Reply to the comments of RC1

RC1: General comment

The manuscript, "Spatiotemporal optimization of NO_x and VOC emissions using a hybrid inversion framework and its implication for ozone sensitivity diagnosis", presents a methodology for optimizing emission inventory data for regional air quality model simulations in South Korea. This is achieved using hybrid data assimilation approach, which first adjusts the emissions of NO_x and VOC using TROPOMI observations with the finite difference mass balance, and then refines the emission optimization with 4D-Var data assimilation of ground-based air quality monitoring stations (NO_2 and O_3) and Pandora spectrometers (HCHO). The conducted study covers a two-week period in May 2022, for which the emission adjustment as well as the simulation analysis improvement are evaluated. Additionally, spatially resolved ozone sensitivity regimes are derived for South Korea, showing a shift towards larger regions becoming NO_x sensitive with the newly derived emissions. Finally, the change in O_3 response to NO_x and VOC emissions is evaluated over time, showing the different precursors' individual behavior and their daytime dependences.

Overall, the manuscript's scientific topic is very important for the air quality community. Since uncertainties in emission inventories remain the primary source of uncertainty in model predictions, the proposed approach offers a sophisticated methods to address the existing challenges in accurately representing ozone concentrations and its primary precursors, NO_x and VOC. Furthermore, the presented hybrid inverse modelling method provides an advanced top-down approach that jointly optimizes NO_x and biogenic and anthropogenic VOC emission inventories. The current aim of reducing NO_x emissions in East Asia results in a shift in the ozone sensitivity regime, as demonstrated here. In terms of guidance for O_3 policy management, the manuscript proposes a time and ozone regime dependent analysis of adjoint sensitivities for O_3 responses to the precursor emissions. With these comprehensive methods and analysis, the manuscript complies well with the scope of ACP.

Although, I consider the methods and analysis presented in the manuscript to be of high scientific significance and rate the presentation quality as high, I have major concerns about parts of the applied methods, which I will discuss in more detail in the following section. In conclusion, I recommend accepting the manuscript for publication in ACP once my major and minor concerns have been addressed.

Response: We thank the reviewer for the constructive and helpful comments on our manuscript. We have revised the manuscript accordingly. Below, we provide point-by-point responses to each comment. Reviewers' comments are shown in plain text, our responses are shown in bold, and revised manuscript text is shown in *italics*. Line numbers refer to the revised manuscript.

Major comment 1: The methodology described in section 2.5 and illustrated in Figure 4 is not clear. How exactly is the adjoint model driven by the posterior emissions? The relationship between the "local hour" and the two-week average remains unclear. Is this done for every hour individually within the two-week analysis period? What does "d" represent in equation (9)? What is "D" as a set of analysis days? Is this an average over the two weeks? Is "g" really the grid, or does it refer to individual grid cells? Could you please explain what is meant by "receptor hour" and how the cost function Jr is treated (e.g. minimizing procedure)? Equations (10) to (12) should be explained in words, with the different variables introduced precisely to give the reader a better understanding and ensure reproducibility. In addition, please add information to the text on how the 24-hour simulations are selected, and how t1 and t2 are relating to these 24-hour windows. There is a hint in the caption of Fig. 4, however, this information should also be included in the main text. Finally, please add a paragraph describing Fig. 4 to the section. The current presentation is insufficient for understanding the derivation of ΔO_3 responses.

Response: **We thank the reviewer for their detailed and constructive comments. We agree that Section 2.5 required clearer descriptions, and we have revised the section and Figure 4 to improve clarity and reproducibility.**

To address the reviewer's specific questions, we made the following key revisions:

(1) Adjoint model setup: We added an explicit description at the beginning of Section 2.5, clarifying that a forward CMAQ simulation is first run with the optimized posterior emissions to generate the concentration fields, which then drive the adjoint model.

(2) Notation and variables: We replaced the previous notation (t_1/t_2) with emission hour (h_e) and receptor hour (h_r) to explicitly distinguish them. We also defined all variables in Eq. (9), including d (individual analysis day), D (complete set of 14 analysis days), and g (individual grid cell), and added step-by-step explanations for Eqs. (10)–(12).

(3) Cost function clarification: We clarified that $\mathcal{J}_{Reg}(h_r)$ serves purely as a diagnostic receptor function to evaluate O_3 sensitivity, rather than an optimization cost function like $J(\alpha)$ in Eq. (8), and is therefore not subject to iterative minimization.

(4) Simulation design: We clarified the relationship between the local hour and the two-week average. Because $\mathcal{J}_{Reg}(h_r)$ is defined separately for each receptor hour

h_r, 24 independent adjoint simulations were performed per regime (totaling 48 runs), each spanning the entire two-week analysis period.

(5) Figure 4 Updates: The notation in Figure 4 has been updated, and a descriptive paragraph has been added at the end of Section 2.5 to guide the reader through the schematic.

<Fig. 4 in the revised manuscript>

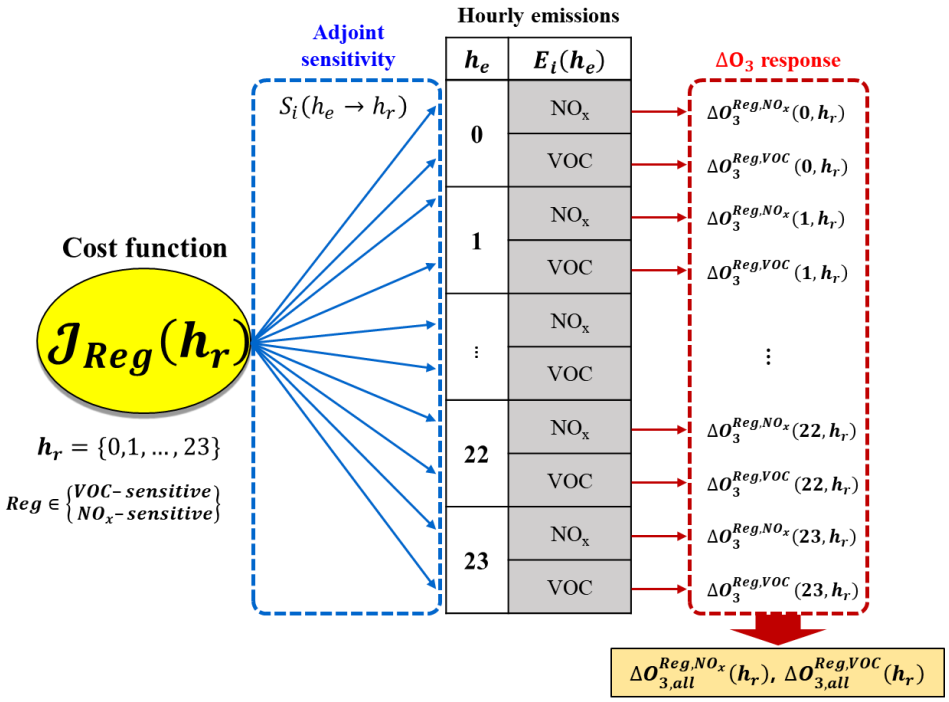

Figure 1: Schematic of the adjoint-based O_3 response calculation. A single adjoint run provides, for a receptor hour h_r, the sensitivities of the regime-mean surface O_3 cost function $\mathcal{J}_{Reg}(h_r)$ to precursor emissions at each emission hour t_1. Multiplying these sensitivities by the corresponding hourly emissions $E_i(h_e)$ ($i \in \{NO_x, VOC\}$) yields the emission-time–specific O_3 response $\Delta O_3^{Reg,i}(h_e, h_r)$. Summing over all emission hours $h_e = 0\text{–}23$ gives the emission-time–integrated response $\Delta O_{3,all}^{Reg,i}(h_r)$. Responses are evaluated over grids classified as regime Reg at hour h_r, enabling regime-by-regime comparison of NO_x and VOC influences.

Major comment 2: Regarding the evaluation presented in section 3.2, I have some concerns about whether this is a fair and robust method of evaluating the benefits of the emission corrections. In general, comparing concentrations of the prior and posterior simulations with observed concentrations is a valid approach. However, as I understand it, the NO_2 and O_3 observations used for the evaluation are the same as those used for the optimization experiments. Therefore, the evaluation data forms part of the model, and it is obvious that the posterior results in better performance and statistics. Consequently, I consider the conclusion drawn in lines 347-348, "indicating enhanced model performance in reproducing observed O_3 concentrations", to be misleading. I would like to suggest using independent observations for validation. This could be achieved by withholding some stations from the network for validation purposes, or by finding additional data that the model has not seen.

Response: **We thank the reviewer for this important comment. We agree that evaluating posterior simulations only against the same AQMS observations used in the inversion does not provide a fully independent assessment of the posterior emissions. To address this concern, we conducted an additional validation experiment using a data-splitting approach, as described in Section 2.2.1 of the revised manuscript. Among the 619 AQMS sites, 496 sites (80%) were randomly selected for the inversion, while the remaining 123 sites (20%) were withheld from the inversion and used exclusively for independent validation. The spatial distribution of the inversion and validation sites has been added as Fig. S2.**

In the revised manuscript, we clarified that the comparison using all AQMS sites in the baseline inversion primarily evaluates the consistency of the posterior simulations with the assimilated observations. We then added the independent validation results using the withheld AQMS sites to assess the robustness of the posterior emissions. The independent validation experiment showed that the Hybrid_NOx+VOC experiment improved the model performance at withheld sites, indicating that the posterior corrections were not solely a result of fitting the assimilated observations. These results are summarized in Table S6 and discussed in Section 3.2.

<Section 2.2.1 in the revised manuscript>

To assess the robustness of the posterior emissions using independent observations, an additional validation experiment was conducted using a data-splitting approach, in which 496

sites (80%) were randomly selected for the inversion and the remaining 123 sites (20%) were reserved for independent validation. The spatial distribution of the selected sites is shown in Fig. S2.

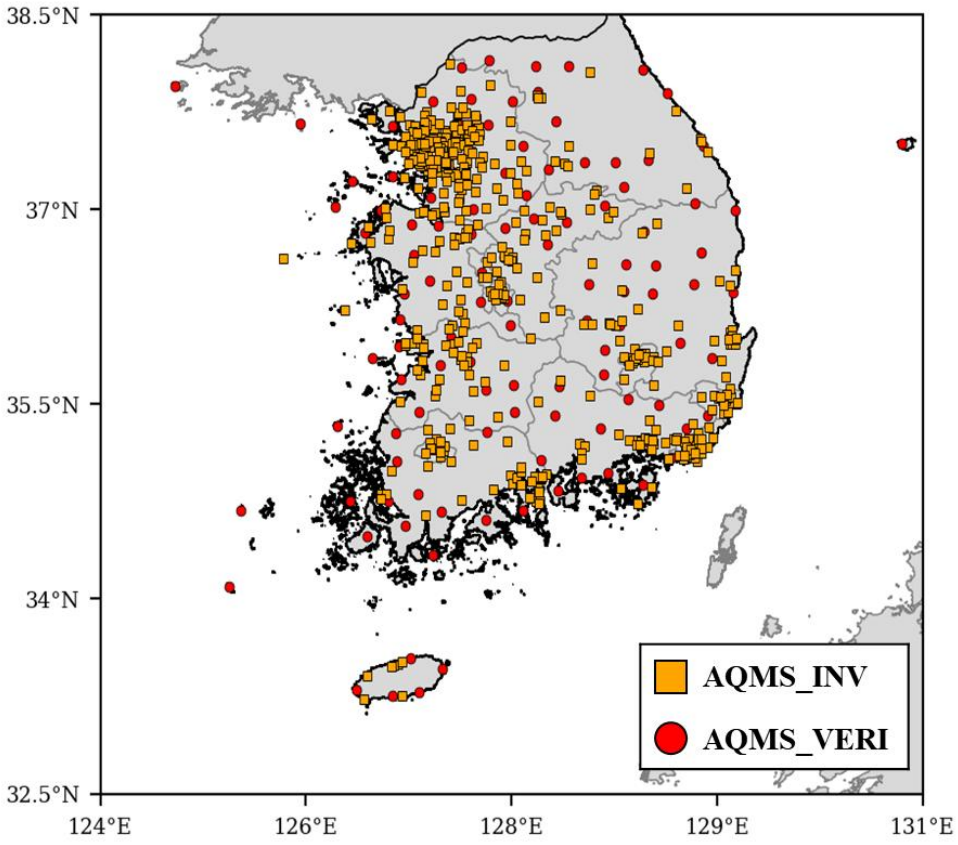

Figure S2: Spatial distribution of AQMS sites used in this study. A total of 619 sites were considered, among which 496 sites (80%) were randomly selected for the inversion (AQMS_INV; orange squares) and the remaining 123 sites (20%) were reserved for independent validation (AQMS_VERI; red circles).

<Section 3.2 in the revised manuscript>

The IOA improved from 0.71 (Prior) to 0.80 (Hybrid_NOx) and 0.83 (Hybrid_NOx+VOC), indicating improved consistency with the AQMS observations used in the baseline inversion. […] As described in Sect. 2.2.1, the independent validation experiment produced similar improvements at the withheld AQMS sites, indicating that the posterior emission corrections were not solely a result of fitting the assimilated observations. Compared with the Prior experiment, the Hybrid_NOx+VOC experiment reduced the NO_2 and O_3 biases and improved the temporal agreement with independent AQMS observations (Table S6). These results support the robustness of the posterior emissions and indicate that the hybrid inversion framework improves O_3 simulations beyond the stations directly used in the inversion.

Table S6. Statistical evaluation of NO_2 and O_3 concentrations for the Prior and Hybrid_NOx+VOC experiments at the independent validation sites. The statistics were calculated over the 123 withheld AQMS_VERI sites shown in Fig. S2 and averaged over the entire study period. Observations were obtained from the AQMS network.

Species	Experiment	Obs.	CMAQ	MBE	RMSE	IOA	r
NO_2 [ppb]	Prior	8.69	9.53	0.84	11.50	0.58	0.44
	Hybrid_NOx+VOC		6.58	-2.11	7.33	0.67	0.48
O_3 [ppb]	Prior	44.48	43.54	-3.02	16.60	0.71	0.53
	Hybrid_NOx+VOC		47.42	0.86	13.65	0.77	0.64

Major comment 3: For the hybrid optimization approaches, it remains unclear how the daily derived emission corrections are taken into account in the analysis. Are the corrected emission profiles from the previous day used as prior for the following day? Or is the same prior temporal emission profile always used? How are the prior emission profiles implemented? What does the prior in Figure 6 base itself on?

Response: **We thank the reviewer for this comment. We agree that the manuscript lacked a clear description of how the prior emission fields are implemented and how emission corrections are carried over across the analysis period, and have revised Section 2.3.3 and the caption of Figure 6 accordingly.**

The prior emission fields used throughout this study are derived from the EDGAR HTAPv3 inventory, temporally disaggregated to hourly resolution using sector-specific temporal allocation profiles (Crippa et al., 2020), and this forms the basis of the "Prior" curve in Figure 6. Within the hybrid inversion, the FDMB step corrects only the spatial distribution of emissions without modifying their temporal allocation, and the spatially corrected FDMB emissions are used as the spatial prior for each 24-hour assimilation window in the 4D-Var step. The optimized emissions from a given day do not serve as the prior for the following day; instead, the optimized concentration fields are carried over as initial conditions for the subsequent day's forward integration, allowing the atmospheric state to evolve continuously across the two-week period.

