We thank the reviewers for their constructive comments and suggestions to improve the quality and clarity of our manuscript. We have made major and careful modifications to the original manuscript according to all the comments and suggestions from the reviewers.

Item-by-item responses to the specific comments are provided below, in which the reviews' comments are in blue, our responses in **black**, and modifications of the original manuscript are indicated by red in the revised manuscript.

Reviewer #1:

The manuscript 'Global and diurnal variations in tropospheric ammonia observed from a constellation of hyperspectral infrared sounders in three different LEO orbits' by Hua et al. presents an evaluation of the global spatial and diurnal variability of NH3 based on observations from the HIRAS-II instruments aboard the Chinese FY-3E and FY-3F satellites along with observations from CrIS on JPSS-1.

The authors utilize a harmonized retrieval approach and the differing overpasses of these three satellites to provide unique insights into the diurnal variability of NH3 which remains broadly uncertain and was previously challenging to capture globally from the existing polar-orbiting instruments. The NH3 observations from these three satellites are combined to yield a quasi-geostationary dataset with observations made roughly every 4 hours throughout the day. Furthermore, a comparison is performed of each polar-orbiting dataset against true geostationary NH3 observations from GIIRS, which serves to provide a consistent intercomparison point. The authors demonstrate that their quasi-geostationary constellation is capable of resolving the seasonal and diurnal variability globally and in several high-emission regions of interest.

Overall, the manuscript is well written and the results are presented in a clear manner. I believe that the manuscript fits well within the scope of AMT. There are some sections of the manuscript that can be improved both in terms of clarity and level of detail, but after these revisions are made, I recommend it for publication.

We thank the referee for the positive evaluation and constructive comments. We have revised the manuscript substantially to improve the descriptions of the methodology and strengthen the robustness of the results. The main revisions include: (1) expanding the retrieval-method section to better document the profile-scaling approach, the selection of the a priori profile, the state vector, and the cross-sensor treatment; (2) adding a priori profile sensitivity analysis; (3) quantifying the impact of quality filters on the number of observations for further analysis; (4) expanding the FY-4B/GIIRS cross-comparison to representative months/seasons with additional statistics, including sample size, intercept, mean bias, and ODR fitting information; and (5) adding comparisons with ground-based FTIR measurements.

Major comments

(1) Section 2.2 (Optimal estimation methods) the authors describe some deviations in their retrieval methods from the previous studies of Zeng et al. (2023) including using a single-value column scaling retrieval as opposed to a

full OEM profile retrieval. However more details are needed here and I have some concerns on the impacts of such choices.

For example, on L127 - L128, the authors state "A profile scaling approach is used to retrieve a single scale factor applied to a fixed a priori NH3 profile." and later that "A fixed a priori profile instead of a variable one is adopted, so that any spatiotemporal changes in retrieved

NH3 column concentrations reflect only information from satellite observations, rather than variations in the prior". However the authors do not describe how the initial prior profile was selected. Additionally, using the same fixed profile shape for both daytime and nighttime retrievals would likely bias the retrieved ammonia total column high or low.

Since this study being considered for AMT and is also partly a presentation of these harmonized retrievals and with a modified approach from previous studies, I think more details should be provided on, for example, the choice of a priori profile shape and how much that impacts the resulting total columns (and potentially also the diurnal cycles). This could be included in the appendix, but this is important to support the robustness of the remainder of the results presented in the manuscript.

Furthermore, a comparison of the retrievals with a ground-based dataset like an FTIR such as the one at Hefei (https://doi.org/10.1016/j.atmosenv.2022.119256) could strengthen the case for the satellite retrievals even more. This could be included as part of Section 3.3 and leads me to my next comment.

We thank the referee for this important comment. We have substantially revised Section 2.2 to make the retrieval description more self-contained, by providing sufficient detail on the modified retrieval setup, especially the use of the profile-scaling approach and the construction of the fixed NH_3 a priori profile.

In the revised manuscript, we now clarify that FY-LeoAIR follows the optimal-estimation framework and forward-model structure of FY-GeoAIR developed by Zeng et al. (2023). Specifically, the retrieval uses the same 955–975 cm^{-1} band, similar auxiliary datasets, the UOW-M land surface emissivity database as the forward-model emissivity, and the GEOS-CF-based fixed NH_3 a priori profile following Zeng et al. (2023). The fixed NH_3 a priori profile is derived from GEOS-CF simulations over representative polluted land regions in East Asia and South Asia, as in Zeng et al. (2023). The parameters in the retrieval state vector are summarized in Table S1.

