We thank both Reviewers for their insightful comments. Our point-by-point response is included below. Changed text is in red, and all line numbers refer to the track changes version.

Reviewer #1

The authors have adequately addressed all of my concerns, except for my comment on the potential effect of nitrate dry deposition on their analysis and conclusions. In their response to my original comment, they cited Dutta and Heald (2023) to support their working hypothesis regarding the relatively minor biases in their results associated with capturing anthropogenic trends on a seasonal and annual scale, as well as their geographic distribution. However, I find the provided reference unsuitable for supporting their claim about the similarity in seasonal variations, annual trends, and geographic distributions between dry and wet nitrate deposition. For instance, Figure 5 in Dutta and Heald (2023) indicates quite large temporal and spatial differences between dry and wet nitrate deposition.

Reference

Dutta, I. and Heald, C. L.: Exploring Deposition Observations of Oxidized Sulfur and Nitrogen as a Constraint on Emissions in the United States, JGR Atmospheres, 128, e2023JD039610, https://doi.org/10.1029/2023JD039610, 2023.

We should have more fully explained our reasoning, as there are different considerations with dry deposition when we are discussing magnitudes versus trends. Dutta and Heald (2023) find similar trends between NWD and dry deposition on an annual basis, with some differences that do not change the overall conclusion (i.e., dry deposition trends are slightly steeper than wet deposition due to capturing urban emissions better). On an annual averaged regional basis, as we use in this analysis, trends in wet deposition are better at capturing total NO_x emissions that include background sources, and thus our national and annual trend analysis is not greatly affected by excluding dry deposition.

However, the Reviewer is correct that magnitudes will be more uncertain in ways that are difficult to assess using current models and understanding, and we did not address this adequately using the Dutta and Heald (2023) reference. There are some geographical differences between dry and wet deposition that may impact our analysis of NO_x overestimates. In general, deposition for both is greater in the eastern US than in the western US. Dry deposition is more prominent in the southeast US, and wet deposition is most prominent in the northeast US. It is possible that this would bias our results to miss some nitrate deposition in the southeast US by just using NWD. An impact of not including dry deposition in our analysis is that discrepancies in parts of the country where dry deposition is more important are more difficult to assess with just wet deposition data. Here, that may mean that potential biases in NO_x emissions in this region would be more apparent in the dry deposition data. This is also true for the Mountain West, where we see evidence for underestimates in NO_x emissions using wet deposition, but dry deposition is more prominent in that region. In contrast, the most prominent wet deposition during summer occurs over the rest of the eastern US, where we see high biases consistent with previous analyses of NO_x estimates. These biases may be easier to see using wet deposition data since nitrate is predominantly deposited in the northeast through precipitation. However, the impact this has on our analysis is difficult to assess quantitatively due to the uncertainties in both dry deposition estimates (also see the communication with Reviewer #2 below) and model predictions of dry deposition in GC, which are known to be biased high. Dry deposition estimates and modeling should be improved to better resolve potential NO_x emissions discrepancies in places where dry deposition dominates. This adds uncertainty to our results, which we now discuss more thoroughly.

We edit the paragraph beginning Line 175 (after Figure 2) to include an acknowledgment of the uncertainty in magnitudes:

"Excluding dry deposition may lead to some bias in capturing anthropogenic trends, as dry deposition tends to be more influenced by urban sources, but natural sources are important for determining regional trends, and these are captured better by wet deposition (Dutta and Heald, 2023). Magnitudes may be more uncertain, as the geographic distribution between wet and dry deposition differs. For example, over the CONUS, wet deposition is most prominent in the northeast US, and dry deposition is most prominent in the southeast US. Thus, dry deposition may be more adept at capturing biases in NO_x emissions in the southeast US, which may not be as apparent in the wet deposition data. This introduces uncertainty to our regional analysis of NO_x magnitudes using just NWD (Section 3.3); this uncertainty is currently difficult to quantify given observational limitations in dry deposition and known model biases in GEOS-Chem."

Reviewer #2

I appreciate the authors' and editor's consideration of my feedback and the authors' constructive effort to respond. I feel that the paper is overall technically sound, and include below a series of minor suggestions that I hope might help to strengthen the paper further.

• Fig. 2 – Consider moving up in the text to where it is first referenced?