<Section 2.3.3 in the revised manuscript>

The prior emissions used in this study are derived from the EDGAR HTAPv3 inventory, temporally disaggregated to hourly resolution using sector-specific temporal allocation profiles (Crippa et al., 2020). […] As the FDMB inversion relies on two-week-averaged satellite column observations, it corrects only the spatial distribution of emissions without modifying their temporal allocation. […] The 4D-Var inversion was applied sequentially using a 24-hour assimilation window over the two-week analysis period, with the FDMB-corrected emissions serving as the spatial prior for each window. The optimized concentration fields from each window were carried over as initial conditions for the subsequent day's forward integration, allowing the atmospheric state to evolve continuously across the analysis period.

< The caption of Fig. 6 in the revised manuscript>

Diurnal variations of NO_x (top), AVOC (middle), and BVOC (bottom) emissions for each experiment (Prior: gray, Hybrid_NOx: blue, Hybrid_NOx+VOC: red), averaged over South Korea during the study period. The Prior represents the hourly emission profile derived from the EDGAR HTAPv3.

Major comment 4: I find the hybrid modelling approach very interesting. However, could you please provide insights into the individual contribution of the FDMB and 4D-var data assimilation to the emission correction? Which optimization approach drives the corrections? Is the hybrid approach the most effective method, or would using one method provide sufficient results at lower computational cost?

Response: **We thank the reviewer for this insightful question regarding the individual contributions of the FDMB and 4D-Var methods. This specific aspect was thoroughly investigated in our previous work (Moon et al., 2024).**

In brief, the FDMB inversion utilizes time-averaged satellite column observations, making it highly effective at correcting the mean spatial distribution of emissions. However, it is fundamentally limited in its ability to correct temporal variations. Conversely, while the 4D-Var approach excels at optimizing temporal (hourly) variations, it is less efficient at correcting large-scale spatial biases.

Therefore, relying on a single method may be insufficient to fully resolve the coupled spatiotemporal discrepancies simultaneously under our modeling configuration. The hybrid approach synergistically combines the strengths of both: the FDMB method establishes an accurate spatial baseline, which the 4D-Var method subsequently refines in the temporal domain. We have explicitly added this justification to the revised manuscript to clarify the necessity and efficiency of the hybrid framework.

<Section 2.3.3 in the revised manuscript>

A model-based twin experiment by Moon et al. (2024) demonstrated that a standalone 4D-Var approach structurally struggles to effectively correct emissions in grid cells with exceptionally large prior spatial errors. By sequentially combining these methods, the FDMB step establishes a robust and accurate spatial baseline, which enables the subsequent 4D-Var step to focus more effectively on optimizing diurnal temporal variations and multi-species chemical feedbacks, while reducing the influence of spatial distribution errors in the prior emissions.

Major comment 5: Regarding the discussion about O_3 policy management in section 3.4, it would be interesting to assess the impact of AVOC emissions on the O_3 responses. As only AVOC emissions can be actively regulated while BVOC emissions are controlled by nature, it would be interesting to see to how efficient possible AVOC controls would actually be.

Response: We thank the reviewer for this constructive suggestion. We agree that distinguishing between AVOCs and BVOCs is crucial for evaluating the practical efficiency of O_3 policy management, as only AVOCs can be actively regulated.

To address this, we have further analyzed the VOC contributions and separated the O_3 responses into AVOC and BVOC components, which are now presented in a newly added supplementary figure (Fig. S7).

Our extended analysis reveals the following:

1) In the NO_x-sensitive regime, the contribution of AVOCs to O_3 formation is small.
2) In the VOC-sensitive regime, however, AVOCs account for approximately 14% to 21% of the total VOC-driven O_3 response, indicating a non-negligible potential for targeted AVOC controls in these specific areas.
3) Nevertheless, BVOCs still predominantly drive the overall O_3 responses across the domain. We attribute this strong BVOC dominance to the specific study period (May), which is characterized by highly active biogenic emissions.

Therefore, the influence of BVOCs observed in this study might be more pronounced than in other seasons, such as autumn or winter, when biogenic activities significantly decrease. Consequently, the potential effectiveness of AVOC controls can vary considerably by season, and our findings regarding the limited impact of AVOC reductions should be interpreted with the seasonal characteristics of the study period in mind.

We have added this discussion and Fig. S7 to Section 3.4 of the revised manuscript to provide a more comprehensive perspective on emission control strategies.

<Section 3.4 in the revised manuscript>

Furthermore, to assess the practical efficiency of VOC emission controls, it is crucial to distinguish between anthropogenic and biogenic sources, as only AVOCs can be actively regulated. As shown in the separated VOC contributions (Fig. S7), the contribution of AVOCs to O_3 formation is small in the NO_x-sensitive regime. In the VOC-sensitive regime, however, AVOCs account for approximately 14% to 21% of the total VOC-driven O_3 response, indicating a non-negligible potential for targeted AVOC controls in these specific areas. Nevertheless, BVOCs still predominantly drive the overall O_3 responses across the domain. We attribute this strong BVOC dominance to the specific study period (May), which is characterized by highly active biogenic emissions. Therefore, the influence of BVOCs observed in this study might be more pronounced than in other seasons, such as autumn or winter, when biogenic activities significantly decrease. Consequently, the potential effectiveness of AVOC controls can vary considerably by season, and our findings regarding the limited impact of AVOC reductions should be interpreted with the seasonal characteristics of the study period in mind.

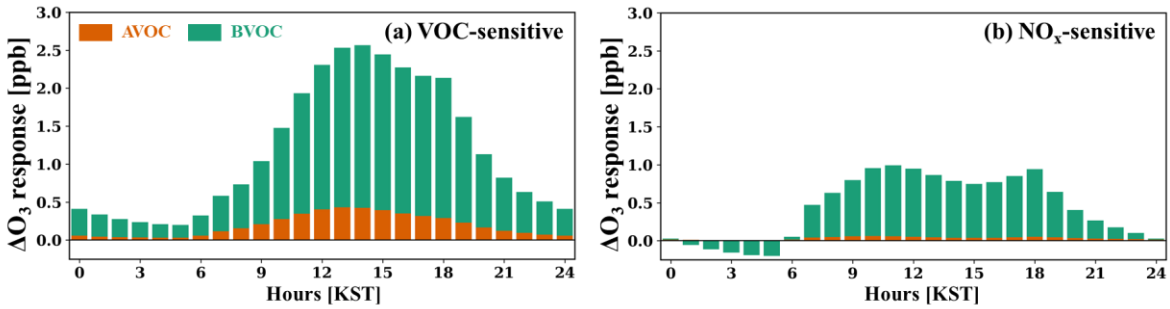

Figure S7: Two-week mean ΔO_3 response (ppb) to anthropogenic VOC (AVOC, orange) and biogenic VOC (BVOC, green) emissions, by local hour, for 1–14 May 2022. For each local hour, responses are summed over all grids classified in each regime at that hour; regimes are diagnosed with the HCHO-to-NO_2 ratios from the Hybrid_NOx+VOC experiment. Panels (a) and (b) show, for VOC-sensitive and NO_x-sensitive regimes respectively, the ΔO_3 response at each local hour to emissions released over all hours (i.e., emission-time–integrated response).

RC1: Specific comments and responses

Comment 1: The abstract misses information about the emission being optimized and that the study focuses on South Korea.

Response: **We have added the phrase "over South Korea" to the Abstract to clearly define the study domain and ensured that the optimized emissions are explicitly mentioned.**

<Abstract in the revised manuscript>

We develop a hybrid inverse modeling framework that couples the Finite Difference Mass Balance (FDMB) method with four-dimensional variational data assimilation (4D-Var) using the Community Multiscale Air Quality (CMAQ) model to jointly constrain spatiotemporal NO_x and VOC emissions over South Korea.

Comment 2: In general, the manuscript uses many abbreviations that are partly not very common. If introduced paragraphs ago, the reader gets distracted with going back and searching the abbreviation. Thus, I recommend reducing those, as much as possible. Examples are: in line 29, "IOA" is not even defined in the abstract; in line 82 "ΔO_3 responses" are used without introduction; in line 235, FNR is introduced – I suggest, just using HCHO-to-NO_2 ratio instead of FNR.

Response: **We agree that excessive or non-standard abbreviations can hinder the reader's comprehension. We have revised the manuscript accordingly to improve clarity and readability:**

(1) Reduction of Abbreviations: We have replaced the abbreviation "FNR" with the full term "HCHO-to-NO_2 ratio" throughout the manuscript, including all figure labels (Fig. 9 and Fig. S6) and the main text.

(2) Proper Definitions: We have thoroughly reviewed the manuscript and ensured that all abbreviations (e.g., IOA, KST) are clearly defined upon their first appearance in both the Abstract and the main text.

(3) Clarification of Terminology: For terms like "ΔO_3 responses," we have ensured that they are clearly introduced or contextually defined at their first usage to avoid ambiguity.

Comment 3: Line 23: insert “data” before assimilation.

Response: We have inserted the word "data" as suggested.

Comment 4: Line 26: In the abstract, you refer to “a period of climatologically high O_3” (line 26). This is unclear. How can a period of two weeks be climatological? Please improve the wording to better relate the O_3 burden of the selected episode to climatological O_3 values.

Response: We agree that the term "climatological" was misleading for a two-week episode. We intended to express that the selected two-week period falls within a month that recorded the highest mean O_3 concentration over the past decade. We have rephrased the sentence to clearly convey this context.

<Abstract in the revised manuscript>

The analysis covers 1–14 May 2022, during a month that exhibited the highest mean O_3 over the past decade.

Comment 5: Line 29: I recommend writing “Optimized emissions” instead of “Constrained emissions”.

Response: We have changed “Constrained emissions” to “Optimized emissions” in the revised manuscript.

Comment 6: Lines 31 and 32: The terms “VOC controls” and “NOx controls” are not clear at this stage of the manuscript. Please rephrase.

Response: **We agree that the shorthand terms "VOC controls" and "NOx controls" could be ambiguous in the abstract. We have rephrased the text to remove these terms, instead explicitly describing the distinct temporal characteristics of precursor influences (titration vs. photochemical production) and stating the implications for emission reductions directly.**

<Abstract in the revised manuscript>

Adjoint-based hourly ΔO_3 responses reveal distinct temporal characteristics: O_3 titration by NO_x is immediate, whereas photochemical O_3 production by VOCs requires a 1–2 hour reaction time. Furthermore, reactive biogenic VOCs act as a notable nighttime O_3 sink under NO_x-sensitive conditions. These findings motivate hour-specific, regime-specific controls rather than uniform daily reductions.

Comment 7: Line 49: Should the first word not be increase instead of decrease? Please double check.

Response: We have double-checked the context, and indeed, "increase" is the correct term. We have revised the sentence accordingly.

<Introduction in the revised manuscript>

In a VOC-sensitive regime, O_3 production increases when VOC emissions rise, but shows little change or may even increase when NO_x emissions are reduced.

Comment 8: Line 52: Please delete the word “paradoxically”.

Response: We have deleted the word “paradoxically” as suggested.

Comment 9: Line 63: Insert “data” before assimilation.

Response: We have inserted the word “data” before “assimilation” as requested.

Comment 10: Line 71: Please replace "combining the mass balance and the 4D-Var have been recently proposed" by "combining mass balance and 4D-Var has recently been proposed".

Response: We have corrected the phrasing and grammar as suggested: "combining mass balance and 4D-Var has recently been applied."

Comment 11: Line 84: "a unit perturbation" is unclear. Please explain.

Response: We have revised the sentence to provide a more detailed explanation of the term.

<Section 1 in the revised manuscript>

Here, ΔO_3 represents the change in O_3 concentration in response to changes in precursor emissions, as diagnosed by the adjoint model.

Comment 12: Line 89: I would like to suggest using the wording "spatiotemporal corrections" here instead of "spatiotemporal changes" in emissions and concentrations.

Response: **We have replaced "spatiotemporal changes" with "spatiotemporal corrections" throughout the manuscript, including the titles of Section 3.1 and 3.2, to more accurately reflect our inversion methodology.**

Comment 13: Line 98: Please provide the horizontal resolution of the modelling domains, at least for D2 in the text.

Response: We have added the horizontal resolution for both modeling domains (D1 and D2) in the text to provide full technical details of the simulation configuration.

<Section 2.1 in the revised manuscript>

The modeling domains consisted of two nested grids: a coarse-resolution outer domain (D1) covering East Asia at a horizontal resolution of 27 km and a finer-resolution inner domain (D2) focusing on South Korea at 9 km resolution (Fig. 1).

Comment 14: Figure 1: The blue triangle ASOS stations are hard to see on the map. Try to remove the province boundaries.

Response: We have updated Figure 1 by adjusting the transparency of the province boundaries to enhance the visibility of the ASOS station markers.

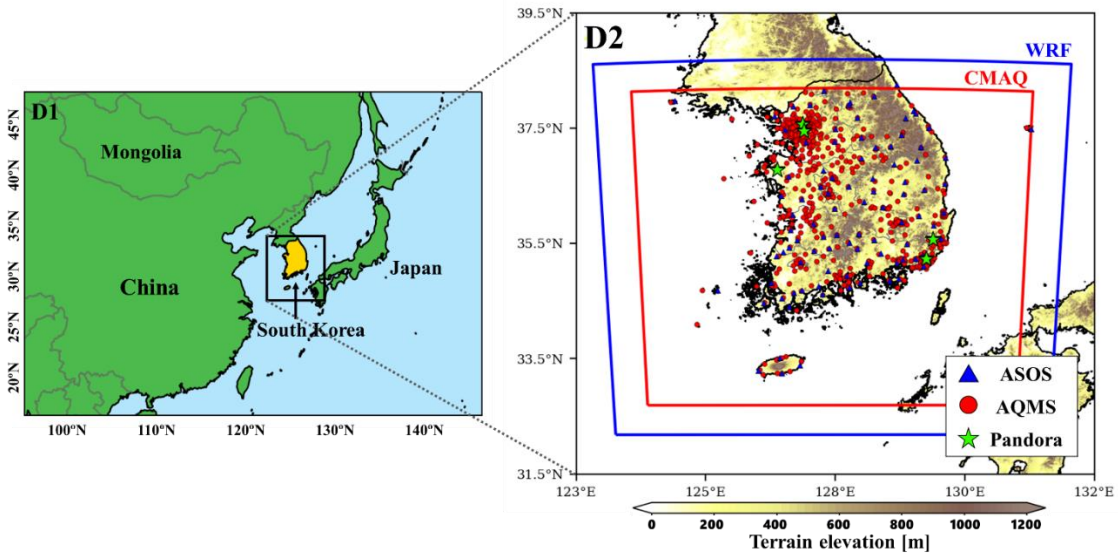

Figure 2: WRF/CMAQ modeling domains and spatial distribution of observational sites (ASOS: blue triangles; AQMS: red circles; Pandora: green stars).