We now explicitly describe the main modification relative to Zeng et al. (2023): NH_3 is retrieved as one multiplicative profile-scaling factor rather than as independent layer-by-layer NH_3 elements. This modification is motivated by the limited NH_3 information content in individual thermal infrared observations, with the degree of freedom for signal generally below 1, and by the need for computationally efficient global multi-sensor processing. In this implementation, H_2O is also retrieved using one multiplicative profile-scaling factor.

We also add a detailed description of the retrieval state vector. The state vector includes the NH_3 profile-scaling factor, the H_2O profile-scaling factor, scaling factors for weakly interfering trace gases, surface skin temperature, atmospheric temperature scaling, and four surface-emissivity adjustment coefficients based on first- to fourth-order Legendre polynomial terms. The UOW-M emissivity spectrum is used as the forward-model emissivity prior and is adjusted within the retrieval window through these four Legendre-polynomial coefficients.

We further clarify the multi-sensor treatment. The same retrieval window, forward model, auxiliary datasets, and quality-control procedure are applied to FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS. Sensor-specific spectral noise estimates are used in the measurement-error covariance matrix. No empirical sensor-specific radiometric bias correction or additional spectral harmonization is applied.

We adopt a fixed a priori profile in the retrieval to reduce the possibility that retrieve spatiotemporal variability is inherited directly from a time-varying model prior. However, we agree with the referee that using the same fixed profile shape for different local-time periods could potentially affect the retrieved total columns, especially under low-sensitivity conditions. To address this concern, we add a new a priori profile sensitivity test in the supplementary material. In this sensitivity test, we repeat the retrievals using a variable GEOS-CF-based NH_3 a priori profile. For each satellite observation, the GEOS-CF NH_3 profile closest in space and time to the observation is selected as the a priori profile shape. We then compare these variable-prior retrievals with the baseline fixed-prior retrievals over the North China Plain and the Indo-Gangetic Plain for January, April, July, and October. For each month, three representative days, namely the 5th, 15th, and 25th, are used. The same strict quality-control and sensitivity filters are applied to both retrieval sets before comparison. Overall, the fixed-prior and variable-prior retrievals show broadly consistent NH_3 columns after strict filtering, indicating that the main spatial and local-time-dependent patterns discussed in this study are not primarily controlled by the choice of fixed versus variable NH_3 a priori profile.

The sensitivity test shows that the fixed-prior and variable-prior retrievals are broadly consistent after strict filtering. The differences are generally small compared with the range of the retrieved columns, and no evident systematic local-time-dependent bias caused by the fixed-prior assumption is found in the tested regions and months. We have added this analysis to the supplementary material and added corresponding discussion in Sect. 2.2 and the results section. We also revised the wording in the manuscript to state more cautiously that the fixed prior reduces prior-driven variability, rather than implying that the retrieval is completely independent of the assumed profile shape.

We also add an independent comparison with ground-based FTIR NH_3 column measurements at Hefei, China, shown in Fig. S14.

(2) Section 3.3 (Comparisons with geostationary NH3 observations from GIIRS) is light on details and could be expanded to better match the scope of the remainder of the manuscript.

The paper places a significant focus on seasonal variability, but the comparisons presented between FY-3E, FY-3F and CrIS versus FY-4B/GIIRS are only shown on the yearly scale. It would be valuable to also show the comparisons monthly and/or seasonally to help identify whether there are any seasonal biases or inconsistencies across the satellite products, as these may be covered up or averaged out when looking at the correlations at the yearly scale.

Additionally, there is a pretty consistent slope on the order of 0.86-0.89 in almost all of the comparisons of FY-3E, FY-3F and CrIS with GIIRS but this is not discussed. Can the authors provide some potential explanations? Is this due to some systematic bias in the GIIRS product and/or the products retrieved in this study?

We agree that the original Section 3.3 was too brief and that annual-scale statistics may mask seasonal or local-time-dependent differences. We have expanded this section by adding monthly/seasonal collocation statistics between FY-3E/HIRAS-II, FY-3F/HIRAS-II, CrIS, and FY-4B/GIIRS. The revised comparison now reports the number of collocations, slope, intercept, correlation coefficient, RMSE, and mean bias.

We also add a discussion of the regression slopes that are generally lower than unity. We do not attribute this behavior to a single systematic bias source. Instead, we discuss several possible contributors, including differences in retrieval algorithms, viewing geometry, footprint size, cloud screening, vertical sensitivity, prior constraints, and representativeness mismatch within the collocation window. We have also replaced “validate” with “evaluate consistency” or “cross-compare” throughout Section 3.3.

(3) In Section 3.1 the authors describe a rigorous filtering approach for the data including filters based on surface temperature, thermal contrast, emissivity and the averaging kernel values. However, the impact of the filtering is not described. On average how many observations are removed/excluded based on the filters?

