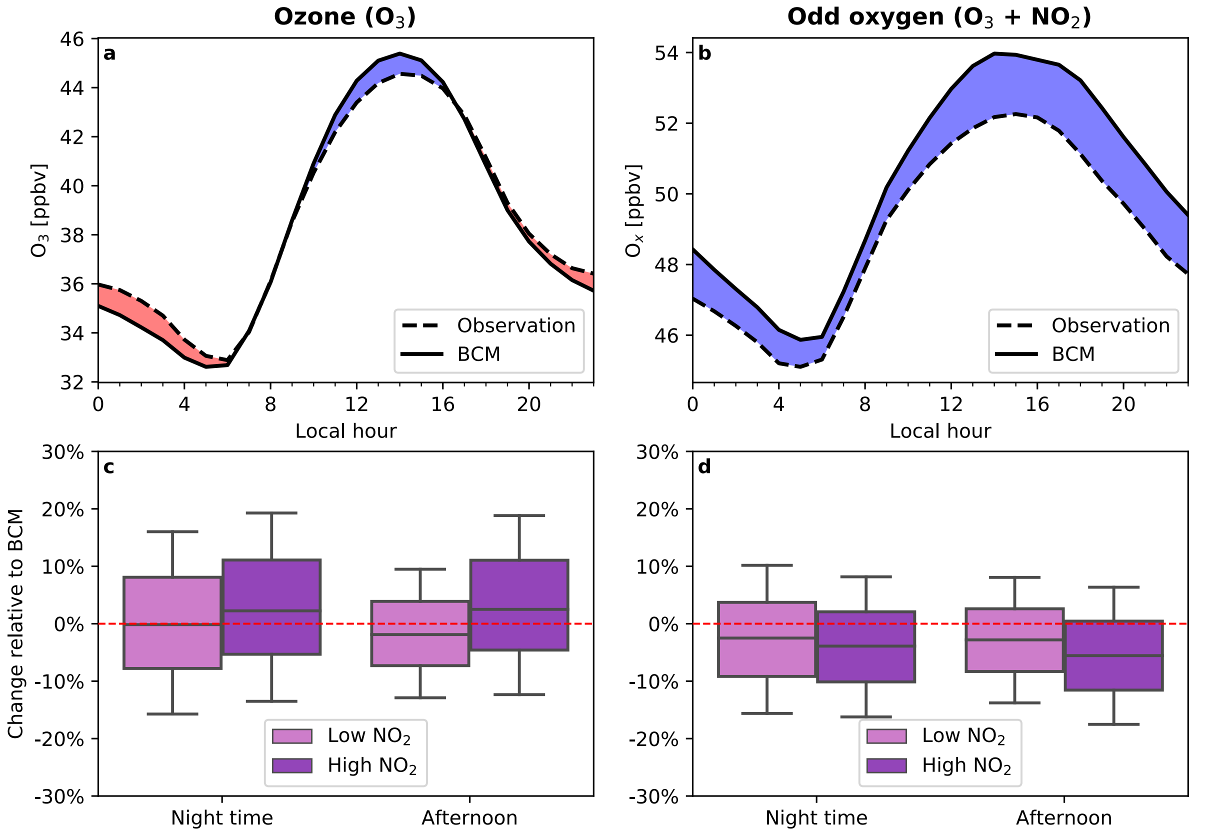


Figure 8: Observed and BCM modelled diurnal cycle of O_3 (a) and O_x (b) averaged across all surface observation sites between February 1, 2020 through June 30, 2020 with estimated corresponding changes in surface O_3 (c) and O_x (d) relative to the BCM. Barplots (c and d) show observed changes during night time (0-5 local time) and the afternoon (12-17 local time) for locations with low (< 15 ppbv) and high (> 15 ppbv) NO_2 concentrations (based on the 2019 average).

3.3 NO_x emission reductions

The NO_2 analysis presented in Section 3.1 implies a stark reduction in NO_x emissions. However, due to the impact of atmospheric chemistry, changes in NO_2 concentrations do not reflect the same relative change in NO_x emissions. Because of this, the NO_2/NO_x ratio and the NO_x lifetime, both of which depend on seasonality and the local chemical environment, need to be taken into account when inferring NO_x emissions from NO_2 concentrations (Lamsal et al., 2011; Shah et al., 2020). To estimate the relationship between changes in NO_x emission and changes in NO_2 concentrations, we conducted a sensitivity simulation for the time period December 1, 2019 to June 8, 2020 using the GEOS-CF model with perturbed anthropogenic emissions. The perturbation simulation uses anthropogenic NO_x emissions scaled based on adjustment factors derived from NO_2 tropospheric columns observed by the NASA OMI instrument (Boersma et al., 2011). Daily scale factors were computed by normalizing coarse-resolution (2x2.5 degrees), 14-day NO_2 tropospheric column moving averages by the corresponding moving average for year 2018 (the emissions base year in GEOS-CF; section 2.2). Forest fire signals were filtered out based on QFED emissions and no scaling was applied over water. This results in anthropogenic emission adjustment factors of 0.3 to 1.4 (Fig. A14), comparable to the magnitude obtained from the observation-BCM comparisons at cities globally (Fig. 5) and capturing the range of expected NO_x emission changes. However, it should be noted that the scale factors do not necessarily coincide in space and time with the ones derived from observations and the BCM, and they do not include any adjustment for the NO_2/NO_x ratio.

Figure 9a shows the response of NO_2 surface concentration to a change in NO_x emissions, derived from the comparison of the sensitivity experiment against the GEOS-CF reference simulation. Our results indicate that NO_2 concentrations drop, on average, by 80% of the fractional decrease in anthropogenic NO_x emission, with a further diminishing effect for emission reductions greater than 50%. This reflects both the buffering effect of atmospheric chemistry and the presence of natural background NO_2. The here derived average sensitivity of 0.8 between a change in surface NO_2 to a change in NO_x emissions is comparable to the value of 0.86 (1/1.16) obtained by Lamsal et al. (2011) for the relationship between NO_x emissions and tropospheric column NO_2 observations.

To infer the reduction in anthropogenic NO_x emissions due to COVID-19 containment measures during the first six months of 2020, we use the best linear fit between the simulated NO_x/NO_2 sensitivity (dashed purple line in Fig. 9a). To do so, we calculate the monthly percentage emission change at each observation site based on the NO_2 anomalies derived in Section 3.1 and the corresponding best fit NO_x/NO_2 sensitivity (Fig. 9a). This is a simplification as the local NO_x/NO_2 sensitivity ratio is highly dependent on the local environment. To account for this uncertainty, we assign an absolute error of 15% to our NO_x/NO_2 sensitivity, as derived from the spread in the NO_x/NO_2 ratio in the sensitivity simulation (Fig. 9a). We then aggregate these estimates to a country-level by weighting them based on average NO_2 concentrations per location, thus giving higher weight to locations with more nearby NO_x emission sources. It should be noted that for some countries, our estimates are based upon a small number of observation sites that might not be representative for the country as a whole. This is particularly true for India and Brazil, where less than 10 observation sites are available. While the smaller observation sample size is reflected in the wider uncertainty associated with these emission estimates compared to countries with a much denser monitoring network (e.g., China or Europe), the applied extrapolation method might incur errors that are not reflected in the stated uncertainty ranges. To obtain absolute estimates in emission changes, the monthly country-level percentage emission changes are convoluted with bottom-up emissions estimates for 2015 from the Emission Database for Global Atmospheric Research (EDGAR v5.0_AP, Crippa et al., 2018, 2020). The choice of EDGAR v5.0 as the bottom-up reference inventory (over e.g., the HTAP emissions inventory used in GEOS-CF) was motivated by the fact that its baseline has been updated more recently and the country emission totals - which our analysis is based on - are readily available.

As summarized in Table 2, we calculate that the total reduction in anthropogenic NO_x emissions due to COVID-19 containment measures during the first six months of 2020 amounted to 3.1 (2.6-3.6) TgN (Fig. 9b and Table 2). This is equivalent to 5.5 (4.7-6.4)% of global annual anthropogenic NO_x emissions (Table 2). Our estimate encompasses 46 countries that together account for 67% of the total emissions (excluding international shipping and aviation). We have no information for significant countries such as Russia, Indonesia, or anywhere in Africa due to the lack of publicly available near real-time air quality information. China accounts for the largest fraction of the total deduced emission reductions (28%), followed by India (25%), the United States (18%), and Europe (12%). While our method does not allow for sector-specific emission attribution, we assume our results to be most representative for changes in traffic emissions (rather than, say, aircraft emissions) given the location of the observation sites. On average, traffic emissions represent 27% of total anthropogenic NO_x emissions (Crippa et al., 2018), and our derived total NO_x emission reduction from Jan-Jun 2020 corresponds to 21 (17-24)% of global annual traffic emissions. The share of transportation on total NO_x emissions is higher in the US and Europe (approx.. 40%) compared to India and China (20-25%). Taking this into account, the derived ratio of NO_x emission reductions to annual traffic emissions is 21 (16-26)% in the US, 25 (20-30)% in Europe, 39 (34-44)% in China, and 62 (55-69)% in India.

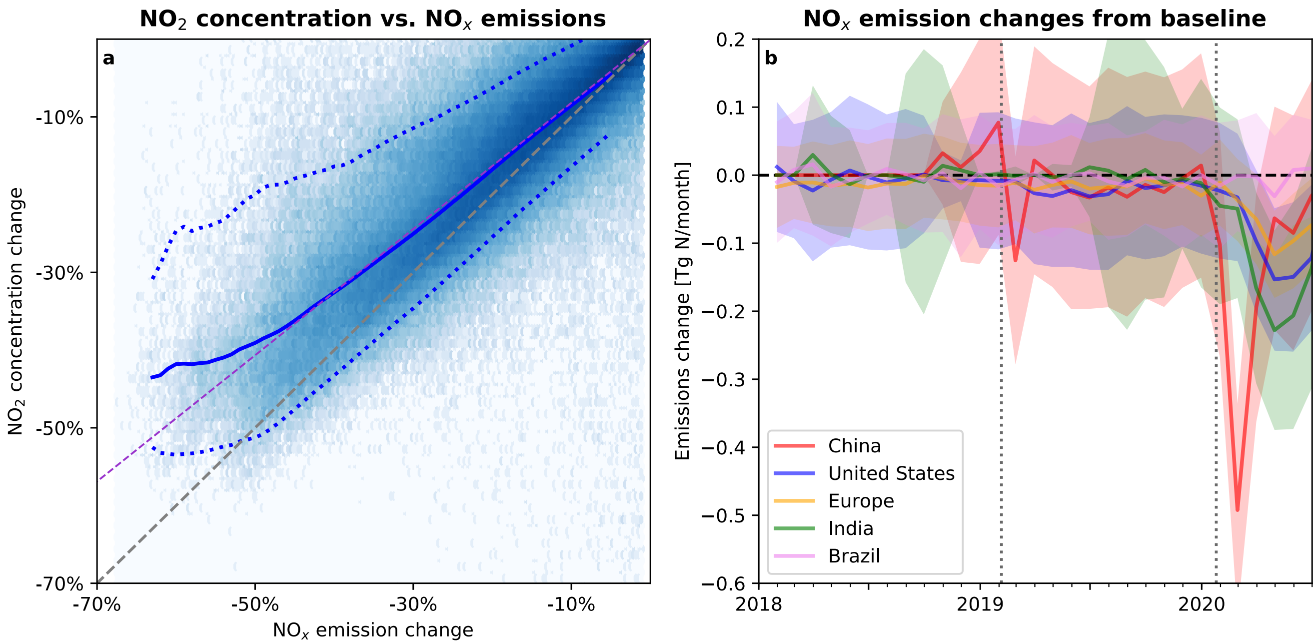


Figure 9: a) Response of NO_2 surface concentration (y-axis) to a change in NO_x emissions (x-axis), as deduced from a model sensitivity simulation (see methods). The solid blue line shows the mean value across all individual grid cells (blue squares) and the dotted blue lines show the 5% and 95% quantiles. The dashed purple line shows the best linear fit. b) Estimated monthly change in NO_x emissions from the baseline since 2018 for China (red), United States (blue), Europe (yellow), India (green), and Brazil (purple), as estimated from observed NO_2 concentration anomalies. Shaded areas indicate estimated emission uncertainties. Dotted grey lines indicate Chinese New Year 2019 and 2020.

Table 2: Anthropogenic NO_x emission reductions in GgN month^{-1} as derived from NO_2 concentration changes.

	Baseline[1]	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Australia	621	-2.9 (-13.8-8.0)	-3.7 (-14.4-6.9)	-8.5 (-18.5-1.6)	-6.1 (-16.1-3.8)	-7.9 (-17.6-1.8)
Austria	73	-0.5 (-1.4-0.5)	-1.7 (-2.7--0.7)	-2.0 (-3.0--1.0)	-1.6 (-2.6--0.6)	-1.8 (-2.8--0.8)
Belgium	98	-1.0 (-2.4-0.3)	-2.0 (-3.4--0.7)	-3.2 (-4.5--1.9)	-2.4 (-3.8--1.1)	-1.8 (-3.2--0.5)
Bosnia and Herzegovina	32	0.05 (-0.47-0.57)	-0.43 (-0.96-0.11)	-0.90 (-1.45--0.35)	-0.39 (-0.99-0.21)	-0.28 (-0.92-0.35)
Brazil	1844	-1.3 (-35.7-33.2)	-1.5 (-37.0-34.0)	-32.0 (-67.2-3.2)	7.2 (-26.1-40.6)	10.3 (-21.7-42.3)
Bulgaria	46	-0.12 (-0.83-0.58)	-0.60 (-1.32-0.12)	-1.20 (-1.93--0.46)	-0.67 (-1.44-0.11)	-0.41 (-1.20-0.37)
Canada	755	-6.3 (-17.1-4.6)	-12.2 (-23.3--1.1)	-19.8 (-31.4--8.1)	-18.5 (-30.5--6.5)	-11.4 (-23.8-1.0)
Chile	202	-0.5 (-3.9-2.9)	-0.7 (-4.0-2.6)	-2.8 (-6.0-0.4)	-1.6 (-4.7-1.4)	-1.0 (-4.1-2.1)

