We thank the reviewers for their time and insightful comments, which have substantially improved the manuscript. We have revised the manuscript and addressed the comments raised by the reviewers. The reviewers raised important comments on the rationales for our hypothesis, and the effects of our findings on future simulations. The main purpose of the study is to investigate the sensitivity of model predictions to the main inputs into the model. We apply different scenarios to evaluate the importance of major sources during the November 2017 extreme pollution episode over northern India. We feel this evaluation of inputs is needed to understand the extent that the forward model can be configured to capture the events. A contemporary way to try to capture such events in prediction mode is to employ data assimilation. The data assimilation results compensate for deficiencies in the inputs as well as structural problems within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability. Below, please find our responses to the reviewer's comments. The reviewer's comments are shown in black, our responses are shown in red, and the modified section of the manuscript is shown in blue.

We appreciate your time and comments and look forward to your decision.

Best Regards,

Behrooz Roozitalab, on behalf of all co-authors

RC1:

In this study, the authors have used the WRF-Chem to simulate the intensive pollution episode in the Indo-Gangetic Plain (IGP) in November 2017. They carried out 14 sensitivity simulations for different scenarios based on biomass burning emissions, chemical boundary conditions, and dust emissions. The model (base scenario) was evaluated for meteorological parameters (10 m wind speed and direction, 2 m temperature, and surface water vapor) with MERRA-2. The simulated AOD and $PM_{2.5}$ were compared with observations from AERONET and CPCB/US Embassy monitors, respectively. The authors have also looked into the daytime variation in ozone. The study is interesting because the authors try to simulate the $PM_{2.5}$ during November 6-13 using the emissions and aerosol-radiation feedbacks but no assimilation. This study is similar to a recent study published in JGR by Kumar et al., 2020 (https://doi.org/ 10.1029/2020JD033019). Overall, the manuscript needs a major revision. There are too many figures, which makes it hard to get the message across the reader. The labels in the figures are difficult to read. Here are my comments:

Authors Response:

We appreciate the reviewer for thoughtful comments. Please find our responses below. We also moved some of the figures to the supplementary documents and improved the labels quality.

Main comments:

RC1-1: I do not understand the hypothesis behind 14 simulations and still not being able to simulate the aerosols. The authors consider FINN_VIIRS_7Xperiod2 (base scenario) as the best scenario but the bias is still high (for AOD and PM2.5) compared to the observations.

Authors Response:

We appreciate the reviewers concerns and we try to clarify them in the followings. (We split the comments and address each part individually)

The main purpose of the study is to investigate the sensitivity of model predictions to the main inputs into the model. Here we take the approach of systematically exploring the impacts of different boundary conditions, dust, fire and anthropogenic emissions on the predictions of the pollution episode in November 2017. Based on literature (e.g. (Beig et al., 2019)), three major sources can play a role in this episode: - long-range transported dust incoming from boundaries, - long-range transported dust emitted inside boundary, and – agricultural fires on north-west India. We apply different scenarios to evaluate the importance of mentioned sources for the specific Nov. 2017 episode. We feel this evaluation of inputs is needed to understand the extent that the forward model can be configured to capture the events. A contemporary way to try to capture such events in prediction mode is to employ data assimilation. The data assimilation results compensate for deficiencies in the inputs as well as structural problems within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability. We clarified these points in the revised paper.

Text:

The main purpose of this study is to investigate the sensitivity of model predictions to the main inputs into the model. Prediction of extreme pollution events is important as they have major impacts on people and also make a strong impression regarding the capabilities of models. However, extreme events are hard to predict because they are often heavily impacted by episodic emission sources. Here we take the approach of systematically exploring the impacts of different boundary conditions, dust, fire and anthropogenic emissions on the predictions of the pollution episode in November 2017. A contemporary way to try to capture such events in prediction models is to employ data assimilation (Kumar et al., 2020). The data assimilation results compensate for deficiencies in the inputs as well as structural problems within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability.

There are different variations of inputs that can be used for each of these sources. In particular:

1. **We investigate four global data (MOZART, CAMChem, CAMS, and MERRA-2) to find how long-range transported dust incoming from boundaries affected air quality in India.**
2. **We modified the speciation of dust in dust emission module in the model to understand how in-boundary dusts played role in this episode.**
3. **We looked at two different biomass burning (B.B.) inventory, representative of two different methods of biomass burning emission inventories i.e. FRP and burned area, to find better B.B. emission inventory for high resolution modeling of agricultural fires.**
4. **We also did some experiments to reveal whether B.B. emission inventories are either systematically or occasionally biased low.**
5. **We also add one more experiment in the revised version to understand the impacts of anthropogenic emissions as suggested by reviewer 2.**

As a result, the large number of different scenarios were inevitable. Regarding the comment on high uncertainty in the base scenario after many experiments, we acknowledge the reviewer's concern. However, these experiments document the extent to which modifying these inputs can improve the prediction for this event.

For example, we found that statistics improved when we switched from a default scenario (ID: FINN_MERRA2) to the base scenario (ID: FINN_VIIRS_7Xperiod2) as shown below, Fig. S2, and discussion of Fig. 12a. For example, NMB were decreased by 62%. Moreover, the results for the base scenario show a fair performance compared to suggested benchmark criteria by (Emery et al., 2017). On the other hand, Fig. 12b and table S3 depict the statistics after excluding pollution episode days and show even better performance and satisfies "the goal criteria" based on Emery et al. (2017).

Table 1 Statistics before (FINN_MERRA2) and after (FINN_VIIRS_7Xperiod2) modifying biomass burning emission inventory

Scenario	**RMSE (μgm^{-3})**	**NMB (%)**	**MB (%)**
FINN_MERRA2	**167.88**	**-44.32**	**-113.14**
FINN_VIIRS_7Xperiod2	**118.47**	**-16.6**	**-42.38**

Furthermore, the aspect we looked to improve the modeling results is very different from optimization aspects. As a result, we did not expect to see as good as data assimilation results that strongly constrain the model. We evaluated our results in contrast with a new recent study by Kumar et al. (2020). They used data assimilation to look at a different time-period (while covering the same pollution episode). After assimilating MODIS AOD, their model performance for $PM_{2.5}$ improved significantly for the first day forecast: Mean bias (from -98.7 to -13.7 vs -42.38 in our study) and RMSE (from 167.4 to 117.3 vs 118.47 in our study); that study has better Mean Bias but RMSE values are very close to our values. We have added this evaluation in the revised paper.

Text:

Kumar et al. (2020) assimilated MODIS AOD to WRF-Chem in order to improve the air quality forecasts over Delhi. In their study, Mean Bias for first-day forecast of $PM_{2.5}$ concentration decreased from -98.7 μgm^{-3} to -13.7 μgm^{-3}. They also showed that RMSE decreased from 167.4 μgm^{-3} to 117.3 μgm^{-3}. Our results from the base scenario (Mean Bias: -42.38 μgm^{-3} and RMSE: 118.47 μgm^{-3}) shows comparable results to the data assimilation technique, while still both models are biased low.

RC1-2: From Fig. 3, the simulated AOD is underestimated over the IGP and overestimated over the rest of India. Comparison with AERONET (Fig. 4) shows that MERRA-2 does better over Jaipur because from Fig. 3b it is evident that AOD over Jaipur is in the range 0.5-0.8. Both WRF-Chem and MERRA-2 should have the same resolution while comparing (Fig. S2) and color bar scale comparable to Fig. 3. In conclusion, I think AOD is better simulated by MERRA-2 at both Jaipur and Kanpur. Please include the statistics for AERONET vs WRFChem and MERRA-2 AOD.

Authors Response:

Thanks for the comment. We completely agree that model is biased low over the IGP and biased high elsewhere and we have exclusively mentioned this point in the revised version. This behavior has been also shown in another new study by Jena et al. (2020): Fig. 4 in Jena et al. (2020), where they looked at December 2017 and January 2018.

Regarding the comparison of AOD over Jaipur, we agree that Fig. 3b shows averaged AOD over Jaipur for November was in the range 0.5-0.8, as the reviewer mentioned, and MERRA-2 results are closer to AERONET for non-episode days (AERONET data is missing for episode days). That is reasonable as MERRA-2 modeling system assimilates satellite AOD. VIIRS data also supports high AOD bias of model in Jaipur. However, it is also important that AERONET is missing data for the pollution episode between Nov. 6th and Nov. 13th as shown in Fig.4a. It suggests, as one possibility, that PM concentrations were too high during this period that the instrument was not able to retrieve data at that specific location. In the revised paper, we also include results where we scaled the particles anthropogenic emissions by a factor of two based on some new emission estimates (ID: Base_Anth2X). Using these anthropogenic emissions, averaged AOD bias for the IGP was reduced. This shows the need for improved estimates of biomass burning as well as anthropogenic emissions.

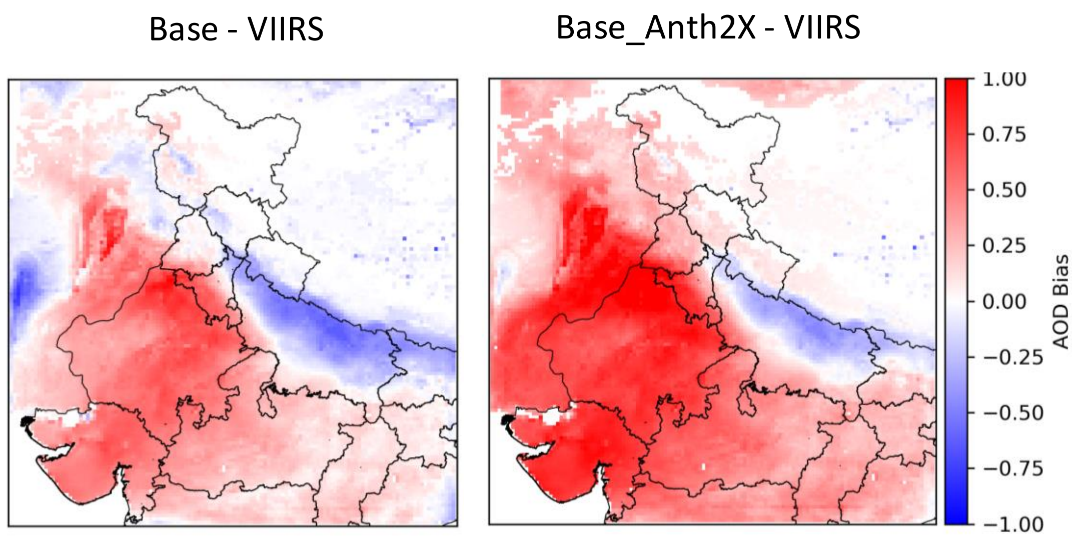


Figure 1 Bias of AOD at 550nm averaged over November 2017 base on a) base scenario b) base scenario with 2 times more anthropogenic particle emissions (ID: Base_Anth2X)

In the revised version, we also added VIIRS retrievals, which supports better performance of MERRA-2. Normalized Mean Bias between the model and AERONET was mentioned in the manuscript (Jaipur: +29.9% and Kanpur: -27.4%) but we added the same metric values between MERRA-2 and AERONET (Jaipur: -20.1% and Kanpur: -1.3%), which supports better performance of MERRA-2 in estimating AOD.

We thank the reviewer for the comments on Fig. S2. We modified the map in the revised document.

Text:

Figure 4 shows time series of modeled, MERRA-2 product, VIIRS retrievals, and observed AOD at the AERONET stations (location shown on Fig.1). AOD values at Kanpur, a station in the eastern IGP, were more than 1.0 before the pollution episode and reached up to 2.0 during the episode days, and decreased to values between 0.5 and 1 for the rest of days. The model captured the general trend although missed high AOD's between Nov. 9th and 13th, while MERRA-2 successfully captured the AOD trend through the whole month, including days with enhanced AOD values. This shows that AOD assimilation in MERRA-2 significantly improves AOD predictions. At Jaipur, located in southern IGP, the model overestimated AOD for the first five days of November. During the pollution episode days, the model is biased high compared to MERRA-2 and VIIRS retrievals. AERONET data showed low AOD values before the pollution episode but did not report values during the pollution episode. It suggests, as one possibility, that PM concentrations were too high during this period that the instrument was not able to retrieve data. After the pollution period, AOD values were lower than 0.5, showing relatively low PM concentrations. In general, MERRA-2 showed better performance in terms of NMB (Kanpur: -1.3% and Jaipur:-20.1%) compared with our model (Kanpur:-27.4% and Jaipur: +29.9%). Comparing averaged AOD with VIIRS retrievals for BASE_ANTHRO2X scenario showed lower bias over the IGP (Fig. S7). These results show the need for improved estimates of biomass burning as well as anthropogenic emissions.

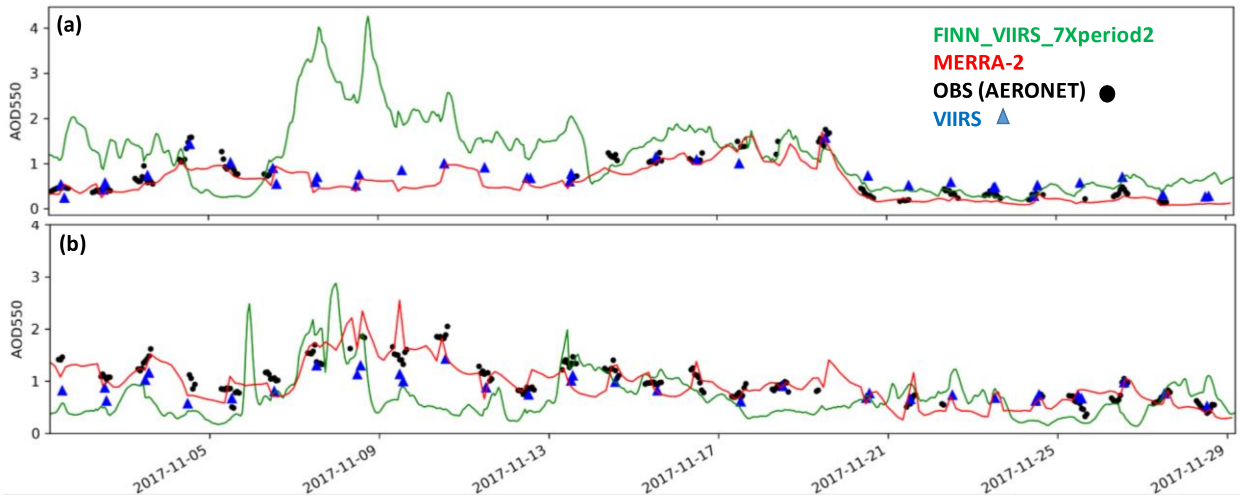


Figure 2 Figure 4 Time series of modeled (green line), VIIRS retrievals (blue triangle), MERRA-2 (red line), and AERONET (black dots) AOD at 550 nm during Nov. 2017 at a) Jaipur, b) Kanpur.

RC1-3: Lines 254-255: There are no major fires during November over western India/Rajasthan (as seen in Fig. 10). Also, there is no major dust event but there is a possibility of anthropogenic dust some of it being unique to the Indian region.
Authors Response:

Thanks for the comment that reveals the sentence was not clear. By western India for major fires, we meant western IGP and specifically Punjab, as it is clear on one-day fire map for Nov. 5th (Fig. 10).

Regarding dust, we have major dust emissions in eastern Pakistan near India, where the PM_{10} concentrations are more than 300 μgm^{-3}, and it can be seen on Fig. 14. However, neither these dust emissions nor long-range transported dust affect Delhi's air quality as explained in sections 3.5 and 3.6. Furthermore, it should be noted that the anthropogenic emissions used in the study include anthropogenic dust.

We have modified line254-255 sentence to clarify these points in the revised version.

Text:

Moreover, AODs were high over western IGP, close to major fires of Punjab, with a gradual gradient towards eastern and central India. Dust emission sources in the border of Pakistan also led to high AODs although they did not affect Delhi as discussed in the supporting document.

RC1-4: Lines 270-271: I do not agree with the authors' explanation. As seen in Fig. 3, WRF-Chem is simulating higher AOD values over western India/Rajasthan. The figures do not completely agree with the statements made by the authors.
Authors Response:

We thank the reviewer and rewrote the paragraph on AERONET data as shown above.

RC1-5: Why are the authors comparing the diurnal variation (Fig. 5a)? Do the emissions have a diurnal variation in the model?

Authors Response:

Thanks for the comment. In this study, all biomass burning emissions have diurnal variation in contrast with monthly anthropogenic emissions. As a result, $PM_{2.5}$ concentrations are subject to both daily atmospheric processes and emissions.

RC1-6: Why wasn't the $PM_{2.5}$ data from CPCB stations used in Fig. 5a?

Authors Response:

Regarding the question about comparing model results with CPCB stations, we compared our results against CPCB stations in Delhi in Fig. 6. In general during the whole paper, we show the results only at one station (i.e. US Embassy) when we look at time-series and we show daily box and whisker plots when we look at all CPCB stations. Our rationale is that: First, averaging data for time-series may remove some information by smoothing data. Second, we intended to show results from MERRA-2 in our time-series plots. Because of lower resolution of MERRA-2 data, almost all measurement stations in Delhi are located in only one grid cell of MERRA-2, which leads to misinformation. As a result, we decided to show time-series for only US Embassy location and box and whisker plots for CPCB stations. Nonetheless, we show timeseries for US-embassy and all CPCB stations in Delhi below, which shows CPCB averaged values have a similar trend to US-embassy data but with lower peaks.

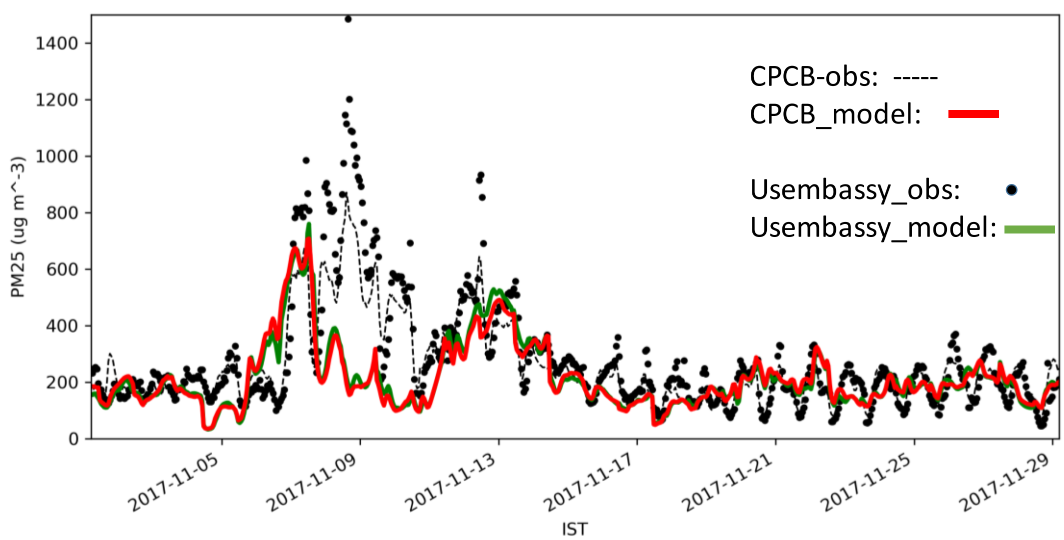


Figure 3 $PM_{2.5}$ timeseries in Delhi based on CPCB and US embassy data

RC1-7: Fig. 6 includes data from all the CPCB stations, how was the quality check performed on the CPCB data? Please add the details on the quality check of CPCB data in the methods section.

Authors Response:

Thanks for the comment. We didn't apply any filter to this data as we relied on quality control done by CPCB (https://cpcb.nic.in/quality-assurance-quality-control/). However, we studied how applying the following filters, done by Jena et al. (2020) and Kumar et al. (2020), change the dataset consisting of total 12768 hourly data points:

Filter 1: Remove less than 10 μgm^{-3} instances: removes 31 data-points

Filter 2: Remove the hourly difference between 100 (or 150 or 200) μgm^{-3} : removes 186 (or 71 or 31) hourly-data

Filter 3: Remove values more than 200 (400) μgm^{-3} right after NAN value: 33 (19). It basically removes data for Nov. 9th as it was applied after filter #2.

We found that the order of applying these filters is important. Below, statistics and timeseries for different orders of filters are presented. Order of filters (1,2,3) removes data for Nov. 9th and significantly improves the model performance over Delhi. We added these findings in the supplementary document and described in the revised version.

Text:

No additional quality control filters, other than the ones by CPCB (https://cpcb.nic.in/quality-assurance-quality-control/), were applied. We evaluated the results after applying the filters proposed by other studies (e.g. Kumar et al. (2020)); they had slight impacts on statistics (shown in the supporting document).

Table 2 Effect of applying filters to CPCB data on $PM_{2.5}$ statistics in Delhi

Province	Hourly Obs. Mean (±std) (μgm^{-3})	Hourly Model Mean (±std) (μgm^{-3})	24-hours NMB (%)	24-hours NME (%)	24-hours R (%)
CPCB-Delhi	255.5 (±146.6)	213.9 (±113.9)	-16.6	27.6	0.48
Only filter 3	248.4 (±140.3)	214.5 (±114.5)	-13.9	26.4	0.49
Filter123	215.5 (±95.5)	214.8 (±115.2)	-1.9	23.6	0.64
Filter132	248.6 (±140.8)	214.6 (±114.5)	-13.9	26.4	0.49

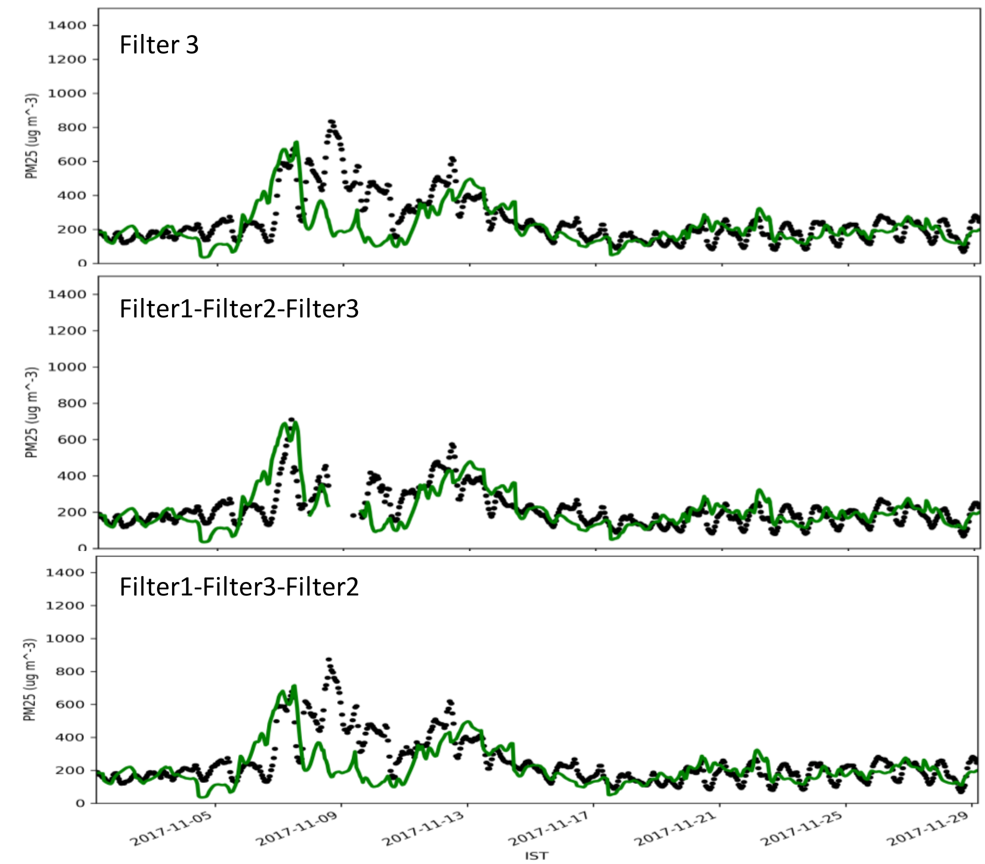


Figure 4 Effect of applying additional filters to CPCB data on averaged $PM_{2.5}$ timeseries in Delhi

RC1-8: It is better to show the spatial plot along with the CPCB and US Embassy observations as a scatter. It will show if the model was able to capture the spatial variation in observed $PM_{2.5}$. How does MERRA-2 compare with the CPCB observations?

Authors Response:

As the model resolution is 15km in this study, we usually see more than one measured CPCB stations are located in one model grid cell. For example, 17 stations in Delhi are located in only 6 grid cells; repetition affects the scatter plot (Fig. 5a below). We also observe lower variability in box and whisker plots of the model compared to observation data due to same reason. In other words, scatter plots will not provide enough insights when considering all individual stations. To show the spatial performance of the model, we plot the scatter plot for averaged concentration of different states. Below, we show the scatter plot for Delhi, Haryana, and Rajasthan, which reveals the good spatial performance of the model (Fig. 5b below). Adding data from Punjab to this plot (Fig. 5c below) ,significantly degrades the performance. The reason is due to extremely high bias in Punjab data. Punjab observation data seem to be very uncertain, as it doesn't show any signal of the pollution episode while satellite data show huge amount of agricultural fires during those days. We added the below scatter plots in the supplementary document. However, we think scatter plots were better tools if we had more spatial data (e.g. a gridded dataset).

Regarding MERRA-2, due to large grid cell size of MERRA-2 (0.625x0.5deg), all CPCB stations of Delhi are in the same grid cell. As a result, we only look at one representative station (i.e. US Embassy station) when using MERRA-2 data.

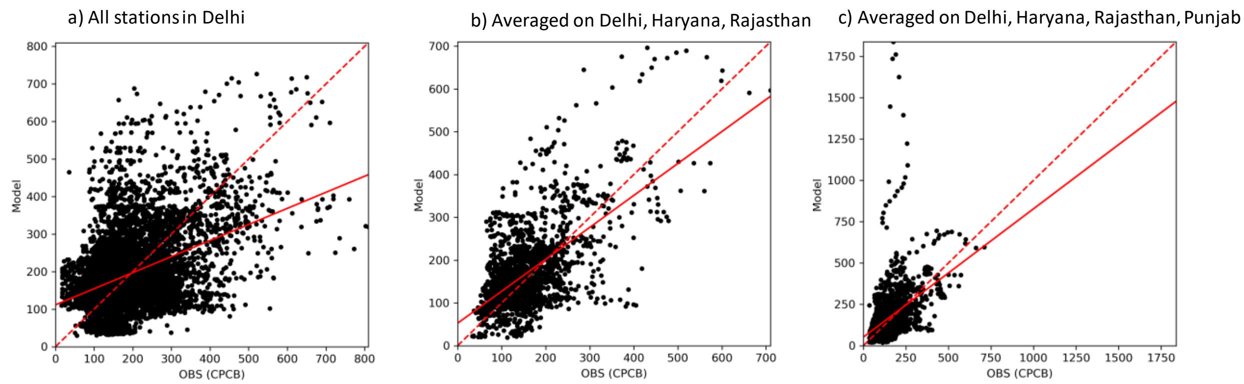


Figure 5 Scatter plots for a) all stations in Delhi combined b) averaged concentrations in Delhi, Haryana, and Rajasthan c) averaged concentrations in Delhi, Haryana, Rajasthan, and Punjab. Filters are applied to CPCB data.

RC1-9: Lines 292-293: I do not completely agree with the explanation of transported dust from the Middle East. Is PM_{10} high over Delhi? Looking at the CALIPSO profile data shown in Beig et al., 2019, it is polluted dust, which is different from desert dust. The authors can also look into the MISR data for dust AOD. The authors have made a statement in section 3.6 that sensitivity tests do not show a major influence of dust being transported from the Middle East.

Authors Response:

We completely agree with the reviewer and sorry for confusion. Our analysis do not show a major influence of long-range transported dust as discussed in sections 3-5 and 3-6. Since lines 292-293, which are in the section we were looking at the whole month and mentioning other studies' views, may be confusing for the reader on our point of view, we removed the following sentence:

"We looked at MERRA-2 surface $PM_{2.5}$ concentration data for the study period to explore if dust was a major source."

RC1-10: Have the authors looked into the PBLH from the model and compared with the observations? You might have to derive the PBLH from radiosonde observations. The authors attribute the low $PM_{2.5}$ on Nov 8-10 to the plume rise in the model. My understanding is half of the fire emissions will be released within PBLH and the rest above it. The model is simulating higher PBLH as seen in Fig. 13. I would suggest instead of comparing at the US Embassy only, include observation data from all CPCB stations. PBLH in Delhi during November is less than 1000 m (Nakoudi et al., 2019, AMT). The days when PBLH was low (less than 1000 m) in the model (Nov 7, 11-13), the simulated $PM_{2.5}$ was comparable to the observations.

Authors Response:

We thank the reviewer for this important comment. Unfortunately, there is not any measured PBLH data available to compare the modeling results. On the other hand, estimating PBLH using sensing data is a challenging task (Nakoudi et al., 2019)and needs specific considerations (Wang and Wang,

2014). As an example, we used specific humidity (and relative humidity) from radiosonde data for Delhi (provided by university of Wyoming at http://weather.uwyo.edu/upperair/sounding.html) and attributed the height with lowest vertical gradient as the PBLH. As the figure below shows, WRF-Chem diagnostic PBLH shows higher values (~150m) for Nov. 6th 12UTC (17:30 IST). This result shows PBLH overestimation. We agree with the reviewer that lower PBLH could entrain more aerosol and increase concentrations and we have added the impacts of PBLH on aerosol loading and the importance of accurate PBLH in the revised version. We also compared our modeling results to Nakoudi et al. (2019)findings, which show comparable results. Our hourly averaged PBL has higher heights during daytime compared to ones in winter by Nakoudi et al. (2019). We should note that their results do not cover October and November (i.e. post-monsoon).

