**Response to reviewers:**

We sincerely thank the reviewers for the time and efforts in reviewing our manuscript and for providing constructive feedback to improve our manuscript. We have revised the manuscript following the suggestions. Below we provide the point-by-point responses with the original comments in black and our responses in blue.

**RC 1:**

This paper focuses on the development of a Unified Inputs for WRF-Chem (UIWRF-Chem) system to support the MAIA satellite mission. The authors propose a framework that integrates NASA's GEOS-FP and MERRA-2 as initial and boundary conditions, incorporates a stand-alone emissions preprocessor, updates land surface properties, and implements a new $NO_x$ emission scheme. They test the system's performance across four MAIA target cities.

Overall, this is a solid and technically sound study. The authors demonstrate a good understanding of the different options available in WRF-Chem and the key differences among them. However, the manuscript currently suffers from a lack of clarity, particularly in the Introduction and Model Description sections. I recommend major revisions before it can be considered for publication.

Response: We thank the reviewer for taking the time to review our manuscript and we truly appreciate the efforts. We are also grateful for the valuable feedback to improve our manuscript.

General Comments:

1. Clarify the link to MAIA mission needs. Since the core purpose of this study is to support the MAIA mission, the paper should better articulate MAIA's specific modeling requirements—e.g., what variables are most relevant, what forecast capabilities are needed, and how UIWRF-Chem is designed to meet those needs.

   Response: Thanks for the valuable feedback and we agree it would be beneficial to clarify the link to the MAIA mission needs. We have added Sect 2.1 to describe the MAIA PM products and what are needed from UI-WRF-Chem to generate the PM products. We have then added Sect 2.2 to provide an overview of the UI-WRF-Chem modeling framework and how it is designed to meet those needs. Below shows Sect 2.1 and please find Sect 2.2 in our response under Specific Comments #3.

   2.1 Overview of MAIA PM products

   *The MAIA PM products to be generated in the PTAs include a Level 2 (L2) PM product and a Level 4 (L4) Gap-Filled PM (GFPM) product. Both L2 and L4 PM products include 24-hr averaged total and speciated PM mass concentration with a spatial resolution of 1 km within bounding boxes measuring 360 km x 480 km (east-west x north-south) size. The L2 PM data are only available for days corresponding to MAIA satellite overpasses (typically 3–4 times per week in the PTAs) at locations with valid MAIA*

*aerosol retrievals. The L4 PM data merge L2 satellite-derived PM concentration with bias-corrected PM concentrations from UI-WRF-Chem outputs and are therefore spatially (covering the whole target area) and temporally (daily) "complete". The L2 PM product is derived using GRMs which take the satellite retrieved aerosol parameters, meteorological variables and total and speciated PM concentrations from UI-WRF-Chem and other ancillary information such as population density data as predictors and surface observations of total and speciated PM concentrations as target variables. GRMs are trained for each PM type and each PTA. For the launch-ready version of the GRMs, four meteorological variables from UI-WRF-Chem are used: 2 m air temperature, 10 m wind speed, surface relative humidity (RH) and planetary boundary layer height (PBLH). To generate the L4 GFPM product, separately trained GRMs are employed to generate a bias-corrected, CTM-based PM product where the primary predictor is the CTM-generated PM concentration, rather than the satellite-retrieved aerosol optical depth. Other predictors and target variables are the same as those used in the generation of L2 PM product. For areas where both satellite-derived L2 PM and CTM-based PM products are available, these two products are then combined using weights derived from a Bayesian Ensemble Averaging model to generate the final L4 GFPM product. More detailed information can be found in Jin et al. (2024).*

*Two versions of the MAIA L2 PM and L4 GFPM products will be generated as part of the routine processing: the "forecast" and the "reanalysis" version. For the forecast product version, GEOS FP meteorology is used for model initial and boundary conditions and GEOS FP fields of aerosols and aerosol precursors will also be used to specify boundary conditions of atmospheric composition. The reanalysis versions replace GEOS FP variables with outputs from MERRA-2 data. Due to the ~ 6 month latency of speciated $PM_{2.5}$ data from surface monitors, the forecast versions will rely on previously available measurements. Generation of the reanalysis products will nominally occur on an annual basis and will benefit from more complete surface monitor datasets. More detailed information about the PM products can be found at https://maia.jpl.nasa.gov/resources/data-and-applications/.*

2. Improve the Introduction. The rationale for modifying WRF-Chem is not clearly laid out. The authors should explain why it is necessary to update land surface properties, emissions modules, and boundary/initial conditions in the context of MAIA. A clearer articulation of these needs would better frame the scientific motivation.

   Response: Thanks for the suggestion. We have revised the introduction to clarify the role of UI-WRF-Chem in the MAIA satellite mission and the motivation for our major updates in UI-WRF-Chem. Please see the updated manuscript for the revised introduction.

3. Reorganize the Model Description section. The current presentation of model improvements is confusing. I suggest breaking it into clearly labeled subsections, each focused on a single enhancement (e.g., emissions, boundary conditions, $NO_x$ scheme, land surface update).

Response: Thanks for the suggestion to improve the Model Description Section. We agree clarification is needed and have reorganized the session following the suggestion. We first added Sect 2.1 to provide an overview of the MAIA PM products and identify variables required from UI-WRF-Chem outputs to generate these products. We next added Sect 2.2 to provide an overview of the UI-WRF-Chem modeling framework, explain the motivation of the model updates and highlight the novelty of developments in the current work. We then have each of the model improvement as a subsection:

2.3 Updates of meteorological and chemical initial and boundary conditions as well as soil properties;

2.4 Updates of land surface properties;

2.5 Development of BDISNP soil NOx emission scheme;

2.6 Development of WRF-Chem Emission Preprocessing System (WEPS);

2.7 Updates of WRF-Chem Chemistry scheme;

2.8 Postprocessing and evaluation codes, and repository management.

4. Quantify significance of improvements. While the paper compares results from different modeling schemes, it does not provide evidence of whether the differences are statistically significant or robust across other regions or time periods.

   Response: Thank you for the suggestion. We agree that it is important to assess whether the differences are statistically significant. We have added Sect 3.1 to describe the methods used for significance testing and apply the tests for case studies conducted over Beijing, Rome, Los Angeles, and Atlanta target areas. Please see our detailed response under Specific Comments #10.

Specific Comments:

1. Why use WRF-Chem v3.8.1? Given that WRF-Chem versions above 4.0 are now available (with improvements such as subgrid-scale chemical transport for KF and GF schemes), the authors should justify why UIWRF-Chem is based on v3.8.1. Even though they mention plans to test GF in the future, a more detailed explanation is needed.

   Response: Thanks for bringing up this important point. The model development presented in this paper spanned several years, during which WRF-Chem v3.8.1 was selected as the base version due to its stability and widespread use at the time. While the newer versions of WRF-Chem (v4.0 or above) include updates relevant to this work such as the subgrid-scale chemical transport using the KF or GF scheme, we have kept the WRF-Chem v3.8.1 throughout the development to ensure consistency and reproducibility of the results. We also note that only the GF scheme can ensure the consistency of the transport of the chemical species and other scalars. Although the KF scheme is widely used and advanced, it has not yet been updated to support the consistent transport of chemical species and other scalars using the same scheme. Nevertheless, we

acknowledge the limitation of using an older version and recognize the benefits introduced in the newer versions. We plan to update the UI-WRF-Chem system with a newer version to incorporate these improvements in the future work.

