

Reviewer 1

General comment:

Overall, this manuscript is built on a highly valuable QAR dataset. The combination of nationwide coverage, second-level trajectories, and engine-level fuel-flow measurements has clear potential to substantially improve aviation emission inventories and to reveal operational features that are difficult to resolve using traditional approaches. However, the current manuscript does not yet fully realize this potential. Several critical steps between the QAR observations and the estimated emissions are either insufficiently documented or methodologically questionable, particularly the implementation of BFFM2, the treatment of humidity and PM, and the uncertainty analysis. In addition, key data-processing procedures, engine matching, and the construction of the final "4D" inventory are not described with sufficient transparency. As a result, some of the strongest conclusions regarding improved accuracy, reduced uncertainty, and newly resolved emission processes are not yet adequately supported. These methodological issues should be resolved before the inventory and its derived insights can be reliably interpreted.

Overall response:

We greatly appreciate the reviewer for all the thorough and constructive suggestions and comments on our manuscript. We agree that the QAR data provide an important opportunity to improve the temporal and operational resolution of aviation emission inventories. We also acknowledge the reviewer's central concern that the scientific value of such a dataset depends not only on the quality and resolution of the observations, but also on the transparency, methodological rigor, and reproducibility of the procedures used to derive emission estimates.

In response to these concerns, we have carefully reconsidered the methodological framework linking the QAR observations to the resulting emission inventory. In particular, we have focused on the implementation and documentation of BFFM2, the treatment of humidity and particulate matter (PM), the quantification of uncertainty, the preprocessing of QAR data, the matching of

aircraft and engine information, and the procedures used to construct the final four-dimensional (4D) emission inventory. We have addressed these issues through targeted revisions to the relevant methodological descriptions, analyses, and interpretations. Where necessary, we have also reconsidered the scope of the conclusions to ensure that claims regarding improved accuracy, reduced uncertainty, and the identification of previously unresolved emission processes are commensurate with the evidence provided. Our detailed responses below address each of the reviewer's specific comments and explain the corresponding revisions to the manuscript.

Comment 1:

The implementation of BFFM2 requires careful verification. Based on the equations currently presented, measured in-flight fuel flow appears to be mapped directly to the EEDB-derived reference EI relationship, followed by atmospheric correction of the EI. The intermediate conversion from actual in-flight fuel flow to equivalent reference fuel flow is not shown, and Mach-number correction is also absent. If this conversion was not performed, the fuel-flow input used to obtain the reference EI would not be consistent with the standard test-bench conditions underlying the EEDB, particularly during cruise. The authors should explicitly present the complete BFFM2 calculation chain and confirm whether the reference fuel-flow correction was applied. If it was omitted, the resulting NO_x, CO, and HC estimates would need to be reassessed.

Response 1:

We sincerely appreciate the reviewer identifying this critical gap in our original BFFM2 implementation. We fully agree that omitting the equivalent reference fuel flow conversion and Mach-number correction breaks the physical alignment with EEDB test-bench conditions. To resolve this, we fundamentally rebuilt the emission calculation chain by introducing the intermediate equivalent reference fuel flow (W_{ff}) transformation, directly incorporating QAR-measured static air temperature, pressure altitude, and instantaneous Mach number. Using this corrected pipeline, we completely recalculated the 2023 national inventory. This physical correction substantially altered our NO_x and PM emission totals and eliminated artificial overestimations during high-load operations. The detailed step-by-step calculations and revised models are now explicitly provided in the manuscript.

In Section 2.1, the research framework was revised to explicitly include the equivalent reference fuel flow conversion step in the methodological chain, reading as follows:

"In the second stage, these datasets are integrated into a high-fidelity emission modeling framework. Aircraft are matched to specific engine types to identify the corresponding EEDB-based emission-index (EI) curves, while QAR-derived fuel-flow measurements are converted to reference fuel flow and combined with environmental factors."

In Section 2.3.2, the core calculation was reconstructed to clearly present the complete BFFM2 formulation, pressure derivation, and the introduction of the actual flight Mach number, reading as follows:

"To maintain consistency between the QAR observations and the EEDB-based reference curves, the instantaneous fuel flow was first converted to an equivalent reference-condition fuel flow, denoted as W_{ff} , following the fuel-flow normalization procedure of the Boeing Fuel Flow Method 2 (BFFM2).

The ambient static temperature was directly obtained from the Static Air Temperature (SAT) recorded by QAR. For ambient pressure, the QAR-recorded Pressure Altitude was used to derive the corresponding atmospheric pressure based on the standard barometric relationship. Specifically, the ambient pressure (P_{amb}) was calculated as:

$$P_{amb} = P_b \times \left(1 - \frac{L \cdot (h - h_b)}{T_b}\right)^{\frac{g_0 \cdot M}{R \cdot L}} \quad (1)$$

where P_b and T_b are the reference baseline pressure and temperature, L is the temperature lapse rate of the troposphere, h is the pressure altitude recorded by QAR, h_b is the baseline height, g_0 is the gravitational acceleration, M is the molar mass of Earth's atmosphere, and R is the universal gas constant. The pressure and temperature ratios required by BFFM2 (δ_{amb} and θ_{amb}) were then calculated by dividing the instantaneous ambient pressure (P_{amb}) and temperature (T_{amb} , which is the QAR-recorded SAT) by their respective standard sea-level reference values.

The measured single-engine fuel flow (W_f) was then converted to the corresponding reference-condition fuel flow according to the BFFM2 formulation:

$$W_{ff} = \frac{W_f \cdot \theta_{amb}^{3.8}}{\delta_{amb}} e^{0.2M^2} \quad (2)$$

where M is the instantaneous flight Mach number. This formulation accounts for the effects of ambient pressure, temperature, and flight speed when transforming the fuel flow from flight conditions to the reference condition. The resulting W_{ff} represents the fuel flow corresponding to the observed operating point after normalization to the reference conditions."

Comment 2:

The humidity treatment is not physically clear from the current description. Temperature and pressure alone are insufficient to determine actual specific humidity. The manuscript should identify the actual humidity information or assumption used.

Response 2:

We agree that relying on temperature and pressure alone without incorporating actual specific humidity compromises the accuracy of the NO_x emission estimates. To resolve this, we eliminated the previous static humidity assumption and integrated the ERA5 meteorological reanalysis dataset to provide high-resolution, trajectory-matched specific humidity. The NO_x atmospheric correction equations in the BFFM2 framework were updated to explicitly incorporate the ERA5-derived specific humidity, and the entire NO_x inventory was recalculated accordingly.

In Section 2.2.4, a new subsection was added to detail the integration of ERA5 meteorological data for determining specific humidity, reading as follows:

"Aircraft engine NO_x emissions are sensitive to ambient atmospheric humidity. In the Boeing Fuel Flow Method 2 (BFFM2) framework, the atmospheric correction of NO_x emission indices requires specific humidity in addition to ambient temperature and pressure. Since the QAR data provide in-flight measurements of static pressure and static air temperature but not specific humidity, the ERA5 reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) was incorporated into the emission calculation (Hersbach et al., 2020). For the spatial and temporal domain of this study, ERA5 meteorological data were extracted for the bounding rectangle covering mainland China throughout 2023. Monthly vertical profiles of specific

humidity (ω) were obtained for 12 months, with 10 selected pressure levels for each month. The resulting dataset describes the seasonal and vertical variations in atmospheric water vapor over the study region.

The extracted ERA5 specific-humidity data were subsequently matched with the QAR trajectories at the individual observation-point level. For each QAR observation, the flight month was identified from its timestamp, while the QAR-recorded standard pressure altitude was used to determine the corresponding pressure level. The horizontal position of each QAR observation was also associated with the corresponding ERA5 grid location. Based on the month, horizontal grid location, and pressure level, a specific humidity value (ω) was assigned to each QAR observation."

In Section 2.3.3, the NO_x atmospheric correction equation was revised to explicitly demonstrate the incorporation of the matched specific humidity parameter, reading as follows:

"The NO_x emission index was corrected for both temperature and pressure effects and ambient humidity:

$$H = -19(\omega - 0.00634) \quad (5)$$

$$\text{EI}_{\text{NO}_x, \text{flight}} = \text{EI}_{\text{NO}_x, \text{ref}} \left(\frac{\delta_{\text{amb}}^{1.02}}{\theta_{\text{amb}}^{3.3}} \right)^{0.5} e^H \quad (6)$$

where ω is the ambient specific humidity."

Comment 3:

The full-flight PM methodology is incomplete. The atmospheric correction equations are provided for NO_x , CO, and HC, but no corresponding procedure is given for converting EEDB PM values from LTO reference conditions to actual full-flight conditions. This is particularly important because the manuscript reports a national full-flight PM inventory and interprets phase-resolved PM behavior in terms of temperature and pressure effects. The authors should provide the full PM calculation framework, clarify its applicability outside the LTO regime, and justify the physical interpretation of the resulting cruise and climb PM emissions.

Response 3:

We thank the reviewer for highlighting the need to formalize the full-flight particulate matter (PM) calculation framework. We agree that extending LTO-based EEDB PM values to full-flight conditions requires explicit atmospheric and phase-specific corrections to ensure physical validity during cruise and climb. To address this, we comprehensively revised the PM methodology to explicitly distinguish and independently quantify its three components: volatile organic PM, volatile sulfate PM, and non-volatile PM (nvPM). Volatile components are now strictly derived using the First Order Approximation (FOA) methodology, while nvPM is dynamically corrected to flight conditions using the MEEM2 altitude correction framework. This comprehensive breakdown and the associated formulations have been transparently integrated into the revised manuscript, ensuring the physical interpretability of PM emissions across all flight phases.

