

Responses to the referee comments:

## Assessing the climate and air quality effects of future aerosol mitigation in India using a global climate model combined with statistical downscaling

*Miinalainen, T., Kokkola, H., Lipponen, A., Hyvärinen, A.-P., Soni, V. K., Lehtinen, K. E. J., and Kühn, T.: Assessing the climate and air quality effects of future aerosol mitigation in India using a global climate model combined with statistical downscaling, Atmos. Chem. Phys. Discuss. [preprint], <https://doi.org/10.5194/acp-2022-513>, in review, 2022.*

We thank Anonymous Referees for the efforts made to review our paper and for the valuable comments. Our responses are written below each comment separately. The referee comments are marked with *yellow color and italic*, and the author replies are marked with gray color. The original manuscript text is marked with *pink color*, and updated text with *dark magenta*. The line numbers refer to the original, submitted version of the manuscript.

### Replies to the comments made by the Anonymous Referee #1:

*This manuscript presents global model simulation results over India from ECHAM for air quality and radiative forcing under present and future emission scenarios (from GAINS) through 2030. For the region covering Delhi, the results were downscaled using random-forest corrections using multiple emission, met, and orography variables.*

*While the radiative forcing calculations from ECHAM is a known path, the use of the same to bias correct and estimate air quality for a city is new. The later methods have been used, but not for model resolutions at 1.9 degrees. Machine Learning (ML) approach is a new and emerging field and the benefits of using a global model for both air quality and climate applications cannot be overlooked. While the methodology is well explained for correcting the model results with biases from ML, the statistics also improved after the corrections, the gaps between the measured and model-corrected numbers is still significant.*

We thank Anonymous Referee #1 for the useful comments.

As the referee mentioned, the RF-corrected PM<sub>2.5</sub> does not fully reproduce the station-averaged PM<sub>2.5</sub> for the testing phase data, especially over short time scales. This is analyzed and discussed in Section 3.1.

However, if the aim is to capture the long-term trends correctly, the RF-corrected PM2.5 performs as adequately as the PM2.5 obtained from the dispersion model System for Integrated modelLling of Atmospheric coMposition (SILAM). In the Table below, we show the comparison statistics between modelled and mean of the New Delhi ground stations for the testing phase period 22.8.2018-31.12.2019.

Compared to the mean of the measurement station PM25 for 22.8.2018-31.12.2019, daily average values			
	ECHAM PM2.5	SILAM PM2.5	RF-corrected PM2.5
<b>RMSE (<math>\mu\text{g}/\text{m}^3</math>)</b>	113	65	56
<b>MRE (%)</b>	-36	27	33
<b>MAE (<math>\mu\text{g}/\text{m}^3</math>)</b>	78	46	39
<b>R<sup>2</sup></b>	-0.63	0.46	0.59
<b>Pearson correlation</b>	0.19	0.79	0.79

The error statistics show that the RF-corrected PM2.5 correlates with the measured data at the same level as SILAM PM2.5 does. Furthermore, the mean absolute error (MAE) and root mean squared error (RMSE) are smaller for RF-corrected PM2.5 than for SILAM data. The mean relative error (MRE) is smaller for SILAM than for RF-corrected data, but the difference is less than 10 percent units. This comparison suggests that the performance of the downscaling approach is comparable to air quality models if the aim is to utilize data to applications where long-term trends in air quality are central.

*The scenario analysis for air quality primarily hinges on the reproductive capacity of the model and the only question that is not clearly answered is why extract air quality data from such a coarse model (when the problem is known that coarser models have hard time replicating high-density urban areas with very distinct emission characteristics)? Especially, since FMI and IMD (author organizations) are known to conduct chemical transport modeling for air quality at much better resolutions globally and in India.*

The referee is correct that the downscaled PM2.5 does not fully correspond to outputs from high resolution models when comparing day-to-day variation. However, this was not the aim of this project, and we presume that the concept of our manuscript was misinterpreted by the referee.