Comment 15: Lines 112 and 113: Please introduce all chemical variables not just by their chemical abbreviation but also by name. Please consider replacing the "or" by an "and".

Response: We have revised the text to provide the full names for all chemical species and have replaced "or" with "and" to improve clarity and precision.

<Section 2.1 in the revised manuscript>

This inventory incorporates national emission estimates from South Korea's Clean Air Policy Support System (CAPSS) and provides monthly averaged emissions at a spatial resolution of 0.1° × 0.1° for nine key air pollutants: black carbon (BC), carbon monoxide (CO), nitrogen oxides (NO_x), sulfur dioxide (SO_2), ammonia (NH_3), organic carbon (OC), non-methane volatile organic compounds (NMVOC), particulate matter with an aerodynamic diameter ≤ 10 μm (PM_{10}), and particulate matter with an aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$) (Crippa et al., 2023).

Comment 16: Line 119: Please provide information about the initial (for D1 and D2) and boundary (for D1) conditions used for the CMAQ simulations. Do you start from a climatology or from some spin-up simulations?

Response: We have clarified the model initialization and boundary conditions in Section 2.1. To ensure numerical stability and minimize the influence of initial conditions, a 4-day spin-up period was included. The initial and boundary conditions were derived from standard CMAQ profiles for the outer domain, with one-way nesting utilized for the inner domain.

<Section 2.1 in the revised manuscript>

The CMAQ simulations were conducted from 27 April to 15 May 2022, including a 4-day spin-up period. The analysis focuses on 1–14 May 2022 (two weeks), selected for computational efficiency during the month that recorded the highest monthly average surface O3 concentrations over South Korea in the past decade (Fig. S1). Detailed model configurations for both WRF and CMAQ are provided in Tables S1 and S2. Initial and boundary conditions for the outer domain (D1) were derived from default CMAQ vertical profile data, while the inner domain (D2) was driven by D1 through one-way nesting.

Comment 17: Line 126: You refer to 619 AQMS sites. Are these divided into data used for the assimilation/inversion and data used for the validation of the posterior analysis? See also major comment (2).

Response: We have clarified how the AQMS observations were used in the revised manuscript. In the baseline inversion, all 619 AQMS sites were used for data assimilation and model evaluation. To address the issue of independent validation, we additionally conducted a data-splitting experiment, in which 496 sites (80 %) were randomly selected for the inversion and the remaining 123 sites (20 %) were withheld for validation. The spatial distribution of the inversion and validation sites was added as Fig. S2, and the corresponding validation statistics are provided in Table S6.

<Section 2.2.1 in the revised manuscript>

In the baseline inversion, all sites were used for both data assimilation and evaluation. To assess the robustness of the posterior emissions using independent observations, an additional validation experiment was conducted using a data-splitting approach, in which 496 sites (80%) were randomly selected for the inversion and the remaining 123 sites (20%) were reserved for independent validation. The spatial distribution of the selected sites is shown in Fig. S2.

Comment 18: Line 129: Please explain the "(0,10)" and "(1,11)". At the moment, these values are not clear.

Response: We have revised the text to clarify that these numbers refer to the Level 2 data quality flags of the Pandora retrievals. Specifically, Pandora NO_2 retrievals with high-quality flags 0 and 10 were used, while Pandora HCHO retrievals with high-quality flags 0 and 10 and medium-quality flags 1 and 11 were retained for the analysis.

<Section 2.2.1 in the revised manuscript>

To ensure data reliability, only Pandora NO_2 retrievals with Level 2 (L2) data quality flags classified as "high" quality (flags 0 and 10) were used. For HCHO, L2 retrievals classified as "high" quality (flags 0 and 10) or "medium" quality (flags 1 and 11) were used in this study (Bae et al., 2025; Fu et al., 2025).

Comment 19: Line 151: Please define AMFapriori in the text.

Response: We have added a definition of AMFapriori in the revised manuscript.

<Section 2.2.2 in the revised manuscript>

Here, AMFapriori denotes the a priori air mass factor provided in the TROPOMI Level 2 product, which is used as the reference AMF in the satellite retrieval.

Comment 20: Line 159: How do you define changes in concentrations and emissions? Is it the correction due to the inversion or is it a change between integration time steps? How do the 10% perturbation of line 164 relate to this? Is this defining ΔE?

Response: We have clarified that ΔΩ and ΔE in the FDMB calculation represent finite-difference differences between the prior simulation and a sensitivity simulation with perturbed emissions, rather than temporal changes between model time steps. In this study, ΔE is defined as a prescribed 10 % increase in the prior emissions, and the resulting ΔΩ is used to estimate the FDMB sensitivity factor.

<Section 2.3.1 in the revised manuscript>

Here, $\Delta\Omega_m$ represents the change in modeled column concentrations in response to a prescribed perturbation in emissions, while ΔE_m denotes the imposed emission perturbation used to estimate the sensitivity. In this study, ΔE_m is defined by a 10% perturbation of the prior emissions. […] The sensitivity factor β is calculated from the prior simulation and a perturbed simulation in which emissions are increased by 10%.

Comment 21: Line 178: It is unclear what "employed it as the constraint" means here.

Response: We have revised the wording to clarify the role of the 3D-Var analysis field in the FDMB inversion. Specifically, the 3D-Var analysis field is used as the observational constraint in place of the raw satellite column observations when calculating the FDMB emission scaling factors. This was intended to reduce the direct propagation of observational noise into the top-down emission corrections.

<Section 2.3.1 in the revised manuscript>

To address these challenges associated with observational uncertainties and noise propagation, we incorporated a 3D-Var data assimilation step into our framework to generate a smoothed analysis field as the observational constraint for the FDMB inversion. In this configuration, the 3D-Var step serves as an observational constraint that provides chemically and physically consistent constraints to mitigate error propagation into the top-down emission fields …

Comment 22: Figure 2: It remains unclear what the input to the FDMB really is (connection between green and blue dashed boxes). Furthermore, it is not explained what "i" is in the equation E(i), because there is no dependence on i here. How does the equation E(i) refer to equation (3) in the text? Why is $\Omega_a - \Omega_o$ used here?

Response: We have revised Fig. 2 to clarify the connection between the 3D-Var and FDMB steps. The FDMB step uses the 3D-Var analysis column field as the observational constraint, together with the prior CMAQ column field and the finite-difference sensitivity derived from the prior and perturbed-emission simulations. We also removed the ambiguous E(i) notation from Fig. 2 and revised the figure notation to be consistent with Eq. (3). The column difference used in the FDMB scaling now refers to the mismatch between the analysis column field and the prior model column field.

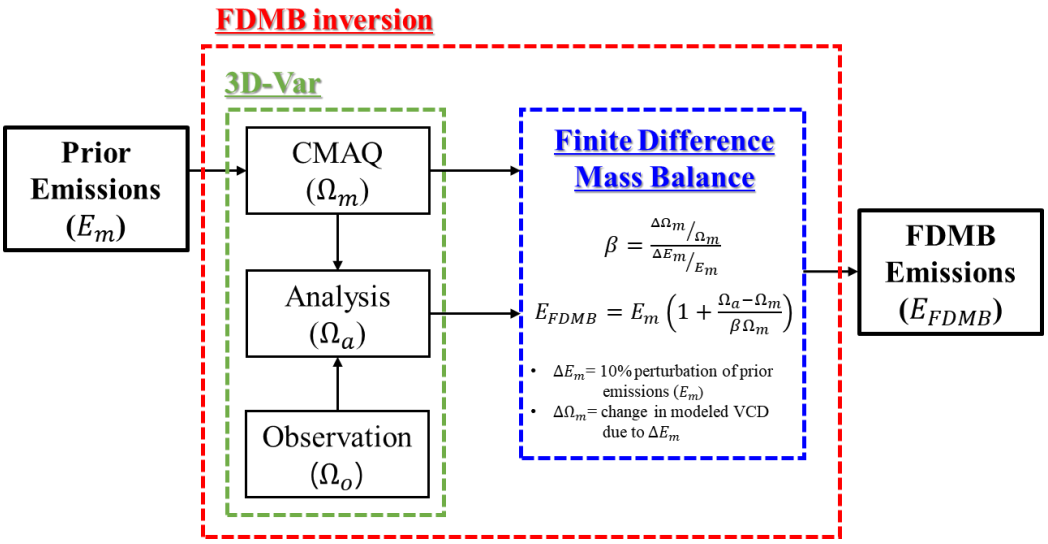

Figure 3: Flowchart of the FDMB inversion framework with 3D-Var assimilation. Red boxes denote the FDMB inversion process, green boxes represent the 3D-Var assimilation (Ω_m: CMAQ VCD, Ω_o: observed VCD, Ω_a: analysis VCD), and blue boxes indicate the FDMB process that updates emissions using finite-difference sensitivities.

Comment 23: Line 187: Please clarify to what the "two-week-averaged column density" refers? Observations or model or both?

Response: We have clarified that the two-week averaging was applied to both the model and observational column densities used in the FDMB inversion.

<Section 2.3.1 in the revised manuscript>

Furthermore, to minimize errors associated with the smearing effect caused by atmospheric transport, the FDMB inversion was performed using two-week-averaged column densities for both the model and observations over the study period.

Comment 24: Line 195: Please check if "(n)" should not be "i=0, n"?

Response: We have revised the notation in Eq. (8) and the accompanying explanation to clarify the time index used in the 4D-Var inversion.

<Section 2.3.2 in the revised manuscript>

Here, n represents the number of hourly time steps within the assimilation window, which was set to 24 in this study. The control variable is the emission scaling factor (α_i) at a specific hourly time step i ($i = 0, \ldots, n-1$).

Comment 25: Line 196 and equation (8): The definition of the observation operator is unclear. The observations are concentrations or column densities, not emissions. Why do you subtract $y_i - H_{i-1}(e_{i-1})$? Why do the indices differ? Please explain i and t.

Response: We have revised Eq. (8) and the accompanying text to clarify the role of the observation operator and the time indices. The previous notation was misleading because it could be interpreted as if the observation was directly compared with emissions. In the revised formulation, the forward model integration operator ($\boldsymbol{M_{i-1\rightarrow i}}$) is explicitly introduced so that the modeled concentration field is first calculated from the previous concentration field and emissions. The observation operator is then applied to this modeled concentration field, not directly to emissions. We also clarified that the concentration field at each time step contains the accumulated effects of earlier emissions, transport, chemical reactions, and initial conditions.

<Equation 8 in the revised manuscript>

$$J(\alpha) = \frac{1}{2}\gamma \sum_{i=0}^{n-1} (\alpha_i - \alpha_b)^T B_i^{-1} (\alpha_i - \alpha_b) + \frac{1}{2} \sum_{i=1}^{n} (y_i - H_i(c_i))^T R_i^{-1} (y_i - H_i(c_i)) \, , where \;\; c_i = M_{i-1\rightarrow i}(c_{i-1}, e_{i-1})$$

<Section 2.3.2 in the revised manuscript>

As explicitly denoted by the forward model integration operator $M_{i-1\rightarrow i}$, the concentration field c_i is obtained by integrating the model from time step $i-1$ to i using the previous concentration field c_{i-1} and emissions e_{i-1}. […] y_i denotes the observation vector at time step i, consisting of hourly AQMS NO_2 and O_3 concentrations, and H_i denotes the observation operator that maps the model concentration field c_i onto the corresponding AQMS observation space. Thus, the observation term compares hourly AQMS observations with the simulated concentrations mapped to the corresponding station locations and times, rather than comparing observations directly with emissions.

Comment 26: Line 198: Please explain the uncertainty of 100%. What does this represent as error considered here?

Response: We have clarified the meaning of the 100 % emission uncertainty in the revised manuscript. In our 4D-Var framework, the control variable is the emission scaling factor (α). The prior scaling factor is uniformly set to 1 ($\alpha_b = 1$). Setting the emission uncertainty to 100% means that we assume a standard deviation of 1.0 (100% of the prior value) for the prior scaling factors. Consequently, the diagonal elements of the background error covariance matrix (B_i) are set to $1.0^2 = 1.0$.

Comment 27: Line 223: Please specify the observations used here.

Response: We have specified the observations in the revised manuscript by explicitly listing the datasets used for the evaluation.

Comment 28: Figure 4: Do not separate "Δ" and "O_3" in a line break (line 265).

Response: We have corrected this in the revised manuscript.

Comment 29: Line 281: The Hybrid_NOx experiment did not change any VOC emissions. Does this mean that there is no adjoint chemistry taken into account in the optimization procedure? Is the optimized emission vector restricted to NOx emissions only in the cost function? Wouldn't it be possible to also derive corrections for VOC emissions when just assimilating NOx observations, due to the chemical coupling in the ozone chemistry in the model? In contrary in line 337 you refer to little improvements for the VOC in the Hybrid_NOx experiment. How is this corrections induced and why is there no correction in Figure 5?

Response: We appreciate the reviewer's insightful questions regarding the experimental design and chemical coupling.

The purpose of the Hybrid_NOx experiment was to isolate the effect of optimizing NOx emissions alone. Therefore, in this experiment, the control vector in the 4D-Var cost function included only NO_x emissions, and VOC emissions were kept unchanged from the prior inventory. Although the CMAQ adjoint framework represents the chemical coupling among NO_x, VOC, and O_3 during the sensitivity calculation, VOC-related gradients were not used to update VOC emissions in the Hybrid_NOx experiment.

Regarding the confusion in Line 337 (now Line 436): the slight change in HCHO VCD observed in the Hybrid_NOx experiment (from an MBE of -2.47 to -2.45) was not due to an update in VOC emissions. Instead, it was a secondary chemical effect: the changes in NO_x emissions altered the atmospheric oxidative capacity (e.g., OH radical concentrations), which in turn slightly modified the photochemical degradation rate of the unaltered prior VOCs, resulting in a marginal change in ambient HCHO concentrations.

We acknowledge that our phrase "little improvement" was misleading and implied a correction in VOC emissions. We have revised the text in Section 3.2 to state that the Hybrid_NOx experiment "showed marginal changes" rather than "little improvement," clarifying that no meaningful correction to the VOC bias occurred until VOC emissions were actively constrained in the Hybrid_NOx+VOC experiment.

<Section 3.2 in the revised manuscript>

The Hybrid_NOx experiment showed marginal changes in HCHO VCDs (MBE = -2.45×10^{15} molec cm^{-2}) because VOC emissions were held constant; the slight difference from the Prior is attributable to secondary changes in atmospheric oxidative capacity driven by the NO_x emission updates.

Comment 30: Figure 5: I would like to suggest difference plots for the analysis increments, because it is not possible to identify if there are only emission increases or also decreases in the posteriors.

Response: We agree that providing difference plots (analysis increments) is essential for clearly identifying the spatial patterns of emission increases and decreases.