Providing more information on the impact of the filters is important, particularly since the paper presents night-time retrievals which typically are far more challenging due to poorer observational conditions, and these filters likely impact the data density. Additionally, it seems a different filtering criterion is applied in Section 3.2, but it is not immediately clear to me why a different criterion is necessitated and also if it is applied on top of the earlier criterion or not.

We thank the referee for this helpful comment. We have revised the manuscript to clarify the purpose of the different filtering criteria and to quantify their impact on data retention. The absolute TC> 3 K and surface AVK > 0.1 criteria used in Section 3.1 are applied to the global spatial-distribution maps, where the goal is to demonstrate broad constellation coverage and global NH_3 spatial patterns while retaining sufficient sampling density. For the hotspot diurnal analysis in Section 3.2, we apply stricter sensitivity criteria, tightening the thresholds to absolute TC > 5 K and surface AVK > 0.3. These stricter criteria are used to reduce the influence of local-time-dependent retrieval sensitivity on the inferred diurnal cycles, because TC itself has a pronounced diurnal variation.

To quantify the impact of filtering, we add a sensitivity analysis using three filtering configurations. A loose filter with surface AVK > 0.1 and absolute TC > 3 K, the current hotspot filter with surface AVK > 0.3 and absolute TC > 5 K, and a strict filter with surface AVK > 0.4 and absolute TC > 7 K. The remaining quality-control criteria are kept consistent among the three configurations. The resulting NH_3 spatial distributions over the North China Plain and the Indo-Gangetic Plain, the corresponding diurnal cycles, and the total numbers of retained quality-controlled retrievals for each satellite are now provided in the supplementary material.

The sensitivity analysis shows that stricter filtering reduces the number of valid retrievals, as expected, but the main spatial hotspot patterns remain similar and the regional diurnal cycles are not substantially distorted. We have added this discussion to Sections. 3.1–3.2 and the supplementary material.

(4) L151-152: Why were periods of only one week used for the seasonal variability investigation? Why not use the full seasons? The reasoning behind only choosing one week periods is not provided and by including longer periods, it would give a more robust picture of the true seasonal variability.

By limiting it to single weeks, the results are more subject to things such as transient events (e.g., biomass burning, manure application etc) that could skew or bias the seasonal results. This is already apparent in the results, for example the large wildfire over North America in July. I suggest revising the seasonal analysis to include averages over more months. Later on in Section 3.2 full months are seemingly used, so this would make it more consistent with that Section and the results there as well.

We thank the referee for this helpful comment. We agree that one-week averages cannot fully represent robust seasonal means and may be influenced by short-lived episodic events, such as biomass burning, fertilizer application, or wildfire plumes. Our intention in Figs. 2–3 is not to derive climatological seasonal averages, but to provide an illustrative demonstration of the global sampling capability of the three-satellite LEO constellation and to show representative global NH_3 spatial distributions at six local overpass times.

To avoid overinterpreting these weekly maps, we have revised Section 3.1 to clarify that Figs. 2–3 are representative seven-day examples rather than full seasonal means. We now explicitly state that these maps are intended to demonstrate the constellation-derived global coverage and broad spatial contrasts, and that they may still be affected by transient events. Accordingly, we have revised statements referring to "seasonal variability" in Section 3.1 to "representative weekly global distributions" or "broad seasonal contrasts." The figure captions have also been revised.

We further clarify that the quantitative regional diurnal and seasonal analyses in Section 3.2 are based on full-month averages using all available observations in each selected month. Therefore, while Figs. 2–3 serve as illustrative global examples, the subsequent hotspot analyses are based on more robust monthly statistics.

Minor comments:

(1) L41: Dammers et al., 2017 is cited here for the general background on NH3, but the references therein are likely more appropriate for citation here as the focus of that paper is on satellite validation.

We agree that Dammers et al. (2017) is more focused on satellite validation and is not the most appropriate citation for general NH_3 background. We have revised the citation and now cite references that more directly address NH_3 sources, atmospheric chemistry, and environmental impacts.

(2) L47-48: "A number of in situ monitoring sites equipped with Fourier transform infrared (FTIR) spectrometers have been established...", there are also ground/surface measurement stations that make in-situ measurements that are not FTIRS e.g., via DOAS measurements or passive samplers. It is not also 100% correct to call an FTIR site an "in situ" site, since they technically measure on a very long slant-path through the atmosphere. In situ methods for NH3 are typically made at the point via an inlet based or some other measurement method.

We thank the referee for this clarification. We have deleted the phase “equipped with Fourier transform infrared (FTIR) spectrometers”.