We have made some revisions and added one paragraph in Sect 3 to provide more detailed explanation as follows:

*With the current version (WRF-Chem v3.8.1) of the code, chemical species are transported using the G3D scheme, regardless of which cumulus scheme is used, while other scalars are transported with the selected cumulus scheme. Therefore, the G3D scheme is used to ensure the consistency between chemistry and physics. Additionally, WRF-Chem v3.8.1 was selected as the base version at the beginning of this project due to its stability. We have maintained this version over the course of the project to ensure the consistency and reproducibility of the results. Although there are several scale-aware cumulus schemes available in WRF-Chem such as the Kain-Fritsch scheme (KF, (Kain, 2004)) and the Grell-Freitas scheme (GF, (Grell and Freitas, 2014)), only the GF scheme has been updated to ensure the consistent transport of both chemical species and other scalars, as described by Li et al. (2018, 2019). We acknowledge the limitation of using only the G3D scheme in this work and plan to update the UI-WRF-Chem modelling framework to a newer version to enable the use of the GF scheme and incorporate other recent improvements as well.*

2. Summarize model setup in a table. Please consider adding a summary table listing the model configuration (e.g., resolution, land surface model, physics schemes, emissions setup, etc.), or update Table S2 accordingly.

Response: Thanks for the suggestion. We have added Table 1 to summarize the model configurations regarding physics, chemistry and emissions for the four target areas.

Table 1. A summary of model physics, chemistry and emissions configurations for CHN-Beijing, ITA-Rome, USA-LosAngeles, and USA-Atlanta target areas.

| Category | Model component | CHN-Beijing | ITA-Rome | USA-Los Angeles | USA-Atlanta |
|---|---|---|---|---|---|
| Physics | Microphysics | Lin | Morrison | Lin | Lin |
| | Cumulus | G3D | G3D | G3D | G3D |
| | Longwave radiation | RRTMG | RRTMG | RRTMG | RRTMG |
| | Shortwave radiation | RRTMG | RRTMG | RRTMG | RRTMG |
| | Planetary boundary layer | YSU | YSU | YSU | YSU |
| | Surface layer | Revised MM5 | | | |
| | Land surface model | NOAH | NOAH | NOAH | NOAH |
| Chemistry | Gas-phase | RADM2 | RADM2 | RADM2 | RADM2 |
| | Aerosols | MADE/SORAGM-DustSS | | | |
| | Photolysis | Madronich F-TUV | | | |

| | | | | | |
|---|---|---|---|---|---|
| Emissions | Anthropogenic emissions | MEIC 2016 | HTAP v3 (2018) | NEI 2017 | NEI 2017 |
| | Dust emissions | GOCART with AFWA modifications | | | |
| | Biogenic emissions of VOCs | MEGAN | MEGAN | MEGAN | MEGAN |
| | Soil NOx emissions | BDISNP | BDISNP | BDISNP | BDISNP |
| | Wildfire emissions | FLAMBE | FLAMBE | FLAMBE | FLAMBE |

3. Highlight novelty of new modules. Some of the newly added components appear to be simple integrations into WRF-Chem rather than innovations. The paper should more clearly highlight what is original and novel in this system.

   Response: We appreciate the suggestion and agree that highlighting the originality of our work is important. We have added section 2.2 to provide an overview of the UI-WRF-Chem modeling framework and clearly outline the novelty of the new modules we have developed. In addition, we have included a flowchart of the UI-WRF-Chem modeling framework (Fig 1) to visually emphasize these updates.

   2.2 Overview of UI-WRF-Chem modeling framework

   *To meet these needs, UI-WRF-Chem is designed to operate in both forecasting (or near real time, NRT) and reanalysis modes. We use the NASA GEOS model data: GEOS FP in forecasting or NRT mode and MERRA-2 in reanalysis mode to drive WRF-Chem simulations by providing self-consistent and unified meteorological and chemical initial and boundary conditions, referred to as the Unified Inputs (of initial and boundary conditions) for meteorology and chemistry. Figure 1 presents the flowchart of the UI-WRF-Chem modeling framework. Here, we provide a brief description of the UI-WRF-Chem framework, outline the components included in the standard WRF-Chem model and highlight the major updates we have introduced.*

   *Compared with the standard WRF-Chem model, the UI-WRF-Chem modeling framework incorporates new modules and significant modifications to enable the seamless use of NASA GEOS data, updates of land surface properties with recent available MODIS land data and expanded emission capabilities. First, we incorporate the GEOS2WRF module from NASA'S Unified-Weather Research and Forecasting model (NU-WRF) (Peters-Lidard et al., 2015), which functions similarly to the standard ungrib process, by converting GEOS FP or MERRA-2 data to an intermediate file format. We also develop the LDAS2WRF module, adapted from the GEOS2WRF module to convert the GLDAS or NLDAS data into the same intermediate file format. The standard metgrid process then converts these intermediate files into meteorological files in the NetCDF format (met_em.d*.nc), respectively. These two NetCDF files are subsequently merged to generate the final meteorological files for the real process. Second, to integrate the MODIS land data into the static geographical datasets, we develop the conv_geo Python-based module, where we convert the MODIS land data into the standard binary file formats required by the geogrid process. This enables updates of land surface properties*

*with recent available MODIS land data, not available in the standard WRF-Chem model. Additionally, we develop the GEOSBC module, by modifying the standard mozbc module to use GEOS FP or MERRA-2 data for updating both chemical initial and boundary conditions, which improves the consistency between meteorology and chemistry inputs. Additionally, we modify WRF-Chem's chemistry scheme to ensure compatibility between dust fields from GEOS FP or MERRA-2 and the dust representation in the chemistry scheme itself (see Sect 2.7 for more information).*

*For emissions, we develop the BDISNP scheme for soil NOx emissions by extending the workflow of the standard MEGAN-based biogenic VOC calculation. Same as the MEGAN process, we first use the standard bio_emiss module to read the MEGAN emission input datasets (e.g., isoprene emission factor) and then convert them into the wrfbiochemi_d0* files for the real process. We then apply the add_fert module that we have developed here to incorporate emission input datasets (e.g., fertilizer data), specific to the BDISNP scheme into wrfbiochemi_d0* files. Additionally, we modify WRF-Chem codes to calculate soil NOx emissions. We also develop the WEPS module to process both anthropogenic and fire emissions, adopting some functionalities from the widely used anthro_emiss and EPA_ANTHRO_EMISS utilities in the WRF-Chem community. This provides flexibility for incorporating additional emission inventories into the WEPS. Lastly, we develop a Python-based postprocessing module to calculate selected WRF-Chem variables and compile hourly WRF-Chem output files into daily files in the formats required by the GRMs.*