To address this, we have reorganized our spatial distribution maps:

(1) In the revised manuscript, Figure 5 now presents the spatial distributions of the Prior emissions, the final Hybrid_NOx+VOC emissions, and their corresponding analysis increments (Hybrid_NOx+VOC – Prior) for NO_x, AVOC, and BVOC.

(2) Additionally, we have provided a new figure in the Supplement (Figure S4) to show the Prior emissions, the intermediate Hybrid_NOx emissions, and their analysis increments (Hybrid_NOx – Prior) specifically for NO_x.

We have also updated the text in Section 3.1 to discuss these spatial patterns in detail.

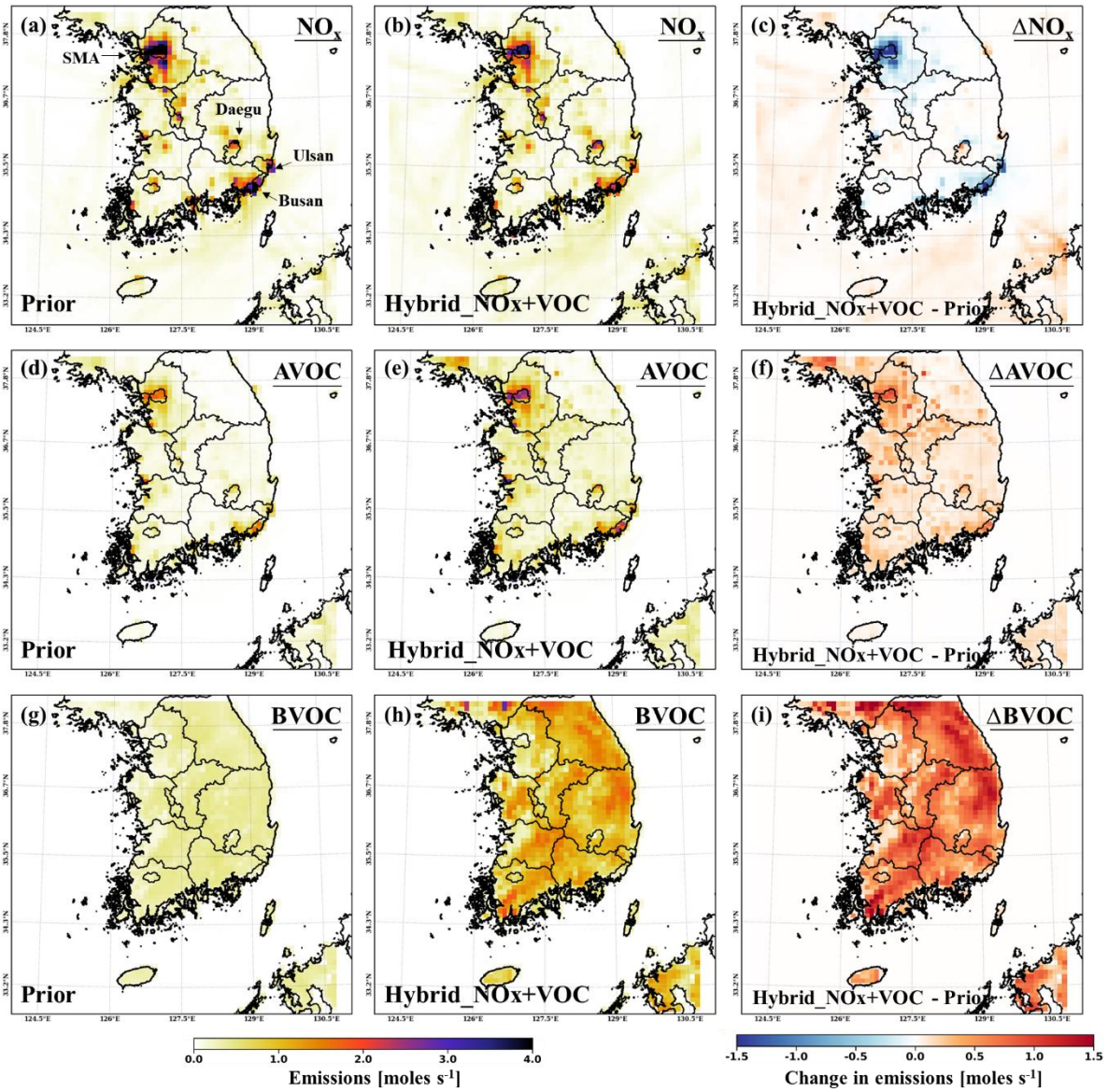

Figure 4: Spatial distributions of (a, d, g) the Prior emissions, (b, e, h) the Hybrid_NOx+VOC emissions, and (c, f, i) the corresponding analysis increments (Hybrid_NOx+VOC – Prior) for NOx, AVOC, and BVOC, respectively, averaged over the study period.

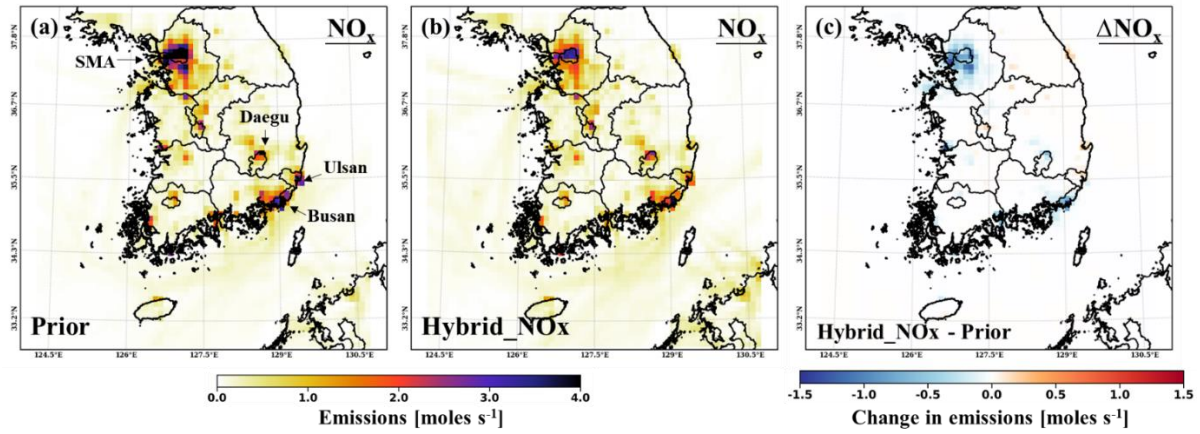

Figure S4: Spatial distributions of (a) the Prior emissions, (b) the Hybrid_NOx emissions, and (c) the corresponding analysis increments (Hybrid_NOx – Prior) for NO_x, averaged over the study period.

Comment 31: Table 1: Please clarify if the values presented are mean values averaged over the all stations and over the analysis period.

Response: We have clarified this in the caption of Table 1.

Comment 32: Figure 7: The caption misses the "comparison to the observations". Also, is the data really averaged over South Korea or over all station locations?

Response: We thank the reviewer for pointing this out. The data are indeed averaged over all station locations, not over the entire geographical area of South Korea. We have corrected this inaccuracy and added the phrase "in comparison to the observations" to the caption of Figure 7.

Comment 33: Line 339: Please add a short explanation of the index of agreement, as this is not a very common and widely used statistical value.

Response: We thank the reviewer for this suggestion. We have added a brief explanation of the Index of Agreement (IOA) in Section 3.2. We clarified that IOA is a standardized measure of the agreement between model simulations and observations, ranging from 0 to 1, where 1 indicates perfect agreement.

<Section 3.2 in the revised manuscript>

In contrast, the Hybrid_NOx+VOC experiment more effectively reduced the bias (MBE = -1.22×10^{15} molec cm^{-2}) and achieved the highest index of agreement (IOA = 0.67), a standardized measure of agreement between model simulations and observations, where a value of 1 indicates perfect agreement.

Comment 34: Line 361: Do you mean the Prior experiment of the Hybrid_NOx approach?

Response: Yes, the reviewer is correct. We referred to the initial Prior experiment (i.e., the baseline simulation before any data assimilation was applied). To avoid any confusion, we have explicitly clarified this by changing the text to "the initial Prior experiment" in the revised manuscript.

Comment 35: Line 368: Please add a reference to Figure S3 here. Otherwise, the reader is wondering why you compare the posterior to TROPOMI and not the Prior.

Response: We have added a reference to Figure S3 (now Figure S6) in the revised manuscript to clearly direct readers to the initial baseline comparison between the Prior experiment and TROPOMI observations.

Comment 36: Line 394: Is it the mean of 1-14 May 2022?

Response: Yes, the reviewer is correct. To be precise, the values represent the hourly ΔO_3 responses that are averaged for each specific hour of the day over the 14-day period (1–14 May 2022). We have clarified this in the revised manuscript to avoid any misinterpretation.

<Section 3.4 in the revised manuscript>

Figure 10 shows regime-stratified hourly ΔO_3 responses to NO_x and VOC emissions, averaged by hour of the day over the period of 1–14 May 2022.

Comment 37: Lines 394-395: "Panels (a) and (c) show, for each local hour, the ΔO_3 at that hour resulting from emissions integrated over all emission times (Eq. (12))." Please explain, what this means. What is the local hour?

Response: We apologize for the ambiguity. "Local hour" refers to the specific time of day in Korean Standard Time (KST). The phrase "integrated over all emission times" means that the ΔO_3 at a given hour is the cumulative response to all precursor emissions released prior to and during that specific hour. We have rephrased this sentence in the revised manuscript to clearly explain this cumulative physical effect.

Comment 38: Figure 9: Why is there missing data in the model analysis? Additionally, is there no data over the ocean?

Response: We have addressed the reviewer's questions regarding the data coverage in Figure 9. The missing pixels in both the TROPOMI observations and the model simulation are due to the application of strict quality assurance (QA) filters. Furthermore, our analysis focused on the O_3 sensitivity regimes driven by terrestrial emissions over the Korean Peninsula; therefore, we applied a land-mask and excluded oceanic regions from the comparison. We have clarified these points in the revised caption of Figure 9 and Figure S6.

Comment 39: Line 411: Please double check if the first sentence of this paragraph can be expressed clearer after addressing my concerns described in the major comment (1).

Response: We have revised the sentence to explicitly state that 15:00 KST was selected as the receptor time because it represents the hour of maximum O_3 concentration. The updated text clarifies that Panels (b) and (d) illustrate the sensitivity of O_3 concentration at 15:00 KST to the timing of precursor emissions, allowing for a direct comparison of how hourly emissions contribute to the O_3 concentration at that specific hour.

<Section 3.4 in the revised manuscript>

Panels (b) and (d) focus on 15:00 KST, the hour of maximum O_3 concentration, to illustrate the sensitivity of O_3 at this hour to the timing of precursor emissions.

Comment 40: Line 433: Please explain how this approach reduces data and why the computational demands decrease?

Response: We agree with the reviewer that further clarification on the efficiencies of the top-down approach is beneficial. Compared with conventional bottom-up inventories, which require the labor-intensive collection of detailed activity data and source-specific emission factors, the top-down approach directly utilizes atmospheric observations to constrain precursor emissions. This bypasses the exhaustive primary data-gathering stage and significantly lowers overall time and computational costs by employing inverse modeling-based estimation. We have revised the text to explicitly explain these advantages.

<Section 3.4 in the revised manuscript>

Compared with conventional bottom-up inventories, the top-down approach reduces the need for detailed activity data and source-specific emission factors by directly using atmospheric observations to constrain precursor emissions, thereby lowering time and computational costs through inverse modeling-based estimation.

Comment 41: Line 444: Are their any ideas to what could be expected from a Hybrid_VOC experiment?

Response: We appreciate the reviewer's comment. A Hybrid_VOC experiment would be expected to mainly influence daytime O_3 concentrations, particularly in VOC-sensitive urban regions where O_3 production is strongly controlled by VOC availability. In contrast, its effect on nighttime O_3 would likely be smaller because nighttime O_3 is more directly affected by NO titration than by local photochemical production. We have clarified this point by noting that the limited daytime improvement in the Hybrid_NOx experiment highlights the need for jointly constraining NO_x and VOC emissions.

Comment 42: Lines 453 and 454: Please add "emissions" after "NO_x" and "AVOC and BVOC".

Response: We have added "emissions" after NO_x, AVOC, and BVOC to improve clarity.

Comment 43: Line 470: Please revise "NO_x contributes" with "NO_x emissions contribute".

Response: We have changed "NO_x contributes" to "NO_x emissions contribute" to improve clarity.

References

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., Schieberle, C., Friedrich, R., Janssens-Maenhout, G.: High resolution temporal profiles in the Emissions Database for Global Atmospheric Research, Sci. Data, 7 (1), 121, http://dx.doi.org/10.1038/s41597- 020-0462-2, 2020.

Moon, J., Choi, Y., Jeon, W., Kim, H. C., Pouyaei, A., Jung, J., Pan, S., Kim, S., Kim, C.-H., Bak, J., Yoo, J.-W., Park, J., and Kim, D.: Hybrid IFDMB/4D-Var inverse modeling to constrain the spatiotemporal distribution of CO and NO_2 emissions using the CMAQ adjoint model, Atmos. Environ., 327, 120490, https://doi.org/10.1016/j.atmosenv.2024.120490, 2024.

Reply to the comments of RC2

RC2: General comment

The manuscript by Moon et al. presents the application of a recently developed hybrid (mass balance / 3D-Var / 4D-Var) inverse modeling scheme to constrain emissions of NO_x and VOC emissions in Korea. The work is new and noteworthy, and includes an analysis of O_3 regimes using adjoint sensitivities. Some of the results, such as the agreement of the posterior simulations with satellite-based assessments of NO_x vs VOC limited regimes, are really quite impressive.

The inversion itself could use a bit more evaluation, given that there aren't any independent measurements of NO_2 or O_3 used, and the VOC assessment is limited to 5 Pandora sites. Meanwhile, the posterior simulation shows dramatic BVOC emission increases throughout the country that aren't really evaluated — checking isoprene levels would seem to be an essential step here. These and some other issue are described in my comments below.

Response: We thank the reviewer for the constructive and helpful comments on our manuscript. We have revised the manuscript accordingly. Below, we provide point-by-point responses to each comment. Reviewers' comments are shown in plain text, our responses are shown in bold, and revised manuscript text is shown in *italics*. Line numbers refer to the revised manuscript.

Major comment 1

Section 2.3.1:

- Why apply FDMB to NO_x and VOCs sequentially, rather than simultaneously, and iteratively? See e.g. Choi, J., D. K. Henze, K. C. Wells, and D. B. Millet, Joint inversion of satellite-based isoprene and formaldehyde observations to constrain emissions of non-methane volatile organic compounds, J. Geo-phys. Res., 130, 13, e2024JD042070, https://doi.org/10.1029/2024JD042070.

- As noted above, why not iterate the FDMB, to minimize smearing and nonlinear errors and also pickup some feedback of the VOC constraints on the NOx?

- For the FDMB part, what drove the decision to assimilate NO_2 first and then HCHO, rather than the other way around?

