This is a response to reviewer 1, who provided some really useful feedback, while we wait for comments from reviewer 2. Below we include a point-by-point response (in blue) to reviewer 1 comments, and describe the corresponding revisions to the manuscript (in red).

Reviewer #1

This paper promises to "critically assess the efficacy of these [NOx/CO2, CO/CO2] ratios to isolate the combustion contribution of CO2". This is an interesting and timely goal. After reading the paper, which contains high-quality figures and is written fluently, I can only conclude that paper mostly fails to reach this goal.

Methodologically, the paper lacks proper hypotheses about what to expect, and has a substantial number of flaws. The GEOS-Chem model does a poor job to simulate NO2 columns over major CO2 emitting areas within Europe. This makes the analysis of the paper rather speculative. I attach an annotated pdf with comments, and below I summarize the main points that need to be addressed before it can be considered for publication.

We appreciate the reviewer's critical and constructive comments. With respect, we refute the criticism of the paper lacking a hypothesis: we want to evaluate the efficacy of using $NO_x:CO_2$ and $CO:CO_2$ ratios to quantify fossil fuel CO_2. These ratios are often used without understanding how useful they are or whether they produce consistent results. Here we lay out in detail the differences in these ratios that highlight their strengths and weaknesses.

We have now addressed the major concerns by 1) extending the study period to include two additional (transitional) months: September 2018 and April 2019; 2) including a model evaluation against in-situ measurements (AirBase, following Visser et al. (2019)); and 3) updating the calculation of column concentration. These updates provide more confidence in the model performance against measurements and in general the chain of analysis. More in-depth address to each specific concern is detailed below.

1, Formaldehyde and December introduced but not presented.

In the introduction, the authors introduce formaldehyde as a species that will be studied: "we explore the agreement between model and observed ratios of NO_2, CO, and HCHO with CO_2."

They write: "Formaldehyde (HCHO) is another proxy for incomplete combustion (e.g., Fu et al. (2007a); Gonzi et al. (2011))". However, later HCHO is discarded: "Given the limited use of HCHO as a tracer of combustion we do not pursue this tracer any further." Why include this tracer anyhow? I understand that the authors worked on HCHO before (they refer to their papers), but the added value of showing HCHO in this paper is limited, because this paper links emission ratios to modelled atmospheric ratios. Since HCHO does not have substantial emissions, I would recommend the authors to leave this tracer out altogether. As a side note, one could think of a way to use HCHO to quantify the interfering contribution of VOCs to the CO budget (an issue that is left largely open). The authors write on page 9: "Using GEOS-Chem, we find that the secondary source is typically 10–20% of the total CO source in winter months but in July can be as much as 75% of the total CO source over Europe. This secondary source will therefore need to be considered if CO is to be used to isolate combustion CO2.". One reason to leave HCHO in, would be an analysis of how HCHO could be used to quantify the indirect CO source.

Formaldehyde was initially selected as one of the candidates to help identify hotspots of incomplete fossil fuel combustion. While we eventually decided not to use this compound in the reported analysis, we thought the responsible approach was to highlight our reasons to deter other other researchers from repeating our aborted analysis. During winter months, the speed at which HCHO is produced from primary emissions (direct HCHO source is comparatively small) via atmospheric chemistry is slow, making it difficult to relate elevated HCHO values to emission sources. During summertime, we find that the secondary source of HCHO from biogenic VOCs precludes HCHO from being used as an effective tracer of fossil fuel combustion.

We discuss our reasoning on Page 3 Line 66 (and P6 L171). Even though it is a negative result, we think it is worthwhile to present it in the manuscript. We agree with Reviewer 1 that there is limited added value in reporting results on HCHO, so we have relegated that analysis to the Appendix.

On a similar note, satellite data are not useable in December 2018, because there are insufficient valid OCO-2 observations. The presentation of the December emissions only does not supply much added value. In fact, the analysis is based on only one month of model simulation and satellite data, which is rather limited.

December was selected as the contrasting month to July. Originally intended to emphasize the differences in emissions and atmospheric columns due to seasonal photochemistry. Instead, we showed that satellite observations are limited, and data quality is low during this period (Figure A5, A6 in the updated manuscript). We agree with Reviewer 1 that December data did not provide substantial weight to our results. To address this point, we have extended our study period to include two transitional months September 2018 and April 2019. We have included satellite measurements and the corresponding model simulation results from these two months in our revised manuscript.

2, CO_2 modelling.

Concerning the modelling of CO_2, the description is very brief. It seems that the authors do not include the exchange with the biosphere (or oceans). Yet, in figure 5 they compare OCO-2 with the model. Given the large role of the biosphere in July, I would argue that this analysis is therefore flawed.

This is an oversight for which we apologize. CO_2 fluxes in both the global simulation, that determine lateral boundary conditions, and the nested grid simulation are documented extensively in other papers. Both simulations include fossil fuel emissions, pyrogenic emissions, terrestrial exchange, and ocean fluxes. We have now added a more comprehensive description of the CO_2 fluxes used in our model simulation.

3, Sharp gradients in the emission gradients.

