Response to Referee #1:

We would like to thank the referee for the useful comments and constructive suggestions. In the following, we address the referee's comments and describe corresponding changes we have made to the manuscript. The referee's comments are listed in blue, followed by our response in black. New/modified text in the manuscript is in *dark orange italics*.

The authors clearly demonstrate the significance and necessity of improved NH_3 flux estimates, and the results present interesting spatial and seasonal patterns that are valuable to the community. However, the methodological description is currently insufficient and relies too heavily on briefly reproducing elements from Ayazpour et al. (2024) without providing enough self-contained derivation or detailed explanation. To ensure clarity and reproducibility, I strongly recommend expanding Section 2 with a more thorough presentation of the directional derivative framework, explicit definitions of all key variables and terms, and clearer justification for the assumptions made. Addressing these issues will greatly strengthen the scientific rigor and standalone value of the manuscript. Furthermore, emissions are presented in quite some detail, but the deposition fluxes seem to draw the short straw. The lack of dry-deposition measurements of course does not help in being able to make a nice comparison, but some discussion or comparison with past modelled results would increase the value of the deposition results. Subsequently linking those results to for example critical load limits could greatly improve the overall value of the manuscript and enhance the impact of all the work that is already performed.

We thank the referee for the constructive suggestions. In the revised manuscript, we have expanded Section 2 to provide a more detailed description of the directional derivative analysis (DDA) framework. We have also strengthened the discussion of deposition fluxes in Section 4.2 (Line 466 in the original manuscript) by comparing with past modelled deposition results: "*Similar hotspots of deposition downwind of intensive agricultural regions have been reported in model-based studies (Ellis et al., 2013; Hu et al., 2021; Zhang et al., 2012). These hotspots often coincide with high NH_3 loadings, large leaf area indices, and micrometeorological conditions favorable for stomatal and cuticular uptake (Sutton et al., 2009).*"

Main comments:

1. Section 2.2 — Please define "DD" clearly before using it throughout the paper. Also clarify the meanings of "∇" and "Ω" (column density) upon first mention.

We thank the referee for pointing this out. We do agree that the terms "DD", "∇", and "Ω" should be more clearly stated. We have revised the description of Eq.1 in Section 2.2.1 (Lines 154-168) as follows:

"The estimation of emissions (E) from satellite-observed VCDs (Ω) is grounded in the principle of mass conservation as in Eq. (1), which is in the same form as presented in the previous DDA literature (Sun, 2022; Ayazpour et al., 2025). The DDA considers the physical and chemical processes affecting gas distribution, incorporating horizontal transport, topography, and chemical transformation. Three estimators within the DDA framework are labeled in Eq. 1 as DD, DD_topo, and DD_chem, representing the directional derivative of column densities, the directional derivative with consideration of topography, and the directional derivative with consideration of both topography and chemistry. DDA considers the physical and chemical processes affecting gas distribution, incorporating horizontal transport, topography, and chemical transformation. The DD estimator ($\vec{u} \cdot (\nabla\Omega)$) captures the horizontal advection of NH_3, representing the directional derivatives of the VCDs with respect to horizontal wind vectors representing the planetary boundary layer ($\vec{u}$, 100 m winds). $\nabla = (\partial/\partial x, \partial/\partial y)$ is the horizontal vector differential operator. The wind height choice is supported by Ayazpour et al. (2025), who demonstrated that winds from ~100 m yield strong agreement between DD estimator and emissions while maintaining robustness across boundary layer depths. The DD_topo estimator accounts for the topography term ($X\Omega\overrightarrow{u_0} \cdot (\nabla z_0)$), which is driven by directional derivatives of the surface altitudes (z_0, obtained from Level 2 satellite data) relative to near-surface wind vectors ($\overrightarrow{u_0}$, 10 m winds). This component captures the influence of terrain on NH_3 movement. For example, variations in elevation can create localized gradients that resemble NH_3 fluxes. The DD_chem estimator considers the chemistry term ($k\Omega$), representing chemical interactions between NH_3 and atmospheric acids which result in the formation of particulate matter."

2. Line 136 vs. Line 223 — There appears to be an inconsistency regarding the emission inventory year: Line 136 mentions using HEMCO for 2016, while Line 223 states the focus is on September 2019 to April 2021. Please clarify which year(s) of the emission inventory were used and how they relate to the period analyzed.

We thank the referee for pointing out this potential inconsistency. We clarify that the gridded monthly HEMCO emission inventory is only available to us for the year 2016, and this is consistently used throughout the study for two purposes: (1) To identify low-emission regions (where $E < 1\times10^{-9}$ mol m^{-2} s^{-1}) that are suitable for fitting the topography and chemistry terms under the assumption of negligible emissions. (2) For comparison with the satellite-derived flux estimates from IASI and CrIS during the analysis period of September 2019 to April 2021. We have updated the manuscript to clearly clarify the use of HEMCO data in Section 2.1 and the role of the emission inventory in the fitting process in Section 2.2.1 as follows:

In Section 2.1 (Lines 140-142): *"We use this bottom-up inventory from HEMCO to compare with satellite-derived fluxes, providing insights into their consistency and helping to assess the utility of satellite-based estimates."*

In Section 2.2.1 (Lines 187-189): *"In regions where emissions are negligible ($E < 1 \times 10^{-9}$ mol m^{-2} s^{-1}, based on bottom-up inventory from HEMCO; see Fig. S1 for spatial distribution), Eq. (1) can be reformulated into a multilinear regression form model by omitting the emission term".*

3. Line 143-144: Why are the wind data at 100-10m used when the observed air masses are clearly well mixed within the mixing layer, would a wind speed more representative of the mixing layer (500m-1000m) not make more sense? Wind speeds closer to the surface can also be expected to be much smaller, which seems essential to for the resulting fluxes. Either add a reference for this value or better explain the expected uncertainties.

We thank the referee for this thoughtful comment. We acknowledge the importance of wind height selection in estimating surface fluxes. In the DDA framework (Eq. 1), we use 100m winds for the horizontal wind $\vec{u}$ in DD estimator ($\vec{u} \cdot (\nabla \Omega)$). The 100 m wind reflects horizontal advection near the surface but still within the well-mixed lower boundary layer. Fig. 3c of Ayazpour et al. (2025) below shows that using winds from layer 4 (~100 m) yields strong agreement between DD estimates and true emissions, and that the DD is robust to wind height over layers 4–7 (~100–800 m).