On the other hand, we added the PBLH line to the cross sections (white line) in Fig. 8. The line is not obvious on 00 UTC times due to very low extent of the PBL. On Nov. 6th-12, we see very low PBL upwind of Delhi and significant amount of smoke above boundary layer. Therefore, these findings accompanied with Fig. 13b supports the argument that the plume rise in the model released the emissions too high or the model did not mix the smoke sown fast enough (Plume rise module does not have a constrain to release half of the emissions below PBLH and the rest above). We have used new figure in the revised version.

However as mentioned earlier, the main purpose of this study is to investigate the sensitivity of model predictions to the main inputs into the model. We believe another study is required to look at the structure of the model and study the effects of different PBL parameterization modules on PBLH, which is beyond the scope of this paper.

Regarding CPCB stations, as discussed earlier, one MERRA-2 grid cell includes all CPCB stations; as a result, including them will not provide any insights. However, we agree with the reviewer's comment that the model showed comparable $PM_{2.5}$ concentrations with lower PBLH as discussed above.

Text:

Increasing emissions also indirectly influenced modeled air quality over Delhi. As our model configuration included feedbacks, absorbing aerosols in the atmosphere (products of fire emissions) decreased the surface solar radiation budget, changed the dynamics of the atmosphere, reduced the Planetary Boundary Layer (PBL) height, and increased aerosol concentrations. In other words, higher PBLH leads to lower concentrations. For example, Murthy et al. (2020)found that $PM_{2.5}$ concentration decreased up to 14 μgm^{-3} for 100m increase in PBLH. Figure 12 shows the interactions between PBLH and $PM_{2.5}$ concentration at the location of the US embassy. By increasing FINN inventory by 7 times, the PBL height decreased by ~50% on Nov. 6th, (compare FINN_VIIRS_7Xperiod2 and FINN_MERRA-2 panels in Fig. 12). However, a measured PBLH dataset can provide better insights. As a result, another study is required to compare modeled PBL heights to observed data (e.g. Nakoudi et al., 2019) and study the effects of different PBL parameterization modules on aerosol concentrations.

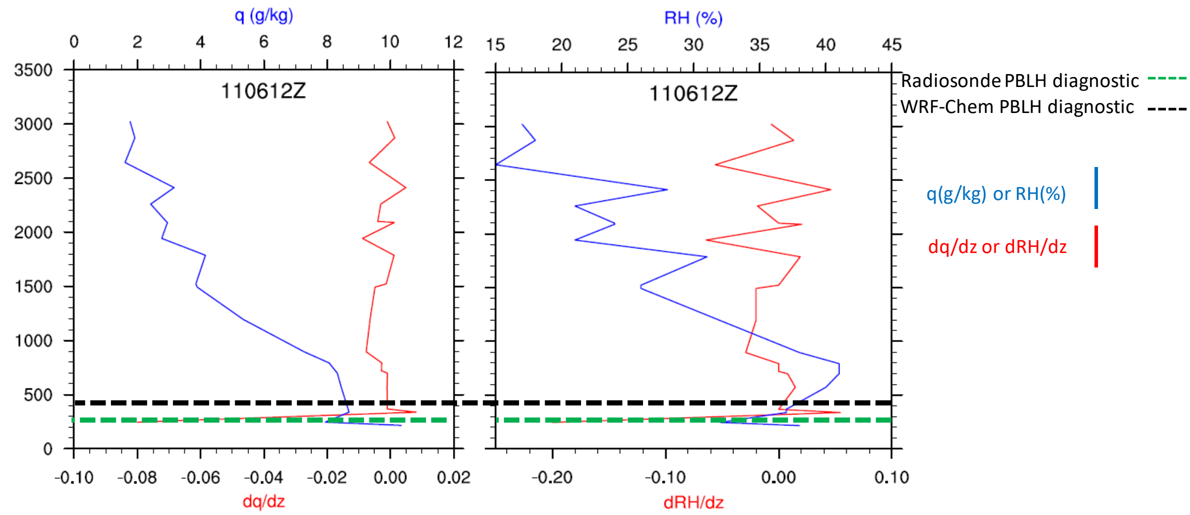


Figure 6 The vertical blue lines represent the quantity (left: specific humidity, right: relative humidity) and vertical red lines show the vertical gradients for mentioned parameters. We defined radiosonde diagnosed PBL height as the minimum gradient (dashed green line) and WRF-Chem diagnosed PBLH is shown in dashed black line.

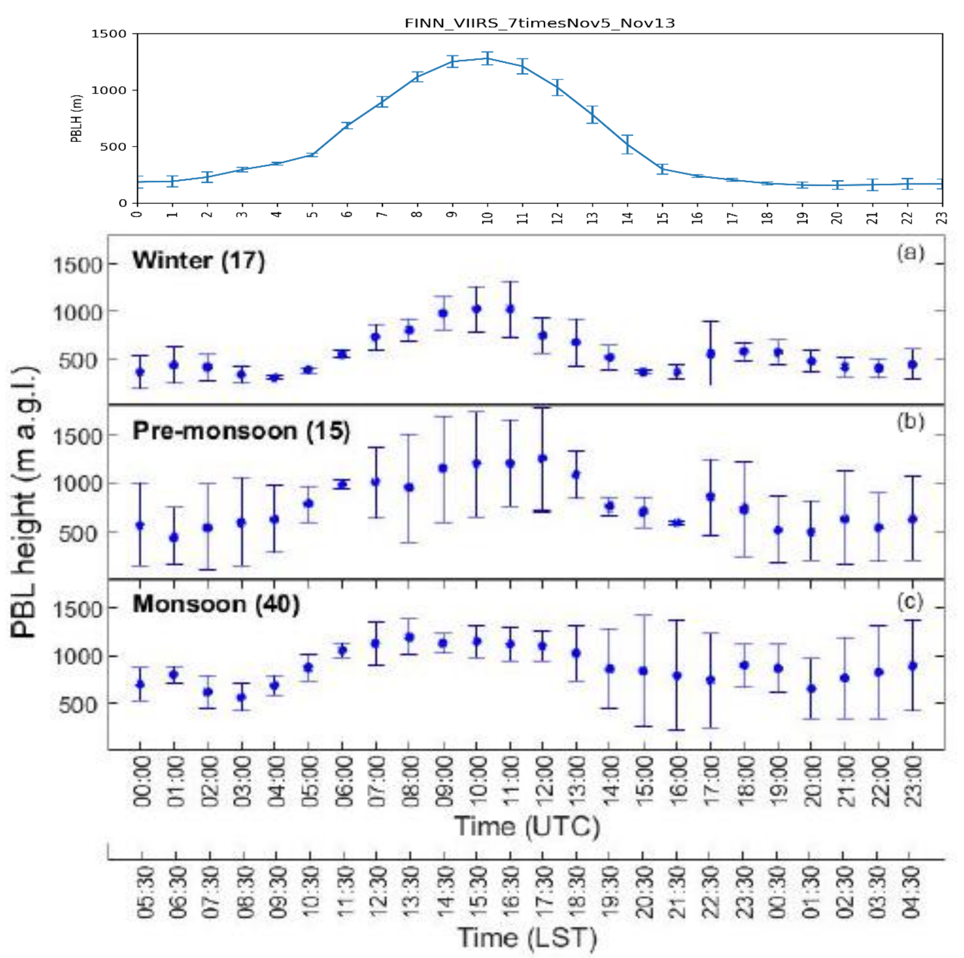


Figure 7 Averaged hourly PBLH during November (top panel) and other different seasons (screenshot from Nakoudi et al. 2019)

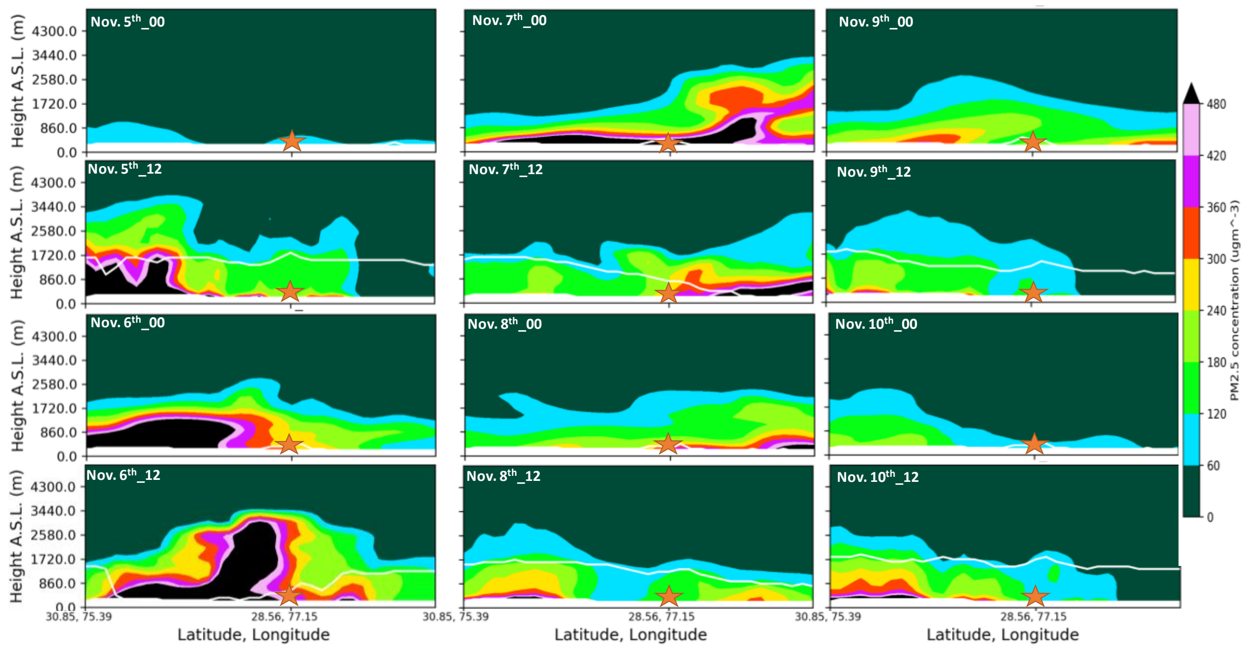


Figure 8 Figure 8 Vertical cross section of $PM_{2.5}$ concentration through the path shown in Fig. 1 for the days between Nov. 5th and Nov. 10th. For each day, two snapshots are shown at 00UTC (5:30AM local time) and 12UTC (5:30PM local time). The orange star shows the location of Delhi through the path. White line shows the PBL height across the path

RC1-11: According to me, sections 3.3, 3.5 and, 3.6 do not add anything new to the paper. The inclusion of missing fire emissions is an important part of the simulation. Also, it is worth to include a comparison with the Kumar et al., 2020 study.

Authors Response:

We thank the reviewer for the comment. We moved sections 3-5 and 3-6 to the supplementary document and replace them with one brief discussion at the end of section 3-4 (sensitivities on biomass burning). These two sections show why long-range transported dust both outside and inside of the domain did not influence air quality in Delhi during pollution episode of November 2017. However, we decided to keep section 3-3 (PM speciation) as it conveys important information on secondary aerosols.

We appreciate the reviewer for introducing the paper by Kumar et al., 2020. We compared our results to their findings as discussed earlier.

Specific Comments:

RC1-12: Line 33: Ghude et al., 2016 do not mention the long-term health impacts due to an increase in emissions based on current policies.

Authors Response:

We thank the reviewer for catching this mistake. We removed that reference.

RC1-13: Lines 42-45: David et al., 2019 show the impact of both transport and emissions on PM2.5 in different regions in India. The authors can add results from the study.

Authors Response:

We added David et al. (2019)findings to the revised version.

Text:

Studies show that ozone and particulate matter with diameter less than 2.5 micron ($PM_{2.5}$), are attributed to more than one million individual premature deaths in India (Cohen et al., 2017;HEI, 2018). David et al. (2019) found anthropogenic emission within India led to about 80% of the total premature death due to $PM_{2.5}$ in India. Furthermore as industrial activities are growing, emissions are increasing too; health impacts attributed to long-term exposure air pollution are predicted to increase based on current policies (Conibear et al., 2018a).

Text:

David et al. (2019) attributed about 16% of total premature $PM_{2.5}$-related death to emissions outside India.

RC1-14: Line 78: Add reference - Kumar et al., 2020 (JGR)

Authors Response:

We added this reference.

RC1-15: Lines 128-129: Studies by Conibear et al., 2018, Venkataraman et al., 2018, and David et al., 2019 have shown that residential energy use is the main source of PM2.5 in India.

Authors Response:

Thanks. We emphasized the reviewer's point in the revised version.

Text:

Biomass and biofuel use in residential sector for heating and cooking purposes have significant contributions to air quality in India (Conibear et al., 2018a;David et al., 2019;Venkataraman et al., 2018)

RC1-16: Lines 203-204: "Irrespective …" From where did the authors get this information?

Authors Response:

Thanks for the comment. This was from personal inspections but to be clear we modified the sentence.

Text:

Irrespective of this condition, some of CPCB stations are placed on top of the buildings with restricted clean flow of air (personal inspections).

RC1-17: Multiple places the authors mention "the model was able to capture ..." (For example, Lines 230, 266, 343) – please support your statements with statistics (MB, RMSE).

Authors Response:

Thanks. We added statistics when it was not mentioned.

RC1-18: Replace provinces with states (Lines 311, 315).

Authors Response:

Thanks. We replaced all 'provinces' with 'states'

RC1-19: Line 374: Change "Fig. 1" to "Fig. 8".

Authors Response:

Thanks. We clarified it.

RC1-20: Line 452: What are some of the other meteorological phenomena?

Authors Response:

Thanks for the comment. We mostly meant thick fires can be identified as clouds in retrieval algorithm or they may be actual clouds and large amount of water vapors, leading to biases. Bright surfaces (in deserts) are other uncertainty sources. We modified the sentence:

Text:

Some studies have shown that thick fires can be identified as clouds in retrieval algorithms (Dekker et al., 2019;Huijnen et al., 2016).

RC1-21: Table 2: Arrange the table as explained in the text (section 2.2).

Authors Response:

Thanks for the comment. We rearranged it to first show the experiments on anthropogenic and biomass burning emissions, then on boundary conditions, and lastly on dust emissions, to follow section 2.2. We had to add the new sensitivity test to the bottom of the table as to keep the format.

References

Beig, G., Srinivas, R., Parkhi, N. S., Carmichael, G., Singh, S., Sahu, S. K., Rathod, A., and Maji, S.: Anatomy of the winter 2017 air quality emergency in Delhi, Science of The Total Environment, 681, 305-311, 2019.

David, L. M., Ravishankara, A., Kodros, J. K., Pierce, J. R., Venkataraman, C., and Sadavarte, P.: Premature mortality due to PM2. 5 over India: Effect of atmospheric transport and anthropogenic emissions, GeoHealth, 3, 2-10, 2019.

Dekker, I. N., Houweling, S., Pandey, S., Krol, M., Röckmann, T., Borsdorff, T., Landgraf, J., and Aben, I.: What caused the extreme CO concentrations during the 2017 high-pollution episode in India?, Atmospheric Chemistry and Physics, 19, 3433-3445, 2019.

Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., and Kumar, N.: Recommendations on statistics and benchmarks to assess photochemical model performance, Journal of the Air & Waste Management Association, 67, 582-598, 2017.

Huijnen, V., Wooster, M. J., Kaiser, J. W., Gaveau, D. L., Flemming, J., Parrington, M., Inness, A., Murdiyarso, D., Main, B., and Van Weele, M.: Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997, Scientific reports, 6, 26886, 2016.

Jena, C., Ghude, S. D., Kulkarni, R., Debnath, S., Kumar, R., Soni, V. K., Acharja, P., Kulkarni, S. H., Khare, M., and Kaginalkar, A. J.: Evaluating the sensitivity of fine particulate matter (PM 2.5) simulations to chemical mechanism in Delhi, Atmospheric Chemistry and Physics Discussions, 1-28, 2020.

Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., Kulkarni, S., Sperati, S., Soni, V. K., and Nanjundiah, R. S.: Enhancing Accuracy of Air Quality and Temperature Forecasts During Paddy Crop Residue Burning Season in Delhi Via Chemical Data Assimilation, Journal of Geophysical Research: Atmospheres, 125, e2020JD033019, 2020.

Murthy, B., Latha, R., Tiwari, A., Rathod, A., Singh, S., and Beig, G.: Impact of mixing layer height on air quality in winter, Journal of Atmospheric and Solar-Terrestrial Physics, 197, 105157, 2020.

Nakoudi, K., Giannakaki, E., Dandou, A., Tombrou, M., and Komppula, M.: Planetary boundary layer height by means of lidar and numerical simulations over New Delhi, India, Atmospheric Measurement Techniques, 12, 2019.

Wang, X., and Wang, K.: Estimation of atmospheric mixing layer height from radiosonde data, Atmospheric Measurement Techniques Discussions, 7, 2014.

RC2:

This study investigates the processes causing severe air pollution episodes in New Delhi, India by focusing on one such event observed during November 2017. Specifically, the authors evaluate the impact of biomass burning emissions, long-range transport of dust, and dust emissions on WRF-Chem simulated $PM_{2.5}$. The model captured the day to day variability but missed the peak pollution peak during 7-10 Nov. Secondary Inorganic Aerosols and Secondary Organic Aerosols are estimated to contribute 30% and 27% of total $PM_{2.5}$ concentrations in Delhi. Back trajectories showed influence of agricultural fires in Punjab on $PM_{2.5}$ in Delhi. Long-range transport of dust is not found to affect air quality in Delhi during this time. High biases in model AOD were observed over central India and low biases over the eastern IGP.

While such studies are very important as they provide important information about the sources leading to dangerous air pollution episodes and inform the mitigation strategies, unfortunately this study does not consider all the key sources of uncertainties in the model simulations and may misinform the mitigation strategies. I am particularly concerned about the ignorance of anthropogenic emission uncertainties that were left out irrespective of several evidences pointing to their key role in the analysis presented in the paper itself. The authors should also provide a clear description of the rationale behind selecting biomass burning and dust aerosols as the most important sources of uncertainties in the model simulations. Below I provide my major and minor comments.

Authors Response:

We appreciate the reviewer for pointing to important issues. We addressed the comments and concerns here and below:

We share the view about the critical role of anthropogenic emissions roles in air quality over the IGP. The uncertainty in anthropogenic emissions lead to concentration biases for typical days. Moreover, we acknowledge the importance of anthropogenic emissions since emissions due to heating also increase as the weather gets cold during Oct. and Nov. We added a scenario in which we increased all particles anthropogenic emissions by a factor of 2 based on recent emission work in Delhi (ID: Base_Anth2X). In following comments, we present its results.

However, agricultural fires have a more significant contribution in post-monsoon extreme pollution events in Delhi (Kulkarni et al., 2020). Moreover, (Beig et al., 2019) showed that extreme pollution episodes during November 2017 was mainly due to agricultural fires and long-range transported dust. Lines 46-52 discuss these points although we agree that there are some exceptions, too. For example, extensive use of firecrackers and fireworks in the Diwali festival on October 20th in 2017 led to $PM_{2.5}$ concentrations above 600 (μgm^{-3}) Therefore, we focused only on November to exclude that episode. On the other hand, our simulation results after excluding extreme pollution days show fair statistics (Table S3). We highlighted the importance of anthropogenic emissions in the revised paper and tried to express the reviewers point in the study limitations section in the revised version:

Text:

During this study, we did not primarily focus on improving anthropogenic emissions over the region in order to capture extreme pollution episode. However, anthropogenic emissions are low in global emission inventories and needed to be improved (Jat et al., 2020). Moreover, very low biased concentrations for some days and trajectory results suggest the existence of some other sources, primarily anthropogenic sources, upwind of Delhi that should be studied more.

Main comments:

RC2-1: Figure 3 shows that increasing the fire emissions by a factor of 7 is too high and leads to large overestimation of AOD especially in the western part of the domain. Large underestimation in the IGP is reflecting the underestimation of anthropogenic emissions but no sensitivity experiment was designed to look into that. So, the "base" configuration might be showing good performance in Delhi for wrong reasons.

Authors Response:

We appreciate reviewer's genuine and important concerns. Please find our responses, below. (We split the comments and address each part individually)

Regarding Biases in Fig. 3: As you and reviewer 1 mentioned, we completely agree that model is biased low over the IGP and biased high elsewhere and we have exclusively mentioned this point in the revised version. We acknowledge that uncertainty of anthropogenic emissions is playing an important role in these biases. We did another experiment where we increased anthropogenic emissions for all the particles by a factor of 2 (ID: Base_Anth2X). This modification increased $PM_{2.5}$ concentrations in Delhi up to ~150 μgm^{-3}, which led to overestimation (in contrast to underestimation in base scenario) at most of non-episode days (time-series shown below). Although this scenario did not help capturing concentrations during the episode, it confirms the need for better anthropogenic emissions. On the other hand, it increased the AOD bias over southern IGP while reduced the bias over IGP (bias map shown below). These results suggest anthropogenic emission inventories have higher bias over IGP compared with non-IGP regions. However, we acknowledge the importance of having dynamic (daily) anthropogenic emission inventory.

Text:

Although different meteorological parameters can be responsible for the biases, accuracy of anthropogenic emissions is important. For example, recent local anthropogenic emission inventories developed for Delhi have higher particle emissions than in the regional inventory used in this study, which impacts modeled $PM_{2.5}$ concentrations for typical days (Kulkarni et al., 2020). We conducted BASE_ANTHRO2X scenario to investigate the effect of uncertainties in the anthropogenic emissions. This scenario increased $PM_{2.5}$ concentrations in Delhi up to ~150 μgm^{-3}, which led to overestimation (in contrast to underestimation in base scenario) at many of non-episode days (Fig. S7). Although this scenario did not help in capturing the high concentrations during the episode, it confirms the need for better anthropogenic emissions. On the other hand, it reduced the bias over IGP (Fig. S7). These results point out the need for best estimates of emissions of both anthropogenic and biomass.

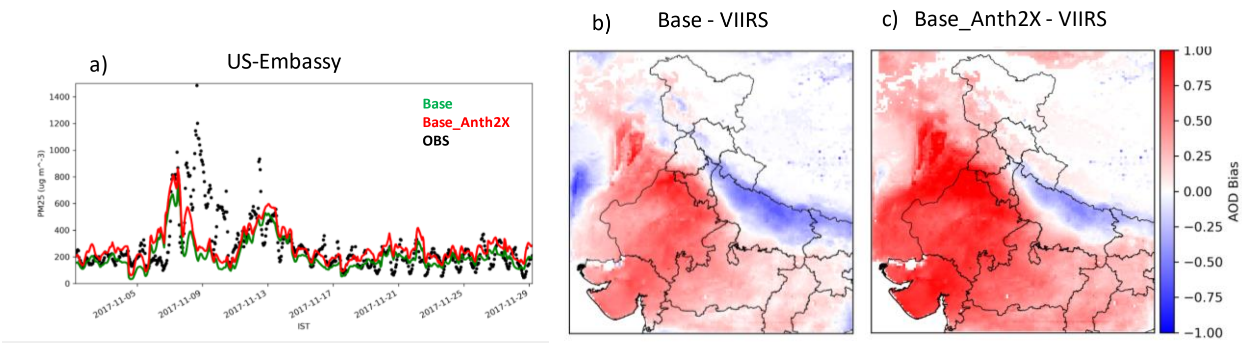


Figure 1 a) Timeseries for $PM_{2.5}$ concentration at the location of US embassy using Base scenario and Base_Anth2X scenario B) Bias of AOD at 550nm averaged over November 2017 base on b) base scenario c) base scenario with 2 times more anthropogenic particle emissions (ID: Base_Anth2X)

In addition, our experiments were primarily focused to capture the extreme pollution episode over Delhi as the reviewer pointed out. On the other hand, we would like to mention an important point regarding the accuracy of the base scenario for other locations:

Below, we show the AOD biases for our base scenario (as in Fig3.c) on left panel, FINN_MERRA2 scenario (a scenario without any enhancement on fire emissions) on middle, and the difference between these two scenarios on the right panel.

The bias pattern of FINN_MERRA2 has also been reported in another study by Jena et al. (2020). They looked at a different time period (Dec. 2017 to Jan. 2018) but they show same pattern with lower values (which is most possibly due to lower concentrations in their period of interest). Their results (specifically Fig.4 in Jena et al., 2020) support the importance of anthropogenic emissions as the reviewer mentioned, and we acknowledge that as discussed above.

On the other hand, looking at base and FINN_MERRA2 reveals that we clearly improved the AOD results for Punjab. It also shows low bias of FINN_MERRA2 shifted to high bias of base scenario for Haryana. The difference between base scenario and FINN_MERRA2 scenario (right panel) shows the impact of increasing FINN emissions by 7 times for a 8-days period; it increased the mean AOD biases over the whole domain by 0.09 (±0.23).

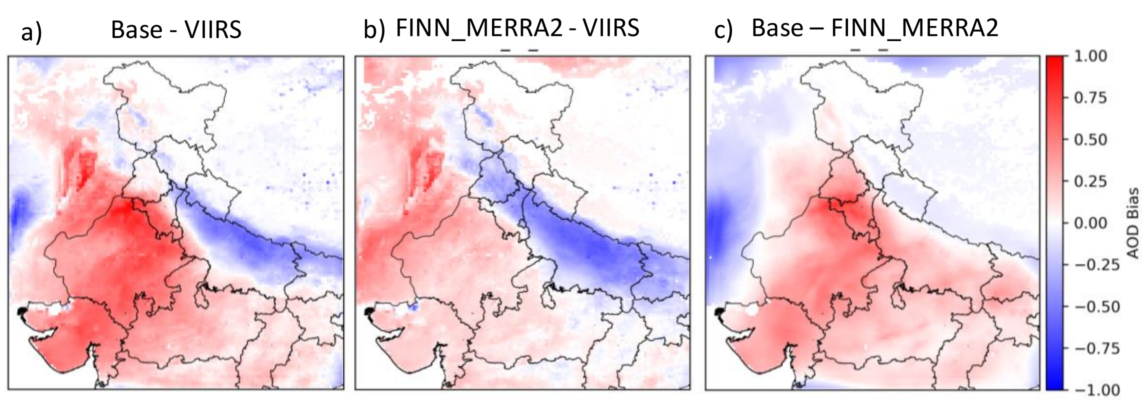


Figure 2 Bias of AOD at 550nm averaged over November 2017 base on a) base scenario b) a scenario without any modifications on biomass burning emissions (ID: FINN_MERRA2), c) difference between Base and FINN_MERRA2

RC2-2: Fig. 4a shows an AOD of 4 which is unrealistic for Jaipur. It looks like the authors paid all the attention to getting $PM_{2.5}$ in Delhi correct simply by upscaling the emissions in the upwind regions but no care was taken to maintain the model performance in the upwind regions. Consequently, the model shows a positive bias in $PM_{2.5}$ in Punjab with a spatial variability (reflected by standard deviation in Table 3) that is nearly 3.5 times higher than the observed variability in Punjab. Nov. 24 case (no fire day) also supports the idea that the anthropogenic emissions are substantially underestimated.

Authors Response:

Regarding Fig4.a, we agree that the model is generally biased high over the Jaipur. Moreover, VIIRS data also show low AOD values for Jaipur during episode days. That is reasonable as MERRA-2 modeling system assimilates satellite AOD. However, it is also important that AERONET is missing data for the pollution episode between Nov. 6th and Nov. 13th as shown in Fig.4a. It suggests, as one possibility, that PM concentrations were too high during this period that the instrument was not able to retrieve data at that specific coordinates. We modified the discussion on Fig.4 in the revised version.

Text:

Figure 4 shows time series of modeled, MERRA-2 product, VIIRS retrievals, and observed AOD at the AERONET stations (location shown on Fig.1). AOD values at Kanpur, a station in the eastern IGP, were more than 1.0 before the pollution episode and reached up to 2.0 during the episode days, and decreased to values between 0.5 and 1 for the rest of days. The model captured the general trend although missed high AOD's between Nov. 9th and 13th, while MERRA-2 successfully captured the AOD trend through the whole month, including days with enhanced AOD values. This shows that AOD assimilation in MERRA-2 significantly improves AOD predictions. At Jaipur, located in southern IGP, the model overestimated AOD for the first five days of November. During the pollution episode days, the model is biased high compared to MERRA-2 and VIIRS retrievals. AERONET data showed low AOD values before the pollution episode but did not report values during the pollution episode. It suggests, as one possibility, that PM concentrations were too high during this period that the instrument was not able to retrieve data. After the pollution period, AOD values were lower than 0.5, showing relatively low PM concentrations. In general, MERRA-2 showed better performance in terms of NMB (Kanpur: -1.3% and Jaipur:-20.1%) compared with our model (Kanpur:-27.4% and Jaipur: +29.9%). Comparing averaged AOD with VIIRS retrievals for BASE_ANTHRO2X scenario showed lower bias over the IGP (Fig. S7). These results show the need for improved estimates of biomass burning as well as anthropogenic emissions.

