Point-by-point response to the editor's comments

(original comments in black, our responses in blue, changes in manuscript in green)

We thank the editor for their thorough and constructive assessment of our revised manuscript. We provide a point-by-point response to each of the three required revisions, along with descriptions of all corresponding changes made to the manuscript text and supplementary materials.

1.Non-linearity in Aerosol Optical Properties (Reviewer #3, Point 1)

The Issue:
The authors rely on a scaling factor (Equation 7) to close the mass budget for source attribution. While this is a valid tool for PM2.5 mass, Reviewer #3 correctly identified that small changes in emissions (particularly BC and sulfate) yield highly non-linear changes in extinction (AOD) due to mixing state and hygroscopicity (Cohen and Wang, 2013). The current response focuses solely on mass closure, but does not address the point that the reviewer made about optical discrepancy.

Required Action:
1. The manuscript must include a quantitative discussion of how a change in AOD relates to change in PM2.5 mass derived from the 20% perturbation experiments. How much AOD response is observed from a 20% cut in emissions? I believe that you can use your existing model runs to address this issue.
2. If this ratio is non-linear (as expected from, e.g., Wang et al., 2021), the authors must add a limitation stating that the linear scaling factor applied to source-attributed PM2.5 mass cannot be directly extrapolated to represent source-attributed AOD biases. This limitation must be noted in Section 5 (Uncertainty and Limitation).

We thank the editor for raising this point and appreciate the opportunity to clarify it.

We note that our study does not perform source attribution of AOD, nor does it apply the scaling factor to AOD. However, we acknowledge that our Supplementary Information includes a qualitative discussion attributing the model's AOD underestimation primarily to missing coarse PM mass from unaccounted emission sources (e.g., construction and industrial fugitive dust). We recognize that this discussion implicitly assumes an approximately linear response of column-integrated AOD to near-surface emission changes, and should therefore be interpreted with caution.

In response to Action 1, we have compared the surface $PM_{2.5}$ and column AOD responses from our 20% versus 100% emission reduction experiments for the residential sector. We use the residential sector as an illustrative case because it is one of the dominant sources of BC, for which, as the editor noted, small changes in emissions

can yield highly non-linear changes in extinction. At the national scale, surface $PM_{2.5}$ concentrations decrease by 1.04 $\mu g/m^3$ and 5.6 $\mu g/m^3$ under the 20% and 100% residential reduction scenarios respectively, yielding a ratio of 5.4, which is close to the expected linear scaling factor of 5. In contrast, AOD decreases by 0.04 and 0.06 under the same set of scenarios, yielding a ratio of 1.5, far below the linear expectation of 5. This confirms Reviewer #3's point regarding the non-linearity of AOD response to emission perturbations, and supports our methodological choice to apply the scaling factor only for surface $PM_{2.5}$ mass closure rather than extrapolating it to AOD.

This non-linearity in AOD arises from multiple factors, including (1) the non-linear relationship between aerosol mass and extinction efficiency due to varying aerosol source types and core-shell structures (Wang et al., 2021), (2) the non-linear relationship between surface aerosol mass and column-integrated AOD due to varying vertical aerosol profiles and aerosol composition (Zhu et al., 2024), and (3) responses of natural dust emissions to perturbed meteorology resulting from aerosol emission reduction (as WRF-Chem is a fully coupled chemistry–meteorology model), which in turn affects column AOD but not necessarily surface $PM_{2.5}$ concentrations.

In response to Action 2, we have added the following paragraph in Section 5 (Uncertainty and Limitation):

"Finally, the qualitative attribution of AOD underestimation to missing coarse PM mass in the **Supplementary Information 2.2** should be interpreted as a diagnostic discussion of likely bias sources, not as a quantitative source apportionment of AOD. Given the non-linear relationship between surface aerosol mass and column-integrated AOD resulting from variations in aerosol composition and vertical profiles (Wang et al., 2021; Zhu et al., 2024), the mass-based surface $PM_{2.5}$ source attribution results presented in this study should not be directly extrapolated to infer source contributions to AOD or AOD bias."

2. Spatio-Temporal Validation Standards (Reviewer #3, Point 3)

The Issue:
The current analysis and evaluation use large-area daily means (e.g., Figures 3e, 3g) and annual spatial scatter plots (Figure 3d). As noted, and as established in model evaluation protocols (Cohen and Prinn, 2011; Solazzo & Galmarini, 2016), this approach fails to diagnose whether the model is correctly placing pollution in the specific grid cell on the specific day. While high correlation in the IGP daily mean (R=0.93) is a necessary condition, it is also insufficient to demonstrate model skill.