(3) L63-65: "However, the lack of geostationary orbit observations across most global regions especially in the Southern Hemisphere prevents multiple daily observations from being achieved." the wording of this sentence is a bit strange. "Multiple daily observations" can technically be achieved with 1 to 2 polar orbiting satellites no?

We agree that the original wording was imprecise. We revise the sentence to clarify that multiple daily observations can be achieved with polar-orbiting satellites, but dense geostationary-like diurnal sampling remains unavailable over most global regions.

(4) General comment on Section 1: was there a reason to exclude IASI from the main analysis? Was it just that its overpass time was too close to FY-3F and thus did not provide more info? Or something else? IASI is utilized at some points in the analysis as a comparison point (e.g., in Figure 4), but it isn't really mentioned on why its excluded from the main analysis.

We thank the referee for pointing this out. We have added a clarification in Section 2.1. “Note that IASI’s overpass times, approximately 09:30 and 21:30 LST, are close to those of FY-3F/HIRAS-II at 10:00 and 22:00 LST and thus can provide additional mid-morning orbit observations. For this concept study, however, only FY-3F/HIRAS-II is used.”

L82: "It detects upwelling infrared radiative signals across the short-wave.." the phrasing and word choice is a bit strange here. I would probably rephrase it as "It measures upwelling infrared radiation across the short-wave...".

Revised as suggested.

L84-87: very minor comment, but this sentence describing the overpasses of the instruments is slightly repetitive with the second last paragraph of Section 1. In my opinion, the description of the overpass times definitely fits better here in the Methodology section, but if the authors feel it is important to also keep it in Section 1 then it is fine as well.

We reduce the repeated description of satellite overpass times in the Introduction and kept the detailed description in Section 2.1, where it fits better methodologically.

L93: Why was the 930 cm-1 NH3 band not utilized in the retrievals? Was the performance less consistent across the instruments for this band? Or was there a different reason?

We thank the referee for raising this point. We have added an explanation for the choice of the retrieval window. Although NH_3 has spectral features near 930 cm^{-1}, this band is not used because the present study focuses on a harmonized multi-sensor retrieval. The 955–975 cm^{-1} band provides more stable spectral residuals and more consistent sensitivity across FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS.

L117-119: the authors do not describe what the actual cloud-filtering criteria/threshold was that they applied here. Was it also the same exact criterion of <0.4 that was used in Wells et al. (2020)? Even if so, it should be mentioned here in the text. Additionally, I am curious how the cloud flagging applied here compares to, for example, that from the CrIS CFPR product. Is it consistent?

We thank the referee for pointing out that the cloud-screening criterion was not sufficiently described. We have revised Section 2.2 to clarify the actual cloud-screening method used in this study. We do not apply a fixed cloud-fraction threshold such as cloud fraction < 0.4. Instead, following the brightness-temperature-difference method of Wells et al. (2020), we compare the observed brightness temperature near 900 cm^{-1} with the ERA5 surface skin temperature and use an ERA5-water-vapor-dependent threshold to identify cloud-contaminated pixels. For each observation, the observed brightness temperature near 900 cm^{-1} is compared with the nearest hourly matched ERA5 surface skin temperature. The cloud-screening threshold is adjusted according to the ERA5 water vapor column and no fixed cloud-fraction threshold is used. Observations for which the 900 cm^{-1} brightness temperature is lower than this water-vapor-dependent threshold relative to the surface skin temperature are flagged as cloud-contaminated and excluded. The same procedure is applied to FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS to maintain internal consistency across the constellation.

L158-L159: What is the approximate height above the surface of the lowest atmospheric layer here? This is an important clarification for understanding how rigid of a thermal contrast filter 3 kelvin is. For example, if the first layer is at a height of 500m above the surface, then 3K is a relatively loose filter, if its 50m then its a strict filter.

We thank the referee for pointing out that the definition of the lowest atmospheric layer is not sufficiently clear. We have revised the text to clarify that the TC is calculated as the temperature difference between the surface skin temperature and the temperature at the center of the lowest retrieval layer. The lowest retrieval layer is bounded by the local surface pressure and the first pressure level above the surface. Its center height varies with surface pressure, but is typically within the lowest several hundred meters above the surface. We have added this clarification to Section 3.1.

L173-L174: "In particular, large-scale wildfires prevalent in summer North America..." -> "... during the summer in North America..."

Revised as suggested.

L175-L177: I think another good citation to include here would be Lutsch et al. (2019; https://doi.org/10.1029/2019JD030419) as they discuss the long-range transport of NH_3 in wildfire plumes.

Revised as suggested.

L196-L197: its unclear what is meant here by "valid data defined as having more than 10 effective observation points to ensure statistical robustness". Do the authors mean a valid overpass must have more than 10 effective observations? Or the monthly averages themselves must consist of 10 points or more?