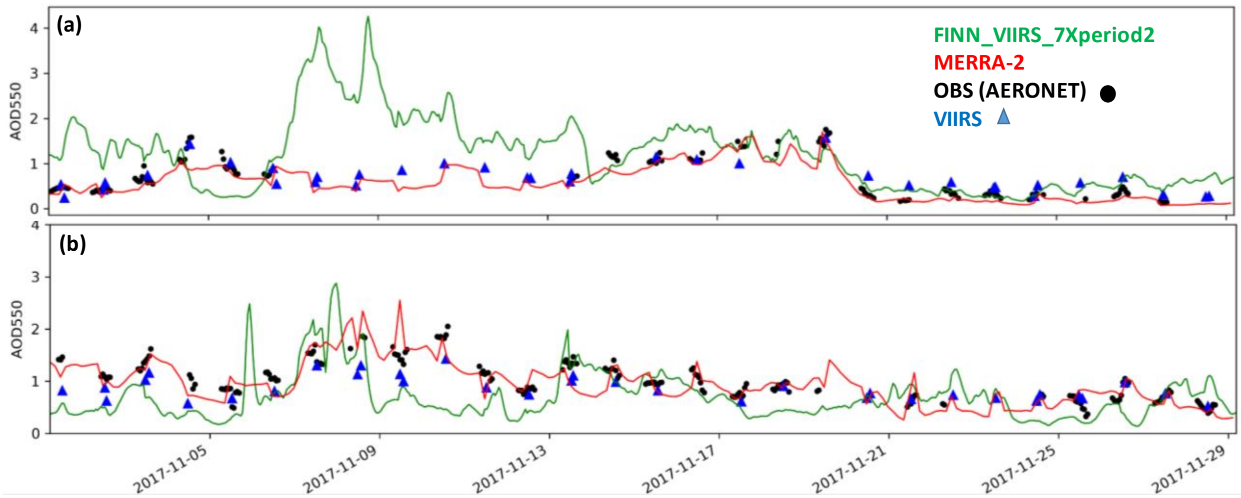


Figure 3 Figure 4 Time series of modeled (green line), VIIRS retrievals (blue triangle), MERRA-2 (red line), and AERONET (black dots) AOD at 550 nm during Nov. 2017 at a) Jaipur, b) Kanpur.

RC2-3: Figure S3 shows that $PM_{2.5}$ concentrations in Punjab were lower than those in Haryana and increasing the fire emissions by a factor of 7 introduced large uncertainties in model simulations as the model $PM_{2.5}$ in Punjab became nearly a factor of 4 higher than the observations. If crop residue burning was the major source of this air pollution episode, one must see the highest observed concentrations in Punjab followed by Haryana and Delhi. Such a pattern exists in the model but not in the observations reflecting that the increasing fire emissions by a factor of 7 is not a reasonable choice. The authors have used back air trajectory to corroborate their assumption that crop residue burning is the major source but backward trajectories only show that the air masses passed over the fire region before arriving at Delhi and are possibly influenced by the fire emissions but they do not tell that agricultural fires are the main source of $PM_{2.5}$ during this episode. Backward trajectory analysis in Figure 7 also shows that $PM_{2.5}$ during the pollution episode was driven by a combination of both the anthropogenic and fire emissions. Thus, this approach presents the danger of attributing missing anthropogenic sources to fire sources and may misinform the mitigation strategies if used for that purpose. Therefore, I recommend the authors to include additional sensitivity simulations exploring the role of anthropogenic emission uncertainties.

Authors Response:

Regarding the discussion about low measured concentrations in Punjab, VIIRS satellite images clearly show massive agricultural fires in this state during November (e.g. Fig.10d). However, we do not see any $PM_{2.5}$ enhancement in observation data over Punjab as the reviewer mentioned (Fig. S3). As a result, we believe the observed values during episode days in Punjab have high uncertainty. We have emphasized this point in the revised version. As discussed above and in back trajectory analysis, all the evidences show that extreme pollution episode has been due mainly to agricultural fires but we have pointed out the importance of anthropogenic emissions too. For example, it was mentioned in the manuscript that short-term increase in anthropogenic emissions (due to social reasons) may have intensified the pollution but quantifying those sources can be the subject of another whole study.

Text:

In Punjab, measured data did not report $PM_{2.5}$ enhancement during the extreme episode, while the model showed very high concentrations after scaling fire emissions by a factor of 7. However, VIIRS satellite images (e.g. Fig. 9d) clearly show massive agricultural fires in this state during November and its signals were expected in the measured data.

RC2-4: Fig 4 and related discussion: In addition to the AOD, could you please evaluate the Angstrom exponent to examine if there any differences in the abundance of fine and coarse mode particles and if the model was able to capture those variations. Can you also plot VIIRS AOD in Figure 4 to see if the satellite observed an AOD of 4 in Jaipur?

Authors Response:

We thank the reviewer for the comment. VIIRS AOD is added to Fig. 4 in the revised version and show low AOD values over Jaipur. We modified the text as described above.

Regarding Angstrom Exponent, we added the following discussion to the paper and added the figures (shown below) to the supplementary document:

Text:

We also looked at Angstrom Exponent (AE) at Jaipur and Kanpur to understand if the model captured the mode of the particles (Fig. S8). Over Jaipur the model is biased high compared to AERONET data (NMB: 30%) and shows more finer aerosols. After Nov. 20th, both AERONET and VIIRS retrievals suggest the dominance of coarser aerosols, while the AE for the model does not follow the same trend. However, $PM_{2.5}/PM_{10}$ ratio shows more coarse aerosols compared to the rest of the month (Fig. S9). Over Kanpur, the model AE is biased high (NMB: 50.8%). On the other hand, the model shows closer AE values to VIIRS retrievals. For example, both the model and VIIRS retrieval show similar reduction in AE on Nov. 8th and 9th. Kumar et al. (2014) also reported slight AE overestimation in WRF-Chem during a pre-monsoon dust storm at Kanpur and Jaipur. Furthermore, model and AERONET have variational trend while MERRA-2 is smooth during the whole month at both Jaipur and Kanpur.

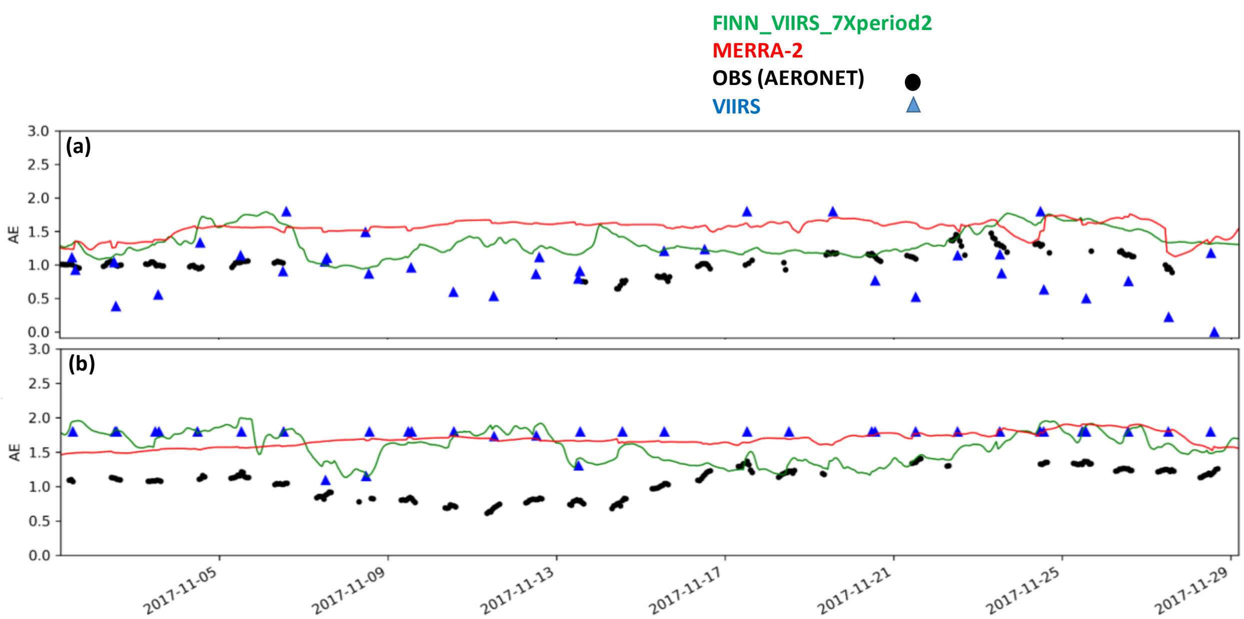


Figure 4 Time series of modeled (green line), VIIRS retrievals (blue triangle), MERRA-2 (red line), and AERONET (black dots) Angstrom Exponent during Nov. 2017 at a) Jaipur, b) Kanpur.

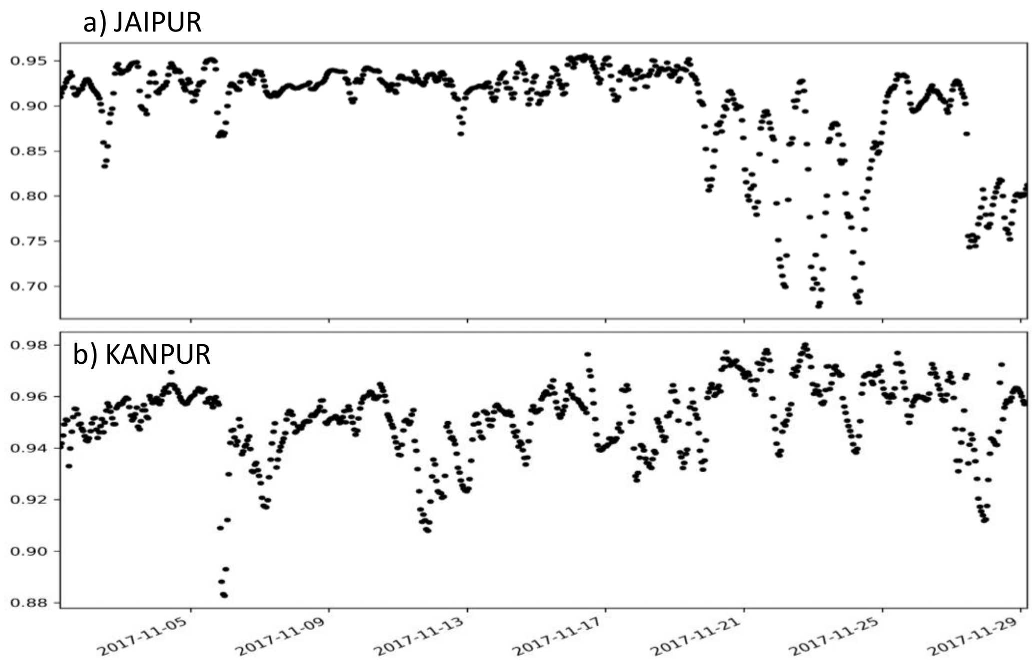


Figure 5 Modeled PM_{25}/PM_{10} ratio (Base scenario) at a) Jaipur and b) Kanpur

RC2-5: Fig 5/Table 3: Could you please add a few panels in Figure 5 showing the evaluation against the CPCB data?

Authors Response:

Fig.6 in the paper shows the box and whisker plots for CPCB stations in Delhi. In general during the whole paper, we show the results only at one station (i.e. US Embassy) when we look at time-series and we show daily box and whisker plots when we look at all CPCB stations. As the model resolution is 15km, we usually see more than one measured CPCB stations are located in one model grid cell. For example, 17 stations in Delhi are located in only 6 grid cells; repetition affects the scatter plot (Fig. 6a below). We also observe lower variability in box and whisker plots of the model compared to observation data due to same reason. In other words, scatter plots will not provide enough insights when considering all individual stations. To show the spatial performance of the model, we plot the scatter plot for averaged concentration of different states. Below, we show the scatter plot for Delhi, Haryana, and Rajasthan, which reveals the good spatial performance of the model (Fig. 6b below). Adding data from Punjab to this plot (Fig. 6c below) significantly degrades the performance. The reason is due to extremely high bias in Punjab data. Punjab observation data doesn't seem to be right as it doesn't show any signal of the pollution episode while satellite data show huge amount of agricultural fires during those days. We added the below scatter plots in the supplementary. However, we think scatter plots were better tools if we had more spatial data (e.g. a gridded dataset).

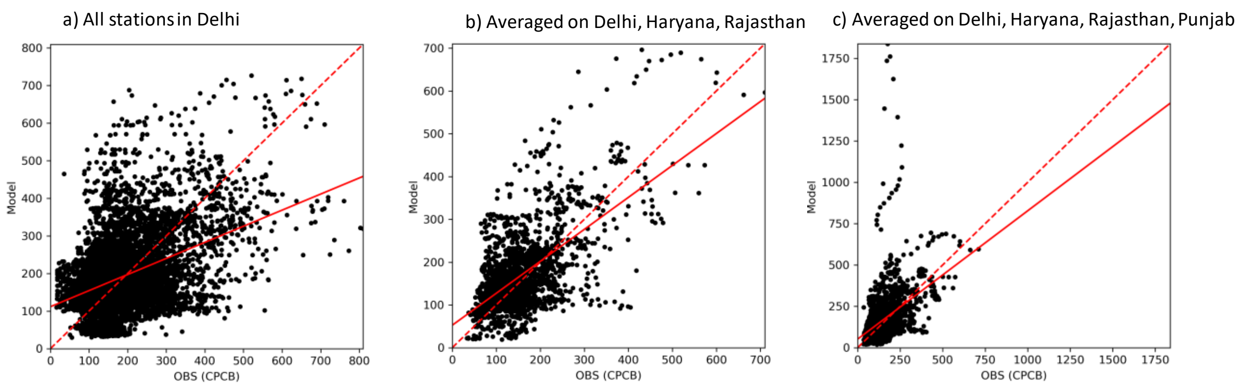


Figure 6 Scatter plots for a) all stations in Delhi combined b) averaged concentrations in Delhi, Haryana, and Rajasthan c) averaged concentrations in Delhi, Haryana, Rajasthan, and Punjab. Filters are applied to CPCB data.

RC2-6: Line 301: I think the model observation comparison for the non-episode periods looks good because of the scale of Figure 5a. A zoom into the figure 5a shows that on several occasions, the model showed a bias of up to 100 ug/m3 even in the non-episode period.

Authors Response:

We appreciate the comment. We agree that for some typical days the error is high, which can be related to the accuracy of anthropogenic emissions and we have mentioned that when presenting the results between lines 320-326. However, statistics for the whole November after excluding days between Nov. 7th and Nov. 10th (4 days), also show fair results as shown below.

Table 1 Statistics for all days in November 2017 after excluding extreme days of Nov. 7th, 8th, 9th, 10th compared with data from CPCB stations in Delhi

Scenario	Hourly Mean	Hourly Standard Deviation	24-hours R	24-hours RMSE	24-hours NMB	24-hours NME	24-hours MB	24-hours ME
CPCB Obs data	215.26	97.58						
FINN_VIIRS_7Xperiod2	209.91	104.94	0.7	55.11	-2.44	18.96	-5	38.94

RC2-7: Line 308: Are you referring to the model biases relative to MERRA-2 here? If yes, is it reasonable to do so given large biases in MERRA-2 simulated PM2.5 itself as shown in Figure 5a?
Authors Response:

We appreciate the reviewer's concern. We agree that MERRA-2 may have large biases, as we saw in Fig. 5a. However, in this paper, we use MERRA-2 as an observation package when we do not have any other data to evaluate our results. Looking at domain wide $PM_{2.5}$ concentrations is one of those cases. We assume that enhancing MERRA-2 modeling system by data assimilation makes it a fair benchmark.

RC2-8: Line 368-369: Why do you attribute this error only to transport and not to uncertainties in anthropogenic emissions or other physical processes in the model.

Authors Response:

We thank the comment. For Nov. 8th, the back trajectory was passing through anthropogenic sources; so, we hypothesized that the model may have missed major fire emission due to transport. But, we agree with the reviewer that other mentioned factors can be important, as well. We have modified that sentence to:

Text:

The model underestimated $PM_{2.5}$ concentrations on Nov. 8th, which can be partly related to errors in transport as the trajectories for Nov. 8th_12 crossed eastern parts of Punjab. However, other physical processes or lower anthropogenic emissions can also be responsible for low bias.

RC2-9: Figure 8: Could you please add PBL height to these panels to help understand whether the smoke was injected in the free troposphere.
Authors Response:

We thank the reviewer for the comment. Below, we added the PBLH line to the cross sections (white line). The line is not obvious on 00 UTC times due to very low extent of the PBL. On Nov. 6th-12, we see very low PBL upwind of Delhi and significant amount of smoke above boundary layer. Therefore, these findings accompanied with Fig. 13b supports the argument that the plume rise in the model released the emissions too high or the model did not mix the smoke down fast enough. We modified the text in the revised version.

Text:

To further understand the regional scale transport of the smoke plumes, we plotted cross section of $PM_{2.5}$ over the path from Punjab through Delhi (Fig. 8, path line shown in Fig. 1). $PM_{2.5}$ concentrations showed typical values on Nov. 5th_00 although they still exceeded the standard limits. On Nov. 5th_12, concentrations significantly increased over Punjab area because of fires and the winds brought them on a path towards Delhi. The Punjab's smoke did not completely cross Delhi yet on Nov. 6th as back trajectories for 00 and 12 UTC hours also showed the effects of anthropogenic emissions and fires in eastern Delhi. On the other hand, a significant amount of smoke was above the boundary layer as shown in Nov. 6th_12 panel. Due to shifting winds on Nov. 7th (as shown in Fig. 2), the smoke upwind of Delhi blew over Delhi and led to extremely high concentrations. Although the model captured the median in Nov. 7th, it missed the maximum extent of observed values. Cross sections on Nov. 8th, 9th, and 10th show the residual Punjab's smoke in the boundary layer, while we saw the model underestimated $PM_{2.5}$ concentrations on these days. Measured $PM_{2.5}$ concentrations over Delhi show a decreasing trend between Nov. 8th and Nov. 10th (Fig. 6). Vertical profiles for the base scenario also show the model captured high biomass burning emission period on Nov. 6th (Fig. 12). However, it also showed high amounts of smoke above the PBL. Cross sections for Nov. 11th to Nov 14th can be found in the supporting document (Fig. S12). These results suggest that plume rise in the model released the emissions too high or the model did not mix the smoke down fast enough. Vijayakumar et al. (2016) showed agricultural fires can transport via upper troposphere and subside over Delhi using ECMWF map. Social reasons can also be important as the first reaction of people during hazy days is to drive to work which directly (exhaust emission) and indirectly (road dusts) worsen air pollution.

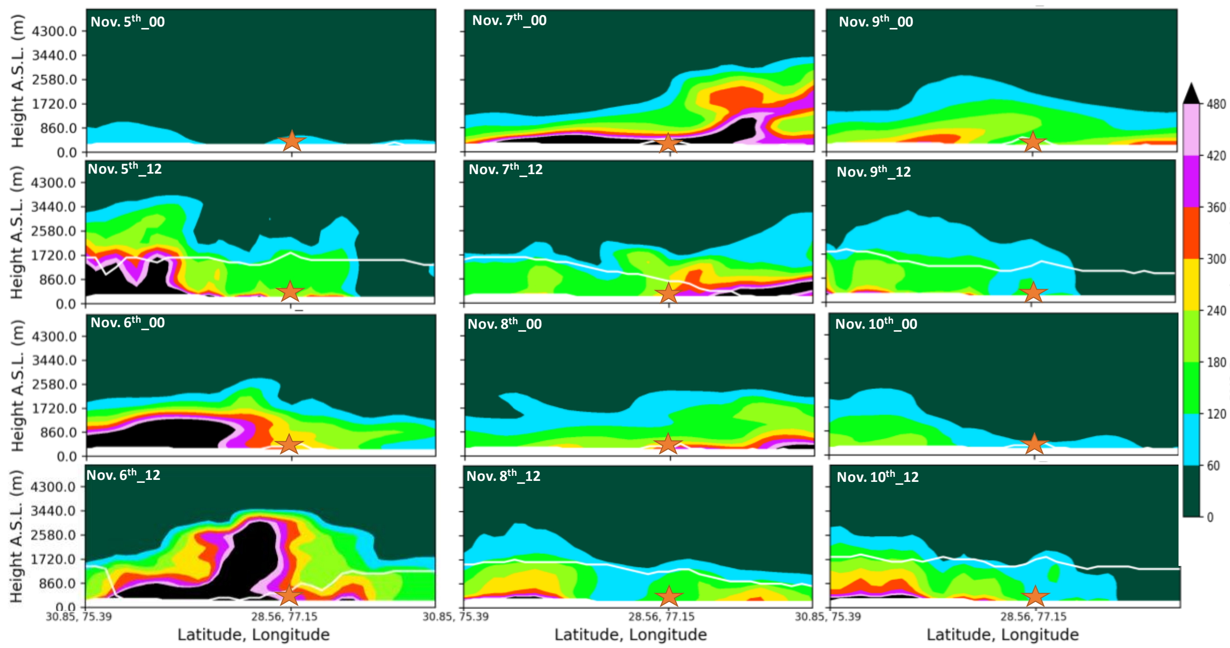


Figure 7 Figure 8 Vertical cross section of $PM_{2.5}$ concentration through the path shown in Fig. 1 for the days between Nov. 5th and Nov. 10th. For each day, two snapshots are shown at 00UTC (5:30AM local time) and 12UTC (5:30PM local time). The orange star shows the location of Delhi through the path. White line shows the PBL height across the path

Minor comments:

RC2-10: Line 100: Replace '*' with the 'x' and also elsewhere in the paper where you describe the resolution.

Authors Response:

We replaced all of them.

RC2-11: Line 194-195: Have you applied any filtering criteria to the CPCB data?

Authors Response:

We didn't apply any filter to this data as we relied on quality control done by CPCB (https://cpcb.nic.in/quality-assurance-quality-control/). However, we studied how applying the following filters, done by Jena et al. (2020) and Kumar et al. (2020), change the dataset consisting of total 12768 hourly data points:

Filter 1: Remove less than 10 μgm^{-3} instances: removes 31 data-points

Filter 2: Remove the hourly difference between 100 (or 150 or 200) μgm^{-3}: removes 186 (or 71 or 31) hourly-data

Filter 3: Remove values more than 200 (400) μgm^{-3} right after NAN value: 33 (19). It basically removes data for Nov. 9th as it was applied after filter #2.

We found that the order of applying these filters is important. Below, statistics and timeseries for different orders of filters are presented. Order of filters (1,2,3) removes data for Nov. 9th and

significantly improves the model performance over Delhi. We added these findings in the supplementary document and described in the revised version.

Text:

No additional quality control filters, other than the ones by CPCB (https://cpcb.nic.in/quality-assurance-quality-control/), were applied. We evaluated the results after applying the filters proposed by other studies (e.g. Kumar et al. (2020)); they had slight impacts on statistics (shown in the supporting document).

Table 2 Effect of applying filters to CPCB data on $PM_{2.5}$ statistics in Delhi

Province	Hourly Obs. Mean (±std) (μgm^{-3})	Hourly Model Mean (±std) (μgm^{-3})	24-hours NMB (%)	24-hours NME (%)	24-hours R (%)
CPCB-Delhi	255.5 (±146.6)	213.9 (±113.9)	-16.6	27.6	0.48
Only filter 3	248.4 (±140.3)	214.5 (±114.5)	-13.9	26.4	0.49
Filter123	215.5 (±95.5)	214.8 (±115.2)	-1.9	23.6	0.64
Filter132	248.6 (±140.8)	214.6 (±114.5)	-13.9	26.4	0.49

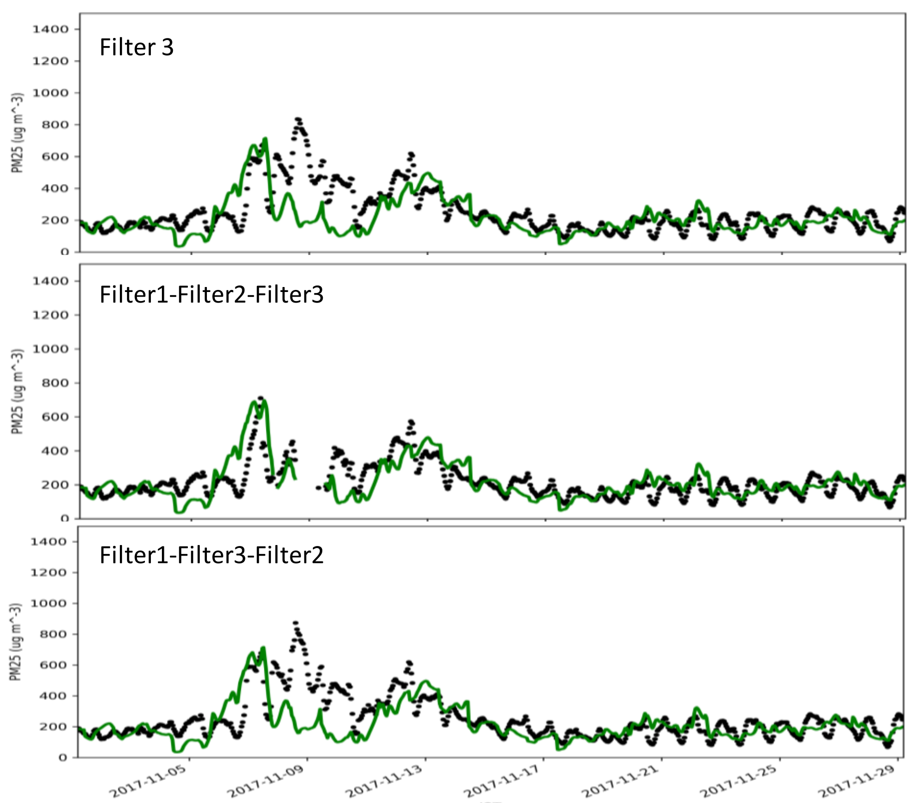


Figure 8 Effect of applying additional filters to CPCB data on averaged $PM_{2.5}$ timeseries in Delhi

RC2-12: Equation (1): I assume this equation is used to calculate MERRA-2 $PM_{2.5}$ and not WRF-Chem.

Authors Response:

Yes, it is to calculate $PM_{2.5}$ for MERRA-2 and we used WRF-Chem diagnosed $PM_{2.5}$ variable directly.

RC2-13: Line 288-289: But the underestimation could also be because of the underestimation of emissions from Delhi.

Authors Response:

Yes, we modified the sentence:

Text:

This suggests either low local anthropogenic emissions in Delhi or some missing pollution sources upwind of Delhi that were not included in the emission estimates led to underestimation.

RC2-14: Line 322-333: This is not true as EDGAR-HTAP provides monthly varying emissions with higher emissions in winter.

Authors Response:

We appreciate the reviewer and apologize for this mistake. All the experiment have been done using monthly EDGAR-HTAP data and it was just a drafting mistake. It has been removed in the revised version.

RC2-15: Figure 6: It would be useful to mark period 1 and period 2 in the figure.

Authors Response:

Thanks for the comment. We added period1 and period2 that have been used for emission modifications to Fig. 6.

RC2-16: Line 365: Change “lower” to “smaller”.

Authors Response:

We changed that.

RC2-17: Line 567: change “intensify” to “accuracy”.

Authors Response:

We changed the “intensify the accuracy” to “improve the accuracy”

References

Beig, G., Srinivas, R., Parkhi, N. S., Carmichael, G., Singh, S., Sahu, S. K., Rathod, A., and Maji, S.: Anatomy of the winter 2017 air quality emergency in Delhi, Science of The Total Environment, 681, 305-311, 2019.
Jat, R., Gurjar, B. R., and Lowe, D.: Regional pollution loading in winter months over India using high resolution WRF-Chem simulation, Atmospheric Research, 105326, 2020.
Jena, C., Ghude, S. D., Kulkarni, R., Debnath, S., Kumar, R., Soni, V. K., Acharja, P., Kulkarni, S. H., Khare, M., and Kaginalkar, A. J.: Evaluating the sensitivity of fine particulate matter (PM 2.5) simulations to chemical mechanism in Delhi, Atmospheric Chemistry and Physics Discussions, 1-28, 2020.
Kulkarni, S. H., Ghude, S. D., Jena, C., Karumuri, R. K., Sinha, B., Sinha, V., Kumar, R., Soni, V., and Khare, M.: How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi?, Environmental Science & Technology, 54, 4790-4799, 2020.
Kumar, R., Barth, M., Madronich, S., Naja, M., Carmichael, G., Pfister, G., Knote, C., Brasseur, G., Ojha, N., and Sarangi, T.: Effects of dust aerosols on tropospheric chemistry during a typical pre-monsoon season dust storm in northern India, Atmospheric Chemistry & Physics, 14, 2014.
Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., Kulkarni, S., Sperati, S., Soni, V. K., and Nanjundiah, R. S.: Enhancing Accuracy of Air Quality and Temperature Forecasts During Paddy Crop Residue Burning Season in Delhi Via Chemical Data Assimilation, Journal of Geophysical Research: Atmospheres, 125, e2020JD033019, 2020.
Vijayakumar, K., Safai, P., Devara, P., Rao, S. V. B., and Jayasankar, C.: Effects of agriculture crop residue burning on aerosol properties and long-range transport over northern India: a study using satellite data and model simulations, Atmospheric research, 178, 155-163, 2016.

RC3:

In this study, the authors used the WRF-Chem to simulate the pollution episode during Nov 2017 over New Delhi and evaluated the impacts of biomass burning emissions, long-range transport of dust, and dust emissions on simulated $PM_{2.5}$. The model was evaluated by comparing simulated meteorological parameters and simulated AOD and $PM_{2.5}$ with MERRA2 data and observational data. This study provides information on the sources that contribute to the severe PM pollution during Nov 2017. The paper is well organized but improvements in presentation are needed.

Authors Response:

We appreciate the reviewer for pointing to important issues. We try to address the comments and concerns here and below:

My comments are as follows.