We have moved Figure 2 up to Line 170, right below the paragraph where we introduce site locations.

• Fig. 3 – Just flagging that at a quick glance, Fig 3a seems to undermine the conclusion that NOx emissions are underestimated, until the reader looks further to see the offsetting regional differences in biases in Fig 6a.

In our paper, we conclude that NO_x emissions are overestimated in some regions and seasons over the CONUS. The Reviewer here mentions underestimates, which we mention briefly later on in Section 3.3 when discussing winter and spring in certain regions. To avoid confusion, we add a sentence in Section 3.2 to clarify that, on average, model NWD agrees well with observations, but some regions during certain seasons do not agree. We do this in Section 3.2 since Section 3.1 is focused solely on observational trends (paragraph beginning Line 379).

"Trends and magnitudes in NWD are well-reproduced by GC over the CONUS (Fig. 3). During the largest decrease, 2000-2010, GC shows a significantly ($p<0.05$) decreasing trend of -3.4 ± 0.8%/yr, which agrees with significant observed decreases in NWD of -4.1 ± 1.2%/yr. While annual NWD values predicted by GC agree with observations when taken over the entire CONUS domain, regional NWD predictions do not agree in certain regions and seasons. This issue is explored further in Section 3.3."

• L145 – The citation for NADP is not consistent with their use conditions. https://nadp.slh.wisc.edu/data-and-information-use-conditions/

Thank you for bringing this important issue to our attention. We have fixed this to be consistent with the NADP use conditions.

• L171 – I appreciate the authors' consideration of my feedback regarding this paragraph, which has improved substantially since the first draft. But I disagree with the characterization of CASTNET dry deposition as "measurements." As expressed in this paragraph, CASTNET dry deposition is based on a modeled deposition velocities. I feel that it's important to clarify that these are estimates, and not direct observations, to be clear that there is a dearth of this data with which to constrain models.

The reviewer makes a good point. We have changed this paragraph (now starting Line 175) to specify that dry deposition is an estimate rather than a direct observation. It now reads:

"We do not consider dry deposition in this analysis due to timeframe limits and inherent limitations in its determination. Dry deposition estimates are available only after 2000 over the CONUS from the Clean Air Status and Trends Network (CASTNET), and there are only four sites over Europe (EMEP) with analysis timeframes long enough to include. There are also many uncertainties with regard to both estimates and model representation of dry deposition. Dry deposition estimates are not direct observations, and are instead based on modeled deposition velocities. These estimates are typically done by using a multi-layer model (Finkelstein et al., 2000; Meyers et al., 1998), which calculates deposition velocity as a function of chemical composition, meteorology, and vegetation. Limitations to this method include the lack of meteorological measurements co-located with observation sites, requiring the use of a chemical transport model to estimate deposition velocities. These velocities are uncertain, and different velocity estimation methods can result in fluxes that differ by ~1.6x (Schwede and Lear, 2014). A recent study also shows a high bias in GEOS-Chem for nitrate dry deposition that persists throughout seasons and across multidecadal timeframes (Dutta and Heald, 2023), largely due to a model overestimate of dry deposition velocity of HNO_3. The uncertainty inherent to the dry deposition estimates, the limitations of these estimates, and the known bias in GEOS-Chem make dry deposition a more uncertain comparator for NO_x trends than wet deposition. Excluding dry deposition may lead to some bias in capturing anthropogenic trends, as dry deposition tends to be more influenced by urban sources, but natural sources are important for determining regional trends, and these are captured better by wet deposition (Dutta and Heald, 2023)."

• L191 – It may be worth reminding the reader that ozone is only assessed for Europe (and maybe include a rationale for that?).

We have clarified this. It now reads:

"Surface ozone data from 1990-2014 over Europe were obtained from the Tropospheric Ozone Assessment Report (TOAR) Surface Ozone Database (Schultz et al., 2017), which has been compiled and processed by the TOAR Database team and made public via https://doi.org/10.1594/PANGAEA.876108."

• L210 – Citation appears to be missing a year.

The names in parentheses here are actually names of ozonesonde site locations, not citations. We have made this clearer by removing the parentheses:

"Over Europe, these homogenization efforts impact 3 of the 7 ozonesonde sites: De Bilt, Hohenpeissenberg, and Uccle."

