Response to Referee #1 Comments

We are grateful to the anonymous referee for its thoughtful and in-depth review of our manuscript, which has helped to strengthen the manuscript.

We have now modified the manuscript and adopted most of the recommendations from the reviewer. We have substantially expanded the discussion of results in a new section and have included new and relevant references both in the introduction and discussion sections. Below you can find our point-by-point response to the comments.

Recommendation: Major Revision

The manuscript is well-written, comprehensible and a good fit for GMD. The study presents the first implementation of the InMAP reduced complexity model (RCM) for Bogotá, supported by a WRF-Chem base simulation. The central finding that InMAP's baseline fields agree well with the driving CTM WRF (R^2 = 0.92), while marginal sensitivities to emission-reduction scenarios are overestimated by a factor of 2 (primary PM) to 5 (gas-phase precursors), is an important and relevant finding regarding the use of RCMs for policy assessment. The study is also one of the first InMAP implementations in the Global South and makes a useful methodological contribution through its locally derived emission seasonality using public-transit data as an activity proxy.

Before publication, however, (1) a discussion section (incl. some critical points to be discussed, see below) should be added to interpret the findings and put them in context, as at present there is essentially no discussion between the Results and Conclusions sections. Also, (2) some inconsistencies between the manuscript and the referenced data archives should be resolved, and (3) an analysis that is already implemented in the code and announced in the introduction (population-weighted exposure and mortality) should either be added to the manuscript or the scope in the introduction adjusted. These points are addressed below under major comments.

(1) A discussion section was included as recommended by the reviewer. The introduction and the bibliography were also expanded to include other RCM architectures.
(2) The inconsistencies in the text were clarified. Some were simple typos and are now completely corrected.
(3) This was an oversight. It has now been corrected.

Besides, there are also some minor comments.

Major comments

1. Missing discussion of the results

The manuscript moves directly from Results to Conclusions without interpreting the findings or putting them in context. But taking into account the differences between WRF simulations and the InMAP-BOG results, the following points should be included in a dedicated discussion section:

We appreciate the reviewer's comment regarding the need to expand the discussion and implications of the results. We have now included such section in the revised version of the manuscript. Although we incorporated as many recommendations as we could, new CTM simulations, however, were not feasible for our group at this time. Detailed responses below:

1a. Comparison with other reduced-complexity and/or source-receptor models.

The manuscript cites InMAP-US, InMAP-China, and Global-InMAP in the introduction but does not discuss alternative RCM architectures that have already been well validated for the specific task addressed here (sensitivity to emission scenarios), such as EASIUR, APEEP, RIAT+ or GAINS (https://pmc.ncbi.nlm.nih.gov/articles/PMC6911961/ gives already a good starting point for InMAP and other RCM validation against CTMs). Maybe also to name one very policy related example from Europe in detail: The SHERPA tool by EC JRC (https://knowledge4policy.ec.europa.eu/fairmode/sherpa_en), which operates at 7 x 7 km over Europe using a statistical source-receptor approach (trained on CHIMERE CTM runs) and achieves a bias below +/-10% for sectoral reduction scenarios up to 50% reduction (e.g. https://gmd.copernicus.org/articles/13/5725/2020/gmd-13-5725-2020.pdf), which is an order of magnitude better than the factors of 2-5 found with InMAP-BOG in this study. A discussion of why InMAP's physically based, steady-state transport approach performs worse here than a statistical SRR approach would substantially strengthen the interpretation of the findings and also support the use of RCMs in general.

We have expanded the introduction to include a broader range of RCM architectures, including EASIUR, APEEP, and SHERPA, as suggested. A discussion of InMAP's physically based approach and relatively poor performance is now also discussed.

1b. Possible causes of the overestimation.

The consistent factor of roughly 2 for primary PM across all scenarios might be a systematic bias in InMAP's simplified vertical mixing/dilution scheme rather than a random error. The additional, larger overestimation for gas-phase precursors (SOA, pNO3) could come from the fact that InMAP's equilibrium partitioning coefficients were originally calibrated for North American conditions and may not transfer readily to a high altitude (2600 m?) Andean basin with cooler and more humid conditions? The authors should discuss this or maybe alternative hypotheses regarding the source of the overestimation rather than only quantifying it.

This topic is now included in the discussion. We consider that there is perhaps an underestimation of vertical mixing in InMAP relative to WRF-Chem, leading to the overestimation in concentration when comparing both models.