We thank the referee for pointing out this ambiguity. We have revised the sentence to clarify that the threshold refers to the number of quality-controlled retrievals used to calculate each monthly mean. Specifically, for each hotspot region and each satellite overpass time, a monthly mean is calculated only when more than 10 valid retrievals are available within the 2.5° × 2.5° box during that month. Monthly means based on 10 or fewer valid retrievals are excluded from the analysis.

L197-L198: I think it would be useful to (in a single sentence) explain why higher TC = better retrieval accuracy/better retrievals. It is also unclear whether the TC criterion is based on the absolute value of TC (thus allowing more strongly negative TC's), or if only positive TC values > 3K are allowed.

We thank the referee for pointing out this ambiguity. We have revised the text to clarify that the TC criterion is based on the absolute value of thermal contrast, rather than only positive TC. And we add description in Section 3.2 to explain why higher TC = better retrieval accuracy/better retrievals. “When TC is close to zero, the NH_3 spectral signal becomes weak and the retrieval sensitivity decreases, whereas a larger absolute TC generally enhances the spectral contrast and improves the retrieval sensitivity.”

L203: What version of the IASI retrievals was used here? It is never mentioned nor in the data availability section. I assume ANNI V4.0, but it should be noted in the text.

The IASI NH_3 retrievals we use here is ANNI V4.0 product. We add the IASI NH_3 product version in the main text and include the corresponding data source in the Data availability section.

L205: "a pattern that consists with Clarisse et al.’s (2021) findings from GIIRS observations and aligns with broader research on NH3 dynamics" -> "a pattern that is consistent with the findings of Clarisse et al. (2021) from GIIRS..."

Revised as suggested.

Figure 4: A more differentiated line color could be chosen for the NH3 and BLH lines to make it more colorblind friendly.

Revised as suggested.

Figure 5 and Figure 6: The aspect ratio of the individual subpanels looks stretched in the vertical.

Revised as suggested.

Figure 7: Overlay outlines of the country borders and the US state and Canadian province borders to make it easier to interpret the geographical area depicted in the figures.

Revised as suggested.

General comment on Section 3.3 and Figure 8: The co-location criterion is only described in the Figure 8 caption, but not in the main text. Please add this to the text. I am also curious how this co-location criterion was selected initially? The choice is not motivated or explained.

We thank the referee for this helpful suggestion. We have added the collocation criteria to the main text of Section 3.3 and explained this selection. Specifically, a polar-orbiting retrieval and a FY-4B/GIIRS retrieval are considered collocated when their longitude and latitude differences are both less than 0.5° and their observation time difference is less than 0.5 h. These thresholds are chosen as a compromise between maintaining a sufficient number of matched samples and reducing spatial and temporal representativeness differences.

L286-287: specify in the text the fitting method used for the linear regressions. Was it ordinary least squares or something more robust like reduced major axis?

We thank the referee for this helpful comment. We have clarified in the revised manuscript that the linear regression in Fig. 8 is performed using orthogonal distance regression (ODR), rather than ordinary least-squares regression. ODR is selected because both the polar-orbiting satellite retrievals and the FY-4B/GIIRS retrievals contain retrieval uncertainties, whereas ordinary least-squares regression assumes uncertainty only in the dependent variable. We have added this information to Section 3.3 and revised the Figure 8 caption accordingly.

Figure 8: It would be good to include information on the total number of observations, the fit intercepts, and also possibly the mean biases for each case somewhere on the figure subpanels or in the text.

Revised as suggested.

L315: Not 100% true that it follows a pure inverse relationship. For example, Boynard et al. 2014 (https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1002/2013GL058333) show that as thermal contrast goes largely negative (e.g., during winter), you regain sensitivity since your absorption features become emission features. So it is not fully correct to claim it is a pure inverse relationship with TC. Even here in Fig. 9 it is not always strictly the case.

We thank the referee for this important clarification. We agree that retrieval uncertainty should not be described as following a pure inverse relationship with TC. We have revised Section 3.4.

“Column retrieval uncertainty generally decreases under conditions with larger absolute TC and higher surface AVK, but this relationship is not purely inverse and should not be interpreted as strictly monotonic. In particular, strongly negative TC can also enhance thermal infrared sensitivity because absorption-like spectral features may become emission-like features, as shown for lower-tropospheric pollution retrievals by Boynard et al. (2014). Therefore, TC should be interpreted together with AVK and retrieval uncertainty, rather than as a standalone monotonic sensitivity indicator. This behavior is also visible in Fig. 9, where retrieval uncertainty does not always vary inversely with TC at all regions, months, and overpass times.”

