Point-to-point response

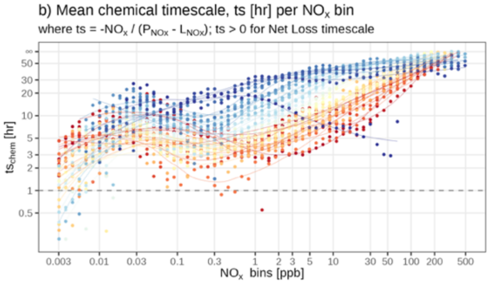
We thank two reviewers for their constructive comments and have addressed all comments with additional analyses and clarifications to the manuscript. We start the response with a summary of changes for both reviewers. Our point-to-point responses are highlighted in blue with track changes in blue/red and reviewers' comments in black.

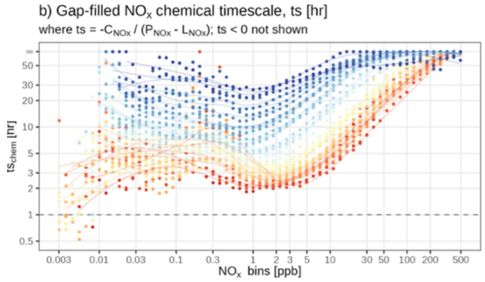
General responses to both reviewers

Two main criticisms from the two reviewers include:

1) Omission of chemical pathways and processes (heterogeneous NOx chemistry, NO + HO2/RO2);
2) Lagrangian atmospheric mixing (mixing scales in PBL, mixing between FT and ML).

We also identified a minor bug in the estimated NOx net timescales and reran simulations. The main difference from the initial submission (**upper right panel**) is the timescale over NOx-limited regimes at high SZA bins (NOx < 0.1 ppb and SZA > 50 degrees in the **lower right panel**).





As changes in the TROPOMI version and model parameters may affect model-data comparisons, we summarize the model-data slopes for all overpasses over the New Madrid power plant based on each configuration (**Fig. R1 below**). All simulations presented here used EPA emissions and 3km HRRR meteorological fields. Similar to **Fig. 6**, not accounting for NOx chemistry results in a positive model bias of tNO2 (**blue crosses in Fig. R1**); and using the latest TROPOMI version (v2.3 from v1.3) has largely reduced the model-data mismatches (**orange circles to orange dots**). The minor correction in the derived net loss timescale and the mixing time scale (from 3 to 1hr) only slightly alter the model-data slopes (**orange dots vs. purple dots vs. purple crosses**).