The description of the emission ratios applied in the model reveals sharp boundaries between the countries (also in Super et al. 2020). This can partly be understood ("The differences between countries for the two months reflect the relative importance of individual sectors"), but the authors should at least mention that there are still significant uncertainties in reporting of national scale emissions, their allocation, and downscaling. It is hard to believe that emission ratios suddenly change when you cross a border.

We appreciate the reviewer's comments. Sharp gradients and boundaries between the countries are present in emission inventory. We agree that there is significant uncertainty in reporting of national scale emissions, estimation of sector emission using proxies, and downscaling. We have now included this discussion in the revised manuscript.

4, Strategy and results.

The authors mention a high spatial correlation between GEOS-Chem and TROPOMI (R=0.60 for NO2 and R=0.82 for CO). However, this spatial correlation in µmole/m2 includes a spurious correlation driven by orography in the domain. I suggest to recalculate the correlation in mixing ratio units.

Thanks for the reviewer's suggestion. We think the reviewer is referring to Figure 5, where spatial correlations between satellite and model for CO and NO_2 with the unit of mol m^{-2} are calculated and discussed in P8 L247. We have now updated the correlation coefficients using mol/mol as unit:

P8 L247:
"The model generally reproduces well the observed monthly spatial variations for CO_2 (R=0.67, calculated in unit mol/mol), CO (R=0.48) and NO_2 (R = 0.59) across Europe, but …"

The follow-up question is then of course: how good is this? Visually, it does not look great. The fact that important hotspots, like Madrid, Paris, and London, are misrepresented by the model does not give large credibility to the simulations. Has this model version been validated against surface observations? How can I be sure that the chemistry is not totally flawed over urban environments?

Simply giving reasons is not enough ("...likely reflect over-reporting of NOx emissions from rural areas of France, Germany, Poland and other eastern European countries, i.e., errors in emission inventory, temporal profile and errors in vertical mixing and lifetime of NOx against chemical oxidation.", …" which is due to some combination of underestimating emissions from these large urban sources, errors in the model description of NOx photochemistry, and the low sensitivity of averaging kernels to lower levels of the atmosphere".). Basically, we are dealing with largely unvalidated model results.

We have now evaluated our model simulation for July/September 2018 and for April 2019 using AirBase in-situ measurements. Figure 1 (below) presents the comparison of GEOS-Chem simulated monthly average surface NO_2 and sampled at UTC 13:00-14:00 against AirBase (EEA, 2018) rural background stations. Performance statistics are R = 0.48, RMSE = 5.54 and R = 0.27, RMSE = 5.75 for monthly and UTC 13:00-14:00, respectively.