1c. Framing of the study's motivation

The Introduction explicitly justifies InMAP as a tool for regions with limited resources and expertise for running full CTMs (including the Global South). The paper's conclusion that InMAP's sensitivities must first be calibrated against a full CTM run before being trusted challenges this idea, since exactly the regions lacking CTM capacity cannot perform this calibration. This tension should be discussed. Some ideas to do so:

We deeply appreciate the suggestions from the reviewer. We emphasized two aspects of our manuscript: (1) many implementations of RCMs do not evaluate their ability to reproduce scenarios. But we decided to explicitly test it so we could quantify its uncertainties and range of applicability. (2) Even with the uncertainties involved, which are not uncommon in RCM applications, we think there is potential for the model to be of use for marginal benefits analysis, as the spatial pattern of emission reduction benefits would still provide valuable information for decision makers. This is now included in the revised version of the manuscript.

- A more realistic framing of the workflow: not "no CTM required," but "run a CTM once (baseline plus a few representative scenarios) to get calibration factors, then use InMAP to run many additional scenarios."

From the reviewer suggestions, we adopted this approach and now explicitly describe it in the text, as we believe it is the most reasonable use of the model.

- Testing the transferability of the correction factors by applying the InMAP-BOG workflow to a second, contrasting city in SA (different altitude/topography/climate) to determine whether the factors of 2-5 are generic to InMAP or specific to Bogotá's Andean basin setting (could also help with 1b?).

Although this is an excellent point, we consider that it is beyond the scope of the current paper.

- Mechanistic diagnosis of the error (vertical mixing, partitioning coefficients) directly within the InMAP code rather than a purely empirical correction.

Although this is an excellent point, we consider that it is beyond the scope of the current paper. We have included a paragraph in the paper stating this as a potential avenue for future research.

- Testing whether a shorter calibration CTM run (e.g., 1-2 representative months rather than a full year) yields similar correction factors, which would lower the computational demand for calibration.

1d. Role of the regional background in the scenarios.

The manuscript notes that biomass-burning-driven background PM2.5 from the eastern lowlands is significant but not affected by the local emission-reduction scenarios (Sections 2.1.1, 2.1.3), and that InMAP's over- or underestimations are spatially linked to this east-west topographic pattern (Fig. 3b). However, there is no discussion of how this fixed regional background is treated in InMAP's simplified transport relative to the boundary conditions of WRF runs from higher nests, and whether inconsistent treatment contributes to the observed sensitivity mismatch.

InMAP's purpose is to calculate marginal concentration changes resulting from within-domain emission changes. We also consider that the biomass burning signal is most important for controlling the seasonality of air pollution, rather than the annual mean changes resulting from local anthropogenic emissions. For this reason, we consider unlikely that the marginal sensitivity overestimation is related to the regional background, as the reductions analyzed where only applied within the city-limits of Bogotá. We now discuss this more thoroughly in the manuscript.

The background is not fixed. Fires are common in the lowlands from January to April, but during the wet season the fire activity stops entirely and so does its contribution to air pollution in the region.

1e. Justification of the chosen resolution.

The 3 km x 3 km resolution is justified by the need to resolve the sharp gradient between dense urban and rural land use. However, since PM2.5 in Bogota appears to be strongly influenced by large-scale, regional transport (especially biomass burning), a discussion of how much added value the high resolution provides specifically for PM2.5 (as opposed to more locally dominated pollutants such as NO2 or fresh primary PM especially from road traffic) would be useful (also linked to 1d).

Points are well taken and are now discussed in the manuscript. Part of the reason for the 3x3 km resolution is that there are sharp gradients within the urban area (with concentrations increasing rapidly from the south-east to north-west). This is mostly due to local emissions. There are also sharp population-density gradients in the region. The biomass burning component, which is indeed very relevant for the region, largely controls the seasonality of PM pollution, but not necessarily its intra-urban distribution.

Additionally, we chose the highest resolution we could reliably run WRF-Chem at with the computational resources available. We tried 1km resolution runs but those would often stop due to numerical instabilities. This is now explicitly stated in the manuscript.

2. Data and code availability

There are some inconsistencies between the manuscript and the referenced data and code archives that should be resolved prior to publication:

- WRF Version conflict: The Zenodo dataset (10.5281/zenodo.19654867, describing the InMAP-Bog dataset) states WRF-Chem v3.9.1 as the basis, whereas the manuscript consistently states WRF-Chem version 4.1 throughout (Abstract, Section 2.1.2, Table 1). Neither the namelist nor the DataHub metadata (10.71590/9RESNN) contain a version statement that resolves this. Please clarify and make consistent.

The WRF-Chem version used was indeed v3.9.1 and this is now correctly stated across the manuscript. The links to the datasets were updated, and more data included to ensure the reproducibility of the work.