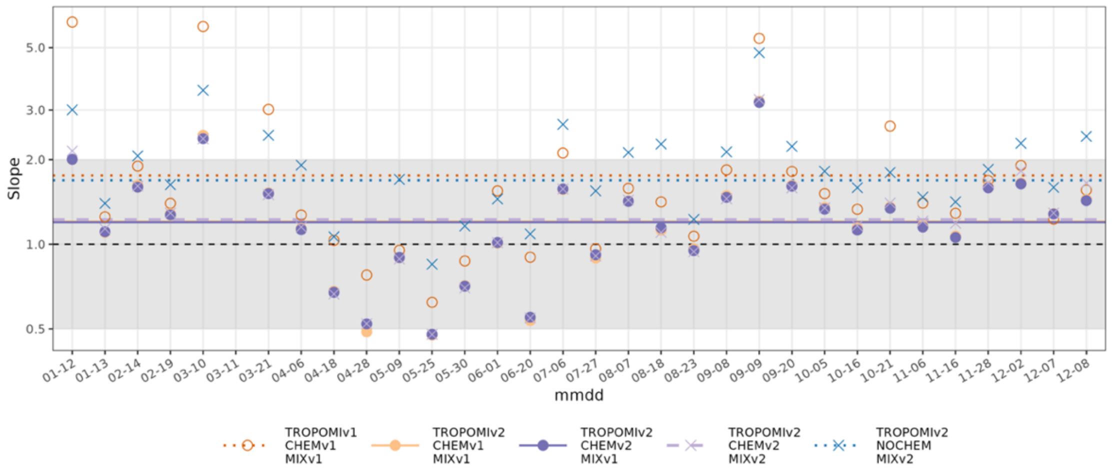


Figure R1. Model-data comparisons informed by the linear slope between the two using different versions (v1/v2) are explained as follows: **TROPOMI**: v1.3 or v2.3 of the TROPOMI L2 NO_2 retrieval. **CHEM**: v1 vs. v2 for runs using NOx curves from the initial vs. revised manuscript (as shown in the above comparison). **MIX**: inter-parcel mixing with two different horizontal mixing timescales tested (3 vs. 1 hr) over a 1km box. Tests with a spectrum of mixing parameters have been conducted in response to the 1st comment of reviewer 2.

Comments from reviewer #2

This paper introduces STILT-NOx, a Lagrangian chemical transport model, evaluates it against satellite-based column observations of NOx, and presents various sensitivity studies. The paper is well written, and I recommend publishing after the following minor comments are addressed.

We thank reviewer #2 for the positive feedback and have tried to address all the comments. Please also refer to the general comments for both reviewers on the first page if neccesary.

General comments:

Interparcel-mixing: I was a bit surprized to see the 3-hour timescale for mixing within a volume with a horizontal of 1km x 1km and a vertical extend from surface to PBL top described as "relatively fast mixing". Given the fact that the plume emitted from a power plant, when transported over 3 hours (or 30 km for typical winds speeds of 10 m/s within a well-developed PBL), and given the shape of typical plumes seen from satellite imagery, can the mixing time scale not be estimated from that? Based on that I would expect that within three hours mixing likely occurs over much larger "boxes". This is along the thought that dispersion/mixing is similar when running LPDMs backward and forward as shown in the Lin et al. 2003 paper, so forward mixing as needed for chemistry should be similar to backward mixing (or spreading of particles emitted at the same time). Would a smaller mixing timescale increase the impact of mixing, as with a 3-hour time the impact was found to be less than 5% (line 356)? Was that assessed when comparing no-mixing (i.e. infinite timescale) with mixing turned on? I think this deserves a bit of attention, as it keeps puzzling me that given the quite nonlinear property of NOx chemistry mixing at those scales does not seem to matter in chemistry-transport simulations.

What matters here is the horizontal turbulent mixing/diffusion scales as the rapid vertical mixing in STILT is realized by diluting emissions from the surface to the mixed layer. If horizontal mixing is not accounted for, individual air parcels will not "communicate" with their neighbor parcels for mass exchange. Indeed, previous studies have used a Gaussian plume model to fit the observed plumes, especially of CO, to determine the horizontal diffusion rate from observed concentrations and their horizontal distribution (e.g., along the across-wind direction). However, the chemical decay and column observations of NO_2 make it challenging to accurately separate diffusion timescales from chemical timescales.

To address the question of inter-parcel mixing, we have added a sensitivity test of modeled tropospheric NO_2 to various horizontal mixing length- or time- scales. Three mixing length scales of 1, 3, and 10 km are tested, considering the typical km scale of satellite sounding, emission grid, or high-resolution meteorological fields.

According to Seinfeld and Pandis (2016) and several practical horizontal diffusion treatments (e.g., either the simple constant diffusion coefficient or the Smagorinsky scheme in WRF), the typical diffusion time scale ranges from minutes to a few hours for a mixing box of 1 km. For example, Seinfeld and Pandis (2016) suggested that under stable met conditions, the horizontal diffusion coefficient K_{yy} is on the order of 50 to 100 $m^2 s^{-1}$, which translates to approximately 2.7 hrs in terms of a characteristic diffusion timescale (dx^2 / K_{yy}).

Furthermore, we tested a range of time scales from 0.1 to 100 hours over three receptors from a summertime New Madrid power plant overpass (**see plot shown below, now as Fig. S20**). Three receptors differ in the number of trajectories influenced by the single power plant emission. Since STILT-NOx tracks modeled tNO2, xCO, and xCO$_2$ with mixing and without mixing, we can calculate the normalized difference in the modeled tNO2 between the mixing runs and the non-mixing runs, i.e., = (tNO2_mix – tNO2_nomix) / tNO2_nomix.