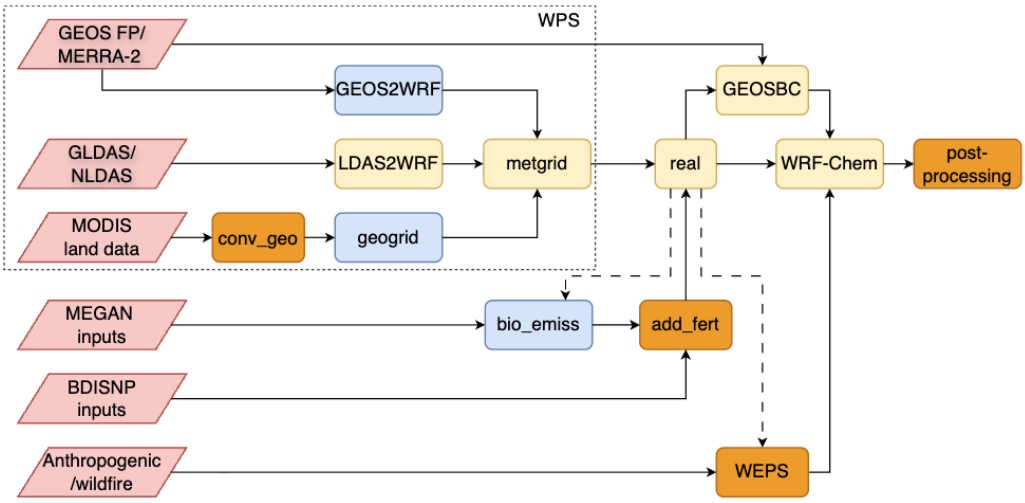


*Figure 1. Flowchart of UI-WRF-Chem modeling framework. Pink parallelogram represent input datasets used, including meteorological, land surface and emission data. Rounded rectangles represent different modules and processes within the UI-WRF-Chem framework. Blue rounded rectangles denote standard WRF-Chem components without any changes, except for GEOS2WRF, which is from NASA's NU-WRF framework. Yellow round rectangles represent modified modules based on standard WRF-Chem components, except for LDAS2WRF, which is adapted from GEOS2WRF. Orange rounded rectangles indicate new modules developed in this work. The input datasets and modules enclosed within the dashed box corresponds to the WPS in the standard WRF-Chem model, where meteorological files (met_em.d*.nc) are generated. The conv_geo process converts MODIS land data into binary files, for the geogrid process.*

*Both GEOS2WRF and LDAS2WRF convert input data in the NetCDF file format to an intermediate file format, equivalent to the ungrib process. GEOSBC is adapted from the mozbc module, where GEOS FP and MERRA-2 data are used to update chemical initial and boundary conditions. The bio_emiss module reads MEGAN emission input datasets (e.g., isoprene emission factor) and generates files (wrfbiochemi_d0*) for WRF-Chem to calculate biogenic emissions. The add_fert module is used to add the BDISNP input datasets (e.g., fertilizer data) into the wrfbiochemi_d0* files for the real process. WEPS processes both anthropogenic and fire emission datasets and converts them into WRF-Chem-ready emission files (*wrfichemi*). Dashed lines from real to bio_emiss and WEPS indicate that real needs to be executed once before running the full flow to generate wrfinput_d0* files, which provide domain information to these two modules.*

4. Table S1. Please consider highlighting the best-performing configurations for easy comparison.

   Response: Thanks for the suggestion and we have added an asterisk mark to denote the best-performing configuration for each PTA as seen in Table S1.

Table S1. A suite of UI-WRF-Chem sensitivity simulations with different options of physics schemes over CHN-Beijing, ITA-Rome, USA-LosAngeles and USA-Atlanta target areas.

| Target area | Simulation number | Microphysics | Longwave | Shortwave | PBL |
|---|---|---|---|---|---|
| Beijing | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Morrison | RRTMG | RRTMG | YSU |
| | 3 | Lin | RRTMG | RRTMG | MYJ |
| | 4 | Lin | RRTM | Goddard | YSU |
| Rome | 1 | Lin | RRTMG | RRTMG | YSU |
| | 2* | Morrison | RRTMG | RRTMG | YSU |
| | 3 | WSM6 | RRTMG | RRTMG | YSU |
| | 4 | Morrison | RRTMG | RRTMG | MYJ |
| | 5 | Morrison | RRTMG | RRTMG | MYNN2.5 |
| | 6 | Morrison | RRTM | Goddard | YSU |
| Los Angeles | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Lin | RRTMG | RRTMG | MYJ |
| | 3 | Lin | RRTM | Goddard | YSU |
| Atlanta | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Morrison | RRTMG | RRTMG | YSU |
| | 3 | WSM6 | RRTMG | RRTMG | YSU |
| | 4 | Lin | RRTMG | RRTMG | MYJ |
| | 5 | Lin | RRTMG | RRTMG | MYNN2.5 |
| | 6 | Lin | RRTM | Goddard | YSU |

*These are the final configurations selected for each target area.

5. Figure 5: The comparison may be misleading due to resolution differences—MERRA-2 is coarse and likely underestimates high $PM_{2.5}$ values, whereas WRF-Chem has higher resolution and better captures spatial variability. Consider interpolating WRF-Chem output to the MERRA-2 grid for a fair comparison or include scatterplots at matched resolution.

Response: We agree that it is a fair comparison to interpolate the WRF-Chem output to the MERRA-2 grid. We have regridded the WRF-Chem outputs into the MERRA-2 grid and average the data of the sites that fall into the same MERRA-2 grid. We have updated the figure as follows:

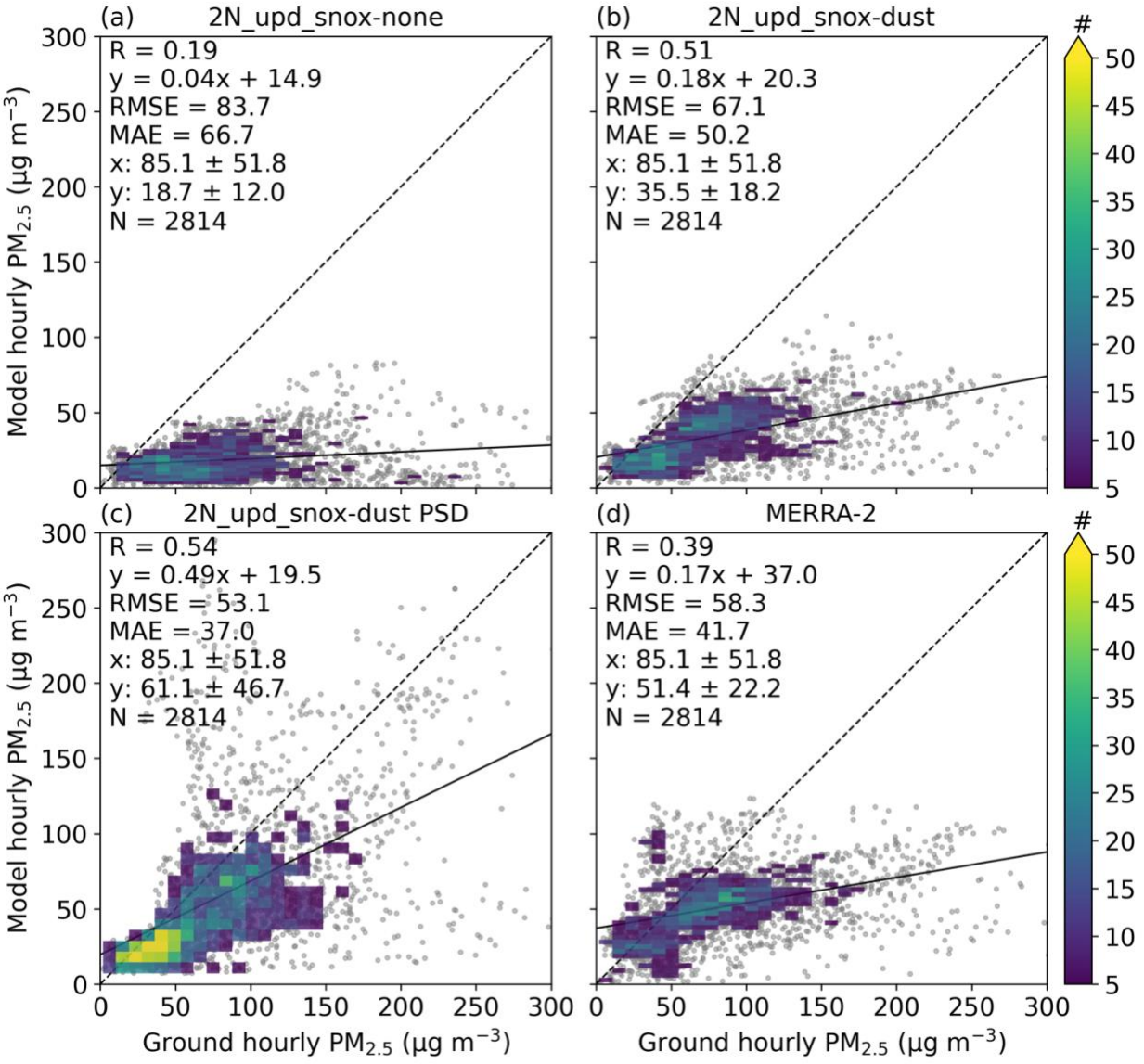


*Figure 6. Scatter plot of hourly surface $PM_{2.5}$ concentration between model (y axis) and ground observation (x axis) for surface sites in the inner domain (D2) of CHN-Beijing for 24–31 March 2018. (a)–(c) refer to the UI-WRF-Chem sensitivity simulations with different chemical boundary conditions being considered using MERRA-2 data (Table 2). (a) no chemical species, (b) dust and other aerosols and (c)*