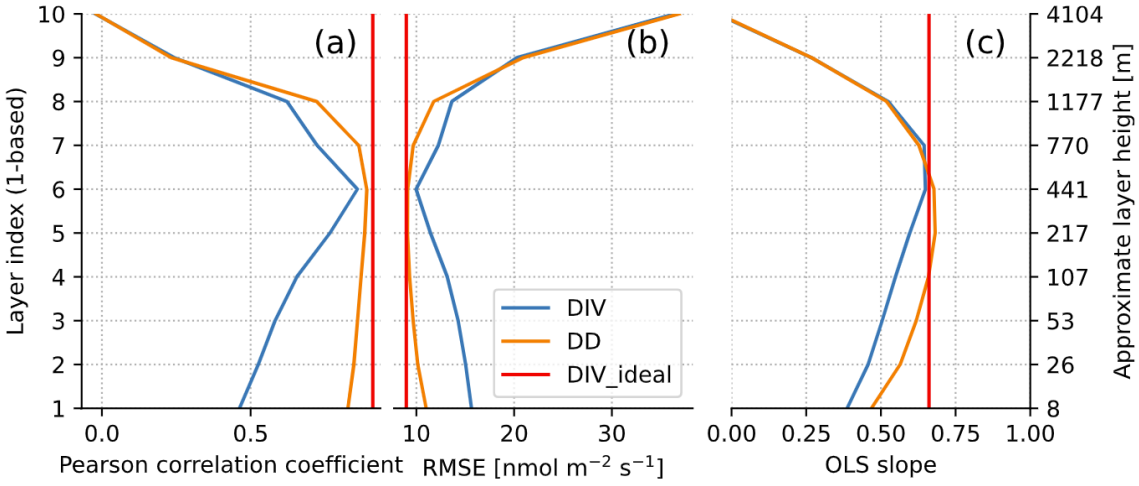

Figure 3 of Ayazpour et al. (2025). (a) Pearson correlation coefficient, (b) root mean square error, and (c) slope of linear regression of DIV_ideal and DIV estimators (two estimators in the flux divergence approach, not relevant for this work) of the divergence method and the DD estimator of the directional derivative approach framework when

compared with model emissions from WRF-CMAQ using winds at different layers. The approximate layer center altitude is shown on the right vertical axis.

We use the 10m winds for the near-surface winds $\overrightarrow{u_0}$ in the topography term ($X\Omega\overrightarrow{u_0} \cdot (\nabla z_0)$). Because this process occurs very close to the ground, using near-surface (10 m) winds is physically appropriate for this component of the model. We have revised the text in Section 2.2.1 (Lines 159-160) and added a reference to Ayazpour et al. (2025) to support the choice of wind layer: *"The wind height choice is supported by Ayazpour et al. (2025), who reconstructed emissions using model (WRF-CMAQ) column and winds at different model layers and found that winds from ~100 m yield strong agreement between the DD estimator and model-ingested emissions while maintaining robustness across wind heights up to ~800 m."*

4. Line 197–199 (Fitting criterion) — The statement "This fitting was limited to rough terrains…" conflicts with the description "This step was conducted in flat terrains…" and also appears to misrepresent the original criterion (Ayazpour et al., 2024, Sect. 3.2: "which eliminates open water and very rough terrain"). Please clarify whether the fitting excludes both open water and very rough terrain, or whether it is limited to rough or flat terrains, and explain whether the fitted parameters from flat terrain are appropriate for application in mountainous areas or for the entire CONUS domain.

We thank the referee for this helpful comment. Consistent with the approach in Ayazpour et al. (2025), our fitting process excludes both open water bodies and very rough terrain to ensure the reliability of the parameter estimates. Specifically, we performed a two-step fitting: (1) the scale height (X) was fitted in moderately rough terrains, where elevation gradients are sufficiently present to constrain the relationship, and (2) the chemical term (k) was fitted in flat terrains, where topographic influences are minimal. Moderately rough terrains are broadly distributed across the CONUS and provide representative conditions for fitting X, while flat terrains are best suited for isolating k. We have clarified the terrain selection criteria in the revised manuscript below.

In Section 2.2.1 (Lines 195-200), we have added the description of the fitting process as follows: *"We conducted a two-step fitting process to estimate X and k following Lonsdale and Sun (2023). The two fitting processes exclude open water bodies and very rough terrains. The first fitting step focused on β_1 since the fitting results for β_2 are usually noisy. The first-round fitting for β_1 was limited in moderately rough terrains with $0.001\ m\ s^{-1} < \langle\overrightarrow{u_0} \cdot (\nabla z_0)\rangle < 0.1\ m\ s^{-1}$. Once β_1 was determined and then fixed, and the second-round fitting for β_2 was conducted in flat terrains ($\langle\overrightarrow{u_0} \cdot (\nabla z_0)\rangle < 0.001\ m\ s^{-1}$) with moderate NH_3 VCDs (Ω > 2.5 × 10^{-5} mol m^{-2}) and minimal emissions (E < 1 × 10^{-9} mol m^{-2} s^{-1}) to isolate chemical transformation."*

5. Scale Height Assumption — The manuscript assumes a regionally constant scale height. Considering the substantial local variability in boundary layer depth and surface conditions, could the authors discuss how this assumption affects the flux estimation, particularly over complex terrain, and whether a spatially or seasonally varying scale height was tested?

We thank the reviewer for pointing this out. Fundamentally, the inverse scale height and chemical reactivity in DDA are empirical fitting parameters that ensure the emissions center around zero in low-emission regions. They are tied to physically meaningful parameters through the theoretical derivation of the DDA equations, but they should not be constrained or prescribed. This is because the fittings are already "optimal" as they are directly informed by satellite data and auxiliary variables (wind, surface altitude). Prescribing these numbers, potentially in a complicated, spatiotemporally resolved way (i.e., from a model), will deviate the derived emissions over confidently low-emission regions from zero and thus bias the emissions over source regions. This also allows the large number of assumptions and approximations in various model settings to influence the outcomes of DDA and negates our aims to provide lightweight and timely observation-based emission estimates. The downside of always fitting the emission estimator to zero in low-emission regions through linear regression is the constraints from the quality and quantity of available satellite data. In the case of IASI and CrIS, we do expect the quality of emission estimation to degrade over complex terrain because individual IASI and CrIS pixels are separated by gaps (in contrast to tiled pixels like TROPOMI) and therefore undersample the topography. Disaggregating satellite data spatially and/or temporally may better resolve the fitted parameters but at the expense of further thinning the data and losing the signal under the noise. The clarify these points, we rewrite Lines 201-205 of Section 2.2.1 in the original manuscript as separate paragraph:

"X and k represent the inverse scale height and chemical reactivity, in which scale height represents the characteristic height of the species' vertical distribution, and chemical reactivity represents the inverse of average time before the species being removed by chemical reactions. To improve the performance of the flux estimates, we treat the X and k as fundamentally empirical fitting parameters within a data-driven approach to ensure the resultant emission estimator centers around zero where emissions are negligible. Although in theory X and k are tied to physically meaningful quantities, their main purpose is to enhance emission estimators in the presence of topography and chemistry using information contained by satellite observations. As such, the quality of fitted X and k is subject to the quality and quantity of available satellite data. Because of the gaps in individual IASI and CrIS pixels that lead to undersampling of topography, we expect that X cannot fully account for topography effects over complex terrains, an inherent limitation for scanning Fourier Transform Spectrometers like IASI and CrIS."

To identify the optimal spatial granularity, we divided the domain into quantile-based bins according to topographic conditions, represented by $\langle \Omega \overrightarrow{u_0} \cdot (\nabla z_0) \rangle$. Within each bin, the scale height X was fitted independently. This binning approach allows regions with similar topographic forcing to be treated consistently, but it also increases the noise level when the number of bins becomes large. X remains positive when using a single bin (Fig. S2), whereas using multiple bins can introduce excessive noise, sometimes leading to negative fitted values. Temporal variability was assessed by fitting with different temporal aggregation windows, where individual flux estimates were averaged over fixed periods before fitting (e.g., 1-, 3-, 6-, and 9-month intervals). Longer windows (>6-months) produce stable and positive fitted X values, while shorter intervals yield larger noise and sometimes negative fitting results (Fig. S3).