In Section 2.3.3, the PM methodology was substantially expanded to explicitly define the resolution of total PM into its three components and present the corresponding FOA and MEEM2 formulations, reading as follows:

"For particulate matter, the total particulate emission index was resolved into non-volatile particulate matter (nvPM), volatile organic particulate matter, and volatile sulfate particulate matter. The nvPM component was determined from the engine-specific EEDB-based parameterization using the equivalent reference fuel flow (W_{fr}), whereas the volatile components were estimated following the First Order Approximation (FOA) framework (ICAO, 2011). The total particulate emission index was therefore expressed as Eq. (7):

$$EI_{PM} = EI_{nvPM} + EI_{vPM, organic} + EI_{vPM, sulfate} \quad (7)$$

The volatile organic particulate component was estimated from the flight-condition HC emission index using the phase-specific empirical relationship adopted in the FOA methodology:

$$EI_{vPM, organic} = \delta_m EI_{HC, flight} \quad (8)$$

where EI_{HC} is the HC emission index under actual flight conditions (in g/kg fuel) and δ_m is the phase-specific empirical ratio (in mg/g). The values of δ_m are 115, 76, 56.3, and 6.2 for take-off, climb-out, approach, and taxi, respectively. These phase-specific ratios relate the volatile organic particulate emission index to the HC emission index for the corresponding operating mode.

The volatile sulfate particulate emission index was estimated from the fuel sulfur content and the sulfur conversion fraction following the FOA-based formulation:

$$EI_{\text{vPM,sulfate}} = 10^6 \left(\text{FSC} \cdot \varepsilon \frac{MW_{\text{SO}_4}}{MW_{\text{S}}} \right) \quad (9)$$

where FSC is the fuel sulfur content, prescribed as 500 ppm by mass, and ε is the fraction of fuel sulfur converted from S(IV) to S(VI), which was set to 0.024 following the value adopted in the FOA methodology. MW_{SO_4} and MW_{S} are the molecular weights of sulfate and sulfur, respectively, with values of 96 and 32.

For nvPM, the dynamically determined reference emission index ($EI_{\text{nvPM,ref}}$) was obtained from the engine-specific EEDB-based parameterization using the equivalent reference fuel flow calculated for each QAR observation. The reference nvPM emission index was then corrected to flight conditions using the Aircraft nvPM Mission Emissions Estimation Methodology, Version 2 (MEEM2) of Ahrens et al, (2025):

$$EI_{\text{nvPM,flight}} = EI_{\text{nvPM,ref}} \frac{\delta_{\text{amb}}^{1.377}}{\theta_{\text{amb}}^{4.455}} \times 1.1^{2.5} \quad (10)$$

where $EI_{\text{nvPM,ref}}$ is the engine-specific reference nvPM emission index and δ_{amb} and θ_{amb} are the ambient pressure and temperature ratios, respectively. The MEEM2 correction accounts for the effects of ambient conditions on nvPM emissions through fuel-flow- and altitude-dependent relationships consistent with the BFFM2-based treatment of flight conditions."

Comment 4:

The Monte Carlo analysis should be reframed as uncertainty propagation under prescribed input distributions rather than as an assessment of model accuracy. The resulting CO₂ CVs of approximately 2%, 15%, and 30% essentially reproduce the fuel-flow uncertainties assigned a priori to QAR, ADS-B, and SFD. Likewise, the SFD systematic underestimation is also partly imposed by setting the trajectory-distance factor to a mean of 0.93. These results demonstrate how the assumed error budget propagates through each framework, but they do not independently establish which inventory is more accurate.

Response 4:

We sincerely thank the reviewer for pointing out that the Monte Carlo analysis should represent uncertainty propagation rather than an independent assessment of model accuracy. We fully agree with this perspective and have fundamentally restructured our uncertainty analysis. First, we explicitly defined the prescribed input distributions for fuel flow, environmental, emission-index, and operational uncertainties across the QAR, ADS-B, and SFD frameworks, clarifying that the resulting coefficients of variation (CVs) demonstrate how these assumed errors propagate through the modeling chain. Second, to objectively quantify the relative contribution of different uncertainty sources without subjective residual variance assignment, we introduced a formal Shapley-based variance decomposition. Finally, the interpretations in the Results section were completely revised to reflect this perspective, avoiding unsupported claims regarding which inventory is strictly more accurate.

In Section 2.4, the Monte Carlo analysis was reframed to explicitly focus on uncertainty propagation under prescribed input distributions, reading as follows:

"To characterize the propagation of prescribed input uncertainties through the aircraft emission calculation chain described in Section 2.3, Monte Carlo simulations were conducted for the QAR-, ADS-B-, and SFD-based inventory frameworks. The prescribed uncertainties were grouped into four major sources: fuel-flow uncertainty (FF), environmental uncertainty (Env), emission-index uncertainty (EI), and operational uncertainty (Ops). For strictly positive multiplicative variables, such as fuel-flow bias, distance and time factors, and emission indices, mean-normalized log-normal distributions were used to avoid non-physical negative values. Ambient temperature was perturbed using an additive Gaussian distribution. For fuel-flow uncertainty, the bias was combined with temporally correlated multiplicative and additive perturbations. The temporal correlation of these perturbations was represented using first-order autoregressive AR(1) processes, allowing the simulated fuel-flow uncertainty to vary along the flight profile while retaining temporal dependence between consecutive observations. The corresponding uncertainty parameters for the QAR-, ADS-B-, and SFD-based frameworks are summarized in Table 3.

Table 3: Configuration of uncertainty input parameters

Parameter	Distribution / Perturbation Mode	QAR-based	ADS-B-based	SFD-based
Fuel-flow bias	Mean-normalized log-normal	1.5%	10%	20%
Temporal fuel-flow noise	Correlated multiplicative AR(1)	0.5%, $\rho=0.85$	6%, $\rho=0.65$	10, $\rho=0.85$
Additive fuel-flow residual	Correlated additive AR(1)	0.2%, $\rho=0.80$	1.0%, $\rho=0.55$	1.5%, $\rho=0.80$
Ambient pressure	Mean-normalized log-normal	0.5%	1.5%	3%
Ambient temperature	Additive Gaussian	0.5 K	2 K	5 K
Distance factor	Mean-normalized log-normal	—	—	8%
Airborne-time factor	Mean-normalized log-normal	—	—	8%
Taxi-time factor	Mean-normalized log-normal	—	—	30%
EI _{NOx}	Mean-normalized log-normal	8%	8%	8%
EI _{CO}	Mean-normalized log-normal	12%	12%	12%
EI _{HC}	Mean-normalized log-normal	12%	12%	12%
EI _{nvPM}	Mean-normalized log-normal	18%	18%	18%
Fuel sulfur content (FSC)	Truncated Gaussian	500±100ppm	500±100ppm	500±100ppm

The prescribed uncertainty structures, as summarized in Table 3, strictly reflect the varying

information fidelity across the three frameworks. In the QAR framework, uncertainties stem exclusively from the inherent instrument precision of the onboard systems. Conversely, the ADS-B and SFD frameworks lack direct thermodynamic observations, necessitating progressively larger perturbations to account for the reconstruction of fuel flow from kinematic proxies and the reliance on meteorological reanalysis rather than in-situ measurements (Hersbach et al., 2020; ICAO, 2011). For the SFD framework specifically, the complete absence of actual flight trajectories dictates the widest uncertainty bounds. Consequently, ambient pressure and temperature were assigned maximum perturbations of 3% and 5 K (ICAO, 1993), while the reconstructed fuel-flow profile was subjected to a 20% bias alongside temporally correlated multiplicative (10%) and additive (1.5%) components. Furthermore, macroscopic operational variables—such as distance, airborne time, and taxi time—required independent perturbations to capture unobserved routing deviations and ground congestion (Simaiakis & Balakrishnan, 2016; Teoh et al., 2024).

Independent of the activity and environmental data sources, uncertainties inherent to the EEDB parameterizations and fuel sulfur content (FSC) were applied uniformly across all frameworks. Mean-normalized log-normal perturbations were assigned to the emission indices of NO_x , CO, HC, and nvPM to represent the uncertainty in extrapolating certification data to in-service operating conditions (Wasiuk et al., 2015). FSC was sampled from a truncated Gaussian distribution and propagated through the sulfur oxidation pathway to quantify the variance in SO_2 and sulfate particulate emissions (Kapadia et al., 2016).

For each framework, 10,000 Monte Carlo realizations were propagated through the complete emission calculation procedure described in Section 2.3. The output uncertainty was quantified using the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean of the simulated emission estimates.

To quantify the relative contribution of different uncertainty sources without assigning residual variance subjectively, a Shapley-based variance decomposition was further applied. Four uncertainty sources (FF, Env, EI, and Ops) were considered as players in the variance attribution. For each framework, all $2^4 = 16$ possible combinations of these uncertainty sources were evaluated using independent Monte Carlo simulations. Let $V(S)$ denote the variance of the simulated emission estimates when the subset S of uncertainty sources is activated. The Shapley

contribution of source k was calculated as:

$$\phi_k = \sum_{S \subseteq N \setminus \{k\}} \frac{|S|!(n-|S|-1)!}{n!} [V(S \cup \{k\}) - V(S)] \quad (12)$$

where N denotes the complete set of uncertainty sources and $n = 4$. This formulation accounts for the marginal contribution of each uncertainty source across all possible source combinations and therefore incorporates the effects of interactions among uncertainty sources. The resulting contributions were normalized by the total output variance to obtain the relative variance contribution of each source."

Building on this restructured framework, the revised Section 3.2 now explicitly details the results of this uncertainty propagation experiment. As our Shapley decomposition mathematically demonstrates, once high-fidelity QAR observations constrain the physical flight trajectories and onboard thermodynamic states, the kinematic and environmental variance for incomplete-combustion products (such as CO) is compressed to just 1.4%. Consequently, 98.6% of the residual uncertainty is inherently dominated by the empirical EEDB parameterizations. This finding directly answers the reviewer's concern by objectively quantifying how the rate-limiting step of inventory precision shifts from activity data to emission-index models, rather than subjectively claiming one inventory's absolute accuracy over another.

Comment 5:

The LTO/CCD altitude definition in Figure 6 needs immediate verification. The manuscript labels LTO as below 3000 m and CCD as above 3000 m, whereas the conventional ICAO LTO boundary is 3000 ft, approximately 915 m. If 3000 m was actually used in the calculations rather than being a labeling error, the reported LTO/CCD distance and emission fractions would need to be recalculated. This also affects the interpretation of the reported low-altitude emissions.