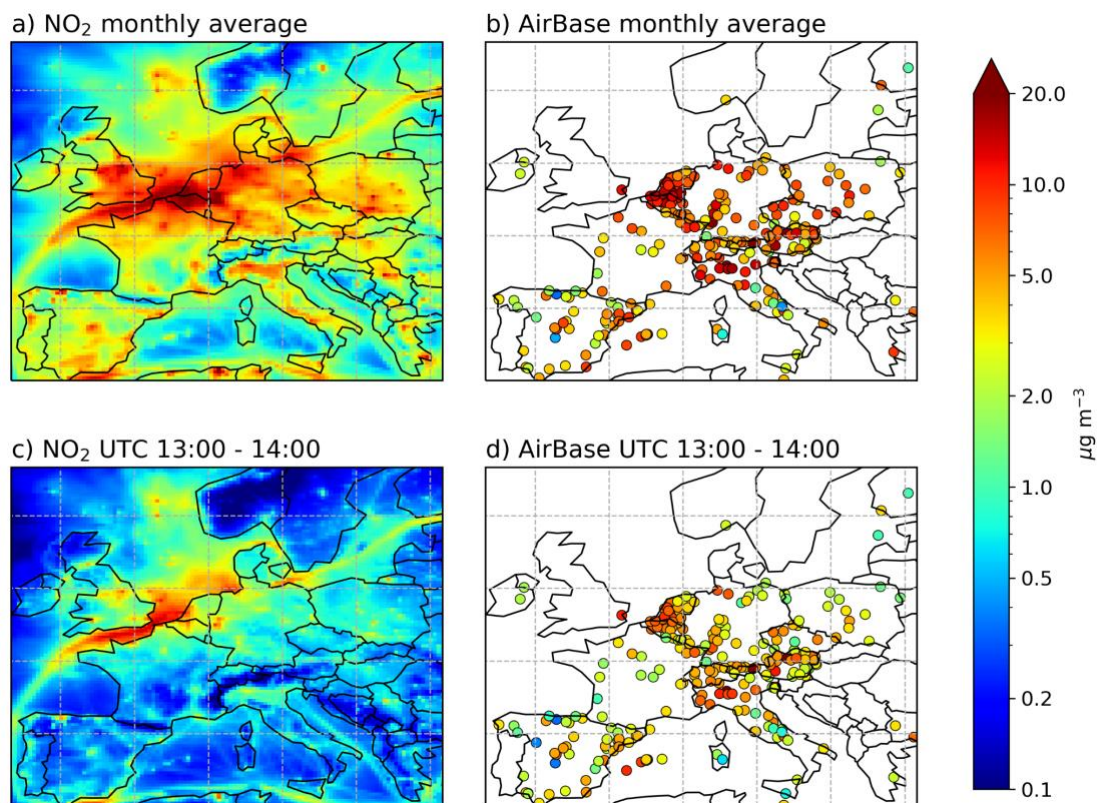
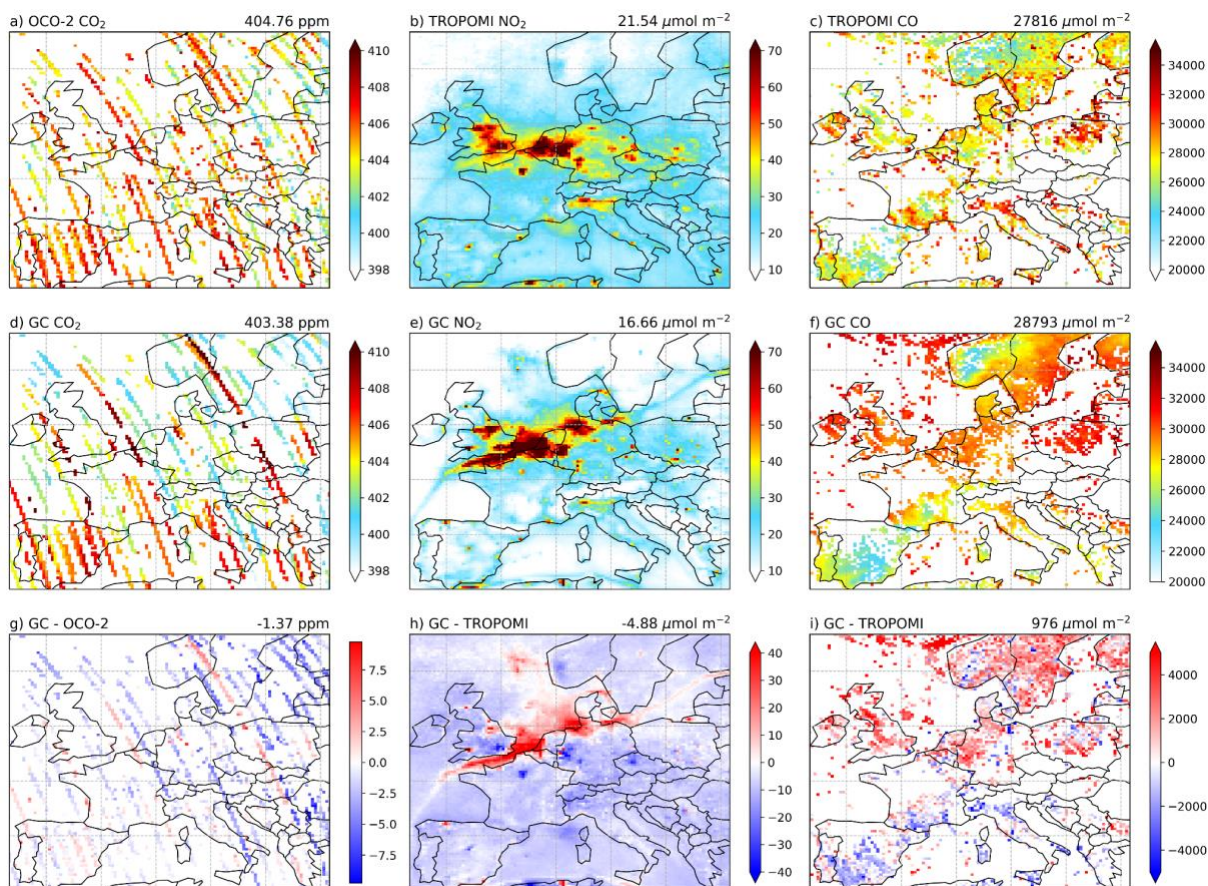



Figure 1. Monthly mean surface NO_2 and simulated by GEOS-Chem (a, c) and observed at AirBase stations (b, d). Panels (a) and (b) are monthly averages, and panels (c) and (d) are sampled at UTC 13:00 - 14:00.

Following the suggestion from Reviewer 1, we have also updated the calculation of column concentrations by replacing the a priori TM5-MP NO_2 and CO profiles with GEOS-Chem profiles to remove inconsistencies in the model-satellite comparison. Figure 2 (below) is the updated Figure 5 in the manuscript. The GEOS-Chem model does indeed capture most of the important urban hotspots, including Madrid, Paris, and London. Model performance in terms of correlation coefficient and model bias has improved significantly. We have also revised the discussion to reflect the changes we have made in our analysis.

Updated Figure 5. Satellite measurements and GEOS-Chem model columns of CO_2 (ppm), CO (μ mol m^{-2}), NO (μ mol m^{-2}), described on a 0.25° latitude × 0.3125° longitude resolution. The top row shows observed distributions from OCO-2 and TROPOMI, the middle row shows the corresponding GEOS-Chem distributions, and the bottom row shows GEOS-Chem minus observed distributions. Domain-mean values and units are shown in the titles of each panel.