We have added the sensitivity tests of fitted X in the end of Section 2.2.1 as follows: *"We tested different spatial groupings, temporal averaging windows, and stricter maximum emission thresholds to refine the fitted X and k. For X, fitting the entire domain as a single group produces consistently positive and stable values, whereas subdividing the domain often introduces excessive noise (Fig. S2). Similarly, varying the temporal aggregation affects the stability of the fits (Fig. S3): shorter intervals produce noisier estimates, while longer intervals yield more robust results. We adopted a six-month aggregation interval as it provides reliable estimates across the entire period while retaining seasonal variability."*

The following figures have been added to the Supplementary Materials:

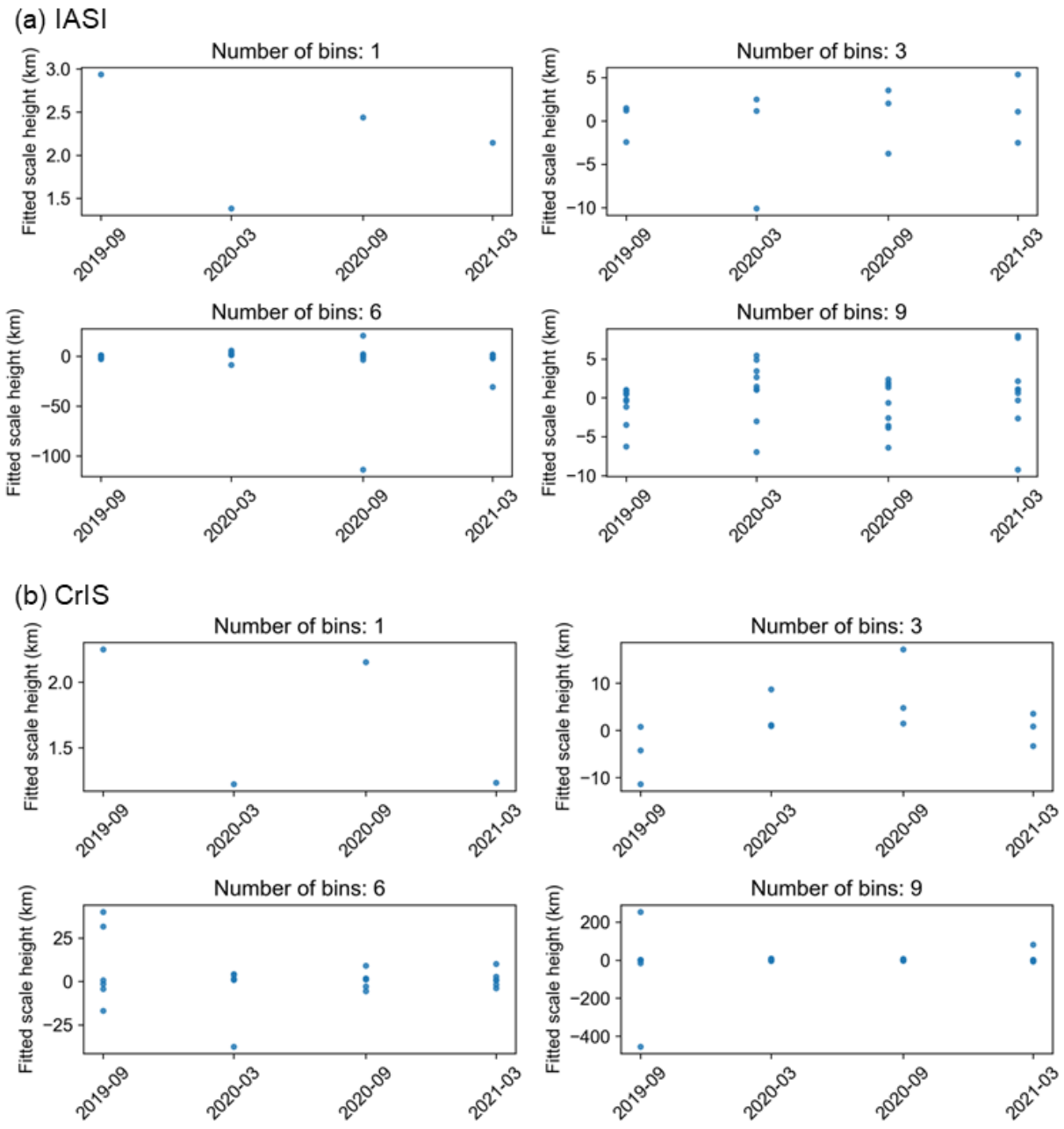

Figure S2. Spatial sensitivity tests for the topography term fitting to derive DD_topo from IASI (a) and CrIS (b). The number of bins indicates the wind-topography ($\Omega\overrightarrow{u_0} \cdot (\nabla z_0)$) sectors used in the fitting. The y-axis shows the fitted X.