RC3-1: In the design of the simulations, why increasing the emissions by 5, 7 or 10 times? Are these numbers chosen only to get a better simulation of PM in Delhi?

Authors Response:

We thank the reviewer for the point. Yes, we used simulation results in Delhi as the criteria for choosing the proper scaling factor. Moreover, the increasing factors were chosen arbitrarily and we have mentioned this as the limitation of our study in the revised version. In this study, we intended primarily to show the bias in biomass burning emission inventory is not systematic and clarify high uncertainty for extremely polluted days. Another bias correction study is required to find the relation between highly polluted days and optimized increasing factor to modify biomass burning emission inventories.

Text:

The choice of the scaling factor for increasing fire emissions was arbitrary in this study. Due to scarcity of observation data, we were not able to apply complicated mathematical scaling techniques based on data assimilation to scale the fire emissions (Saide et al., 2015).

RC3-2: Why this study chose to evaluate the impacts from only biomass burning and dust? How about other anthropogenic emissions which is also important source to severe $PM_{2.5}$ events.

Authors Response:

We focused on biomass burning and dust emissions in this study based on previous studies during this period (e.g. Beig et al. (2019)). However, we acknowledge the importance of anthropogenic emissions since emissions due to heating also increase as the weather gets cold during Oct. and Nov. We conducted another scenario, in which we increased the particles anthropogenic emissions by a factor of 2 (ID: BASE_ANTHRO2X). This modification increased $PM_{2.5}$ concentrations in Delhi up to ~150 μgm^{-3}, which led to overestimation (in contrast to underestimation in base scenario) at most of non-episode days (time-series shown below). Although this scenario did not help capturing concentrations

during the episode, it confirms the need for better anthropogenic emissions. On the other hand, it increased the AOD bias over southern IGP while reduced the bias over IGP (bias map shown below). These results suggest anthropogenic emission inventories have higher bias over IGP compared with non-IGP regions. However, we acknowledge the importance of having dynamic (daily) anthropogenic emission inventory.

Text:

Although different meteorological parameters can be responsible for the biases, accuracy of anthropogenic emissions is important. For example, recent local anthropogenic emission inventories developed for Delhi have higher particle emissions than in the regional inventory used in this study, which impacts modeled $PM_{2.5}$ concentrations for typical days (Kulkarni et al., 2020). We conducted BASE_ANTHRO2X scenario to investigate the effect of uncertainties in the anthropogenic emissions. This scenario increased PM2.5 concentrations in Delhi up to ~150 µgm-3, which led to overestimation (in contrast to underestimation in base scenario) at many of non-episode days (Fig. S7). Although this scenario did not help in capturing the high concentrations during the episode, it confirms the need for better anthropogenic emissions. On the other hand, it reduced the bias over IGP (Fig. S7). These results point out the need for best estimates of emissions of both anthropogenic and biomass. .Maps also show that averaged PM2.5 concentrations over most of India were higher than the air quality standard.

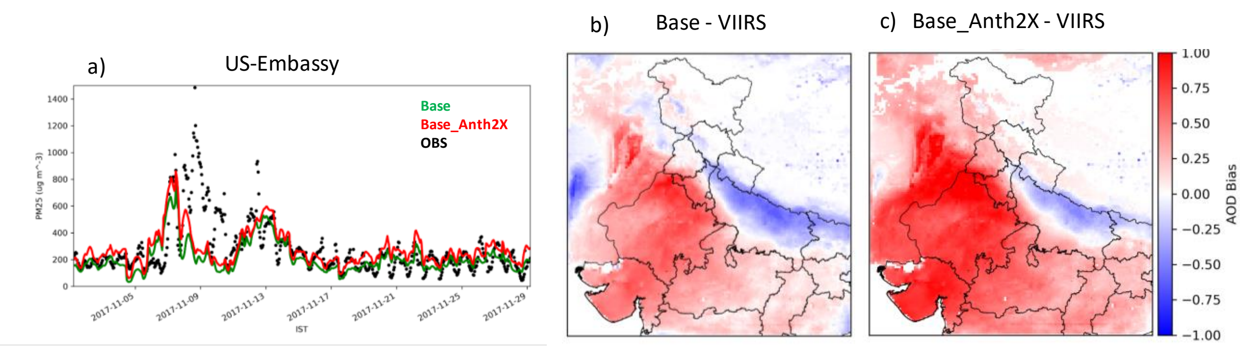


Figure 1 a) Timeseries for $PM_{2.5}$ concentration at the location of US embassy using Base scenario and Base_Anth2X scenario B) Bias of AOD at 550nm averaged over November 2017 base on b) base scenario c) base scenario with 2 times more anthropogenic particle emissions (ID: Base_Anth2X)

RC3-3: This study simulates the haze event during Nov 2017 by adjusting boundary conditions and emissions, how about other haze events in India? How to apply the findings of this study in the simulation of other haze events in India?

Authors Response:

We appreciate the reviewer for this important point. The main purpose of the study is to investigate the sensitivity of model predictions to the main inputs into the model. Prediction of extreme pollution events is important as they have major impacts on people and also make a strong impression regarding the capabilities of models. However, extreme events are hard to predict because they are often heavily impacted by episodic emission sources. Here we take the approach of systematically exploring the impacts of different boundary conditions, dust, fire and anthropogenic emissions on the predictions of the pollution episode in November 2017. We feel this evaluation of inputs is needed to understand the extent that which the forward model can be configured to capture the events. A contemporary way to try to capture such events in prediction mode is to employ data assimilation. The data assimilation results compensate for deficiencies in the inputs as well as structural problems

within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability.

Below, please also find other findings:

- **We showed that biomass burning emission inventories miss some small fire emission and introduced a new technique to use satellite data to fill these missing sources.**
- **We showed that biomass burning emission inventories occasionally underestimate emissions in hazy events up to 7 times lower, where bias correction techniques need to be applied.**
- **We showed either the plume rise in the model release the agricultural fire emissions too high or the model does not mix the smoke down fast enough. These should be considered in future hazy event simulations.**
- **We found Secondary aerosols comprise more than half of the particles in Delhi. It suggests simple aerosol modules like GOCART cannot simulate the actual speciation of particles in Delhi.**

Text:

The main purpose of this study is to investigate the sensitivity of model predictions to the main inputs into the model. Prediction of extreme pollution events is important as they have major impacts on people and also make a strong impression regarding the capabilities of models. However, extreme events are hard to predict because they are often heavily impacted by episodic emission sources. Here we take the approach of systematically exploring the impacts of different boundary conditions, dust, fire and anthropogenic emissions on the predictions of the pollution episode in November 2017. A contemporary way to try to capture such events in prediction models is to employ data assimilation (Kumar et al., 2020). The data assimilation results compensate for deficiencies in the inputs as well as structural problems within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability.

References

Beig, G., Srinivas, R., Parkhi, N. S., Carmichael, G., Singh, S., Sahu, S. K., Rathod, A., and Maji, S.: Anatomy of the winter 2017 air quality emergency in Delhi, Science of The Total Environment, 681, 305-311, 2019.

Kulkarni, S. H., Ghude, S. D., Jena, C., Karumuri, R. K., Sinha, B., Sinha, V., Kumar, R., Soni, V., and Khare, M.: How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi?, Environmental Science & Technology, 54, 4790-4799, 2020.

Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., Kulkarni, S., Sperati, S., Soni, V. K., and Nanjundiah, R. S.: Enhancing Accuracy of Air Quality and Temperature Forecasts During Paddy Crop Residue Burning Season in Delhi Via Chemical Data Assimilation, Journal of Geophysical Research: Atmospheres, 125, e2020JD033019, 2020.

Saide, P. E., Peterson, D. A., da Silva, A., Anderson, B., Ziemba, L. D., Diskin, G., Sachse, G., Hair, J., Butler, C., and Fenn, M.: Revealing important nocturnal and day-to-day variations in fire smoke emissions through a multiplatform inversion, Geophysical Research Letters, 42, 3609-3618, 2015.

Improving regional air quality predictions in the Indo-Gangetic Plain - Case study of an intensive pollution episode in November 2017.

Behrooz Roozitalab[1,2], Gregory R. Carmichael[1,2], Sarath K. Guttikunda[3]

[1]Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA, USA,
[2]Chemical and Biochemical Engineering, University of Iowa, Iowa City, IA, USA
[3]Urban Emissions, New Delhi, India

Correspondence to: B. Roozitalab (behrooz-roozitalab@uiowa.edu) and G. R. Carmichael (gcarmich@engineering.uiowa.edu)

Field Code Changed

Abstract. Indo-Gangetic Plain (IGP) experienced an intensive air pollution episode during November 2017. Weather Research and Forecasting model with Chemistry (WRF-Chem), a coupled meteorology–chemistry model, was used to simulate this episode. In order to capture $PM_{2.5}$ peaks, we modified input chemical boundary conditions and biomass burning emissions. CAM-Chem and MERRA-2 global models provided gaseous and aerosol chemical boundary conditions, respectively. We also incorporated VIIRS active fire points to fill missing fire emissions in FINN and scaled by a factor of seven for an 8-days period. Evaluations against various observations indicated the model captured the temporal trend very well although missed the peaks on Nov. 7th, 8th, and 10th. Modeled aerosol composition in Delhi showed Secondary Inorganic Aerosols (SIA) and Secondary Organic Aerosols (SOA) comprised 30% and 27% of total $PM_{2.5}$ concentration, respectively, during November, with a modeled OC/BC ratio of 2.72. Back trajectories showed agricultural fires in Punjab were the major source for extremely polluted days in Delhi. Furthermore, high concentrations above the boundary layers in vertical profiles suggested either the plume rise in the model released the emissions too high, or the model did not mix the smoke down fast enough. Results also showed long-range transported dusts did not affect Delhi's air quality during the episode. Spatial plots showed averaged Aerosol Optical Depth (AOD) of 0.58 (±0.4) over November. The model AODs were biased high over central India and low over eastern IGP, indicating improving emissions in eastern IGP can significantly improve the air quality predictions. We also found high ozone concentrations over the domain, which indicates ozone should be considered in future air quality management strategies alongside particulate matters.

1. Introduction

Ambient air pollution remains a major environmental issue, even after significant worldwide efforts starting after the deadly smog of London in 1952. It is the fifth-ranking risk of death and a major threat to climate and ecosystem (Cohen et al., 2017;Ramanathan and Carmichael, 2008;Sitch et al., 2007). Air pollution contains many species; particulate matter (PM) is currently the air pollutant of most concern, especially in developing countries like India. India is an emerging economy with burgeoning population that has accelerated its industrial activities in the last three decades, leading to wide spread air pollution and resulting adverse health effects. There are many Indian cities on the list of most polluted cities of the world (World-Bank, 2018;Guttikunda et al., 2014;WHO, 2016). Studies show that ozone and particulate matter with diameter less than 2.5 micron ($PM_{2.5}$), are attributed to more than one million individual premature deaths in India (Cohen et al., 2017;HEI, 2018).~~ ~~David et al. (2019) found that anthropogenic emissions within India led to about 80% of the total premature death due to $PM_{2.5}$ in India. Furthermore as industrial activities are growing, emissions are increasing too; health impacts attributed to long-term exposure air pollution are predicted to increase based on current policies (Conibear et al., 2018a) ~~(Conibear et al., 2018a;Ghude et al., 2016)~~.

Short-term extreme pollution events lead to increased hospital admissions and mortalities (Anenberg et al., 2018;Rajak and Chattopadhyay, 2019). Forest and agricultural fires, dust storms, increased local activities, and stagnant meteorological conditions are major contributing factors in these air pollutions episodes (Beig et al., 2019;Jethva et al., 2018). While forecasting models help authorities to notify people of these extreme pollution events, hindcasting models help scientists improve the capabilities of the models to predict pollution events, identify the main responsible factors causing these events, and inform policy makers as they develop pollution control strategies. However, the ability of air quality models for simulating short-term events highly depends on the quality of input chemical data (i.e. emissions). For example, the total amount of global fire emissions can differ by a factor of 3-4 based on the emission inventory used (Pan et al., 2020). Furthermore, dust storms can travel long distances and influence another region's air quality (Ashrafi et al., 2017;Beig et al., 2019). David et al. (2019) attributed about 16% of total premature $PM_{2.5}$-related death to emissions outside India. Moreover, studies of ~~-~~ Black Carbon (BC) in southern Asia revealed that local emissions in western India can affect eastern and southern regions' air quality (Kumar et al., 2015a). As a result, global models, which provide boundary conditions needed by regional air quality models, can significantly affect the simulated results (He et al., 2019).

The Indo-Gangetic Plain (IGP) experiences high levels of air pollution during the post monsoon season (October to early December) due to stagnant meteorological conditions and higher air pollution emissions (Adhikary et al., 2007;Marrapu et al., 2014). Figure 1a shows the averaged Aerosol Optical Depth (AOD) retrieved from Visible Infrared Imaging Radiometer Suite (VIIRS) remote sensing instrument during November 2017 over northern India. The IGP region has the highest AOD values with the largest values in the north-western parts, which is mostly due to crop residue burning (Beig et al., 2020;Jethva et al., 2018;Liu et al., 2018;Venkataraman et al., 2018;Vijayakumar et al., 2016). Kulkarni et al. (2020) found India's north-western agricultural fires could contribute up to 75% of Delhi's $PM_{2.5}$ concentration.

Not only is there significant spatial variation over the IGP, but also $PM_{2.5}$ concentrations change on a daily basis (Fig. 1c). Delhi, the capital of India with annual average $PM_{2.5}$ concentration of 120 μgm^{-3} (Amann et al., 2017), experienced a severe extreme air pollution during November 2017. Figure 1c shows the daily averaged $PM_{2.5}$ concentrations measured with the US-EPA instrument located at the US Embassy in Delhi. Daily $PM_{2.5}$ concentrations reached values more than 900 μgm^{-3}, 15 (37.5) times higher than 24-hour averaged Indian standards (World Health Organization (WHO) guidelines) (WHO, 2006). However, it is clear that no day is compliant with the air quality standard values. After this extreme pollution episode, the Indian government officially initiated a comprehensive air quality plan called the National Clean Air Programme (NCAP) to reduce the air pollution (MoEF&CC, 2019). Different groups have studied this period. Dekker et al. (2019) attributed carbon monoxide (CO) accumulation between Nov. 11th and Nov. 14th to stagnant meteorological conditions; specifically, low wind speeds and shallow atmospheric boundary layers. Moreover, they argued regional air pollution transport was mostly responsible for this extreme pollution episode (Dekker et al., 2019). However, Beig et al. (2019) concluded biomass burning emissions after post-monsoon crop productions, accompanied with long range transported dust from Middle East led to very high pollution levels although stagnant condition favoured it.

While the current focus of research groups and governments is on PM, ozone concentrations also show high values during the post monsoon season. Figure 1d shows measured daily ozone concentrations at one CPCB station in Delhi; concentrations exceeded India's ozone air quality guidelines. Moreover, the ozone concentrations followed a similar daily variation as PM during November 2017 (Fig. 1d). As a result, extreme pollution episodes cause not only PM-related health issues but also increase the risk of Chronic Obstructive Pulmonary Disease (COPD) (the most important health outcome of ozone pollution) (Conibear et al., 2018b;EPA, 2013).

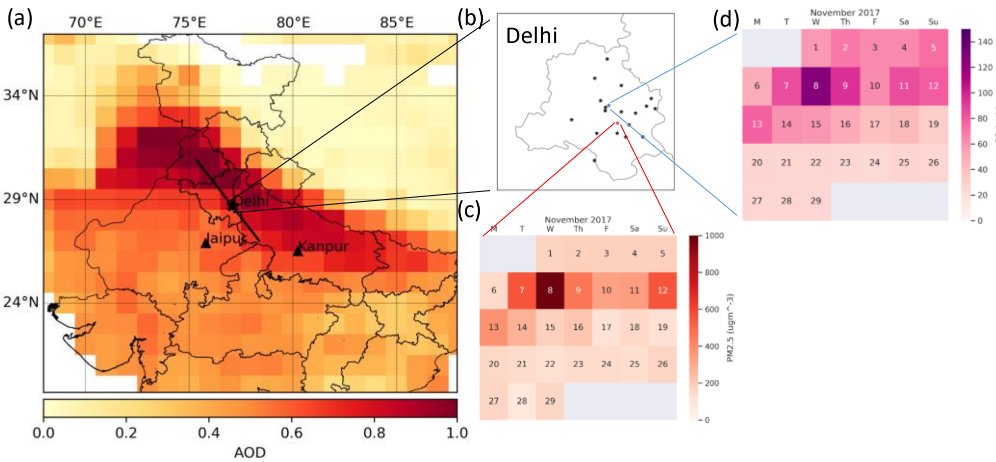


Figure 1 WRF-Chem modeling domain, ground measurement stations, and observed air quality: a) modeling domain and location of Delhi (∗) and AERONET stations at Jaipur and Kanpur (▲), and underlying VIIRS AOD (550nm) averaged over November 2017, the black line also shows the path that was used for vertical cross section analysis. b) location of CPCB stations (black stars) and US Embassy station (red star), c) Calendar map of averaged daily $PM_{2.5}$ concentration measured at US Embassy, d) Calendar map of averaged daily ozone concentration measured at North Campus, DU

Models usually underestimate the concentrations during extreme pollution periods unless they apply chemical data assimilation (Dekker et al., 2019;Kulkarni et al., 2020;Kumar et al., 2015b;Kumar et al., 2020) ~~(Dekker et al., 2019;Kulkarni et al., 2020;Kumar et al., 2015b)~~. Moreover, there are different input data in terms of chemical boundary conditions and fire emissions that can affect air quality modeling results (He et al., 2019). The main purpose of this study is to investigate the sensitivity of model predictions to the main inputs into the model. Prediction of extreme pollution events is important as they have major impacts on people and also make a strong impression regarding the capabilities of models. However, extreme events are hard to predict because they are often heavily impacted by episodic emission sources. Here we take the approach of systematically exploring the impacts of different boundary conditions, dust, fire and anthropogenic emissions on the predictions of the pollution episode in November 2017. A contemporary way to try to capture such events in prediction model is to employ data assimilation (Kumar et al., 2020). The data assimilation results compensate for deficiencies in the inputs as well as structural problems within the models. But the effectiveness of data assimilation improves as the capabilities of the forward model improves. Therefore, our results are also important for those using data assimilation to improve predictability. ~~The main objective of this study is to investigate the impacts of different global input data sets on improving modeled $PM_{2.5}$ concentrations over the IGP during Nov. 2017. We hypothesize that incorporating different available datasets with each other can improve modeling results.~~ In this study, we use the Weather Research and Forecasting model coupled to Chemistry (WRF-Chem). Through a series of sensitivity experiments, we evaluate the impacts of biomass burning emissions coming from FINN and QFED, chemical boundary conditions retrieved from MOZART, CAM-chem, CAMS, and MERRA-2 global models, role of incorporating VIIRS active fire hot spots to improve biomass burning emission inventories and global models to improve chemical boundary conditions, and changes in dust and anthropogenic emissions, on modeled $PM_{2.5}$ concentration during November 2017. We also evaluate ozone predictions.

This paper is organized as follows. First, the WRF-Chem model configuration, sensitivity experiments, and the observation datasets, including ground measurements and satellite data, are described. Then, after evaluating the model performance for the best experiment, the impacts of using different datasets as input data on modeled $PM_{2.5}$ concentrations during November 2017 are analyzed and discussed.

2. Methods

2.1. WRF-Chem configuration:

WRF-Chem is a numerical modeling framework that solves transport, chemistry, and physics of the atmosphere (Grell et al., 2005). The online interaction between meteorology, thermodynamic processes, and atmospheric chemistry makes it a powerful and reliable model in the community. WRF-Chem model (Version 4.0) with 1-domain centered on Delhi with 15 km horizontal grid resolution and 39 vertical levels was used in this study. The domain was set to be big enough to include the north-west biomass burning and urban emission sources in the simulation process as they are shown to be contributors to poor air quality in the region in previous studies (Amann et al., 2017). In the following, we present the model configuration for the base scenario (ID: FINN_VIIRS_7Xperiod2).

National Center for Environmental Prediction (NCEP) Global Forecasting System (GFS-FNL) 1~~*~~x1-degree and 6-hours spatial and temporal resolution meteorological-fields (https://rda.ucar.edu/datasets/ds083.2/) were used as initial and boundary conditions for the meteorology. Community Atmosphere Model with Chemistry (CAM-chem) data (Buchholz, 2019) with horizontal resolution of 0.9~~*~~x1.25 degree and 56 vertical levels provided chemical boundary conditions for gaseous species. MERRA-2 reanalysis data with 0.625~~*~~x0.5 degree horizontal and 72 vertical model levels were used for aerosol species (Bosilovich et al., 2015). However, input data have uncertainties and small uncertainties in nonlinear governing equations of numerical weather predictions can lead to non-negligible errors in results (Xiu, 2010). As a result, re-initialization of NWP models is suggested instead of free runs (Abdi-Oskouei et al., 2020). In this study, the model ran for 30 hours each day starting at 00Z while the first 6 hours data were discarded to account for daily spin-up. Meteorological initial and boundary conditions, and chemical boundary conditions were re-initialized daily at 00Z using global models. However other than for the first cycle in which global models provided initial chemical conditions data, chemical fields from the previous cycle were used as the next cycle's initial chemical conditions. Table 1 summarizes the WRF-Chem physical and chemical configuration options.

Studies have shown improvements for ozone simulations in Delhi using more complicated chemistry mechanisms like Model for Ozone and Related chemical Tracers (MOZART) and CBMZ comparing to simple mechanisms like RACM and RADM (Gupta and Mohan, 2015;Sharma et al., 2017). MOZART gas phase chemistry mechanism and the four-bin Model for Simulating Aerosol Interactions and Chemistry (MOSAIC-4bin) were used for modeling atmospheric chemistry and aerosol properties as suggested in previous studies over India (Kumar et al., 2015a). MOZART, version 4 mechanism was initially developed for global modeling of ozone and other tracers in the troposphere (Emmons et al., 2010). Although it includes 97 gas-phase and bulk aerosol, all monoterpenes, which are important in ozone chemistry, were lumped together. As a result, Hodzic et al. (2015) added a detailed treatment of monoterpenes and Knote et al. (2014) updated the isoprene oxidation scheme in the MOZART mechanism in WRF-Chem. MOSAIC is an aerosol model that considers a wide range of aerosol species that are important in regional scale and treats the chemical and microphysical processes between them including nucleation, coagulation, thermodynamics and phase equilibrium, and gas-particle partitioning (Zaveri et al., 2008). Hodzic and Jimenez (2011) updated secondary organic aerosol (SOA) formation mechanism and the updated version is available in WRF-Chem for doing regional air quality modeling studies. MOSAIC-4bin, used in current study, calculates all the above-mentioned aerosol physics and chemistry in four sectional aerosol size bins with the assumption that each bin is internally mixed and all the particles within a bin have the same chemical composition (Zaveri et al., 2008).

Field Code Changed

In India, both anthropogenic and natural sources have important impacts on air quality. Biomass and biofuel use in residential sector for heating and cooking purposes have significant contributions to air quality in India (Conibear et al., 2018a;David et al., 2019;Venkataraman et al., 2018). Moreover, ~~T~~there are more than 1000 power plants and brick kilns in India that are major anthropogenic sources for SO_2 and particulate matters, respectively (Guttikunda and Calori, 2013). Other than these industrialized sources, literature shows that biomass burning (e.g. agricultural waste burning) contributes to 37 percent of air pollution over sub-continent (Kumar et al., 2015a). Hemispheric Transport of Air Pollution (HTAP v2.2) (Janssens-Maenhout et al., 2015) emission inventory of 2010 with 0.1 degree horizontal resolution, mapped to MOZART-MOSAIC mechanism (https://www2.acom.ucar.edu/wrf-chem/wrf-chem-tools-community), was used as the base anthropogenic emission inventory. Although accuracy of urban anthropogenic emission inventories have significant effects on air quality modeling studies (Gupta and Mohan, 2015;Kumar et al., 2012;Sharma et al., 2017), the focus of this paper is to capture the air pollution due to regional sources; we didn't use higher resolution emission inventories for Delhi.

Field Code Changed

Fire INventory from NCAR, version 1.5 (FINNv1.5) and Model of Emissions of Gases and Aerosols from Nature (MEGAN v. 2.0.4) were used as biomass burning emission and biogenic emission inventories, respectively (Guenther et al., 2006;Wiedinmyer et al., 2011). However, other studies have noticed that uncertainties in FINN emissions can significantly modify the results (Kulkarni et al., 2020). Therefore, two modifications were applied to FINN data to provide better input data: filling missing fires using VIIRS Fire Radiation Power (FRP) data and scaling the fire emissions (scaling procedure described in detail later). Liu et al. (2018) used FRP values to approximate the stubble burning areas affecting Delhi's air quality. In their statistical study, 99% of post monsoon FRP values were attributed to agricultural fires (Liu et al., 2018). In this study, we used FRP values to improve fire emissions. Specifically, we first regridded VIIRS 375m resolution FRP data to our domain. Then at each hour, for all grid cells that have FINN emissions, we find the corresponding mean VIIRS FRP, and do a linear regression between FRP and emission flux. Afterwards, we apply the regression line parameters on VIIRS FRP for the grid cells that don't have any FINN emission, to estimate the flux. It should be mentioned, all the available FRP data were utilized disregarding the retrieval's confidence level. Moreover, we used VIIRS instead of MODIS data as it provided higher resolution active fire points data (375m vs. 1km), which is an important point for small fires. For example, no active fire points in Moderate Resolution Imaging Spectroradiometer (MODIS) instrument were reported in 2018 post-monsoon for Uttar Pradesh (Kulkarni et al., 2020). Figure 9~~Figure 10~~ shows more fire grid cells in eastern IGP and central India when incorporating VIIRS data to FINN inventory. We acknowledge that this technique is a first-order approximation and can have large errors as FINN is based on burned area algorithms from MODIS retrieved data; more detailed research is required to improve the idea.

Dust storms are an important natural pollution source that have caused many pollution events over some parts of India (Kumar et al., 2014a). Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) mechanism was used to calculate the threshold wind velocity and total dust emission, which about 70 percent of total mass was then distributed in different bins of the other inorganics (OIN aerosol component in WRF-Chem; OIN represents all primary inorganic PM) component in the model with the assumption that the rest are larger than PM_{10} (Zhao et al., 2010). This is based on the study in Northern Africa, where Zhao et al. (2010) allocated about 1% in bins with diameter less than 2.5 micron and 69% of the dust in bin 4(2.5-10 microns), and assumed the rest were bigger than 10 micron and will not remain in the atmosphere for an influential period.

Table 1 Details of WRF-Chem physical and chemical setup configuration

Process	**Method**
Domain	1domain (15km horizontal resolution)
Landuse	MODIS 20-category

TimeStep	60 seconds based on CFL stability criterion (Courant et al., 1928)
Vertical	39 (top at 5hpa)
Microphysics	Morrison double-moment scheme (Morrison et al., 2005)
Longwave Radiation	RRTMG, called every 5 minutes
Shortwave Radiation	Goddard, called every 5 minutes
Planetary Boundary Layer	MYNN-level3 (Nakanishi and Niino, 2009)
Land Surface	Noah Land Surface Model (Wang et al., 2018)
Gas-Phase Chemistry	MOZART-4, called every 10 minutes
Photolysis Scheme	New TUV, called every 10 minutes
Aerosol Scheme	MOSAIC 4-bin (no aqueous phase chemistry), called every 10 minutes
Dust	GOCART (Ginoux et al., 2001)
Initial and Boundary meteorology	NCEP FNL

2.2. Sensitivity experiments

Three sets of experiments were performed to explore the impact of using different global data, as either boundary conditions or emissions, and dust emission formulation on $PM_{2.5}$ and AOD predictions (~~Table 2~~). It should be mentioned that all the modeling options and other input data remained unchanged unless specified.

One set of experiments focused on the sensitivity of the predictions to biomass burning emissions. First, we compared the impacts of two different biomass burning emission inventories, namely FINN and Quick Fire Emission Dataset (QFED) (Darmenov and da Silva, 2013). Specifically, simulations using QFED (ID: QFED_CAMCHEM) and FINN (ID: FINN_CAMCHEM) were performed to understand the impact of different fire detection algorithms. When using QFED, it should be mentioned that we mapped total CO values to VOC species in MOZART chemistry mechanism based on emission factors provided in literature instead of using VOCs emissions directly from QFED (Akagi et al., 2011). Second, we investigated whether FINN fire emissions were underestimated for all the days (ID: FINN_10Xall), some days (ID: FINN_10Xperiod1), or just one day before the pollution episode on Nov. 5^{th} (ID: FINN_10Xday). Then after modifying FINN using VIIRS FRP data, we did a sensitivity test with changing the period for scaling fire emissions. Specifically, we scaled fire emissions for a 15-days period between Nov. 3^{rd} and 17^{th} (ID: FINN_VIIRS_10Xperiod1) and an 8-days period between Nov. 5^{th} and 13^{th} (ID: FINN_VIIRS_10Xperiod2). ~~Finally, we~~We also evaluated the performance for a scaling factor of 10 in comparison with 7 (ID: FINN_VIIRS_7Xperiod2). Anthropogenic emissions over India also have high uncertainties (Saikawa et al., 2017). As a result, we studied how increasing the anthropogenic aerosol emissions by a factor of 2 affect the results (ID: BASE_ANTHRO2X).

