

## Review #2

The authors developed a unified deep learning framework to generate a high-spatial-resolution, full-coverage dataset of six major air pollutants across China for the period 2019–2022, which provides a systematic view of how different air pollutants, and their combined index, AAQI, responded during periods of rapid emission reductions and the recovery phase. I think this work is meaningful, as it offers valuable insights into underlying recovery patterns among pollutants, as well as the key drivers behind the transformations in PM<sub>2.5</sub> and O<sub>3</sub> levels. These results can provide important scientific evidence and data support for future air quality management and policy design.

However, I have several major concerns, particularly regarding the consistency and reliability of the reconstructed historical predictions, which are critical to the robustness of the study. Additional minor comments are also provided below.

### Major comments:

1. The input variables used for SHAP originate from multiple datasets (Line265-270: CHAP, CAMS reanalysis, ERA5). Do these datasets have different spatial resolutions? If so, I am concerned that directly applying them without harmonization could introduce substantial bias into the 3.5 section results. Were all inputs resampled to a common resolution before analysis? If yes, please detail the preprocessing steps in the Methods section. Otherwise, it may be necessary to re-run the experiments with consistent spatial resolution to ensure robustness.

**Response:** We agree and have harmonized all input datasets prior to the SHAP analysis. Specifically, auxiliary datasets with coarser spatial resolutions (e.g., CAMS reanalysis and ERA5 meteorological data) were resampled to a common 1 km × 1 km grid using bilinear interpolation to match the CHAP pollutant maps. This has been clarified in the revised Methods section.

2. The author repeatedly refers to population-weighted concentrations (e.g., Lines 568, 581), yet neither the main text nor the supplementary materials provide the corresponding formula or weighting methodology. The explicitly define the calculation approach are needed, and ensure that all derived metrics are clearly documented. In addition, specify the population dataset used, including its reference year. Given the known uncertainties in population data, please clarify whether any harmonization or bias correction was applied?

**Response:** We have clarified the calculation of population-weighted concentrations in the revised manuscript. Specifically, the population-weighted concentration ( $PW_{r,p}$ ) for pollutant  $p$  in region  $r$  is defined as:

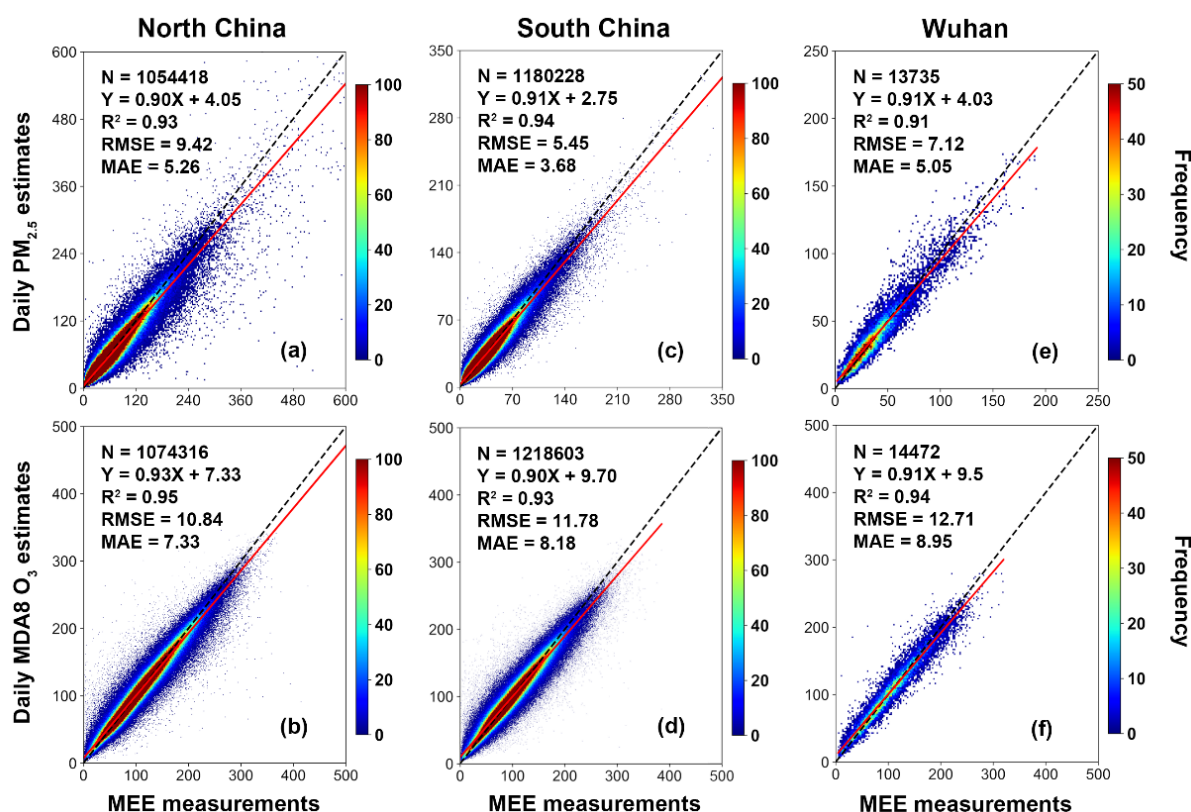
$$PW_{r,p} = \frac{\sum_{i \in c} (C_{i,p} \times P_i)}{\sum_{i \in c} P_i}, \quad (11)$$