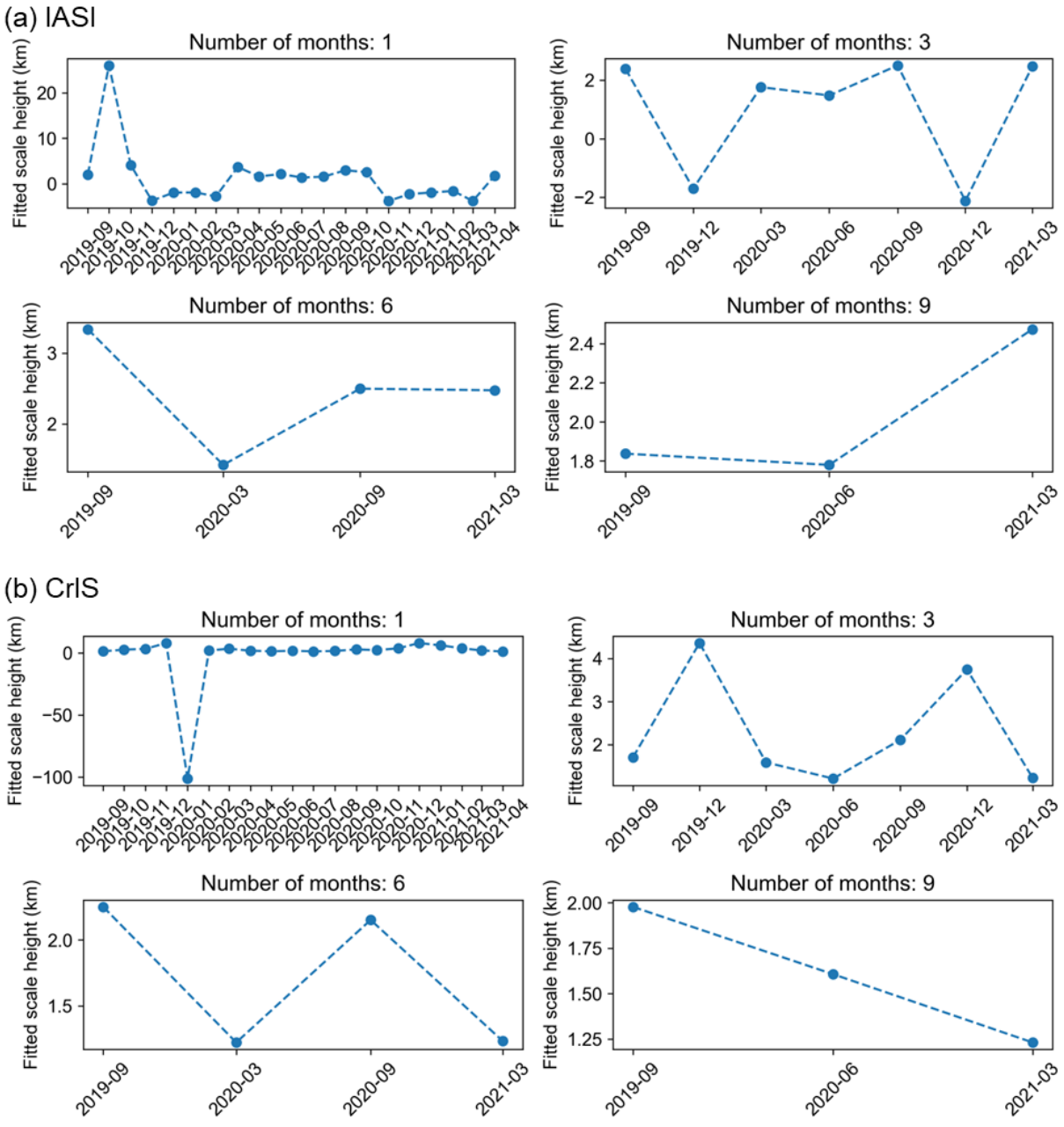

Figure S3. Temporal sensitivity tests for the topography term fitting to derive DD_topo from IASI (a) and CrIS (b). The number of months indicates the temporal aggregation interval used for the fitting. The y-axis shows the fitted X.

6. Inverse scale height – As stated the scale height has a direct relation to the mixing layer, wouldn't it make more sense to use a boundary layer height product to substitute into these functions instead of deriving them from the satellite observations? This will probably remove potential artifacts from spotty spatial/temporal measurement records

and smooth out the resulting fields. (Based on L253-255 this does not seem to be a bad idea).

We thank the reviewer for this thoughtful comment. The fitted inverse scale height (X) in the DDA framework serves a different purpose than boundary layer height. Rather than representing the physical mixing depth alone, X captures the integrated vertical structure of the observed column, which extends to the top of the atmosphere and also encodes the vertical sensitivity of each instrument. In this sense, X acts as an empirical correction that improves the fidelity of the DD estimator—particularly in heterogeneous terrain and under variable atmospheric conditions. The data-driven approach of fitting X can also absorb some of the errors caused by the simplifications and the choice of near-surface wind. The rationale of these fittings is to explain the residual values of the DD over locations where the emissions are negligible. We have clarified that the data-driven fitting approach aims to enhance the DDA estimators in Section 2.2.1, which is detailed in our response to main comment #5. Specifically, to improve the performance of the flux estimates, we treat the X and k as fundamentally empirical fitting parameters within a data-driven approach to ensure the resultant emission estimator centers around zero where emissions are negligible.

7. Chemical Loss Term — dropping the lifetime term because of a bad fit seems a bit easy and one that potentially has a large impact on the resulting emission and deposition fluxes, especially when moving away from the strongest emission gradients.

We thank the reviewer for this important insight. We agree that omitting the chemical loss term due to a poor fit is a debatable choice, particularly in areas with abundant columns. In response to this concern, we have revised our analysis to retain the chemistry term in the final flux estimator, using the DD_chem estimator as our primary flux estimate throughout the manuscript.

According to Ayazpour et al. (2024), a stricter maximum emission threshold is necessary for fitting *DD_chem* than for *DD_topo*. Could the poor fitting performance of the chemical loss term (line 247) be related to an insufficiently strict threshold? Please clarify and discuss whether further refinement of the X and k estimates is planned.

We conducted sensitivity tests to evaluate the fitted X and k, w. For X fitting, we tested both spatially and temporally varying fittings, as described in our previous response to main comment #5. For k fitting, we tested spatially and temporally varying k, as well as stricter maximum emission thresholds, as suggested by the referee. The spatial and temporal variability tests indicate that allowing k to vary does not improve the fitting performance (Figs. S4-S5). Likewise, applying stricter emission thresholds for k fitting does not improve the fitting performance (Fig. S6). Nevertheless, changing the maximum emission threshold from 1×10^{-9} to 3×10^{-10} mol/s alters the area-integrated emission rates

by less than ~7% across regions, indicating that the results are relatively insensitive to this parameter choice. We have added the sensitivity tests of fitted k in the end of Section 2.2.1 as follows: *"The same settings were applied to k for consistency, though its performance was largely unaffected by these changes (Figs. S4-S5). Stricter emission thresholds for the chemistry term also had little impact (Fig. S6)"*.

The following figures have been added to the Supplementary Materials:

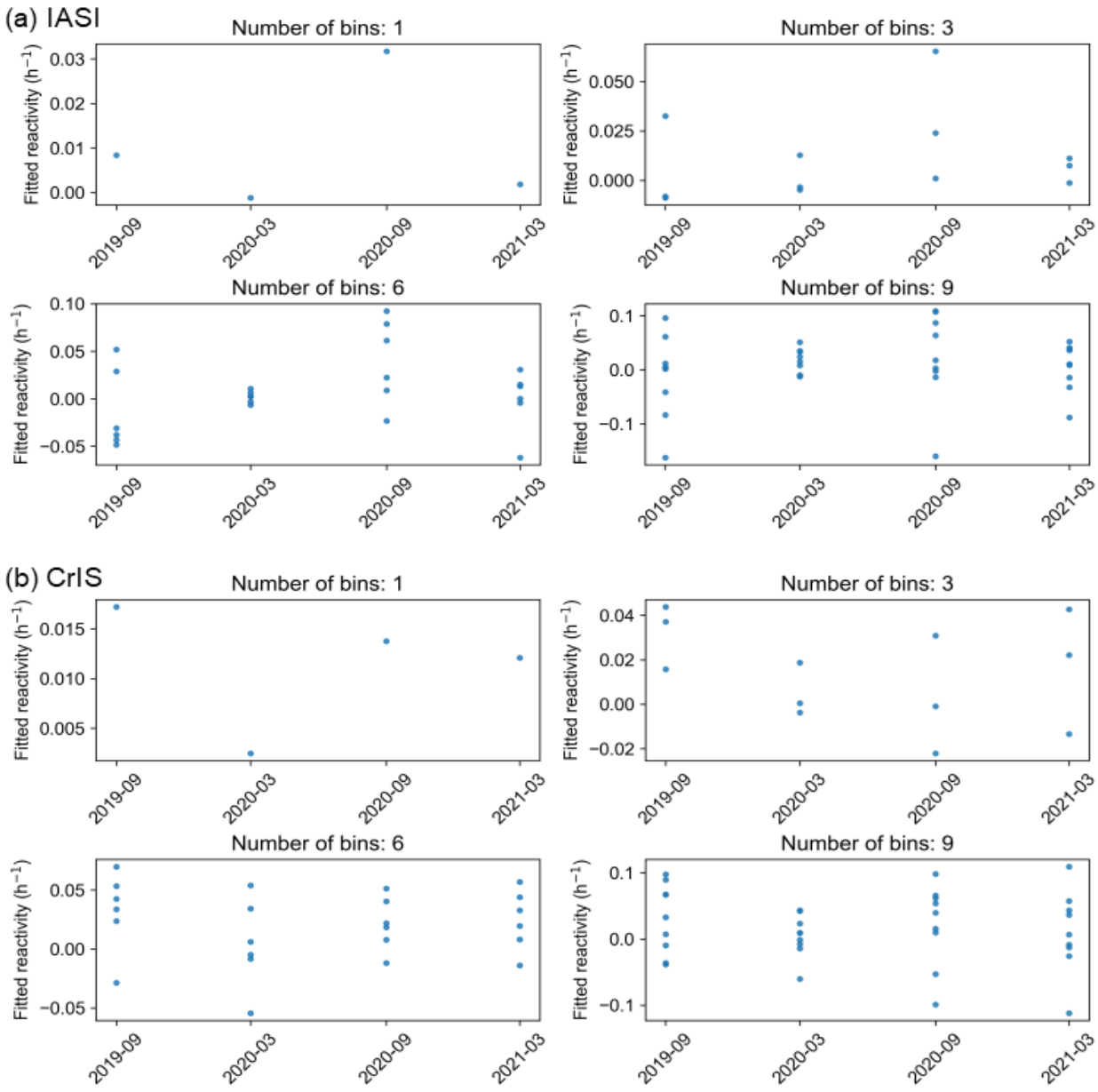

Figure S4. Spatial sensitivity tests for the topography term fitting to derive DD_chem from IASI (a) and CrIS (b). The number of bins indicates the column (Ω) sectors used in the fitting. The y-axis shows the fitted k.

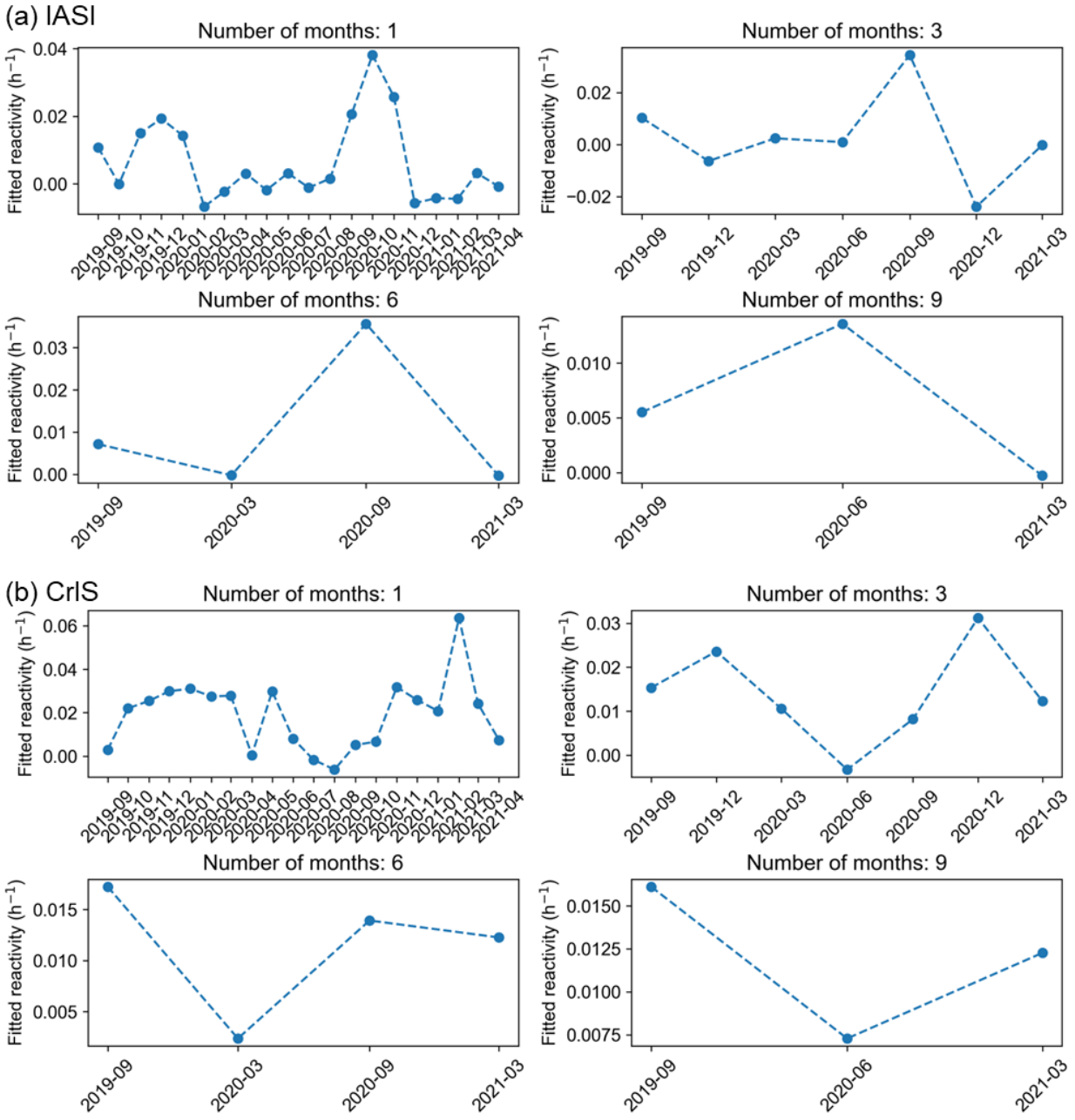

Figure S5. Temporal sensitivity tests for the chemistry term fitting to derive DD_chem from IASI (a) and CrIS (b). The number of months indicates the temporal aggregation interval used for the fitting. The y-axis shows the fitted k.