Another set of experiments evaluated the impacts of chemical boundary conditions. Many global datasets can be used in regional air quality modeling. Simulations were performed using CAM-chem (ID: FINN_CAMCHEM), MOZART (ID: FINN_MOZART), Copernicus Atmosphere Monitoring Service (CAMS, ID: FINN_CAMS.), and a combination of CAM-chem for gaseous and MERRA-2 for aerosol species (ID: FINN_MERRA2) global modeling systems. It is important to note that CAMS and MERRA-2 are reanalysis models and use observed data to improve the results. CAMS assimilates MODIS and Advanced Along-Track Scanning Radiometer (AATSR) satellite instruments AOD (Inness et al., 2019). MERRA-2 assimilates AOD from multiple sources including MODIS, Multiangle Imaging Spectroradiometer (MISR), Advanced Very High Resolution Radiometer (AVHRR), and Aerosol RObatic NETwork (AERONET) although assimilating some products have been stopped after 2014 (Randles et al., 2017).

Field Code Changed

Formatted: Superscript

Finally, simulations were conducted for various dust emission modifications. In one simulation, we turned off dust emission option in the model (ID: NO_DUST), while in another simulation, we increased total dust emissions by 5 times to explore if dust emission were underestimated in the model (ID: DUST_5X). Moreover, we changed the allocation of total dust in different bins of MOSAIC module to see whether different allocation of aerosols can contribute to the observed extreme pollution in Delhi (ID: DUST_allocation). Specifically, we reduced 30% from the 4^{th} bin (2.5-10 micron) and distributed 25% of it in 3^{rd} bin (0.625-2.5 micron) and 5% in 2^{nd} bin (0.156-0.625 micron). More allocation to bins 2 and 3 were not considered, as it was not realistic to the large-size nature of dust aerosols. FINN_10Xall scenario represents the simulation with turned-on dust option (original allocation) in the model. The detailed results from experiments on boundary conditions and dust emissions can be found in the supporting document.

~~**Table 2** List of scenarios performed in this study~~

~~**Simulation ID**~~	~~**Initial/Boundary Chemical (Gaseous / Aerosol) Condition**~~	~~**Biomass Burning Emission Inventory**~~	~~**DUST**~~
~~**FINN_VIIRS_7Xperiod2 (base scenario)**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~7 times higher (FINN+VIIRS) for Nov 5^{th} to Nov 13^{th}~~	~~GOCART~~
~~**FINN_VIIRS_10Xperiod2**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher (FINN+VIIRS) for Nov 5^{th} to Nov 13^{th}~~	~~GOCART~~
~~**FINN_VIIRS_10Xperiod1**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher (FINN+VIIRS) for Nov 3^{rd} to Nov 17^{th}~~	~~GOCART~~
~~**FINN_10Xperiod1**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN for Nov 3^{rd} to Nov 17^{th}~~	~~GOCART~~
~~**FINN_10Xday**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN for Nov 5^{th}~~	~~GOCART~~
~~**FINN_10Xall**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN~~	~~GOCART~~
~~**NO_DUST**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN~~	~~Turned Off~~
~~**DUST_5X**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN~~	~~5 times higher GOCART emission~~
~~**DUST_allocation**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~10 times higher FINN~~	~~GOCART – put 30% of bin 4 in bins 2 and 3~~
~~**FINN_MERRA2**~~	~~CAMchem (gas) + MERRA2 (aerosol)~~	~~FINN~~	~~GOCART~~
~~**FINN_MOZART**~~	~~MOZART (gas + aerosol)~~	~~FINN~~	~~GOCART~~
~~**FINN_CAMS**~~	~~CAMS (gas + aerosol)~~	~~FINN~~	~~GOCART~~
~~**FINN_CAMCHEM**~~	~~CAMchem (gas + aerosol)~~	~~FINN~~	~~GOCART~~
~~**QFED_CAMCHEM**~~	~~CAMchem (gas +aerosol)~~	~~QFED~~	~~GOCART~~

Table 2 List of scenarios performed in this study

Simulation ID	**Initial/Boundary Chemical (Gaseous / Aerosol) Condition**	**Biomass Burning Emission Inventory**	**DUST**
Reference Scenario			
FINN_VIIRS_7Xperiod2 (base scenario)	CAMchem (gas) + MERRA2 (aerosol)	7 times higher (FINN+VIIRS) for Nov 5^{th} to Nov 13^{th}	GOCART
Biomass Burning Emission Sensitivities			

QFED_CAMCHEM	CAMchem (gas +aerosol)	QFED	GOCART
FINN_CAMCHEM	CAMchem (gas + aerosol)	FINN	GOCART
FINN_10Xall	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN	GOCART
FINN_10Xday	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN for Nov 5^{th}	GOCART
FINN_10Xperiod1	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN for Nov 3^{rd} to Nov 17^{th}	GOCART
FINN_VIIRS_10Xperiod1	CAMchem (gas) + MERRA2 (aerosol)	10 times higher (FINN+VIIRS) for Nov 3^{rd} to Nov 17^{th}	GOCART
FINN_VIIRS_10Xperiod2	CAMchem (gas) + MERRA2 (aerosol)	10 times higher (FINN+VIIRS) for Nov 5^{th} to Nov 13^{th}	GOCART
Boundary Condition Sensitivities			
FINN_MOZART	MOZART (gas + aerosol)	FINN	GOCART
FINN_CAMS	CAMS (gas + aerosol)	FINN	GOCART
FINN_MERRA2	CAMchem (gas) + MERRA2 (aerosol)	FINN	GOCART
Dust Emission Sensitivities			
NO_DUST	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN	Turned Off
DUST_5X	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN	5 times higher GOCART emission
DUST_allocation	CAMchem (gas) + MERRA2 (aerosol)	10 times higher FINN	GOCART put 30% of bin 4 in bins 2 and 3
Anthropogenic Emission Sensitivity			
BASE_ANTHRO2X	Similar to Base scenario (ID: FINN_VIIRS_7Xperiod2) except anthropogenic aerosol emissions increased by a factor of 2		

Formatted: Normal

2.3. Observation data

The model performance was evaluated using ground measurements, space-borne instruments, and global reanalysis data. Specifically, we used data collected by the Central Pollution Control Board (CPCB) over the domain for doing statistics. It includes stations over Delhi (19 stations), Rajasthan (10 stations), Haryana (4 stations), and Punjab (3 stations). No additional quality control filters, other than the ones by CPCB (https://cpcb.nic.in/quality-assurance-quality-control/), were applied. We evaluated the results after applying the filters proposed by other studies (e.g. Kumar et al. (2020)); they had slight impacts on statistics (shown in the supporting information). $PM_{2.5}$ data measured by an US EPA instrument at the US embassy in Delhi was used as the reference station. Level-2 VIIRS remote sensing instrument data onboard Suomi-National Polar-Orbiting Partnership (S-NPP) was used for comparing the spatial pattern of AOD and fire counts over the domain. Specifically, aerosol products with around 6km horizontal resolution based on the Deep Blue algorithm (Hsu et al., 2019) and 375 m active fire products based on VNP14IMG algorithm (Schroeder et al., 2014) were used. There are only two AERONET stations in the domain (Fig. 1a). AERONET data at these two sites confirmed the reliability of VIIRS retrieved data (Fig. S1). MERRA-2 gridded data was also used to evaluate the model performance. MERRA-2 reanalysis is based on the assimilation of many meteorological data and the assimilation of AOD from multiple satellites (Gelaro et al., 2017). The on-ground continuous monitoring stations guidelines state that instruments should sample at heights between 3-10m. Irrespective of this condition, some of CPCB stations are placed on top of the buildings with restricted clean flow of air (personal inspections). ~~some blockades to 360-degree clean view~~. While we observed little impact of this situation on the concentrations in a well-mixed layer, a meteorological parameter like wind speed data can show~~s~~ erratic behaviour. As a result, we used MERRA-2 meteorological data to evaluate the WRF-Chem simulations using 10m wind speed and direction, 2m temperature, surface water vapor mixing ratio variables.

We also compared MERRA-2 AOD (at 550nm) and $PM_{2.5}$ predictions with WRF-Chem results to evaluate how the assimilation of AOD affected the predictions. The MERRA-2 $PM_{2.5}$ was based on the mass mixing ratios of black carbon, organic carbon, dust, sea-salt, and sulfate. Since ammonium concentration is not available, it is common in the literature to assume that sulfate ion will be completely neutralized by ammonium and form ammonium sulfate and therefore a factor of 1.375 was assumed in calculating inorganic aerosol concentrations (Buchard et al., 2016;He et al., 2019;Provençal et al., 2017). On the other hand, literature suggest organic carbon concentration should be multiplied by 1.4 to compensate for other missing organic compounds to estimate the organic mass (Buchard et al., 2016;Chow et al., 2015;He et al., 2019;Provençal et al., 2017;Turpin and Lim, 2001). However, Turpin and Lim (2001) argued that this scaling factor should be 2.6 for biomass burning particles; we used 2.6 according to our studied time period and potential black carbon sources:

$$[PM] = [BC] + 2.6 * [OC] + 1.375 * [SO_4] + [DUST] + [SS] \quad (1)$$

Where BC is black carbon, OC is organic carbon, SO_4 is sulfate, DUST is dust, and SS is sea salt concentrations. As dust and sea salt data are reported in multiple bins, different bins should be used for different particle diameters.

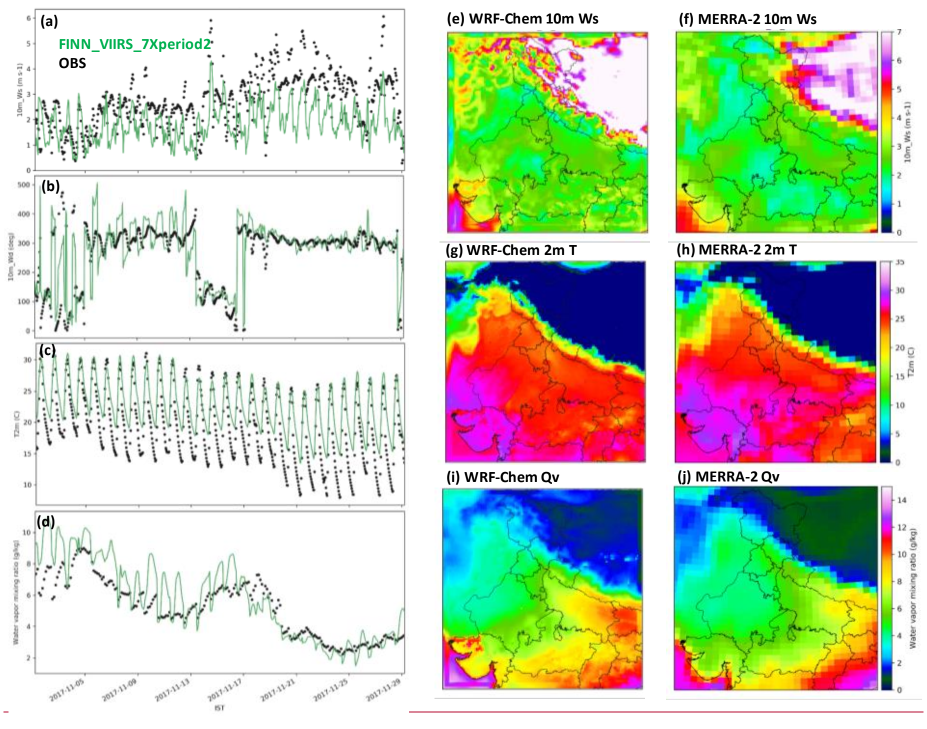
The metrics we used to assess the performance of the simulations are Root Mean Squared Error (RMSE), Mean Error (ME), Normalized Mean Bias (NMB), Normalized Mean Error (NME) and correlation coefficient (R) as defined in supporting ~~information~~ document (Emery et al., 2017;Emery et al., 2001). Since low values can have significant impacts on normalized values, which are used in mean normalized metrics, normalized mean values are better metrics and used in this study (Emery et al., 2017).

3. Results and discussions

3.1. Model performance

Our analysis between different simulations revealed that FINN_VIIRS_7Xperiod2 scenario had the best statistical performance of the configurations studied. This scenario is called the base scenario and we evaluate it in this section. Performance of the base model in capturing the meteorological parameters was evaluated using MERRA-2 data for 10m wind speed and direction, 2m temperature, and surface water vapor mixing ratio. Figure 2 (a, b, c, d) show these comparisons at the location of the US Embassy in Delhi (28.59° N, 77.19° E). The model was able to capture the general diurnal trend for all these variables and the sharp shift in wind direction between Nov. 13th and Nov. 17th, after the extreme pollution episode. Negatively biased wind speed with ME of 1.1 m/s and RMSE of 1.28 m/s shows the model generally underestimated wind speed and it was most predominant between Nov. 17th and Nov. 25th. Figure 2c shows the model did not accurately capture nighttime 2m temperature minima but captured the maximum values with overall overestimated ME of 3.52° C and RMSE of 4.01° C. The wind speed satisfied the benchmark RMSE value of 2.0 m/s, while temperature was higher than the targeted ME goal of 2.0° C (Emery et al., 2001). The representation error plays an important role in evaluating results due to different horizontal resolutions in the model and MERRA-2 dataset (~0.15~~*~~x0.15 vs 0.625~~*~~x0.5 degree), specifically in urban areas. For instance, the same statistics for a rural area in Rajasthan (27.0° N, 73.0° E; not shown) have smaller biases and are compliant with benchmark values (RMSE of 0.99 m/s for wind speed and ME of 1.08° C for 2 m temperature). For water vapor mixing ratio the model clearly captured the daily variations; specially, the increase after the pollution episode (Nov. 13th). However, it showed a very sharp day-to-night shift during the pollution episode days. The spatial performance of the model averaged over November during daytime hours (8AM to 6PM) is shown in Fig. 2 (panels e, f, g, h, i, j).

The sharp gradient between Himalayas and IGP regions in the north-east, the gradient between land and sea in the south-west, and the slight gradient between different land types in north-west of the domain for both 10m wind speed and 2m temperature were captured well. Overall, the model was able to capture the general daily variations and spatial trends when compared to MERRA-2 data.



(a)
FINN_VIIRS_7Xperiod2
OBS
(b)
(c)
(d)
(e) WRF-Chem 10m Ws
(f) MERRA-2 10m Ws
(g) WRF-Chem 2m T
(h) MERRA-2 2m T
(i) WRF-Chem Qv
(j) MERRA-2 Qv

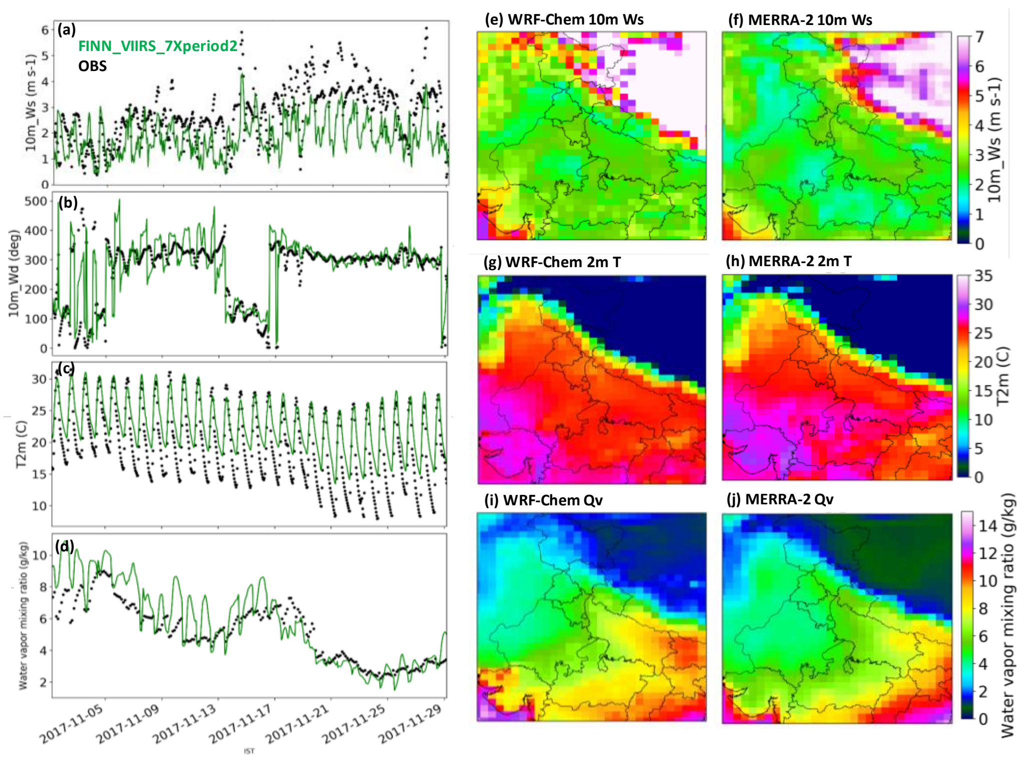


Figure 2 Temporospatial meteorological performance of base scenario simulation: Time series of simulated (green line) and MERRA-2 (black dots) hourly a) 10 m wind speed, b) 10 m wind direction, c) 2m temperature, d) surface water vapor at US Embassy coordinates. e,f) Averaged daytime (8AM-6PM) 10 m wind speed maps of modeled (e) and MERRA-2 (f). g,h) Averaged daytime (8AM-6PM) 2 m temperature maps of simulated (g) and MERRA-2 (h). i,j) Averaged daytime (8AM-6PM) surface water vapor mixing ratio (g/kg) maps of model (i) and MERRA-2 (j)

Figure 3 shows spatial distribution of the base scenario, VIIRS data, and the bias for 550 nm AOD, averaged over all the days in Nov., Nov. 5th as a day with intensive fire emissions, and Nov. 24th as an illustrative day after the extreme pollution episode. Model showed high AODs over Delhi and Punjab, confirming satellite data. Moreover, AODs were high over western IGP, close to major fires of Punjab, with a gradual gradient towards eastern and central India. Dust emission sources in the border of Pakistan also led to high AODs although they did not affect Delhi as discussed in the supporting document. ~~Moreover, AODs were high over western India, close to major fire and dust emission sources, with a gradual gradient towards eastern and central India. However, the values were lower in eastern IGP compared to VIIRS data.~~ In general, the model underestimates AOD over IGP and overestimates elsewhere. WRF-Chem predicted the averaged AOD over the whole domain for Nov. 2017 to be 0.58 (±0.4), while VIIRS data showed 0.43(±0.26). AOD maps for Nov. 5th show the model generally underestimated AOD values for the entire IGP region, except for Punjab. Moreover, the model underestimated aerosol loadings over central India. Other studies have reported biased low AOD and corresponding $PM_{2.5}$ concentrations over other polluted regions (He et al., 2019;Song et al., 2018). Nov. 24th, on the other hand, represented a day with no significant fire emission. The model did a good job capturing AOD values in the central parts of India and around Delhi. However, the model missed high AOD values in eastern IGP. MERRA-2 data also did not show high AODs over the border and did not capture extremely high AODs over Punjab, although it showed ~~some~~ AOD enhancements (Fig. S6~~2~~).

Figure 4 shows time series of modeled, MERRA-2 product, VIIRS retrievals, and observed AOD at the AERONET stations (location shown on Fig.1). AOD values at Kanpur, a station in the eastern IGP, were more than 1.0 before the pollution episode and reached up to 2.0 during the episode days, and decreased to values between 0.5 and 1 for the rest of days. The model captured the general trend although missed high AOD's between Nov. 9th and 13th, while MERRA-2 successfully captured the AOD trend through the whole month, including days with enhanced AOD values. This shows that AOD assimilation in MERRA-2 significantly improves AOD predictions. At Jaipur, located in southern IGP, the model overestimated AOD for the first five days of November. During the pollution episode days, the model is biased high compared to MERRA-2 and VIIRS retrievals. AERONET data showed low AOD values before the pollution episode but did not report values during the pollution episode. It suggests, as one possibility, that PM concentrations were too high during this period that the instrument was not able to retrieve data. After the pollution period, AOD values were lower than 0.5, showing relatively low PM concentrations. In general, MERRA-2 showed better performance in terms of NMB (Kanpur: -1.3% and Jaipur:-20.1%) compared with our model (Kanpur:-27.4% and Jaipur: +29.9%). Comparing averaged AOD with VIIRS retrievals for BASE_ANTHRO2X scenario showed lower bias over the IGP (Fig. S7). These results show the need for improved estimates of biomass burning as well as anthropogenic emissions. We also looked at Angstrom Exponent (AE) at Jaipur and Kanpur to understand if the model captured the mode of the particles (Fig. S8). Over Jaipur the model is biased high compared to AERONET data (NMB: 30%) and shows more finer aerosols in the model. After Nov. 20th, both AERONET and VIIRS retrievals suggest the dominance of coarser aerosols, while the AE for the model does not follow the same trend. However, $PM_{2.5}/PM_{10}$ ratio shows more coarse aerosols compared to the rest of the month (Fig. S9). Over Kanpur, the model AE is biased ~~very~~ high (NMB: 50.8%). On the other hand, the model shows closer AE values to VIIRS retrievals. For example, both the model and VIIRS retrieval show similar reduction in AE on Nov. 8th and 9th. Kumar et al. (2014b) also reported slight AE overestimation in WRF-Chem during a pre-monsoon dust storm at Kanpur and Jaipur. Furthermore, model and AERONET have variational trend while MERRA-2 is smooth during the whole month at both Jaipur and Kanpur.

~~Figure 4 shows time series of modeled, MERRA-2 product, and observed AOD at the AERONET stations, located on Fig.1. AOD values at Kanpur, a station on eastern IGP, were more than 1.0 before the pollution episode and reached up to 2.0 during the episode days, and decreased to values between 0.5 and 1 for the rest of days. The model captured the general trend although missed high AOD's between Nov. 9th and 13th, which led to negative NMB of -27.4%. MERRA-2 successfully captured the AOD trend through the whole month, including days with enhanced AOD values. It shows how AOD assimilation in MERRA-2 significantly improves AOD predictions. Jaipur AERONET station, located in southern IGP, showed low AOD values before the pollution episode but did not report values during the pollution episode. MERRA-2 shows very similar values to AERONET measured AODs. However, MERRA-2 values are low during the extreme pollution episode, which may be related to missing values in assimilated products. After the pollution period, AOD values were lower than 0.5, showing relatively low PM concentrations. The model overestimated AOD for the first five days of Nov., which led to overall high biased NMB of 29.9%. However, it is important that the error for days after the pollution episode were small.~~

Figure 5a shows time series plot of base scenario and observed $PM_{2.5}$ concentration at the US Embassy station. Observed values were high throughout the month on the order of 200 μgm^{-3} with ~~a~~ diurnal variations due to changes in the mixing heights. The extreme pollution episode began on Nov. 7th, when $PM_{2.5}$ concentrations increased to more than 800 μgm^{-3}. On Nov. 8th, the values increased even more to about 1000 μgm^{-3}. $PM_{2.5}$ concentrations started decreasing on Nov. 9th and continued through~~on~~ Nov. 10th. However, values increased again and were high between Nov. 11th and Nov. 13th. Afterwards, they returned to ~ 200 μgm^{-3} for the rest of the month. The model accurately captured the magnitude and diurnal cycle for $PM_{2.5}$ for non-episodic days. Moreover, the

Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Subscript
Formatted: Subscript
Formatted: Not Highlight
Field Code Changed
Formatted: Not Highlight

model was able to see the sharp increase in concentration in the beginning of the episode starting from Nov. 7^{th} with reported $PM_{2.5}$ concentrations of ~650 µgm^{-3}. This sharp increase was captured after incorporating VIIRS data into FINN emissions accompanied by scaling the emissions by a factor of 7. In fact, increased emissions from fires in agricultural fields in the north-west on previous days and favorable north-westerly winds, as shown on Fig.2 explain this concentration hike. However, the model underestimated the concentrations for the next three days. Then, the model captured the second enhancement. Although wind direction showed good agreement with MERRA-2 dataset and wind speed was biased low and even more favorable for stagnant conditions, modeled $PM_{2.5}$ concentrations had a large negative bias for the period between Nov. 8^{th} and Nov. 10^{th}. This ~~suggests~~suggests either low local anthropogenic emission within Delhi or~~either low anthropogenic emissions in Delhi or~~ some missing pollution sources upwind of Delhi that were not included in the emission estimates led to underestimation.

Dekker et al. (2019) studied CO concentrations during Nov. 2017 using satellite observations and they reported low emissions as one of the reasons for large negative concentration biases, although they proposed unfavorable meteorological condition as the main reason for high CO concentrations in Delhi, between Nov. 11^{th} and Nov 14^{th}. Moreover, Cusworth et al. (2018) reported that MODIS based biomass burning emission inventories miss many small fires over India. Beig et al. (2019) concluded that long range transported dust coming from Middle East was a major source for this extreme pollution episode. ~~We looked at MERRA-2 surface $PM_{2.5}$ concentration data for the study period to explore if dust was a major source.~~ Figure 5a~~ also~~ shows that MERRA-2 did not capture high surface $PM_{2.5}$ concentrations after Nov. 8^{th} either. Navinya et al. (2020) reported that MERRA-2 underestimates $PM_{2.5}$ over India, especially during post-monsoon season. More discussions on the extreme pollution episode are provided in the following section.

Starting on Nov. 13^{th}, the modeled concentrations went down as winds shifted to easterlies and wind speed increased. Beig et al. (2019) found $PM_{2.5}$ concentrations after the pollution episode were lower compared with similar periods in previous years. Thereafter, the concentrations went back to values for Delhi before the episode. The model did a fairly good job in capturing the trend during non-episode periods (Table S4). Increasing anthropogenic emissions (ID: BASE_ANTHRO2X) on simulation results overestimated $PM_{2.5}$ concentrations in the US embassy location during non-episode days (Fig. S7).~~, indicating that the anthropogenic emissions provided by of the HTAP emission inventory are reasonable.~~

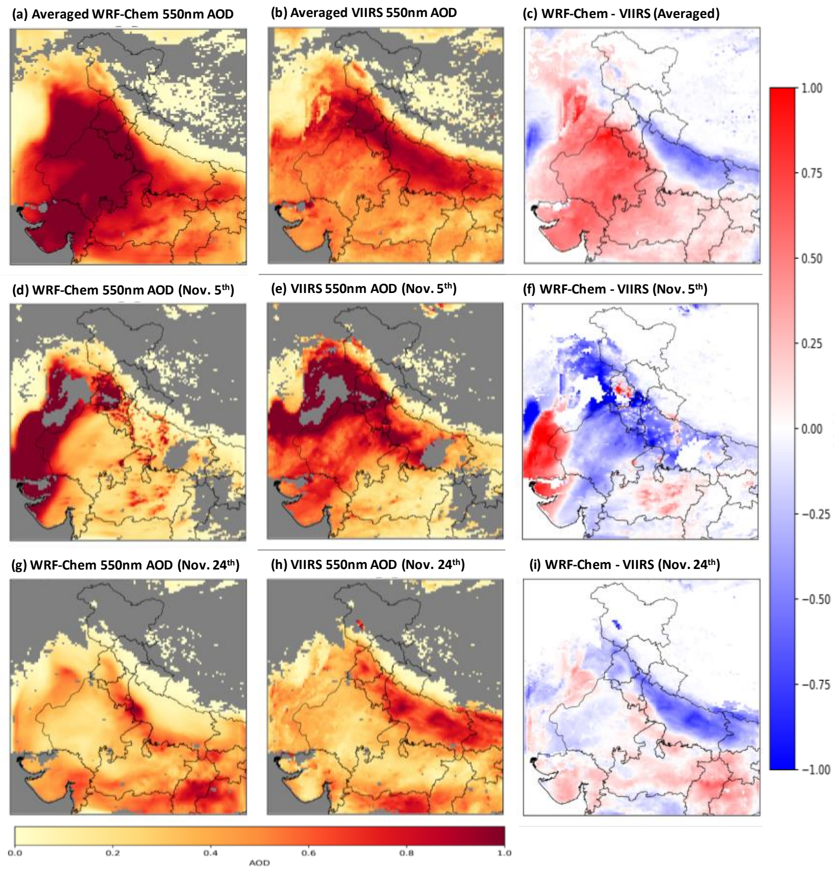
Figure 5b and Figure 5c show the averaged $PM_{2.5}$ maps for all the hours over the studied region in the base simulation and MERRA-2 dataset. The model was able to capture higher concentrations over north-western India and the border with Pakistan where agricultural fire and dust emissions play the most important role for extreme pollution episodes over IGP. However, the model showed higher values than MERRA-2 over southern Punjab, the region with high biomass burning emissions (Kulkarni et al., 2020). Since MERRA-2 assimilates satellite AOD data as its major aerosol forcing, it will not be able to capture high concentrations if satellite retrieval algorithms miss corresponding high AODs.