Required Action:
1. Figure 3d (or a new Supplementary Figure) must be made. Please include a density scatter plot of daily grid-cell/Obs pairs (i.e., Model vs. Obs for all valid days and all grid cells combined on a single plot). This reveals the actual model skill at native resolution.

2. A map of the grid-level temporal Pearson correlation (R) for daily PM2.5 must be provided. This will reveal the spatial boundaries of where the model successfully tracks the timing of pollution events, versus where it merely reproduces climatology.
3. The text in Section 3.2 must be revised: at the present time the claim that "emission magnitude is reasonable" is only based on annual/monthly performance. This is not physically supported given the short aerosol lifetime over India (e.g., David et al., 2018). The text should instead note something along the lines that - while annual budgets are captured, the model's inability to reproduce the magnitude of extreme daily events (Fig. S2) indicates that the temporal allocation (diurnal/daily profiles) of the emission inventory remains a source of significant uncertainty.

We agree that higher-resolution validation is necessary to demonstrate model skill beyond aggregated metrics, and we appreciate this suggestion to supplement our current model evaluation.

- In response to Action 1, we have added a density scatter plot of daily grid-cell model–observation pairs as new Supplementary Figure S2 (attached below). The density scatter plot of all daily grid-level model-observation pairs yields an overall R of 0.7.
- In response to Action 2, we have included a map of grid-level temporal Pearson correlation for daily $PM_{2.5}$ in Supplementary Figure S2 (attached below). Temporal correlations are generally highest ($R > 0.8$) across the Indo-Gangetic Plain (IGP) and parts of central India, while lower correlations are found at scattered sites in southern and western India.
- We changed the Main Text to add this sentence in the caption of Figure 3: "See **Supplementary Figure 2** for a scatter plot of daily modeled versus observed $PM_{2.5}$ for all valid grid-day pairs and a map of grid-level temporal Pearson correlation coefficients (R) for daily $PM_{2.5}$."
- In response to Action 3, we have added a new paragraph in Section 5 (Uncertainty and Limitations): "Our baseline emission inventory has limitations in capturing temporal variations of real-world emission patterns. While annual emission budgets at the regional scale are reasonably represented (**Figure 3 a-b**), the model cannot fully reproduce extreme daily episodes (**Figure 3 e-g** and **Supplementary Figure 3**). This stems from how temporal allocations of emissions are handled in WRF-Chem: for anthropogenic sources (industry, residential, power, transport, and agriculture), monthly totals are distributed evenly across days with prescribed sectoral diurnal profiles, smoothing out episodic spikes. For biomass burning, we use the satellite-based FINN dataset at daily resolution, which better captures day-to-day variability; however, satellite-derived inventories can still underestimate emissions from small-scale fires or be degraded by cloud cover and thick haze during the most intense pollution episodes. Natural dust and biogenic emissions are calculated online within WRF-Chem using real-time meteorology, but carry uncertainties from static input parameters (e.g., land use type, surface erodibility, leaf area index) and biases in simulated meteorological fields (e.g., wind speed)."

