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)