Figure 5b and Figure 5c also show that the model was biased high over central India and biased low over eastern IGP~~.~~. These results indicate improving emissions in eastern IGP can significantly improve the simulation results. Conibear et al. (2018a) also reported limited success of models to capture the spatial variability of $PM_{2.5}$ over India in 2016, specifically during winter. Table 3 provides statistics for 24-hours averaged $PM_{2.5}$ concentrations for base scenario simulation for Delhi and its western states~~provinces~~. Statistics for Delhi show NMB of -16.6%, which passes the "criteria" benchmark of 30% , while NME of 27.6% shows better performance and complies with the benchmark "goal" of 35% for the whole month (Emery et al., 2017). Correlation coefficient of 0.48 is also higher than the benchmark criteria of 0.4. Statistics significantly improve after excluding the four extremely polluted days between Nov. 7^{th} and Nov. 10^{th} and all are within benchmark goals (Table S4~~3~~). Kumar et al. (2020) assimilated MODIS AOD to WRF-Chem in order to improve the air quality forecasts over Delhi. In their study, Mean Bias for first-day forecast of $PM_{2.5}$ concentration decreased from -98.7 µgm^{-3} to -13.7 µgm^{-3}. They also showed that RMSE decreased from

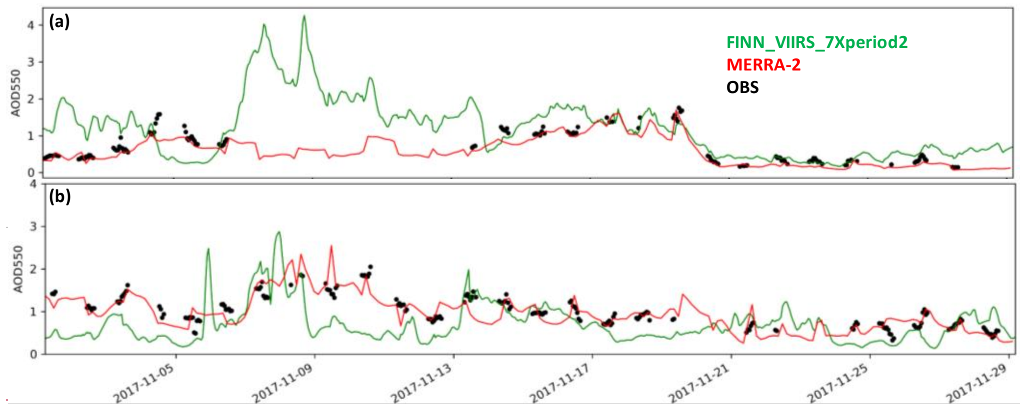
167.4 µgm^{-3} to 117.3 µgm^{-3}. Our results from the base scenario (Mean Bias: -42.38 µgm^{-3} and RMSE: 118.47 µgm^{-3}) shows comparable results to the data assimilation technique, while still both models are biased low.

~~-~~Statistics for Haryana ~~province~~state (4 stations) show good performance (NMB of -7.5% and correlation coefficient of 0.4). The model was biased high for Rajasthan (10 stations NMB: 15.5%) and Punjab (3 stations NMB: 17%). The model slightly overestimated $PM_{2.5}$ concentrations during the episode days in Rajasthan but captured the concentrations during the rest of the month (Fig. S10~~3~~). In Punjab, measured data did not report $PM_{2.5}$ enhancement during the extreme episode, while the model showed very high concentrations after scaling fire emissions by a factor of 7. However, VIIRS satellite images (e.g. Fig. 9d) clearly show massive agricultural fires in this state during November and its signals were expected in the measured data. The overall scatter plots including the averaged values for each state shows good spatial performance of the base scenario (Fig. S11).

Although different meteorological parameters can be responsible for the biases, accuracy of anthropogenic emissions is important. For example, recent local anthropogenic emission inventories developed for Delhi have higher particle emissions than in the regional inventory used in this study, which impacts modeled $PM_{2.5}$ concentrations for typical days (Kulkarni et al., 2020). We conducted BASE_ANTHRO2X scenario to investigate the effect of uncertainties in the anthropogenic emissions. This scenario increased $PM_{2.5}$ concentrations in Delhi up to ~150 µgm^{-3}, which led to overestimation (in contrast to underestimation in base scenario) at many of non-episode days (Fig. S7). Although this scenario did not help in capturing the high concentrations during the episode, it confirms the need for better anthropogenic emissions. On the other hand, it reduced the bias over IGP (Fig. S7). These results point out the need for best estimates of emissions of both anthropogenic and biomass. ~~Moreover, HTAP provides annual averaged emission inventories, which allocates the same amount of emission to each month, while colder seasons have generally more residential emitted pollutants. Furthermore, consumption and emission rates of 2010 needed to be tuned for a more recent year in order to get better results.~~ .Maps also show that averaged $PM_{2.5}$ concentrations over most of India were higher than the air quality standard.



390 **Figure 3** Spatial distributions of AOD at 550 nm averaged over whole November (top panel), Nov. 5th (middle panel), and Nov. 24th (bottom panel). WRF-Chem maps represent base scenario results. Differences between model and VIIRS are also shown.



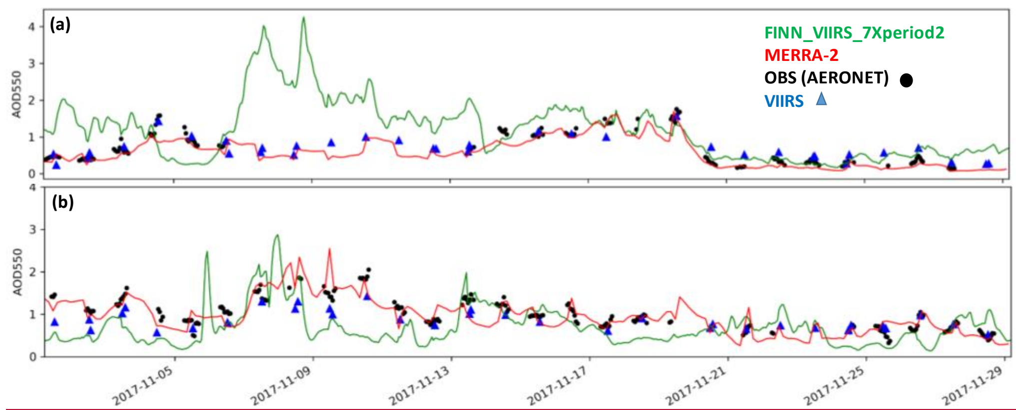
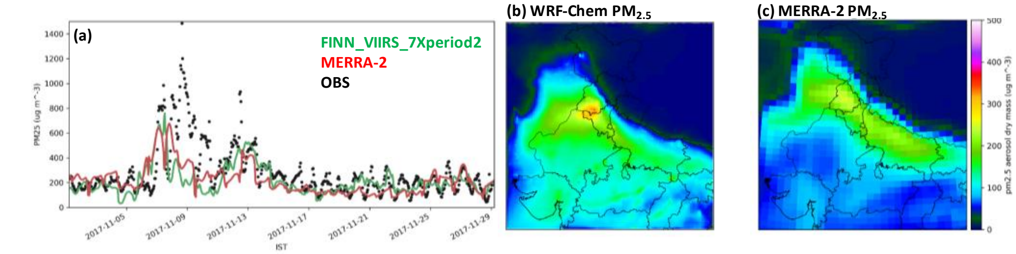


Figure 4 Time series of modeled (green line), VIIRS retrievals (blue triangle), MERRA-2 (red line), and AERONET (black dots) AOD at 550 nm during Nov. 2017 at a) Jaipur, b) Kanpur.



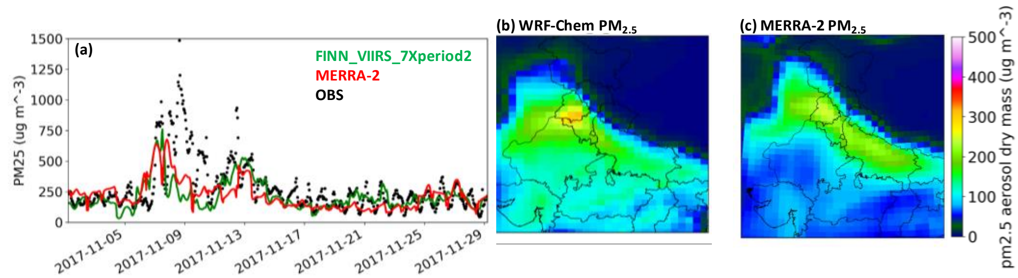


Figure 5 Temporospatial air quality performance of base scenario simulation: a) Time series of simulated (green line), MERRA-2 (red line), and ground measurement (black dots) hourly $PM_{2.5}$ concentration at US Embassy coordinates. b, c) Hourly averaged $PM_{2.5}$ concentration maps of model regridded to MERRA-2 resolution~~ed~~ (b) and MERRA-2 (c).

Table 3 Mean (±standard deviation), Normalized Mean Bias (NMB), Normalized Mean Error (NME), and Pearson Correlation Coefficient (R) averaged for all CPCB stations in different ~~provinces~~ states during Nov. 2017. Model values are for base scenario (FINN_VIIRS_7Xpeiod2). Mean values are for hourly data, while NMB, NME, and R relates 24-hours averaged values**.**

~~Province~~State	Hourly Obs. Mean (±std) (μgm^{-3})	Hourly Model Mean (±std) (μgm^{-3})	24-hours NMB (%)	24-hours NME (%)	24-hours R (%)
Delhi	255.5 (±146.6)	213.9 (±113.9)	-16.6	27.6	0.48
Haryana	177.7 (±77.6)	165.8 (±89.9)	-7.5	29.5	0.40
Punjab	139.9 (±54.7)	166.7 (±198.3)	17	55.5	0.24
Rajasthan	123.4 (±62.7)	147.7 (±62.7)	15.5	34.4	0.22

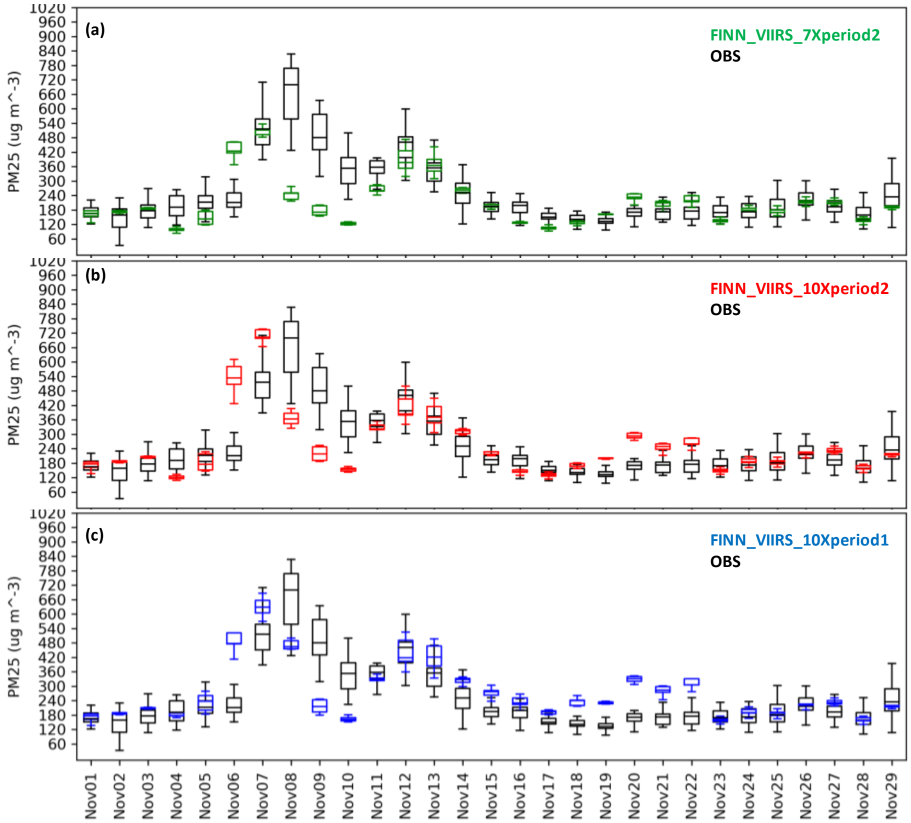
3.2. Extreme pollution episode analysis

Figure 6a shows the box and whisker plots for daily $PM_{2.5}$ concentration for the base scenario and all of the CPCB stations in Delhi. 24-hour averaged measured values over all CPCB stations in Delhi for $PM_{2.5}$ ranged between 133 μgm^{-3} and 664 μgm^{-3}, which is about 2 and 11 times, respectively, higher than India 24-hours standard value of 60 μgm^{-3}. The model showed overall good performance for daily $PM_{2.5}$ concentrations for typical days (Table S4: NMB of -2.44 and R of 0.7) and followed the observed trend in the extreme pollution episode (Fig. 6), which suggests the overall meteorology and transport patterns were captured by the simulations. However, the model started the episode on Nov. 6th and significantly overestimated the concentrations. The model captured the median for Nov. 7th very well, although measured values span a wider range. The model missed the high concentrations on Nov. 8th, which led to underestimations on Nov. 9th and Nov. 10th, as well, regardless of capturing the decreasing trend. However, the model was able to simulate the second wave of the episode starting on Nov. 11th and accurately captured the median and range of $PM_{2.5}$ concentrations on Nov. 12th and Nov. 13th. It is important to point out that the underestimation of PM on the 9th and 10th persisted for all of the sensitivity cases performed. This suggests either the transport in the model during these days missed high source regions and/or significant emission sources for these days were not included in the inventories.

Back trajectories can be used to provide insights into modelled concentrations during the extreme pollution episode. Back trajectories were calculated for releasing 10,000 air parcels at 100 m above ground level and over eastern Delhi using the FLEXible PARTicle dispersion model (FLEXPART) with inputs from WRF-Chem model output (Brioude et al., 2013). Figure 7 shows 72-hours mean back trajectory maps for Nov. 6th, 7th, 8th, 9th, and 10th. The releasing times are 00 (red line) and 12 (blue line) UTC on each day. Also plotted are the fire (grey line) and anthropogenic (black line) emissions along the trajectory. The model started to build up $PM_{2.5}$ concentrations on Nov. 6th and was biased high (Fig. 6~~7~~a). Back trajectories for Nov. 6th_00 show $PM_{2.5}$ concentrations were majorly due to anthropogenic emissions (Fig. 7a). However, Nov. 6th_12 trajectories in Fig. 7c show a spike in fire emissions on the previous hours (backward hours: 5 and 30), which immediately led to high $PM_{2.5}$ concentrations. Moreover, trajectory paths for this day reveal that emissions belonged to fires east of Delhi. Figure 9~~Figure 10~~ shows that the fires east of Delhi in the base scenario are due to incorporating VIIRS data into the fire emissions. Therefore, high biased $PM_{2.5}$ concentration may be related to the scaling factor applied to eastern Delhi fires. On Nov. 7th, the model perfectly captured $PM_{2.5}$ median (Fig. 6~~7~~a). Back trajectories for Nov. 7th_00 (Fig. 7d,e) show the beginning of a shift in wind direction and $PM_{2.5}$ concentration was exclusively due to fire emissions on Nov. 5th (backward hour: 40). Compared to Nov. 7th_00, fire emission footprints for Nov. 7th_12 trajectories are ~~lower~~smaller, while higher for local anthropogenic emissions. Back trajectories for Nov. 8th show the northern parts' contribution for both releasing times, although trajectories for Nov. 8th_00 crossed through central parts of Punjab. Moreover, local anthropogenic emission sources affected Nov. 8th_00 trajectories. The model underestimated $PM_{2.5}$ concentrations on Nov. 8th, which can be partly related to errors in transport as the trajectories for Nov. 8th_12 crossed eastern parts of Punjab. However, other physical processes or lower anthropogenic emissions can also be responsible for low bias. ~~However, the model underestimated $PM_{2.5}$ concentrations on Nov. 8th, which can be related to errors in transport as the trajectories for Nov. 8th_12 crossed eastern parts of Punjab.~~ Delhi's air quality in Nov. 9th_00 was still getting affected by northern parts, while trajectories for Nov. 9th_12 shifted towards the east. Since, trajectories for Nov. 9th do not show any fire or anthropogenic emissions' pulse, either the model missed the dynamics of that day or emission sources. Nov. 10th trajectories show eastern flow, again, and no fire emission contribution.

Formatted: Font color: Accent 5

To further understand the regional scale transport of the smoke plumes, we plotted cross section of $PM_{2.5}$ over the path from Punjab through Delhi (Fig. 8, path line shown in Fig. 1). $PM_{2.5}$ concentrations showed typical values on Nov. 5th_00 although they still exceeded the standard limits. On Nov. 5th_12, concentrations significantly increased over Punjab area because of fires and the winds brought them on a path towards Delhi. The Punjab's smoke did not completely cross Delhi yet on Nov. 6th as back trajectories for 00 and 12 UTC hours also showed the effects of anthropogenic emissions and fires in eastern Delhi. On the other hand, a significant amount of smoke was above the boundary layer as shown in Nov. 6th_12 panel.~~The smoke was mixed in the boundary layer and reached to altitudes more than 2km as shown in Nov. 6th_12 panel.~~ Due to shifting winds on Nov. 7th (as shown in Fig. ~~7~~2), the smoke upwind of Delhi blew over Delhi and led to extremely high concentrations. Although the model captured the median in Nov. 7th, it missed the maximum extent of observed values. Cross sections on Nov. 8th, 9th, and 10th show the residual Punjab's smoke in the boundary layer, while we saw the model underestimated $PM_{2.5}$ concentrations on these days. Measured $PM_{2.5}$ concentrations over Delhi show a decreasing trend between Nov. 8th and Nov. 10th (Fig. 6). Vertical profiles for the base scenario also show the model captured high biomass burning emission period on Nov. 6th (Fig. 12~~3~~). However, it also showed high amounts of smoke above the PBL. ~~Vertical c~~Cross sections for Nov. 11th to Nov 14th can be found in the supporting information (Fig. S12~~4~~).These results suggest that plume rise in the model released the emissions too high or the model did not mix the smoke down fast enough. Vijayakumar et al. (2016) showed agricultural fires can transport via upper troposphere and subside over Delhi using ECMWF map. Social reasons can be also important as the first reaction of people during hazy days is to drive to work which directly (exhaust emission) and indirectly (road dusts) worsen air pollution.


(a)
FINN_VIIRS_7Xperiod2
OBS
(b)
FINN_VIIRS_10Xperiod2
OBS
(c)
FINN_VIIRS_10Xperiod1
OBS
PM25 (ug m^-3)
Nov01 Nov02 Nov03 Nov04 Nov05 Nov06 Nov07 Nov08 Nov09 Nov10 Nov11 Nov12 Nov13 Nov14 Nov15 Nov16 Nov17 Nov18 Nov19 Nov20 Nov21 Nov22 Nov23 Nov24 Nov25 Nov26 Nov27 Nov28 Nov29

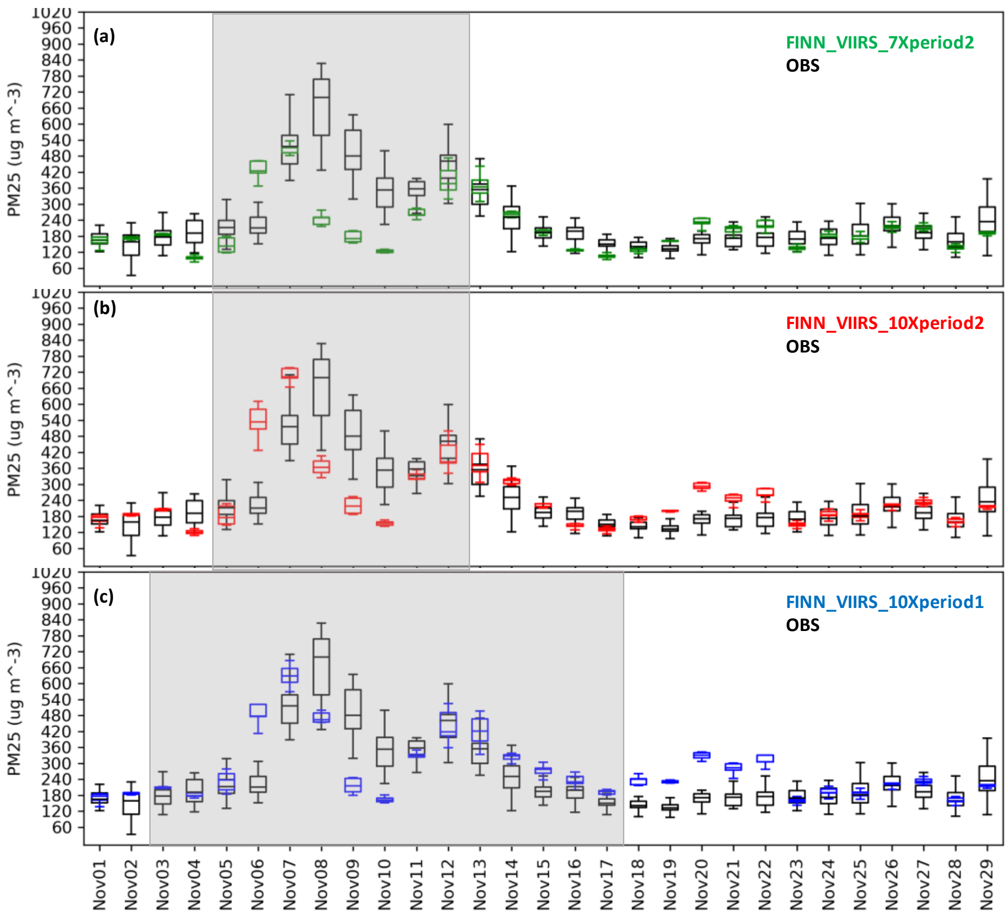


Figure 6 Box and Whisker plots of observed (black) and modeled daily $PM_{2.5}$ concentration averaged over all CPCB stations in Delhi: a) FINN_VIIRS_7Xperiod2, b) FINN_VIIRS_10Xperiod2, c) FINN_VIIRS_10Xperiod1. Shaded area show the time window that biomass burning emissions were increased.

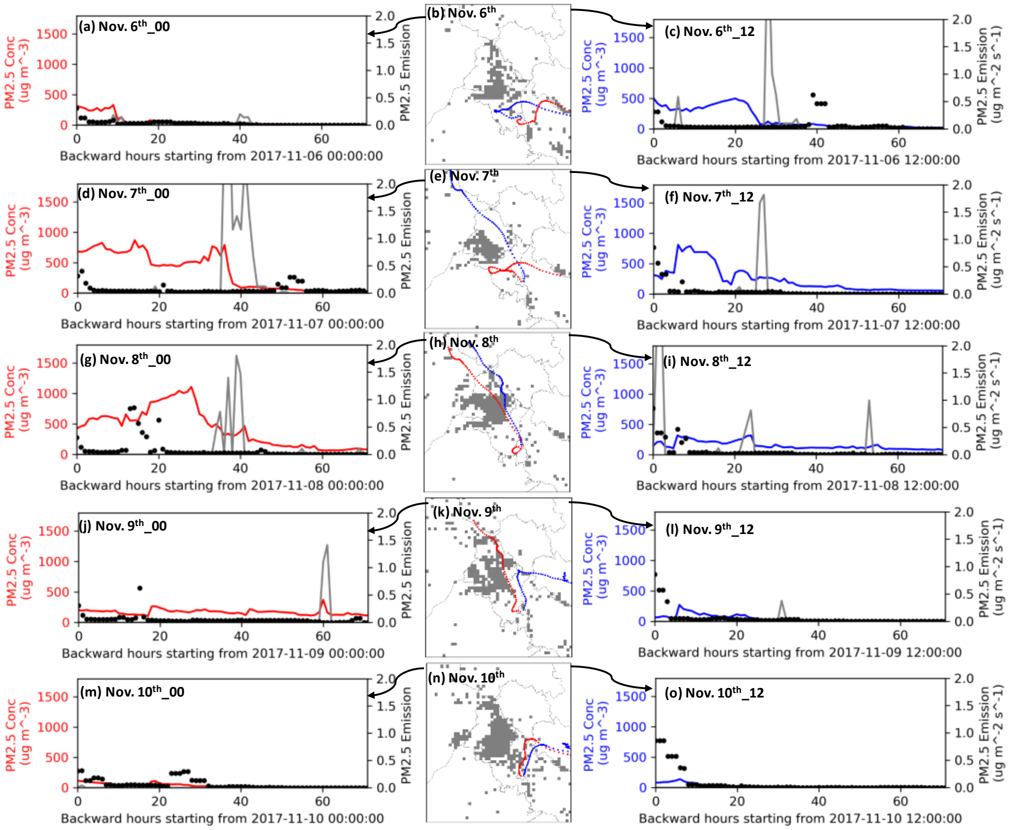
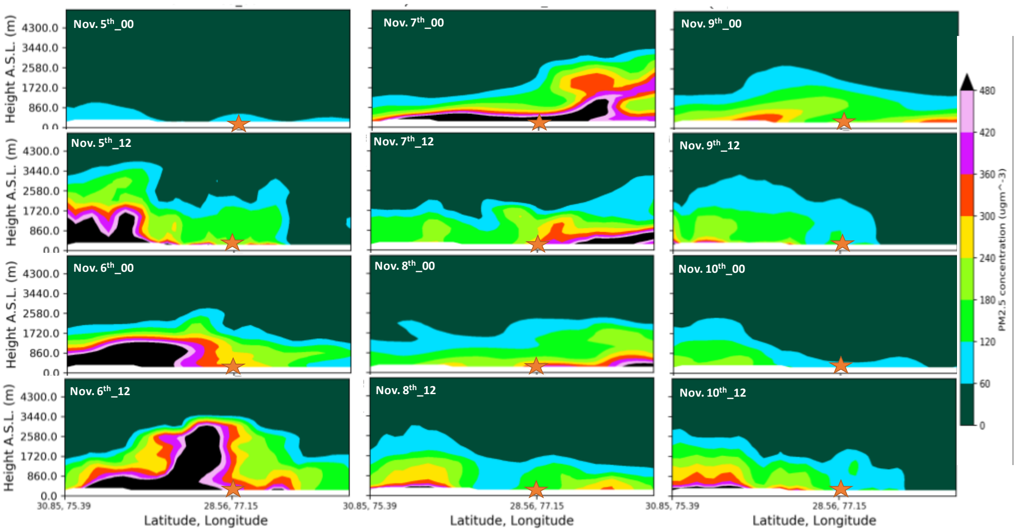


Figure 7 Back trajectory plots of $PM_{2.5}$ concentration (base scenario) on different days for 72 hours: In each row, the map shows back trajectory path for the mass mean location for releasing 10000 particles on eastern Delhi at 00 UTC (red line) and 12 UTC (blue line), where the underlying map shows FINN grid cells on that day. Time series show $PM_{2.5}$ concentrations (primary y-axis) and emissions (secondary y-axis) for Anthropogenic (black dots) and FINN (gray line) inventories along the path. a, b, c) November 6th. d, e, f) November 7th. g, h, i) November 8th. j, k, l) November 9th. m, n, o) November 10th.00 and 12 UTC denote 5:30 AM and 5:30 PM local time, respectively.



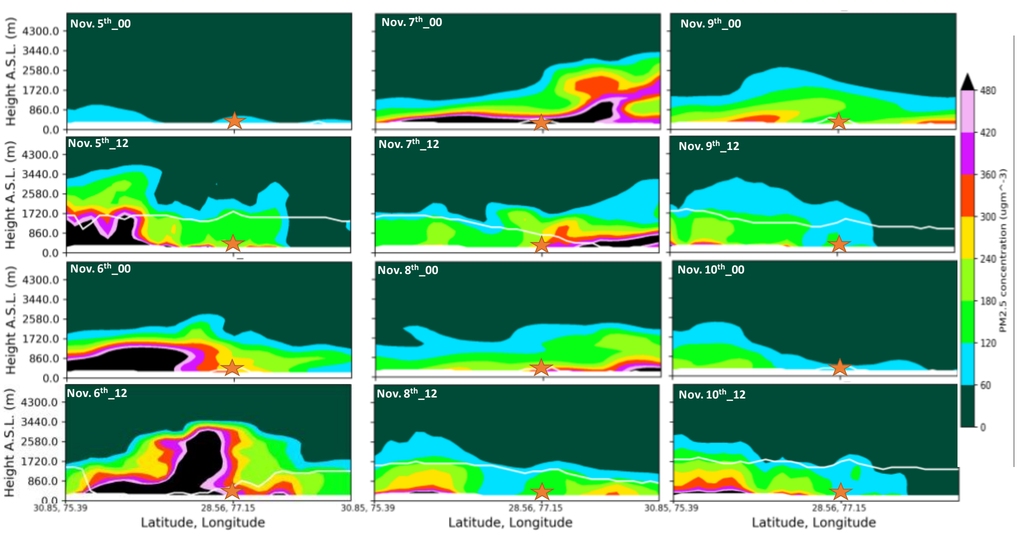


Figure 8 Vertical cross section of $PM_{2.5}$ concentration through the path shown in Fig. 1 for the days between Nov. 5th and Nov. 10th. For each day, two snapshots are shown at 00UTC (5:30AM local time) and 12UTC (5:30PM local time). The orange star shows the location of Delhi through the path. White line shows the PBL height across the path.