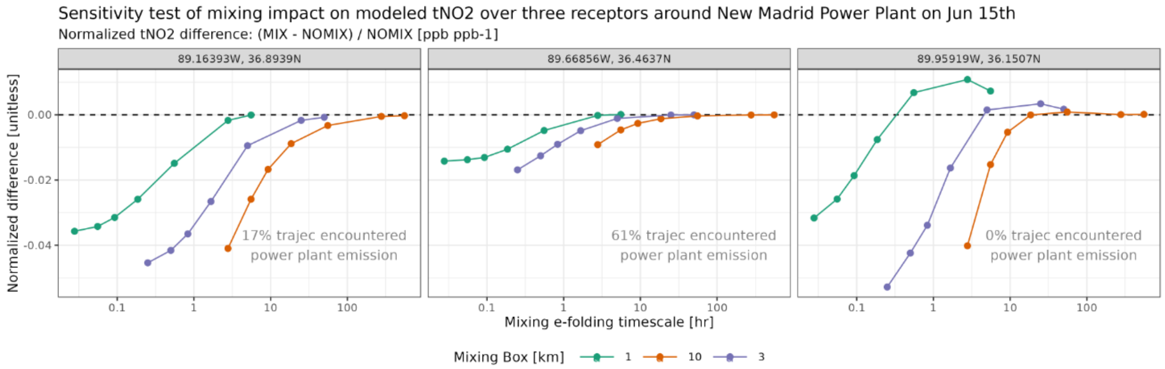


Figure S20. Normalized difference in modeled tropospheric NO$_2$ between the mixing and non-mixing runs as a function of the horizontal mixing timescale [hr, x-axis] and mixing length scale [km, in colors]. The normalized difference is calculated as $(\mathrm{tNO}_{2,\mathrm{MIX}} - \mathrm{tNO}_{2,\mathrm{NOMIX}})/\mathrm{tNO}_{2,\mathrm{NOMIX}}$. Three examples for three receptors/soundings on June 15th for the New Madrid case are shown here and they differ by the fraction of model trajectories that “hit” the power plant emission. For receptors where some trajectories encountered the emissions, a faster mixing reducing the spatial gradient in NO$_x$ leads to a reduced final tNO$_2$ at the receptor (left two panels).

As a result, for the receptor close to the power plant emission (where 61% of the trajectories were affected by power plant emissions for at least 1 min, **in the middle panel in Fig. S20**), the reduction in tNO2 is as minimal as about 2% regardless of the mixing length scale. In contrast, when trajectories are not influenced by power plant emissions (i.e., receptor/sounding sitting on the edge of the plume, **in the left- or right-most panels in Fig. S20**), the changes between mixing and non-mixing simulations appear to be larger, i.e., reaching 5%. The larger discrepancy is reasonable as mixing would exchange the NOx concentration between trajectories experiencing the plume and trajectories in the background.

In summary, mixing tends to “smooth” the gradient in NOx concentrations among air parcels but to a less extent if a larger fraction of the total 2000 air parcels is contaminated by the plume.

We now added relevant discussions of horizontal PBL mixing and other mixing in Sect. 5.3.2 (see track changes attached on the following page).

Accounting for inter-parcel mixing was an important aspect when developing Lagrangian chemical models. Omitting inter-parcel mixing makes solving for non-linear processes (such as chemical NO_x loss) problematic. On the contrary, Eulerian models suffer from excessive numerical diffusion. Mixing that is too strong smooths the spatial gradient of concentration and can lower the concentration within the fixed model grids, which may cause slight shifts in NO_x regimes. Valin et al. (2011) suggested that a spatial resolution of 4 to 12 km is sufficient to capture the non-linearity in NO_x loss rate. As for Lagrangian models, efforts can be made to enable the flux exchange between air parcels via deformations (Konopka et al., 2019; McKenna et al., 2002). ~~STILT realized the~~ In addition to the inter-pacel mixing within the mixed layer (ML), several other turbulent mixing processes require future investigation, including (1) horizontal mixing in the free troposphere (FT), (2) vertical mixing between the ML and FT, and (3) mixing between tracked air parcels with the untracked surrounding background. For example, Real et al. (2008) utilized a linear relaxation with exponential decay of the plume concentrations towards the background based on a timescale of 2 days to address the third mixing process. The second mixing process requires future modifications involving the determination of entrainment zones and mixing hyperparameters for such ML-FT exchange.