*same as (b) except that the dust concentration is scaled based on constraining MERRA-2 dust PSD data with AERONET PSD climatology data. (d) is from MERRA-2 simulated surface $PM_{2.5}$ concentration. Also shown on the scatter plot is the correlation coefficient (R), the root-mean-square error (RMSE), the mean absolute error (MAE), the mean ± standard deviation for observed (x) and model-simulated surface $PM_{2.5}$ (y), the number of collocated data points (N), the density of points (the color bar), the best fit linear regression line (the solid black line) and the 1:1 line (the dashed black line). WRF-Chem PM data are regridded onto the MERRA-2 grid, and when multiple surface sites fall within the same MERRA-2 grid, the observations are then averaged to represent a single collocated site.*

6. Figure 6: Please emphasize the observational data (e.g., bold lines or larger markers) to improve readability.

Response: Thanks for the suggestion and we have increased the size of the markers for observational data as follows:

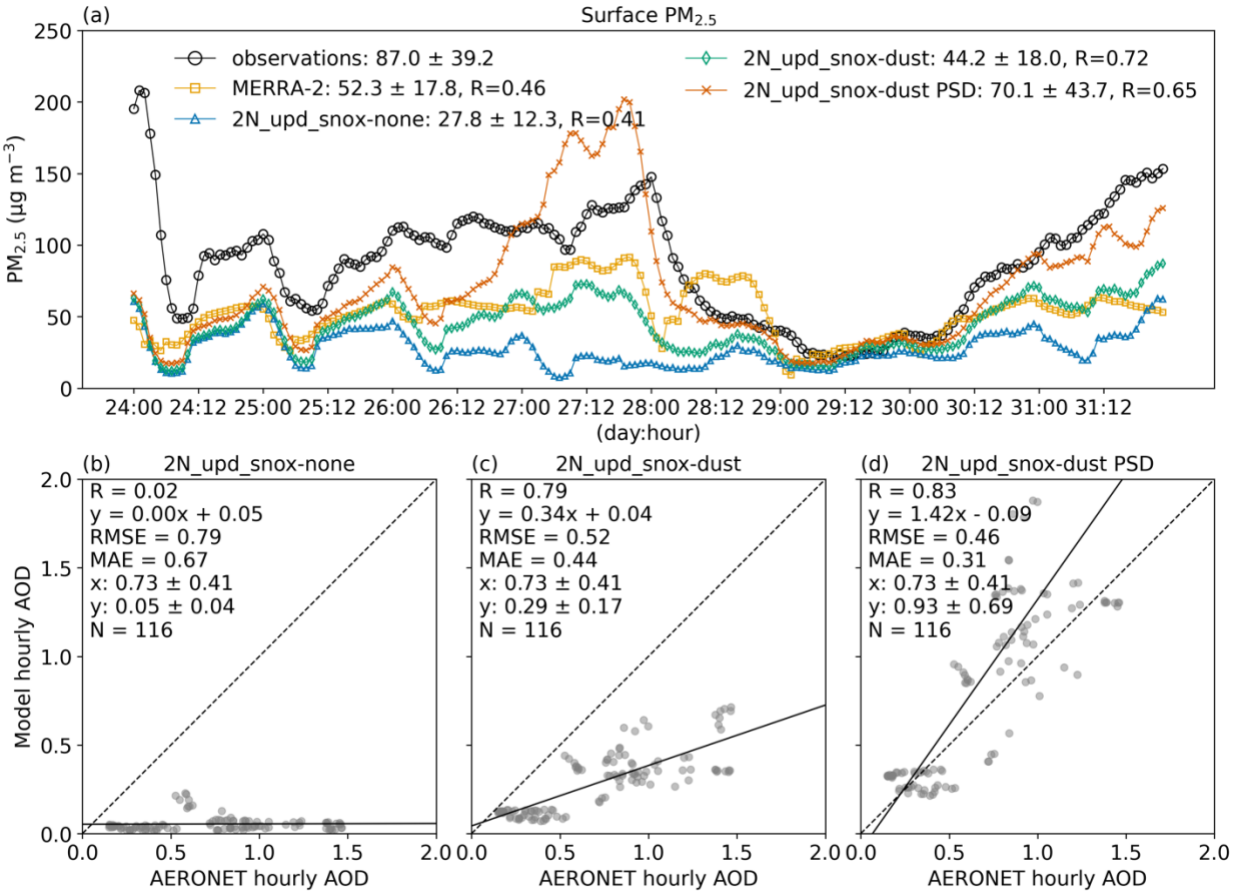


*Figure 7. (a) time series of hourly surface $PM_{2.5}$ concentration averaged over surface sites in the inner domain (D2) of CHN-Beijing for 24–31 March 2018, from model simulations and ground observations. 2N_upd_snox-none/dust/dust PSD refer to the UI-WRF-Chem sensitivity simulations with different chemical boundary conditions being considered using MERRA-2 data (Table 2): no chemical species; dust and other aerosols; dust concentration is scaled based on constraining MERRA-2 dust PSD data with AERONET PSD climatology data. Also shown on the plot is the mean ± standard deviation of surface PM2.5 for model simulations or observations as well as the correlation coefficient (R). (b)–(d): scatter plot of hourly AOD between model (y axis) and AERONET observation (x axis) for 24–31 March 2018. Also shown on the scatter plot is R, the root-mean-square error (RMSE), the mean absolute error (MAE), the*

*mean ± standard deviation for observed (x) and model-simulated AOD (y), the number of collocated data points (N), the best fit linear regression line (the solid black line) and the 1:1 line (the dashed black line).*

7. Line 720: Consider discussing why the updated system better captures the observed $PM_{2.5}$ peaks. This would strengthen the case for the model improvements.

   Response: Thanks for the suggestion and we have added discussion on why the updated system better captures the observed $PM_{2.5}$ peaks as follows:

   *Time series of UI-WRF-Chem simulated hourly speciated $PM_{2.5}$ (e.g., OC, EC, sulfate, nitrate) and dust components in both $PM_{2.5}$ and $PM_{10}$ from the two sensitivity simulations (2N_upd_snox_dust and 2N_upd_snox_dust PSD) (not shown here) indicate that only the dust components exhibit similar peaks as in the total $PM_{2.5}$ and $PM_{10}$, while other speciated $PM_{2.5}$ components do not follow the same temporal pattern. This demonstrates that the observed peaks in both $PM_{2.5}$ and $PM_{10}$ are primarily driven by the dust intrusion event. Moreover, the magnitude of the peak from the sensitivity simulation – 2N_upd_snox_dust PSD is larger and matches better with surface observations, especially for $PM_{10}$, than that of the 2N_upd_snox_dust. This further highlights the effectiveness of our method in improving the representation of dust size distribution in MERRA-2 data.*

8. Line 1035: It would be helpful to summarize the sensitivity tests in a table for easier interpretation.

   Response: Thanks for the suggestion and we have added a table (Table 3) to summarize the sensitivity tests.