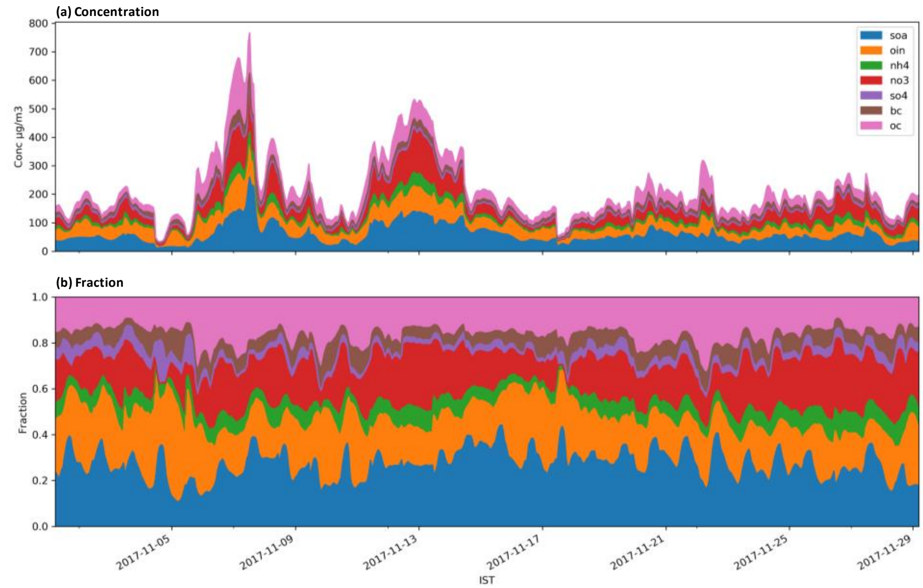
~~3.3.1.1. Aerosol composition in Delhi~~

~~Figure 9 shows the modeled $PM_{2.5}$ composition, both in concentrations and by fraction, at the location of the US Embassy in Delhi. Secondary aerosols (secondary organic aerosols (SOAs) + secondary inorganic aerosol (SIA) consisting of Ammonium (NH_4) + Nitrate (NO_3) + Sulfate (SO_4)) comprised 57% of the total averaged $PM_{2.5}$ concentration, whereas primary aerosols (BC + Organic Carbon (OC) + OIN) constituted the rest. Gani et al. (2019) measured PM_1 in Delhi and reported 50-70% for secondary aerosols,~~

Field Code Changed

~~and PM_1 constituted ~85% of $PM_{2.5}$ concentration. SOAs, individually, comprise 27% of the aerosol mass, while SIAs account for 30% of the mass. Amongst inorganic species, NO3, NH4, and SO4 comprise 19%, 7%, and 4%, respectively. Gani et al. (2019) reported the same ranked order but with different percentages. Major contribution of NO_3 in winter is also reported in other studies (Pant et al., 2015). BC fraction was 7%, which is very close to the measured fraction of 6.4% in wintertime PM_1 (Gani et al., 2019). Pant et al. (2015) reported averaged OC and elemental carbon concentrations of 104.4 μgm^{-3} and 46.3 μgm^{-3}, respectively, which is consistent with our OC/BC ratio of 2.72. Comparing modeled BC1 data with available data for this period (Gani et al., 2019) shows an overall measured to modeled ratio of 1.22, which is consistent with the range other studies reported (Kumar et al., 2015b;Moorthy et al., 2013).~~

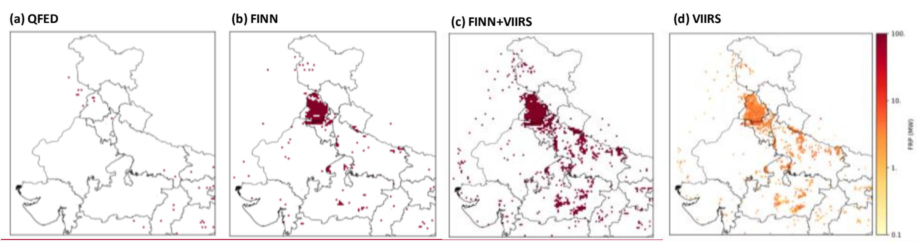
Field Code Changed



~~**Figure 9** $PM_{2.5}$ composition at US Embassy coordinates in base scenario: a) Concentration values, b) Fractional values.~~

~~3.4.~~3.3. Sensitivity to changes in biomass burning emission inventories

Biomass burning emissions used in the base scenario in order to capture the extreme pollution episode were tuned after exploring how these inventories influenced $PM_{2.5}$ concentrations (Table S2). First, we looked at two different emission inventories based on different methodologies and horizontal resolutions; Specifically, FINNv1.5 and QFEDv2.5r1. Both inventories rely on MODIS data; FINN is based on active fire points and estimates of burned area, whereas QFED uses FRP approach (Pan et al., 2020). Figure 9~~Figure 10~~ shows the grid cells with biomass burning emissions based on QFED (panel a), FINN (panel b), and FINN_VIIRS composition (panel c), used in base scenario, accompanied with active fire points seen by VIIRS (panel d) based on the Fire Information for Resource Management System (FIRMS) product for Nov. 5th. It shows FINN captured more fire points in the domain although missing some in eastern IGP and central India while QFED missed almost all of the fire points in Punjab on that specific day. As a result, the QFED simulation did not show any major signal for $PM_{2.5}$ concentration on Nov. 7th, whereas the experiment using FINN inventory (ID: FINN_CAMCHEM) followed the measured start of the episode period, regardless of its low bias (Fig. S13~~5~~). In general, results using QFED inventory had worse statistics (Fig. 11~~2~~ and Table S3~~2~~), which is mostly due to the inability of the inventory to capture the fire points over the domain and it can be attributed to both the technique and the resolution as QFED data have ~10 km resolution, whereas FINN data has ~1 km resolution. Pan et al. (2020) found high uncertainty between different biomass burning emission inventories over Southeast Asia. They showed FINN is, in general, a better dataset for tropical regions as its 2-days continuous fire emission compensates for the lack of daily MODIS coverage used in QFED (Pan et al., 2020;Wiedinmyer et al., 2011). Dekker et al. (2019) increased the GFAS biomass burning emission inventory by 5 times and did not see any improvement on CO simulation and reported about 2 percent contribution from fires on Delhi's air quality. Our results confirms that FINN provides better biomass burning emissions for India for this period and sheds light on the importance of choosing proper biomass burning emission inventory for a specific domain.



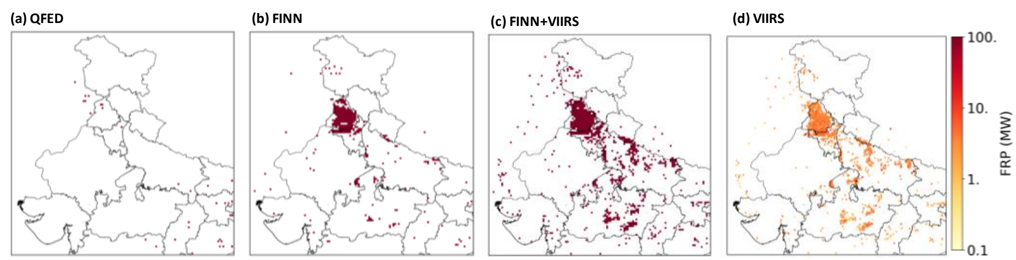


Figure 9~~10~~ Spatial fire coverage in different datasets for Nov. 5th: a) QFED fire grid cells b) FINN fire grid cells c) FINN fire grid cells filled missing points with VIIRS d) VIIRS active fire points and corresponding FRP values.

However, the signals from the simulation using FINN biomass burning emission inventory were not high enough as it recorded a maximum concentration of 400 μgm^{-3} while the corresponding measured value was 680 μgm^{-3}. Since observation data are sporadic over India and there were not many ground measurement stations available, sophisticated techniques such as inversion modeling were not feasible (Saide et al., 2015). As a result, manual tuning of the emission data was performed. The first attempt was to understand if FINN required to be increased for the whole month, a 15-days period around the episode, or just on Nov. 5th, which had many fire points in Punjab (Fig. 9~~10~~). Figure 10~~Figure 11~~ shows $PM_{2.5}$ time series averaged over all CPCB stations based on these scenarios. Increasing FINN emissions for the whole month (ID: FINN_10Xall) led to an overestimation in the first 5 days of November but it significantly helped capturing high peaks on Nov. 7th and 8th. Moreover, it increased the concentrations on Nov 12th and 13th regardless of missing the peaks. However, it did not show any improvements between Nov. 9th and 12th, which suggests that the included fires did not influence Delhi's air quality during this period. On the other hand, increasing FINN emissions data by 10 times for all days led to very high $PM_{2.5}$ concentrations on later days (Nov 20th-27th). It showed that FINN data were not systematically biased low. In other words, these results suggest that FINN algorithm underestimated the magnitude of only some fires emission amounts. Some studies have shown that thick fires can be identified as clouds in retrieval algorithms ~~or some other meteorological phenomenon may cause biases~~ (Dekker et al., 2019;Huijnen et al., 2016). As another experiment, we increased FINN emission only on Nov 5th since that day had original high values in the inventory (ID: FINN_10Xday). This experiment resulted in better $PM_{2.5}$ concentrations on the last third of November. However, it captured only high concentrations of

Nov. 7th and missed the peak of Nov. 8th as well as underestimated on some other days including Nov 13th. Finally, we multiplied the fire emissions by 10 for a 15 days period between Nov. 3rd and 17th (ID: FINN_10Xperiod1). In this way, we were able to capture the peaks on Nov. 7th and 8th, see major contributions between Nov. 12th and 14th, and realistic values between Nov 19th and 27th. It should be mentioned this 15-days period was chosen arbitrary and better scaling factors could be achieved by implementing tracers in the model, which is ~~out of~~beyond the scope of this paper.

Although this experiment significantly improved the statistical performance for the CPCB stations, the model was spatially underestimating concentrations over eastern IGP (Fig. S6). Moreover, we observed lower fire grid cells in FINN inventory comparing to VIIRS active fire points (Fig. 10). As a result, we tested how incorporating VIIRS data into FINN in order to fill missing fire grid cells would improve the results. Figure 10~~Figure 11~~ shows how incorporating VIIRS data improved the performance on Nov. 12th and Nov. 13th (ID: FINN_VIIRS_10Xperiod1). However, the model started the episode too early on Nov. 6th and overestimated $PM_{2.5}$ concentrations after the episode. This suggests that the 15-days period for increasing FINN emissions could be too long; we then changed the scaling period from 15 to 8 days between Nov. 5th and Nov. 13th (ID: FINN_VIIRS_10Xperiod2). This modification led to higher $PM_{2.5}$ concentrations on Nov. 7th as Fig. 6 shows. Moreover, the model was still biased high on after-episode days, which led to choosing the increasing factor of 7 (FINN_VIIRS_7Xperiod2) as our best experiment.

Figure 11~~Figure 12~~a shows the Taylor diagram for hourly $PM_{2.5}$ concentrations based on the studied experiments, representing their statistical performance, for all the days in November. It shows that switching between different experiments mostly improved the standard deviation values. The QFED_CAMCHEM experiment had the lowest standard deviation, but missed high $PM_{2.5}$ concentration values. On the other hand, three experiments using VIIRS-integrated fire emissions had closer standard deviations to measured value. Although the base scenario had good statistical metrics, standard deviation and correlation coefficient were lower, compared to other two VIIRS-included and BASE_ANTHRO2X scenarios for all the days. The reason is that overestimation of other scenarios for Nov. 7th and after-episode days compensate for underestimation between Nov. 8th and Nov. 10th. Figure 11~~Figure 12~~b shows the same variables for all the days except Nov. 8th, 9th, and 10th. It shows that the base scenario had the best statistical performance for non-extremely polluted days.

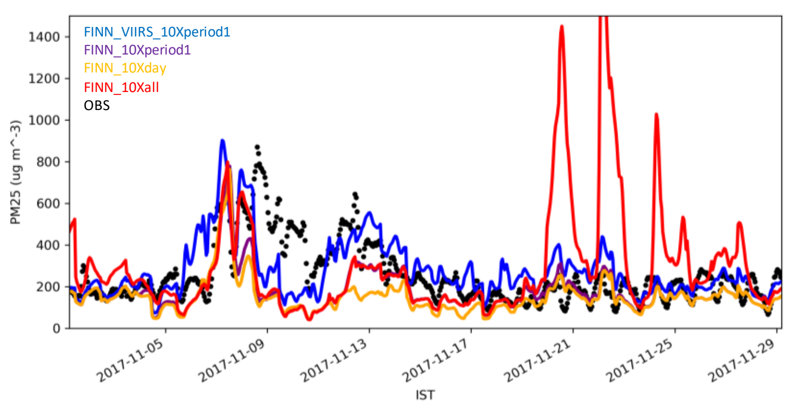


Figure 10~~11~~ Time series of model $PM_{2.5}$ sensitivity to 10 times increase in FINN emission inventory for different periods: FINN_VIIRS_10Xperiod1: Nov. 3rd to 17th, FINN_10Xperiod1: Nov. 3rd to 17th, FINN_10xday: Nov 5th, FINN_10Xall: all days. Black dots presents ground measurements data averaged over all CPCB stations in Delhi.

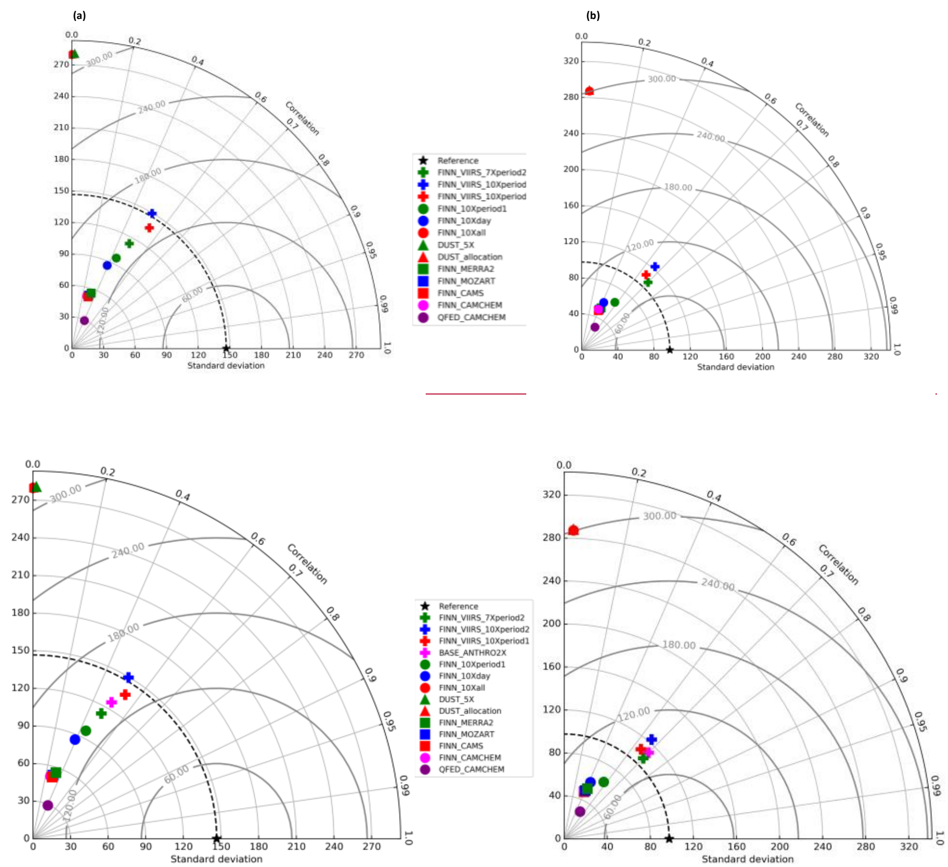


Figure 11~~12~~ Taylor Diagram of hourly $PM_{2.5}$ concentration based on different simulation scenarios for: a) all the days, b) all the days except Nov. 8th, 9th, 10th. Plus signs denote experiments after incorporating VIIRS data to FINN, Circles denote different experiments on biomass burning emission inventory, triangles denote dust emission experiments, and squares denote chemical boundary condition experiments. Green colors show best performance in each experiment. Black star denotes standard deviation of all CPCB stations averaged in Delhi. The statistics plotted are standard deviation on the radial axis and Pearson correlation coefficient (r) on the angular axis, and the gray lines indicate normalized Centered RMSE (CRMSE).

Figure 12~~Figure 13~~ shows vertical ~~distributions~~profiles of $PM_{2.5}$ as a function of time at the US Embassy coordinates for variations using the FINN inventory and MERRA-2. Increasing the emissions in FINN inventory significantly increased $PM_{2.5}$ concentrations both vertically and temporally. Although the concentrations got closer to MERRA-2 data, the timing for the peak of~~there was a shift in~~the boundary layer on Nov. 6th was different. By incorporating VIIRS data to FINN and adding more fire emissions, the boundary layer values peaked on Nov. 6th earlier and looks much more like MERRA-2 data. Figure 12~~Figure 13~~ also shows more particles at altitudes above the boundary layer, which do not influence surface concentrations. Furthermore, it suggests the scaling factor for the base scenario could be smaller than 7 if the aerosols had been in the boundary layer. It can be partly related to the plume rise module in the WRF-Chem model that may have emitted species at too high altitudes. Increasing emissions also indirectly influenced modeled air quality over Delhi. As our model configuration included feedbacks, absorbing aerosols in the atmosphere (products of fire emissions) decreased the surface solar radiation budget, changed the dynamics of the atmosphere, reduced the Planetary Boundary Layer (PBL) height, and increased aerosol concentrations. In other words, higher PBLH leads to lower concentrations. For example, Murthy et al. (2020) found that $PM_{2.5}$ concentration decreased up to 14 μgm^{-3} for 100m increase in PBLH. Figure 12~~Figure 13~~ also shows the interactions between PBLH and $PM_{2.5}$ concentration at the location of the US embassy.~~these phenomena.~~ By increasing FINN inventory by 7 times, the PBL height decreased by ~50% on Nov. 6th, (compare FINN_VIIRS_7Xperiod2 and FINN_MERRA-2 panels in Fig. 1~~3~~2). However, a measured PBLH dataset can provide better insights. As a result, another study is required to compare modeled PBL heights to observed data (e.g. Nakoudi et al. (2019)) and study the effects of different PBL parameterization modules on aerosol concentrations. Vertical ~~time series plots~~profiles for other experiments using FINN can be found in the supporting ~~information~~document (Fig. S~~7~~15).

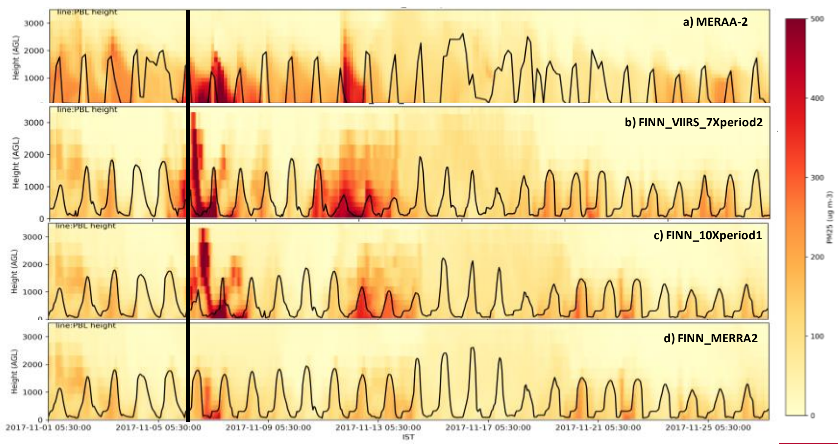
We also did two sets of experiments to understand if long-range transported dust from middle east or in-boundary dust emissions impacted air quality in Delhi. Our sensitivity tests suggest that dusts had very low contribution to air quality in Delhi during Nov. 2017. Detailed discussion on their impacts have been presented in the supporting document.

Formatted: Subscript

Formatted: Subscript

Formatted: Superscript

Formatted: Font: Bold, Font color: Blue, Complex Script Font: Bold



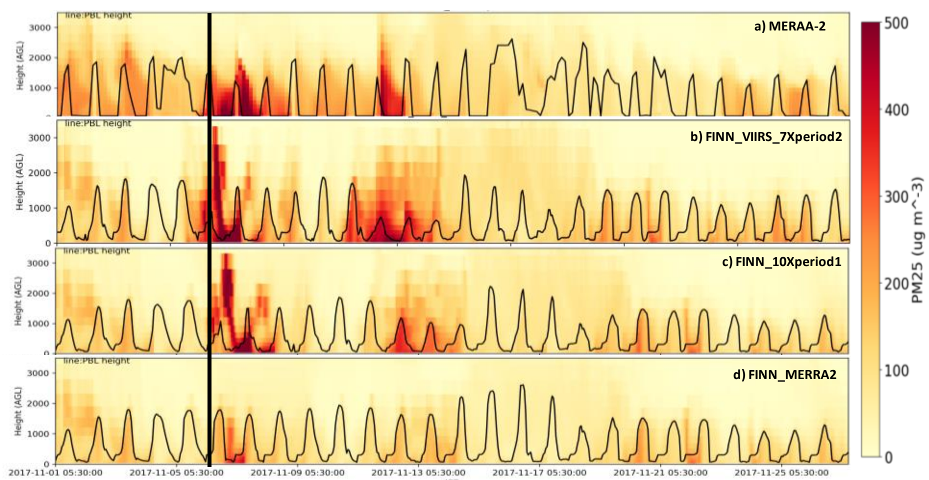


Figure 12~~13~~ Vertical cross section of $PM_{2.5}$ sensitivity to major changes in FINN emission inventory at US Embassy coordinates: a) MERRA-2 data as true values, b) FINN_VIIRS_7Xperiod2, c) FINN_10Xperiod, d) FINN_MERRA2. Black lines present planetary boundary layer height. Vertical black line crossing all panels shows boundary layer peak on Nov. 6th for MERRA-2 and other experiments.

~~3.5. Sensitivity to changes in boundary conditions data~~

~~Figure 14 shows the hourly averaged $PM_{2.5}$ and PM_{10} concentration maps during the studied period using four different boundary conditions as described in methods section. The major difference between these maps is on the western parts of the domain. The conceptual model in Beig et al. (2019) suggested that long range transported dust coming from Pakistan and Middle East influenced air quality in northern India during this period. FINN_MERRA2 simulation had the highest values for both $PM_{2.5}$ and PM_{10}, which shows that some parts of the domain were affected by pollution from the boundaries. FINN_CAMS simulation shows lower concentrations, which can be attributed to CAMS assimilation technique. On the other hand, FINN_MOZART and FINN_CAMCHEM scenarios are very similar to each other. Overall, data assimilation as applied in MERRA-2 can improve the~~

~~regional modeling features for the domains that get affected by long-range transported dust. However, pollutants coming from boundaries had small influences on Delhi region's air quality, during the studied period (Fig. 12 and Table S2)~~

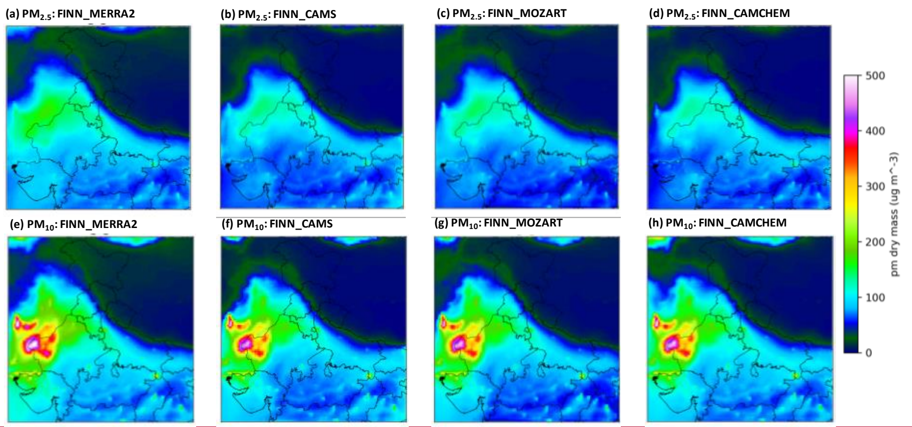
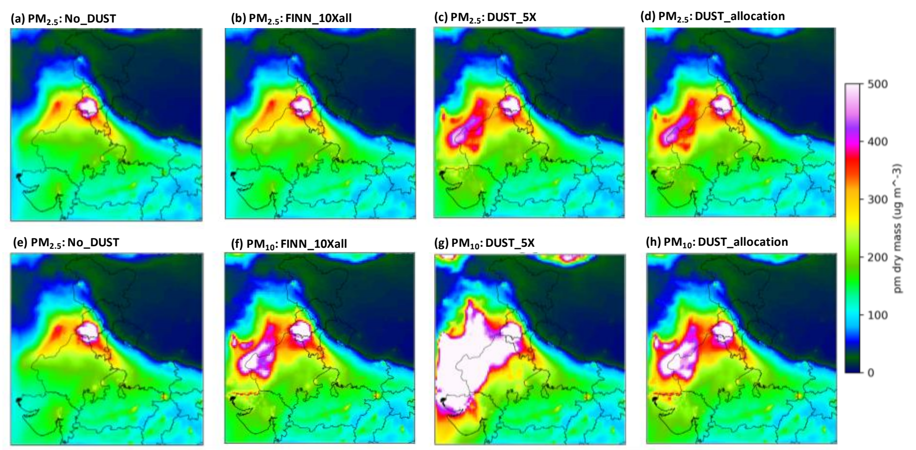


Figure 14 ~~Responses of $PM_{2.5}$ and PM_{10} to changes in boundary conditions coming from: a, e) MERRA-2, b, f) CAMS, c, g) MOZART, and d, h) CAM-Chem.~~

~~3.6. Sensitivity to changes in dust emissions~~

~~Sensitivity tests using different boundary conditions showed that long range transported dust coming from Middle East did not majorly influence air quality in Delhi. However, our domain covers some desert regions in eastern Pakistan and their dust emission impacts were evaluated. Figure 15 shows the response of hourly averaged $PM_{2.5}$ and PM_{10} concentrations to changes in the dust emissions. Turning on the dust option affected $PM_{2.5}$ and PM_{10} concentrations in eastern parts of Pakistan and some parts of the borders between Pakistan and India, but did not affect Delhi (Fig. S8). In another experiment, we increased the total dust emissions by 5 times, which increased PM_{10} concentrations significantly over western parts of the domain, close to source. This small-range transport is due to the mass of large dust particles and accompanying higher dry deposition rates. It also increased $PM_{2.5}$ concentrations and influenced some western parts of India with smaller size aerosols. However, they did not reach Delhi region, as the statistics over Delhi show no improvements (Fig. 12 and Table S2). In another experiment, changing the allocation of total dust in different bins as explained in methods section changed the aerosol regime in the west parts of the domain. Specifically, larger areas were effected by small size aerosols. Changing allocation of dusts, directly affected $PM_{2.5}$ concentrations in Delhi during the extreme pollution episode. Specifically, it increased $PM_{2.5}$ concentrations by ~20 µgm^{-3} on Nov. 8th. However, it was less than 5% contribution (Fig. 12 and Table S2). Moreover, increasing dust emissions had both positive and negative effects on concentrations (e.g. positive effect on Nov. 20th and negative effect on Nov. 28th), which are due to indirect effects of aerosols (Fig. S8). We did not perform more experiments as these tests suggest that in-domain dust sources were not a major source of extreme pollution episode in Delhi during November 2017. It should be mentioned that dust experiments, had lowest correlation coefficients, since the fire emissions were significantly high for all the days in all of them (Fig. 12).~~



~~**Figure 15** Responses of $PM_{2.5}$ and PM_{10} to changes in dust emissions: a, e) dust is turned off, b, f) dust is turned on, c, g) dust emissions are increased by 5 times, and d, h) dust emissions with different allocation in MOSAIC bins.~~

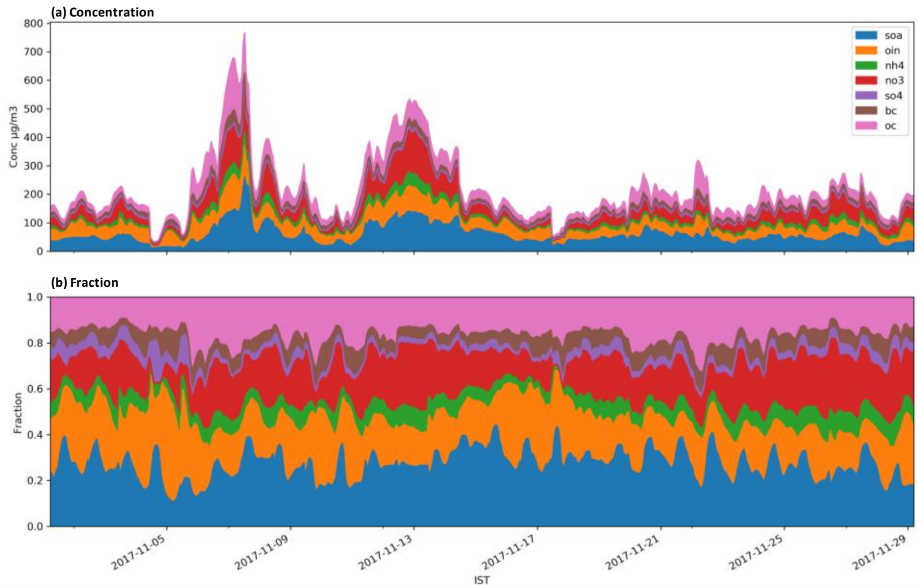
Formatted: No bullets or numbering

Formatted: Font: Bold

3.4. Aerosol composition in Delhi

~~Using MOSAIC aerosol module enabled the model to track speciation of aerosols during Nov. 2017. In this section~~ ~~Figure 9 shows the m~~We analyzed the modeled $PM_{2.5}$ composition, both in concentrations and by mass fraction, at the location of the US Embassy in Delhi~~was analysed~~ (Fig. S16). Secondary aerosols (secondary organic aerosols (SOAs) + secondary inorganic aerosol (SIA) consisting of Ammonium (NH_4) + Nitrate (NO_3) + Sulfate (SO_4)) comprised 57% of the total averaged $PM_{2.5}$ concentration, whereas primary aerosols (BC + Organic Carbon (OC) + OIN) constituted the rest. Gani et al. (2019) measured PM_1 in Delhi and reported 50-70% for secondary aerosols, and PM_1 constituted ~85% of $PM_{2.5}$ concentration. SOAs, individually, comprise 27% of the aerosol mass, while SIAs account for 30% of the mass. Amongst inorganic species, NO3, NH4, and SO4 comprise 19%, 7%, and 4%, respectively. Gani et al. (2019) reported the same ranked order but with different percentages. Major contribution of NO_3 in winter is also reported in other studies (Pant et al., 2015). BC fraction was 7%, which is very close to the measured fraction of 6.4% in wintertime PM_1 (Gani et al., 2019). Pant et al. (2015) reported averaged OC and elemental carbon concentrations of 104.4 μgm^{-3} and 46.3 μgm^{-3}, respectively, which is consistent with our OC/BC ratio of 2.72. Comparing modeled BC1 data with available data for this period (Gani et al., 2019) shows an overall measured to modeled ratio of 1.22, which is consistent with the range other studies reported (Kumar et al., 2015b;Moorthy et al., 2013).