China	11876	-517 (-669--366)	-191 (-342--39)	-63 (-215-89)	-82 (-235-70)	-30 (-182-123)
Colombia	207	1.2 (-2.5-4.9)	-0.2 (-3.8-3.4)	-12.0 (-15.5--8.5)	-5.5 (-9.1--1.9)	-4.4 (-8.0--0.7)
Croatia	24	-0.25 (-0.64-0.14)	-0.55 (-0.95--0.15)	-1.03 (-1.44--0.63)	-0.96 (-1.37--0.55)	-0.90 (-1.31--0.48)
Czech Republic	108	-1.0 (-2.5-0.4)	-1.3 (-2.8-0.2)	-1.8 (-3.2--0.3)	-1.3 (-2.8-0.2)	-1.6 (-3.1--0.1)
Denmark	48	-0.5 (-1.3-0.3)	-0.8 (-1.6--0.1)	-1.4 (-2.1--0.6)	-1.0 (-1.8--0.2)	-0.8 (-1.5--0.0)
Ecuador	133	0.5 (-1.6-2.6)	-3.8 (-5.9--1.8)	-8.9 (-11.0--6.9)	-7.5 (-9.6--5.4)	-4.0 (-6.1--2.0)
Estonia	13	-0.20 (-0.44-0.05)	-0.16 (-0.41-0.10)	-0.28 (-0.54--0.02)	-0.29 (-0.54--0.03)	-0.20 (-0.45-0.04)
Finland	77	-1.1 (-2.3-0.1)	-0.8 (-2.0-0.4)	-2.3 (-3.6--1.1)	-2.0 (-3.3--0.8)	-2.0 (-3.3--0.8)
France	337	-3.2 (-7.6-1.2)	-9.1 (-13.5--4.7)	-15.7 (-20.1--11.3)	-12.7 (-17.1--8.2)	-6.9 (-11.3--2.4)
Germany	494	-3.0 (-9.4-3.4)	-7.1 (-13.5--0.7)	-11.5 (-17.9--5.1)	-8.3 (-14.7--1.9)	-9.2 (-15.7--2.8)
Greece	101	0.1 (-1.4-1.5)	-0.5 (-1.9-1.0)	-2.9 (-4.4--1.5)	-1.5 (-3.0--0.0)	-1.3 (-2.8-0.1)
Hong Kong	90	-1.5 (-2.8--0.2)	-0.2 (-1.6-1.1)	-0.4 (-1.7-1.0)	-0.3 (-1.6-1.0)	-1.2 (-2.6-0.2)
Hungary	55	-0.3 (-1.1-0.5)	-0.4 (-1.2-0.4)	-1.0 (-1.8--0.2)	-1.0 (-1.8--0.1)	-1.0 (-1.9--0.2)
Iceland	2	-0.04 (-0.08-0.01)	-0.04 (-0.09-0.01)	-0.09 (-0.14--0.04)	-0.07 (-0.12--0.01)	-0.04 (-0.10-0.01)
India	4693	-52 (-125-21)	-161 (-234--88)	-232 (-307--157)	-202 (-280--125)	-140 (-220--59)
Ireland	35	-0.3 (-1.0-0.3)	-0.8 (-1.4--0.2)	-1.3 (-1.9--0.7)	-1.4 (-2.0--0.8)	-1.2 (-1.8--0.6)
Italy	357	-1.9 (-6.5-2.7)	-9.7 (-14.4--5.1)	-15.6 (-20.2--10.9)	-12.4 (-17.1--7.8)	-7.7 (-12.3--3.0)
Japan	996	-4.1 (-17.2-9.0)	-12.6 (-25.7-0.6)	-23.4 (-36.7--10.2)	-28.7 (-41.9--15.4)	-18.0 (-31.3--4.7)
Latvia	14	-0.08 (-0.38-0.22)	-0.16 (-0.45-0.12)	-0.37 (-0.67--0.06)	-0.44 (-0.74--0.13)	-0.18 (-0.48-0.12)
Luxembourg	12	-0.17 (-0.48-0.15)	-0.27 (-0.58-0.05)	-0.50 (-0.82--0.18)	-0.40 (-0.72--0.07)	-0.32 (-0.64-0.01)
Lithuania	20	0.01 (-0.18-0.19)	-0.23 (-0.42--0.05)	-0.47 (-0.65--0.29)	-0.31 (-0.49--0.13)	-0.23 (-0.42--0.05)
Macedonia	10	-0.00 (-0.17-0.16)	-0.08 (-0.25-0.09)	-0.33 (-0.50--0.16)	-0.28 (-0.46--0.10)	-0.06 (-0.24-0.12)
Malta	3	-0.00 (-0.06-0.06)	-0.07 (-0.13--0.01)	-0.13 (-0.20--0.07)	-0.11 (-0.18--0.05)	-0.11 (-0.18--0.05)
Netherlands	121	-1.4 (-3.1-0.4)	-2.6 (-4.3--0.8)	-3.4 (-5.1--1.7)	-2.1 (-3.9--0.4)	-2.0 (-3.8--0.2)
Norway	63	-0.8 (-1.7-0.0)	-1.5 (-2.4--0.6)	-1.9 (-2.8--1.0)	-1.5 (-2.5--0.6)	-1.1 (-2.0--0.2)
Poland	284	-3.3 (-7.1-0.5)	-3.2 (-7.0-0.5)	-5.9 (-9.7--2.1)	-4.0 (-7.8--0.2)	-5.1 (-9.0--1.3)
Portugal	70	-0.4 (-1.4-0.6)	-1.2 (-2.2--0.2)	-2.6 (-3.6--1.5)	-1.9 (-2.9--0.9)	-1.6 (-2.6--0.5)
Romania	102	0.5 (-0.9-1.8)	-1.1 (-2.5-0.3)	-2.5 (-3.9--1.2)	-1.6 (-3.0--0.2)	-1.0 (-2.4-0.4)
Serbia	63	-0.48 (-1.54-0.59)	-1.77 (-2.87--0.68)	-3.83 (-4.94--2.72)	-2.24 (-3.44--1.04)	-0.82 (-2.07-0.43)
Slovakia	33	-0.20 (-0.67-0.28)	-0.43 (-0.90-0.05)	-0.61 (-1.09--0.13)	-0.54 (-1.04--0.05)	-0.57 (-1.07--0.07)
Spain	333	-2.9 (-7.2-1.5)	-8.6 (-13.0--4.2)	-17.0 (-21.3--12.6)	-13.9 (-18.3--9.5)	-10.0 (-14.5--5.6)
Sweden	85	-1.0 (-2.2-0.2)	-1.3 (-2.5--0.0)	-2.0 (-3.3--0.8)	-1.9 (-3.1--0.6)	-1.6 (-2.9--0.4)
Switzerland	36	-0.25 (-0.77-0.26)	-0.65 (-1.16--0.14)	-0.94 (-1.45--0.43)	-0.83 (-1.36--0.30)	-0.84 (-1.37--0.31)
Taiwan	371	-3.7 (-8.6-1.2)	-1.5 (-6.4-3.5)	-1.3 (-6.2-3.7)	-1.7 (-6.7-3.4)	-3.9 (-8.9-1.2)
Thailand	458	-2.6 (-9.2-4.0)	-4.8 (-11.7-2.0)	-10.7 (-17.6--3.8)	-11.6 (-18.6--4.7)	-8.5 (-15.6--1.4)
United Kingdom	390	-3.0 (-8.3-2.2)	-6.8 (-12.0--1.6)	-16.4 (-21.6--11.2)	-16.4 (-21.6--11.1)	-12.8 (-18.0--7.5)
United States	6243	-34 (-116-48)	-94 (-177--11)	-155 (-239--72)	-147 (-231--64)	-123 (-207--40)
Other countries[2]	1307	n/a	n/a	n/a	n/a	n/a
Shipping and Aviation	671	n/a	n/a	n/a	n/a	n/a
Total	**4681**	**-651 (-843--460)**	**-553 (-745--360)**	**-692 (-885--498)**	**-603 (-798--408)**	**-418 (-615--222)**

[1] EDGAR v5.0_AP 2015 annual emissions expressed as GgN month^{-1} (Crippa et al., 2020)
[2] Primarily Indonesia, Iran, Mexico, Pakistan, Russia, Saudi Arabia, South Africa, South Korea, Vietnam

3.4 Long-term impact of reduced NO_x emissions on surface O_3

The response of O_3 to NO_2 declines in the wake of the COVID-19 outbreak is complicated by the competing influences of atmospheric chemistry. From February through June 2020, the diurnal observation-BCM comparisons suggest that the reduction in photochemical production was offset by a smaller loss from titration, as described in Section 3.2. This resulted in a flattening of the diurnal cycle and an insignificant net change in surface O_3 over a diurnal cycle. The competing impacts of reduced NO_x emissions on O_3 production and loss are dependent on the local chemical and meteorological environment. This is reflected in the variable geographical response of O_3 following the implementation of COVID-19 restrictions (Le et al., 2020; Dantas et al., 2020). Moreover, as atmospheric reactivity increases through the Northern Hemisphere spring and summer, the relative importance of photochemical production is expected to increase in the Northern Hemisphere.

To assess the potential seasonal-scale impact of reduced anthropogenic emissions on O_3, we conducted two free-running forecast simulations between June 8 through August 31, 2020, initialized from the GEOS-CF simulation and the sensitivity simulation described in Section 3.3, respectively. Both simulations use the same biomass burning emissions based on a historical QFED climatology. For the forecast sensitivity experiment, we assume a sustained, time-invariant 20% reduction in global anthropogenic emissions of NO_x, carbon monoxide (CO), and VOCs. We chose to alter not only the anthropogenic emissions of NO_x but also of other pollutants whose anthropogenic emissions are highly correlated to NO_x, as a reduction in NO_x emissions without corresponding declines in CO and VOC emissions seems unrealistic.

Figure 10 shows the differences between the reference forecast and the sensitivity simulation over the United States, Europe and China. Our results indicate that sustained lower anthropogenic emissions lead to a general decrease in surface O_3 concentrations of 10-20% over Eastern China, Europe, and the Western and Northeast US during July and August relative to the business-as-usual reference forecast simulation. However, it is also notable that in some locations the model forecast O_3 concentrations increase by an equivalent amount (e.g., Scandinavia, South Central US and Mexico, Northern India), reflecting the high nonlinearity of atmospheric chemistry. This highlights the complex interactions between emissions, chemistry, and meteorology and their impact on air pollution on different time scales.

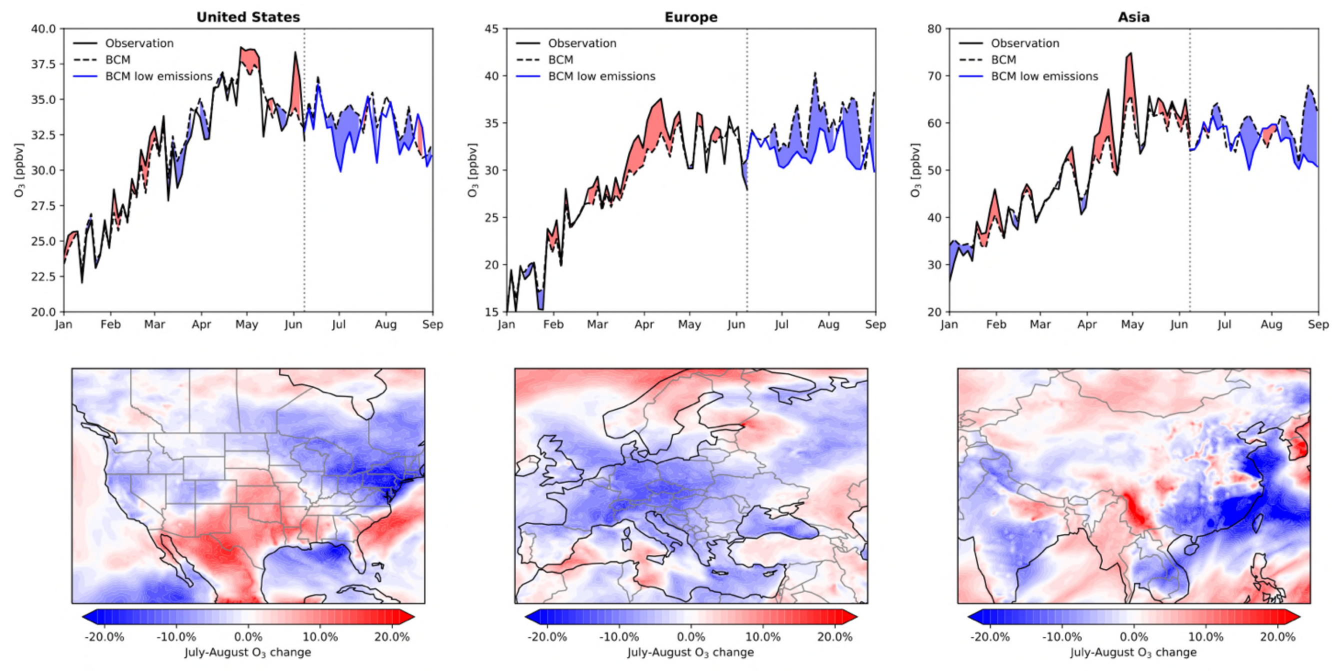


Figure 10: Change in mean surface O_3 over the United States, Europe, and Asia for a sensitivity simulation with altered anthropogenic emissions. Top panels show daily average O_3 concentrations at all observation sites within the given region (solid black, Jan-Jun), the bias-corrected GEOS-CF model ("BCM", solid black, Jan 1st - Jun 8th) continued with a business-as-usual GEOS-CF forecast from Jun 9th - Aug 31st, and GEOS-CF forecast assuming sustained 20% anthropogenic emission reduction (blue). Bottom panels show mean changes in surface O_3 for July and August for the low emissions simulation relative to the business-as-usual forecast.

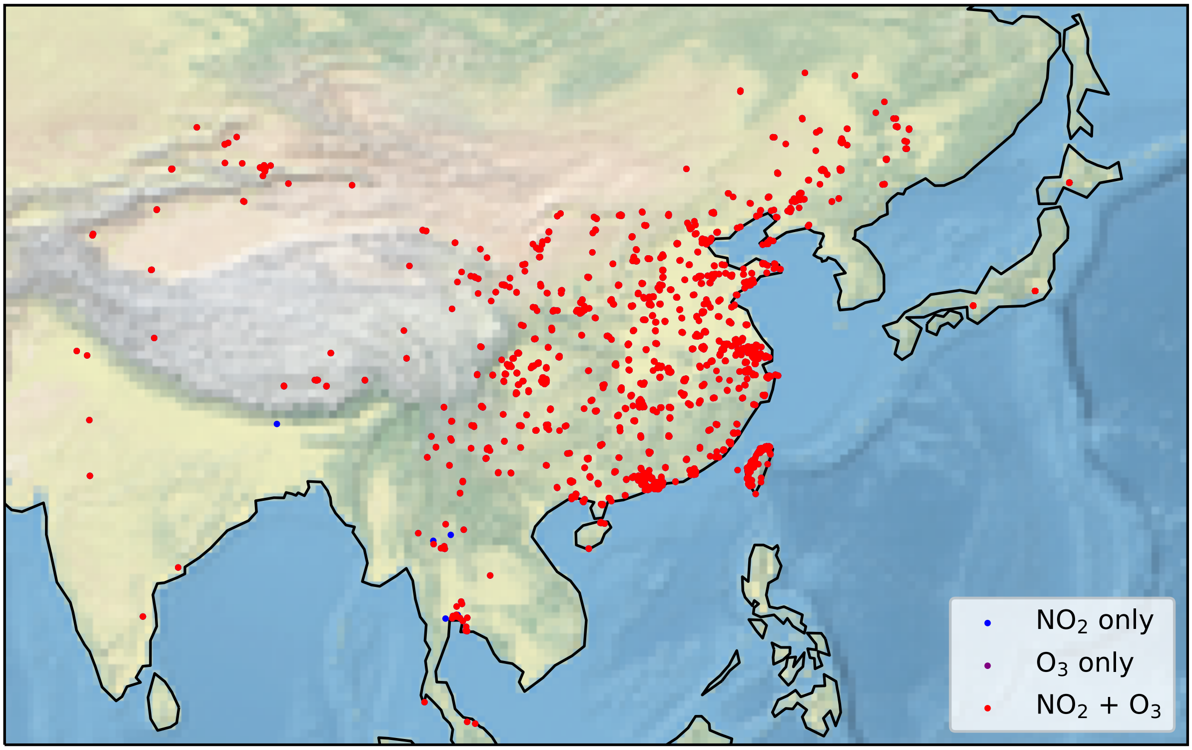
4 Conclusions

The combined interpretation of observations and model simulations using machine learning can be used to remove the compounding effect of meteorology and atmospheric chemistry, offering an effective tool to monitor and quantify changes in air pollution in near real-time. The global response to the COVID-19 pandemic presents a perfect testbed for this type of analysis, offering insights into the interconnectedness of human activity and air pollution. While national mitigation strategies have led to substantial regional NO_2 concentration decreases over the past decades in many places (e.g., Hilboll et al., 2013; Russell et al., 2013; Castellanos and Boersma, 2012), the widespread and near-instantaneous reduction in NO_2 following the implementation of COVID-19 containment measures indicates that there is still large potential to lower human exposure to NO_2 through reduction of anthropogenic NO_x emissions.

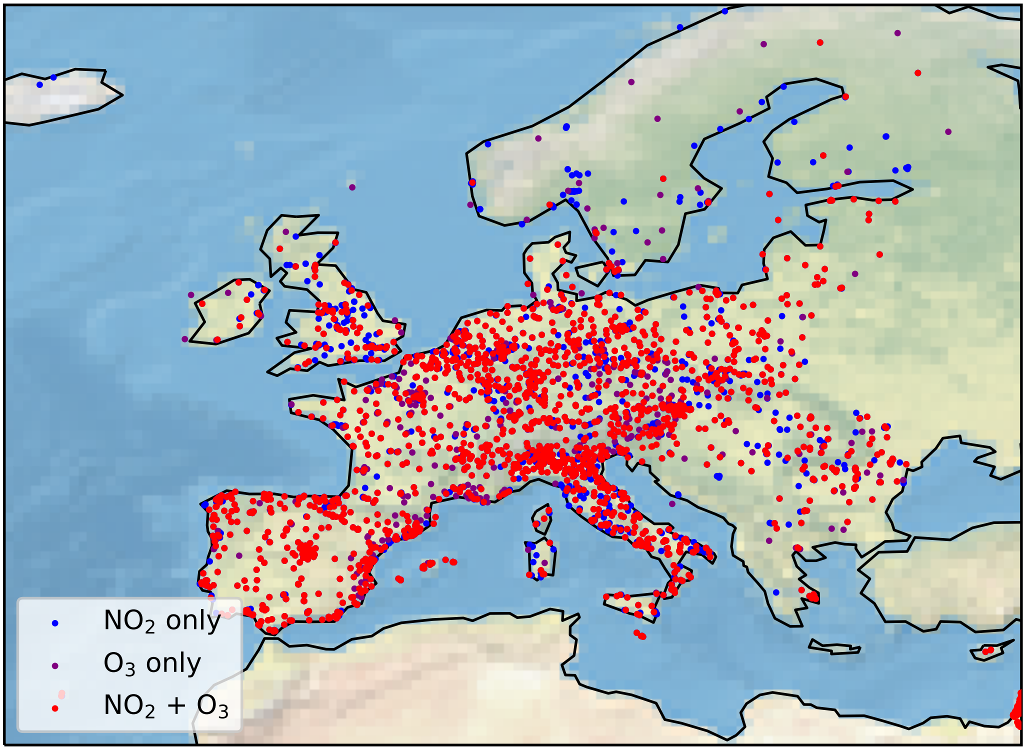
The here derived NO_2 reductions are in good agreement with other emerging estimates. For instance, we determine an 18% decline over China for the 20 days after Chinese New Year relative to the preceding 20 days, consistent with the 21% reduction reported in Liu et al. (2020a). Similarly, our estimated 22% reduction over China for January to March 2020 is in excellent agreement with the 21-23% reported by Liu et al. (2020b). For Spain, we obtain an NO_2 reduction of 46% between March 14 to April 23, again in close agreement with the values reported in Petetin et al. (2020).