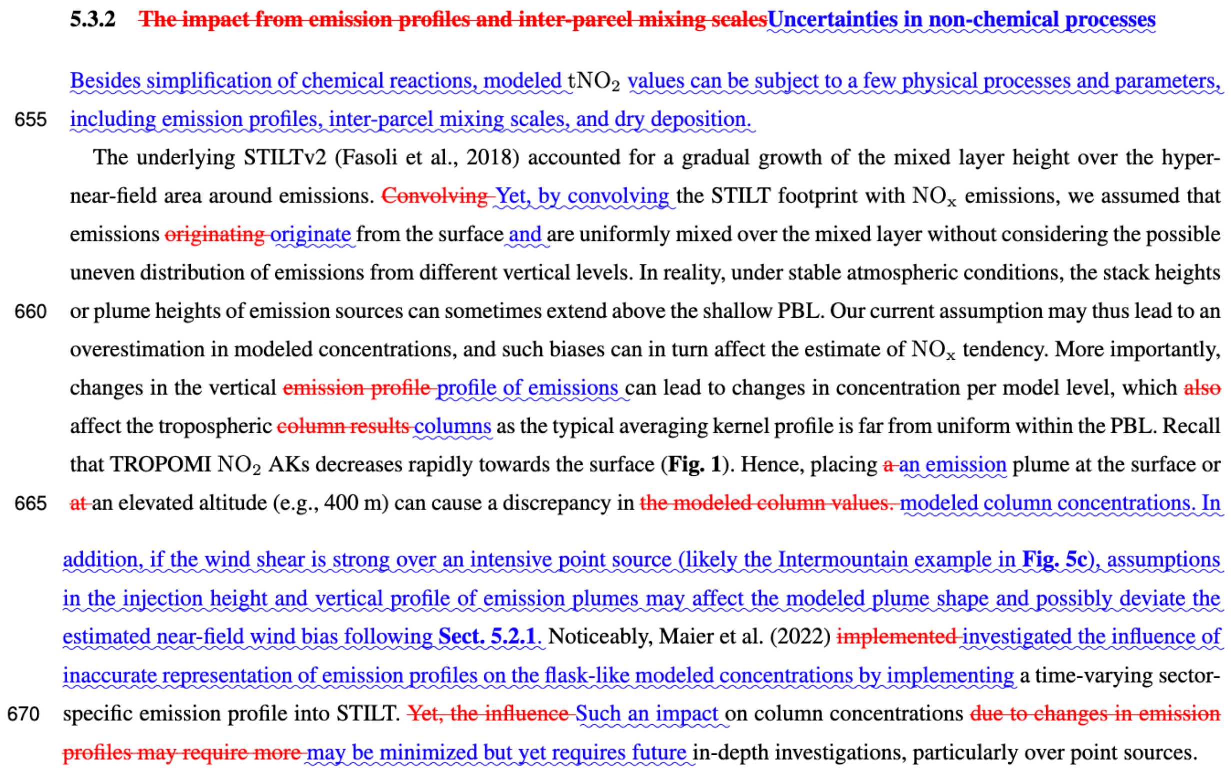
The original STILT model realized vertical mixing by diluting surface emissions ~~over~~ across the ML height (Lin et al., 2003) ~~. We~~ and we further enabled an exchange in pollutants' concentrations with prescribed mixing length- and ~~time- scales following (Wen et al., 2012). As a final sensitivity test~~time-scales representing typical horizontal mixing rates (**Sect. 2.3**). As final sensitivity tests, we simulated tNO_2 ~~using~~ based on a spectrum of mixing hyperparameters ~~over~~ for the New Madrid power plant~~and found minimal influence on modeled values per sounding (uncertainty < 20%,~~. Uncertainties in the mixing parameters result in minimal uncertainties on the sounding-level modeled tNO_2 values (**Supplement Fig. S19**). For example, differences in modeled tNO_2 between the mixing and non-mixing simulations become larger as mixing becomes faster and for receptors/soundings located on the edge of the plume (i.e., only a small fraction of the trajectories encountered power plant emission in **Supplement Fig. S20**). Uncertainties in the prescribed mixing hyperparameters contribute even less to the modeled values over urban areas (i.e., < 10% for Phoenix cases), where emissions are generally better mixed than at power plants. In addition, such mixing influence can vary with the spatial resolution of the emission inventory used in the simulations.