Response 5:

We have corrected the threshold to the standard ICAO definition of 3,000 ft (approximately 914 m) and completely recalculated the corresponding LTO and CCD distance and emission fractions.

The text and Figure 6 have been consistently revised to reflect this standardized boundary, ensuring our physical interpretations of low-altitude emissions are accurate and aligned with established conventions.

In Section 3.4.1, the text was revised to reflect the corrected 3,000 ft boundary and the recalculated emission and distance fractions, reading as follows:

"The vertical distribution of CO₂ emissions along the Chengdu–Beijing and Beijing–Chengdu routes exhibits pronounced phase-specific characteristics (Figure 6(c)). Using 3,000 ft (approximately 914 m) as the boundary between the LTO and CCD phases, the LTO phase accounts for only 1.9% of the total flight distance but contributes 7.1% of total CO₂ emissions. In contrast, the CCD phase above 914 m accounts for 98.1% of the flight distance and 92.9% of total CO₂ emissions. The disproportionately higher emission share of the LTO phase relative to its distance share reflects the elevated fuel burn rates per unit distance associated with takeoff, initial climb, approach, taxi, and other low-altitude operations. The vertical distribution also exhibits pronounced emission peaks at several cruise altitudes, particularly around 9,600–10,000 m, reflecting the concentration of flight operations within assigned cruise flight levels. Here, the resolved CO₂ profile serves as a direct indicator of fuel-burn intensity and vertical operational clustering. Because the intense low-altitude fuel burn during the LTO phase occurs near airports and populated terminal areas, the co-emitted criteria pollutants (specifically NO_x and PM) within this layer are of particular relevance as inputs for subsequent dispersion modeling and local air-quality assessments (Unal et al., 2005; Wang et al., 2023).

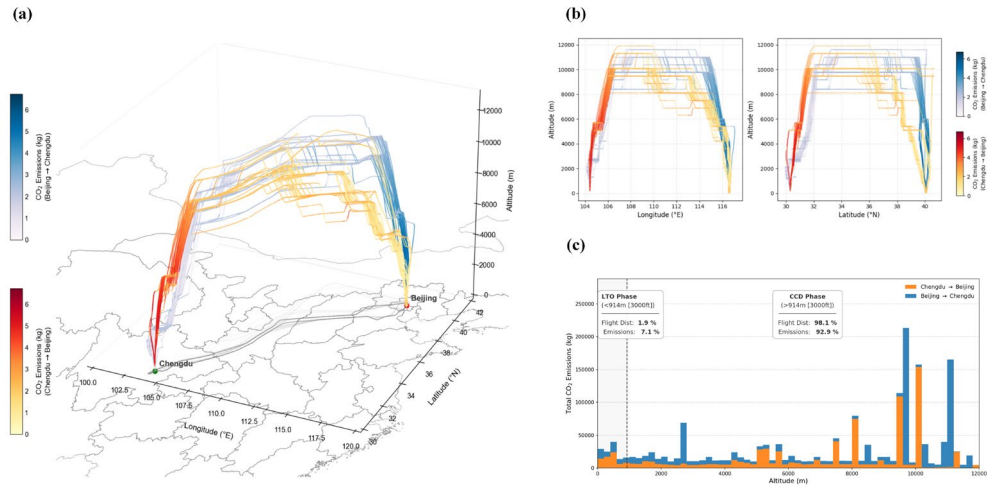


Figure 6: (a) Three-dimensional visualization of real-world flight trajectories between Beijing (ZBAA) and Chengdu (ZUTF), colored by instantaneous CO₂ emission rates. (b) Orthogonal projections of flight trajectories for the ZBAA–ZUTF route onto longitude-altitude and latitude-altitude planes, where blue and orange-red spectra denote outbound (to Chengdu) and inbound (to Beijing) flights, respectively. (c) Vertical profile of total accumulated CO₂ emissions for the ZBAA–ZUTF route, aggregated in 200 m altitude bins."

Comment 6:

The QAR preprocessing and reconstruction procedure needs substantially more documentation. The manuscript should report the fractions of flights and second-level observations that are retained as original measurements, reconstructed by short-gap interpolation, and reconstructed using the similarity-based method. For the similarity-based reconstruction, the matching criteria, reconstruction algorithm, validation error, and contribution of reconstructed data to the final inventory are necessary to evaluate how much of the inventory is genuinely observation-based.

Response 6:

We fully agree that explicitly reporting the proportions of retained versus reconstructed data, alongside robust error validation, is essential for evaluating the observation-based fidelity of the inventory. In the revised manuscript, we have substantially expanded the documentation in Section 2.2.1 and introduced Supplementary Tables S1–S3. We now clearly detail the stepwise reconstruction algorithm, which prioritizes short-gap local linear interpolation (<30 s) followed by

historical similarity-based matching (constrained by identical route and engine type) for longer gaps. Furthermore, we implemented a rigorous point-level mask-based validation, confirming that the vast majority of data required no or only minor interpolation, and that the maximum absolute deviation in total emissions introduced by our reconstruction procedure remains below 0.01% across all pollutants.

In Section 2.2.1, we condensed the detailed statistical reporting in the main text to maintain narrative flow while directing readers to the comprehensive breakdown in the supplementary tables. The revised text reads as follows:

"The data foundation of this study stems from a massive airborne QAR dataset provided by the Civil Aviation Flight Quality Monitoring Base Station of China. The initial archive contained approximately 4.85 million flight records for the entire year of 2023. As this study aims to construct a civil aviation emission inventory for strictly domestic flights within mainland China, flights with departure or arrival airports outside mainland China, as well as flights involving Hong Kong, Macao, and Taiwan, were excluded. After flight-scope screening and the removal of invalid records, approximately 4.52 million valid domestic flight records were retained, accounting for 93.2% of the original archive (details of the initial data screening are provided in Table S1). All retained flights contained complete FF records for the subsequent emission calculations.

Table S1. Summary of QAR Flight Screening and Data Retention

Data Processing Stage	Proportion of Raw Records
	(%)
Raw QAR records	100.0%
Flights beyond the study scope	6.2%
Domestic flight candidate records	93.8%
Invalid records	0.6%
Final valid domestic flights	93.2%

To ensure consistency among the 1-Hz QAR parameters, a common observation index was used

as the temporal reference, and all target variables were aligned on this native 1-Hz sequence. Missing observations arising from incomplete QAR records were identified at the observation level, while values violating parameter-specific physical constraints were treated as invalid and converted to missing observations. The screening included negative fuel-flow values, altitudes above 15,000 m, and SAT values outside the physically plausible atmospheric range, together with corresponding constraints for other target variables (Gariel et al., 2011). The observation-level missing rates were then calculated from the combined effects of originally missing and quality-control-screened observations, providing the basis for subsequent missing-data reconstruction.

Following the identification of missing observations, a stepwise reconstruction procedure was applied according to the duration and position of individual gaps. Internal gaps of no more than 30 s, bounded by valid observations on both sides, were recovered using local linear interpolation on the native 1-Hz observation sequence. Remaining gaps, including longer gaps and leading or trailing gaps without valid observations on both sides, were reconstructed using historical similar-flight matching. Historical records from the same route and engine type were preferentially searched, followed by same-route records when no qualified engine-specific match was available; the first-ranked matched flight was then used to reconstruct the missing observations. At the flight level, 68.4% of the retained flights required no reconstruction, 21.2% were fully resolved by short-gap interpolation, and 10.4% additionally required historical-flight matching after the interpolation step. The detailed 1-s observation-level reconstruction proportions for all critical parameters are comprehensively summarized in Table S2.

Table S2. 1-s Observation-Level Data Reconstruction Proportions

Variable	Original valid rate	After 30-s interpolation	After historical-flight matching
MACH	99.93%	100.00%	—
ALT_STD	87.28%	93.67%	100.00%
SAT	91.77%	98.75%	100.00%
FF1	99.36%	100.00%	—

FF2	99.43%	100.00%	—
GPS_LONG_CA	92.67%	98.27%	100.00%
GPS_LAT_CA	92.46%	98.31%	100.00%

The effects of the reconstruction procedures were evaluated using point-level mask-based validation. For each variable, known valid 1-s observations were masked according to the observed missing rate while preserving the gap structure of the corresponding missing observations. The masked observations were then reconstructed using the same interpolation and historical-flight matching procedures applied to the operational data. The reconstructed values were compared with the original observations at the masked locations using R2, RMSE, and bias. The reconstructed variables were subsequently propagated through the complete emission calculation framework, and the resulting total emissions were compared with those calculated from the original observations. The validation results and corresponding total emission deviations are presented in Table S3, with the absolute deviations in total emissions remaining below 0.01% for all evaluated pollutants."

Table S3. Point-Level Mask-Based Validation of Data Reconstruction and Its Impact on Total Emissions

Variable	Masking rate	Reconstruction method	R ²	RMSE	Bias	NO _x deviation (%)	CO deviation (%)	HC deviation (%)	PM deviation (%)
MAC	0.07%	30-s interpolation	0.9998	0.008	+0.001	-0.0001	+0.0000	+0.0001	+0.0001
ALT	6.39%	30-s interpolation	0.9988	0.66	-0.02	-0.001	+0.000	+0.005	+0.004
SAT	6.98%	30-s interpolation	0.9992	0.62	+0.01	-0.001	+0.000	+0.004	+0.003
FF1	0.64%	30-s interpolation	0.9980	1.18	-0.04	-0.003	+0.001	+0.004	+0.003
FF2	0.57%	30-s	0.997	1.26	-0.05	-0.003	+0.001	+0.005	+0.004

		interpolation	8							
		Historical-flt								
ALT			0.996							
	6.33%	g-ht		2.01	+0.13	+0.007	-0.002	+0.006	+0.005	
STD			8							
		matching								
		Historical-flt								
			0.994							
SAT		g-ht		1.96	+0.11	+0.002	-0.002	+0.005	+0.004	
	1.25%		8							
		matching								

Comment 7:

The QAR-schedule fusion and engine assignment are critical to the entire EI calculation but are currently treated as a black box. The external schedule database is not identified, and aircraft registration alone does not uniquely identify an individual flight. The manuscript should provide the database source and version, the complete linkage procedure, the origin of the ENGTYPE field, the mapping from ENGTYPE to EEDB engine identifiers, and diagnostics for matching quality. In particular, the fractions of flights assigned to an exact engine variant, an engine-family fallback, or an unresolved category should be reported, together with the sensitivity of emissions to ambiguous assignments.