• L272 to end of paragraph – This is a nice introduction to why this simulation is included.

Thank you. We are glad it is a helpful clarification.

• Line 299 – Should this be "short lifetime of NO"?

The NO_2 is correct. This refers specifically to the daytime lifetime of NO_2, which is estimated to be 2-8 hours (Goldberg et al., 2021). We add that we are referring specifically to daytime NO_2.

"NO_2 measurements are commonly used to infer NO_x concentrations due to the short lifetime of daytime NO_2 (2-8 hours), which results in robust correlations between NO_x emissions and NO_2 column amounts (Goldberg et al., 2021)."

Goldberg, D. L., Anenberg, S. C., Lu, Z., Streets, D. G., Lamsal, L. N., E McDuffie, E., and Smith, S. J.: Urban NO $_x$ emissions around the world declined faster than anticipated between 2005 and 2019, Environ. Res. Lett., 16, 115004, https://doi.org/10.1088/1748-9326/ac2c34, 2021.

• I struggled a bit with the organization of section 3.1 and think that it would benefit from some restructuring. Most importantly I think that it would help to state earlier the link between NWD and satellite NO2 for reflecting background NOx as some form of thesis.

We have moved around a few sentences and added extra context to help make the connection between the NWD observations and satellite NO_2 clearer. First, we add our thesis statement in the first paragraph of Section 3.1:

"To establish how well NWD trends capture NO_x trends, we first compare observed NWD trends to those from satellite NO_2 measurements previously reported. NO_2 measurements are commonly used to infer NO_x concentrations due to the short lifetime of daytime NO_2 (2-8 hours), which results in robust correlations between NO_x emissions and NO_2 column amounts (Goldberg et al., 2021). This is especially useful in rural areas, such as where NWD observations sites are located, as the influence of background (e.g., non-anthropogenic) NO_x is more prominent for both satellite and NWD observations. We find that over the CONUS, the strongest decreases in NWD occur from 2000-2010 and average -4.1 ± 1.2%/yr (mean ± standard deviation). Prior to 2010, there is generally good agreement between NWD and measurements of NO_2 from surface stations and satellites when analyzed on a regional scale. The EPA's Air Quality System (AQS) surface NO_2 trends decrease by -6.6 ± 1.4 %/yr from 2005-2009 and satellite NO_2 trends decrease by -6 ± 0.5%/yr (Silvern et al., 2019), both of which are in good agreement with NWD measurements over that timeframe (-5.7 ± 1.9%/yr). It should be noted NWD is most useful as a constraint on regional spatial scales, as NWD observations are located in rural areas influenced by transport and background emissions, which can show different trends from urban areas (Silvern et al, 2019)."

We also add in a sentence to the paragraph beginning in Line 332:

"After 2010 over the CONUS, decreases in NWD observations slow down, averaging -1.2 ± 2.9%/yr from 2011-2019. This slowdown in trends is consistent with satellite measurements of NO_2, which also record a flattening of the trend from 2011-2015 at -1.7%/yr (Jiang et al., 2018). This slowdown in satellite NO_2 has been attributed to the increasing sensitivity of satellite measurements to free tropospheric NO_2, which in recent years has contributed an increasingly larger portion of column NO_2 as emissions of anthropogenic NO_x have declined (Silvern et al., 2019). NWD observations may reflect satellite NO_2 trends, as these sites are primarily rural and thus also influenced by background, non-anthropogenic NO_x, similar to satellite measurements. Another reason for the similarity between NWD and satellite NO_2 trends is that the NWD measurements capture NO_2 concentrations through the precipitation column, which extends into the free troposphere. Consistent with this hypothesis, NWD and satellite NO_2 trends over the CONUS do not agree with surface AQS NO_2 measurements, which have decreased by -4.5%/yr since 2010 (Silvern et al., 2019); the CEDS inventory also shows strong decreases after 2010, averaging -4.3 ± 0.6%/yr (Fig. S4). Over Europe, although NWD trends are noisier due to a smaller number of sites compared to the CONUS, we also find a leveling off of trends since 2010. NWD trends level off from -1.9 ± 2.2%/yr from 2000 to 2010 to -1.4 ± 3.6%/yr from 2010-2019. Again, this is in contrast with the CEDS inventory (Fig. S4), which shows a ~2 times faster decrease (an average decrease of -4.0 ± 0.9%/yr) since 2010."