While large reductions in NO_2 concentrations are achievable and immediately follow curtailments in NO_x emissions, the O_3 response is more complicated and can be in the opposite direction, at times by as much as 50% (Jhun et al., 2015, Le et al., 2020). The O_3 response is dependent on season, time scale, and environment, with an overall tendency to lower surface O_3 under a scenario of sustained NO_x emission reductions. This shows the complexities faced by policy makers in curbing O_3 pollution.

Appendix



525 **Figure A1:** Close up of Chinese observation sites included in the analysis. Red points indicate sites with both NO2 and O3 observations, purple points show locations with O3 observations only and blue points show locations with NO2 observations only.



530 **Figure A2:** Close up of European observation sites included in the analysis. Red points indicate sites with both NO2 and O3 observations, purple points show locations with O3 observations only and blue points show locations with NO2 observations only.

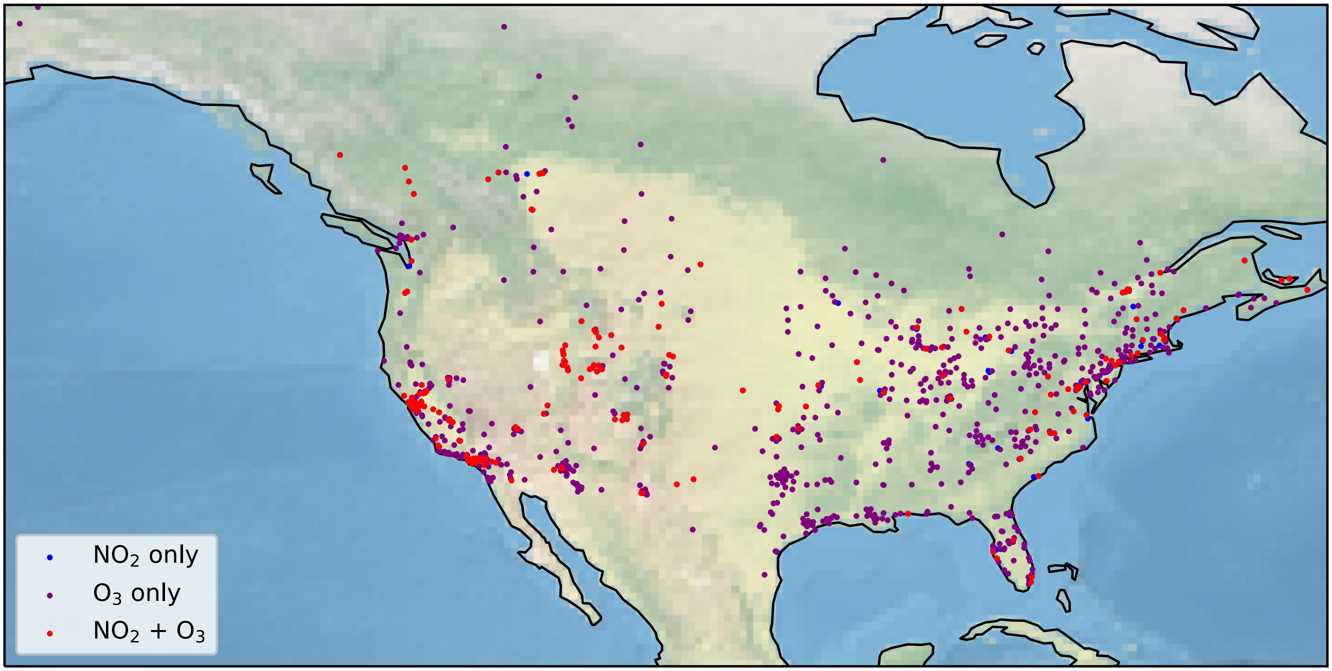


Figure A3: Close up of North American observation sites included in the analysis. Red points indicate sites with both NO2 and O3 observations, purple points show locations with O3 observations only and blue points show locations with NO2 observations only.

535

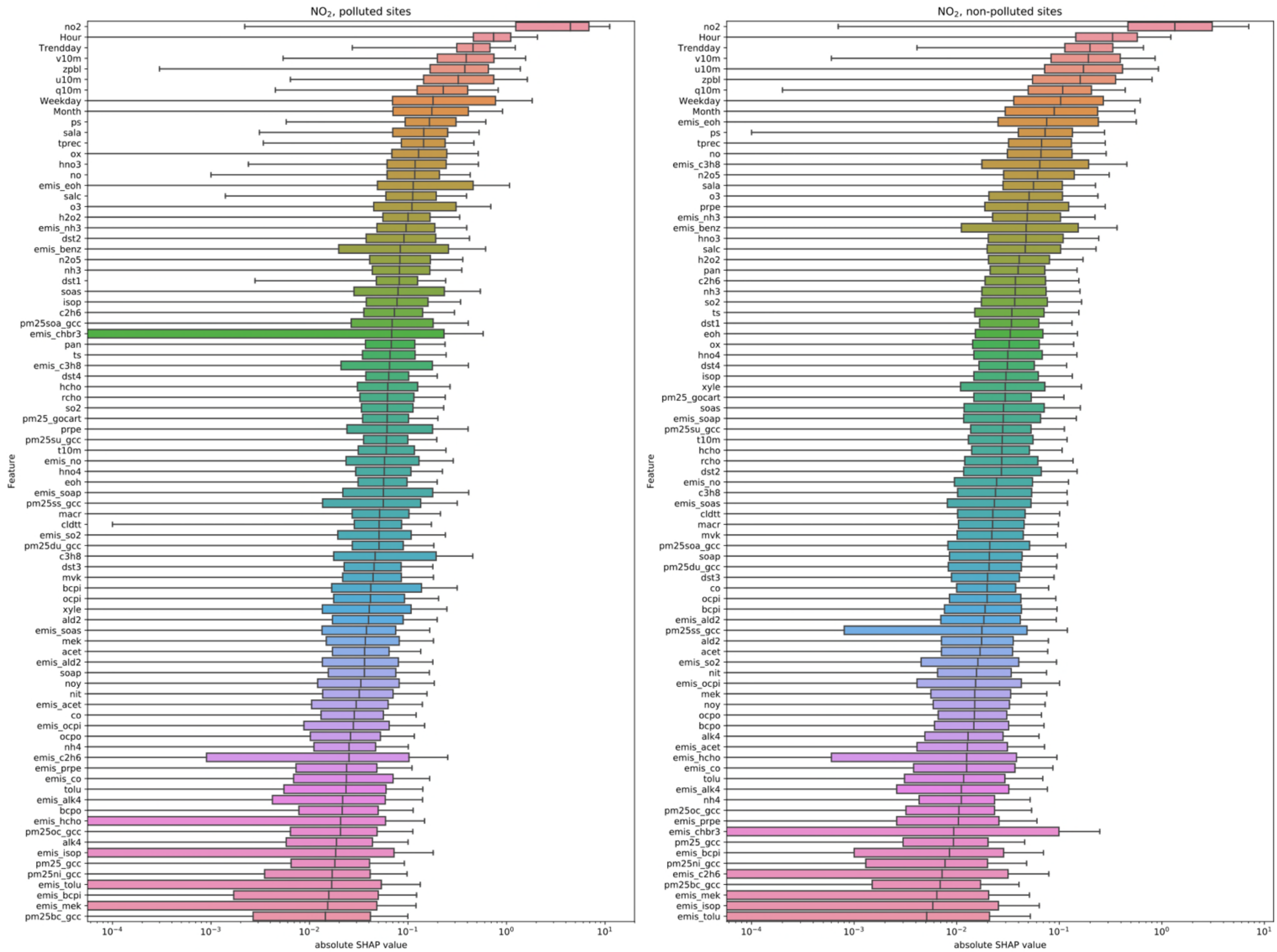


Figure A4: Importance of input variables (features) for the XGBoost models trained to predict NO_2 model bias. Shown are the distribution of the absolute SHAP values for each feature, ranked by the average importance of each feature. Higher SHAP value indicates higher feature importance. Left panel shows results for polluted sites (average annual NO_2 concentration > 15ppbv) and right panel shows results for all other sites.

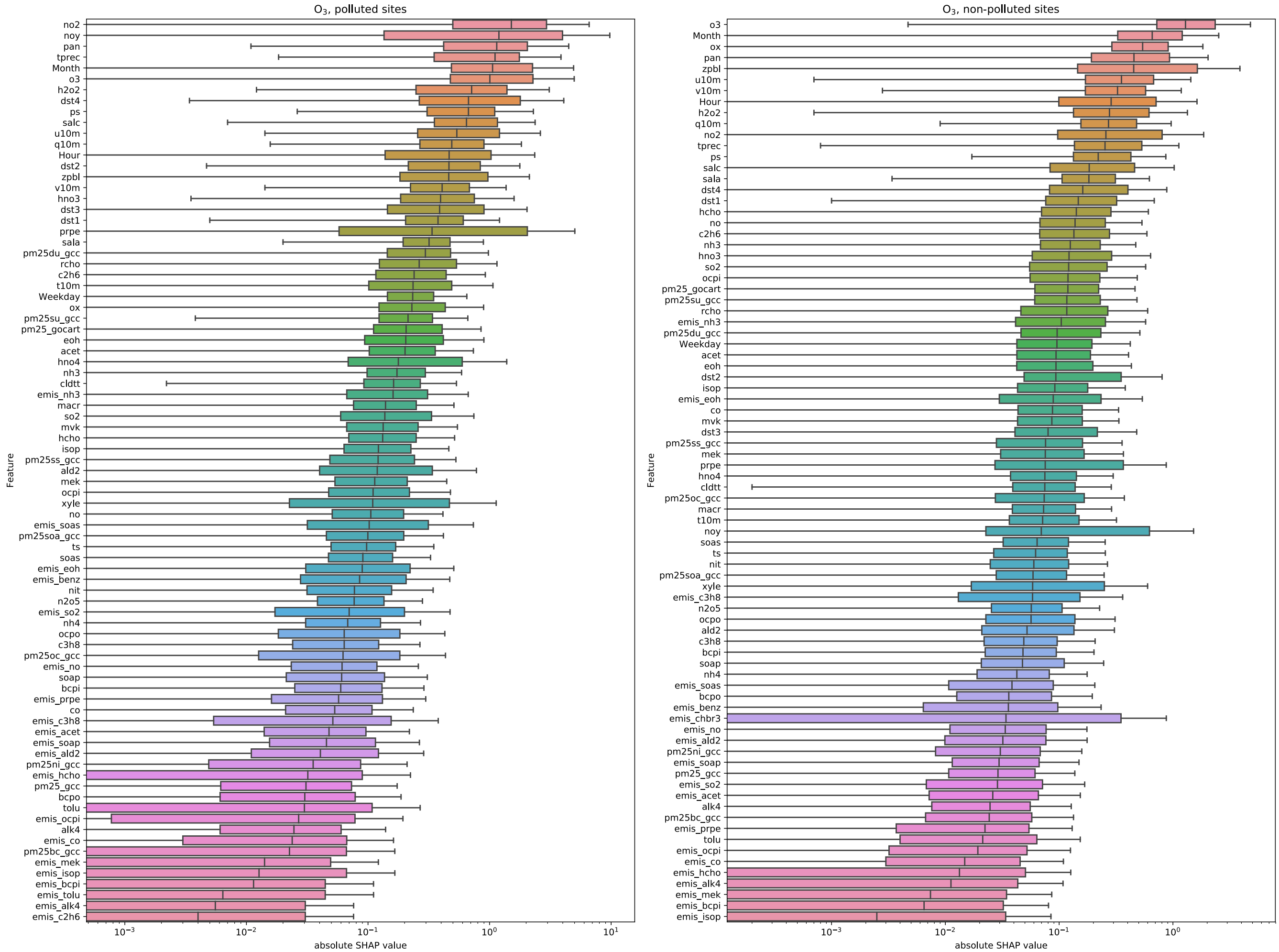


Figure A5: As Figure A4 but for O_3.

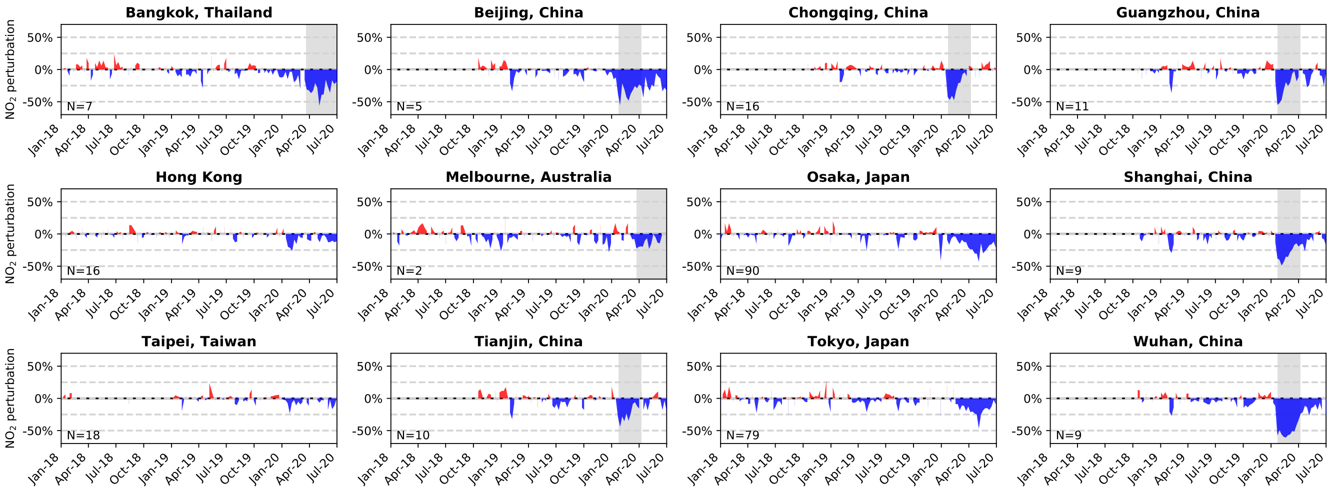


Figure A6: Normalized fractional NO_2 perturbations (observation - bias-corrected model, normalized by the bias-corrected model prediction) from Jan 1, 2018 through June 2020 for selected cities in Asia and Australia. Grey shaded areas indicate COVID-19 lockdown periods. Number of sites per city are shown in the bottom left of each panel.