Response 7:

We fully agree that treating engine assignment as a "black box" undermines the reproducibility of the emission inventory. To address this, we have comprehensively detailed the complete linkage procedure in the revised manuscript. We clarified that the linkage is fundamentally anchored by the unique FLIGHT_INFO_ID rather than relying solely on aircraft registration, ensuring a one-to-one flight mapping. Furthermore, we explicitly documented a robust seven-level hierarchical matching procedure to link the identified engine types with the ICAO EEDB. To maintain narrative flow, the specific criteria for each matching level have been detailed in the notes of Supplementary Table S5. The matching quality diagnostics, demonstrating that over 82.0%

of engines were resolved through exact or standardized exact matching, are now transparently reported alongside detailed matching coverage rates in the body of Supplementary Table S5.

In Section 2.3.1, the text was refined to document the database linkage and final assignment diagnostics, with the detailed matching rules successfully moved to the supplementary table notes, reading as follows:

"Before constructing the high-fidelity emission model, each QAR trajectory record was mapped to its corresponding aircraft and engine type. The QAR trajectory data were first linked to the flight schedule data using the unique identifier FLIGHT_INFO_ID, resulting in complete matching of the basic flight information. For the successfully linked records, the aircraft registration number (AC_TAIL) and aircraft type (AC_TYPE) were extracted and linked to the aircraft static database to identify the engine types installed on the corresponding aircraft. Because engine nomenclature varied across the databases, we developed a seven-level hierarchical matching procedure to link the identified engine types with the International Civil Aviation Organization (ICAO) Aircraft Engine Emissions Databank (EEDB). The procedure was applied sequentially, with the next level activated only when the preceding level failed to return an acceptable match. The corresponding coverage rates for each matching level are summarized in Table S5, while the detailed criteria for the seven-level matching procedure are explicitly documented in the table notes. Among the 4.52 million retained flight records, 98.7% were assigned two engines, whereas the remaining 1.3% corresponded to aircraft types inherently equipped with a single engine. More than 82.0% of the successfully resolved records were obtained through exact or standardized exact matching. A small number of unresolved records were subsequently verified against aircraft registration records and corrected manually. This procedure established an engine-specific reference assignment for the retained QAR records."

Table S5. Hierarchical aircraft-engine matching rules and corresponding flight coverage rates

Matching level	Matching method	Proportion
1	Exact	59.9904%

2	Normalized_Exact	22.0960%
3	Series_Fallback	0.0264%
4	Contains	9.0603%
5	Fuzzy	7.7783%
6	Core_Number_Fallback	0.0158%
7	Information_Inference	1.0328%
	Total	100.0000%

Notes:

Detailed criteria for each matching level are as follows:

(1) Direct exact matching. The original engine designation was directly compared with the EEDB engine identifier for an exact string match.

(2) Standardized exact matching. If no direct match was found, both engine designations were standardized by harmonizing letter case, removing spaces, and normalizing symbols such as underscores and hyphens, followed by another exact comparison.

(3) Engine-series matching. If the complete engine subtype remained unresolved, the available engine series was used to identify a corresponding generic engine designation within the same engine family.

(4) Containment matching. The matching was further relaxed by checking whether the core engine designation appeared as a complete substring within the EEDB designation.

(5) Fuzzy matching. For records that remained unresolved, fuzzy matching based on Levenshtein similarity was applied, with a similarity threshold of 0.80.

(6) Numerical-sequence matching. If fuzzy matching also failed, the characteristic numerical sequence contained in the engine designation was extracted and used to identify the corresponding EEDB candidate.

(7) Historical-information matching. When the engine description was missing or insufficient for direct EEDB matching, historical engine information was used sequentially, first from the same aircraft registration, then from the historically dominant engine associated with the same aircraft type, and finally from the representative engine assigned to that aircraft type.

Comment 8:

The SO₂ assumptions appear internally inconsistent. The nominal inventory uses a fixed EI of 1 g/kg, as reflected by the direct correspondence between 28.4 Mt of fuel and 28.4 kt of SO₂, whereas the Monte Carlo analysis assumes a mean fuel sulfur content of 600 ppm with a standard deviation of 150 ppm. These assumptions do not appear to represent the same nominal sulfur condition under a simple sulfur mass balance. The point-estimate and uncertainty assumptions should be reconciled, and the sulfur-to-SO₂ conversion should be stated explicitly.

Response 8:

We fully agree that the point-estimate calculation and the uncertainty analysis must be strictly aligned under the exact same nominal sulfur condition. In the revised manuscript, we explicitly clarified that under a stoichiometric sulfur mass balance, the nominal SO₂ emission index of 1.0 g/kg fuel directly corresponds to a nominal fuel sulfur content (FSC) of 500 ppm by mass, based on the molecular weight ratio of SO₂ to atomic sulfur (64/32). To ensure rigorous internal consistency across the entire modeling framework, the Monte Carlo uncertainty parameters were corrected to utilize this aligned nominal value of 500 ppm as the mean of the FSC distribution, with a standard deviation of 100 ppm.

In Section 2.3.1, the text was revised to explicitly state the fixed SO₂ emission index applied to the point-estimate calculations, reading as follows:

"Following a fuel mass-balance formulation, CO₂ and SO₂ emissions were calculated using fixed EIs rather than thrust-dependent EEDB curves. In accordance with the values adopted in previous aviation emission inventories, fixed EIs of 3150 g kg⁻¹ for CO₂ and 1 g kg⁻¹ for SO₂ were used (Chen et al., 2021; Howitt et al., 2011). The EEDB reference data were therefore used to parameterize the EIs of NO_x, CO, HC, and non-volatile particulate matter (nvPM), whose emission characteristics vary with engine operating conditions."

In Section 2.3.3, the prescribed baseline FSC utilized for sulfate estimation is explicitly stated as 500 ppm, formally establishing the scaling baseline, reading as follows:

"...where FSC is the fuel sulfur content, prescribed as 500 ppm by mass, and ϵ is the fraction of

fuel sulfur converted from S(IV) to S(VI), which was set to 0.024 following the value adopted in the FOA methodology. MW_{SO_4} and MW_S are the molecular weights of sulfate and sulfur, respectively, with values of 96 and 32."

In Section 2.4, the Monte Carlo parameterization was updated to strictly align with the nominal 500 ppm FSC baseline, eliminating the previous discrepancy.

Comment 9:

The PM fit for the CF6-80A engine deserves specific verification. The reported parameters and are extreme outliers relative to the other engine models and imply an exceptionally strong dependence of EI on fuel flow. The underlying EEDB anchor points, units, fitted curve, and influence of this engine-specific relationship on the national PM total should be shown. This is particularly important given that the reported PM inventory is substantially higher than most previous estimates.

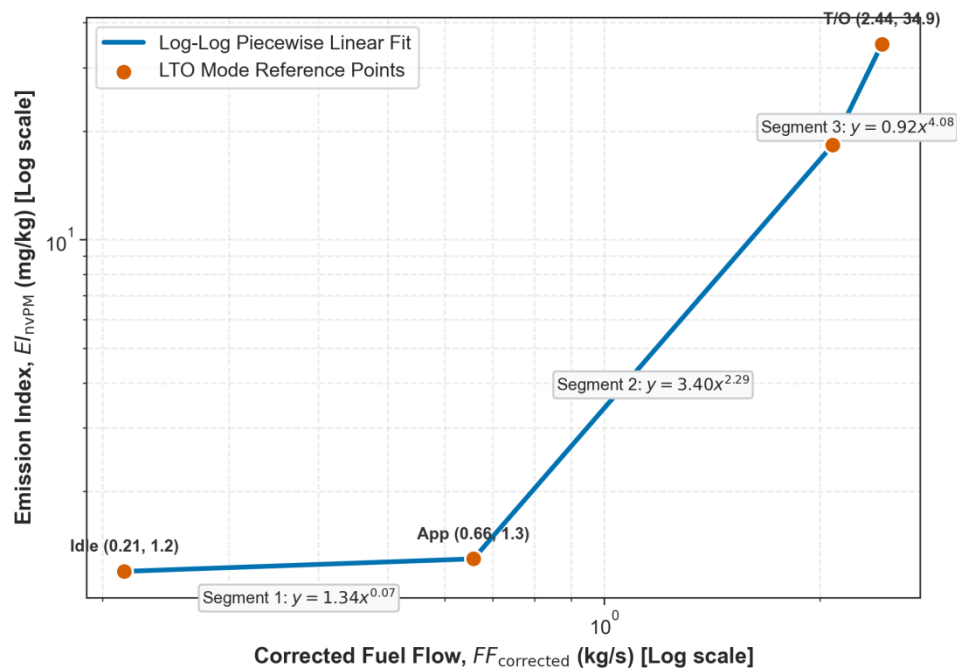
Response 9:

We revisited the original EI parameterization and agree that the previous single-function fitting approach had limitations. For some engine types, the distribution of the four EEDB operating points could lead to exceptionally large fitting parameters and, in some cases, fitted EI values outside a physically reasonable range. We therefore considered that adjusting the CF6-80A parameters alone would not adequately address the underlying methodological issue. The revised EI parameterization is now described in detail in the new Section 2.3.1, where the distinct log-domain fitting approaches for NO_x , CO, HC, and nvPM are explicitly presented.