• Paragraph starting line 314 – Here it's expressed that the trends in NWD and satellite NO2 mirror one another, but that they do not reflect AQS NO2. The explanation is that the satellite is picking up free tropospheric NO2. It's not clear to me that we'd expect the same reason for the NWD trend (I believe that Silvern et al. 2019 alludes to a deeper column of precipitation scavenging?). I think that it would be worth either clarifying this link or separating the rationale for the NWD and satellite NOx trends in line 318.

We clarify this connection with the edits made in the prior bullet point (please see paragraphs amended in the previous point). We have moved our explanation to the beginning of Section 3.1, as well as have added a clarifying sentence in former Line 318, now Line 336. Our hypothesis is that, because both satellites and NWD are influenced by background emissions, they will show a decreased sensitivity to anthropogenic NO_x in recent years, unlike the AQS sites which are located in urban areas. We also state that NWD reflects NO_2 concentrations through the precipitation column, not just the surface.

• Line 331: What timeframe do these percentages pertain to?

We add in the timeframe:

"In contrast, the magnitude of lightning, soil, and biomass burning NO_x emissions have remained relatively steady from 1980-2017; together they make up 34% of the total NO_x emissions profile in 2017 in the CONUS and 17% in Europe in GEOS-Chem (Fig. S5)."

• Line 333: The sentence starting "Since NWD sites are primarily rural…" is helpful for parsing the connection and distinction between the drivers of satellite NO2 and NWD trends. Consider moving this to earlier in the section?

We have included the information in this sentence in prior paragraphs in Section 3.1 (see response to bullet points relating to organization of Section 3.1 and "paragraph starting Line 314").

• L360 – Suggest cutting the sentence: "NWD magnitudes predicted by GC typically fall within measurement uncertainty ranges (<3% difference) when aggregated over the entire US domain." I don't think that this is an appropriate application of measurement uncertainty given that the bias can be larger at individual sites, and the aggregate bias is low because of the offsetting latitudinal gradient.

We change this sentence to better state our findings and clarify that some regions and seasons do not agree well (beginning of Section 3.2, now Line 379):

"Trends and magnitudes in NWD are well-reproduced by GC over the CONUS (Fig. 3). During the largest decrease, 2000-2010, GC shows a significantly ($p<0.05$) decreasing trend of -3.4 ± 0.8%/yr, which agrees with significant observed decreases in NWD of -4.1 ± 1.2%/yr. While annual NWD values predicted by GC agree with observations when taken over the entire CONUS domain, regional NWD predictions do not agree in certain regions and seasons. This issue is explored further in Section 3.3."

• L363: Consider adding quantitative details on the post-2010 trend (as for pre-2010).

We have added in summary values for the trends in the first paragraph of Section 3.2:

"Trends and magnitudes in NWD are well-reproduced by GC over the CONUS (Fig. 3). During the largest decrease, 2000-2010, GC shows a significantly ($p<0.05$) decreasing trend of -3.4 ± 0.8%/yr, which agrees with significant observed decreases in NWD of -4.1 ± 1.2%/yr. While annual NWD values predicted by GC agree with observations when taken over the entire CONUS domain, regional NWD predictions do not agree in certain regions and seasons. This issue is explored further in Section 3.3. GC v10-01 also shows a significant decrease of -2.7 ± 1.8%/yr for the same period. The post-2010 leveling off of the decreasing trend in NWD described earlier (-1.2 ± 2.9%/yr), also observed in satellite NO_2 measurements (-1.7%/yr) (Jiang et al., 2018), is captured by GC (-1.8 ± 2.4%/yr), suggesting that the CEDS NO_x trends are correct with the constraints provided by NWD observations."

• Fig. 4 – Helpful to have clarification. I find it a little unintuitive to name the simulation where emissions are changed 'meteorology' and vice versa but it is a matter of preference at this point. With red, green and brown/gold on the same plot, it may be worth checking the colors on this figure for whether they are friendly to colorblind readers.

We change the Base simulation color to be black to keep in line with colorblind-friendly palettes using the Okabe-Ito palette as a guide.