Table 3. A suite of UI-WRF-Chem sensitivity simulations performed over PTA-Atlanta with different setups of microphysics and cumulus schemes for Domain 1 (D1) and Domain 2 (D2), respectively.

| | mp2cu5 | mp2cu5bothon | mp2cu3bothon | mp10cu5 | mp10cu5bothon | mp10cu3bothon |
|---|---|---|---|---|---|---|
| Microphysics-D1 | Lin | Lin | Lin | Morrison | Morrison | Morrison |
| Microphysics-D2 | Lin | Lin | Lin | Morrison | Morrison | Morrison |
| Cumulus-D1 | G3D | G3D | GF | G3D | G3D | GF |
| Cumulus-D2 | off | G3D | GF | off | G3D | GF |

9. MAIA compositional data: Since MAIA will retrieve PM component information, the paper should demonstrate how UIWRF-Chem simulates PM species. It would be useful to show comparisons against ground-based observations (e.g., from the IMPROVE network).

   Response: Thank you for bringing up this point. We agree that it is important to demonstrate how UI-WRF-Chem simulates PM species. As an example, we have evaluated the UI-WRF-Chem simulated sulfate + nitrate, EC, OC and dust against both

IMPROVE and CSN networks over the Atlanta PTA. At the time of writing, speciation data for some non-US PTAs are not yet available and we do not have sufficient modeling data to evaluate the model performance either. We have recently generated longer datasets of UI-WRF-Chem simulations for each PTA. Our ongoing work focuses on evaluating model performances for total PM, speciated $PM_{2.5}$ and AOD, which will provide a more comprehensive assessment of model performance. We have added Sect 4.4.2 to present the composition analysis as follows and the updated supplement includes Figures S17 and S18.

4.4.2 Evaluation of model simulated $PM_{2.5}$ composition

*Surface measurements of total and speciated $PM_{2.5}$ mass concentration from the Interagency Monitoring of Protected Visual Environments (IMPROVE) (Malm et al., 1994; Solomon et al., 2014) and the Chemical Speciation Network (CSN) (Solomon et al., 2014) networks (see Fig S17 for sites location information) are used to evaluate model performance. We compare UI-WRF-Chem simulated speciated $PM_{2.5}$ (OC, EC, Sulfate + Nitrate, Dust) and total $PM_{2.5}$ against these observations. Figure S18 shows the comparison of daily speciated $PM_{2.5}$ between the model and ground observations for the six different sensitivity simulations (Table 3), while Fig S17 shows the spatial distribution of total and speciated $PM_{2.5}$ for the "mp2cu5" sensitivity simulation only.*

*During this month of June, both surface observations and model simulations indicate that OC, sulfate and dust are the dominate components of total $PM_{2.5}$, consistent with previous studies, which show that OC and sulfate are the primary contributors to total $PM_{2.5}$ in the Southeastern U.S. (Hand et al., 2024; Zhu et al., 2024). Prescribed burns in the Southeastern U.S. including the states of Alabama and Georgia are a major source of OC emissions in this region (Li et al., 2023; Cummins et al., 2023), some of which are represented by the FLAMBE emission inventory in this work. All the model sensitivity simulations for OC show good correlation (0.45–0.60, Fig S18) and reasonable MB values (–1.13 to –0.36 $\mu g\ m^{-3}$). Model simulated EC concentrations also show good correlation (0.45–0.72, Fig S18) but underestimate ground observations with MB from –0.28 to –0.21 $\mu g\ m^{-3}$. For the dust component, correlation ranges from 0.42–0.73 (Fig S18) but all the model sensitivity simulations overestimate ground observations with MB from 0.42–1.6 $\mu g\ m^{-3}$. In contrast, the combined sulfate + nitrate for all the sensitivity simulations show relatively lower correlation (–0.03 to 0.23) and varying levels of MB (–0.49 to 0.01 $\mu g\ m^{-3}$). The nitrate concentration from ground observations is low in this region with an average value of 0.198 $\mu g\ m^{-3}$ for this month, which makes it challenging for the model to reproduce such a low level. Also due to limited samples used for comparison here, sulfate and nitrate are combined for evaluation. Overall, the "mp2cu5" sensitivity simulation (Table 3) yields the best performance.*

*Figure 15 shows the variability in total and speciated $PM_{2.5}$ mass concentration from model simulation mp2cu5 (Table 3) compared with surface observations as well as the ratio of model simulation to observation. The simulated-to-observed ratio for dust (1–5.8) exhibits much larger variability than other $PM_{2.5}$ components (0–2), with model simulation consistently overestimating dust. During this month, PTA-Atlanta may have*

*been affected by long-range transport of Sahara dust in the model simulation. These biases are likely due to uncertainties in the MERRA-2 simulated dust particle size distribution, as also demonstrated by the case studies over CHN-Beijing and ITA-Rome, motivating future work to tune the dust particle size distribution of MERRA-2 data for this region. Ratios for other components mostly remain below 2. Both sulfate and nitrate aerosols are predominantly secondary aerosols in the atmosphere, formed through chemical reactions and are also highly water-soluble, making them sensitive to uncertainties in the aerosol chemistry and wet deposition schemes. As discussed earlier, this month experiences some convective precipitation events, which likely contributes to the uncertainty and large variability in the simulated speciated $PM_{2.5}$ concentration.*

*Although our analysis here is limited to one month and one PTA, it provides a valuable case study of how the UI-WRF-Chem modeling framework simulates speciated $PM_{2.5}$. Moreover, previous work by Jin et al. (2024) using the same UI-WRF-Chem framework demonstrated its broader robustness over the Boston PTA. It illustrated the feasibility of the MAIA modeling framework for generating L2 and L4 PM products with a full year (2018) of UI-WRF-Chem outputs of total and speciated $PM_{2.5}$ mass concentrations and showed the correlation of evaluating model total and speciated $PM_{2.5}$ mass concentrations against ground observations ranging from 0.40 to 0.73 (Table S1 therein). Together, these results suggest that while the single-month evaluation such as the case study here only provides a partial picture of model performance, the framework has been shown to produce reliable and robust results for longer time periods. Future work will therefore focus on a more comprehensive assessment of model performance with respect to the PM composition using longer datasets across different PTAs.*