where  $C_{i,p}$  is the pollutant concentration  $P_i$  is the population in grid cell  $i$ . Population data are derived from the LandScan™ dataset and adjusted to match United Nations annual total population estimates.

3. The discussion emphasizes the case of Wuhan, however, the validation appears to be conducted primarily at the national scale, which remains unclear whether the dataset maintains its reliability at finer spatial scales, particularly for urban areas such as Wuhan. I believe a dedicated validation specifically for Wuhan is necessary to demonstrate the dataset’s capability to support high-precision, city-scale air quality management. Similarly, since the manuscript discusses differences between northern and southern regions, region-specific validation should also be conducted to confirm the dataset’s applicability across these contrasting environments.

**Response:** We agree and have conducted additional cross-validation at both regional (Northern and Southern China) and urban (Wuhan) scales for PM<sub>2.5</sub> and O<sub>3</sub> in the revision as follows:

“At the regional scale (Fig. S5), both pollutants show strong performance, with sample-based CV-R<sup>2</sup> (slope) exceeding 0.93 (0.90) in both Northern and Southern China, and RMSE (MAE) ranges from 5.45–9.42 (3.68–5.26)  $\mu\text{g m}^{-3}$  for PM<sub>2.5</sub> and 10.84–11.78 (7.33–8.18)  $\mu\text{g m}^{-3}$  for O<sub>3</sub>. At the city scale, Wuhan also demonstrates high accuracy, with CV-R<sup>2</sup> of 0.91 (slope = 0.91) for PM<sub>2.5</sub> and 0.94 (0.91) for O<sub>3</sub>, along with average RMSE (MAE) values of 7.12 (5.05) and 12.71 (8.95)  $\mu\text{g m}^{-3}$ , respectively. These results confirm the model’s robustness across spatial scales, including its applicability to fine-scale urban analyses.”



**Figure S5.** Density scatter plots of satellite-derived daily concentrations of (a–c) PM<sub>2.5</sub> ( $\mu\text{g m}^{-3}$ ) and (d–f) O<sub>3</sub> ( $\mu\text{g m}^{-3}$ ) from the unified four-dimensional spatiotemporal deep forest (4D-STDF) model against ground-based measurements in Northern China (left), Southern China (middle), and Wuhan (right), based on sample-based cross-validation. The black dashed lines represent the 1:1 reference lines, and the red lines indicate the fitted linear regression lines.

**Minor comments:**

1. There are several instances of awkward phrasing and incomplete expressions, like in Line 311, "... can effectively **estimate**?". It is recommended that the manuscript undergo professional language editing by a native English speaker or editing service.

**Response:** We have corrected them throughout the revised manuscript. In addition, the manuscript has been edited and polished by a native English speaker.

2. There are inconsistencies between American and British spelling, for example, "modeling" in Section 2.1 versus "modelling" in Table S1. Please standardize spelling throughout the manuscript.

**Response:** We have standardized the spelling throughout the manuscript.

3. The subscripts are incorrectly formatted (e.g., Line 571: "NO<sub>x</sub>"; Line 664: "NO<sub>2</sub>"). Please ensure proper chemical notation throughout the manuscript. Similar issues are present in Figure S9 (e.g., "5 PM<sub>10</sub>" in the caption)

**Response:** We apologize for these formatting issues. We have corrected the subscripts throughout the manuscript and supplementary materials.

4. In Table S1 of the Supplement, PM<sub>10</sub> is reported in  $\mu\text{g}/\text{m}^3$ , whereas the main text uses  $\mu\text{g m}^{-3}$ . Please standardize all units according to journal guidelines.

**Response:** We have corrected and double-checked the units throughout the manuscript and supplementary materials to ensure consistency with the journal guidelines.