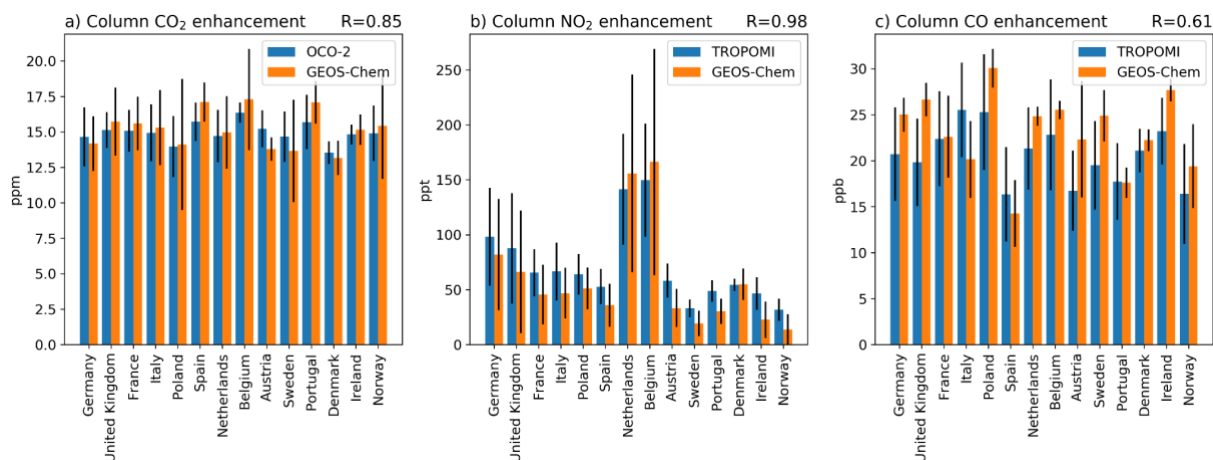
Nevertheless, the authors proceed ("….provide us with some confidence in our ability to use enhancement ratios of column CO and NO_2"). After some manipulation (subtracting background from model and satellite data) they present in Figure 6 regional enhancements on the national scale. The good correlation they find for CO2 and CO on country level is reassuring, although it remains unclear why they moved to the country scale.

We have now updated both grid-cell level and national level correlation analysis in the manuscript. This does not change our main conclusions.

However, for NO_2 the modeled columns are too small in the Netherlands, Belgium, and Germany, and too high for Ireland. This calls at least for some interpretation. Likely, the smearing of the high NOx emissions leads to a too-high OH regime. Also, one should question the use of the standard TROPOMI averaging kernel (from the coarse-scale TM5 model) for the higher-resolution simulations (see abundant literature on this subject).

We appreciate the reviewer's comments. As mentioned above, we have now updated the computation of column concentration by replacing the a priori TM5-MP NO_2 and CO profile with GEOS-Chem profiles to remove inconsistencies in the model-satellite comparison. This led to significant improvement in model performance.

The authors now take the next step in figure 7, in which they show the ratio of the NO2 and CO enhancement relative to the CO2 enhancements on the grid-scale (back to the grid scale?). My problem with this procedure is that the spatial variations in NO2 (typically 25%, from figure 6) and CO (typically 15%) are dominating the CO2 signal (variations < 10%). A more serious problem is that the errors associated with the modeled NO2 variations are larger than the signal in CO2, making the analysis that follows flawed in my opinion.

While we do not disagree with this Reviewer about the dominance of the reactive trace gases, the analysis of these ratios is the point of the study so it is instead a strength of the analysis. Model errors in the NO_2 distribution are now comparable with previous model studies, with national-scale correlations R = 0.98 and differences on the national scale typically < 20 ppt (Figure 6, below). Our study highlights the need for careful analysis of these ratios if they are to be used effectively to infer $ffCO_2$. We have addressed this point in the revised manuscript by acknowledging the point about the role of model error and the importance of reactive gases in the interpretation of the ratios.

Updated Figure 6. Observed and GEOS-Chem monthly mean and standard deviation of (a) ΔXCO_2 (ppm), and (b) ΔNO_2 (ppt) and (c) ΔCO columns (ppb) during July 2018 for the top 14 CO_2 emitting countries, in descending order. Correlation coefficients from linear regression are shown on subpanel title.

In the next step, the authors proceed one step further, by correlating (figure 8a) the combustion ratios NOx/CO2 with CO/CO2. There is, however, not a prior hypothesis why there should be a correlation. Do countries with a large proportion of sectors with high CO/CO2 emission ratios, also have high temperature combustion (with high NOx/CO2 ratios? Apparently not, because there is too much scatter to draw a correlation line. Figure 8a does not bring relevant information here, and would belong to section 3.1. Concerning the unit of the slope: I think it is not unitless (kg NOx/kg CO, and a factor 10 somewhere).