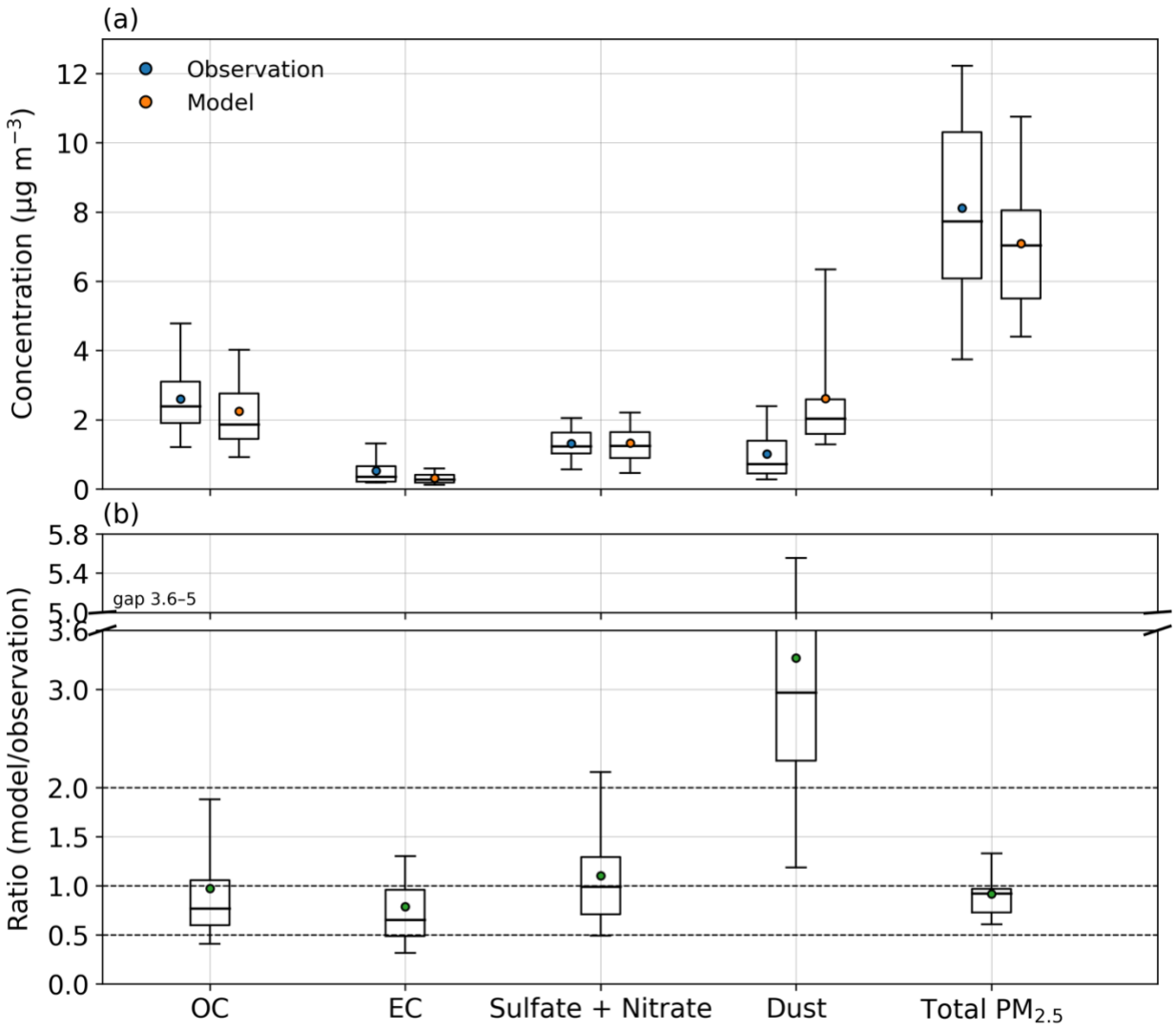


*Figure 15. Box-whisker plots of (a) total and speciated $PM_{2.5}$ concentrations from UI-WRF-Chem simulation (mp2cu5 in Table 3) and surface observations from IMPROVE and CSN sites over the inner domain (D2) of PTA-Atlanta for June 2022, and (b) the ratio of model simulated to observed $PM_{2.5}$. Speciated $PM_{2.5}$ include OC, EC, dust and the combined sulfate + nitrate. Also Shown on the boxer plot are the 5th and 95th percentiles (the whiskers), the interquartile range (the boxes), the median (the black lines) and the mean (the filled circles). Note on (b), the y-axis is truncated between 3.6–5.0 for improved visualization.*

10. Significance testing: The paper discusses improved performance for certain configurations but lacks significance tests to demonstrate that the improvements are statistically meaningful. This is important to ensure the optimal setup is not case-specific.

    Response: Thank you for the suggestion to strengthen our analysis with significance testing. We have first added Sect 3.1 *Evaluation statistics* to describe the methods used to conduct significance tests. Briefly, we apply the paired t-test and add the non-parametric Wilcoxon signed rank test when the sample size is small to ensure a robust test. In addition, we use either the Bonferroni correction procedure or the Benjamini-Hochberg

false discovery rate correction when multiple tests are considered. These tests are applied in different case studies to assess whether improvements are statistically significant, focusing on the mean absolute error (MAE) metric.

Overall, the improvements for the case studies over Beijing, Rome, Los Angeles, and Atlanta are statistically significant, indicating that the results are not random. We further emphasize that incorporating the improved dust size distribution in MERRA-2 data, constrained by AERONET observations, enhance simulations of surface PM and AOD for both Rome and Beijing. This provides additional evidence that the improvements are robust across target areas. Below is Sect 3.1 and details of the significance testing for each case study can be found in the updated manuscript.

3.1 Evaluation statistics

*Several statistics are used to evaluate the model performance against ground and satellite observations, including linear correlation coefficient (R), root mean square error (RMSE), mean bias (MB), normalized mean bias (NMB), mean absolute error (MAE), normalized standard deviation (NSD) and normalized centered root mean square error (NCRMSE). NSD is the ratio of the standard deviation of the model simulation to the standard deviation of the observation. NCRMSE is like RMSE except that the impact of the bias is removed. Some of these statistics are summarized in a Taylor Diagram (Taylor, 2001), which includes R (shown as the cosine of the polar angle), NSD (shown as the radius from the quadrant center), and NCRMSE (shown as the radius from the expected point, which is located at the point where R and NSD are unity).*

*To determine whether the performances among model sensitivity simulations for different case studies over different target areas are statistically significant, we conduct the paired t-test on collocated model-observation samples or between model simulations. We focus on the MAE as the evaluation metric. For comparison of hourly data, we account for the temporal autocorrelation by estimating the lag-1 autocorrelation and applying the effective sample size adjustment (Wilks, 2011). For cases with smaller sample size, we also apply the non-parametric Wilcoxon signed rank test (e.g., Menut et al., 2019; Tao et al., 2025) to ensure the robustness of our test. In addition, when multiple model sensitivity simulations are evaluated, we apply a Bonferroni correction procedure (SIMES, 1986) to both paired-t and Wilcoxon tests, following previous work (Crippa et al., 2017). Under this approach, the null hypothesis is rejected if* $p \leq \frac{\alpha}{m}$*, where* $p$ *is the raw* $p$ *value,* $\alpha$ *is the significance level (0.05 in this study) and m is the number of hypothesis tests. For testing the significance over spatial maps, where a large number of tests are performed simultaneously, we instead apply the Benjamini-Hochberg false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). We hence report adjusted* $p$*-value throughout this work unless noted otherwise.*

**RC2:**

Zhang et al. present UI-WRF-Chem, a set of unified inputs (initial and boundary conditions) for WRF-Chem in support of the MAIA satellite mission. UI-WRF-Chem provides meteorological inputs as well as emissions; land surface data; and a new soil NOx emissions scheme. A new chemistry scheme based on MADE/SORGAM is also developed, MADE/SORGAM-DustSS, to incorporate GOCART-AFWA emission scheme for matching with the MERRA-2/GEOS-FP dust size bins.

The manuscript details many improvements to provide inputs to WRF-Chem. Of particular note is the development of WEPS as an emissions pre-processor which resolves a point of frustration in offline emissions processing for WRF-Chem. In addition, to support the MAIA satellite mission, UI-WRF-Chem extensively incorporates data from GEOS-FP and MERRA-2 products to the WRF-Chem model pipeline and evaluates many of these developments. These improvements have also been validated in four extensive case studies across the globe. The manuscript is well written and I recommend its publication.

Response: We sincerely thank the reviewer for spending the time and effort providing valuable feedback to improve our manuscript and we truly appreciate the recognition of our work.

Major comments:

1. The authors evaluate throughout the effect of incorporating "MERRA-2 chemical boundary conditions" into the simulation. Just to confirm that the simulations marked "none" mean zero boundary conditions are input; is this usual in WRF-Chem simulations? Does the common approach of using CAM-chem/WACCM outputs as boundary conditions provide no information for the dust and other aerosols in a WRF-Chem simulation? If CAM-chem/WACCM, as global models, can provide some kind of information, I think it would be a more fair comparison as to whether these conditions can/can not help the regional model capture the long-range transport event.