Figure 9: in my opinion, more information is needed either on the panels or in the text on the total number of observations used to generate the diurnal plot here. Are there comparable number of observations used for January versus June? Or is it quite different?

We thank the referee for this helpful suggestion. We agree that the number of observations used to generate the monthly diurnal averages should be reported, especially because cloud screening and sensitivity-related quality filters can affect data availability differently across months, regions, and local overpass times. We have added Table A2, which lists the number of quality-controlled retrievals used to calculate each monthly mean in Fig. 9 for each region, month, and overpass time. We also revise the text in Section 3.4 and the caption of Fig. 9 to explicitly refer to this table. Monthly means are calculated only when more than 10 quality-controlled retrievals are available.

The Table A2 shows that the number of valid retrievals is not always comparable across months and overpass times. Some daytime bins contain more than 1000 valid retrievals, whereas some nighttime or evening bins contain far fewer samples.

L334: What is meant by "performance across different observational periods" here? The comparison with GIIRS was not separated into different periods/months seasons, so it is difficult to judge whether the performance is consistent across the sensors in the different periods. We can only really say that they compare across the full year.

We thank the referee for pointing out this ambiguity. By “different observational periods,” we intended to refer to the different matched local overpass periods of the three polar-orbiting sensors, namely CrIS at 01:30/13:30, FY-3E/HIRAS-II at 05:30/17:30, and FY-3F/HIRAS-II at 10:00/22:00, rather than to different seasons or months. We have revised the conclusions to make this explicit.

L336-L337: "anthropogenic activities (e.g., agricultural practices) govern diurnal and seasonal cycles, while regional variations reflect differences in emission intensity and atmospheric chemistry." but anthropogenic activities and differences in emission intensity are quite directly linked here for NH3 in most cases no?

We thank the referee for pointing out this redundancy. We agree that anthropogenic agricultural activities and emission intensity are closely linked for NH_3. We have revised the sentence to avoid treating them as independent drivers. "Key findings indicate that temperature-dependent volatilization, boundary-layer dynamics, and agricultural activity patterns jointly regulate the diurnal and seasonal variability of NH_3. Differences among regions are further modulated by source strength, fertilizer and livestock management practices, meteorological conditions, and atmospheric chemical processing."

Data availability: It seems the CrIS NH3 retrieval dataset is missing, it should also be included here same as the FY-3E and FY-3F datasets.

The link to the CrIS NH_3 retrieval dataset: https://disk.pku.edu.cn/link/AA444B51D57822489AAA3CAAF2CD3C5C1F

The CrIS retrievals are generated using the same retrieval framework as the FY-3E/HIRAS-II and FY-3F/HIRAS-II retrievals and are used in this study to complete the three-orbit LEO constellation analysis instead of using FY-3D at similar overpass time. However, no routine future updates of this CrIS dataset are currently planned.

We thank the reviewers for their constructive comments and suggestions to improve the quality and clarity of our manuscript. We have made major and careful modifications to the original manuscript according to all the comments and suggestions from the reviewers.

Item-by-item responses to the specific comments are provided below, in which the reviews' comments are in blue, our responses in **black**, and modifications of the original manuscript are indicated by red in the revised manuscript.

Reviewer #2:

This manuscript presents a potentially valuable study on global and diurnal tropospheric NH3 monitoring using a constellation of polar-orbiting hyperspectral infrared sounders: FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS. The idea of combining complementary local overpass times to achieve quasi-geostationary-like global sampling is interesting and important, especially for NH3, whose short lifetime and strong diurnal variability make sparse temporal sampling a major limitation.

The manuscript shows promising results, including global seasonal maps, regional diurnal cycles over major source regions, and comparison with FY-4B/GIIRS. But I still have some suggestions for the authors to consider.

We thank the referee for the positive evaluation and constructive comments. We have revised the manuscript substantially to improve the descriptions of the methodology and strengthen the robustness of the results. The main revisions include: (1) expanding the retrieval-method section to better document the profile-scaling approach, the selection of the a priori profile, the state vector, and the cross-sensor treatment; (2) adding a priori profile sensitivity analysis; (3) quantifying the impact of quality filters on the number of observations for further analysis; (4) expanding the FY-4B/GIIRS cross-comparison to representative months/seasons with additional statistics, including sample size, intercept, mean bias, and ODR fitting information; and (5) adding comparisons with ground-based FTIR measurements.