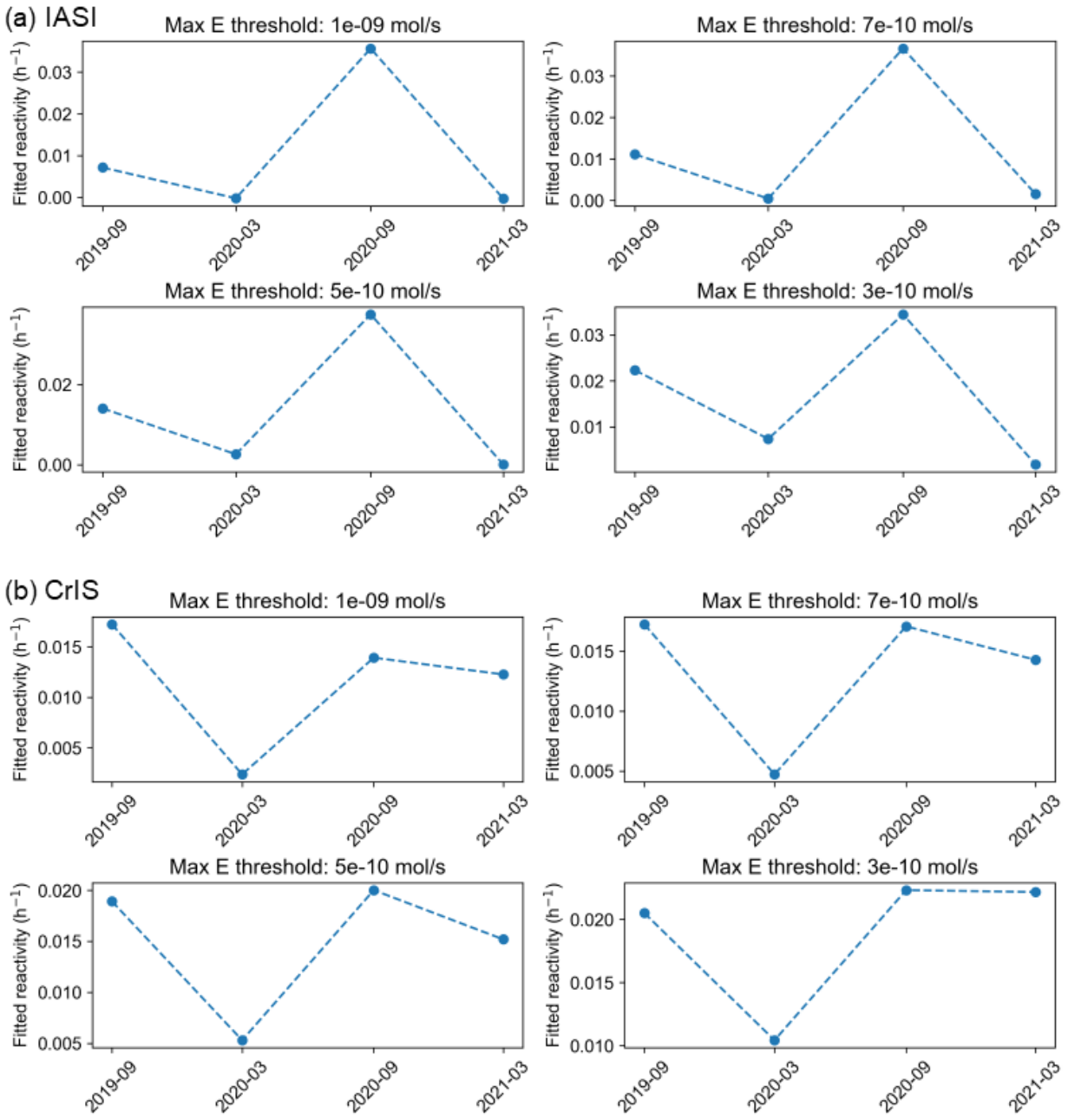

Figure S6. Maximum emission (E) threshold sensitivity tests for the chemistry term fitting to derive DD_chem from IASI (a) and CrIS (b). The y-axis shows the fitted k.

Additionally, the fit seems to have been performed on the whole of the CONUS, whereas lifetime will vary strongly depending on the local pollution levels of other species (produced hno3/h2so4). A switch to locally varying fits would make sense from a chemistry point of view.

Thanks for the suggestion. We have addressed this concern in our previous response to main comment #7. Specifically, we have tested spatially and temporally varying k fittings, and the fitting performance was not improved by these changes (Figs. S4-S5). Although in theory k is tied to physically meaningful quantities, its main purpose is to enhance emission estimators in the presence of chemistry using information contained by satellite observations. As such, the quality of k fits is subject to the quality and quantity of available satellite data. Given the data gaps in individual IASI and CrIS pixels, as well as the reduced sample size when applying varying fits, introducing additional spatial or temporal variations led to larger noise in the fitted values without improving performance.

Alternatively, the authors could add an Alinea what the expected lifetime to chemistry is for typical hno3/h2so4 concentrations, and discuss from that point of view if chemistry is important or not. In its current form it's not convincing.

Thanks for the suggestion. As detailed in our response to main comment #7, we conducted sensitivity tests on chemistry term fitting.In the revised manuscript, we have included the chemistry term and use DD_chem estimator as our flux estimates. We do agree that omitting the chemistry term is debatable, and its spatial pattern indicates that it plays a larger role in column-abundant regions (Fig. S8c, d). The following figure has been added to the Supplementary Materials:

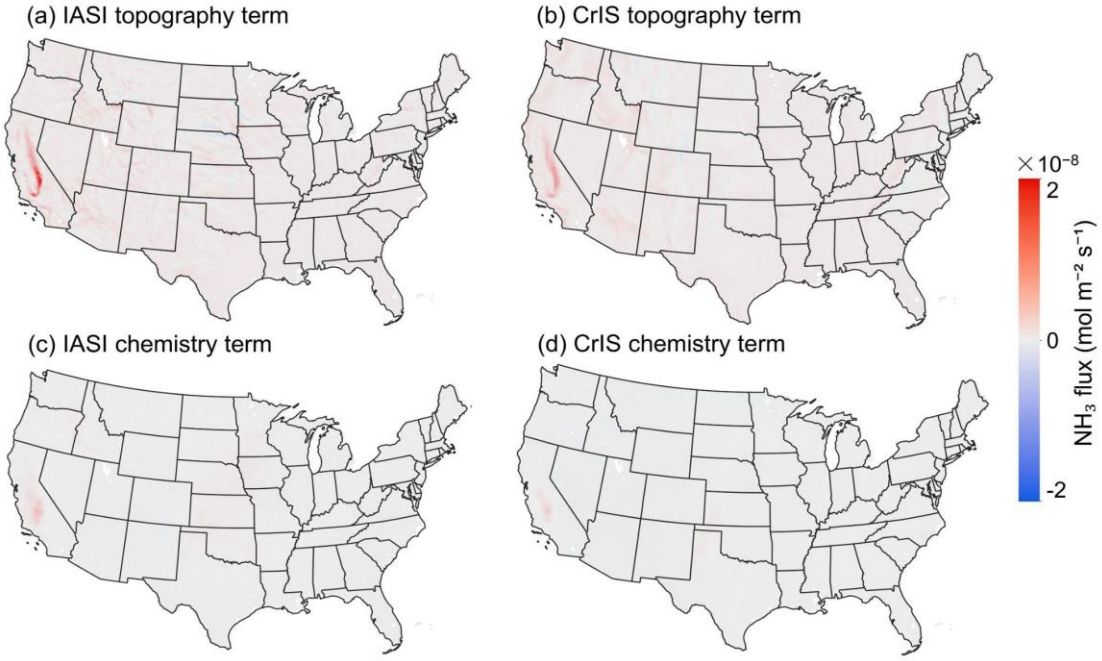

Figure S8. Topography and chemistry terms from IASI and CrIS over Sep 2019 to Apr 2021.

1. Daily Data and Bias Correction — Using only daytime data may overestimate NH_3 due to satellite sampling biases. Have you considered applying a satellite bias correction, as recommended in Ayazpour et al. (2024, Sect. 2.2.4)?