550

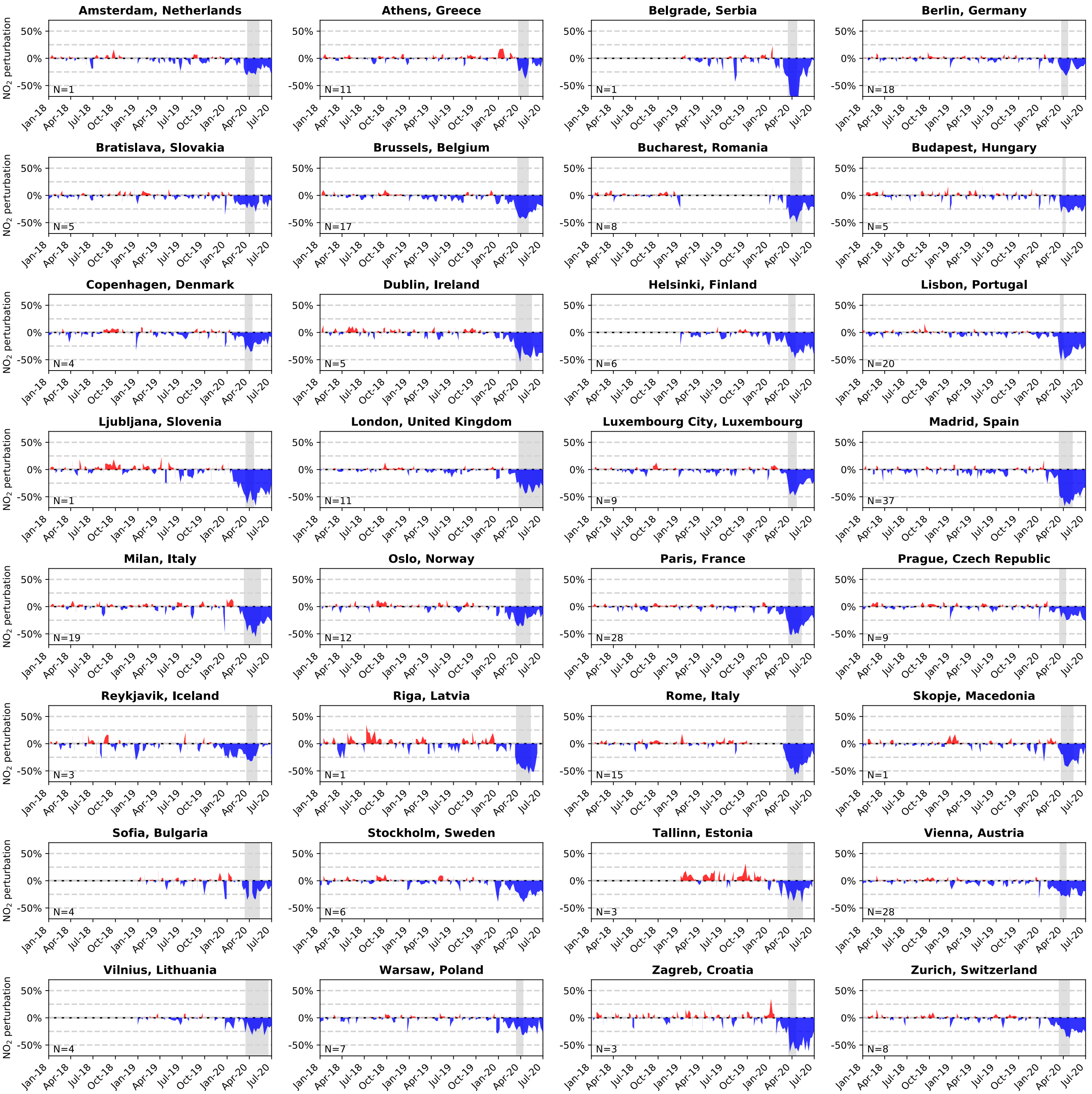


Figure A7: As figure A6 but for Europe.

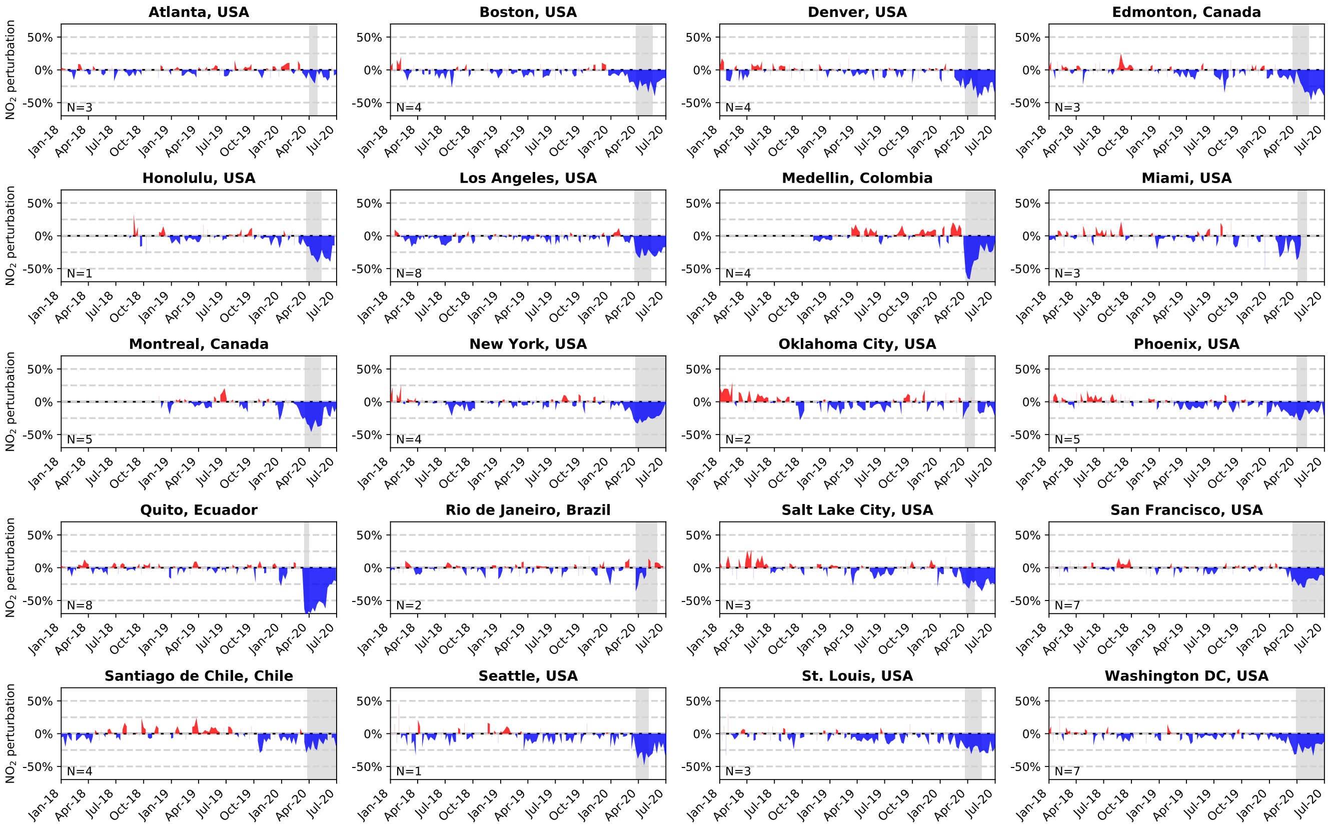


Figure A8: As Figure A6 but for North and South America.

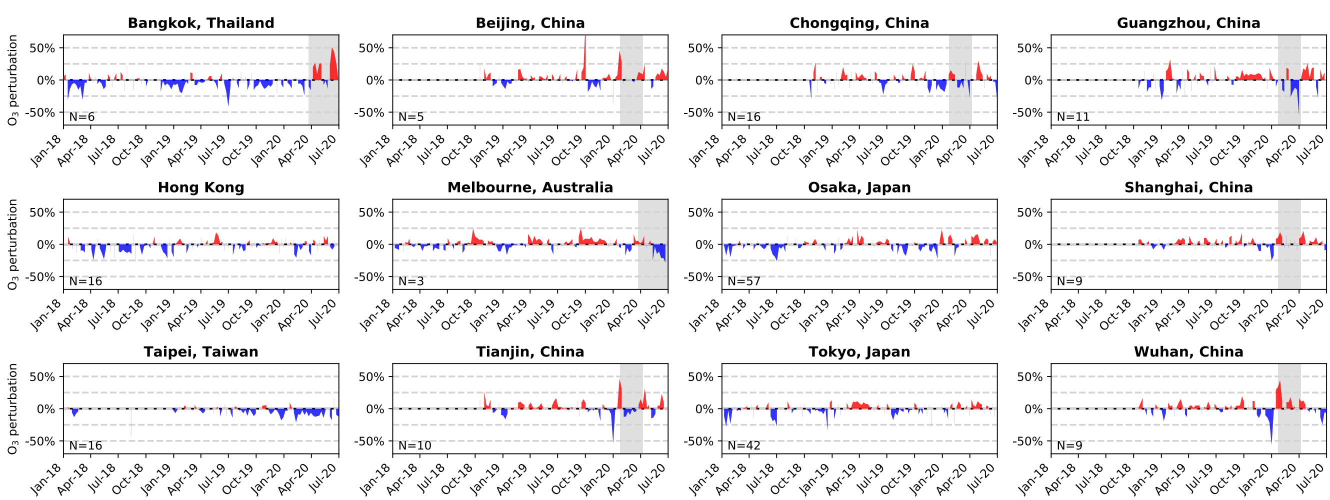


Figure A9: Normalized fractional O_3 perturbations (observation - bias-corrected model, normalized by the bias-corrected model prediction) from Jan 1, 2018 through June 2020 for selected cities in Asia and Australia. Grey shaded areas indicate COVID-19 lockdown periods. Number of sites per city are shown in the bottom left of each panel.

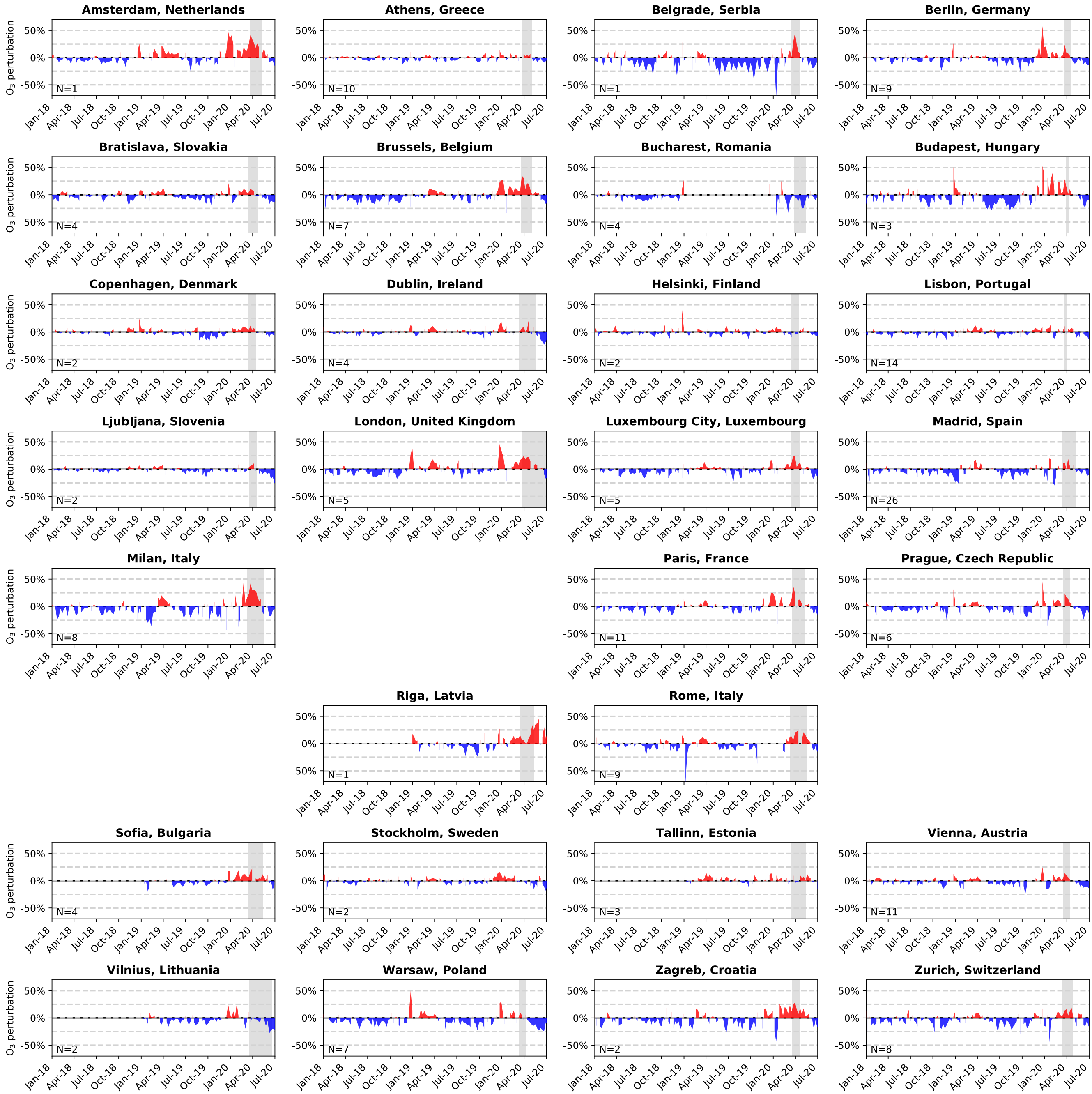
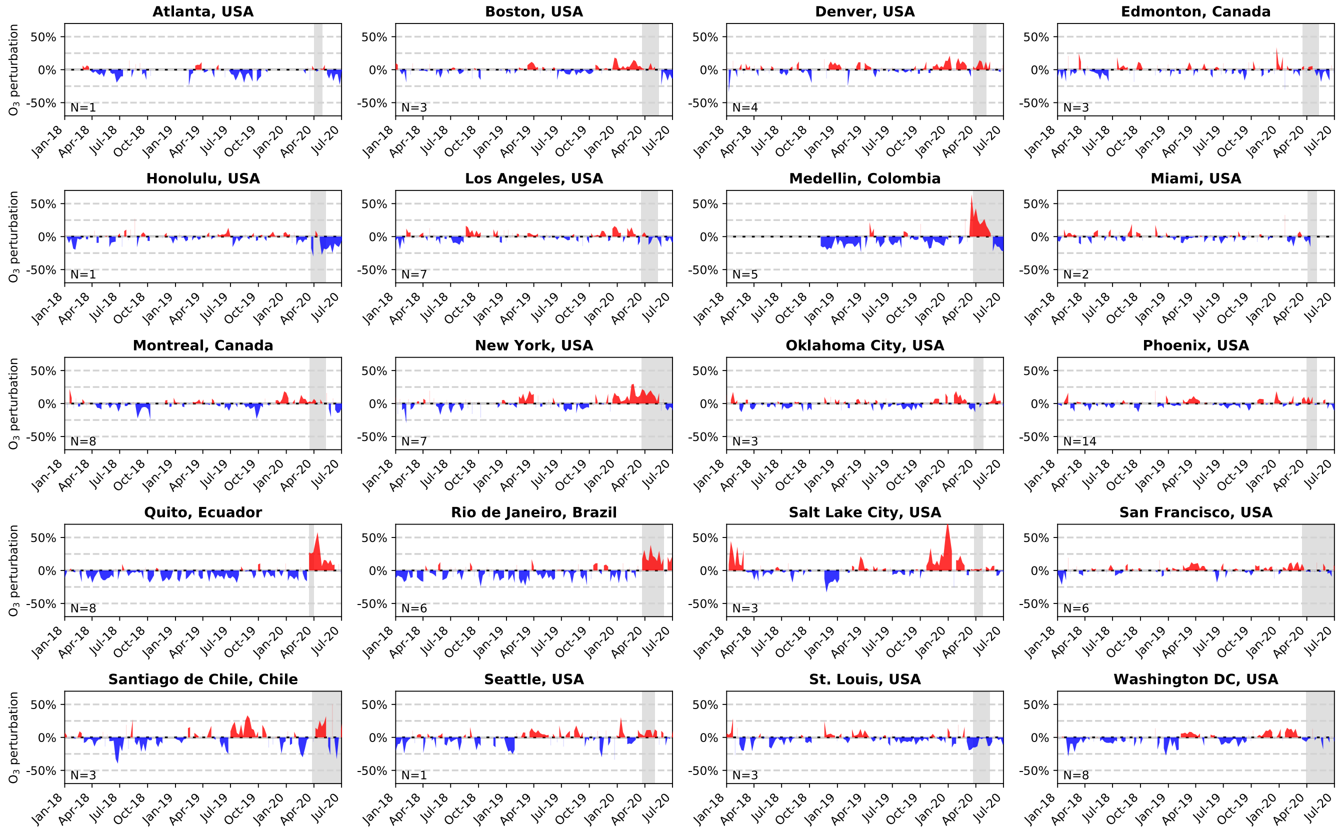


Figure A10: As Figure A9 but for Europe. No observations are available for Reykjavik, Oslo, and Skopje.



565 **Figure A11:** As Figure A9 but for North and South America.