Major Comments:

1) The manuscript states that the retrieval is based on the FY-LeoAIR optimal estimation framework and that a profile scaling approach is used, but the actual NH3, retrieval setup is not sufficiently documented in this paper. Here are some of the contents that are suggested to include into the manuscript: the exact state vector elements retrieved jointly with NH3, whether surface temperature and emissivity are simultaneously fitted or fixed, whether the three sensors are spectrally harmonized before inversion, any sensor-specific bias correction or radiometric adjustment. Since this paper’s novelty rests on a multi-sensor integrated retrieval, the retrieval description must be more self-contained. Referring readers to previous papers is not enough.

We thank the referee for pointing out that the NH_3 retrieval setup should be documented more explicitly. We agree that a self-contained description is necessary because the novelty of this work relies on applying a harmonized retrieval framework to FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS.

We have revised Section 2.2 to describe the FY-LeoAIR NH_3 retrieval configuration in detail. The revised text now specifies the state-vector elements retrieved jointly with NH_3, including one multiplicative scaling factor for the NH_3 a priori profile, one multiplicative scaling factor for the H_2O a priori profile, scaling factors for weakly interfering trace gases, surface skin temperature, atmospheric temperature scaling, and four surface-emissivity adjustment coefficients based on first- to fourth-order Legendre polynomial terms. We also summarize the parameters in the retrieval state vector in the supplementary material.

We also clarify that surface skin temperature is initialized from ERA5 and retrieved simultaneously with the gas-scaling parameters. The forward-model emissivity prior is taken from the UOW-M global infrared land surface emissivity database, following Zeng et al. (2023). The a priori emissivity spectrum is adjusted within the retrieval window using four Legendre-polynomial coefficients.

Regarding the multi-sensor treatment, we now clarify that the same 955–975 cm^{-1} retrieval window is used for FY-3E/HIRAS-II, FY-3F/HIRAS-II, and CrIS. HIRAS-II and CrIS have the same nominal unapodized spectral resolution of 0.625 cm^{-1}, allowing the same retrieval window to be used consistently across the three sensors. Sensor-specific spectral noise estimates are used to construct the diagonal measurement-error covariance matrix. No empirical sensor-specific radiometric bias correction or additional spectral harmonization is applied. The same forward model, retrieval window, auxiliary datasets, and quality-control procedure are used for all three sensors to maintain internal consistency.

We further clarify that the fixed NH_3 a priori profile follows Zeng et al. (2023), based on GEOS-CF NH_3 simulations over representative polluted land regions in East Asia and South Asia. The main modification relative to Zeng et al. (2023) is that NH_3 is retrieved as a single multiplicative profile-scaling factor rather than as independent layer-by-layer NH_3 state-vector elements. This approach is adopted because individual thermal infrared spectra generally provide limited vertical information for NH_3 and because it improves computational efficiency for global multi-sensor processing.

2) The main consistency check is a comparison with FY-4B/GIIRS over the Indo-Gangetic Plain and North China Plain. This is useful, but I think it is not sufficient to support the broader conclusion that the constellation can robustly monitor global and diurnal NH3 variability.

We thank the referee for this important comment. We agree that the original consistency evaluation based only on the annual FY-4B/GIIRS comparison over the Indo-Gangetic Plain and North China Plain is not sufficient to support a broad statement about global retrieval robustness. In the revised manuscript, we have strengthened the evaluation in two ways.

First, we add seasonal comparisons between FY-3E/HIRAS-II, FY-3F/HIRAS-II, CrIS, and FY-4B/GIIRS, shown in Fig. A2. The results show that the seasonal slopes mostly fall within the range of 0.7–1.1, supporting broad consistency between the polar-orbiting retrievals and FY-4B/GIIRS.

Second, we add an independent comparison with ground-based FTIR NH_3 column measurements at Hefei, China, shown in Fig. A3. This provides an additional evaluation against ground-based measurements. We acknowledge that this comparison is still subject to representativeness differences and sensitivity differences, but it strengthens the evidence that the constellation retrievals capture the main temporal and spatial variability of NH_3.

We have also revised Section 3.3 and the conclusions to avoid overstatement. The revised manuscript now describes these analyses as consistency evaluations and states that they support the capability of the constellation to capture major NH_3 spatial and temporal patterns.

3) Another major concern is that retrieval sensitivity is explicitly controlled by thermal contrast (TC), and the authors apply thresholds such as TC > 3 K globally and TC > 5 K in hotspot analyses. Since TC itself has a diurnal cycle, the filtering preferentially retains daytime observations and may distort the apparent amplitude and phase of the NH3 diurnal cycle.

We thank the referee for raising this important concern. We agree that TC-based filtering can affect the local-time sampling of the retrievals because TC itself has a strong diurnal cycle. We have revised the manuscript to clarify that the filter is applied to the absolute TC, rather than to positive TC only. The absolute TC > 3 K threshold is used in Section 3.1 for the global spatial-distribution examples, where the aim is to retain sufficient sampling density while excluding retrievals with very weak thermal contrast. For the hotspot diurnal analysis in Section 3.2, we use a stricter threshold of absolute TC > 5 K together with surface AVK > 0.3 to reduce the influence of local-time-dependent retrieval sensitivity on the inferred diurnal variations.