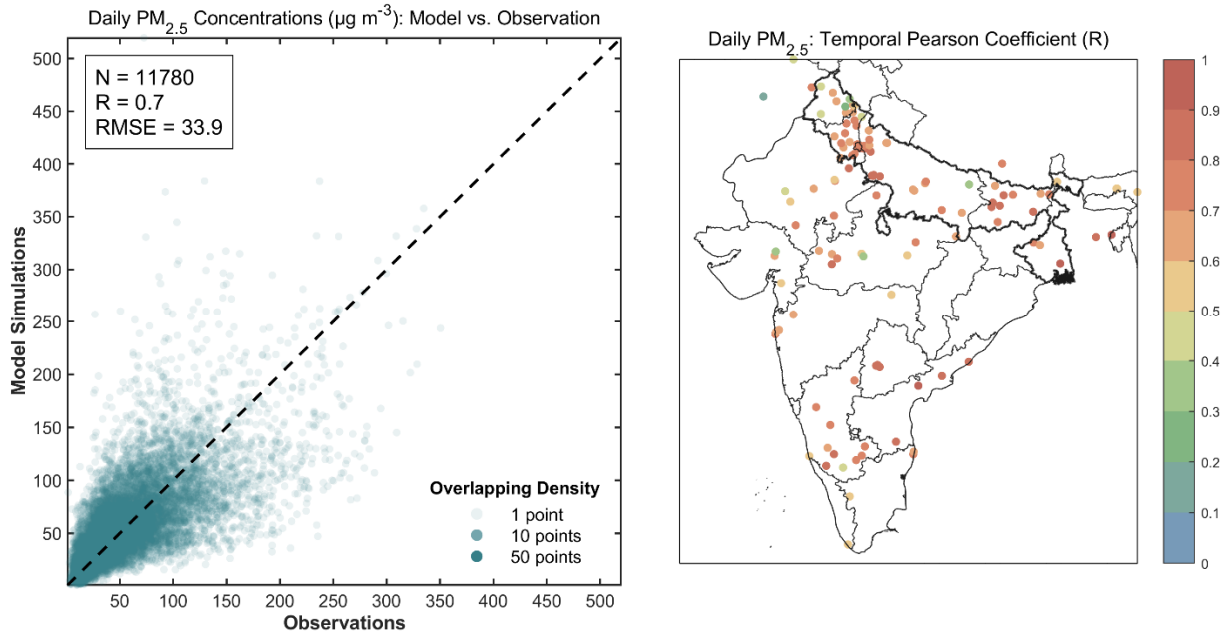

Figure S2. Daily grid-cell-level model evaluation of surface $PM_{2.5}$ concentrations (μg/m^3) for 2022. (a) Density scatter plot of daily modeled versus observed $PM_{2.5}$ for all valid grid-day pairs (N = 11,780), with the dashed line indicating the 1:1 reference. (b) Map of grid-level temporal Pearson correlation coefficients (R) for daily $PM_{2.5}$, where each dot is colored by R value. Thick black lines represent the boundary of the Indo-Gangetic Plain (IGP). Lower temporal R in northwest India likely stem from model bias in simulating natural dust emissions.

3. Seasonal Representation Bias (Reviewer #3, Point 4)

The Issue:

The selection of January for Winter and April for Pre-Monsoon deliberately excludes months with peak biomass burning and aerosol loading over your model region (e.g., February-March for fires in Myanmar and December for BC in Bangladesh) respectively.

Required Action:

1. The authors must add text to Section 5 (Uncertainty and Limitation) acknowledging this selection bias. A possible text could be: "We further acknowledge that using single months (January, April) to represent entire seasons may lead to an underestimate of the contribution of transboundary biomass burning and regional haze, particularly as peaks in AOD and BC loading in the eastern IGP and neighboring countries are observed in February-March and December respectively (i.e., Sembhi et al., 2020; Shaik et al., 2019). Consequently, our annual estimates for Organic Aerosol and Black Carbon from these episodic sources should be considered conservative lower bounds."

We acknowledge the limitation of representing seasons by single months. We focus the discussion on the pre-

monsoon season as it exhibits the largest intra-seasonal variability in fire emissions, while biases in other seasons are smaller and vary in direction. To address the uncertainty, we added the following text to Section 5 (Uncertainty and Limitation):

"In addition, using single months (January, April, July, and October) to represent entire seasons may result in biases in air quality impacts of emission sources that exhibit large intra-seasonal variability. For example, biomass burning emissions in 2022 exhibit the largest intra-seasonal variability during the pre-monsoon season, peaking in April in the IGP but in March in other regions (based on FINN v2.5.1). Selecting April to represent the pre-monsoon season on average underestimates national seasonal mean fire emissions by ~30%, with regional bias ranging from −70% (northeast India) to +50% (IGP). As a result, our estimates of biomass burning-related $PM_{2.5}$ (specifically organic carbon and black carbon) concentrations during the 2022 pre-monsoon season should be considered conservative lower bounds at the national scale, while an upper bound in the IGP."

Reference:

[1] Wang, S., Wang, X. Y., Cohen, J. B., and Qin, K.: Inferring Polluted Asian Absorbing Aerosol Properties Using Decadal Scale AERONET Measurements and a MIE Model, Geophysical Research Letters, 48, 2021.

[2] Zhu, H., Martin, R. V., van Donkelaar, A., Hammer, M. S., Li, C., Meng, J., et al.: Importance of aerosol composition and aerosol vertical profiles in global spatial variation in the relationship between PM2.5 and aerosol optical depth, Atmospheric Chemistry and Physics, 24, 11565-11584, 2024.