The dry deposition of NO_2 was not factored into this study, which could lead to an overestimation of modeled NO_2. For future model implementations, it is possible to track loss of NO_2 concentrations due to dry deposition by calculating "dry deposition velocities" (e.g., **?**) when air parcels descend close to the surface, e.g., 50 meters above the surface (Wen et al., 2012).

Rotation of wind: As the wind changes significantly within the atmospheric boundary layer with height (the Ekman spiral), differences between modelled wind direction and the direction apparent from the observed plume can also be related to inaccuracies in the plume release height distribution, potentially associated with plume rise of the buoyant exhaust. I would recommend this to be discussed.

We have overlooked the impact of inaccurate emission profiles on the modeled plume shapes (not just on column concentrations) and quantified near-field wind directional biases. One can examine the vertical wind shear from trajectories output for the near-field region between receptor and emission sources for more clues.

We now add the following text in Sect. 5.3.2 (the limitation section on emission profiles and mixing).



5.3.2 ~~The impact from emission profiles and inter-parcel mixing scales~~Uncertainties in non-chemical processes

Besides simplification of chemical reactions, modeled tNO_2 values can be subject to a few physical processes and parameters, including emission profiles, inter-parcel mixing scales, and dry deposition.

The underlying STILTv2 (Fasoli et al., 2018) accounted for a gradual growth of the mixed layer height over the hyper-near-field area around emissions. ~~Convolving~~ Yet, by convolving the STILT footprint with NO_x emissions, we assumed that emissions ~~originating~~ originate from the surface and are uniformly mixed over the mixed layer without considering the possible uneven distribution of emissions from different vertical levels. In reality, under stable atmospheric conditions, the stack heights or plume heights of emission sources can sometimes extend above the shallow PBL. Our current assumption may thus lead to an overestimation in modeled concentrations, and such biases can in turn affect the estimate of NO_x tendency. More importantly, changes in the vertical ~~emission profile~~ profile of emissions can lead to changes in concentration per model level, which ~~also~~ affect the tropospheric ~~column results~~ columns as the typical averaging kernel profile is far from uniform within the PBL. Recall that TROPOMI NO_2 AKs decreases rapidly towards the surface (**Fig. 1**). Hence, placing ~~a~~ an emission plume at the surface or ~~at~~ an elevated altitude (e.g., 400 m) can cause a discrepancy in ~~the modeled column values.~~ modeled column concentrations. In addition, if the wind shear is strong over an intensive point source (likely the Intermountain example in **Fig. 5c**), assumptions in the injection height and vertical profile of emission plumes may affect the modeled plume shape and possibly deviate the estimated near-field wind bias following **Sect. 5.2.1**. Noticeably, Maier et al. (2022) ~~implemented~~ investigated the influence of inaccurate representation of emission profiles on the flask-like modeled concentrations by implementing a time-varying sector-specific emission profile into STILT. ~~Yet, the influence~~ Such an impact on column concentrations ~~due to changes in emission profiles may require more~~ may be minimized but yet requires future in-depth investigations, particularly over point sources.

Specific comments:

Fig. 3: which WRF-Chem runs were used? Before three different cities were mentioned, are all those simulations included in Fig. 3?