To directly evaluate whether the absolute TC and surface AVK filtering distorts the retrieved diurnal cycles, we add a sensitivity analysis using three filtering configurations: surface AVK > 0.1 and absolute TC > 3 K, surface AVK > 0.3 and absolute TC > 5 K, and surface AVK > 0.4 and absolute TC > 7 K. We compare the resulting spatial distributions over the North China Plain and Indo-Gangetic Plain, the regional diurnal cycles, and the total number of retained retrievals for each satellite. These results are provided in the supplementary material Figs. S7-S10.

The sensitivity test shows that the number of retained retrievals decreases as the filters become stricter, but the spatial distributions remain broadly similar and the main diurnal patterns are not obviously distorted by the choice of absolute TC and surface AVK threshold. We now state that the exact column magnitude and sampling density remain filter-dependent, but the principal diurnal features discussed in the manuscript are not primarily an artifact of the TC-based filtering.

Minor Comments:

1) Section 3 is titled "Results and Discussions," while Section 4 is said to be "the discussions" in the introduction, but there is no standalone Section 4 in the provided text.

We thank the referee for pointing out this inconsistency. We have revised the section description in the Introduction to match the actual manuscript structure. Section 3 now presents and discusses the results, and the conclusions are given in Section 4.

2) The title and text alternate between "tropospheric ammonia," "NH3 column," and "total column." Since the retrieval is limited to 11 layers from the surface to 200 hPa, please define precisely what column is being reported and use consistent terminology.

We thank the referee for pointing out the need for a precise definition of the reported column quantity. We have revised the manuscript to clarify that the retrieval state vector adjusts NH_3 in 11 layers from the surface to 200 hPa. The reported NH_3 total column is then calculated as the sum of the retrieved surface-to-200 hPa column and the fixed a priori contribution above 200 hPa. Since the NH_3 abundance above 200 hPa is generally small and is not independently constrained by the spectra, the spatial and temporal variability discussed in this study is primarily determined by the retrieved surface-to-200 hPa component. We now define this explicitly in Section 2.2 and use "NH_3 column" consistently throughout the manuscript to refer to this reported total column.

3) A 2.5°×2.5° box is fairly large, especially in regions with heterogeneous sources. Please explain why this spatial extent was selected and how sensitive results are to box size.

We thank the referee for this comment. We have added a discussion explaining that the 2.5° × 2.5° box is chosen to balance spatial representativeness and sample size. Because the analysis is separated by month and overpass time and uses strict quality filters, smaller boxes can result in too few valid retrievals, especially for nighttime observations. We also acknowledge that the selected box size may smooth sub-regional source heterogeneity, and we identify box-size sensitivity as a limitation to be examined in future work.

4）The manuscript states that the emissivity database is “Monthly Global 0.05° V003” and then says it provides data “at a 0.5° spatial resolution.” This should be checked.

We thank the referee for noting this inconsistency. We have corrected the description of the emissivity product. The Combined ASTER and MODIS Emissivity over Land Database Monthly Global V003 product has a spatial resolution of 0.05°.

5）Section 3.3 title says “Comparisons with geostationary NH3 observations,” but the text uses “validate.” I suggest replacing “validate” with “evaluate consistency” or “cross-compare.”

We agree with the referee that FY-4B/GIIRS provides an independent satellite retrieval product rather than an absolute validation reference. We have therefore replaced “validate” with “evaluate consistency” or “cross-compare” throughout Section 3.3 and revised the section title to “Cross-comparison with geostationary FY-4B/GIIRS and ground-based FTIR NH_3 observations”

6）The collocation uses <0.5° in latitude/longitude and <0.5 h in time. Given the gradients and diurnal variability of NH3, please justify these thresholds and discuss representativeness mismatch.

We thank the referee for this suggestion. We have added the collocation criteria to the main text and explained the rationale for the selected thresholds. A polar-orbiting retrieval and a FY-4B/GIIRS retrieval are considered collocated when their longitude and latitude differences were both less than 0.5° and their observation time difference is less than 0.5 h. These criteria are selected as a compromise between retaining enough matched samples and reducing representativeness differences.

7) The manuscript uses both “PBL” and “BLH.” Please make it consistent.

We thank the referee for pointing this out. We have standardized the terminology throughout the manuscript. The term “boundary layer height (BLH)” is defined at its first occurrence, and “BLH” is used consistently thereafter.