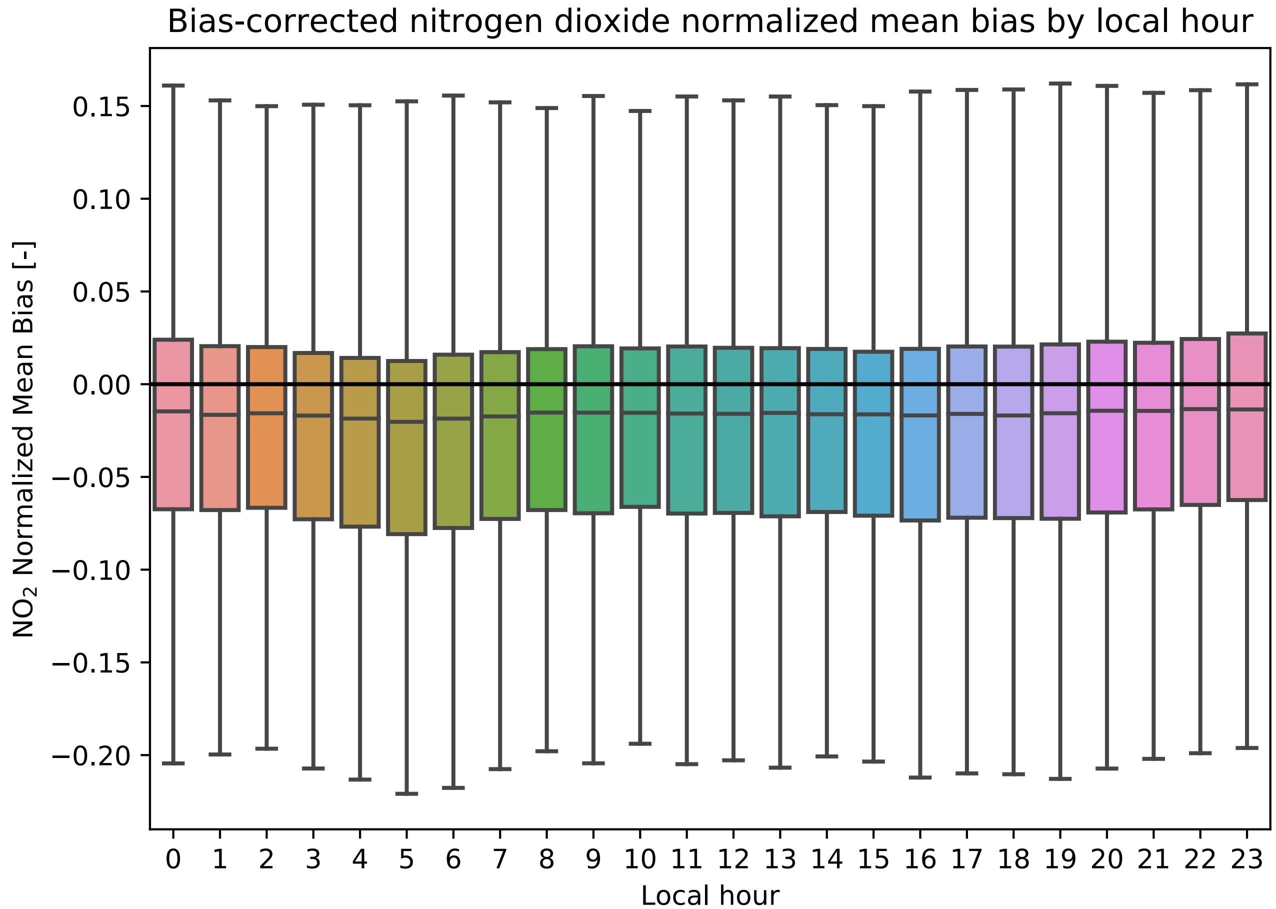


Figure A12: Distribution of normalized mean bias of the machine-learning corrected nitrogen dioxide concentrations at all available observation sites as a function of local hour. Shown are the results for the test dataset.

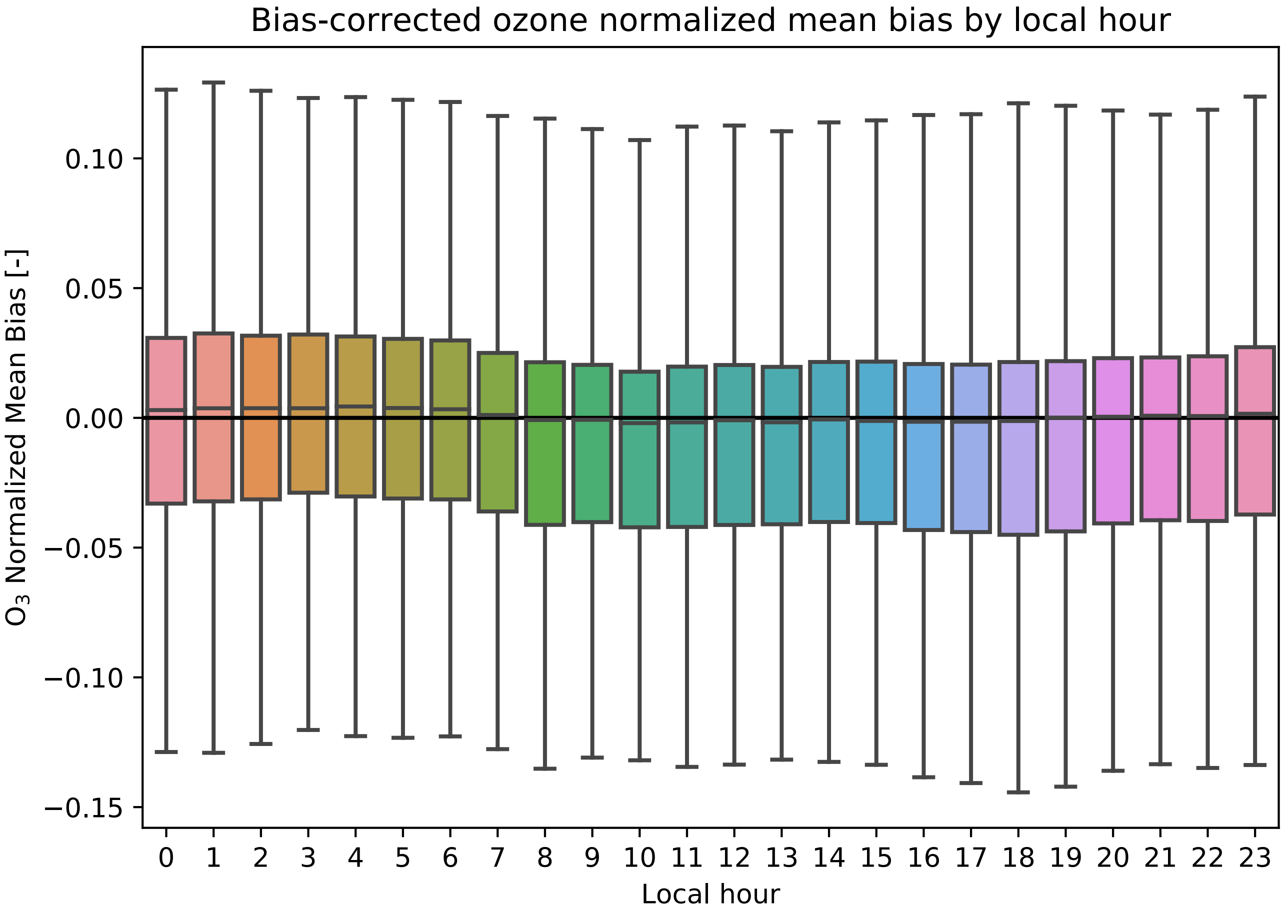


Figure A13: As Figure A12 but for ozone.

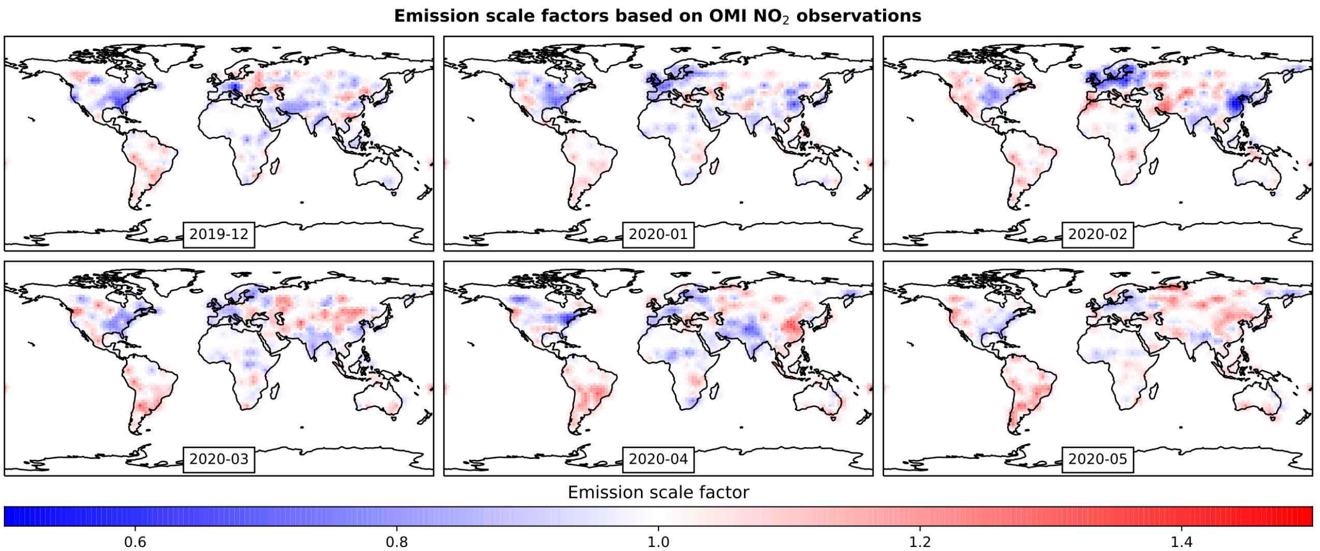


Figure A14: Emission scale factors used for model sensitivity simulation. Shown are the monthly average perturbations applied to the GEOS-CF anthropogenic base emissions.

Table A1: List of input features fed into the XGBoost machine learning model.

Short name	Description	Unit
no2	nitrogen dioxide	ppbv
no	nitrogen oxide	ppbv
noy	reactive nitrogen (no + no2 + nitrates)	ppbv
o3	ozone	ppbv
co	carbon monoxide	ppbv
acet	acetone	ppbv
alk4	alkanes	ppbv
ald2	acetaldehyde	ppbv
hcho	formaldehyde	ppbv
c2h6	ethane	ppbv
c3h8	propane	ppbv
bcpi	hydrophilic black carbon	ppbv
bcpo	hydrophobic black carbon	ppbv
ocpi	hydrophilic organic carbon	ppbv
ocpo	hydrophobic organic carbon	ppbv
eoh	ethanol	ppbv
dst1	dust with diameter of 0.7 microns	ppbv
dst2	dust with diameter of 1.4 microns	ppbv
dst3	dust with diameter of 2.4 microns	ppbv
dst4	dust with diameter of 4.5 microns	ppbv
h2o2	hydrogen peroxide	ppbv
hno3	nitric acid	ppbv
hno4	peroxynitric acid	ppbv
isop	isoprene	ppbv
macr	methacrolein	ppbv
mek	methyl ethyl ketone	ppbv
mvk	methyl vinyl ketone	ppbv
n2o5	dinitrogen pentoxide	ppbv
nh3	ammonia	ppbv
nh4	ammonium	ppbv
nit	inorganic nitrates	ppbv
pan	peroxyacetyl nitrate	ppbv
prpe	alkenes	ppbv
rcho	aldehyde	ppbv
sala	fine sea salt aerosol	ppbv
salc	coarse sea salt aerosol	ppbv
so2	sulfur dioxide	ppbv
soap	secondary organic aerosol precursor	ppbv
soas	simple secondary organic aerosol	ppbv
tolu	toluene	ppbv
xyle	xylene	ppbv
ox	odd oxygen (o3 + no2)	ppbv
pm25_gcc	total PM2.5	µg m-3
pm25ni_gcc	nitrate PM2.5	µg m-3
pm25su_gcc	sulfate PM2.5	µg m-3
pm25ss_gcc	sea salt PM2.5	µg m-3
pm25du_gcc	dust PM2.5	µg m-3
pm25bc_gcc	black carbon PM2.5	µg m-3
pm25oc_gcc	organic carbon PM2.5	µg m-3
pm25soa_gcc	secondary organic aerosol PM2.5	µg m-3
pm25_gocart	total PM2.5 as calculated by the GOCART model	µg m-3

Table A1: cont.

Short name	Description	unit
Hour	hour of day	-
Weekday	day of the week	-
Month	month of the year	-
Trendday	days since Jan 1, 2018	days
cldtt	total cloud fraction	unitless
ps	surface pressure	Pa
q10m	specific humidity at 10m	kg/kg
t10m	temperature at 10m	K
tprec	total precipitation	mm
ts	skin surface temperature	K
u10m	10m East-West wind-speed	m/s
v10m	10m North-South wind speed	m/s
zpbl	planetary boundary layer height	m
emis_no	nitroxen oxide emissions	µg m-2 s-1
emis_co	carbon monoxide emissions	µg m-2 s-1
emis_acet	acetone emissions	µg m-2 s-1
emis_ald2	acetaldehyde emissions	µg m-2 s-1
emis_alk4	alkanes emissions	µg m-2 s-1
emis_benz	benzene emissions	µg m-2 s-1
emis_c2h6	ethane emissions	µg m-2 s-1
emis_prpe	alkenes emissions	µg m-2 s-1
emis_tolu	toluene emissions	µg m-2 s-1
emis_xyle	xylene emissions	µg m-2 s-1
emis_isop	isoprene emissions	µg m-2 s-1
emis_bcpi	hydrophilic black carbon emissions	µg m-2 s-1
emis_bcpo	hydrophobic black carbon emissions	µg m-2 s-1
emis_ocpi	hydrophilic organic carbon emissions	µg m-2 s-1
emis_ocpo	hydrophobic organic carbon emissions	µg m-2 s-1
emis_sala	fine sea salt aerosol emissions	µg m-2 s-1
emis_salc	coarse sea salt aerosol emissions	µg m-2 s-1
emis_so2	sulfur dioxide emissions	µg m-2 s-1
emis_soap	secondary organic aerosol precursor emissions	µg m-2 s-1
emis_soas	simple secondary organic aerosol emissions	µg m-2 s-1
emis_chbr3	bromoform emissions	µg m-2 s-1

Table A2: National lockdown dates used for visualizations.

Country	Start Date	End Date
Australia	March 23	-
Austria	March 16	April 13
Belgium	March 18	April 4
Bosnia and Herzegovina	March 10	April 27
Brazil	March 13	June 2
Bulgaria	March 13	May 13
Canada	March 18	May 11
Chile	March 26	-
China	Jan 23	April 8
Colombia	March 24	June 1
Croatia	March 18	April 19
Czech Republic	March 16	April 12
Denmark	March 13	April 13
Ecuador	March 16	March 31
Estonia	March 13	May 15
Finland	March 16	April 15
France	March 17	May 11
Germany	March 23	April 20
Greece	March 23	May 4
Hungary	March 28	April 10
Iceland	March 21	May 4
India	March 25	-
Ireland	March 12	May 18
Italy	March 9	May 18
Latvia	March 13	May 12
Lithuania	March 16	June 18
Luxembourg	March 15	April 20
Macedonia	March 22	April 12
Malta	March 12	May 4
The Netherlands	March 23	May 11
Norway	March 12	May 11
Poland	March 13	April 11
Portugal	March 19	April 2
Romania	March 25	May 12
Serbia	March 15	April 21
Slovakia	March 16	April 22
Slovenia	March 16	April 20
Spain	March 14	May 9
Switzerland	March 13	April 27
Thailand	March 25	-
United Kingdom	March 23	-
United States	March 19	April 13

Table A3: Monthly changes in NO_2 concentrations relative to the bias-corrected model predictions for cities in Asia & Australia. Values in *italic* denote values that are statistically different from business-as-usual ($p<0.001$ based on Kolmogorov-Smirnov test).