- Scope of the archived WRF-Chem data: The README of the DataHub archive (10.71590/9RESNN) describes the data provided there as daily means of the first model layer only, for aerosol species only and split by trimester. These are the data used for model evaluation (Figs. 1-2), not the full hourly 3D fields across all 41 vertical layers and the full species list from Table D1 that were used for the InMAP pre-processing. The data availability statement should clarify that not the full dataset is uploaded and available.

We have now stated this explicitly. The hourly 3D fields are a large dataset (+2.7 GB per day and around 1 TB for the full year). The size of the data set meant we were unable to store it in any of the freely available repositories, such as the recommended Zenodo, which has a storage cap of 150 GB.

- Missing reference to the InMAP-Bog archive: The Zenodo dataset containing the InMAP configuration files and the modified source code is not referenced anywhere in the preprint's data availability section and should be added. Following the interactive discussion, this is probably already in preparation.

Thanks for pointing this out. This was already corrected in the early discussion. Please see the modified manuscript and the track-changes version we prepared.

- Only one example run configuration (RPM-100) is archived, despite three scenarios are analysed in the manuscript. But the archived configuration should provide out-of-the-box run scripts before publication, given that providing a usable, reproducible implementation is one of the central contributions of this work and important for GMD.

Thanks for pointing this out. The sample run config was added as only the path to the input files needs to be changed. As we expanded the results sections and included a cell-by-cell reduction in emissions to demonstrate the potential to build source-receptor matrices, we have now included detailed scripts for configuring any emission reduction scenarios.

- Also, scripts to reproduce the plots and statistical analysis are missing in the repository. Not sure if these are necessary (following GMD rules) but they would be helpful to reproduce your results.

We have made sure that all the data used in the figures is accessible. We hope this is sufficient.

3. Announced but unreported analysis

The Introduction (lines 76 ff.) promises an analysis of "population-weighted PM2.5 reductions" resulting from the scenarios. And the archived InMAP run configuration (config_run_d03.toml) already defines an output variable (TotalPop) that computes population-weighted excess mortality via a standard log-linear concentration-response function, using the population and baseline mortality-rate data that are also archived. But neither this calculation nor any PWE result appears anywhere in the manuscript. The authors should either add this (as it's implemented in the code) analysis or adjust the corresponding statement in the paper's introduction. And if the analysis was omitted by choice, it should be stated why it was omitted, as this is directly relevant to assessing the tool's practical applicability for one of the main use cases of RCMs stated in the Introduction.

We have now expanded the results to include a demonstration of its use: reducing 1 ton of primary PM in each cell and estimating the benefits (i.e., the summation of the product of the population times PM reduction for each grid cell). In this way we wat to address the reviewer's comments and show the potential applications of the model. This was done not using the InMAP built-in capabilities, but could also be used in that way. In the current state, we wanted to stay clear of mortality estimates as spatially resolved baseline mortality data was not available.

However, the original focus of this manuscript was mainly the evaluation of the WRF-Chem run, its use to build InMAP, and the evaluation of some "examples" of reduction scenarios. The addition we have made to the manuscript addresses the reviewer's comment.

Minor Comments

- Section 3.3 (lines 237–239): The correlation coefficients of 0.60 (RPM-100) and 0.59 (Diesel-100) for the marginal PM2.5 fields are used in the text to support the claim that, "with a proper calibration," absolute reductions could also be estimated. Values of roughly 0.6 should be characterized as moderate rather than high.

This was corrected. And additional statistics describe the association between marginal concentrations from the two models are now included.

- It should be clarified consistently throughout whether reported "correlation coefficient" values are Pearson r or R^2: the Abstract states R^2 = 0.92 for the WRF-Chem vs. In-MAP agreement, but Section 3.2 reports a "correlation coefficient of 0.99" (urban) and "0.91" (entire D03 domain) without specifying whether these are r or R^2, or exactly which spatial subset the Abstract figure refers to.

This is now clarified. The metric used was the Pearson r. The number reported in the abstract is for the entire domain. The exact value of the metric was 0.913. The number to two decimal places (0.91) is now correctly reported everywhere in the manuscript.

- None of the reported factor estimates (the 2x and 5x sensitivity overestimation) or performance statistics (NMB, NME, R^2) are accompanied by uncertainty ranges or confidence intervals, making it difficult to judge whether the differences between scenarios, or between InMAP and WRF-Chem, are statistically robust or are within the models' uncertainty.

This issue was corrected. We have included more detailed statistics to quantify the uncertainties in our estimates.