Accordingly, we revised the EI-fuel-flow parameterization for all four pollutants that require EEDB-based fitting, namely NO_x , CO, HC, and nvPM. The revised framework is directly constrained by the four certified EEDB operating points (7%, 30%, 85%, and 100% thrust) and uses two distinct log-domain approaches. CO and HC are represented using a bilinear log-log parameterization, with the low-thrust branch defined by the 7% and 30% points and the

high-thrust branch constrained by the 85% and 100% points. NO_x and nvPM are instead treated using point-to-point piecewise interpolation in log-log space, with each segment determined only by two adjacent EEDB anchor points.

The below figure illustrates the revised point-to-point log-log parameterization for nvPM using the CF6-80A EEDB reference points, with each local segment constrained by two adjacent certification points.



The above figure provides a more data-constrained representation of the EEDB relationships than the previous global fitting approach. In particular, the revised nvPM parameterization does not impose a single functional form across all thrust conditions and therefore preserves the local variation of the certified EEDB data, including cases where the EI-fuel-flow relationship is strongly nonlinear or non-monotonic. The calculation is also restricted to the valid EEDB range of 7-100% thrust, with no additional functional extrapolation beyond the certified operating points. These constraints substantially reduce the possibility that an anomalous global fitting parameter, such as that previously obtained for CF6-80A, disproportionately affects the estimated nvPM emissions.

We have also clarified that PM is not independently fitted as a single pollutant. Total PM is

calculated by combining the EEDB-based nvPM component with the HC-derived volatile component and the sulfur-related sulfate component. The revised treatment is applied consistently across all engine types rather than being introduced as an engine-specific adjustment. We believe this methodological revision provides a more robust and physically constrained basis for the engine-specific EI estimates and, consequently, for the national PM inventory.

In Section 2.3.1, the text was extensively revised to explicitly define the distinct log-domain fitting strategies for gaseous pollutants and nvPM, eliminating the previous global fitting approach. The revised text reads as follows:

"For the gaseous pollutants, the EEDB provides engine-specific reference data at four standardized thrust settings corresponding to 7%, 30%, 85%, and 100% of maximum rated thrust. These certification points are used as the reference basis for establishing continuous EI–fuel-flow relationships under the fuel-flow-based emission modeling framework (DuBois & Paynter, 2006). Following the established log-log parameterization used in fuel-flow emission models, different fitting strategies were applied according to the observed characteristics of the emission indices. For CO and HC, a bilinear relationship was constructed in log-log space. The first segment was defined by the 7% and 30% certification points, while the second segment was represented by a horizontal line at the mean EI level of the 85% and 100% points; the two segments were extended to their intersection to form the complete reference curve. For NO_x, a point-to-point piecewise linear relationship was established between adjacent certification points in log-log space."

"The available EEDB nvPM data were treated separately from the classical BFFM2 gaseous-emission formulation. The EEDB reports engine-specific nvPM mass and number EIs at the same four ICAO certification points for applicable engine types. In this study, the available nvPM reference points were connected using a local point-to-point interpolation scheme in log-log space to provide a continuous empirical EI–fuel-flow relationship over the valid 7–100% reference range. This treatment preserves the local variation represented by the certification data without imposing a single global functional form on the nvPM emission profile. The resulting engine-specific reference EI relationships were subsequently combined with the instantaneous single-engine fuel-flow measurements from QAR to determine the reference EIs corresponding to

actual flight operating conditions."

Comment 10:

The 13.5% difference between the QAR-derived fuel consumption of 28.4 Mt and the CAAC-derived estimate of 25.03 Mt requires a more rigorous validation. The current explanation based mainly on accounting definitions is plausible but not sufficient, especially because the QAR dataset covers only 96.8% of reported commercial flight movements yet already produces a larger national fuel total. The comparison should first harmonize system boundaries and then examine possible contributions from scope differences, duplicate records, fuel-flow integration, unit conversion, and other systematic effects.

Response 10:

We agree that the difference between the QAR-derived and CAAC-reported fuel consumption required a more rigorous re-examination. Following the reviewer's suggestion, we revisited the system boundary and the entire calculation workflow, including flight screening, QAR preprocessing, fuel-flow integration, unit conversion, aircraft-engine matching, reference fuel-flow conversion, engine-specific EI parameterization, and atmospheric corrections. The revised comparison of CO₂, NO_x, SO₂, PM, CO, and HC with the corresponding CAAC estimates is now presented in Section 3.1.

The most important issue identified was the original inclusion of a small number of international flights and domestic flights involving Hong Kong, Macao, and Taiwan in the national inventory. Their emissions and fuel consumption were consequently included in the original national totals, leading to an overestimation of both fuel consumption and CO₂ emissions. We therefore redefined the study boundary as commercial domestic flights operating within mainland China in 2023, explicitly excluding international/regional inbound and outbound flights. All the emission estimates were then recalculated using this harmonized boundary.

We also rechecked the QAR fuel-flow integration and unit conversions, as well as the

aircraft-engine matching and subsequent reference fuel-flow and emission-index calculations. These checks were conducted to ensure that the revised difference was not caused by duplicate records, integration errors, unit inconsistencies, or systematic errors in the engine-specific calculation procedure.

After applying the revised flight boundary and recalculating the inventory, national CO₂ emissions decreased from 89.3 Mt to 74.6 Mt. The corresponding fuel consumption derived from the carbon balance is approximately 23.7 Mt, compared with the CAAC estimate of 25.03 Mt. Thus, the original approximately 13.5% higher fuel estimate is no longer present; after harmonizing the flight scope, the QAR-derived estimate is approximately 5.3% lower than the CAAC estimate.

The remaining difference is considered more likely to reflect differences in data coverage and accounting definitions between the QAR-based operational calculation and the CAAC statistical estimate, rather than a single computational bias.

In Section 3.1, the text was comprehensively revised to explicitly define the harmonized domestic boundary and present the recalculated fuel consumption and emission totals, reading as follows:

"Derived from 1-Hz operational observations across approximately 4.52 million QAR-recorded flights, this study compiled a four-dimensional emission inventory for purely domestic civil aviation flights within mainland China in 2023, excluding flights involving Hong Kong, Macao, and Taiwan as well as international inbound and outbound operations. Under this system boundary, total annual fuel consumption is estimated at 23.7 Mt, corresponding to total emissions of 74.6 Mt for CO₂, 311.84 kt for NO_x (reported as NO₂ mass equivalent), 73.5 kt for CO, 6.02 kt for HC, 23.69 kt for SO₂, and 2.00 kt for PM. To evaluate the macroscopic consistency of the inventory, the QAR-based fuel consumption was compared against official industry statistics released by the Civil Aviation Administration of China (CAAC, 2024). Based on the annual transport turnover and average fuel efficiency reported by the CAAC, national domestic fuel consumption in 2023 is estimated at approximately 25.03 Mt, yielding a relative difference of 5.3% compared with our bottom-up summation. Here, the CAAC statistical value serves as an independent national-scale benchmark of overall magnitude. While the retained QAR archive captures approximately 96.8% of the commercial flight movements reported by the CAAC, flight-count coverage does not

correspond directly to fuel-burn coverage because individual flights vary in aircraft type, stage length, and payload. Overall, the 5.3% difference indicates good agreement in magnitude between the integration of 1-Hz onboard fuel-flow measurements and top-down statistical accounting. The statistical distribution of major engine models and departure airports comprising this domestic fleet is detailed in Figure S2. Furthermore, monthly variations in flight activity and emission totals across the six species are presented in Figure S3.

Compared with previous aviation inventories, the estimates in this study demonstrate strong macroscopic alignment in species directly proportional to fuel burn. Our total fuel consumption of 23.7 Mt and the corresponding CO₂ emissions of 74.6 Mt are broadly consistent with top-down statistical baselines and recent discrete-flight assessments for the national fleet. Because CO₂ generation is strictly governed by stoichiometric fuel mass balance, this baseline agreement validates the comprehensive capture of flight activities and the overall representation of route network intensity.

Building upon this validated baseline, meaningful divergences emerge in thermodynamically sensitive pollutants and incomplete combustion products (Table 4). Specifically, the estimated CO emissions (73.5 kt) are higher than previous full-flight values (39.7–59.6 kt), whereas NO_x emissions (311.84 kt) are substantially lower than trajectory-modeled upper estimates (up to 477.2 kt). Rather than mere methodological noise, these discrepancies highlight the inherent limitations of relying on idealized operating profiles. Schedule- and trajectory-based models often underrepresent prolonged low-power ground operations, such as extended taxiing and holding which disproportionately elevate CO generation due to incomplete combustion. Conversely, generic performance models frequently assume maximum rated thrust during takeoff and climb, leading to a systematic overestimation of peak combustor temperatures and thermal NO_x formation (Koudis et al., 2017). By coupling 1-Hz measured fuel flow with actual in-flight ambient temperature, pressure, and specific humidity, our QAR-based framework directly captures real-world transient operations (e.g., reduced-thrust takeoffs) and dynamic environmental effects. This approach transitions the emission estimates from model-parameterized approximations to a physically constrained operational baseline, with the corresponding uncertainty propagation formally evaluated in Section 3.2."

Table 4: Comparison of estimated annual civil aviation emissions in China across different

inventory studies.

Inventory	Source	Year	CO/kt	HC/kt	PM/kt	SO ₂ /kt	NO _x /kt	CO ₂ /Mt
Our study	QAR	2023	73.5	6.02	2.0	23.69	311.84	74.6
Ma et al. (2024)	AIP	2019	58.2	5.03	3.2	36.2	349.9	86.7
Zhang et al.(2022)	ADS-B	2018	-	14.3	1.64	21.3	477.2	81.3
Lang et al. (2024)	ADS-B	2023(LTO)	46.1	3.2	1.1	18.4	62.3	-
Fan et al. (2012)	Great-circle	2010	39.7	4.6	-	9.7	154.1	38.21
Liu et al. (2019)	Great-circle	2015	59.6	4.77	3.32	19	304.8	59.96

Comment 11:

The unusually low CO estimate also warrants closer examination. The reported 2023 full-flight CO total of 44.3 kt is lower than the 46.1 kt reported by another 2023 study for the LTO phase alone. This is too substantial to dismiss as a generic difference in taxi or idle treatment.