Location	#	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Bangkok, Thailand	7	-7.8% (-10.9%--4.6%)	-11.5% (-15.2%--7.8%)	*-20.4% (-25.9%--14.9%)*	*-27.4% (-31.9%--22.9%)*	*-39.1% (-43.5%--34.6%)*	*-25.7% (-30.5%--20.9%)*
Beijing, China	5	*-21.1% (-23.2%--19.0%)*	*-38.8% (-41.3%--36.4%)*	*-31.7% (-34.6%--28.7%)*	*-25.0% (-28.5%--21.6%)*	*-23.8% (-27.4%--20.3%)*	*-20.4% (-24.1%--16.6%)*
Chongqing, China	16	-11.1% (-13.0%--9.2%)	*-38.4% (-40.5%--36.3%)*	*-11.6% (-13.6%--9.5%)*	-0.1% (-2.1%-1.9%)	4.3% (2.0%-6.5%)	3.5% (1.1%-6.0%)
Guangzhou, China	11	-16.6% (-19.5%--13.7%)	*-31.8% (-35.1%--28.4%)*	*-13.6% (-16.7%--10.5%)*	-5.6% (-8.7%--2.5%)	-11.5% (-15.6%--7.5%)	-5.8% (-11.1%--0.6%)
Hong Kong	16	-6.9% (-8.8%--4.9%)	-13.9% (-16.0%--11.8%)	-3.7% (-5.8%--1.6%)	-4.1% (-6.1%--2.1%)	-4.7% (-6.9%--2.4%)	*-12.8% (-15.7%--10.0%)*
Melbourne, Australia	2	4.6% (-5.4%-14.6%)	1.5% (-9.0%-12.0%)	-7.6% (-16.6%-1.5%)	*-13.9% (-21.9%--5.8%)*	-11.9% (-19.1%--4.7%)	-7.8% (-14.2%--1.5%)
Osaka, Japan	90	-14.4% (-15.4%--13.4%)	-6.4% (-7.3%--5.5%)	-13.4% (-14.3%--12.5%)	*-25.7% (-26.8%--24.7%)*	*-28.1% (-29.2%--27.0%)*	-17.4% (-18.6%--16.3%)
Shanghai, China	9	-14.4% (-16.5%--12.3%)	*-39.8% (-42.1%--37.5%)*	*-23.3% (-25.7%--20.8%)*	*-19.3% (-21.7%--17.0%)*	-5.5% (-8.4%--2.7%)	-1.9% (-5.0%-1.2%)
Taipei, Taiwan	18	*-7.8% (-9.4%--6.2%)*	-6.4% (-8.0%--4.8%)	-5.1% (-6.7%--3.6%)	-2.0% (-3.6%--0.3%)	-3.0% (-4.7%--1.2%)	*-11.1% (-13.0%--9.3%)*
Tianjin, China	10	-7.2% (-8.7%--5.8%)	*-32.1% (-34.0%--30.3%)*	*-14.5% (-16.5%--12.4%)*	-8.4% (-10.9%--5.9%)	-1.0% (-3.7%-1.7%)	-2.5% (-5.5%-0.6%)
Tokyo, Japan	79	1.6% (0.5%-2.7%)	-2.5% (-3.6%--1.4%)	-13.9% (-15.1%--12.8%)	-21.3% (-22.5%--20.1%)	*-26.7% (-28.0%--25.5%)*	-12.0% (-13.3%--10.6%)
Wuhan, China	9	*-20.4% (-23.6%--17.3%)*	*-56.3% (-59.4%--53.2%)*	*-51.4% (-54.6%--48.2%)*	*-22.0% (-25.3%--18.7%)*	-15.0% (-18.7%--11.2%)	-11.1% (-15.9%--6.2%)

Table A4: Monthly changes in NO_2 concentrations relative to the bias-corrected model predictions for cities in Europe. Values in *italic* denote values that are statistically different from business-as-usual ($p<0.001$ based on Kolmogorov-Smirnov test).

Location	#	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Amsterdam, Netherlands	1	-1.5% (-8.4%-5.5%)	-8.5% (-17.8%-0.8%)	*-16.2% (-25.8%--6.5%)*	*-27.2% (-36.9%--17.5%)*	*-18.4% (-29.9%--6.9%)*	-17.2% (-28.0%--6.5%)
Athens, Greece	11	*13.8% (11.4%-16.1%)*	1.9% (-0.2%-4.1%)	-5.0% (-7.1%--3.0%)	*-26.4% (-28.7%--24.2%)*	-8.8% (-10.9%--6.7%)	*-12.7% (-15.0%--10.5%)*
Belgrade, Serbia	1	2.8% (-5.4%-11.0%)	-10.6% (-20.0%--1.1%)	*-40.9% (-50.6%--31.2%)*	*-71.7% (-81.5%--61.9%)*	*-31.0% (-44.2%--17.7%)*	-17.0% (-33.1%--0.9%)
Berlin, Germany	18	2.1% (0.4%-3.8%)	*-7.3% (-9.4%--5.2%)*	-9.7% (-11.6%--7.8%)	*-25.3% (-27.3%--23.3%)*	*-12.6% (-15.0%--10.2%)*	*-15.4% (-17.9%--12.9%)*
Bratislava, Slovakia	5	0.3% (-2.9%-3.4%)	*-15.7% (-19.8%--11.5%)*	*-18.3% (-22.2%--14.5%)*	*-18.2% (-22.3%--14.2%)*	*-12.4% (-17.7%--7.0%)*	*-14.9% (-20.5%--9.3%)*
Brussels, Belgium	17	-5.2% (-7.1%--3.4%)	*-11.8% (-14.2%--9.3%)*	*-25.4% (-27.4%--23.4%)*	*-40.8% (-42.6%--39.0%)*	*-30.7% (-33.0%--28.5%)*	*-19.7% (-22.3%--17.0%)*
Bucharest, Romania	8	-1.7% (-4.5%-1.0%)	*-9.9% (-12.9%--7.0%)*	*-21.6% (-24.6%--18.6%)*	*-39.8% (-43.0%--36.5%)*	*-23.3% (-26.9%--19.8%)*	*-23.0% (-26.5%--19.5%)*
Budapest, Hungary	5	-2.2% (-5.3%-0.8%)	-9.2% (-12.9%--5.4%)	-12.0% (-15.5%--8.6%)	*-23.6% (-27.1%--20.1%)*	*-24.4% (-28.8%--20.0%)*	*-25.7% (-30.4%--21.1%)*
Copenhagen, Denmark	4	-1.7% (-6.2%-2.7%)	*-9.6% (-14.9%--4.3%)*	*-19.0% (-23.5%--14.6%)*	*-25.7% (-30.2%--21.3%)*	*-18.7% (-24.1%--13.3%)*	*-13.8% (-18.5%--9.0%)*
Dublin, Ireland	5	*-8.7% (-14.7%--2.7%)*	*-9.3% (-17.2%--1.5%)*	*-28.8% (-34.4%--23.1%)*	*-39.2% (-43.7%--34.8%)*	*-38.0% (-44.2%--31.8%)*	*-40.8% (-47.3%--34.2%)*
Helsinki, Finland	6	-9.1% (-13.2%--4.9%)	*-15.2% (-19.2%--11.1%)*	*-18.2% (-22.0%--14.5%)*	*-38.8% (-43.0%--34.5%)*	*-33.0% (-37.5%--28.6%)*	*-25.9% (-29.7%--22.2%)*
Lisbon, Portugal	20	-6.1% (-8.1%--4.1%)	-3.4% (-5.3%--1.5%)	*-26.9% (-29.4%--24.3%)*	*-41.3% (-44.1%--38.6%)*	*-32.5% (-35.4%--29.6%)*	*-28.1% (-32.0%--24.1%)*
Ljubljana, Slovenia	1	-3.8% (-11.4%-3.9%)	*-26.0% (-33.4%--18.7%)*	*-47.9% (-56.0%--39.7%)*	*-49.8% (-59.2%--40.3%)*	*-41.4% (-52.9%--29.9%)*	*-42.6% (-54.1%--31.1%)*
London, United Kingdom	11	*-3.6% (-6.0%--1.1%)*	-4.3% (-7.3%--1.2%)	*-17.0% (-19.7%--14.3%)*	*-35.9% (-38.1%--33.6%)*	*-35.1% (-37.8%--32.3%)*	*-28.8% (-31.9%--25.8%)*
Luxembourg City, Luxembourg	9	3.7% (1.0%-6.4%)	-3.4% (-6.7%--0.1%)	*-24.4% (-27.4%--21.4%)*	*-40.2% (-42.9%--37.6%)*	*-27.0% (-30.1%--23.9%)*	*-19.2% (-22.7%--15.7%)*
Madrid, Spain	37	0.9% (-0.5%-2.4%)	-9.4% (-10.8%--8.0%)	*-31.7% (-33.6%--29.8%)*	*-60.6% (-62.6%--58.6%)*	*-51.8% (-54.0%--49.6%)*	*-40.4% (-42.8%--38.0%)*
Milan, Italy	19	*10.0% (8.8%-11.2%)*	-2.0% (-3.6%--0.3%)	*-25.7% (-27.2%--24.2%)*	*-41.6% (-43.3%--39.9%)*	*-34.7% (-36.9%--32.6%)*	*-24.4% (-26.6%--22.1%)*
Oslo, Norway	12	-6.9% (-9.3%--4.5%)	*-16.0% (-18.5%--13.5%)*	*-25.4% (-28.1%--22.6%)*	*-26.3% (-29.4%--23.2%)*	*-18.8% (-22.4%--15.2%)*	*-12.2% (-15.7%--8.7%)*
Paris, France	28	-3.2% (-4.5%--1.8%)	*-12.7% (-14.4%--11.0%)*	*-28.0% (-29.5%--26.6%)*	*-46.6% (-48.0%--45.2%)*	*-35.7% (-37.3%--34.1%)*	*-21.1% (-22.9%--19.3%)*
Prague, Czech Republic	9	3.9% (1.6%-6.2%)	*-9.1% (-12.1%--6.0%)*	-10.1% (-12.6%--7.5%)	*-16.6% (-19.0%--14.2%)*	*-16.7% (-19.5%--13.9%)*	*-17.7% (-20.6%--14.8%)*
Reykjavik, Iceland	3	-20.9% (-28.1%--13.7%)	*-17.4% (-23.6%--11.1%)*	*-20.0% (-27.1%--12.9%)*	*-28.1% (-36.1%--20.1%)*	-6.7% (-17.0%-3.5%)	-0.2% (-9.9%-9.4%)
Riga, Latvia	1	-5.7% (-17.2%-5.9%)	*-7.6% (-20.2%-4.9%)*	*-23.9% (-34.9%--12.9%)*	*-45.5% (-60.3%--30.8%)*	*-48.9% (-62.6%--35.2%)*	-12.6% (-24.7%--0.5%)
Rome, Italy	15	N/A	-0.9% (-3.3%-1.5%)	*-32.3% (-34.3%--30.2%)*	*-50.0% (-52.1%--47.8%)*	*-39.3% (-41.6%--36.9%)*	*-24.2% (-26.8%--21.7%)*
Skopje, Macedonia	1	-9.5% (-17.0%--2.0%)	5.6% (-2.5%-13.8%)	-8.1% (-16.3%-0.0%)	*-35.5% (-43.9%--27.2%)*	*-32.2% (-41.5%--22.9%)*	*-18.0% (-27.5%--8.5%)*
Sofia, Bulgaria	4	8.5% (3.9%-	-3.4% (-8.2%-	*-16.5% (-*	*-27.3% (-*	-20.3% (-	*-10.7% (-*

		13.1%)	1.4%)	*21.6%--11.4%)*	*34.5%--20.1%)*	29.1%--11.4%)	*18.2%--3.3%)*
Stockholm, Sweden	6	-11.1% (-15.4%--6.8%)	*-11.8% (-15.9%--7.7%)*	*-18.0% (-22.0%--14.1%)*	*-33.0% (-37.2%--28.8%)*	*-23.7% (-28.0%--19.4%)*	*-20.9% (-25.4%--16.5%)*
Tallinn, Estonia	3	-3.2% (-11.8%-5.4%)	*-12.3% (-21.1%--3.5%)*	*-14.8% (-24.1%--5.4%)*	*-25.2% (-34.1%--16.2%)*	*-22.3% (-31.3%--13.4%)*	-9.7% (-17.3%--2.1%)
Vienna, Austria	28	-2.3% (-3.6%--0.9%)	*-14.4% (-16.3%--12.6%)*	*-20.6% (-22.3%--18.9%)*	*-26.2% (-27.9%--24.5%)*	*-16.5% (-18.8%--14.2%)*	*-25.4% (-27.6%--23.1%)*
Vilnius, Lithuania	4	-11.8% (-16.5%--7.2%)	-3.5% (-8.5%-1.5%)	-10.1% (-14.9%--5.3%)	*-24.4% (-29.7%--19.0%)*	-16.2% (-22.1%--10.3%)	*-20.3% (-26.8%--13.8%)*
Warsaw, Poland	7	-10.4% (-13.1%--7.6%)	*-15.7% (-18.7%--12.6%)*	*-15.2% (-18.0%--12.4%)*	*-26.0% (-28.8%--23.2%)*	*-18.3% (-21.6%--15.1%)*	*-17.5% (-21.3%--13.7%)*
Zagreb, Croatia	3	8.0% (2.9%-13.1%)	-15.2% (-20.9%--9.6%)	*-32.0% (-37.9%--26.0%)*	*-51.5% (-57.5%--45.4%)*	*-50.0% (-56.4%--43.6%)*	*-42.3% (-50.0%--34.6%)*
Zurich, Switzerland	8	2.2% (0.0%-4.5%)	*-11.8% (-14.8%--8.7%)*	*-15.1% (-17.9%--12.3%)*	*-28.0% (-30.8%--25.3%)*	*-23.4% (-26.9%--19.9%)*	*-25.5% (-29.1%--22.0%)*

Table A5: Monthly changes in NO_2 concentrations relative to the bias-corrected model predictions for cities in North and South America. Values in *italic* denote values that are statistically different from business-as-usual (p<0.001 based on Kolmogorov-Smirnov test).

Location	#	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Atlanta, USA	3	9.7% (4.7%-14.8%)	0.8% (-4.4%-6.1%)	-3.3% (-8.5%-2.0%)	*-12.4% (-17.6%--7.2%)*	-6.7% (-12.3%--1.0%)	-7.1% (-13.3%--0.9%)
Boston, USA	4	-8.4% (-12.3%--4.6%)	-5.1% (-9.0%--1.2%)	*-20.7% (-25.6%--15.9%)*	*-23.4% (-29.0%--17.7%)*	*-27.0% (-33.1%--20.9%)*	-14.7% (-20.8%--8.6%)
Denver, USA	4	-4.5% (-7.3%--1.7%)	-7.5% (-10.9%--4.1%)	-20.5% (-24.4%--16.7%)	*-21.2% (-25.9%--16.5%)*	*-35.8% (-41.1%--30.6%)*	*-24.1% (-29.3%--18.9%)*
Edmonton, Canada	3	-4.0% (-7.4%--0.6%)	-11.5% (-15.1%--7.9%)	-15.2% (-19.9%--10.5%)	-18.0% (-24.4%--11.5%)	*-34.9% (-43.2%--26.6%)*	*-32.3% (-40.9%--23.7%)*
Honolulu, USA	1	-3.1% (-16.3%-10.2%)	-0.9% (-15.8%-14.1%)	-11.7% (-26.4%-3.0%)	*-28.3% (-42.4%--14.1%)*	*-30.6% (-47.3%--14.0%)*	*-26.7% (-43.1%--10.2%)*
Los Angeles, USA	8	4.7% (2.7%-6.8%)	-4.1% (-6.3%--1.9%)	-11.9% (-15.1%--8.7%)	*-28.2% (-31.6%--24.9%)*	*-27.5% (-31.0%--24.1%)*	*-23.4% (-27.6%--19.3%)*
Medellin, Colombia	4	-2.0% (-6.5%-2.5%)	*11.6% (7.3%-15.8%)*	-7.1% (-11.9%--2.4%)	*-51.9% (-56.2%--47.7%)*	*-24.8% (-29.3%--20.3%)*	*-21.1% (-26.1%--16.0%)*
Miami, USA	3	-4.1% (-14.2%-6.0%)	*-22.6% (-32.6%--12.6%)*	*-12.9% (-21.5%--4.4%)*	*-29.2% (-40.8%--17.5%)*	N/A	N/A
Montreal, Canada	5	-5.4% (-8.8%--2.0%)	-4.4% (-8.2%--0.7%)	*-22.9% (-27.3%--18.6%)*	*-36.4% (-42.3%--30.5%)*	*-25.8% (-32.3%--19.3%)*	*-14.7% (-21.2%--8.2%)*
New York, USA	4	-5.5% (-8.9%--2.2%)	-8.0% (-11.3%--4.7%)	*-20.8% (-24.4%--17.2%)*	*-29.7% (-34.0%--25.4%)*	*-25.1% (-30.3%--19.9%)*	*-17.1% (-23.5%--10.7%)*
Oklahoma City, USA	2	-0.9% (-6.6%-4.7%)	-3.5% (-10.0%-3.0%)	-7.1% (-13.3%--1.0%)	-2.8% (-9.2%-3.5%)	-13.9% (-20.0%--7.7%)	-8.0% (-13.7%--2.4%)
Phoenix, USA	5	-9.5% (-11.9%--7.1%)	-12.6% (-15.7%--9.5%)	*-19.8% (-23.5%--16.1%)*	*-22.2% (-26.1%--18.3%)*	-13.8% (-18.1%--9.5%)	-11.0% (-15.9%--6.2%)
Quito, Ecuador	8	-10.7% (-13.9%--7.6%)	2.5% (-0.5%-5.4%)	*-34.7% (-37.5%--31.8%)*	*-64.3% (-67.2%--61.4%)*	*-55.8% (-58.8%--52.7%)*	*-24.9% (-27.9%--21.8%)*
Rio de Janeiro, Brazil	2	-4.9% (-11.2%-1.4%)	0.8% (-5.4%-6.9%)	-5.2% (-11.7%-1.3%)	-15.9% (-22.2%--9.6%)	6.5% (1.2%-11.9%)	4.4% (-0.3%-9.2%)
Salt Lake City, USA	3	-1.4% (-4.9%-2.1%)	-5.5% (-9.5%--1.6%)	*-15.4% (-20.9%--10.0%)*	*-24.7% (-31.7%--17.7%)*	*-28.3% (-36.3%--20.2%)*	-20.6% (-29.3%--11.9%)
San Francisco, USA	7	4.6% (2.1%-7.2%)	-1.8% (-4.4%-0.7%)	*-14.8% (-18.1%--11.5%)*	*-26.7% (-30.8%--22.6%)*	*-18.4% (-22.7%--14.0%)*	*-15.7% (-20.6%--10.8%)*
Santiago de Chile, Chile	4	-6.6% (-16.6%-3.3%)	-10.5% (-19.1%--1.9%)	-8.4% (-16.0%--0.7%)	*-20.9% (-26.7%--15.0%)*	*-11.8% (-16.6%--7.1%)*	-6.9% (-11.9%--1.8%)
Seattle, USA	1	-8.6% (-18.2%-1.1%)	-2.5% (-12.0%-7.0%)	-12.7% (-22.7%--2.7%)	*-34.0% (-45.0%--22.9%)*	*-28.9% (-41.4%--16.5%)*	-19.8% (-33.9%--5.8%)
St. Louis, USA	3	-4.4% (-9.2%-0.4%)	-11.9% (-17.1%--6.8%)	*-16.4% (-21.6%--11.2%)*	*-24.8% (-30.9%--18.8%)*	*-23.9% (-30.4%--17.4%)*	*-23.8% (-29.9%--17.7%)*
Washington DC, USA	7	-7.0% (-9.9%--4.0%)	-10.0% (-13.0%--7.0%)*	*-18.4% (-21.8%--14.9%)*	*-22.3% (-26.4%--18.2%)*	*-25.2% (-29.9%--20.4%)*	*-14.6% (-19.4%--9.9%)*