The introduction outlines the importance of CO and NO_x as tracers of combustion that are often used in one form or another to infer CO_2 emissions or interpret CO_2 observations. There appears

to be a misconception that using NO_x/CO_2 or CO/CO_2 will lead to the same results. We show that it isn't true. Even on a national scale, spatial variations in the two inventory-based ratios are not correlated, reflecting differences in the magnitude and distribution of CO and NO_x emissions. However, the two corresponding atmospheric ratios are much more correlated suggesting a role for atmospheric physics (mixing) and chemistry from low- and high- emitting regions, which could lead to an overly optimistic assessment of the equivalency of these two ratios to infer fossil fuel emissions. We now clarify these points in the revised manuscript.

Figure 8b does bring interesting information. But, if you divide both the x-axis and the y-axis by the same number (ΔCO2), would it not be more straightforward to simply study the ΔNOx–ΔCO relation? Earlier identified model flaws in correctly modelling TROPOMI NO2 mostly explain the difference in slope between model and satellites. The added value of dividing by the CO2 enhancement is not clear, nor quantified. The authors write: "This suggests both observed and model columns reproduce the positive relationship based on the inventory estimates.", and basically confirm that CO and NOx emissions are correlated both in the inventory and in the modeled columns (since CO2 divides out).

The purpose of Figure 8 was to compare the behaviour of the two ratios across European countries. Since we find a stronger relationship between the two atmospheric ratios than for the same ratios inferred from the prior emissions it suggests that atmospheric physics (mixing) and chemistry result in an overestimation of the ability of the atmospheric ratios to inform on combustion emissions. We find a much wider (relative) range for $\Delta NO_2 : \Delta CO_2$ than $\Delta CO : \Delta CO_2$. Given that all three atmospheric tracers are subject to the same meteorology and since CO_2 is not subject to atmospheric chemistry, the differences must be due to a greater importance of atmospheric chemistry for ΔNO_2, as expected.

As an aside, while it is tempting to simply cancel the ΔCO_2 in the figure this cannot be done for the following reason: there are differences in the magnitude and distribution of ΔNO_x and ΔCO emission and this will be reflected in the national mean values for each ratio. If we simply plotted ΔNO_x against ΔCO that would defeat the purpose of our analysis - researchers commonly use $\Delta NO_2:\Delta CO_2$ rather than $\Delta CO:\Delta CO_2$ so we are investigating their validity to infer fossil fuel emissions.