Response: We sincerely thank the reviewer for this insightful comment and for bringing the recent study by Choi et al. (2025) to our attention. We agree that simultaneous and iterative inversion approaches can be valuable for reducing smearing errors and accounting for nonlinear feedbacks when the observational constraints and target emissions are sufficiently well defined. In the present study, however, the FDMB step was intentionally designed as a single-step spatial preconditioning procedure for the subsequent 4D-Var optimization, rather than as a fully iterative standalone inversion. We have clarified this point in the revised manuscript.

(1) Sequential application of FDMB

We applied the FDMB correction sequentially because NO_2 and HCHO provide fundamentally different types of observational constraints. Tropospheric NO_2 columns are more directly related to NO_x emissions, whereas HCHO is a secondary product of VOC oxidation, with production yields that depend non-linearly on NO_x-dependent photochemistry and oxidative capacity (Wolfe et al., 2016). Therefore, establishing an accurate NO_x baseline first builds a more consistent chemical background, thereby reducing the risk of attributing NO_x-related HCHO biases to VOC emissions.

Structurally, the conventional FDMB method relies on a localized, diagonal one-to-one proportional assumption between a precursor emission and its primary column indicator. It does not explicitly account for cross-sensitivities between precursor emissions and secondary products (e.g., the sensitivity of HCHO columns to NO_x emissions). If we were to perform a simultaneous joint inversion or a reversed sequence within the FDMB step, this highly non-linear relationship would be approximated within a simplified finite-difference scaling framework, which could misattribute chemically coupled HCHO responses and potentially amplify top-down emission errors. By pre-conditioning NO_x emissions first, the forward model explicitly updates the chemical baseline, providing a more consistent background field before evaluating the HCHO-to-VOC mass balance.

(2) Difference from Choi et al. (2025)

We note that the study by Choi et al. (2025) differs fundamentally from ours in its inversion target. Their joint inversion used isoprene and HCHO observations to constrain NMVOC emissions, whereas our framework targets both NO_x and VOC emissions. Because simultaneous NO_x and VOC correction within FDMB would likely be more affected by shifting NO_x regimes, it could increase the risk of misattributing chemically coupled HCHO responses to VOC emissions.

(3) Single-step rather than iterative FDMB

We did not iterate the FDMB step because the method can be vulnerable to the amplification of satellite retrieval noise and smearing effects, particularly at the fine 9 km resolution used in this study. Furthermore, previous pseudo-observation tests conducted under the same model domain and horizontal resolution indicated that iterative FDMB (IFDMB) corrections can increase emission errors when residual observational noise is propagated through repeated updates (Moon, 2025). Therefore, we utilized FDMB strictly as a single-step spatial preconditioning tool. The remaining multi-species non-linear coupling among NO_x, VOCs, and O_3 is then further constrained in our subsequent 4D-Var inversion step using the full CMAQ chemistry.

We have revised Section 2.3.1 to clarify the rationale for applying the NO_2-based NO_x correction before the HCHO-based VOC correction, and to explain why the

FDMB step was applied as a single-step spatial preconditioning procedure rather than as a simultaneous iterative inversion.

<Section 2.3.1 in the revised manuscript>

In this study, we applied the FDMB inversion to constrain NOx and VOC emissions. To mitigate potential inversion errors arising from the highly non-linear dependence of HCHO production on background NO_x concentrations (Wolfe et al., 2016), we sequenced the initial mass-balance step by optimizing NO_x prior to VOCs. Specifically, we first derived an emission factor for NO_x based on the sensitivity of NO_2 columns to NO_x emissions (Eq. (5)). Establishing this updated NO_x baseline in the forward model effectively reduces the non-linearities associated with HCHO yields. Following this, we corrected VOC emissions using HCHO columns as observational constraints. For this step, total VOC emissions were separated into anthropogenic (AVOC) and biogenic (BVOC) categories, as their distinct species compositions lead to different HCHO responses (Millet et al., 2006; Choi et al., 2022; Oomen et al., 2024).

While iterative applications of the mass balance method can theoretically reduce spatial smearing errors, localized scaling factors become highly vulnerable to overfitting this observation noise when applied at fine model resolutions, such as our 9 km grid. A pseudo-observation test conducted by Moon (2025) demonstrated that iterating the FDMB framework at this high resolution paradoxically increased emission errors due to the propagation of residual noise through repeated updates.

Major comment 2

Eq 8: It seems to me that e_i is an emission within some time window defined as t to $t-1$, and n is the total number of time windows. It would be useful to clarify this, and also state what is the time window (1 hour? 1 day?) and how many are there in total (n)? Also, relating y_i to $H_{i-1}(e_{i-1})$ doesn't seem correct, unless the time window between i and i-1 is longer than the lifetime of the species in question. Otherwise, the estimate of the observation at y_i would be impacted by emissions at times i-1, i-2, i-3, etc.

I guess from line 201 that the difference between time i and i-1 is 24 hours, in which case it is very incorrect when assimilating AQMS O_3 observations to assume that only emission e_{i-1} would impact the model estimate that corresponds to observation y_i. I don't know if this is just an issue with how the authors wrote out this equation, or if they actually conducted the inversion this way (which wouldn't seem correct); hopefully just the former.

Response: **We sincerely thank the reviewer for this important and insightful comment. We agree that the original notation in Eq. (8) was misleading and did not sufficiently describe how the temporal evolution of emissions and concentrations was represented in the 4D-Var inversion. In particular, the previous expression could be interpreted as if the simulated observation at time i depended only on the emission at the immediately preceding time step, e_{i-1}. This was an issue with the notation and description, not with the actual implementation of the inversion.**

In the actual 4D-Var framework, the simulated observation is obtained from the evolving CMAQ concentration field, which is generated through forward model integration. Therefore, the concentration field at each observation time includes the accumulated effects of previous emissions, atmospheric transport, chemical transformation, and initial conditions. To clarify this point, we revised Eq. (8) by explicitly introducing the forward model integration operator, $M_{i-1\rightarrow i}$, such that the concentration field at time step i, c_i, is calculated from the previous concentration field and emissions through model integration. This revised notation makes clear that the observation term is evaluated using the modeled concentration field, rather than assuming a direct one-step relationship between the observation and only the immediately preceding emission.

We also clarified the meaning of the time index and assimilation window. In this study, the assimilation window was set to 24 hours, and the control variables were defined at hourly time steps within each window. Thus, n represents the number

of hourly time steps within the assimilation window, with n = 24 for each window. The observation term is evaluated from i = 1 to n, corresponding to concentrations generated by emissions from i = 0 to n−1 through the forward model integration. This structure allows the adjoint backward integration to account for the effects of atmospheric transport and chemical processing over the full assimilation window.

Accordingly, we revised Eq. (8) and the accompanying text in Section 2.3.2 to clarify the definitions of n, e_i, α_i, c_i, and the forward model operator. We also added an explanation that the concentration field contains the cumulative effects of previous emissions, transport, chemistry, and initial conditions, thereby addressing the reviewer's concern.

<Section 2.3.2 in the revised manuscript>

$$J(\alpha) = \frac{1}{2}\gamma \sum_{i=0}^{n-1}(\alpha_i - \alpha_b)^T B_i^{-1}(\alpha_i - \alpha_b) + \frac{1}{2}\sum_{i=1}^{n}(y_i - H_i(c_i))^T R_i^{-1}(y_i - H_i(c_i)), where\ \ c_i = M_{i-1\to i}(c_{i-1}, e_{i-1}) \qquad (8)$$

Here, n represents the number of hourly time steps within the assimilation window, which was set to 24 in this study. The control variable is the emission scaling factor (α_i) at a specific hourly time step i ($i = 0, \ldots, n-1$). The emission scaling factor ($\alpha_i = e_i/e_i^b$) represents the ratio between the updated and prior emissions at time step i. e_i denotes the updated emissions at time step i, and e_i^b denotes the corresponding prior emissions. c_i represents the simulated concentration field at time step i. As explicitly denoted by the forward model integration operator $M_{i-1\to i}$, the concentration field c_i is obtained by integrating the model from time step $i-1$ to i using the previous concentration field c_{i-1} and emissions e_{i-1}. Because c_{i-1} already contains the accumulated effects of earlier emissions, transport, chemical transformation, and initial conditions, c_i reflects the cumulative atmospheric history up to time step i rather than the influence of e_{i-1} alone. Consequently, to appropriately account for the transport and chemical processing time in the adjoint backward integration, the observation term is evaluated from $i = 1$ to n, corresponding to the concentrations driven by emissions from $i = 0$ to $n-1$.

Major comment 3

General: Why is this hybrid approach better than just applying 4D-Var to assimilate satellite and AQMS measurements simultaneously? It seems like the addition of a 3D-Var and FDMB inversion as a precursor to this add computational cost and complexity, but I'm not sure if the value has been explained or demonstrated. If, for example, this approach for example allows the 4D-Var inversion to converge in fewer iterations, that should be stated (and demonstrated).

Response: We sincerely thank the reviewer for this sharp and insightful assessment regarding the value of our hybrid framework compared to a standalone 4D-Var approach. We agree that introducing 3D-Var and FDMB as pre-conditioning steps adds computational complexity, and therefore, their methodological value and effects must be clearly stated. We have detailed our rationale below and added these justifications to the revised manuscript:

(1) Role of FDMB as spatial preconditioning before 4D-Var

In our previous observing system simulation experiment (OSSE) (Moon et al., 2024), we quantitatively compared Mass balance, 4D-Var, and hybrid inversions. The results demonstrated that a standalone 4D-Var approach struggles structurally to effectively correct emissions in grids where the spatial error of the prior emissions is exceptionally large. Because this limitation was evident even in experiments using gridded synthetic observations, directly assimilating actual satellite and AQMS data into a standalone 4D-Var system may face similar challenges. In contrast, the Mass balance approach is highly effective at correcting the spatial distribution of emissions but inherently lacks the ability to optimize temporal variations. By sequentially combining them, the FDMB step establishes an improved spatial baseline. This preconditioning allows the subsequent 4D-Var step to focus more effectively on temporal variability and multi-species chemical feedbacks while reducing the influence of large prior spatial errors.

(2) Role of 3D-Var in reducing observational noise before FDMB

As mentioned in our response to Comment 1, while the FDMB method excels at spatial scaling, it is highly vulnerable to the amplification of satellite retrieval noise and smearing effects at our fine 9 km resolution. If raw satellite data were fed directly into the inversion, this observational noise would propagate directly into

the top-down emission fields, ironically increasing the overall error. To resolve this, we adopted the approach of East et al. (2022) by introducing a 3D-Var assimilation step. 3D-Var acts as a pre-filter that provides chemically and physically consistent constraints for the FDMB calculation, thereby mitigating the propagation of observational uncertainties (East et al., 2022). Furthermore, our pseudo-observation experiments (Moon, 2025) quantitatively confirmed that applying 3D-Var to block noise propagation effectively reduces the final emission errors.

We have added detailed descriptions of the rationale for the hybrid approach in Section 2.3.3 and the error reduction achieved by incorporating 3D-Var in Section 2.3.1 of the revised manuscript.

<Section 2.3.1 in the revised manuscript>

To address these challenges associated with observational uncertainties and noise propagation, we incorporated a 3D-Var data assimilation step into our framework to generate a smoothed analysis field as the observational constraint for the FDMB inversion. In this configuration, the 3D-Var step serves as an observational constraint that provides chemically and physically consistent constraints to mitigate error propagation into the top-down emission fields (East et al., 2022), allowing the FDMB inversion to establish a robust spatial baseline (Fig. 2).

<Section 2.3.3 in the revised manuscript>

A model-based twin experiment by Moon et al. (2024) demonstrated that a standalone 4D-Var approach structurally struggles to effectively correct emissions in grid cells with exceptionally large prior spatial errors. By sequentially combining these methods, the FDMB step establishes a robust and accurate spatial baseline, which enables the subsequent 4D-Var step to focus more effectively on optimizing diurnal temporal variations and multi-species chemical feedbacks, while reducing the influence of spatial distribution errors in the prior emissions.

Major comment 4

225: I suppose it will be mentioned later what observations are used for the evaluation (presumably some that aren't used for the assimilation…) but maybe that could be stated here as well? Ok, I see later in section 3.2 that they are evaluating against AQMS NO_2, O_3, and Pandora HCHO. But since AQMS NO_2 and O_3 were used in the assimilation, this is not a good test of the robustness of the posterior emissions. It does show that the modeling framework was correctly implemented, but one should evaluate against independent data. It seems that was only done with Pandora HCHO.

Pandora would have NO_2 as well — why was this not used?

Also, are there not any other VOC measurements in Korea to compare against? What about from KORUS-AQ or other? The BVOC emission changes shown in Fig 5 seem very large — can these be evaluated against any isoprene observations?

From CrIS? In Choi et al. (2025), Fig 10, it was found that HCHO assimilation led to very high isoprene concentrations compared to KORUS-AQ aircraft observations; however, assimilation of isoprene did not lead to these large over-estimates, nor did joint isoprene-HCHO inversions. I thus am concerned that that the large BVOC emissions found here might be compensating for a different issue in the model.

Response: **We thank the reviewer for this important comment. The key revisions are as follows.**

(1) Independent validation using withheld AQMS sites

We agree that evaluating posterior simulations only against the same AQMS observations used in the inversion does not provide a fully independent assessment of the posterior emissions. To address this concern, we conducted an additional validation experiment using a data-splitting approach, as described in Section 2.2.1 of the revised manuscript. Among the 619 AQMS sites, 496 sites (80%) were randomly selected for the inversion, while the remaining 123 sites (20%) were withheld from the inversion and used exclusively for independent validation. The spatial distribution of the inversion and validation sites has been added as Fig. S2.

<Section 2.2.1 in the revised manuscript>

To assess the robustness of the posterior emissions using independent observations, an additional validation experiment was conducted using a data-splitting approach, in which 496 sites (80%) were randomly selected for the inversion and the remaining 123 sites (20%) were

reserved for independent validation. The spatial distribution of the selected sites is shown in Fig. S2.

<Section 3.2 in the revised manuscript>

The IOA improved from 0.71 (Prior) to 0.80 (Hybrid_NOx) and 0.83 (Hybrid_NOx+VOC), indicating improved consistency with the AQMS observations used in the baseline inversion. […] As described in Sect. 2.2.1, the independent validation experiment produced similar improvements at the withheld AQMS sites, indicating that the posterior emission corrections were not solely a result of fitting the assimilated observations. Compared with the Prior experiment, the Hybrid_NOx+VOC experiment reduced the RMSEs for both NO_2 and O_3 and improved their temporal agreement with independent AQMS observations (Table S6). These results support the robustness of the posterior emissions and indicate that the hybrid inversion framework improves O_3 simulations beyond the stations directly used in the inversion.