Table A6: Monthly changes in O_3 concentrations relative to the bias-corrected model predictions for cities in Asia & Australia. Values in *italic* denote values that are statistically different from business-as-usual (p<0.001 based on Kolmogorov-Smirnov test).

Location	**#**	**20-Jan**	**20-Feb**	**20-Mar**	**20-Apr**	**20-May**	**20-Jun**
Bangkok, Thailand	6	-6.1% (-9.4% - -2.8%)	-2.8% (-6.1% - 0.5%)	-10.2% (-13.9% - -6.5%)	*11.2% (7.4% - 15.1%)*	3.6% (-2.1% - 9.2%)	*34.7% (28.0% - 41.3%)*
Beijing, China	5	10.9% (6.0% - 15.8%)	1.4% (-2.4% - 5.1%)	1.7% (-1.3% - 4.7%)	*11.7% (9.2% - 14.2%)*	-4.4% (-6.3% - -2.5%)	9.5% (7.9% - 11.2%)
Chongqing, China	16	-3.6% (-7.0% - -0.3%)	-2.4% (-4.9% - 0.1%)	4.3% (2.6% - 5.9%)	3.1% (1.6% - 4.6%)	6.9% (5.8% - 8.1%)	2.1% (0.8% - 3.3%)
Guangzhou, China	11	-0.9% (-3.4% - 1.7%)	5.0% (2.3% - 7.8%)	-15.4% (-18.0% - -12.9%)	4.8% (2.6% - 7.0%)	11.3% (9.4% - 13.2%)	-1.5% (-3.8% - 0.9%)
Hong Kong	16	-0.1% (-2.7% - 2.6%)	-8.3% (-10.9% - -5.6%)	-2.9% (-5.6% - -0.2%)	-2.3% (-4.8% - 0.1%)	9.0% (5.0% - 12.9%)	-4.5% (-10.3% - 1.4%)
Melbourne, Australia	3	1.1% (-3.9% - 6.2%)	-8.9% (-14.6% - -3.2%)	3.7% (-1.8% - 9.1%)	2.7% (-2.7% - 8.1%)	-6.1% (-11.8% - -0.5%)	-19.6% (-25.4% - -13.8%)
Osaka, Japan	57	*12.7% (11.3% - 14.2%)*	-3.6% (-4.8% - -2.4%)	1.4% (0.4% - 2.3%)	*11.4% (10.7% - 12.2%)*	3.3% (2.5% - 4.0%)	1.2% (0.3% - 2.1%)
Shanghai, China	9	-2.9% (-5.2% - -0.5%)	5.8% (3.7% - 7.9%)	0.6% (-1.1% - 2.3%)	11.4% (9.9% - 12.9%)	2.1% (0.8% - 3.4%)	1.5% (-0.0% - 3.0%)
Taipei, Taiwan	16	-0.9% (-2.4% - 0.6%)	-9.9% (-11.3% - -8.6%)	-8.0% (-9.2% - -6.8%)	-9.6% (-10.6% - -8.6%)	-6.6% (-8.0% - -5.1%)	-5.2% (-7.1% - -3.3%)
Tianjin, China	10	5.9% (2.6% - 9.2%)	-8.1% (-10.5% - -5.8%)	-1.4% (-3.4% - 0.6%)	*14.1% (12.4% - 15.8%)*	*-8.4% (-9.7% - -7.2%)*	7.3% (6.1% - 8.5%)
Tokyo, Japan	42	-4.5% (-6.3% - -2.7%)	1.7% (0.2% - 3.3%)	0.3% (-0.9% - 1.4%)	6.4% (5.4% - 7.4%)	3.8% (2.8% - 4.8%)	-1.4% (-2.6% - -0.3%)
Wuhan, China	9	8.6% (4.7% - 12.4%)	12.3% (9.6% - 15.0%)	4.3% (2.3% - 6.3%)	10.4% (8.7% - 12.0%)	-2.4% (-3.7% - -1.1%)	-8.2% (-9.7% - -6.7%)

Table A7: Monthly changes in O_3 concentrations relative to the bias-corrected model predictions for cities in Europe. Values in *italic* denote values that are statistically different from business-as-usual ($p<0.001$ based on Kolmogorov-Smirnov test).

Location	#	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20
Amsterdam, Netherlands	1	*21.6% (11.4% - 31.8%)*	11.4% (5.7% - 17.1%)	*22.5% (16.7% - 28.3%)*	*26.4% (20.9% - 31.8%)*	7.1% (2.5% - 11.6%)	-3.7% (-8.3% - 0.9%)
Athens, Greece	10	1.0% (-1.3% - 3.2%)	4.3% (1.9% - 6.6%)	0.0% (-1.9% - 2.0%)	4.4% (2.6% - 6.1%)	-3.9% (-5.5% - -2.2%)	-5.7% (-7.3% - -4.0%)
Belgrade, Serbia	1	-25.2% (-39.8% - -10.6%)	0.9% (-7.5% - 9.3%)	15.2% (8.8% - 21.7%)	*15.0% (10.0% - 20.0%)*	-5.5% (-10.7% - -0.3%)	-14.3% (-19.9% - -8.7%)
Berlin, Germany	9	8.2% (5.4% - 11.1%)	*7.3% (5.5% - 9.2%)*	7.4% (5.7% - 9.1%)	1.8% (0.6% - 3.1%)	-6.2% (-7.4% - -4.9%)	-9.0% (-10.2% - -7.7%)
Bratislava, Slovakia	4	-7.1% (-11.3% - -2.9%)	8.1% (5.5% - 10.7%)	6.1% (4.0% - 8.3%)	1.3% (-0.4% - 3.0%)	-3.2% (-4.9% - -1.4%)	-11.2% (-13.0% - -9.4%)
Brussels, Belgium	7	8.7% (5.6% - 11.9%)	*11.2% (9.2% - 13.3%)*	*16.7% (14.6% - 18.7%)*	*18.3% (16.5% - 20.1%)*	2.9% (1.4% - 4.4%)	-5.9% (-7.4% - -4.3%)
Bucharest, Romania	4	-12.3% (-17.0% - -7.6%)	0.4% (-3.3% - 4.0%)	-11.8% (-14.6% - -9.0%)	-9.1% (-11.4% - -6.8%)	-13.3% (-15.7% - -11.0%)	-3.6% (-6.0% - -1.2%)
Budapest, Hungary	3	18.2% (8.9% - 27.5%)	*20.1% (15.5% - 24.7%)*	*16.1% (12.1% - 20.0%)*	4.4% (1.4% - 7.5%)	-2.7% (-5.6% - 0.2%)	-13.1% (-16.0% - -10.1%)
Copenhagen, Denmark	2	4.9% (1.9% - 7.9%)	5.1% (2.8% - 7.5%)	7.5% (5.5% - 9.6%)	1.8% (-0.0% - 3.6%)	-2.4% (-4.1% - -0.6%)	-5.2% (-6.9% - -3.4%)
Dublin, Ireland	4	*5.6% (2.3% - 8.8%)*	-2.0% (-4.7% - 0.7%)	2.7% (-0.1% - 5.5%)	*8.7% (5.7% - 11.6%)*	-1.9% (-4.5% - 0.8%)	*-13.8% (-16.7% - -10.9%)*
Helsinki, Finland	2	5.5% (0.7% - 10.3%)	-0.3% (-4.7% - 4.2%)	-3.7% (-7.8% - 0.3%)	-2.2% (-5.8% - 1.4%)	3.1% (0.4% - 5.8%)	-4.7% (-7.4% - -2.0%)
Lisbon, Portugal	14	6.6% (4.8% - 8.3%)	-3.8% (-5.5% - -2.2%)	2.6% (1.5% - 3.7%)	4.8% (3.8% - 5.8%)	*-6.4% (-7.4% - -5.5%)*	-4.3% (-5.4% - -3.1%)
Ljubljana, Slovenia	2	-4.9% (-8.7% - -1.1%)	0.2% (-2.9% - 3.3%)	0.8% (-1.6% - 3.3%)	3.5% (1.4% - 5.6%)	-6.4% (-8.6% - -4.3%)	*-12.4% (-14.5% - -10.3%)*
London, United Kingdom	5	*8.6% (4.3% - 12.9%)*	*5.6% (2.8% - 8.3%)*	*14.9% (12.4% - 17.3%)*	*20.3% (17.9% - 22.7%)*	*6.7% (4.5% - 9.0%)*	-2.5% (-4.7% - -0.3%)
Luxembourg City, Luxembourg	5	0.8% (-2.6% - 4.3%)	3.4% (1.1% - 5.7%)	13.0% (10.8% - 15.3%)	10.2% (8.2% - 12.2%)	-0.2% (-2.0% - 1.5%)	-9.5% (-11.4% - -7.7%)
Madrid, Spain	26	0.7% (-1.3% - 2.7%)	-9.2% (-11.1% - -7.3%)	4.7% (3.7% - 5.7%)	6.4% (5.5% - 7.3%)	-2.2% (-3.0% - -1.5%)	-6.1% (-6.8% - -5.4%)
Milan, Italy	8	5.3% (-8.6% - 19.2%)	14.6% (7.9% - 21.3%)	*22.6% (19.1% - 26.1%)*	*25.6% (23.3% - 27.9%)*	0.7% (-1.1% - 2.6%)	-10.3% (-12.0% - -8.6%)
Paris, France	11	5.7% (2.9% - 8.5%)	*3.8% (2.1% - 5.6%)*	*14.5% (12.8% - 16.3%)*	8.3% (6.9% - 9.6%)	2.5% (1.2% - 3.7%)	-6.2% (-7.5% - -4.9%)
Prague, Czech Republic	6	-11.0% (-14.6% - -7.3%)	10.8% (8.5% - 13.0%)	9.3% (7.2% - 11.4%)	6.9% (5.2% - 8.6%)	-2.6% (-4.2% - -1.0%)	-12.6% (-14.2% - -11.1%)
Riga, Latvia	1	-1.3% (-6.5% - 3.9%)	8.1% (3.6% - 12.7%)	*11.7% (7.6% - 15.8%)*	3.7% (0.4% - 7.0%)	*32.6% (27.8% - 37.5%)*	*20.6% (14.2% - 27.0%)*
Rome, Italy	9	N/A	-5.1% (-8.4% - -1.7%)	8.2% (6.1% - 10.3%)	*15.4% (13.4% - 17.3%)*	*12.5% (10.8% - 14.3%)*	-5.8% (-7.4% - -4.3%)
Sofia, Bulgaria	4	3.7% (0.2% - 7.3%)	*11.1% (8.3% - 13.9%)*	*10.6% (8.3% - 12.9%)*	4.0% (1.6% - 6.3%)	4.7% (2.6% - 6.9%)	-3.8% (-5.7% - -1.9%)
Stockholm, Sweden	2	5.8% (2.4% - 9.1%)	5.8% (2.7% - 8.9%)	0.6% (-2.0% - 3.2%)	-3.6% (-5.9% - -1.3%)	-8.1% (-10.7% - -5.6%)	-2.2% (-4.6% - 0.1%)
Tallinn, Estonia	3	3.1% (-0.7% - 6.9%)	2.1% (-1.4% - 5.5%)	1.1% (-2.1% - 4.3%)	*0.2% (-2.6% - 3.0%)*	7.0% (3.9% - 10.2%)	-0.5% (-3.8% - 2.9%)
Vienna, Austria	11	-8.3% (-10.8% -	*8.0% (6.4% -*	5.6% (4.2% -	3.9% (2.8% -	-4.5% (-5.6% - -	-11.1% (-12.2%

		-5.8%)	*9.6%)*	6.9%)	5.0%)	3.4%)	- -10.0%)
Vilnius, Lithuania	2	6.8% (1.5% - 12.0%)	1.2% (-2.7% - 5.1%)	-3.9% (-7.1% - - 0.6%)	-2.7% (-5.2% - -0.1%)	-5.5% (-8.3% - - 2.6%)	*-20.3% (-23.1% - -17.5%)*
Warsaw, Poland	7	3.3% (-0.2% - 6.8%)	8.9% (6.5% - 11.4%)	1.7% (-0.3% - 3.7%)	-2.5% (-4.1% - -0.8%)	*-13.9% (-15.5% - -12.4%)*	*-21.3% (-22.8% - -19.8%)*
Zagreb, Croatia	2	-12.3% (-22.8% - -1.8%)	17.3% (10.9% - 23.7%)	*17.8% (13.0% - 22.6%)*	*15.3% (11.1% - 19.4%)*	7.6% (3.6% - 11.7%)	-7.6% (-11.6% - -3.5%)
Zurich, Switzerland	8	-12.6% (-16.7% - -8.6%)	2.8% (0.5% - 5.1%)	7.3% (5.1% - 9.6%)	*12.4% (10.6% - 14.2%)*	-1.2% (-2.8% - 0.5%)	-6.9% (-8.6% - - 5.2%)

Table A8: Monthly changes in O_3 concentrations relative to the bias-corrected model predictions for cities in North and South America. Values in *italic* denote values that are statistically different from business-as-usual ($p<0.001$ based on Kolmogorov-Smirnov test).