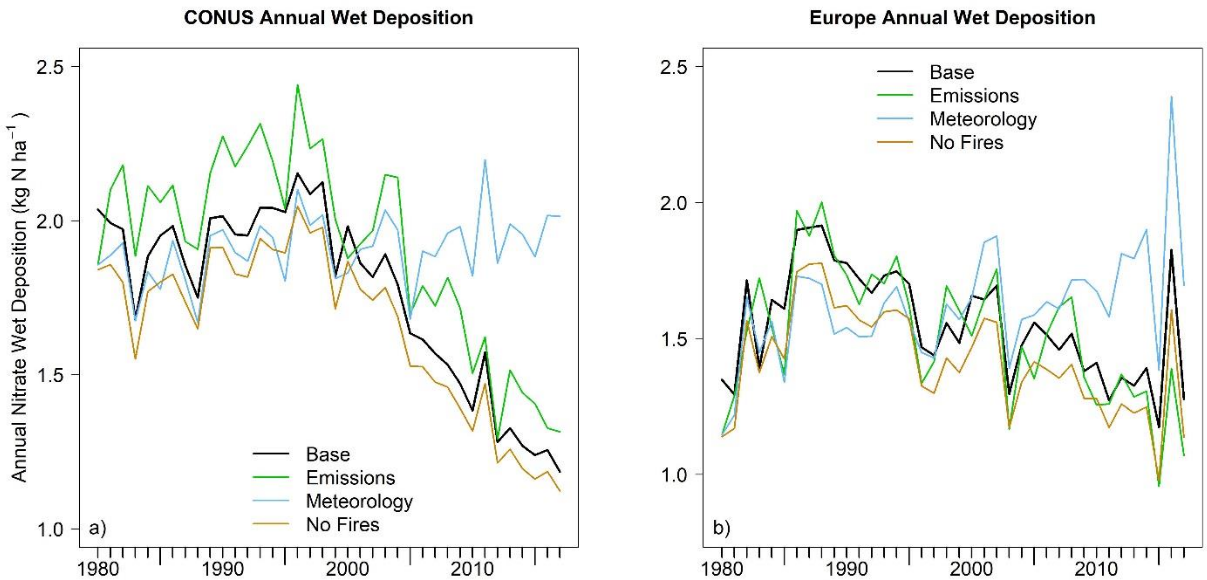

Figure 4. Results of annual NWD (kg N/ha) from each sensitivity simulation over (a) the CONUS and (b) Europe. The base GC simulation is shown in black, the simulation with changing emissions but constant meteorology (Emissions) is shown in green, the simulation with changing meteorology but constant anthropogenic emissions (Meteorology) is shown in blue, and the no biomass burning simulation (No Fires) is shown in gold. See Sect. 2.2 for a detailed description of these sensitivity runs.

• L441 – Why is summer defined as JAS? Is there an effect of considering different seasonal periods for Fig. 6 (JAS vs Aug-Sept)?

We used this definition to be able to compare our results to previous papers that have reported NEI overestimates of NO_x during summer. Each of these papers (Castellanos et al., 2011; Anderson et al., 2014; Goldberg et al., 2016; Souri et al., 2016; Travis et al., 2016) used different timeframes during summer, but all looked at timeframes occurring sometime during July-September. We used July-August-September to cover all those timeframes and make for a better comparison to previous literature. We also directly compare our results to defined timeframes, such as the August-September 2013 timeframe used in Travis et al. (2016) that we discuss in the paragraph immediately following. We have clarified this in the paragraph beginning Line 465:

"Similar to previous analyses of NEI inventories (Castellanos et al., 2011; Anderson et al., 2014; Goldberg et al., 2016; Souri et al., 2016; Travis et al., 2016), we find evidence of an overestimate of summertime (July-August-September or JAS) CEDS-estimated NO_x across the entire CONUS from 1980-2017 (12% on average), but most prominently in the eastern US after 2000 (Fig. 6a). Overestimates in the eastern US average ~20% and range up to 77% after 2000. This implies that NO_x emissions may be overestimated over the eastern US during JAS in the CEDS inventory and is consistent with previous analyses of inventories with similar emissions. We use JAS as a summertime definition for a better comparison to these previous analyses using one or multiple of these months to assess NO_x overestimates. The CEDS inventory estimates that NO_x emissions over the eastern US during JAS are 0.7 Tg N on average (0.9 Tg N in 1980 decreasing to 0.3 Tg N in 2017). These overestimates are likely present in other emissions inventories, such as the NEI 2017 and MACCity, as their emission trends and sizes over CONUS are similar across inventories (Fig. S4)."