Response: We thank the reviewer for this insightful comment and the opportunity to clarify. In our work, the simulations marked “none” indicate that no chemical boundary conditions from MERRA-2 are used. Instead, the model applies its default chemical boundary conditions which represent a clean North American summery day and includes a limited number of mostly gas-phase species. This default setting was originally developed for tropospheric ozone forecast. For aerosol species, the concentration values are close to zero. We have added this clarification to the caption of Table 2. We also note that the common approach of using CAM-Chem/WACCM outputs as boundary conditions does provide information for dust and other aerosols.

We agree with the reviewer that comparing MERRA-2 chemical boundary conditions to an alternative global model such as CAM-Chem/WACCM could provide a more direct assessment of how using different global models influence the representation of long-range transport. In this study, we chose the “none” case as the baseline run to directly quantify the contribution from MERRA-2 chemical boundary conditions. Our focus was on improving the representation of dust size distribution in MERRA-2. We recognize that different studies have adopted different

global models as chemical boundary conditions depending on the research scope and a comprehensive assessment of WRF-Chem's sensitivity to these choices would provide valuable insights to the community.

We have also included the following sentences in Sect 5 to clarify this point:

*Since our work mainly focuses on improving the representation of the dust size distribution in MERRA-2 data, we recognize that other global models such as CAM-Chem may also provide useful information for chemical boundary conditions in different applications. While a comprehensive understanding of how different global models affect WRF-Chem simulations of special events such as the dust long-range transport, would provide valuable insights to the community, our work here demonstrates an efficient way for improving the simulation of dust transport using WRF-Chem.*

2. UI-WRF-Chem extensively integrates outputs from GEOS-FP and MERRA-2 as inputs for WRF-Chem; many of these improvements are not trivial, e.g., the updates to land input data, a new soil NOx emissions scheme, etc... Do the authors plan to contribute this capability to WPS and the WRF mainline model code in the future?

Response: Thank you for the affirmation and bringing up this point. We are planning to update the UI-WRF-Chem modeling framework with a newer version of the WRF-Chem code, and we consider sharing the updates such as the soil NOx emissions scheme with the mainline model code.

3. WEPS builds on the existing WRF-Chem emissions processing tools to incorporate several global inventories, as well as allowing NEI and MEIC inventories to replace the global inventory. Is the process of regional inventories to override the global inventory an automated process (i.e., a regional netCDF file can be supplied and it'll overwrite the global inventory?) like in GEOS-Chem's emissions tool, HEMCO, or code changes will be needed? How extensible is WEPS to update with further inventories, and how easy is it to update inventores in the future? For example, I noted that FINN v1.01 is supported but not the more recent FINN v2.5 - will WEPS enable an easier update of the inventores for ingestion into WRF-Chem?

Response: Thank you for the question. The approach used by WEPS is like HEMCO to some extent. In WEPS, a namelist file is used to specify the emission inventory to be used, whether it is a global emission inventory or a specific regional emission inventory such as NEI. The extent of code modifications required depends on the format of the raw emission inventories. For example, we have successfully ingested the global emission inventory EDGAR-HTAP 2010, EDGAR 2015 and EDGAR-HTAP 2020 using the same code with minimal modifications as they share similar formats. Currently, WEPS has the capability to ingest emission inventories in both netcdf and text file formats. We anticipate that WEPS will allow an easy integration of an alternative wildfire or anthropogenic emission inventory by adapting the current code.

4. I also suggest some presentation improvements: organize the best configuration (of model physics) for each case study domain in a table; also label in the figures the D1 and D2 domains

for each case study; at times D1 is the whole region and D1 is marked by a rectangle and inset text could help the reader.

Response: Thank you for the suggestion to improve the presentation of the manuscript. We have first improved Table S1 by adding asterisk marks to denote the best-performing configuration of physics scheme for each target area. We have also added a table (Table 1) to summarize the best configuration of physics scheme selected together with other model set up such as land surface model and emission schemes for each target area studied. We have also improved figures with maps to clearly denote D1 and D2 (please see all figures in the updated manuscript and supplement).

Table S1. A suite of UI-WRF-Chem sensitivity simulations with different options of physics schemes over CHN-Beijing, ITA-Rome, USA-LosAngeles and USA-Atlanta target areas.

| Target area | Simulation number | Microphysics | Longwave | Shortwave | PBL |
|---|---|---|---|---|---|
| Beijing | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Morrison | RRTMG | RRTMG | YSU |
| | 3 | Lin | RRTMG | RRTMG | MYJ |
| | 4 | Lin | RRTM | Goddard | YSU |
| Rome | 1 | Lin | RRTMG | RRTMG | YSU |
| | 2* | Morrison | RRTMG | RRTMG | YSU |
| | 3 | WSM6 | RRTMG | RRTMG | YSU |
| | 4 | Morrison | RRTMG | RRTMG | MYJ |
| | 5 | Morrison | RRTMG | RRTMG | MYNN2.5 |
| | 6 | Morrison | RRTM | Goddard | YSU |
| Los Angeles | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Lin | RRTMG | RRTMG | MYJ |
| | 3 | Lin | RRTM | Goddard | YSU |
| Atlanta | 1* | Lin | RRTMG | RRTMG | YSU |
| | 2 | Morrison | RRTMG | RRTMG | YSU |
| | 3 | WSM6 | RRTMG | RRTMG | YSU |
| | 4 | Lin | RRTMG | RRTMG | MYJ |
| | 5 | Lin | RRTMG | RRTMG | MYNN2.5 |
| | 6 | Lin | RRTM | Goddard | YSU |

*These are the final configurations selected for each target area.

Table 1. A summary of model physics, chemistry and emissions configurations for CHN-Beijing, ITA-Rome, USA-LosAngeles, and USA-Atlanta target areas.

| Category | Model component | CHN-Beijing | ITA-Rome | USA-Los Angeles | USA-Atlanta |
|---|---|---|---|---|---|
| Physics | Microphysics | Lin | Morrison | Lin | Lin |
| | Cumulus | G3D | G3D | G3D | G3D |
| | Longwave radiation | RRTMG | RRTMG | RRTMG | RRTMG |
| | Shortwave radiation | RRTMG | RRTMG | RRTMG | RRTMG |
| | Planetary boundary layer | YSU | YSU | YSU | YSU |
| | Surface layer | Revised MM5 | | | |
| | Land surface model | NOAH | NOAH | NOAH | NOAH |
| Chemistry | Gas-phase | RADM2 | RADM2 | RADM2 | RADM2 |
| | Aerosols | MADE/SORAGM-DustSS | | | |
| | Photolysis | Madronich F-TUV | | | |
| Emissions | Anthropogenic emissions | MEIC 2016 | HTAP v3 (2018) | NEI 2017 | NEI 2017 |
| | Dust emissions | GOCART with AFWA modifications | | | |
| | Biogenic emissions of VOCs | MEGAN | MEGAN | MEGAN | MEGAN |
| | Soil NOx emissions | BDISNP | BDISNP | BDISNP | BDISNP |
| | Wildfire emissions | FLAMBE | FLAMBE | FLAMBE | FLAMBE |

Specific/Minor comments:

L40: "because of" -> I suggest "enabled by".

Response: Thank you for the suggestion. We have replaced “because of” with “enabled by”.