Location	**#**	**Jan-20**	**Feb-20**	**Mar-20**	**Apr-20**	**May-20**	**Jun-20**
Atlanta, USA	1	N/A	7.4% (-8.8% - 23.5%)	-3.4% (-8.4% - 1.6%)	-0.9% (-4.9% - 3.2%)	-3.9% (-8.2% - 0.4%)	-9.9% (-14.6% - -5.2%)
Boston, USA	3	6.1% (2.9% - 9.4%)	*8.4% (5.3% - 11.4%)*	6.2% (3.9% - 8.4%)	1.9% (-0.2% - 4.1%)	-4.5% (-6.7% - -2.2%)	-7.9% (-10.2% - -5.5%)
Denver, USA	4	7.3% (4.0% - 10.7%)	13.5% (10.3% - 16.7%)	-1.0% (-3.5% - 1.6%)	7.6% (5.3% - 9.9%)	6.1% (4.0% - 8.2%)	-0.6% (-2.5% - 1.4%)
Edmonton, Canada	3	0.5% (-4.2% - 5.2%)	1.0% (-2.6% - 4.5%)	-2.9% (-5.4% - -0.4%)	3.1% (1.0% - 5.3%)	-7.7% (-9.8% - -5.6%)	-5.3% (-7.9% - -2.7%)
Honolulu, USA	1	1.0% (-2.6% - 4.7%)	-6.7% (-10.0% - -3.5%)	-9.2% (-12.6% - -5.8%)	-6.4% (-10.7% - -2.1%)	*-18.0% (-22.2% - -13.9%)*	*-15.9% (-21.0% - -10.8%)*
Los Angeles, USA	7	1.0% (-1.9% - 3.9%)	10.0% (7.5% - 12.4%)	-2.5% (-4.5% - -0.6%)	-4.9% (-6.6% - -3.2%)	-5.7% (-7.3% - -4.1%)	-3.6% (-5.2% - -2.0%)
Medellin, Colombia	5	1.3% (-3.6% - 6.3%)	*-4.2% (-8.1% - -0.2%)*	*19.0% (15.0% - 23.0%)*	*26.0% (21.0% - 31.0%)*	8.5% (3.1% - 13.9%)	-14.8% (-20.2% - -9.4%)
Miami, USA	2	3.2% (-3.9% - 10.2%)	-5.8% (-10.7% - -1.0%)	1.2% (-1.8% - 4.2%)	-8.2% (-12.5% - -3.9%)	N/A	N/A
Montreal, Canada	8	5.0% (2.5% - 7.5%)	*6.0% (3.8% - 8.1%)*	3.1% (1.4% - 4.7%)	3.6% (2.0% - 5.2%)	-2.8% (-4.3% - -1.3%)	-9.2% (-10.8% - -7.6%)
New York, USA	7	6.5% (2.9% - 10.1%)	*18.5% (14.9% - 22.1%)*	*15.1% (12.8% - 17.4%)*	15.3% (13.4% - 17.3%)	6.7% (4.7% - 8.7%)	-4.4% (-6.3% - -2.5%)
Oklahoma City, USA	3	7.6% (3.0% - 12.1%)	*9.1% (6.0% - 12.3%)*	-4.0% (-6.6% - -1.5%)	-3.8% (-5.9% - -1.6%)	1.8% (-0.4% - 4.0%)	6.0% (3.8% - 8.3%)
Phoenix, USA	14	-2.5% (-4.8% - -0.3%)	3.7% (2.0% - 5.3%)	3.0% (1.6% - 4.4%)	4.6% (3.5% - 5.8%)	0.3% (-0.7% - 1.3%)	-2.0% (-3.0% - -1.0%)
Quito, Ecuador	8	-9.8% (-13.2% - -6.5%)	-7.1% (-10.1% - -4.0%)	6.3% (3.4% - 9.1%)	*37.4% (33.7% - 41.1%)*	*11.7% (8.1% - 15.2%)*	*5.9% (2.4% - 9.4%)*
Rio de Janeiro, Brazil	6	-7.2% (-11.3% - -3.2%)	-5.7% (-9.8% - -1.6%)	*4.0% (-0.2% - 8.2%)*	*23.9% (19.6% - 28.2%)*	*17.6% (13.4% - 21.9%)*	*11.3% (7.0% - 15.7%)*
Salt Lake City, USA	3	*29.6% (22.5% - 36.7%)*	*12.3% (7.4% - 17.3%)*	-1.4% (-4.5% - 1.6%)	1.1% (-1.6% - 3.7%)	3.1% (0.7% - 5.6%)	-2.0% (-4.4% - 0.5%)
San Francisco, USA	6	-0.9% (-3.5% - 1.8%)	8.5% (5.9% - 11.0%)	7.1% (5.3% - 8.9%)	-2.2% (-3.9% - -0.5%)	3.6% (1.8% - 5.5%)	-3.4% (-5.6% - -1.3%)
Santiago de Chile, Chile	3	-2.0% (-6.0% - 1.9%)	*-14.5% (-18.6% - -10.4%)*	-6.9% (-11.1% - -2.6%)	12.3% (6.9% - 17.6%)	6.7% (0.1% - 13.3%)	3.1% (-7.5% - 13.7%)
Seattle, USA	1	*8.0% (2.5% - 13.5%)*	-8.3% (-13.9% - -2.8%)	4.9% (0.0% - 9.7%)	6.7% (2.1% - 11.3%)	5.2% (0.9% - 9.5%)	-3.4% (-8.9% - 2.0%)
St. Louis, USA	3	-5.3% (-12.9% - 2.3%)	9.5% (5.2% - 13.8%)	*-13.9% (-17.0% - -10.8%)*	-7.3% (-10.0% - -4.6%)	-6.6% (-9.3% - -3.9%)	-3.5% (-6.0% - -1.0%)
Washington DC, USA	8	8.5% (5.2% - 11.9%)	*11.1% (8.7% - 13.5%)*	-2.2% (-3.8% - -0.6%)	0.2% (-1.3% - 1.7%)	-3.6% (-5.1% - -2.1%)	-1.3% (-2.8% - 0.2%)

Data availability. The model output and air quality observations used in this study are all publicly available (see methods). The output from the GEOS-CF sensitivity simulation as well as the bias-corrected model predictions are available from CAK per request.

Author contributions. CAK and MJE designed the study and conducted the main analyses. CAH and SM contributed OpenAQ observations. TO provided observations and interpretations for Japan. FCM and BBF provided observations and interpretations for Rio de Janeiro, and MVDS provided observations and interpretations for Quito. RGR provided observations for Melbourne and helped analyze results for Australia. KEK and RAL conducted the GEOS-CF simulations. KEK and CAK conducted the GEOS-CF sensitivity experiments and forecasts. LHF conducted NO_x to NO_2 sensitivity simulations. SP contributed to overall study design and context discussion. All authors contributed to the writing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Resources supporting the model simulations were provided by the NASA Center for Climate Simulation at the Goddard Space Flight Center (https://www.nccs.nasa.gov/services/discover). We thank Jenny Fisher (U. Wollongong, Australia) for helpful discussions. CAK, KEK and SP acknowledge support by the NASA Modeling, Analysis and Prediction (MAP) Program. MJE and LHF are thankful for support from the University of York's Viking, HPC facility.

References

Boersma, K.F., H.J. Eskes, R. J. Dirksen, R. J. van der A, J. P. Veefkind, P. Stammes, V. Huijnen, Q. L. Kleipool, M. Sneep, J. Claas, J. Leitao, A. Richter, Y. Zhou, and D. Brunner. An improved retrieval of tropospheric NO2 columns from the Ozone Monitoring Instrument, Atmos. Meas. Tech. 4, 1905-1928, doi:10.5194/amt-4-1905-2011, 2011.

Castellanos, P., and F. Boersma. Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession, Sci. Rep., 265, doi:10.1038/srep00265, 2012.

Chen, T. and C. Guestrin. XGBoost: A Scalable Tree Boosting System, CoRR, abs/1603.02754, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Crippa, M., Solazzo, E., Huang, G. et al. High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. Sci Data 7, 121 (2020). https://doi.org/10.1038/s41597-020-0462-2, 2020.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U., Olivier, J. G. J., Pagliari, V., and G. Janssens-Maenhout. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2, Earth Syst. Sci. Data, 10, 1987–2013, https://doi.org/10.5194/essd-10-1987-2018, 2018.

Dantas, G., Siciliano, B., França, B.B., da Silva, C.M., and G. Arbilla. The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil, Science of The Total Environment, Volume 729, https://doi.org/10.1016/j.scitotenv.2020.139085, 2020.

Darmenov, A.S., and A. da Silva. The Quick Fire Emissions Dataset (QFED)—Documentation of versions 2.1, 2.2 and 2.4. Technical Report Series on Global Modeling and Data Assimilation. NASA//TM-2015-104606, Vol. 38, 212 pp., 2015.

Djalalova, I., Delle Monache, L., and J. Wilczak. PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model, Atmospheric Environment, Volume 108, Pages 76-87, https://doi.org/10.1016/j.atmosenv.2015.02.021, 2015.

Duncan, B. N., L. N. Lamsal, A. M. Thompson, Y. Yoshida, Z. Lu, D. G. Streets, M. M. Hurwitz, and K. E. Pickering. A space-based, high-resolution view of notable changes in urban NOx pollution around the world (2005–2014). J. Geophys. Res. 121, 976–996, 2016.

Eslami, E., Salman, A.K., Choi, Y., Sayeed, A. and Y. Lops. A data ensemble approach for real time air quality forecasting using extremely randomized trees and deep neural networks, 518 Neural Comput. Appl., 1-17, 2019.

Frery, J., Habrard, A., Sebban, M., Caelen, O., and L. He-Guelton. Efficient Top Rank Optimization with Gradient Boosting for Su- pervised Anomaly Detection, Machine Learning and Knowledge Discovery in Databases, Ecml Pkdd 2017, Pt I, 10534, 20–35, https://doi.org/10.1007/978-3-319-71249-9_2, 2017.

Friedman, J. H.: Greedy function approximation: A gradient boosting machine., The Annals of Statistics, 29, 1189–1232, https://doi.org/10.1214/aos/1013203451, http://projecteuclid.org/euclid.aos/1013203451, 2001.

Grange, S. K. and D. C. Carslaw. Using meteorological normalisation to detect interventions in air quality time series, Science of The Total Environment, 653, 578–588, https://doi.org/10.1016/j.scitotenv.2018.10.344, https://linkinghub.elsevier.com/retrieve/pii/S004896971834244X, 2019.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and C. Hueglin. Random forest meteorological normalisation models for Swiss PM 10 trend analysis, Atmospheric Chemistry and Physics, 18, 6223–6239, https://doi.org/10.5194/acp-18-6223-2018, https://www.atmos-chem-phys.net/18/6223/2018/, 2018.

Hilboll, A., A. Richter, and J. P. Burrows. Long-term changes of tropospheric NO2 over megacities derived from multiple satellite instruments, Atmos. Chem. Phys., 13, 4145–4169, doi:10.5194/acp-13-4145-2013, 2013.

Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov, A. L., Travis, K. R., Grange, S. K., Evans, M. J., and D. J. Jacob. Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model Dev., 11, 4603–4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

Ivatt, P. D. and M. J. Evans. Improving the prediction of an atmospheric chemistry transport model using gradient boosted regression trees, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-753, in review, 2019.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411–11432, https://doi.org/10.5194/acp-15-11411-2015, 2015.

Jhun I, Coull BA, Zanobetti A, P. Koutrakis. The impact of nitrogen oxides concentration decreases on ozone trends in the USA. Air Qual Atmos Health. 2015;8(3):283-292. doi:10.1007/s11869-014-0279-2, 2015.

Kang, D., C. Hogrefe, K. Foley, S. Napelenok, R. Mathur, and S. T. Rao. Application of the Kolmogorov-zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model, Atmos. Environ., 80, 58-69, 2013.

Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A. M., Pawson, S., and D. J. Jacob. HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models, Geosci. Model Dev., 7, 1409–1417, https://doi.org/10.5194/gmd-7-1409-2014, 2014.

Keller, C.A., Knowland, K.E., Duncan, B.N., Liu, J., Anderson, D.C., Das, S., Lucchesi, R.A., Lundgren, E.W., Nicely, J.M., Nielsen, J.E., et al. Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, Earth and Space Science Open Archive, 38, https://doi.org/10.1002/essoar.10505287.1, 2020.

Kleinert, F., Leufen, L. H., and Schultz, M. G.: IntelliO3-ts v1.0: A neural network approach to predict near-surface ozone concentrations in Germany, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-169, in review, 2020.

Knowland, K.E., C.A. Keller, and R. Lucchesi, File Specification for GEOS-CF Products. GMAO Office Note No. 17 (Version 1.1), 37 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes, 2020.

Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P., and M. J. Newchurch. Application of satellite observations for timely updates to global anthropogenic NOx emission inventories, Geophys. Res. Lett., 38, L05810, doi:10.1029/2010GL046476, 2011.

Lamsal, L. N., R. V. Martin, D. D. Parrish, and N. A. Krotkov. Scaling relationship for NO2 pollution and urban population size: A satellite perspective, Environ. Sci. Technol., 47(14), 7855–7861, doi:10.1021/es400744g, 2013.

Le, T., Wang, Y., Liu, L., Yang, J., Yung, Y.L., Li, G., and J.H. Seinfeld. Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China, Science, doi:10.1126/science.abb7431, 2020.

Le Quéré, C., Jackson, R.B., Jones, M.W. et al. Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement. Nat. Clim. Chang., https://doi.org/10.1038/s41558-020-0797-x, 2020.

Liu F., et al. Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19, Science Advances, DOI: 10.1126/sciadv.abc2992, 2020a.

Liu, Z. et al. COVID-19 causes record decline in global CO2 emissions. http://arxiv.org/abs/2004.13614, 2020b.

Long, M. S., Yantosca, R., Nielsen, J. E., Keller, C. A., da Silva, A., Sulprizio, M. P., Pawson, S., and D.J. Jacob. Development of a grid-independent GEOS-Chem chemical transport model (v9-02) as an atmospheric chemistry module for Earth system models, Geosci. Model Dev., 8, 595–602, https://doi.org/10.5194/gmd-8-595-2015, 2015.

Lucchesi, R., File Specification for GEOS-5 FP-IT. GMAO Office Note No. 2 (Version 1.4) 58 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes.php, 2015.

Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30, 4768–4777, 2017.

Lundberg, S.M., Erion, G., Chen, H. et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Ministry of the Environment (MOE), Government of Japan. The Atmospheric Environmental Regional Observation System (AEROS), data available at http://soramame.taiki.go.jp/Index.php (last access: July 3, 2020), 2020. (in Japanese)

Pagowski, M., Grell, G.A., Devenyi, D., Peckham, S.E., McKeen, S.A., Gong, W., Delle Monache, L., McHenry, J.N., McQueen, J., and P. Lee. Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts, Atmospheric Environment, Volume 40, Issue 18, 3240-3250, https://doi.org/10.1016/j.atmosenv.2006.02.006, 2006.

Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., and Pérez García-Pando, C.: Meteorology-normalized impact of the COVID-19 lockdown upon NO2 pollution in Spain, Atmos. Chem. Phys., 20, 11119–11141, https://doi.org/10.5194/acp-20-11119-2020, 2020.

Russell, A. R., L. C. Valin, and 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(24), 12,197–12,209, 2012.

Shah, V., Jacob, D. J., Li, K., Silvern, R. F., Zhai, S., Liu, M., Lin, J., and Zhang, Q. (2020). Effect of changing NOx lifetime on the seasonality and long-term trends of satellite-observed tropospheric NO2 columns over China, *Atmos. Chem. Phys.,* 20, 1483–1495, https://doi.org/10.5194/acp-20-1483-2020.

Seinfeld, J.H. and S.N. Pandis. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. John Wiley & Sons, Hoboken, 2016.

Streets, D.G., Canty T., Carmichael, G.R. , de Foy, B. Dickerson, R.R, Duncan, B.N., Edwards, D.P. , Haynes, J.A., Henze, D.K., Houyoux, M.R., Jacob, D.J., Krotkov, N.A., Lamsal, L.N., Liu, Y., Lu, Z., Martin, R.V., Pfister, G.G., Pinder, R.W., Salawitch, R.J., and K.J. Wecht, Emissions estimation from satellite retrievals: A review of current capability, Atmospheric Environment, Volume 77, Pages 1011-1042, https://doi.org/10.1016/j.atmosenv.2013.05.051, 2013.

Tobías, A., Carnerero, C., Reche, C., Massagué, J., Via, M., Minguillón, M. C., Alastuey, A., and X. Querol. Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-CoV-2 epidemic, Science of The Total Environment, 726, 138 540, 540 https://doi.org/10.1016/j.scitotenv.2020.138540, 2020.

Wargan, K., Pawson, S., Olsen, M.A., Witte, J.C., Douglass, A.R., Ziemke, J.R., Strahan, S.E., and J. E. Nielsen. The Global Structure of Upper Troposphere-Lower Stratosphere Ozone in GEOS-5: A Multi-Year Assimilation of EOS Aura Data. J. Geophys. Res. - Atmos, 120, 2013-2036. doi: 10.1002/2014JD022493, 2015.