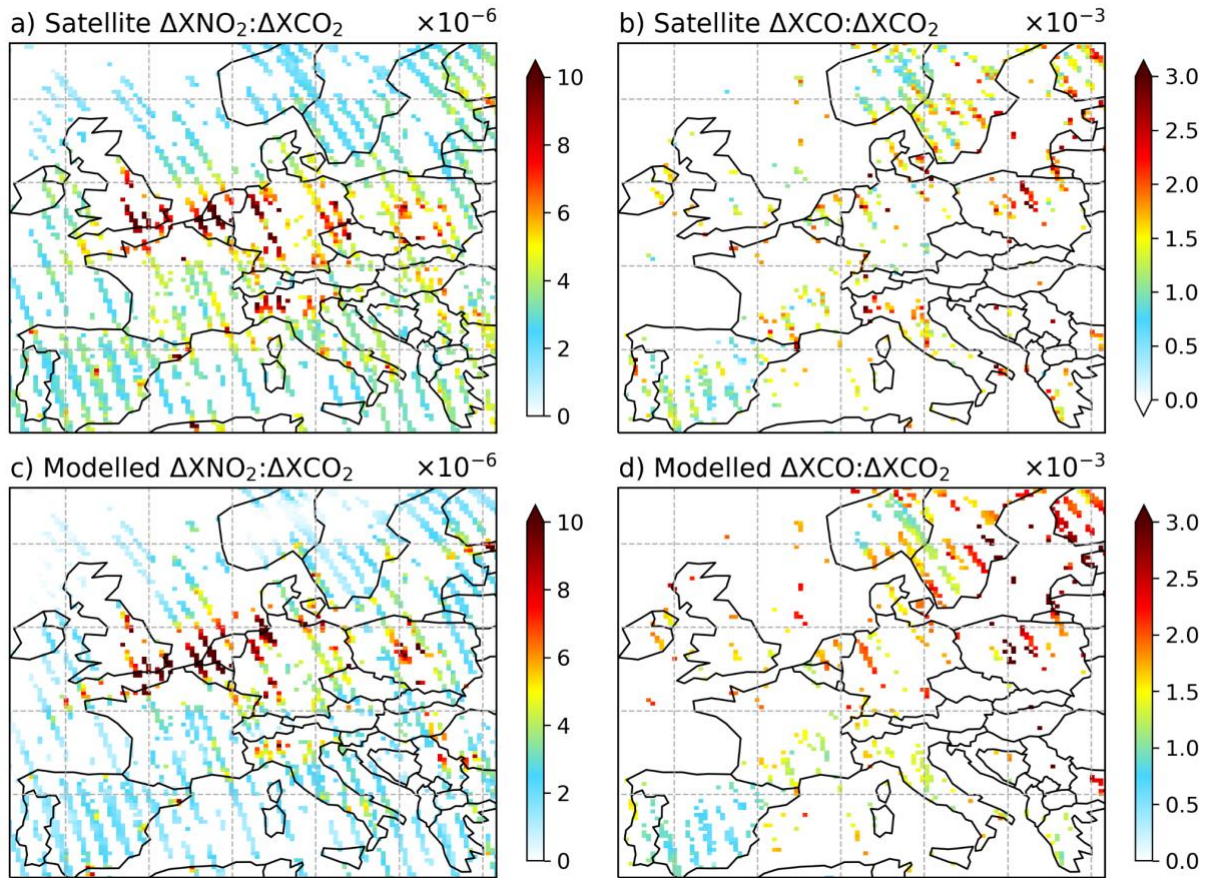
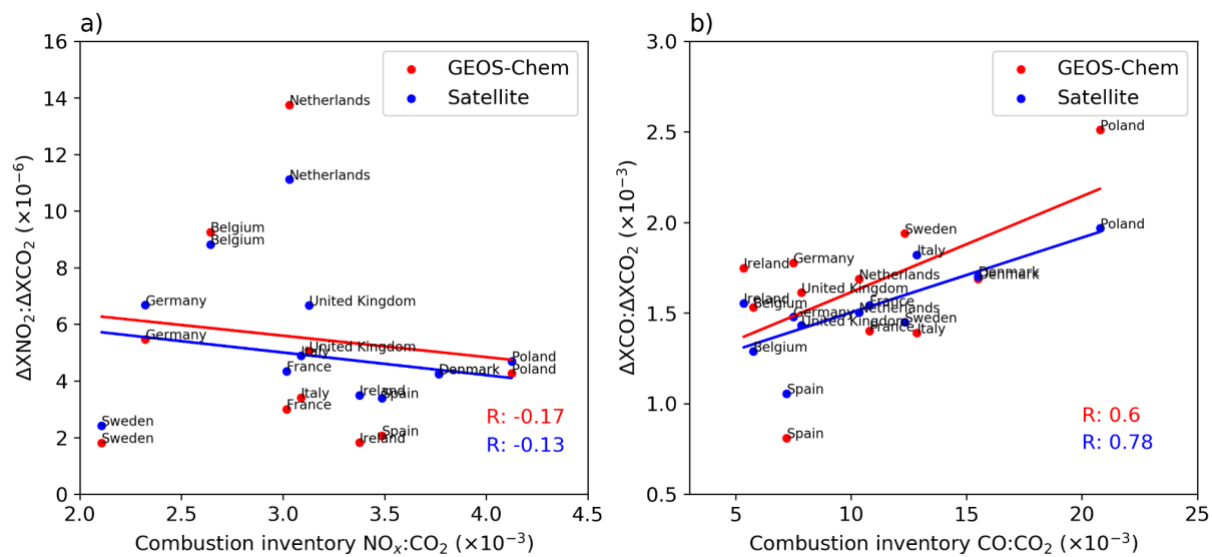
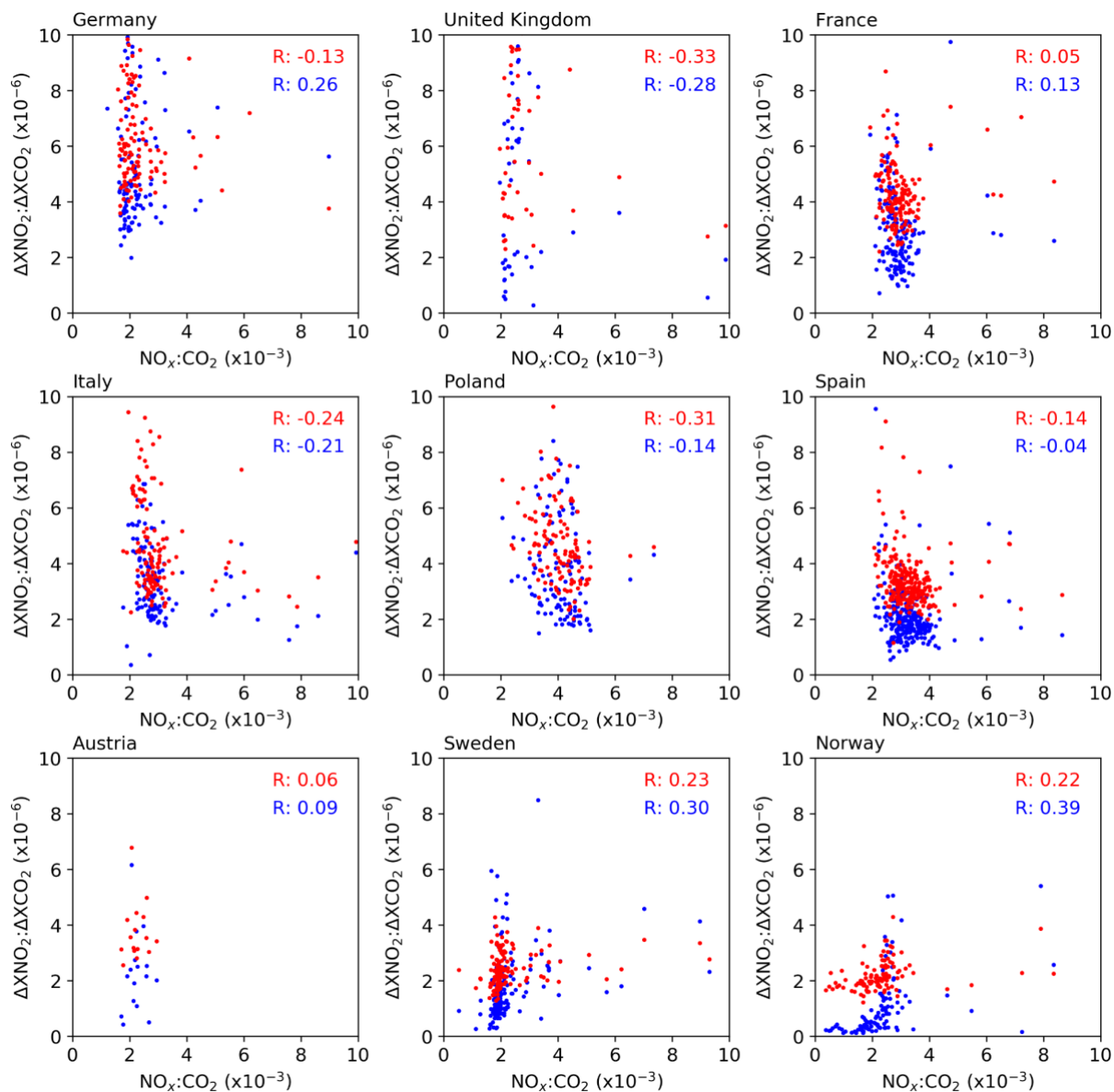