We thank the referee for the comment. It is true that using only daytime satellite observations may overestimate NH_3 levels due to diurnal variability and the lack of nighttime measurements. However, this limitation is inherent to thermal infrared sounders like IASI and CrIS, which only provide good quality data under clear-sky conditions with adequate thermal contrast between the surface and the lower atmosphere. We acknowledge our DDA-based NH_3 flux estimates are based on daytime data and reflect daytime-effective fluxes, and we added discussion about this limitation in Section 4.1 (Lines 446-448): *"retrievals are limited to daytime clear-sky conditions, potentially biasing results due to NH_3's strong diurnal cycle (Blanes-Vidal et al., 2008)".*

Section 2.2.4 of Ayazpour et al. (2025) aims to address retrieval biases related to spatial artifacts (e.g., surface reflectance), not temporal sampling biases. They proposed their use as a potential future enhancement, particularly for longer-lived species like CO, CH_4, and CO_2. We do not include similar bias-sensitive variables in our NH_3 estimates. This is because (1) there is no clear evidence of systematic spatial biases in our NH_3 retrievals that could be robustly modeled via ∇a-type predictor fields, and (2) our regression fitting over low-emission regions inherently absorbs baseline retrieval errors into the fitted parameters for topography and chemistry. Since our primary goal is to estimate NH_3 flux rather than diagnose retrieval errors, we focus on the DD_chem estimator. We acknowledge the potential for retrieval biases and the value of incorporating more advanced correction terms in future studies, as suggested by Ayazpour et al. (2025).

2. Line 301-305: I completely disagree that the directional derivative minimizes the impacts of offsets and scaling differences between the products. As the authors know the bias in the satellite products are not spatially and temporally independent, which means that any offset/scaling difference will vary from point to point, especially around larger shifts in concentration levels. If anything the scaling between the satellite product biases will be enhanced by the derivative. Additionally, the detection limit of both satellites will of course also play an important role in the limitations to detectable gradients. For example the early spring peaks detected by CrIS but not by IASI could be a sign of detection limit (or a strong diurnal variability in emissions of course). Some validation studies are available for both satellite products. These can be used as a basis for an error/uncertainty estimate of the effects. Please show what is the expected uncertainty of the current product bias/scaling on the resulting fluxes. An update to the limit used later in the manuscript (i.e. +-2σ) might be needed.

We thank the referee for raising these insightful points. We have revised the relevant text in Section 3.2 (Lines 299-305) as follows: "*The VCD comparison reveals systematic differences between IASI and CrIS as different slopes and offsets in different regions (Fig. 3a). These offsets and proportional biases can propagate into flux calculations, where they manifest as amplified variability and reduced correlation (Fig. 3b). The lower correlation in flux estimates compared to VCDs likely reflects compounded noises from derivative-based flux estimates, which relies on external datasets, assumptions, and signal differentiation.*"

We acknowledge the potential for product-specific biases to affect flux estimates, particularly near the detection limits of CrIS and IASI. The detection limit of NH_3 depends on both thermal contrast and the vertical distribution of NH_3. We have added the uncertainties of IASI and CrIS NH_3 VCDs from previous validation studies in Section 4.1 (Line 448): *"IASI columns differ from ground-based measurements by −32 ± 56% (Dammers et al., 2016), with errors highly dependent on thermal contrast (Van Damme et al., 2014). CrIS retrievals show ~10 – 30% error in total columns and larger uncertainty at low concentrations (Shephard et al., 2020)."*

To mitigate these effects, we focus on hotspot source and sink regions with relatively strong NH_3 signals, where retrieval uncertainties are lower. We also average over multi-month periods and constrain fitting procedures to low-emission regions where the signal-to-noise ratio is more stable and biases are more consistent. Regarding the ±2σ threshold, we have not revised this threshold at this stage, as the regional-scale aggregation and conservative filtering already help reduce the influence of noises. Future work incorporating satellite-specific error propagation into the DDA framework could further refine this threshold.

3. Suggestion for further validation: The comparison with inventories is interesting, but more direct evidence of the value of satellite based emission and deposition estimates would be a comparison with in-situ data, which currently is missing. Feeding back the emissions into a CTM and comparing the resulting concentrations with in-situ data would strengthen the case that satellite based emissions are an improvement over current inventories. Additionally, a comparison of the simulated deposition data (based on the updated emissions) with the satellite derived estimates would further show the value of those deposition estimates. It is quite an effort though so I would understand if the authors state it's beyond the scope of this manuscript.

We thank the referee for the thoughtful suggestion. We agree that using CTM and in-situ measurements for validation would provide valuable, independent evidence of the utility of satellite-based estimates. However, such model integration and evaluation require substantial computational effort, detailed configuration, and rigorous validation protocols,

which are beyond the scope of the current study. We encourage future studies to explore this direction.

Minor edits/comments:

Abstract L29: "atmospheric" instead of "atmosphere"

We've revised as suggested.

L45-46, I'd rephrase this sentence. Spatially the emissions seem to align, but seasonally and in amplitude they do not.

Lines 45-46 changed to: *"Compared with bottom-up inventory, satellite-based estimates capture general spatial and seasonal patterns, while also revealing additional insights into key flux hotspots and peak periods."*

L74: a few hours is on the low end of the model and measurement estimates, mostly derived from direct fits on satellite data, which are expected to bias low. Estimates of 8-12 or up to 24 hours seems more reasonable based on literature.

Line 74 changed to: *"less than 24 hours".*

L84: Quite a recent reference, the relation between volatilization and environmental conditions was known much before this point.

Line 84 changed reference to: *"(Sommer et al., 1991)"*.

L103-105: Quite the claim when later analysis mostly focuses on monthly or longer temporal resolutions.

Lines 103-105 changed to: "*We derived top-down NH_3 fluxes at 0.1° resolution using observations from two space-based instruments: the Infrared Atmospheric Sounding Interferometer (IASI) and the Cross-track Infrared Sounder (CrIS)*".

L124 onward: add the observational periods of each satellite after each satellite name, this will make it easier for the reader to follow what satellite is in orbit when.

Line 123 added: *"from Metop-A (Jan 2008 to Oct 2021), Metop-B (Mar 2013 to Dec 2022) and Metop-C (Sep 2019 to Dec 2022)".*

Line 132 changed to: *"CrIS-NPP (Jun 2012 to May 2021) and CrIS-NOAA-20 (Mar 2019 to Dec 2022)".*

Line 182-185: Essentially you are gap-filling the record, but I fail to see the basis for just inflating the pixel size without any smart input of additional data. I can imagine this type of gap-filling introducing stronger or weakening gradients in regions with very localized sources and/or very common wind directions. Please add a few words on potential effects on gradients.

Lines 182-185 changed to: "*This is a critical step to enable spatial gradient calculation when data coverage is sparse (e.g., only a single overpass is available). This also limits the spatial resolving power to about twice the pixel size*".

Figure 1: Whats going on with the few outlier months in the CrIS and IASI records, are these specific periods? And what does excluding these do for your results?

In Fig. 1, a few outlier months with higher-than-expected random errors correspond to periods of low data coverage caused by cloudiness or partial instrument downtime. These months constitute a small fraction of the dataset and have minimal influence on the results, as low-coverage periods are down-weighted in the aggregation.

Line 313-315: what about the instrument detection limit?

We have added "*and differing detection limits*" to Lines 313-315 and moved this sentence to Section 4.1 (Line 428).

Section 4.2: or in discussion: I miss a discussion on the potential effects of instrument/product bias changing over time, and the expected impact compared to the increasing trends you observed here.