• Comment on apparent latitudinal or coastal gradient?

We suspect that this apparent gradient during summer is an impact of dry deposition. Dry deposition is highest in summer in the extreme southeast US, where we see many of our blue points indicating an underestimate of NWD in Figure 6. An impact of not including dry deposition in our analysis is that discrepancies in parts of the country where dry deposition is more important are more difficult to assess with just wet deposition data. Here, that may mean that potential biases in NO_x emissions in this region would be more apparent in the dry deposition data. This is also true for the Mountain West, where we see evidence for underestimates in NO_x emissions using wet deposition, but dry deposition is more prominent. However, this is difficult to assess quantitatively due to the uncertainties in both dry deposition estimates and model predictions in GC, which are known to be biased high. The most prominent wet deposition during summer occurs over the rest of the eastern US, where we see high biases consistent with previous analyses. These biases may be easier to see using wet deposition data since nitrate is predominantly deposited in the northeast through precipitation. Dry deposition estimates and modeling should be improved to better resolve potential NO_x emissions discrepancies in places where dry deposition dominates.

• L598 – I'm confused by this phrase in the conclusions: "NWD magnitudes are consistent with analyses from satellite and surface NO2 measurements, demonstrating the value of NWD in constraining NOx emission changes." It may be that I'm not clear on what "NWD magnitudes" means in this context. Earlier it was suggested that satellite and NWD measured trends largely mirror one another but disagree with surface NO2 observations. This phrase seems inconsistent with that.

Here, we are referring to our finding where the relative (%) change in NWD magnitudes during COVID-19 lockdowns is consistent with other measures of NO_x changes during this time (satellites, ground measurements). We have clarified:

"NWD trends are also capable of reproducing the large drop in NO_x emissions during COVID-19 lockdowns, and relative changes in NWD magnitudes during the lockdown period are consistent with analyses from satellite and surface NO_2 measurements, demonstrating the value of NWD in constraining NO_x emission changes."

• Something that I'm left wondering following this paper is what measurements would be needed to better link NWD to NOx emissions? It seems to me that if NWD reflects total NOx, there are many levers that could be pulled (potentially in competing directions) to reduce the model NWD bias without necessarily confirming which is correct. Would it be possible to expand in more detail what sort of analysis and/or observations would be needed to refine NWD as a constraint on NOx emissions?

The most prominent measurement that would be needed to link deposition to emissions is the dry deposition. Currently, with the associated observational limitations and known model biases in GC, this is not possible, but should these be improved in the future, dry deposition should be included. There is some evidence that additional sinks for NOy would need to be included to address known model biases (Dutta and Heald, 2023). Reliable dry deposition estimates would be especially useful for capturing urban NO_x emissions, as dry deposition is more reflective of anthropogenic emissions, while wet deposition is more diffuse and thus useful for regional analyses. In addition, a better understanding of organic nitrates would prove useful, as these are not captured in inorganic nitrate measurements.

We include this discussion in our methods section beginning Line 187:

"The uncertainty inherent to the dry deposition estimates, the limitations of these estimates, and the known bias in GEOS-Chem makes dry deposition a more uncertain comparator for NO_x trends than wet deposition. Excluding dry deposition may lead to some bias in capturing anthropogenic trends, as dry deposition tends to be more influenced by urban sources, but natural sources are important for determining regional trends, and these are captured better by wet deposition (Dutta and Heald, 2023). Magnitudes may be more uncertain, as the geographic distribution between wet and dry deposition differs. For example, over the CONUS, wet deposition is most prominent in the northeast US, and dry deposition is most prominent in the southeast US. Thus, dry deposition may be more adept at capturing biases in NO_x emissions in the southeast US, which may not be as apparent in the wet deposition data. This introduces uncertainty to our regional analysis of NO_x magnitudes using just NWD (Section 3.3); this uncertainty is currently difficult to quantify given observational limitations in dry deposition and known model biases in GEOS-Chem."