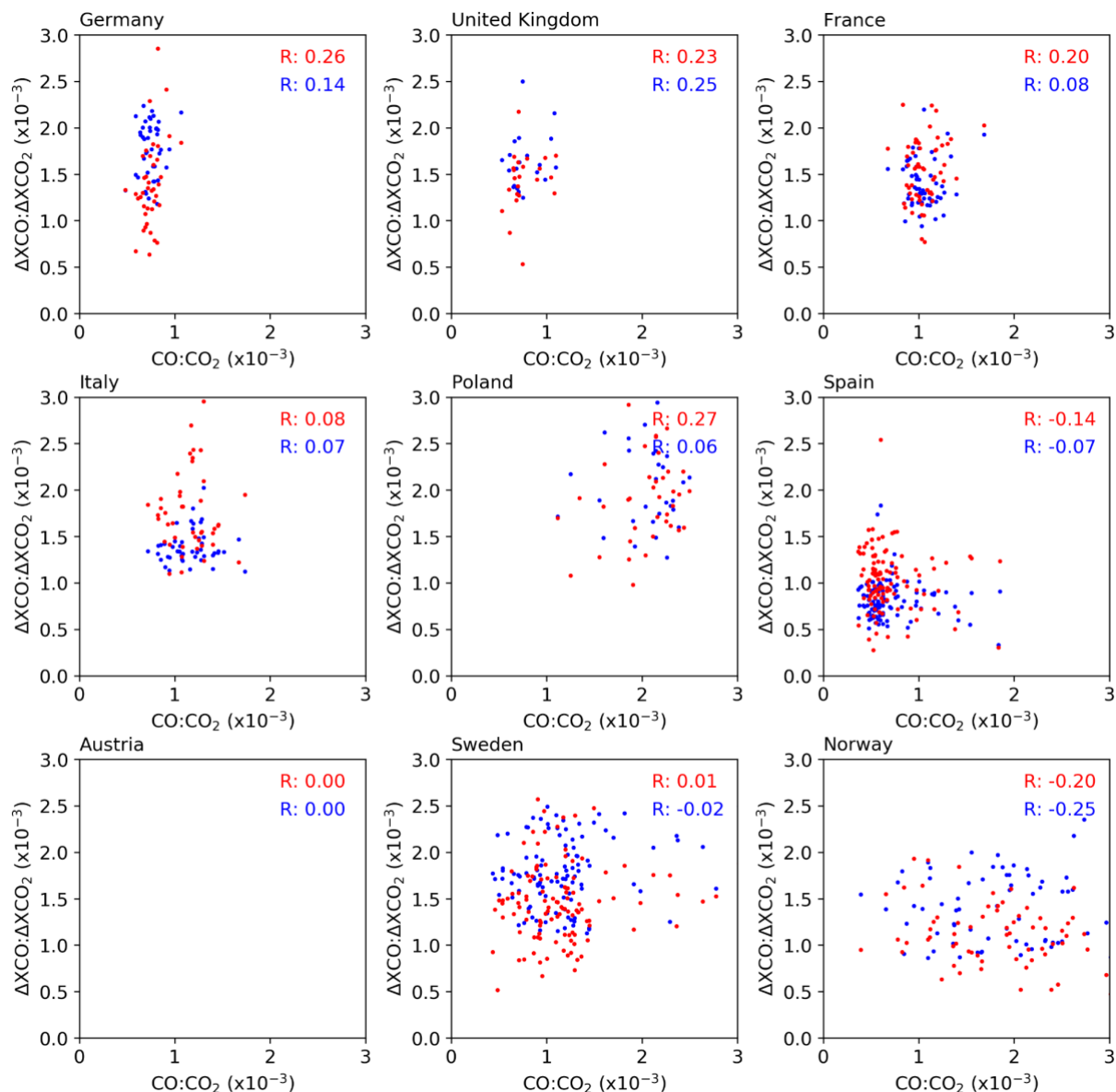
These points are now clarified in the revised manuscript.