We thank the reviewer for this comment. Both IASI and CrIS have operated with multiple instruments over time, but existing studies indicate good temporal consistency For IASI, the ANNI-NH_3 products include empirical corrections to address small instrument- and processing-related differences (Clarisse et al., 2023) and validation shows good agreement with independent datasets in multiple regions (Van Damme et al., 2015). For CrIS, long-term spectral stability is high (Chen et al., 2014) and validation shows strong correlation and minimal bias (Dammers et al., 2017). We therefore consider any time-dependent bias too subtle for the multi-year increases reported here to be attributed to bias drift. To address this potential limitation, we have added the following text to Section 4.1 (Line 448): "*While both IASI and CrIS have demonstrated good long-term stability through calibration monitoring and ground-based validation (Chen et al., 2014; Clarisse et al., 2023; Dammers et al., 2017; Van Damme et al., 2015), subtle time-dependent biases cannot be fully excluded.*"

References:

Ayazpour, Z., Sun, K., Zhang, R., and Shen, H.: Evaluation of the directional derivative approach for timely and accurate satellite-based emission estimation using chemical transport model simulation of nitrogen oxides, J. Geophys. Res., 130, e2024JD042817, https://doi.org/10.1029/2024jd042817, 2025.

Blanes-Vidal, V., Hansen, M. N., Pedersen, S., and Rom, H. B.: Emissions of ammonia, methane and nitrous oxide from pig houses and slurry: Effects of rooting material, animal activity and ventilation flow, Agric. Ecosyst. Environ., 124, 237–244, https://doi.org/10.1016/j.agee.2007.10.002, 2008.

Chen, Y., Han, Y., Jin, X., and Weng, F.: Assessment of S-NPP CrIS Spectral Calibration Accuracy and Stability, the 94 th AMS Annual Meeting, 2014.

Clarisse, L., Franco, B., Van Damme, M., Di Gioacchino, T., Hadji-Lazaro, J., Whitburn, S., Noppen, L., Hurtmans, D., Clerbaux, C., and Coheur, P.: The IASI NH 3 version 4 product: averaging kernels and improved consistency, Atmos. Meas. Tech., 16, 5009–5028, https://doi.org/10.5194/amt-16-5009-2023, 2023.

Dammers, E., Palm, M., Van Damme, M., Vigouroux, C., Smale, D., Conway, S., Toon, G. C., Jones, N., Nussbaumer, E., Warneke, T., Petri, C., Clarisse, L., Clerbaux, C., Hermans, C., Lutsch, E., Strong, K., Hannigan, J. W., Nakajima, H., Morino, I., Herrera, B., Stremme, W., Grutter, M., Schaap, M., Wichink Kruit, R. J., Notholt, J., Coheur, P.-F., and Erisman, J. W.: An evaluation of IASI-NH_3 with ground-based Fourier transform infrared spectroscopy measurements, Atmos. Chem. Phys., 16, 10351–10368, https://doi.org/10.5194/acp-16-10351-2016, 2016.

Dammers, E., Shephard, M. W., Palm, M., Cady-Pereira, K., Capps, S., Lutsch, E., Strong, K., Hannigan, J. W., Ortega, I., Toon, G. C., Stremme, W., Grutter, M., Jones, N., Smale, D., Siemons, J., Hrpcek, K., Tremblay, D., Schaap, M., Notholt, J., and Erisman, J. W.: Validation of the CrIS fast physical NH_3 retrieval with ground-based FTIR, Atmos. Meas. Tech., 10, 2645–2667, https://doi.org/10.5194/amt-10-2645-2017, 2017.

Ellis, R. A., Jacob, D. J., Sulprizio, M. P., Zhang, L., Holmes, C. D., Schichtel, B. A., Blett, T., Porter, E., Pardo, L. H., and Lynch, J. A.: Present and future nitrogen deposition to national parks in the United States: critical load exceedances, Atmos. Chem. Phys., 13, 9083–9095, https://doi.org/10.5194/acp-13-9083-2013, 2013.

Hu, C., Griffis, T. J., Frie, A., Baker, J. M., Wood, J. D., Millet, D. B., Yu, Z., Yu, X., and Czarnetzki, A. C.: A multiyear constraint on ammonia emissions and deposition within the US corn belt, Geophys. Res. Lett., 48, e2020GL090865, https://doi.org/10.1029/2020gl090865, 2021.

Lonsdale, C. R. and Sun, K.: Nitrogen oxides emissions from selected cities in North America, Europe, and East Asia observed by the TROPOspheric Monitoring Instrument (TROPOMI)

before and after the COVID-19 pandemic, Atmos. Chem. Phys., 23, 8727–8748, https://doi.org/10.5194/acp-23-8727-2023, 2023.

Shephard, M. W., Dammers, E., Cady-Pereira, K. E., Kharol, S. K., Thompson, J., Gainariu-Matz, Y., Zhang, J., McLinden, C. A., Kovachik, A., Moran, M., Bittman, S., Sioris, C. E., Griffin, D., Alvarado, M. J., Lonsdale, C., Savic-Jovcic, V., and Zheng, Q.: Ammonia measurements from space with the Cross-track Infrared Sounder: characteristics and applications, Atmos. Chem. Phys., 20, 2277–2302, https://doi.org/10.5194/acp-20-2277-2020, 2020.

Sommer, S. G., Olesen, J. E., and Christensen, B. T.: Effects of temperature, wind speed and air humidity on ammonia volatilization from surface applied cattle slurry, J. Agric. Sci., 117, 91–100, https://doi.org/10.1017/S0021859600079016, 1991.

Sutton, M., Reis, S., and Baker, S. (Eds.): Atmospheric Ammonia: Detecting emission changes and environmental impacts. Results of an Expert Workshop under the Convention on Long-range Transboundary Air Pollution, 1st ed., Springer, New York, NY, 464 pp., https://doi.org/10.1007/978-1-4020-9121-6, 2009.

Van Damme, M., Clarisse, L., Heald, C. L., Hurtmans, D., Ngadi, Y., Clerbaux, C., Dolman, A. J., Erisman, J. W., and Coheur, P. F.: Global distributions, time series and error characterization of atmospheric ammonia (NH3) from IASI satellite observations, Atmos. Chem. Phys., 14, 2905–2922, https://doi.org/10.5194/acp-14-2905-2014, 2014.

Van Damme, M., Clarisse, L., Dammers, E., Liu, X., Nowak, J. B., Clerbaux, C., Flechard, C. R., Galy-Lacaux, C., Xu, W., Neuman, J. A., Tang, Y. S., Sutton, M. A., Erisman, J. W., and Coheur, P. F.: Towards validation of ammonia (NH3) measurements from the IASI satellite, Atmos. Meas. Tech., 8, 1575–1591, https://doi.org/10.5194/amt-8-1575-2015, 2015.

Zhang, L., Jacob, D. J., Knipping, E. M., Kumar, N., Munger, J. W., Carouge, C. C., van Donkelaar, A., Wang, Y. X., and Chen, D.: Nitrogen deposition to the United States: distribution, sources, and processes, Atmos. Chem. Phys., 12, 4539–4554, https://doi.org/10.5194/acp-12-4539-2012, 2012.