Field Code Changed



~~**Figure 9** PM2.5 composition at US Embassy coordinates in base scenario: a) Concentration values, b) Fractional values.~~

Formatted: Normal

~~3.7.~~3.5. Ozone concentration analysis

Figure 13~~Figure 16~~a shows the box and whisker plots for daytime (8AM-6PM) ozone concentration for the base scenario and all of the CPCB stations in Delhi. Observed values ranged between 10 ppb and 110 ppb and the range between lower and upper quartiles was about 20 ppb, showing high ozone variability over Delhi. Moreover, observed values were higher during the extreme pollution episode. It indicates particles are not the only issue during PM pollution episodes in Delhi. The modeled median was in the range of observed values, especially on non-episode days. However, model overestimated ozone concentrations on Nov. 7th. Moreover, the range of observed ozone concentrations were wider than modeled values. In general, model captured the trend fairly good with correlation coefficient of 0.57, but was biased high with NMB of 18% for daytime hours throughout whole November. High biased ozone concentration in Delhi is reported in other studies (Gupta and Mohan, 2015).

Figure 13~~Figure 16~~b shows the daytime ozone concentration maps averaged over November 2017. Central regions of India show higher ozone concentrations compared to northern IGP region. On the other hand, ozone concentration in urban regions were lower than rural areas. This is due to lower ozone production in higher NOx emissions in urban areas (Ghude et al., 2016;Karambelas et al., 2018). Averaged ozone concentration over the domain throughout Nov. 2017 was 77 ppb using the base scenario. Ozone concentrations decreased by up to 27 ppb when using a scenario without any modifications to aerosol emissions (Fig. S~~9~~17). Regardless, the averaged values are 9-17 ppb higher than annual averaged concentration of 60 ppb in year 2011 (Ghude et al., 2016). Overall, high measured and modeled ozone concentrations and positive correlation with $PM_{2.5}$ are concerning and demand for more studies. Moreover, recent observed values over Delhi indicated that, during the COVID-19 pandemic that all activities were suspended, $PM_{2.5}$ concentration went down while the trend and range of ozone concentration remained unchanged (Jain and Sharma, 2020).

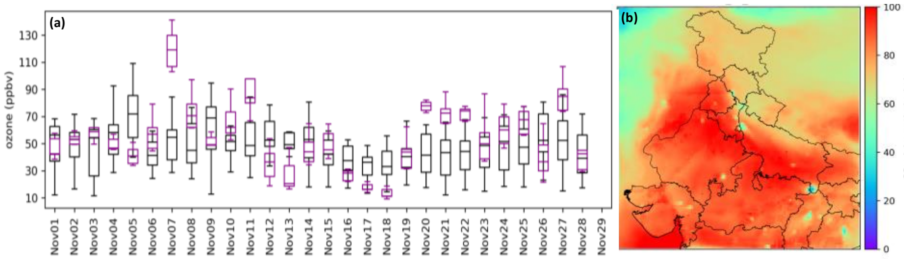


Figure ~~16~~13 Daytime (8AM-6PM) ozone modeling performance in base scenario: a) Box and Whisker plots of observed (black) and modeled (purple) concentrations averaged over all CPCB stations in Delhi. b) November 2017 daytime averaged concentrations.

~~3.8.~~3.6. Study limitations

In this study, we used a simple framework to modify Fire emissions with satellite data. Specifically, we used VIIRS data to fill FINN emissions, which are based on MODIS retrievals. We used VIIRS data as they had higher resolution (375m) for active fire points. Furthermore, we used linear regression to find the relation between VIIRS FRP and FINN emissions of available grid cells and applied that to FRP values in VIIRS to estimate the emissions. We acknowledge these are first estimates and the performance of this technique using MODIS data and more complicated statistical works need to be studied further.

During this study, we did not primarily focus on improving anthropogenic emissions over the region in order to capture extreme pollution episode. However anthropogenic emissions are low in global emission inventories and needed to be improved (Jat et al., 2020). Moreover, very low biased concentrations for some days and trajectory results suggest the existence of some other sources, primarily anthropogenic sources, upwind of Delhi that should be studied more.

Furthermore, geostationary satellites can significantly improve our technique as more retrievals could improve~~intensify~~ the accuracy. In this study, VIIRS or (MODIS) provided only one or two retrievals in one day for each point, while recently~~new~~ launched geostationary satellites, such as GEMS, would provide high temporal frequency data that could improve emission inventories.

The choice of ~~On the other hand, choosing~~ the scaling~~multiplication~~ factor for increasing fire emissions was arbitrary in this study. ~~d~~Due to scarcity of observation data, we were not able to apply complicated mathematical scaling techniques based on data assimilation to scale the fire emissions (Saide et al., 2015). Low number of observation data also limited our statistical assessments. Agricultural fire emissions are small and vary day to day and atmospheric dynamics can significantly change their fate. We ~~didn't~~did not focus on physics and dynamics of the WRF-Chem model as they were ~~out of~~beyond the scope of this study. These are important limitations that readers have to keep in mind when exposed to the results.

Formatted: Font color: Accent 5

4. Summary and conclusion

In this study, we used WRF-Chem model to improve the air quality modeling during extreme pollution episode in November 2017 in the IGP. Various modifications on chemical boundary conditions and biomass burning emissions were tested. Multiple datasets, including ground measurements of $PM_{2.5}$, surface measurements and satellite AOD, and reanalysis models were used to evaluate the model. In our best scenario, CAM-Chem and MERRA-2 global models provided gaseous and aerosol chemical boundary conditions, respectively. Moreover, active fire points in VIIRS remote sensing instrument were used to fill the missing fire emission sources in FINN biomass burning emissions. Furthermore, the modified FINN emissions were scaled by a factor of 7 for an eight-days period to capture peak $PM_{2.5}$ concentrations. 24-hours averaged NMB, NME, and R averaged for all CPCB stations in Delhi during all the days in November were -16.6%, 27.6%, and 0.48, respectively, satisfying suggested benchmark criteria (Emery et al., 2017). These metrics significantly improved when excluding four extremely polluted days between Nov. 7th and Nov. 10th and all were within benchmark goals (Emery et al., 2017). Overall, we improved modeling results by incorporating different available datasets with each other.

The spatial performance of the model was also evaluated using VIIRS AOD. The model overestimated AOD over the domain with monthly averaged value of 0.58 (±0.4), confirming other studies (Kulkarni et al., 2020). Specifically, the model captured high AODs over Delhi and Punjab, overestimated AODs over central India, and underestimated AODs over eastern IGP. Our results indicate improving emissions, mostly anthropogenic emissions, in eastern IGP can significantly improve the air quality predictions. Our modeling results revealed secondary aerosols comprised 57% of total $PM_{2.5}$ concentration during November, confirming measurement studies (Gani et al., 2019). Secondary organic aerosols individually, comprised 27% of the total aerosol mass, while secondary inorganics accounted for 30% of the mass.

Back trajectories and vertical profiles were used to study the extreme pollution episode sources. Back trajectories showed a shift in trajectories from east to north on Nov. 7th. As a result, agricultural fire emissions were transported from Punjab to Delhi. The trajectories remained on north path for 3 days and then shifted again to east. However, the model underestimated the concentrations on these days. Vertical profiles showed a lot of smoke above boundary layers. These results indicated either the plume rise in the model released the emissions too high, or the model did not mix the smoke down fast enough. Social reasons can also add to high $PM_{2.5}$ concentrations during extreme pollution episodes, as people prefer to use their personal vehicles more often.

We also evaluated how QFED and FINN biomass burning emission inventories affected $PM_{2.5}$ concentration results over Delhi. QFED had worse statistics, which is mostly due to the inability of the inventory to capture the fire points over the domain. It can be attributed to both the technique and the resolution as QFED data have ~10 km resolution, whereas FINN data has ~1 km resolution, as other studies have shown FINN provides better data for India (Pan et al., 2020). We also found FINN underestimated fire emissions for some extremely high emission days, and needed to be scaled. It can be mostly because satellite retrievals reported thick smokes as clouds and missed them, as shown in other studies (Dekker et al., 2019).

The base scenario was chosen after evaluating the results for various chemical boundary conditions, including CAM-Chem, MERRA-2, MOZART, and CAMS global models. We found long-range transported dust from middle-east was not affecting Delhi's air quality during the extreme pollution episode. Moreover, we found MERRA-2 provided better aerosol products over India, although studies have shown they underestimate over India (Navinya et al., 2020). We also found in-domain dust emission sources in the border with Pakistan did not affect Delhi's air quality.

While the focus of current study was on PM, we found high ozone concentration in northern India. Averaged daytime ozone concentration over the domain was 77 ppb for November 2017, using the base scenario. Although the model overestimated ozone concentrations in Delhi by NMB of 18%, it indicates ozone is a problem that needs to be considered.

In general, air quality in IGP region is influenced by both local and regional sources. Although availability of new satellites such as GEMS, which covers some parts of India, can improve air quality predictions using data assimilation techniques, local emission inventories can vary day-by-day and significantly affect the modeling results. More works are required to quantify these impacts. Moreover, ozone concentrations showed a positive correlation with $PM_{2.5}$ over IGP. It suggests that control strategies should consider the regional co-benefits of $PM_{2.5}$/ozone perturbations simultaneously, which is the focus of our future work.

Data Availability The WRF-Chem output results for aerosol species are available from Iowa Research Online at https://doi.org/10.25820/data.006126 (Roozitalab et al., 2020).

Author Contributions BR and GRC designed the study; BR performed model simulations and analyzed the data with help from GRC and SKG provided measurements. BR and GRC wrote the paper with inputs from SKG.

Competing Interests The authors declare that they have no conflict of interest.

Acknowledgements We acknowledge the use of data and imagery from LANCE FIRMS operated by NASA's Earth Science Data and Information System (ESDIS) with funding provided by NASA Headquarters. Moreover, this paper contains modified Copernicus Atmosphere Monitoring Service Information 2017, where neither the European Commission nor ECMWF is responsible for any use that may be made of the information it contains. This work was funded in part by NASA HAQAST and ACMAP projects under awards NNX16AQ19G and 80NSSC19K094, respectively. The authors also thank Swagata Payra, Brent_Holben, S._N._Tripathi, and their staff for establishing and maintaining the Jaipur and Kanpur AERONET stations data used in this investigation. We thank Yannick Copin for his python code for plotting Taylor Diagrams. We also thank Dr. Pablo Saide for his comments on using QFED emission inventory.

References

Abdi-Oskouei, M., Carmichael, G., Christiansen, M., Ferrada, G., Roozitalab, B., Sobhani, N., Wade, K., Czarnetzki, A., Pierce, R., and Wagner, T.: Sensitivity of meteorological skill to selection of WRF-Chem physical parameterizations and impact on ozone prediction during the Lake Michigan Ozone Study (LMOS), Journal of Geophysical Research: Atmospheres, e2019JD031971, 2020.

Adhikary, B., Carmichael, G. R., Tang, Y., Leung, L. R., Qian, Y., Schauer, J. J., Stone, E. A., Ramanathan, V., and Ramana, M. V.: Characterization of the seasonal cycle of south Asian aerosols: A regional-scale modeling analysis, Journal of Geophysical Research: Atmospheres, 112, 2007.

Akagi, S., Yokelson, R. J., Wiedinmyer, C., Alvarado, M., Reid, J., Karl, T., Crounse, J., and Wennberg, P.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmospheric Chemistry and Physics, 11, 4039, 2011.

Amann, M., Purohit, P., Bhanarkar, A. D., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Kiesewetter, G., Klimont, Z., and Liu, J.: Managing future air quality in megacities: A case study for Delhi, Atmospheric environment, 161, 99-111, 2017.

Anenberg, S. C., Henze, D. K., Tinney, V., Kinney, P. L., Raich, W., Fann, N., Malley, C. S., Roman, H., Lamsal, L., and Duncan, B.: Estimates of the global burden of ambient PM 2.5, ozone, and NO 2 on asthma incidence and emergency room visits, Environmental health perspectives, 126, 107004, 2018.

Ashrafi, K., Motlagh, M. S., and Neyestani, S. E.: Dust storms modeling and their impacts on air quality and radiation budget over Iran using WRF-Chem, Air Quality, Atmosphere & Health, 10, 1059-1076, 2017.

Beig, G., Srinivas, R., Parkhi, N. S., Carmichael, G., Singh, S., Sahu, S. K., Rathod, A., and Maji, S.: Anatomy of the winter 2017 air quality emergency in Delhi, Science of The Total Environment, 681, 305-311, 2019.

Beig, G., Sahu, S. K., Singh, V., Tikle, S., Sobhana, S. B., Gargeva, P., Ramakrishna, K., Rathod, A., and Murthy, B.: Objective evaluation of stubble emission of North India and quantifying its impact on air quality of Delhi, Science of The Total Environment, 709, 136126, 2020.

Bosilovich, M., Lucchesi, R., and Suarez, M.: MERRA-2: File specification, 2015.

Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., Angevine, W., Evan, S., Dingwell, A., and Fast, J. D.: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1, 2013.

Buchard, V., da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D.: Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States, Atmospheric environment, 125, 100-111, 2016.

Chow, J. C., Lowenthal, D. H., Chen, L.-W. A., Wang, X., and Watson, J. G.: Mass reconstruction methods for PM 2.5: a review, Air Quality, Atmosphere & Health, 8, 243-263, 2015.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., and Dandona, R.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, The Lancet, 389, 1907-1918, 2017.

Conibear, L., Butt, E. W., Knote, C., Arnold, S. R., and Spracklen, D. V.: Residential energy use emissions dominate health impacts from exposure to ambient particulate matter in India, Nature communications, 9, 617, 2018a.

Conibear, L., Butt, E. W., Knote, C., Spracklen, D. V., and Arnold, S. R.: Current and Future Disease Burden From Ambient Ozone Exposure in India, GeoHealth, 2, 334-355, 2018b.

Courant, R., Friedrichs, K., and Lewy, H.: Über die partiellen Differenzengleichungen der mathematischen Physik, Mathematische annalen, 100, 32-74, 1928.

Cusworth, D. H., Mickley, L. J., Sulprizio, M. P., Liu, T., Marlier, M. E., DeFries, R. S., Guttikunda, S. K., and Gupta, P.: Quantifying the influence of agricultural fires in northwest India on urban air pollution in Delhi, India, Environmental Research Letters, 13, 044018, 2018.

Darmenov, A., and da Silva, A.: The quick fire emissions dataset (QFED)–documentation of versions 2.1, 2.2 and 2.4, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2013-104606, 32, 183, 2013.
David, L. M., Ravishankara, A., Kodros, J. K., Pierce, J. R., Venkataraman, C., and Sadavarte, P.: Premature mortality due to PM2. 5 over India: Effect of atmospheric transport and anthropogenic emissions, GeoHealth, 3, 2-10, 2019.
Dekker, I. N., Houweling, S., Pandey, S., Krol, M., Röckmann, T., Borsdorff, T., Landgraf, J., and Aben, I.: What caused the extreme CO concentrations during the 2017 high-pollution episode in India?, Atmospheric Chemistry and Physics, 19, 3433-3445, 2019.
Emery, C., Tai, E., and Yarwood, G.: Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes, Prepared for the Texas natural resource conservation commission, by ENVIRON International Corporation, 2001.
Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., and Kumar, N.: Recommendations on statistics and benchmarks to assess photochemical model performance, Journal of the Air & Waste Management Association, 67, 582-598, 2017.
Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., and Laepple, T.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), 2010.
Gani, S., Bhandari, S., Seraj, S., Wang, D. S., Patel, K., Soni, P., Arub, Z., Habib, G., Hildebrandt Ruiz, L., and Apte, J. S.: Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study, Atmospheric Chemistry and Physics, 19, 6843-6859, 2019.
Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., and Reichle, R.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419-5454, 2017.
Ghude, S. D., Chate, D., Jena, C., Beig, G., Kumar, R., Barth, M., Pfister, G., Fadnavis, S., and Pithani, P.: Premature mortality in India due to PM2. 5 and ozone exposure, Geophysical Research Letters, 43, 4650-4658, 2016.
Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S. J.: Sources and distributions of dust aerosols simulated with the GOCART model, Journal of Geophysical Research: Atmospheres, 106, 20255-20273, 2001.
Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957-6975, 2005.
Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmospheric Chemistry and Physics, 6, 3181-3210, 2006.
Gupta, M., and Mohan, M.: Validation of WRF/Chem model and sensitivity of chemical mechanisms to ozone simulation over megacity Delhi, Atmospheric Environment, 122, 220-229, 2015.
Guttikunda, S. K., and Calori, G.: A GIS based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in Delhi, India, Atmospheric Environment, 67, 101-111, 2013.
Guttikunda, S. K., Goel, R., and Pant, P.: Nature of air pollution, emission sources, and management in the Indian cities, Atmospheric environment, 95, 501-510, 2014.
He, L., Lin, A., Chen, X., Zhou, H., Zhou, Z., and He, P.: Assessment of MERRA-2 surface PM2. 5 over the Yangtze river Basin: Ground-based verification, spatiotemporal distribution and meteorological dependence, Remote Sensing, 11, 460, 2019.
HEI: Burden of Disease Attributable to Major Air Pollution Sources in India, 2018.
Hodzic, A., and Jimenez, J.: Modeling anthropogenically controlled secondary organic aerosols in a megacity: A simplified framework for global and climate models, Geoscientific Model Development, 4, 901-917, 2011.
Hodzic, A., Madronich, S., Kasibhatla, P., Tyndall, G., Aumont, B., Jimenez, J., Lee-Taylor, J., and Orlando, J.: Organic photolysis reactions in tropospheric aerosols: effect on secondary organic aerosol formation and lifetime, Atmospheric Chemistry and Physics, 15, 9253-9269, 2015.
Hsu, N., Lee, J., Sayer, A., Kim, W., Bettenhausen, C., and Tsay, S. C.: VIIRS Deep Blue aerosol products over land: Extending the EOS long-term aerosol data records, Journal of Geophysical Research: Atmospheres, 124, 4026-4053, 2019.
Huijnen, V., Wooster, M. J., Kaiser, J. W., Gaveau, D. L., Flemming, J., Parrington, M., Inness, A., Murdiyarso, D., Main, B., and Van Weele, M.: Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997, Scientific reports, 6, 26886, 2016.
Inness, A., Ades, M., Agusti-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., and Flemming, J.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515-3556, 2019.
Jain, S., and Sharma, T.: Social and Travel Lockdown Impact Considering Coronavirus Disease (COVID-19) on Air Quality in Megacities of India: Present Benefits, Future Challenges and Way Forward, Aerosol and Air Quality Research, 20, 1222-1236, 2020.
Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., and Wankmüller, R.: HTAP_v2. 2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmospheric Chemistry and Physics, 15, 11411-11432, 2015.
Jat, R., Gurjar, B. R., and Lowe, D.: Regional pollution loading in winter months over India using high resolution WRF-Chem simulation, Atmospheric Research, 105326, 2020.
Jethva, H., Chand, D., Torres, O., Gupta, P., Lyapustin, A., and Patadia, F.: Agricultural burning and air quality over northern India: a synergistic analysis using NASA's A-train satellite data and ground measurements, Aerosols and Air Quality Research, 18, 1756-1773, 2018.

Karambelas, A., Holloway, T., Kinney, P. L., Fiore, A. M., DeFries, R., Kiesewetter, G., and Heyes, C.: Urban versus rural health impacts attributable to PM2. 5 and O3 in northern India, Environmental Research Letters, 13, 064010, 2018.
Knote, C., Hodzic, A., Jimenez, J., Volkamer, R., Orlando, J., Baidar, S., Brioude, J., Fast, J., Gentner, D., and Goldstein, A.: Simulation of semi-explicit mechanisms of SOA formation from glyoxal in aerosol in a 3-D model, Atmospheric Chemistry and Physics, 14, 6213-6239, 2014.
Kulkarni, S. H., Ghude, S. D., Jena, C., Karumuri, R. K., Sinha, B., Sinha, V., Kumar, R., Soni, V., and Khare, M.: How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi?, Environmental Science & Technology, 54, 4790-4799, 2020.
Kumar, R., Naja, M., Pfister, G., Barth, M., Wiedinmyer, C., and Brasseur, G.: Simulations over South Asia using the Weather Research and Forecasting model with Chemistry (WRF-Chem): chemistry evaluation and initial results, Geoscientific Model Development, 5, 619-648, 2012.
Kumar, R., Barth, M., Madronich, S., Naja, M., Carmichael, G., Pfister, G., Knote, C., Brasseur, G., Ojha, N., and Sarangi, T.: Effects of dust aerosols on tropospheric chemistry during a typical pre-monsoon season dust storm in northern India, Atmospheric Chemistry & Physics, 14, 2014a.
Kumar, R., Barth, M., Pfister, G., Naja, M., and Brasseur, G.: WRF-Chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget, Atmos. Chem. Phys, 14, 2431-2446, 2014b.
Kumar, R., Barth, M., Nair, V., Pfister, G., Suresh Babu, S., Satheesh, S., Krishna Moorthy, K., Carmichael, G., Lu, Z., and Streets, D.: Sources of black carbon aerosols in South Asia and surrounding regions during the Integrated Campaign for Aerosols, Gases and Radiation Budget (ICARB), Atmospheric Chemistry and Physics, 15, 5415-5428, 2015a.
Kumar, R., Barth, M., Pfister, G., Nair, V., Ghude, S. D., and Ojha, N.: What controls the seasonal cycle of black carbon aerosols in India?, Journal of Geophysical Research: Atmospheres, 120, 7788-7812, 2015b.
Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., Kulkarni, S., Sperati, S., Soni, V. K., and Nanjundiah, R. S.: Enhancing Accuracy of Air Quality and Temperature Forecasts During Paddy Crop Residue Burning Season in Delhi Via Chemical Data Assimilation, Journal of Geophysical Research: Atmospheres, 125, e2020JD033019, 2020.
Liu, T., Marlier, M. E., DeFries, R. S., Westervelt, D. M., Xia, K. R., Fiore, A. M., Mickley, L. J., Cusworth, D. H., and Milly, G.: Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune, Atmospheric environment, 172, 83-92, 2018.
Marrapu, P., Cheng, Y., Beig, G., Sahu, S., Srinivas, R., and Carmichael, G.: Air quality in Delhi during the Commonwealth Games, Atmospheric chemistry and physics, 14, 10619-10630, 2014.
MoEF&CC: NCAP:Nation Clean Air Programme, Ministry of Environment, Forest & Climate Change, India, 2019.
Moorthy, K. K., Beegum, S. N., Srivastava, N., Satheesh, S., Chin, M., Blond, N., Babu, S. S., and Singh, S.: Performance evaluation of chemistry transport models over India, Atmospheric Environment, 71, 210-225, 2013.
Morrison, H., Curry, J., and Khvorostyanov, V.: A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description, Journal of the atmospheric sciences, 62, 1665-1677, 2005.
Murthy, B., Latha, R., Tiwari, A., Rathod, A., Singh, S., and Beig, G.: Impact of mixing layer height on air quality in winter, Journal of Atmospheric and Solar-Terrestrial Physics, 197, 105157, 2020.
Nakanishi, M., and Niino, H.: Development of an improved turbulence closure model for the atmospheric boundary layer, Journal of the Meteorological Society of Japan. Ser. II, 87, 895-912, 2009.
Nakoudi, K., Giannakaki, E., Dandou, A., Tombrou, M., and Komppula, M.: Planetary boundary layer height by means of lidar and numerical simulations over New Delhi, India, Atmospheric Measurement Techniques, 12, 2019.
Navinya, C. D., Vinoj, V., and Pandey, S. K.: Evaluation of PM2. 5 Surface Concentrations Simulated by NASA's MERRA Version 2 Aerosol Reanalysis over India and its Relation to the Air Quality Index, Aerosol and Air Quality Research, 20, 1329-1339, 2020.
Pan, X., Ichoku, C., Chin, M., Bian, H., Darmenov, A., Colarco, P., Ellison, L., Kucsera, T., da Silva, A., and Wang, J.: Six global biomass burning emission datasets: intercomparison and application in one global aerosol model, Atmospheric Chemistry and Physics, 20, 969-994, 2020.
Pant, P., Shukla, A., Kohl, S. D., Chow, J. C., Watson, J. G., and Harrison, R. M.: Characterization of ambient PM2. 5 at a pollution hotspot in New Delhi, India and inference of sources, Atmospheric Environment, 109, 178-189, 2015.
Provençal, S., Buchard, V., da Silva, A. M., Leduc, R., Barrette, N., Elhacham, E., and Wang, S.-H.: Evaluation of PM2. 5 surface concentration simulated by Version 1 of the NASA's MERRA Aerosol Reanalysis over Israel and Taiwan, Aerosol and air quality research, 17, 253, 2017.
Rajak, R., and Chattopadhyay, A.: Short and long-term exposure to ambient air pollution and impact on health in India: a systematic review, International journal of environmental health research, 1-25, 2019.
Ramanathan, V., and Carmichael, G.: Global and regional climate changes due to black carbon, Nature geoscience, 1, 221, 2008.
Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., and Hair, J.: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, Journal of Climate, 30, 6823-6850, 2017.
Saide, P. E., Peterson, D. A., da Silva, A., Anderson, B., Ziemba, L. D., Diskin, G., Sachse, G., Hair, J., Butler, C., and Fenn, M.: Revealing important nocturnal and day-to-day variations in fire smoke emissions through a multiplatform inversion, Geophysical Research Letters, 42, 3609-3618, 2015.

Saikawa, E., Trail, M., Zhong, M., Wu, Q., Young, C. L., Janssens-Maenhout, G., Klimont, Z., Wagner, F., Kurokawa, J.-i., and Nagpure, A. S.: Uncertainties in emissions estimates of greenhouse gases and air pollutants in India and their impacts on regional air quality, Environmental Research Letters, 12, 065002, 2017.

Schroeder, W., Oliva, P., Giglio, L., and Csiszar, I. A.: The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment, Remote Sensing of Environment, 143, 85-96, 2014.

Sharma, A., Ojha, N., Pozzer, A., Mar, K. A., Beig, G., Lelieveld, J., and Gunthe, S. S.: WRF-Chem simulated surface ozone over south Asia during the pre-monsoon: effects of emission inventories and chemical mechanisms, Atmospheric Chemistry and Physics, 17, 14393-14413, 2017.

Sitch, S., Cox, P., Collins, W., and Huntingford, C.: Indirect radiative forcing of climate change through ozone effects on the land-carbon sink, Nature, 448, 791-794, 2007.

Song, Z., Fu, D., Zhang, X., Wu, Y., Xia, X., He, J., Han, X., Zhang, R., and Che, H.: Diurnal and seasonal variability of PM2. 5 and AOD in North China plain: Comparison of MERRA-2 products and ground measurements, Atmospheric Environment, 191, 70-78, 2018.

Turpin, B. J., and Lim, H.-J.: Species contributions to PM2. 5 mass concentrations: Revisiting common assumptions for estimating organic mass, Aerosol Science & Technology, 35, 602-610, 2001.

Venkataraman, C., Brauer, M., Tibrewal, K., Sadavarte, P., Ma, Q., Cohen, A., Chaliyakunnel, S., Frostad, J., Klimont, Z., and Martin, R. V.: Source influence on emission pathways and ambient PM2. 5 pollution over India (2015–2050), Atmospheric Chemistry and Physics Discussions, 18, 8017-8039, 2018.

Vijayakumar, K., Safai, P., Devara, P., Rao, S. V. B., and Jayasankar, C.: Effects of agriculture crop residue burning on aerosol properties and long-range transport over northern India: a study using satellite data and model simulations, Atmospheric research, 178, 155-163, 2016.

Wang, W., Bruyere, C., Duda, M., Dudhia, J., Gill, D., Kavulich, M., Keene, K., Lin, H., Michalakes, J., and Rizvi, S.: User's Guides for the Advanced Research WRF (ARW) Modeling System, version 3, 2018.

WHO: WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment, Geneva: World Health Organization, 2006.

WHO: WHO global urban ambient air pollution database (update 2016), Geneva. Diunduh, 2016.

Wiedinmyer, C., Akagi, S., Yokelson, R. J., Emmons, L., Al-Saadi, J., Orlando, J., and Soja, A.: The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625, 2011.

World-Bank: India: Systematic Countary Diagnostic; Realizing the promise of prosperity, World Bank 126284-IN, 2018.

Xiu, D.: Numerical methods for stochastic computations: a spectral method approach, Princeton university press, 2010.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for simulating aerosol interactions and chemistry (MOSAIC), Journal of Geophysical Research: Atmospheres, 113, 2008.

Zhao, C., Liu, X., Leung, L., Johnson, B., McFarlane, S. A., Gustafson Jr, W., Fast, J. D., and Easter, R.: The spatial distribution of mineral dust and its shortwave radiative forcing over North Africa: modeling sensitivities to dust emissions and aerosol size treatments, Atmospheric Chemistry and Physics, 10, 8821, 2010.