Response to Referee #2 Comments

We are grateful to the anonymous referee for its thoughtful and in-depth review of our manuscript. We enhanced the model evaluation by including meteorological variables and have substantially modified the manuscript to reflect the changes.

We consider that our manuscript has two main strong points that make it novel and thorough: (1) the high-spatial resolution of the parent simulation (using local information) and (2) the explicit evaluation of the RCM results (which is mostly overlooked in other implementations)

In our revised version (and in the point-by-point response below) we aim to clearly explain these key points. We have expanded the discussion of results (including new and relevant references in the introduction and discussion sections) and expanded the evaluation of the base WRF-Chem run.

Although other well-known implementations of InMAP claim to have high-spatial resolution (even the global applications of the model), the information that goes into building the model (i.e., the Chemical Transport Model simulations used to build InMAP and global emission datasets), have much lower resolution that our 3km by 3 km simulation. Therefore, those simulations lack the necessary detail (meteorology, emissions, topography) to simulate emission reductions at the urban scale.

Neisa et al is a relevant paper as it describes a fast tool for a highly populated and polluted city in South America. Bogota is an interesting case because of its high altitude, complex terrain, strong diesel and resuspended-dust sources, and limited computational resources. However, the present manuscript does not yet establish that the Bagota InMAP is reliable enough for policy - InMAP reproduces the annual baseline spatial pattern well but substantially exaggerates the response to emission reductions by twofold for a primary-particle intervention and fivefold for the diesel related intervention- which makes me question the performance and usability.

This point is well taken. We have modified the language use to make it more cautious when describing the policy use potential for InMAP, and we are open about its limitations, and have expanded the discussion on applicability and limitations.

We would like to kindly point out to the reviewer that the literature on the application of InMAP (and that of other Reduced Complexity Models – RCMs) often completely overlooks the evaluation of the performance, as most of the studies simply build InMAP with a base CTM run and then apply it without testing how it performs compares to the same scenario run with the full CTM. We consider a strong point in our manuscript to have included an explicit evaluation of its performance, so the limitations in the approach are clear in any potential application of the tool. These limitations can only be known through the type of study we carried out.

The novelty should be stated more narrowly with respect to the Global InMAP that already provides global-to-urban variable-resolution modelling, with approximately 4-km grids https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268714. Global InMAP has also been applied to source contributions across many cities https://chemrxiv.org/doi/10.26434/chemrxiv-2022-lvgln. So, claims implying that urban-scale or non-US InMAP is novel is probably not valid now, with also ongoing implementations in India https://pavitra.org/. But I am not aware of another Latin American study using INMAP using WRFChem simulation and a local inventory, but the authors should confirm this through a documented literature search and use cautious wording such as 'to our knowledge'

Points are well taken. We are now more narrowly describing the novelty and reach of our work. Of course, our work loses the global reach made possible by Global InMAP but gains insights into urban-scale relevant dynamics.

Regarding the novelty: we would like to point out that the Global InMAP (which we cited in our MS) is built with a CTM run of 2°x2.5° resolution (i.e., roughly 220 km x 275 km at the equator). The emission fields used to feed Global InMAP had better spatial resolution (up to 0.1°x0.1° for NMVOC but 0.25°x0.25° other species) but one that is still too coarse for a compact urban area. In our manuscript we are exploring the potential applicability of InMAP at higher resolution by increasing the parent simulation spatial resolution.

The WRF-Chem methodology is insufficiently evaluated for its role in parameterizing InMAP. PM2.5 evaluation alone is not enough: wind speed and direction, boundary-layer height, temperature and precipitation should be evaluated because they control transport, mixing and deposition. The authors should also evaluate at least nitrate, sulfate, ammonium, organic aerosol and elemental or black carbon where and if measurements are available. WRF-Chem can reproduce total PM2.5 through errors among components, thus supplying incorrect chemical parameters to InMAP.

Following the recommendation we have now expanded the evaluation to include available observations (temperature, wind speed and wind direction). Please see the evaluation section in the Appendix. Unfortunately, no chemical speciation data is available for the city, and no routine measurement of PBLH is available.

Overall, the meteorological fields reproduce the diurnal cycle for wind and temperature. However, the model is biased cold relative to available observations, likely due to excessive IR cooling due to difficulty in simulating the cloud cover. This is not surprising given the location of the city in the equatorial Andes region.

Figure 1 is useful but should show uncertainty and clearer statistical definitions. The time series panel would benefit from station level comparison rather than only city-average lines. Figure 4 may also include the Diesel-10 experiment, although that smaller perturbation is especially important for evaluating marginal linearity. Correlations near 0.6 are moderate and in my opinion must not be evidence for good predictive performance but should be accompanied by slope, intercept, RMSE. Spatial maps of the intervention responses may also be interesting, and their ratios would reveal whether errors are concentrated near sources or downwind.

Points well taken. In Figure 1. To avoid overcrowding the figure we have decided to keep it as it is, but we have included metrics characterizing the spread of the model vs observations (RMSE, r, intercept and slope). Please, be aware that the station-level comparisons were already included in the Appendix section. That section has now been extended to include an evaluation of the meteorological performance of the model.

In Figure 4. We have also included the slope and intercept as suggested. Confidence intervals were included throughout the text when we refer to a mean.

In order to satisfy the claims of InMAP being more efficient, the authors should report WRF-Chem and InMAP wall-clock time, CPU-hours, memory use, storage requirements and preprocessing cost. The authors should calculate the break-even number of policy scenarios after accounting for the expensive annual WRF-Chem run. Otherwise in my opinion, the claim remains qualitative. Transferability should also be discussed as a model parameterized for one meteorological year and one city has not yet been shown to remain valid in another year, during unusual fire conditions. The manuscript section should compare the factor of 2 – 5 biases with other InMAP studies.

We thank the reviewer for the comment. Although we did not explicitly calculate the break-even number of policy scenarios, we have expanded the discussion to include the wall-clock time for InMAPs cell-by-cell emission reduction runs, and the massive difference between storage and computing requirements needed to run each model. We now explicitly acknowledge the considerable effort necessary to build InMAP as the CTM step is unavoidable. In our system, which was heavily limited regarding computational resources, the 1-year WRF-Chem run for the 3 nested domains took the best part of a year to complete. We have expanded the discussion on this topic.

We also included discussion of how the bias reported in our manuscript compares to other reports.

Other issues - Figure 1 refers to 13 monitoring sites, whereas Appendix retains 11 stations with sufficient completeness. The statistics and text must be thoroughly checked especially the R2 and R metrics mentioned.

The statistics were thoroughly checked. As suggested by both reviewers more consistency was applied throughout the entire manuscript. The reported metrics across the manuscript now include the Pearson correlation coefficient, RMSE, NMB, and confidence intervals for reported means.

The main methods describe the 3-km domain as roughly 480 x 480 km, while Appendix A gives 69 x 69 cells at 3 km, corresponding to approximately 207 x 207 km.

The 480 km was a typo, legacy from another configuration, and was corrected.

The introduction mentioned population-weighted PM2.5, but the results do not present population-weighted exposure, health impacts or equity analysis.

We have now addressed this issue.