The final figure (figure 9) correlates the country scale emission ratios to model and satellite ratios. Interestingly, for NO2, a negative correlation is observed. This would imply that countries with high NOx emissions (compared to CO2) show smaller NO2 columns (compared to CO2), which seems rather surprising, and warrants further interpretation. The authors interpret: "the negative relationships between the inventory-based and atmospheric-based ratios reflect a strong non-linearity between NOx emissions and NO2 concentration." In trying to understand this issue better, I went back to figure 4, noting from figure 9 that Norway (2.7x10-3 mole NOx/mole CO2), Spain (2.4x10-3), and Portugal (2.8x10-3 mole NOx/mole CO2) show the largest emission ratios, and that Germany (1.3x10-3 mole NOx/mole CO2) has the lowest ratio in figure 9. Indeed, in figure 4 Portugal has the highest emission ratio, but values and order are distinctly different from figure 9 (Poland is for some reason not shown). Do I interpret the numbers wrong? Anyhow, this issue requires some further scrutiny.

Both country-scale and grid-scale correlation analysis has been updated with the updated column calculations and additional two months of model simulations. Here, we present updated Figure 7,

9, A7 and A8. With updated column calculations and stricter filtering of TROPOMI CO measurements, we noticed a weaker negative correlation between national level column and inventory ratios for NO_x, and much stronger positive correlation for CO. However, the strong negative correlation for NO_x ratios is still observed in grid-cell level regression for the UK, Italy and Poland, which reflects a strong nonlinearity between NO_x emissions and NO_2 concentration.

Poland is now included in updated Figure 9, and the ratios presented in Figure 4 and 9 are now consistent, except for the different scaling used for inventory-based $CO:CO_2$ ratios.

We have revised the results and discussion to reflect the changes in the updated results. We will highlight the importance of the column calculation (particularly NO_x) and the data filtering (particularly CO) when interpreting the two observed ratios.

Updated Figure 7. Observed and GEOS-Chem monthly mean European distributions of (left) $\Delta XNO_2:\Delta XCO_2$ and (right) $\Delta XCO:\Delta XCO_2$ ratios during July 2018, described on 0.25° latitude × 0.3125° longitude resolution.

Updated Figure 9. European country relationships of inventory estimates of combustion NO_x:CO_2 and CO:CO_2 and observed and GEOS-Chem atmospheric ratios of ΔXNO_2:ΔXCO_2 and ΔXCO:ΔXCO_2. Correlation coefficients and slopes from linear regression are shown inset of each panel.

Updated Figure A7. Relationship between combustion emission $NO_x:CO_2$ and satellite measured $\Delta XNO_2:\Delta XCO_2$ (blue) and GEOS-Chem simulated $\Delta XNO_2:\Delta XCO_2$ (red) for nine European countries. Correlation coefficient value of simple linear regression analysis is shown on the sub- panel figure.

Updated Figure A8. Relationship between combustion emission CO:CO₂ and satellite measured ΔXCO:ΔXCO₂ (blue) and GEOS-Chem simulated ΔXCO:ΔXCO₂ (red) for nine European countries. Correlation coefficient value of simple linear regression analysis is shown on the sub- panel figure.

In conclusion, although the subject is interesting and timely, the paper needs substantial revision, scrutiny, and would benefit from a complete redesign. A thorough model validation, including comparisons to surface observations would be needed to gain trust in the model. The severe underestimate of NO2 over major cities in Europe precludes the analysis the authors nevertheless present. Also, I would recommend the authors to improve the comparison to TROPOMI NO2, by recalculating the averaging kernel based on the high-resolution NO2 distribution (see literature).

We have followed the suggestions by the reviewer and substantially revised the manuscript, which addresses the major concerns raised by this reviewer. By extending the study period, including

model evaluation against in-situ measurements and updating the computation of column concentration, model performance has improved significantly. The updated manuscript also provided more confidence in the chain of analysis to investigate the influence of combustion on atmospheric CO_2 over Europe.

References:

Visser, A.J., Boersma, K.F., Ganzeveld, L.N. and Krol, M.C., 2019. European NOx emissions in WRF-Chem derived from OMI: impacts on summertime surface ozone. Atmospheric Chemistry and Physics, 19(18), pp.11821-11841.

Air Quality e-Reporting [Data Set], available at: https:// www.eea.europa.eu/data-and-maps/data/aqereporting-8 (last access: October 2021), 2018.