Response 11:

We thank the reviewer for pointing out this important issue. We agree that the original estimate of 44.3 kt for full-flight CO emissions was unexpectedly low compared with the 46.1 kt LTO estimate reported by an independent 2023 study. We have therefore thoroughly re-examined the emission modeling pipeline in Section 3.1 and recalculated the domestic full-flight inventory using the standardized calculation framework detailed above. Following this revision, the domestic full-flight CO emission total increases from 44.3 kt to 73.5 kt.

Notably, this 65.9% increase in CO occurs despite a stricter domestic flight-scope filtering that reduced total fuel consumption by approximately 16.5%. This upward revision is driven primarily by the refined BFFM2 bilinear fitting in log-log space, which provides a continuous, physically realistic representation of emission index (EI) variation across intermediate engine power settings. Specifically, in the revised bilinear formulation, the log-linear decay segment connecting the idle (7%) and approach (30%) certification points extends across the intermediate equivalent fuel-flow

regime typical of cruise and in-route descent before intersecting the high-thrust plateau (mean of climb-out and take-off). Because the previous calculation did not fully capture the continuous CO contribution outside the LTO regime, incorporating this cruise-phase CO contribution more than offsets the reduction in total flight activity. Consequently, the revised full-flight CO estimate (73.5 kt) exceeds the independent LTO benchmark (46.1 kt), resolving the inconsistency identified by the reviewer.

The adjustments in the remaining pollutants are likewise fully consistent with the updated calculation framework:

(1) CO₂ and SO₂ : CO₂ decreased by 16.5% (from 89.3 Mt to 74.6 Mt) in direct proportion to the reduced fuel consumption resulting from the stricter domestic flight definition. Correspondingly, SO₂ decreased by 16.6% (from 28.4 kt to 23.69 kt), as its point estimate is calculated using the fixed nominal emission index of 1.0 g SO₂ /kg fuel (anchored to a nominal fuel sulfur content of 500 ppm) and therefore directly tracks the revised fuel consumption.

(2) NO_x: NO_x decreased substantially from 634.6 kt to 311.84 kt (-50.9%) because the revised calculation incorporates the updated fuel-flow conversion alongside actual atmospheric temperature, pressure, and specific humidity conditions, resulting in lower corrected NO_x emission indices under actual flight conditions.

(3) HC: Unlike CO, HC does not exhibit an upward shift, instead showing a slight net decrease from 6.35 kt to 6.02 kt (-5.2%). This divergence occurs because the log-log decay slope between idle and approach in the EEDB is much steeper for HC than for CO, causing the low-power regression line for HC to reach its high-thrust minimum floor at a much lower reference fuel flow (just above the approach setting). Although the revised bilinear fitting also provides continuous HC estimates during cruise, the resulting cruise-phase HC increment is very limited and is outweighed by the 16.5% reduction in flight activity and the revised engine operating-condition assignments. Thus, the contrasting responses of CO and HC reflect their distinct emission index behaviors across engine operating regimes rather than an inconsistency in the fitting procedure.

(4) PM: Total PM decreased from 8.46 kt to 2.00 kt (-76.4%). In the revised framework, total PM is resolved into three components including volatile organic PM, volatile sulfate PM, and nvPM following the FOA methodology and EEDB reference data. The substantial reduction is primarily driven by the improved fitting procedure for the nvPM component. Specifically, the original calculation relied on a global power-law fitting ($y=ax^b$), which produced an overly steep curve and generated excessively high emission indices at medium-to-high fuel flows. In the revised calculation, nvPM is modeled using a point-to-point linear connection in log-log space across the four LTO certification modes. This piecewise log-linear approach strictly bounds the interpolated emission indices within a controllable, physically realistic range that closely conforms to the EEDB reference specifications, thereby eliminating the severe high-power overestimation of the previous power-law curve.

The changes in pollutant totals reflect the combined effects of the revised flight scope, fuel consumption, engine assignment, and emission index parameterization, rather than an adjustment targeted at any individual pollutant. The revised CO estimate is consequently consistent with the full-flight emission boundary and with the expected magnitude relative to LTO emissions.

Because these fundamental recalculations and EI parameterization updates substantially altered the magnitude and transient behavior of the emission totals, we have correspondingly updated the downstream temporal and spatial analyses. All high-frequency emission profiles in Section 3.3 and the 3D spatial distribution maps in Section 3.4 have been completely recalculated and redrawn. This systematic update ensures that the newly resolved transient emission peaks (e.g., during low-power ground operations and descent) and the spatial clustering behaviors are strictly and physically consistent with the revised full-flight emission baseline.

Comment 12:

Several interpretations extend beyond what the inventory itself demonstrates. In particular, CO₂ emissions below 3000 m should not be directly interpreted as near-surface air-quality impacts, and CO₂ is not generally the principal pollutant of concern in airport local-air-quality assessments.

These statements should be limited to emission location and potential relevance, while conclusions about local air quality or exposure should be reserved for pollutants such as NO_x and PM or supported by a dedicated dispersion analysis.

Response to Comment 12:

We fully agree that an emission inventory alone characterizes emission sources rather than ambient concentrations or population exposure, and that CO₂, as a greenhouse gas rather than a criteria air pollutant, should not be directly interpreted in terms of local air-quality impacts. Accordingly, we have revised the third paragraph of Section 3.4.1 by removing all direct links between low-altitude CO₂ emissions and near-surface air-quality degradation, while also correcting the standard LTO mixing-height threshold from the misstated 3,000 m to 914 m (3,000 ft). In the revised text, the low-altitude CO₂ results are retained strictly as a proxy for fuel burn to illustrate the three-dimensional spatiotemporal distribution, terminal-area operational intensity, and vertical structure of aviation activity.

In Section 3.4.1, the interpretation of the vertical emission profiles was substantially revised. All unsupported claims regarding CO₂ and air quality were removed, and the boundary conditions for air-quality assessments were explicitly defined, reading as follows:

"Figure 6(c) shows the vertical distribution of CO₂ emissions along the Chengdu–Beijing and Beijing–Chengdu routes. Using 3,000 ft (approximately 914 m) as the boundary between the LTO and CCD phases, the LTO phase accounts for only 1.9% of the total flight distance but contributes 7.1% of total CO₂ emissions. In contrast, the CCD phase above 914 m accounts for 98.1% of the flight distance and 92.9% of total CO₂ emissions. The disproportionately higher emission share of the LTO phase relative to its distance share reflects the elevated fuel burn rates per unit distance associated with takeoff, initial climb, approach, taxi, and other low-altitude operations. The vertical distribution also exhibits pronounced emission peaks at several cruise altitudes, particularly around 9,600–10,000 m, reflecting the concentration of flight operations within assigned cruise flight levels. Here, the resolved CO₂ profile serves as a direct indicator of fuel-burn intensity and vertical operational clustering. Because the intense low-altitude fuel burn during the LTO phase occurs near airports and populated terminal areas, the co-emitted criteria

pollutants (specifically NO_x and PM) within this layer are of particular relevance as inputs for subsequent dispersion modeling and local air-quality assessments (Unal et al., 2005; Y. Wang et al., 2023)."

Reviewer 2

General comment:

I agree with all the comments provided by Referee #1 and would like to add the following comments:

I agree that methodology presented in the manuscript has a certain degree of novelty. However, I have several concerns, ranging from minor to major regarding publication of the manuscript in its current form.

Overall response:

We sincerely thank the reviewer for the constructive assessment of our manuscript, particularly recognizing the methodological novelty of this study. In revising the manuscript, we have carefully considered all the comments and have addressed the methodological and interpretative issues raised here. Our point-by-point responses below explain how each concern has been addressed and identify the corresponding changes made to the manuscript.

Comment 1:

1. Line 38-39: I am not fully convinced by the statement that a "high-resolution aviation emission inventory serves as a critical link connecting aviation activities with atmospheric environmental responses." The manuscript should provide a clearer justification for why such high spatial and temporal resolution is necessary and what specific scientific or practical advantages it offers. This point is particularly important because the conclusion states that "the inventory provides second-level temporal resolution and nationwide coverage, offering a basis for near-airport air

quality assessment and for evaluating the assumptions underlying existing lower-resolution aviation emission inventories." I suggest that the Introduction be revised to better align with this more specific motivation. For example, Yim et al. (2015), which is cited in the manuscript, employed an atmospheric model with a horizontal resolution of approximately $4^{\circ} \times 5^{\circ}$. When an emission inventory with a resolution of $0.1^{\circ} \times 0.1^{\circ}$ is used in such a coarse-resolution model, much of the benefit of the fine spatial resolution may be lost through model-grid aggregation. Therefore, the manuscript should more clearly explain under what applications the proposed high-resolution inventory provides meaningful added value. If the primary advantages of the inventory are improved estimation of total emissions and/or its applicability to regional- or local-scale air quality assessments, particularly near airports, I find these motivations reasonable. However, the current Introduction does not articulate these points sufficiently clearly and should be revised accordingly.

Response 1:

We fully agree that the original statement was overly broad, and we appreciate the critical point regarding model-grid aggregation, a $0.1^{\circ} \times 0.1^{\circ}$ high-resolution inventory indeed loses its spatial advantages when applied to coarse-grid global models like the $4^{\circ} \times 5^{\circ}$ grid used in Yim et al. (2015). Following the exact recommendations, we have thoroughly revised the Introduction to explicitly articulate the specific applications where our high-resolution inventory provides meaningful added value: (1) reducing aggregation biases in total emission estimates, and (2) supporting regional and near-airport local air quality assessments. We have also aligned the Conclusion to ensure perfect consistency with these specific motivations.