Yes – chemical tendency derived from every WRF-Chem model grid around the three cities (LA, Madrid, and Shanghai) from all monthly simulations (4 days in each month across Jan to Dec with the first day as the spin-up time) are included and aggregated into SZA and NOx bins as shown in Fig. 3. We did remove model grids with extreme values (timescale > 72 hours) to avoid skewing the average values per bin.

L200: Were the WRF-Chem simulations also selected for cloud-free conditions?

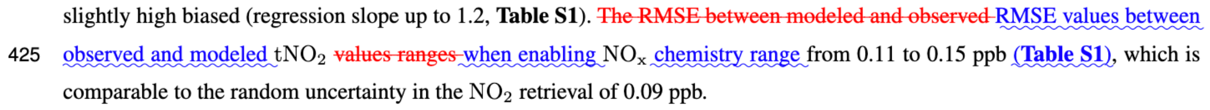
Good question- we did not explicitly remove cloudy scenes from WRF-Chem. During revisions, we re-examined the cloud mixing ratio estimated by WRF-Chem to address the impact of cloud covers on the non-linear RNOx ~ NOx + SZA relationship. For Shanghai or LA, the average fraction of cloudy pixels ranges from 0.03% (or 0.02%) in winter to 3.6% (or 4.8%) in the spring/summer months. On an annual basis, the cloudy pixels only occupy ~1.8% to 2.1% of the select near-field model pixels for three cities. We might expect the impact on NOx curves to be minimal especially since millions of grid cells are considered to create the NOx curves.

Fig. 5 b and d: the color code is missing, or am I overlooking something?

The color legend for Fig. 5b and 5d is labeled at the bottom of Fig. 5d. The four different colors represent linear fit and slope when TROPOMI is compared with four STILT-NOx runs using 1) meteorological fields (GFS vs. HRRR), 2) emission (EDGAR vs. EPA), and 3) chemistry (default chemistry vs. non-chemistry).

L394: in table S1 the RMSE values range from 0.11 to 0.25 ppb

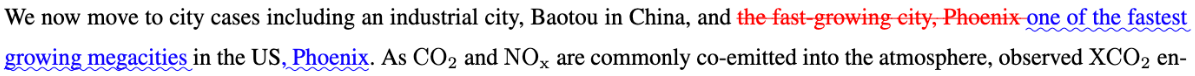
We clarify that range of RMSE is reported for only the simulations with NOx chemistry enabled. The highest RMSE of 0.25 ppb corresponds to simulations without NOx chemistry. **We modify the text as:**



slightly high biased (regression slope up to 1.2, **Table S1**). ~~The RMSE between modeled and observed~~ RMSE values between observed and modeled tNO$_2$ ~~values ranges~~ when enabling NO$_x$ chemistry range from 0.11 to 0.15 ppb (**Table S1**), which is comparable to the random uncertainty in the NO$_2$ retrieval of 0.09 ppb.

L412: “fast-growing” is relative, Baotou certainly has a faster growth in population than Phoenix

Corrected – Phoenix sees a high population growth rate of 2.5% since 2020, one of the “fast-growing” cities in the US. **We now revised the text as:**



We now move to city cases including an industrial city, Baotou in China, and ~~the fast-growing city, Phoenix~~ one of the fastest growing megacities in the US, Phoenix. As CO$_2$ and NO$_x$ are commonly co-emitted into the atmosphere, observed XCO$_2$ en-

Fig. 8a: What does RMSP stand for?

Corrected – the y-axis on Fig. 8a represents the spatial mean of the sounding-wise products of two NO$_2$ plumes with the former one being the initial unrotated observed plume and the latter one being the rotated plume (either from TROPOMI or STILT-NOx using GFS or HRRR). We’ve now fixed the title and caption for Fig 8a.

Table 1 in the supplement should be named “Table S1”
Corrected.

Fig. S6: please use axis titles that clearly indicate v1 and v2 in all figures
Corrected.

Fig. S7: the symbols in the legend don’t quite fit with those in the figure. Triangles should be should only for EDGAR estimates, not for EPA.
Corrected.

Fig. S10 a and b: please use fewer x-axis labels
Corrected.