L223-225: It's not clear what the paragraph is suggesting here. Are you suggesting that the manuscript's use of GEOS-FP and MERRA2 differs from the common practice of using CAM-chem/WACCM outputs as chemical IC/BC (which I believe is the common practice in the WRF-Chem user's guide) or that GEOS-FP and MERRA-2 are different in that they assimilate satellite-based aerosol fields? I would suggest revising this paragraph for clarity.

Response: Thank you for the question and suggestion to make this paragraph clear. We meant the use of the GEOS-FP and MERRA-2 is different from other work in the sense that they assimilate satellite-based aerosol fields. We have revised this paragraph as follows:

*We have developed the capability to use GEOS FP and MERRA-2 data to provide chemical initial and boundary conditions in our UI-WRF-Chem modeling framework. Since WRF-Chem is a regional chemical transport model, time-varying chemical boundary conditions from global*

*chemical transport models are typically used to specify concentrations of different chemical species at the domain boundaries. This is especially important for long-lived chemical species, such as $O_3$, or capturing regional or long-range transport events. The common practice is to use global model outputs such as the Community Atmosphere Model with Chemistry, CAM-Chem (Emmons et al., 2020) for reanalysis or the Whole Atmosphere Community Climate Model (WACCM) (Gettelman et al., 2019) for forecasts. Unlike CAM-Chem or WACCM, which do not assimilate satellite aerosol observations, GEOS FP and MERRA-2 incorporate satellite-based aerosol data assimilation, which provides observational constraints for the day-to-day variations in aerosol concentrations over a given domain. To leverage this unique capability, we have modified the WRF-Chem preprocessor tool – mozbc (https://www2.acom.ucar.edu/wrf-chem/wrf-chem-tools-community) to create the GEOSBC module (Fig 1), enabling direct ingestion of GEOS FP and MERRA-2 data for updating chemical initial and boundary conditions.*

L425: "sea seal" -> "sea salt"?

Response: We have fixed it.

L446: "relative humanity" -> "relative humidity"?

Response: We have fixed it.

L519: "the chemistry will be transported..." -> maybe "the chemical tracers will be transported"?

Response: Yes, we agree. We have changed it to “the chemical species will be transported”.

SI Table S1 Los Angeles Simulation #1: "Li" -> "Lin"

Response: We have fixed it.

References:

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Crippa, P., Sullivan, R. C., Thota, A., & Pryor, S. C. (2017). The impact of resolution on meteorological, chemical and aerosol properties in regional simulations with WRF-Chem. Atmospheric Chemistry and Physics, 17(2), 1511-1528.

Cummins, K., Noble, J., Varner, J. M., Robertson, K. M., Hiers, J. K., Nowell, H. K., and Simonson, E.: The Southeastern U.S. Prescribed Fire Permit Database: Hot Spots and Hot Moments in Prescribed Fire across the Southeastern U.S.A, Fire, 6, 372, 2023.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., Tilmes, S., Vitt, F., Bardeen, C. G., McInerny, J., Liu, H.-L., Solomon, S. C., Polvani, L. M., Emmons, L. K., Lamarque, J.-F., Richter, J. H., Glanville, A. S., Bacmeister, J. T., Phillips, A. S., Neale, R. B., Simpson, I. R., DuVivier, A. K., Hodzic, A., and Randel, W. J.: The Whole Atmosphere Community Climate Model Version 6 (WACCM6), Journal of Geophysical Research: Atmospheres, 124, 12380-12403, https://doi.org/10.1029/2019JD030943, 2019.

Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmospheric Chemistry and Physics, 14, 5233-5250, 2014.

Hand, J., Prenni, A., and Schichtel, B.: Trends in seasonal mean speciated aerosol composition in remote areas of the United States from 2000 through 2021, Journal of Geophysical Research: Atmospheres, 129, e2023JD039902, 2024.

Jin, Z., Pu, Q., Janechek, N., Zhang, H., Wang, J., Chang, H., and Liu, Y.: A MAIA-like modeling framework to estimate PM2.5 mass and speciation concentrations with uncertainty, Remote Sensing of Environment, 303, 113995, https://doi.org/10.1016/j.rse.2024.113995, 2024.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170-181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Li, Y., Pickering, K. E., Barth, M. C., Bela, M. M., Cummings, K. A., and Allen, D. J.: Evaluation of Parameterized Convective Transport of Trace Gases in Simulation of Storms Observed During the DC3 Field Campaign, Journal of Geophysical Research: Atmospheres, 123, 11,238-211,261, https://doi.org/10.1029/2018JD028779, 2018.

Li, Y., Pickering, K. E., Barth, M. C., Bela, M. M., Cummings, K. A., and Allen, D. J.: Wet Scavenging in WRF-Chem Simulations of Parameterized Convection for a Severe Storm During the DC3 Field Campaign, Journal of Geophysical Research: Atmospheres, 124, 7413-7428, https://doi.org/10.1029/2019JD030484, 2019.

Li, Z., Maji, K. J., Hu, Y., Vaidyanathan, A., O'Neill, S. M., Odman, M. T., and Russell, A. G.: An Analysis of Prescribed Fire Activities and Emissions in the Southeastern United States from 2013 to 2020, Remote Sensing, 15, 2725, 2023.

Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., and Cahill, T. A.: Spatial and seasonal trends in particle concentration and optical extinction in the United States, Journal of Geophysical Research: Atmospheres, 99, 1347-1370, 1994.

Menut, L., Tuccella, P., Flamant, C., Deroubaix, A., & Gaetani, M. (2019). The role of aerosol–radiation–cloud interactions in linking anthropogenic pollution over southern west Africa and dust emission over the Sahara. Atmos. Chem. Phys., 19(23), 14657-14676. https://doi.org/10.5194/acp-19-14657-2019

Peters-Lidard, C. D., Kemp, E. M., Matsui, T., Santanello, J. A., Kumar, S. V., Jacob, J. P., Clune, T., Tao, W.-K., Chin, M., Hou, A., Case, J. L., Kim, D., Kim, K.-M., Lau, W., Liu, Y., Shi, J., Starr, D., Tan, Q., Tao, Z., Zaitchik, B. F., Zavodsky, B., Zhang, S. Q., and Zupanski, M.: Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales, Environmental Modelling & Software, 67, 149-159, https://doi.org/10.1016/j.envsoft.2015.01.007, 2015.

SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika, 73(3), 751-754. https://doi.org/10.1093/biomet/73.3.751

Solomon, P. A., Crumpler, D., Flanagan, J. B., Jayanty, R., Rickman, E. E., and McDade, C. E.: US National PM2. 5 chemical speciation monitoring networks—CSN and IMPROVE: description of networks, Journal of the Air & Waste Management Association, 64, 1410-1438, 2014.

Tao, M., Fiore, A. M., Karambelas, A., Miller, P. J., Valin, L. C., Judd, L. M., Tzortziou, M., Whitehill, A., Teora, A., Tian, Y., Civerolo, K. L., Tong, D., Ma, S., Adamo, S. B., & Holloway, T. (2025). Insights Into Summertime Surface Ozone Formation From Diurnal Variations in Formaldehyde and Nitrogen Dioxide Along a Transect Through New York City. Journal of Geophysical Research: Atmospheres, 130(9), e2024JD040922. https://doi.org/https://doi.org/10.1029/2024JD040922

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183-7192, 2001.

Wilks, D. S. (2011). Chapter 9 - Time Series. In D. S. Wilks (Ed.), International Geophysics (Vol. 100, pp. 395-456). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-385022-5.00009-9

Zhu, Q., Liu, Y., and Hasheminassab, S.: Long-term source apportionment of PM2.5 across the contiguous United States (2000-2019) using a multilinear engine model, Journal of Hazardous Materials, 472, 134550, https://doi.org/10.1016/j.jhazmat.2024.134550, 2024.