In Section 1, Paragraph 1 was revised to explicitly differentiate the spatial scales of emission impacts, establishing the foundation for high-resolution requirements, reading as follows:

"...Furthermore, they pose serious threats to public health by deteriorating air quality across multiple spatial scales (Yim et al., 2013). While emissions at cruise altitudes contribute to large-scale background air pollution and premature mortality globally (Barrett et al., 2010), low-altitude emissions during landing and take-off (LTO) operations directly degrade near-surface air quality in densely populated airport vicinity areas (Yim et al., 2015). Given the pronounced

spatiotemporal heterogeneity of aviation activities across flight phases and altitudes, accurately quantifying aircraft emissions and clarifying their complex distributions serve as the scientific foundation for targeted emission control and environmental impact assessment (Lu et al., 2024)."

In Section 1, Paragraph 2 was revised to transition smoothly from these impact scales to the specific application domains where bottom-up, high-resolution modeling is strictly necessary, reading as follows:

"Aviation emission inventories provide the fundamental basis for quantifying emission loads, evaluating mitigation strategies, and supporting atmospheric modeling (Simone et al., 2013; Stettler et al., 2013). Depending on data availability and target application scales, inventory compilation methods are generally categorized into top-down and bottom-up approaches. The top-down approach estimates aggregate emissions primarily from global or national jet fuel sales statistics (e.g., from the IEA). While computationally straightforward and sufficient for macro-level carbon accounting or coarse-grid global modeling, top-down inventories cannot resolve sub-grid spatial gradients or operational dynamics. Consequently, higher spatiotemporal resolution becomes essential when the objective shifts toward reducing aggregation biases in total emission estimates, characterizing emission hotspots along busy flight corridors, or supporting regional and local air quality assessments around airport terminal areas. To meet these demands, compilation methods have increasingly shifted toward the bottom-up approach, which estimates fuel consumption and emissions by aggregating individual flight activities across space and time, thereby providing both improved estimation fidelity and fine-grained spatiotemporal resolution (Cui et al., 2022)."

Furthermore, we refined the final paragraph of the Introduction to explicitly summarize how our dataset fulfills these precise demands:

"To address these limitations, this study constructs a national-scale aviation emission inventory for China's domestic civil aviation sector using 1-Hz QAR records from approximately 4.52 million flights operated in 2023. Synthesizing measured single-engine fuel flow with airborne thermodynamic parameters and reanalysis humidity profiles, we derive a four-dimensional inventory of CO₂, NO_x, SO₂, PM, CO, and HC emissions, and examine how input uncertainties

propagate across QAR, ADS-B, and schedule-based frameworks. This observation-based dataset constrains activity and fuel-flow uncertainties in national emission accounting, resolves phase-level and terminal-area emission patterns relevant to regional and near-airport air-quality assessments, thereby establishing an empirical constraint for regional exposure assessments and evaluating legacy assumptions in lower-resolution models."

Finally, in Section 5 (Conclusion), we explicitly solidified these two core practical advantages to ensure full consistency with the Introduction, reading as follows:

"Ultimately, this dataset provides a robust empirical baseline for improving national total emission estimates and offers a high-resolution physical foundation for regional and near-airport air-quality assessments."

Comment 2:

2. Lines 100 and 143-144: The manuscript states that the dataset covers approximately 4.52 million flights operating to and from civil airports across China in 2023 and later refers to a comprehensive dataset of civil aviation operations in mainland China. However, the exact spatial scope of the dataset is not sufficiently clear. From a data-coverage perspective, it appears that the inventory may be intended to represent aviation emissions over mainland China, but this should be stated explicitly. In particular, the authors should clarify how flight trajectories and emissions were treated at the geographical boundaries of the study domain. For example, for international flights departing from or arriving at airports in mainland China, were emissions included only while the aircraft was located within the defined mainland China domain, or were the full flight trajectories included? Similarly, how were flights crossing the domain boundary handled?

Response 2:

We fully agree that the exact geographical domain and flight inclusion criteria must be explicitly defined to avoid ambiguity. In the revised manuscript, we have clearly specified that the study domain is strictly bounded to purely domestic flights operating exclusively within mainland China.

Flights with departure or arrival airports located outside mainland China, including all international flights as well as domestic flights involving Hong Kong, Macao, and Taiwan, were entirely excluded prior to the emission calculation. Consequently, because the retained dataset consists exclusively of domestic flights connecting two airports within mainland China, their full gate-to-gate trajectories were retained and computed in their entirety, meaning no artificial boundary clipping or partial-flight truncations were required during the emission quantification.

In Section 2.2.1, the text was revised to explicitly define the geographical boundary and the strict exclusion criteria for international and regional flights, reading as follows:

"The initial archive contained approximately 4.85 million flight records for the entire year of 2023. As this study aims to construct a civil aviation emission inventory for strictly domestic flights within mainland China, flights with departure or arrival airports outside mainland China, as well as flights involving Hong Kong, Macao, and Taiwan, were excluded. After flight-scope screening and the removal of invalid records, approximately 4.52 million valid domestic flight records were retained, accounting for 93.2% of the original archive (details of the initial data screening are provided in Table S1)."

Table S1. Summary of QAR Flight Screening and Data Retention

Data Processing Stage	Proportion of Raw Records
	(%)
Raw QAR records	100.0%
Flights beyond the study scope	6.2%
Domestic flight candidate records	93.8%
Invalid records	0.6%
Final valid domestic flights	93.2%

Comment 3:

3. Line 168: Does the term "dynamic environmental correction model" refer to the methodology

described in Section 2.3.2, "Atmospheric Correction of Emission Indices"? If so, I suggest using consistent terminology throughout the manuscript to avoid potential confusion.

Response to Comment 3:

We confirm that the term originally referred to as the "dynamic environmental correction model" describes the exact same methodology detailed in the section for atmospheric corrections. To avoid any potential confusion and ensure clarity, we have standardized the terminology throughout the manuscript, ensuring that all references to this procedure consistently align with the specific title and framework of "Atmospheric Correction of Emission Indices".

In Section 2.3.3, the section heading was explicitly standardized to reflect the exact terminology, reading as follows:

"2.3.3 Atmospheric Correction of Emission Indices"

In Section 2.3.4, the subsequent text was aligned to consistently use this standardized terminology when referring back to the preceding step, reading as follows:

"Following the atmospheric correction of emission indices, pollutant emissions were quantified at the individual-engine level using the 1-Hz QAR observations."

Comment 4:

4. Line 214, Equation (5): Please check Equation (5). It appears that the exponent should be 0.51 and 1.65. It is currently unclear whether this is only a typo in the manuscript or whether the equation was implemented incorrectly in the actual calculations. This should be carefully verified, as it may directly affect the resulting emissions estimates.

Response 4:

We have verified our calculation code and confirm that this discrepancy was purely a typesetting error in the original manuscript, where the outer exponent of 0.5 in the BFFM2 pressure and temperature correction term was inadvertently formatted as a multiplication factor. In the actual

emission calculations, the standard BFFM2 formulation with the correct mathematical scaling was properly implemented. In the revised manuscript, the NO_x atmospheric correction equation is now explicitly written in its standard form, retaining the outer exponent of 0.5 applied to the baseline parameters (which is mathematically equivalent to the expanded exponents of 0.51 and 1.65 noted by the reviewer). We have also eliminated the duplicate typographical error to ensure consistency. In Section 2.3.3, the NO_x atmospheric correction equation was explicitly corrected to accurately display the standard BFFM2 mathematical formulation, reading as follows:

"The NO_x emission index was corrected for both temperature and pressure effects and ambient humidity:

$$H = -19(\omega - 0.00634) \quad (5)$$

$$EI_{NO_x, flight} = EI_{NO_x, ref} \left(\frac{\delta_{amb}^{1.02}}{\theta_{amb}^{3.3}} \right)^{0.5} e^H \quad (6)$$

where ω is the ambient specific humidity."

Comment 5:

5. As also noted by RC #1, I believe that several parts of the manuscript go beyond what can be directly supported by the presented analysis. For example, the discussion beginning around Line 298 appears to attribute the differences between the emission inventories to specific causes that have not been explained in the study. While these explanations look plausible, they remain hypotheses unless authors apply the stated assumptions to the QAR data. Subsequently, the authors should reconstruct the emission inventory accordingly and then compare the resulting estimates with the CAAC inventory. Unless this statement may be removed from the document.

Response 5:

We fully agree that attributing discrepancies to specific operational factors remains speculative without dedicated sensitivity reconstructions. Following your suggestion, we removed all unsupported causal attributions from the manuscript. As detailed in our response to Reviewer 1, after strictly harmonizing the system boundary to purely domestic flights within mainland China,

our estimated fuel consumption of 23.7 Mt yields a 5.3% difference relative to the CAAC statistical estimate of 25.03 Mt. In the revised manuscript, we now present the CAAC figure strictly as an independent macroscopic benchmark of overall magnitude. We limit our discussion to the factual observation that our 96.8% flight-count coverage does not directly translate to an identical fuel-burn coverage due to variations in aircraft type, stage length, and payload, deliberately avoiding any further speculative hypotheses.

In Section 3.1, the discussion comparing the QAR-derived estimates with the CAAC inventory was comprehensively revised to remove speculative causal attributions, reading as follows:

"To evaluate the macroscopic consistency of the inventory, the QAR-based fuel consumption was compared against official industry statistics released by the Civil Aviation Administration of China (CAAC, 2024). Based on the annual transport turnover and average fuel efficiency reported by the CAAC, national domestic fuel consumption in 2023 is estimated at approximately 25.03 Mt, yielding a relative difference of 5.3% compared with our bottom-up summation. Here, the CAAC statistical value serves as an independent national-scale benchmark of overall magnitude. While the retained QAR archive captures approximately 96.8% of the commercial flight movements reported by the CAAC, flight-count coverage does not correspond directly to fuel-burn coverage because individual flights vary in aircraft type, stage length, and payload. Overall, the 5.3% difference indicates good agreement in magnitude between the integration of 1-Hz onboard fuel-flow measurements and top-down statistical accounting. The statistical distribution of major engine models and departure airports comprising this domestic fleet is detailed in Figure S2. Furthermore, monthly variations in flight activity and emission totals across the six species are presented in Figure S3."