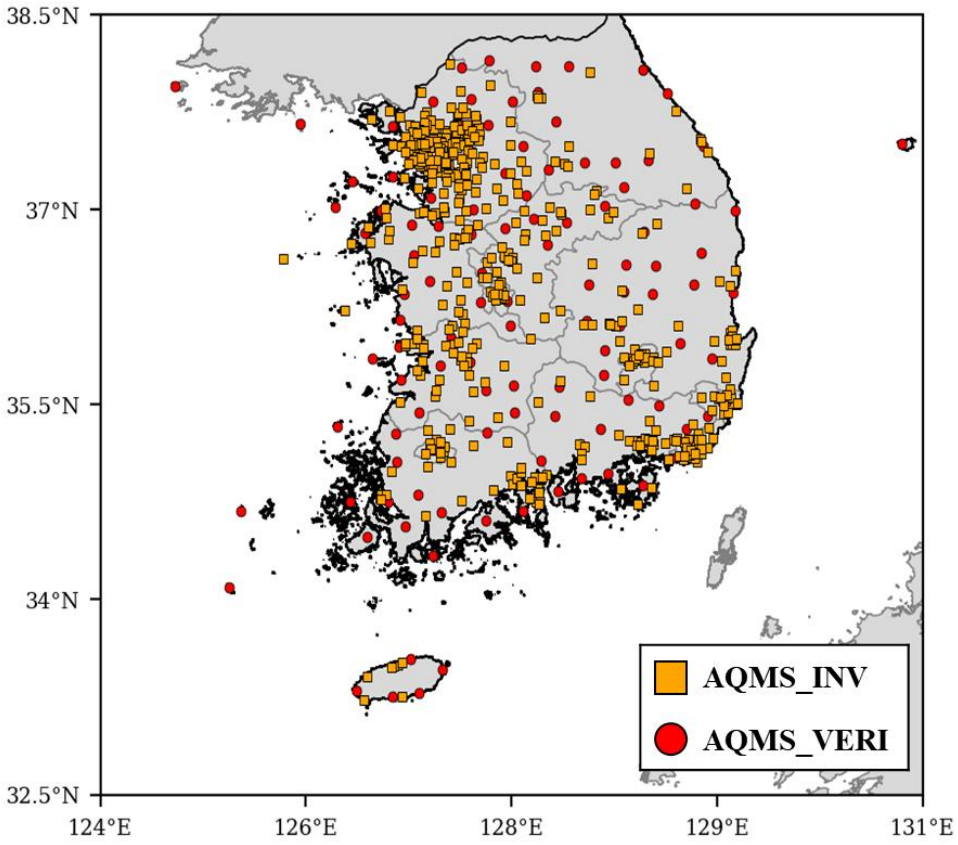

Figure S2: Spatial distribution of AQMS sites used in this study. A total of 619 sites were considered, among which 496 sites (80%) were randomly selected for the inversion (AQMS_INV; orange squares) and the remaining 123 sites (20%) were reserved for independent validation (AQMS_VERI; red circles).

Table S6. Statistical evaluation of NO_2 and O_3 concentrations for the Prior and Hybrid_NOx+VOC experiments at the independent validation sites. The statistics were calculated over the 123 withheld AQMS_VERI sites shown in Fig. S2 and averaged over the entire study period. Observations were obtained from the AQMS network.

Species	Experiment	Obs.	CMAQ	MBE	RMSE	IOA	r
NO_2 [ppb]	Prior	8.69	9.53	0.84	11.50	0.58	0.44
	Hybrid_NOx+VOC		6.58	-2.11	7.33	0.67	0.48
O_3 [ppb]	Prior	44.48	43.54	-3.02	16.60	0.71	0.53
	Hybrid_NOx+VOC		47.42	0.86	13.65	0.77	0.64

(2) Additional evaluation using Pandora NO_2 columns

Following the reviewer's suggestion, we additionally compared the model results with Pandora NO_2 column observations. The Hybrid_NOx+VOC experiment showed a slight reduction in RMSE relative to the Prior experiment, but the MBE increased, and the overall improvement was limited.

The limited response in Pandora NO_2 may reflect differences in representativeness and temporal sampling between Pandora column observations and the surface observations assimilated in the 4D-Var inversion. In particular, the NO_x emission adjustment in this study was constrained by both surface NO_2 and O_3 observations. A substantial part of the NO_x correction was associated with reducing nighttime NO_2 overestimation and weakening nighttime O_3 titration by NO, whereas Pandora direct-sun NO_2 observations mainly provide daytime column constraints. This interpretation is also consistent with the AQMS NO_2 comparison, which showed only minor changes in daytime NO_2 concentrations after the inverse modeling. We have added this comparison and clarified its interpretation in the revised manuscript.

<Section 2.2.1 in the revised manuscript>

In addition, tropospheric NO_2 and HCHO Vertical Column Densities (VCDs) were obtained from Pandora spectrometers at five sites operated by the Pandonia Global Network (PGN) to evaluate the model's performance in simulating NO_x and VOC-related column concentrations. To ensure data reliability, only Pandora NO_2 retrievals with Level 2 (L2) data quality flags classified as "high" quality (flags 0 and 10) were used. For HCHO, L2 retrievals classified as "high" quality (flags 0 and 10) or "medium" quality (flags 1 and 11) were used in this study (Bae et al., 2025; Fu et al., 2025). The Pandora observations were hourly averaged for comparison with model results.

<Section 3.2 in the revised manuscript>

As an additional column-based evaluation, the simulations were compared with Pandora NO_2 VCD observations. The Hybrid_NOx+VOC experiment slightly reduced the RMSE relative to the Prior experiment, but the MBE increased and the IOA and correlation changed only marginally. This limited response is consistent with the comparison against daytime AQMS NO_2 concentrations (Fig. 7), which showed only minor changes after the inverse modeling because the 4D-Var inversion mainly adjusted NO_x emissions to correct nighttime NO_2 overestimation rather than daytime NO_2 levels.

Table 1: Statistical evaluation of AQMS NO_2, AQMS O_3, Pandora NO_2 VCD, and Pandora HCHO VCD for each experiment (Prior, Hybrid_NOx, and Hybrid_NOx+VOC). The statistics were calculated over the corresponding observation locations and averaged over the entire study period. Surface NO_2 and O_3 observations were obtained from the AQMS network, while NO_2 and HCHO VCD observations were obtained from Pandora measurements.

Species	Experiment	Obs.	CMAQ	MBE	RMSE	IOA	r
AQMS NO_2 [ppb]	Prior	12.22	15.56	3.34	15.49	0.59	0.46
	Hybrid_NOx		9.00	-3.22	8.33	0.74	0.60
	Hybrid_NOx+VOC		9.41	-2.81	8.26	0.76	0.61
AQMS O_3 [ppb]	Prior	44.48	38.20	-6.29	18.30	0.71	0.55
	Hybrid_NOx		43.74	-0.74	12.84	0.80	0.70
	Hybrid_NOx+VOC		44.72	0.24	12.34	0.83	0.73
Pandora NO_2 VCD [10^{15} molec cm^{-2}]	Prior	13.01	12.11	-0.91	8.81	0.74	0.66
	Hybrid_NOx		10.64	-2.38	8.59	0.73	0.62
	Hybrid_NOx+VOC		10.70	-2.31	8.35	0.74	0.65
Pandora HCHO VCD [10^{15} molec cm^{-2}]	Prior	7.06	4.59	-2.47	4.40	0.54	0.52
	Hybrid_NOx		4.61	-2.45	4.36	0.55	0.53
	Hybrid_NOx+VOC		5.85	-1.22	3.94	0.67	0.52

(3) Availability of independent VOC observations (KORUS-AQ)

We agree that independent VOC observations would be valuable for evaluation. However, the KORUS-AQ campaign was conducted in 2016, which does not match our analysis period of May 2022. Therefore, we could not utilize these observations for this study.

(4) CrIS isoprene comparison and interpretation of large BVOC adjustments

To further address the reviewer's concern regarding the large BVOC emission increase, we compared the simulated isoprene vertical column densities with the CrIS isoprene retrieval product from Wells and Millet (2022). The CrIS product provides a multi-year May mean for 2012–2020; therefore, we compared the May

mean CrIS isoprene VCDs with the model results. The Prior experiment generally underestimated isoprene VCDs relative to CrIS, whereas the Hybrid_NOx+VOC experiment produced column magnitudes closer to the CrIS observations. However, spatial discrepancies remained, with the model showing stronger localized enhancements over inland South Korea. These discrepancies may reflect differences in temporal representativeness, as CrIS represents a 2012–2020 May mean, whereas the CMAQ simulations represent a specific May 2022 episode, as well as uncertainties in BVOC emissions, chemical oxidation, and satellite retrievals. Therefore, this comparison is interpreted as a qualitative consistency check rather than a direct validation of absolute BVOC emissions.

We agree that the large BVOC adjustments should be interpreted cautiously. HCHO columns reflect the combined effects of VOC emissions, chemical production yields, oxidative capacity, transport, and retrieval uncertainties. Therefore, HCHO-constrained BVOC corrections may partly compensate for uncertainties in model chemistry or oxidation processes, as noted by the reviewer and as also suggested by Choi et al. (2025). We have therefore revised the manuscript to state this limitation explicitly and to clarify that the optimized BVOC emissions should not be interpreted as an independent evaluation of absolute BVOC emission magnitudes.

<Section 3.1 in the revised manuscript>

Nevertheless, the large BVOC adjustment should be interpreted with caution. Because HCHO columns indirectly constrain VOC emissions, the inversion may partly compensate for uncertainties in model chemistry, oxidative capacity, transport, or retrieval errors rather than reflecting BVOC emission errors alone (Choi et al., 2025). To evaluate the consistency of the optimized BVOC emissions, we compared the modeled isoprene VCDs against the multi-year May mean CrIS retrievals for 2012–2020 (Wells and Millet, 2022). Although the Hybrid_NOx+VOC experiment produced isoprene VCD magnitudes closer to the CrIS retrievals than the Prior experiment (MBE decreased from -7.16×1014 molec cm^{-2} in the Prior experiment to -3.73×1014 molec cm^{-2} in the Hybrid_NOx+VOC experiment), spatial differences remained (Fig. S5). These discrepancies are likely related, at least in part, to the temporal mismatch between the multi-year May mean CrIS observations and our specific May 2022 episodic simulation. Therefore, the optimized BVOC emissions in this study should be viewed as HCHO-constrained effective VOC corrections within the model framework, rather than as an independent evaluation of absolute BVOC emission magnitudes.

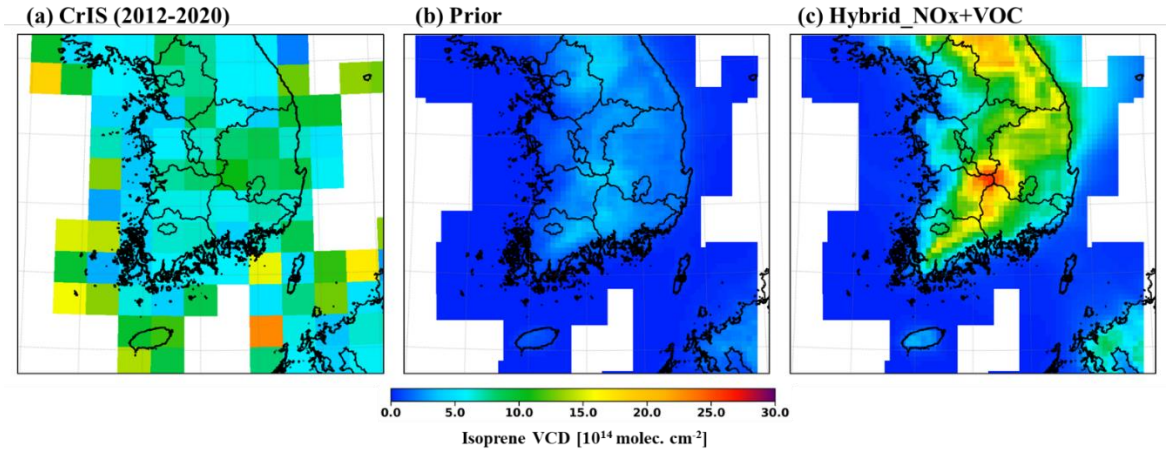

Figure S5: Spatial distributions of isoprene vertical column densities (VCDs) from (a) the May mean CrIS retrievals for 2012–2020, (b) the Prior experiment, and (c) the Hybrid_NOx+VOC experiment. The CMAQ results are averaged over the study period.

Major comment 5

Fig 9: The consistency here is quite impressive! Especially given the differences evident in the prior model (Fig S3). I liked this result a lot.

Response: We thank the reviewer for the positive comment.

Major comment 6

Fig 10 / Conclusions: I also found this figure very interesting and a useful way to present information. Can the authors explain the chemical mechanism whereby emissions of VOCs lead to O_3 depletion under NO_x-sensitive conditions? I also wonder if the authors care to speculate about what might be effective emissions control strategies, given these results? Should urban NO_x emissions controls be stricter at night, for example? Or what if one used the prior model to make a policy recommendation — how incorrect might that have been, compared to using the posterior results

Response: Thank you for your insightful comments and for recognizing the value of our adjoint-based analysis. We fully agree that addressing these points significantly enhances the practical implications of our study.

(1) Mechanism of nighttime VOC-induced O_3 depletion under NO_x-sensitive conditions

The slight O_3 depletion (negative ΔO_3 response) caused by VOC emissions under NO_x-sensitive conditions occurs primarily during the nighttime. To explicitly demonstrate this, we further separated the VOC sensitivity into anthropogenic VOC (AVOC) and biogenic VOC (BVOC) contributions, as now shown in Fig. S7. The results show that both AVOC and BVOC contribute to O_3 production in the VOC-sensitive regime. In contrast, under NO_x-sensitive conditions, the negative nighttime response is mainly associated with the BVOC category, suggesting that reactive biogenic species can act as an O_3 sink during nighttime.

The chemical mechanism driving this response is direct ozonolysis. In our domain, NO_x-sensitive regimes are often associated with vegetated areas where BVOC emissions can be substantial. Under nighttime conditions, when sunlight is

unavailable to drive NO_2 photolysis and regenerate O_3, reactive unsaturated VOCs represented within the BVOC category can react directly with O_3 and act as net gas-phase chemical sinks, leading to O_3 depletion. Such VOC-O_3 reactions are well established in gas-phase tropospheric chemistry, and forest observations have shown that gas-phase reactions involving biogenic hydrocarbons can represent an important O_3 loss pathway (Atkinson and Arey, 2003; Kurpius and Goldstein, 2003). Accordingly, we have added Fig. S7 to the Supplementary Information and expanded the discussion of this nighttime BVOC-related O_3 sink in Section 3.4.

<Section 3.4 in the revised manuscript>

A slight negative VOC response also appears during nighttime under NO_x-sensitive conditions. To clarify this response, we further separated the VOC sensitivity into AVOC and BVOC contributions (Fig. S7). The results indicate that this nighttime O_3 decrease is mainly associated with the BVOC category. This response can be explained by direct ozonolysis of reactive unsaturated VOC species represented within the BVOC category. Under nighttime conditions, when NO_2 photolysis and photochemical O_3 production are suppressed, reactions between O_3 and reactive VOC species can act as a net gas-phase O_3 sink, leading to O_3 depletion. Previous laboratory and field studies provide a chemical basis for this interpretation, showing that reactive BVOCs can undergo ozonolysis and that gas-phase reactions involving biogenic hydrocarbons can contribute to O_3 loss in forest environments (Atkinson and Arey, 2003; Kurpius and Goldstein, 2003).