Comment 6:

6. A similar issue appears in Lines 329-330. In this case, appropriate citations may help support "general" validity of the proposed explanation. However, there should still be a clear distinction between "general" statement and what has been presented by the results of the current manuscript.

Response 6:

We fully agree that cross-study comparisons must be supported by appropriate citations and that general domain knowledge should be clearly distinguished from the specific empirical findings of our study. Accordingly, we have comprehensively revised and condensed the comparative discussion in Section 3.1. We now explicitly discuss the observed magnitude differences across the three pollutant groups by referencing established operational and combustion mechanisms grounded in published literature, strictly avoiding unsupported speculation across external inventories. Meanwhile, quantitative claims regarding error propagation in this model are reserved strictly for the empirical Monte Carlo and Shapley analyses in Section 3.2.

In Section 3.1, the comparative discussion was revised to explicitly support all general domain knowledge and cross-study explanations with robust literature citations, reading as follows:

"Building upon this validated baseline, meaningful divergences emerge in thermodynamically sensitive pollutants and incomplete combustion products (Table 4). Specifically, the estimated CO emissions (73.5 kt) are higher than previous full-flight values (39.7–59.6 kt), whereas NO_x emissions (311.84 kt) are substantially lower than trajectory-modeled upper estimates (up to 477.2 kt). Rather than mere methodological noise, these discrepancies highlight the inherent limitations of relying on idealized operating profiles. Schedule- and trajectory-based models often underrepresent prolonged low-power ground operations, such as extended taxiing and holding which disproportionately elevate CO generation due to incomplete combustion. Conversely, generic performance models frequently assume maximum rated thrust during takeoff and climb, leading to a systematic overestimation of peak combustor temperatures and thermal NO_x formation (Koudis et al., 2017). By coupling 1-Hz measured fuel flow with actual in-flight ambient temperature, pressure, and specific humidity, our QAR-based framework directly captures real-world transient operations (e.g., reduced-thrust takeoffs) and dynamic environmental effects. This approach transitions the emission estimates from model-parameterized approximations to a physically constrained operational baseline, with the corresponding uncertainty propagation formally evaluated in Section 3.2."

Comment 7

7. Is NO_x mass NO₂ equivalent or nitrogen mass?

Response 7:

Throughout this study, all reported NO_x emission indices and total NO_x emission masses are explicitly expressed in terms of NO₂ mass equivalent, strictly adhering to the standard conventions of ICAO Annex 16 Volume II and the ICAO Engine Emissions Databank (EEDB). To ensure there is no ambiguity for readers, we have added explicit clarifications to the manuscript text to formally define this metric.

In Section 3.1, the text was revised to explicitly state the NO₂ equivalent metric alongside the reported totals, reading as follows:

"Under this system boundary, total annual fuel consumption is estimated at 23.7 Mt, corresponding to total emissions of 74.6 Mt for CO₂, 311.84 kt for NO_x (reported as NO₂ mass equivalent), 73.5 kt for CO, 6.02 kt for HC, 23.69 kt for SO₂, and 2.00 kt for PM."

Comment 8:

8. The dataset currently provided is incomplete. As also raised by the editor, the authors should either revise the manuscript to clearly reflect the actual scope of the available dataset or provide the complete dataset described in the manuscript.

Response 8:

We fully agree that the originally provided two-dimensional GeoTIFF files did not adequately represent the full four-dimensional structure of the inventory described in the manuscript. To rectify this, we have comprehensively reorganized the data products and released the complete 4-D gridded emission dataset, incorporating longitude, latitude, altitude, and temporal dimensions, via a finalized Zenodo record. Furthermore, we expanded the methodological documentation to explicitly detail how the high-frequency QAR observations are mathematically aggregated into this four-dimensional spatiotemporal grid.

In Section 2.3.4, the aggregation methodology was expanded to transparently define the four-dimensional grid structure, reading as follows:

"The second-by-second emission records were subsequently aggregated into a four-dimensional spatiotemporal emission grid. The horizontal domain was discretized at a resolution of $0.1^{\circ} \times 0.1^{\circ}$, while the vertical dimension was divided into 1000-m altitude layers. The temporal dimension was aggregated by calendar month. For each 1-Hz QAR observation, the calculated emission mass was assigned to the corresponding horizontal grid cell, altitude layer, and calendar month based on the actual aircraft longitude, latitude, and pressure altitude. Emissions from all trajectory observations falling within the same spatiotemporal grid cell were then summed.

The resulting dataset contains monthly four-dimensional emission fields for CO₂, SO₂, NO_x, PM, CO, and HC. This bottom-up aggregation retains the spatial and temporal distribution of individual flight trajectories while converting the high-frequency emission estimates into a standardized 4D emission inventory."

In Section 4, the Data Availability statement was updated to reflect the finalized 4-D dataset and the permanent repository link, reading as follows:

"The high-resolution 4D gridded aviation emission inventory dataset generated in this study is freely available at Zenodo via <https://doi.org/10.5281/zenodo.23334964> (Lu et al., 2026). The dataset is provided in NetCDF format with a horizontal spatial resolution of $0.1^{\circ} \times 0.1^{\circ}$ and a vertical resolution of 1000 m, including emission fields for CO₂, NO_x, SO₂, PM, CO, and HC."

Minor

Comment 9:

1. There are few instances throughout the manuscript where spaces are either missing or duplicated (maybe more than indicated below, please check!)

a. Space

Line 90, 100, 172, 295, 440, 467

b. Double space

Line 355, 442, 553

Response 9:

We have carefully proofread the entire revised manuscript and eliminated all instances of missing or duplicated spaces, including but not limited to the specific locations pointed out by the reviewer.

Comment 10:

2. Table 3 (also there is no space here)

Font size are different within the same table.

Table 4 -> no space here again in Emission Index(NO_x/others)

Response 10:

We have unified the font sizes and corrected the spacing issues in Table 3 to ensure a standardized presentation. Regarding the original Table 4, please note that as part of the fundamental restructuring of our emission calculation methodology (specifically the revised emission-index parameterization described in Section 2.3.1), this table has been removed from the revised manuscript.

Furthermore, we have conducted a comprehensive proofreading of all remaining tables, figures, and captions across the entire manuscript to eliminate any instances of missing or duplicated spaces and to ensure rigorous academic consistency.

Comment 11:

3. As a reader, I found that there are several points where the terminology or intended reference has to be inferred from the surrounding context. I encourage the authors to review the manuscript with this in mind and define terms and references more explicitly where needed. For example, the "typical flight" denoted in the caption of Figure 5 presumably refers to the B737-800 flight described in Line 422.

Response 11:

We fully agree that readers should not have to infer intended references from the surrounding context. We have thoroughly reviewed the manuscript and made all ambiguous terms explicitly clear. Specifically, the vague reference to "a typical flight" has been updated to explicitly identify the aircraft model, ensuring rigorous and unambiguous presentation.

In Section 3.3, the text and corresponding figure caption were updated to explicitly identify the aircraft type, reading as follows:

"To examine the temporal variations in aircraft emissions under changing flight conditions, a representative B737-800 flight was selected to construct a second-by-second emission profile (Figure 5)."

"Figure 5: High-frequency emission profiles of a representative Boeing 737-800 flight."

Reply to Editor

The manuscript repeatedly describes the product as a four-dimensional aviation emission inventory, and the Data Availability section states that the dataset is provided as GeoTIFF files at $0.1^\circ \times 0.1^\circ$ resolution for CO₂, NO_x, SO₂, PM, CO, and HC. However, based on my inspection of the files currently available on Zenodo, the GeoTIFF files appear to contain only two-dimensional spatial fields, rather than explicit temporal, vertical, or phase-resolved dimensions.

Could the authors please clarify whether this is intentional? If the full four-dimensional dataset is available through additional files, layers, metadata, or a separate data structure, these should be clearly described in the Data Availability section and accompanying documentation. If the public dataset only provides 2-D gridded outputs, the authors should revise the terminology throughout the manuscript to avoid overstating the dimensionality of the released dataset. Given that the "4-D" nature of the dataset is a central claim of the manuscript, this issue is important for

assessing the dataset's completeness, reproducibility, and usability.

Response:

We fully agree that the originally provided two-dimensional GeoTIFF files did not adequately represent the full four-dimensional structure of the inventory described in the manuscript. To rectify this, we have comprehensively reorganized the data products and released the complete 4-D gridded emission dataset, incorporating longitude, latitude, altitude, and temporal dimensions, via a finalized Zenodo record. Furthermore, we expanded the methodological documentation to explicitly detail how the high-frequency QAR observations are mathematically aggregated into this four-dimensional spatiotemporal grid.

In Section 2.3.4, the aggregation methodology was expanded to transparently define the four-dimensional grid structure, reading as follows:

"The second-by-second emission records were subsequently aggregated into a four-dimensional spatiotemporal emission grid. The horizontal domain was discretized at a resolution of $0.1^{\circ} \times 0.1^{\circ}$, while the vertical dimension was divided into 1000-m altitude layers. The temporal dimension was aggregated by calendar month. For each 1-Hz QAR observation, the calculated emission mass was assigned to the corresponding horizontal grid cell, altitude layer, and calendar month based on the actual aircraft longitude, latitude, and pressure altitude. Emissions from all trajectory observations falling within the same spatiotemporal grid cell were then summed.

The resulting dataset contains monthly four-dimensional emission fields for CO₂, SO₂, NO_x, PM, CO, and HC. This bottom-up aggregation retains the spatial and temporal distribution of individual flight trajectories while converting the high-frequency emission estimates into a standardized 4D emission inventory."

In Section 4, the Data Availability statement was updated to reflect the finalized 4-D dataset and the permanent repository link, reading as follows:

"The high-resolution 4D gridded aviation emission inventory dataset generated in this study is freely available at Zenodo via <https://doi.org/10.5281/zenodo.23334964> (Lu et al., 2026). The dataset is provided in NetCDF format with a horizontal spatial resolution of $0.1^{\circ} \times 0.1^{\circ}$ and a

vertical resolution of 1000 m, including emission fields for CO₂, NO_x, SO₂, PM, CO, and HC."