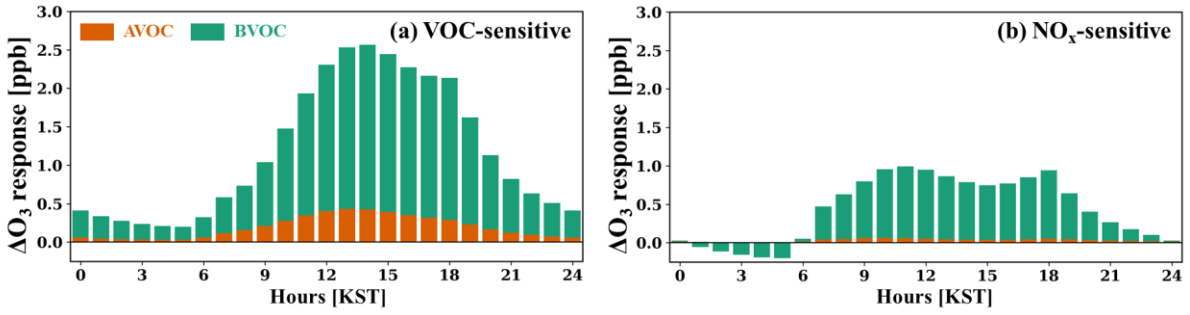

Figure S7: Two-week mean ΔO_3 response (ppb) to anthropogenic VOC (AVOC, orange) and biogenic VOC (BVOC, green) emissions, by local hour, for 1–14 May 2022. For each local hour, responses are summed over all grids classified in each regime at that hour; regimes are diagnosed with the HCHO-to-NO_2 ratios from the Hybrid_NOx+VOC experiment. Panels (a) and (b) show, for VOC-sensitive and NO_x-sensitive regimes respectively, the ΔO_3 response at each local hour to emissions released over all hours (i.e., emission-time–integrated response).

(2) Policy implications of regime- and time-dependent precursor controls

Figure 10 suggests that effective emission control strategies should consider both the chemical sensitivity regime and the timing of precursor emissions. In VOC-sensitive regions, the NO_x response is negative throughout the day, with particularly strong negative responses during nighttime. This indicates that nighttime NO_x emissions reduce O_3 mainly through NO titration. Therefore, stricter nighttime NO_x controls in VOC-sensitive urban areas would not necessarily lead to immediate O_3 reductions; instead, they could weaken NO titration and increase nighttime O_3. In these regions, VOC reductions are expected to be more effective for reducing daytime O_3 formation, as indicated by the positive daytime VOC response in Fig. 10a and the strong contribution of VOC emissions released before the afternoon O_3 peak in Fig. 10b.

In contrast, under NO_x-sensitive conditions, NO_x emissions generally promote daytime O_3 formation, particularly from late morning to afternoon, while nighttime NO_x still contributes to O_3 loss through titration. Therefore, NOx controls are likely to be more effective for reducing daytime O_3 in NOx-sensitive regions, but their timing should be carefully considered to avoid unintended nighttime O_3 increases. These results indicate that O_3 mitigation policies should be regime-specific and time-dependent, rather than relying on uniform daily reductions of NO_x or VOC emissions.

<Section 3.4 in the revised manuscript>

These results also indicate that emission control strategies should not be based solely on daily mean precursor responses. In VOC-sensitive regions, nighttime NO_x emissions decrease O_3 through NO titration; therefore, stricter nighttime NO_x controls may not immediately reduce O_3 and could increase nighttime O_3 by weakening titration. In contrast, VOC reductions are more effective for lowering daytime O_3 under VOC-sensitive conditions, whereas NO_x reductions provide a more direct pathway for reducing daytime O_3 under NO_x-sensitive conditions.

(3) Importance of posterior emissions for policy-relevant sensitivity diagnosis

The comparison between the Prior and posterior (Hybrid_NOx+VOC) simulations also demonstrates how policy recommendations could be misdirected if they were based only on the uncorrected Prior emissions. As shown in Fig. S6, the Prior

experiment classified most of South Korea as VOC-sensitive. If this Prior-based regime diagnosis were used for policy guidance, VOC emission reductions would have been broadly prioritized as the main strategy for reducing O_3. However, the posterior results identified substantially more NO_x-sensitive regions, where NO_x reductions are expected to be more effective than VOC reductions for lowering daytime O_3. Therefore, relying on the Prior simulation alone could lead not only to biased estimates of precursor impacts but also to an incorrect selection of the target precursor for emission control. In other words, areas that should be managed primarily through NO_x reductions could have been incorrectly treated as VOC-control regions. In addition, because the posterior NO_x emissions showed a substantial nighttime reduction relative to the Prior emissions, the inferred timing of effective emission controls could also differ between the Prior and posterior simulations. This demonstrates that posterior, observation-constrained emissions are essential for deriving spatially targeted, temporally resolved, and chemically appropriate O_3 mitigation strategies.

<Section 3.4 in the revised manuscript>

The comparison between the Prior and posterior (Hybrid_NOx+VOC) simulations further demonstrates the value of using observation-constrained emissions to refine policy-relevant O_3 sensitivity diagnostics. As shown in Fig. S6, the Prior experiment classified most of South Korea as VOC-sensitive, suggesting that VOC reductions would be broadly prioritized based on the Prior emissions. In contrast, the posterior results identified substantially more NO_x-sensitive regions, where NOx reductions are expected to be more effective than VOC reductions for lowering daytime O_3. Thus, the posterior simulation complements the Prior inventory by providing a more spatially differentiated diagnosis of target precursors for emission control. Moreover, because the posterior NO_x emissions showed a substantial nighttime reduction relative to the Prior emissions (Fig. 6), the inferred timing of effective controls could also differ between the Prior and posterior simulations. These differences demonstrate that posterior, observation-constrained emissions provide a more reliable basis for deriving spatially targeted, time-specific, and chemically appropriate O_3 mitigation strategies.

<Section 4 in the revised manuscript>

The contrast between the Prior and Hybrid_NOx+VOC experiments further highlights the value of observation-constrained posterior emissions in refining policy-relevant O_3 sensitivity diagnostics. The Prior experiment provided a broad initial indication of VOC-sensitive

conditions over South Korea, while the posterior emissions revealed more spatially differentiated sensitivity regimes, including more extensive NO_x-sensitive regions. This suggests that the hybrid inversion can complement prior emission inventories by improving the representation of both the target precursor and the timing of effective emission controls. Therefore, observation-constrained posterior emissions provide an important basis for developing spatially targeted, time-specific, and chemically appropriate O_3 mitigation strategies.

RC2: Specific comments and responses

Comment 1: L31: Is there anything else that the adjoint sensitivities show, in terms of spatial or hourly characteristics? That VOC controls are most effective under VOC sensitive conditions is rather obvious / circular, as this is explained on lines 48-50.

Response: **We agree that the statement regarding VOC controls under VOC-sensitive conditions is a well-established concept. To address this, we have revised the Abstract to remove the circular statement and instead highlight the novel hour-dependent characteristics revealed by our adjoint analysis: specifically, the 1–2 hour temporal lag for VOC-driven O_3 production compared to immediate NO_x titration, and the nighttime O_3 depletion caused by biogenic VOCs.**

<Abstract in the revised manuscript>

Adjoint-based hourly ΔO_3 responses reveal distinct temporal characteristics: O_3 titration by NO_x is immediate, whereas photochemical O_3 production by VOCs requires a 1–2 hour reaction time. Furthermore, reactive biogenic VOCs act as a notable nighttime O_3 sink under NO_x-sensitive conditions.

Comment 2: 61: Applied, not proposed, as you're citing applications here rather than the first papers to propose new methods (with the exception of Cooper 2017, though if you're referring to FDMB not IFDMB, I believe the correct citation would be Lamsal 2011?).

Response: We have changed the word "proposed" to "applied" to accurately reflect the cited literature. Furthermore, we have included the citation for Lamsal et al. (2011) as the foundational reference for the FDMB method.

<Section 1 in the revised manuscript>

Various inverse modeling techniques have been applied, including mass balance (Lamsal et al., 2011; Cooper et al., 2017; Li et al., 2019; Qu et al., 2019; Momeni et al., 2024) …

Comment 3: 72: To my knowledge the first study to combine 4D-Var and mass balance in a hybrid fashion was: Qu, Z., D. K. Henze, N. Theys, J. Wang, W. Wang, Hybrid mass balance / 4D-Var joint inversion of NO_x and SO_2 emissions in East Asia, J. Geophys. Res., 124, 8203−8224, https://doi.org/10.1029/2018JD030240.

Response: To properly acknowledge their pioneering work in combining the mass balance and 4D-Var methods, we have added the citations for both Qu et al. (2017) and Qu et al. (2019).

<Section 1 in the revised manuscript>

To leverage the strengths of both approaches, a hybrid inverse modeling framework combining the mass balance and the 4D-Var has been recently applied (Qu et al., 2017, 2019; Chen et al., 2021; Choi et al., 2022; Moon et al., 2024).

<References in the revised manuscript>

Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X., Wang, J., and Keller, M.: Monthly top-down NO_x emissions for China (2005–2012): A hybrid inversion method and trend analysis, J. Geophys. Res. Atmos., 122, 4600–4625, https://doi.org/10.1002/2016JD025852, 2017.

Comment 4: 73: The word “Nevertheless” doesn’t really fit here, as the preceding text wasn’t related to the topic of spatial vs temporal emissions adjustments. Also, I’m not sure I believe the “most studies” here — did the authors really do an exhaustive review of all top-down emissions studies? I can think of large numbers that constrain hourly, daily, weekly, or monthly emissions, in applications spanning from NO_x to CO_2. So perhaps overall this sentence / topic just needs to be revised.

Response: The transition word "Nevertheless" and the phrase "most studies" have been removed to avoid overgeneralization. The paragraph has been rewritten to focus strictly on the scientific necessity of optimizing hourly variations and precursor interactions specifically for short-lived, chemically reactive species like O_3.

<Section 1 in the revised manuscript>

Accurately reproducing O_3 concentrations remains a critical challenge due to its short-lived and chemically reactive nature. Because diurnal variations in precursor emissions strongly influence O_3 formation through nonlinear photochemical processes (Wang et al., 2018), it is essential to simultaneously capture the hourly evolution of emissions and constrain the relative contributions of both NO_x and VOCs. Therefore, a comprehensive inverse modeling framework capable of simultaneously optimizing the spatiotemporal distribution of both precursor emissions is required.

Comment 5: 190: I don't see how there's anything intrinsic to FDMB that prevents it from being applied separately to emissions at different times. Emissions in different grid cells in space are already treated separately; it would be trivial to extend to emissions at different times.

Response: We agree that the FDMB method can theoretically be applied to separate time windows. However, because polar-orbiting satellites like TROPOMI provide only a single daily overpass (around 13:30 local time), the temporal resolution is insufficient to independently resolve hourly emissions using FDMB alone. Furthermore, following previous mass balance studies (Lamsal et al., 2011; Chen et al., 2021; Mun et al., 2023), we utilized column data averaged over the study period to minimize transport-induced smearing errors. Therefore, we delegated the hourly diurnal optimization to the subsequent 4D-Var step.

<Section 2.3.1 in the revised manuscript>

While the FDMB method can theoretically be extended to separate time windows, the once-daily overpass of TROPOMI inherently limits its capacity to independently resolve diurnal temporal variability. Furthermore, to minimize transport-induced smearing errors, we utilized column data averaged over the study period following previous methodologies (Lamsal et al., 2011; Chen et al., 2021; Mun et al., 2023). Therefore, the FDMB step is used strictly to establish a robust spatial baseline, while hourly variations are optimized in the subsequent 4D-Var inversion.

Comment 6: 197: Given Eq 8, B is the error covariance of the emission scaling factor, not the emissions.

Response: We have revised the text to clarify that B represents the error covariance matrix of the emission scaling factors rather than the absolute emissions.

<Section 2.3.2 in the revised manuscript>

B_i and R_i are the error covariance matrices for emission scaling factors and observations, respectively.

Comment 7: 198: How is the 100% emission error translated into a value within B, which is the error covariance for the emissions scaling factor? Is a doubling of emissions (alpha = 2) penalized equivalently to a zeroing out of the emissions (alpha = 0) or halving (alpha = 0.5)?

Response: A 100% prior emission error corresponds to assuming a standard deviation of 1.0 for the prior emission scaling factors ($\alpha_b = 1$), which sets the diagonal elements of B to 1.0. Under this Gaussian error assumption, the background penalty is evaluated symmetrically as $(\alpha_i = 1)^2$. Therefore, a doubling ($\alpha_i = 2$) and a complete zeroing ($\alpha_i = 0$) of emissions both result in an absolute deviation of 1.0 and incur the exact same penalty in the cost function. Conversely, halving the emissions ($\alpha_i = 0.5$) yields a smaller deviation of 0.5 and incurs a lower penalty. We have expanded Section 2.3.2 to make this explicit.

<Section 2.3.2 in the revised manuscript>

The prior uncertainty for the emissions was set to 100 %, which corresponds to assuming a standard deviation of 1.0 for the prior emission scaling factors ($\alpha_b = 1$). Consequently, the corresponding variance is 1.0, and the diagonal elements of B_i were set to 1.0. Under this Gaussian error assumption, the background penalty term is symmetrically evaluated; for instance, a doubling of emissions ($\alpha_i = 2$) and a complete zeroing of emissions ($\alpha_i = 0$) incur the same mathematical penalty in the cost function.

Comment 8: 200: Please provide the values of the measurement errors and representativeness errors, and how the latter were obtained. I don't see either of these quantified here nor in section 2.2.1.

Response: We have provided the specific values and parameterizations used to construct the observation error covariance matrix. Following the methodologies of Elbern et al. (2007) and Feng et al., (2018), the measurement and representativeness errors are quantitatively defined based on base error limits, observed concentrations, model grid spacing, and the radius of influence of the observation stations. We have updated Section 2.3.2 to explicitly include these detailed equations and parameter values.

<Section 2.3.2 in the revised manuscript>

For ground-based AQMS observations, the total observational error was estimated as the sum of measurement errors (ϵ_o) and representativeness errors (ϵ_r), such that $\epsilon_{tot} = \sqrt{\epsilon_o^2 + \epsilon_r^2}$ (Elbern et al., 2007; Feng et al., 2018). Specifically, ϵ_o is defined as $\epsilon_{max} + 0.0075 \times y_{obs}$, where ϵ_{max} is a base error limit set to 1.5 $\mu g\ m^{-3}$ for both NO_2 and O_3, and y_{obs} is the observed concentration. The representativeness error (ϵ_r) is parameterized as $\epsilon_r = \kappa\sqrt{\Delta x/L}$, where κ is a tunable scaling factor (0.5), Δx is the grid spacing (9 km), and L represents the radius of influence of an observation (set to 3 km).

Comment 9: 203: Please show the L curve and provide the optimal value of gamma that was found. Ok.. I see this now in the supplemental and later in section 3.1. Well, in Fig S2, I don't believe that J_B or J_O have been defined (I know what they are, in terms of portions of Eq 8, but others may not).

Response: We have revised the caption of the L-curve figure in the Supplementary Information (formerly Fig. S2, now Fig. S3) to explicitly define J_B as the background penalty term and J_O as the observation penalty term corresponding to Eq. (8).

<The caption of Fig. S2(now shown as Fig. S3) in the revised manuscript>

Figure S3: Results of the L-curve test and emission NME [%] used to determine the optimal regularization parameter (γ) for the 4D-Var inverse modeling. J_B represents the background penalty term (the first term on the right-hand side of Eq. (8)) and J_O represents the observation penalty term (the second term on the right-hand side of Eq. (8)). The test was performed for γ values of 0.001, 0.01, 0.1, 1, 10, 100, and 1000, based on the simulation for 1 May 2022. The optimal value was determined to be $\gamma = 10$.

Comment 10: 210: I don't disagree that most studies have focused on a single species; however, there are many that have conducted joint inversions of NO_x along with other species, such as VOCs, CO, NH_3, SO_2…. Some attention to the literature in this regards is warranted.

Response: We thank the reviewer for pointing this out and apologize for the ambiguity in our wording. The phrase "previous studies" was not intended to generalize the entire literature, but rather to refer specifically to our own previous framework (Moon et al., 2024), which focused on single-species (NO_2) corrections. We fully agree that many excellent studies in the field have successfully conducted joint inversions of multiple species. To clarify our original intent and avoid any overgeneralization, we have revised the sentence accordingly.

<Section 2.3.3 in the revised manuscript>

While the original framework primarily focused on single-species corrections such as NO_2, we extended the approach to jointly optimize both NO_x and VOC emissions to better represent the nonlinear photochemical processes governing O_3 formation.

Comment 11: Fig 3: Within 4D-Var, the flowchart doesn't make too much sense to me. I would think that it should be CMAQ fwd —> CMAQ adjoint —> 4D-Var solver (gradient based minimizer) —> an arrow back to CMAQ forward (because 4D-Var requires iteration) and another arrow to 4D-Var posterior emissions. And perhaps another arrow from here to Hybrid Emissions? Because the Hybrid emissions are themselves the emissions that result from the 4D-Var solution, yes? Also, later on e.g. line 244 the authors refer to "posterior emissions" so it would be good, as I've suggested, to specifically identify what these are in the flowchart.

Response: We have redesigned the flowchart in Fig. 3 to clearly represent the iterative nature of the 4D-Var optimization. The revised diagram explicitly shows the closed-loop sequence of the CMAQ forward model, CMAQ adjoint model, and 4D-Var solver, which represents the gradient-based minimization procedure. We have also added explicit labels and arrows indicating that the 4D-Var posterior emissions lead to the final hybrid emissions, consistent with the terminology used in the manuscript.

<Fig.3 in the revised manuscript>

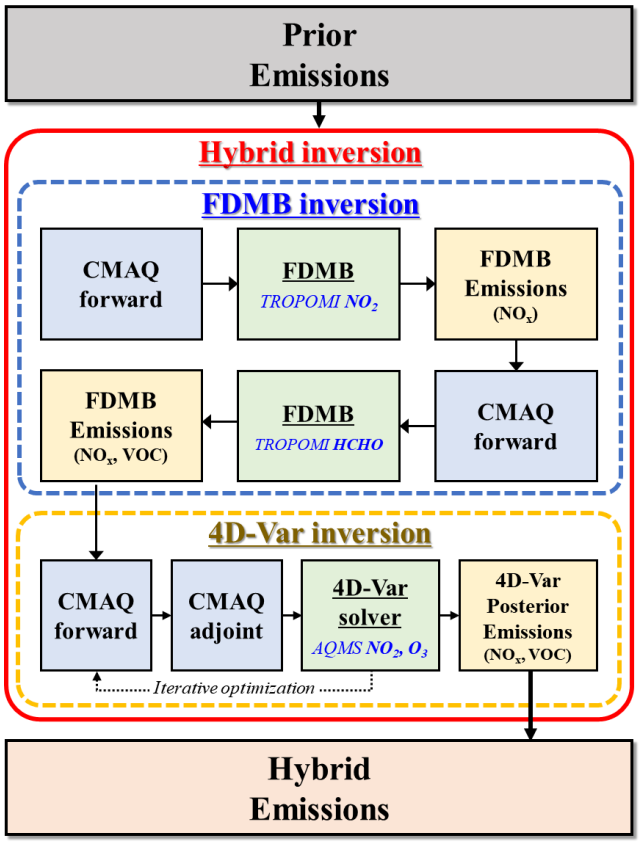

Figure 1: Schematic of the hybrid inverse modeling framework combining FDMB and 4D-Var inversions. In the first step, the FDMB inversion (blue box) corrects the spatial distribution of NO_x and VOC emissions using TROPOMI NO_2 and HCHO VCDs. In the second step, the 4D-Var inversion (Yellow box) constrains the spatiotemporal distribution of NO_x and VOC emissions by assimilating hourly ground-based NO_2 and O_3 observations.

Comment 12: 249: It's a bit confusing as here it seems that the time difference between t1 and t2 is an hour, where in the previous section it was 24 hrs.

Response: We agree that the subscripts "1" and "2" in t_1 and t_2 were misleading, as they inadvertently implied a strict 1-hour consecutive interval (e.g., hour 1 to hour 2), which conflicted with the 24-hour total assimilation window described previously. To avoid this confusion, we have updated the notation to h_e (emission hour) and h_r (receptor hour) in the revised manuscript (Section 2.5).

Comment 13: Fig 5: perhaps show the 2nd and 3rd columns as differences from the prior? or from each other? It's hard to see the NO_x emission changes at all.

Response: Following your suggestion, we have redesigned both Figure 5 and Figure S4. The figures have been updated so that the final column explicitly displays the spatial differences from the prior emissions (i.e., Posterior minus Prior). This modification clearly highlights the specific spatial adjustments made to the precursor emissions across the domain.

<Fig. 5 and Fig.S4 in the revised manuscript>

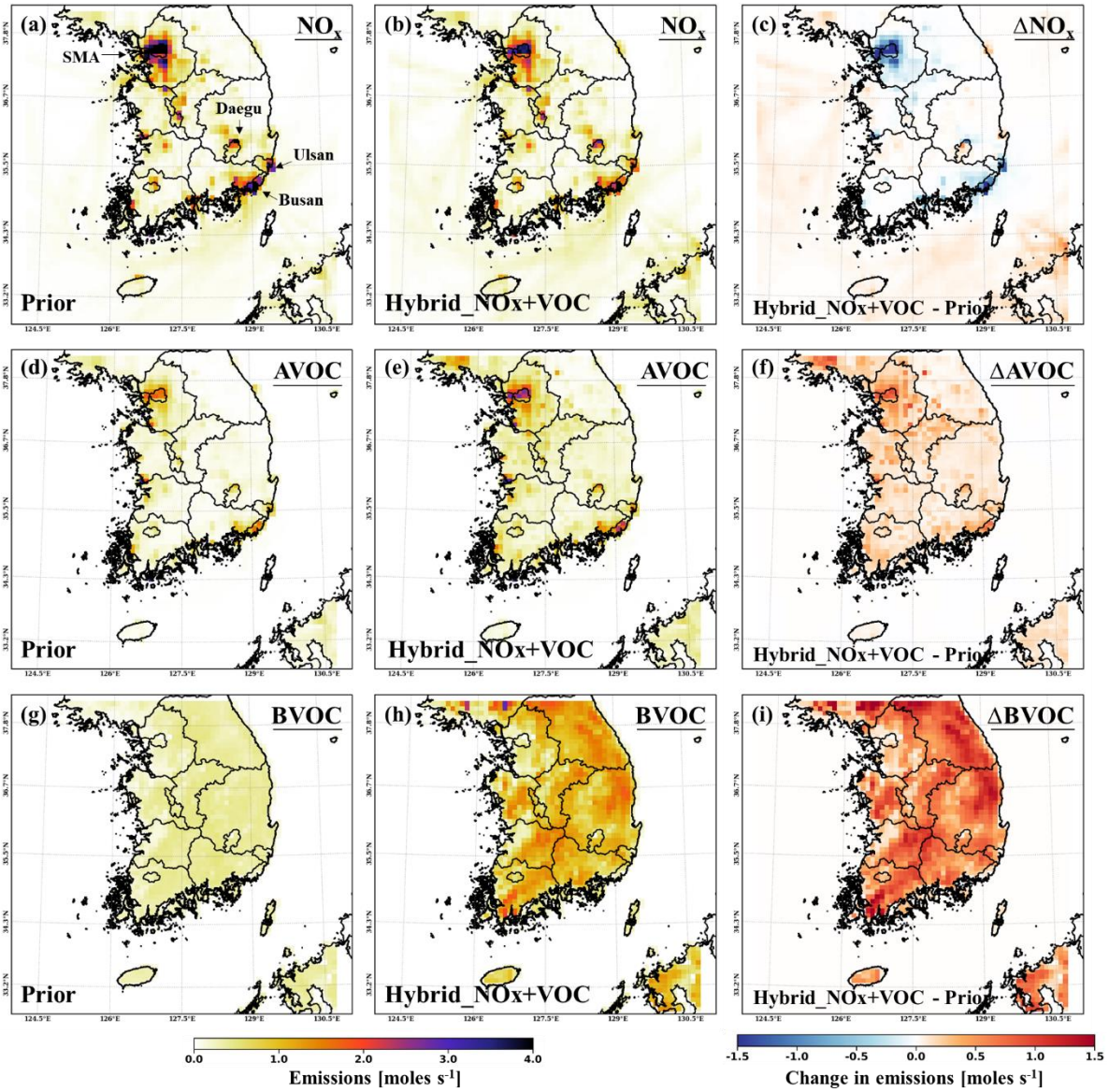

Figure 2: Spatial distributions of (a, d, g) the Prior emissions, (b, e, h) the Hybrid_NOx+VOC emissions, and (c, f, i) the corresponding analysis increments (Hybrid_NOx+VOC – Prior) for NOx, AVOC, and BVOC, respectively, averaged over the study period.

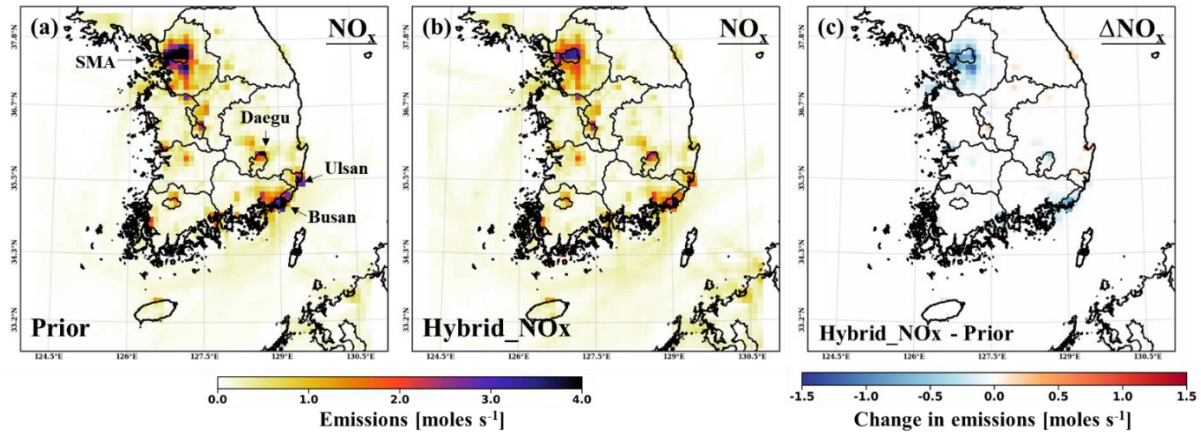

Figure S4: Spatial distributions of (a) the Prior emissions, (b) the Hybrid_NOx emissions, and (c) the corresponding analysis increments (Hybrid_NOx – Prior) for NO_x, averaged over the study period.

References

Atkinson, R. and Arey, J.: Gas-phase tropospheric chemistry of biogenic volatile organic compounds: a review, Atmos. Environ., 37(2), 197–219, https://doi.org/10.1016/S1352-2310(03)00391-1, 2003.

Choi, J., Henze, D. K., Wells, K. C., and Millet, D. B.: Joint inversion of satellite-based isoprene and formaldehyde observations to constrain emissions of nonmethane volatile organic compounds, J. Geophys. Res. Atmos., 130, e2024JD042070, https://doi.org/10.1029/2024JD042070, 2025.

East, J. D., Henderson, B. H., Napelenok, S. L., Koplitz, S. N., Sarwar, G., Gilliam, R., Lenzen, A., Tong, D. Q., Pierce, R. B., and Garcia-Menendez, F.: Inferring and evaluating satellite-based constraints on NOx emissions estimates in air quality simulations, Atmos. Chem. Phys., 22, 15981–16001, https://doi.org/10.5194/acp-22-15981-2022, 2022.

Kurpius, M. R. and Goldstein, A. H.: Gas-phase chemistry dominates O_3 loss to a forest, implying a source of aerosols and hydroxyl radicals to the atmosphere, Geophys. Res. Lett., 30, 1371, https://doi.org/10.1029/2002GL016785, 2003.

Moon, J.: Inverse Modeling for Constraining Spatiotemporal Distributions of Emissions, PhD thesis, Pusan National University, Busan, South Korea, https://doi.org/10.23172/pusan.000000167548.21016.0000787, 2025.

Moon, J., Choi, Y., Jeon, W., Kim, H. C., Pouyaei, A., Jung, J., Pan, S., Kim, S., Kim, C.-H., Bak, J., Yoo, J.-W., Park, J., and Kim, D.: Hybrid IFDMB/4D-Var inverse modeling to constrain the spatiotemporal distribution of CO and NO_2 emissions using the CMAQ adjoint model, Atmos. Environ., 327, 120490, https://doi.org/10.1016/j.atmosenv.2024.120490, 2024.

Wells, K. C. and Millet, D. B.: ROCR isoprene retrievals from the CrIS satellite sensor, Data Repository for the University of Minnesota (DRUM), https://doi.org/10.13020/5n0j-wx73, 2022.

Wolfe, G. M., Kaiser, J., Hanisco, T. F., Keutsch, F. N., de Gouw, J. A., Gilman, J. B., Graus, M., Hatch, C. D., Holloway, J., Horowitz, L. W., Lee, B. H., Lerner, B. M., Lopez-Hilfiker, F., Mao, J., Marvin, M. R., Peischl, J., Pollack, I. B., Roberts, J. M., Ryerson, T. B., Thornton, J. A., Veres, P. R., and Warneke, C.: Formaldehyde production from isoprene oxidation across NOx regimes, Atmos. Chem. Phys., 16, 2597–2610, https://doi.org/10.5194/acp-16-2597-2016, 2016.