The main idea of our study is to expand the possibilities to utilize global model data in additional applications such as local air quality analysis. As we describe in the manuscript text (line 522 and forward), one advantage of using global-scale models for analyzing the effects of aerosol mitigation is that one can simulate fairly long time periods (decades to even a century), and that the simulations typically cover the whole globe. This is especially relevant for future projections of climate and air quality. With the help of downscaling, one can “zoom in” to a

very specific location and analyze how the global or local scale emission mitigation affects. Up to our understanding, air quality models are computationally more expensive due to the high grid resolution, and therefore the simulated time periods are shorter than with GCMs, and the simulation times can be much longer.

Furthermore, another advantage of using downscaling enhanced PM2.5 from a GCM is that one can simultaneously analyze the effects of aerosol mitigation on various other climatic processes. In our case, we focused on the radiative forcing, but one could, for instance, analyze global precipitation patterns or low-level cloud formation at very distinct regions in the globe.

One plausible application of downscaling PM2.5 could be simply to use it as a “quick tool” to evaluate how a mitigation scheme affects future climate or air quality in different parts of globe. The intention of this study is not to present a method which can be used to replace chemical transport models. Instead, the main idea is to provide a relatively light-weight tool that can be used to assess simultaneously the climate and air quality effects of future aerosol emission reduction measures.

Moreover, air quality models require simulated data from global climate models for modelling the future climate. Regional modelling with fine grid resolution requires input for boundary conditions for large scale atmospheric dynamics, and therefore a separate global model simulation is needed before the actual air quality model simulation can be performed. Our intention was to explore if we could use the global model data directly also to analyzing surface air quality, especially if there is no need to study the underlying physical mechanisms in a detailed way.

We have updated the Introduction text by adding the following sentences to line 71:

“The aim of this work is to explore more versatile methods to utilize global climate model data. The combination of ECHAM-HAMMOZ and statistical downscaling could be used as a relatively light-weight tool for applications where one aims to assess the effects of different aerosol mitigation strategies simultaneously on larger scale climate, and on local air quality.”

On line 54 we added:

“Air quality models, on the other hand, are not well suited to assess climate impacts of emission changes. Furthermore, when modelling future emission scenarios, air quality models require input from global climate models to set boundary and initial conditions.”

*Why use a city like Delhi with so many stations with 0% data available in the ML testing phase? Why not use a city in Europe or the US with good availability rates and good representation of the sources, to show that the model is capable of replication after the bias corrections?*

We assume this is a misunderstanding as the majority of the stations used in this study have a coverage of 80 to 90% for the testing phase (see Table S1).

To emphasize this point, we have modified the text at lines 141-142 from

“the final RF bias correction leans more on the year 2018 (and 2019) data since there is a higher number of measurement points for those years.” to

“the final RF bias correction leans more on the year 2018 data since there is a higher number of measurement points for that year.”

and added a sentence after line 147:

“All in all, we had good temporal coverage for the measurement station data for both training and testing phases.”

*The one drawback of the manuscript is the selection of the case study city (Delhi) -- which has strong seasonal trend, strong diurnal trend, and distinct sources (for SO<sub>2</sub>, BC, and OC) over the months. A city(s) or region(s) with consistent emission loads would cut down some uncertainty in the model and corrections methods and then apply to regions like India and China.*

We chose New Delhi as a case study city because of its diurnal and seasonal trends, as it is ideal for this kind of study. The PM<sub>2.5</sub> concentrations are at a high level for most of the year in New Delhi. Furthermore, Delhi National Capital Territory (NCT) is densely populated and relatively large in surface area. In addition, the ground measurement station network in New Delhi is very extensive and there was a sufficient amount of data available for our study.

Furthermore, we respectfully disagree with the referee that the unique characteristics backgrounding New Delhi pollution profile would make Delhi unsuitable for this kind of analysis. On the contrary, as we mention right in the first sentence of the abstract, we aimed to study the potential of the downscaling procedure. As the referee described, the New Delhi PM<sub>2.5</sub> is not constant all year around but has a lot of seasonal variation and strong dependence on anthropogenic sources. This makes New Delhi a good target region to explore how well downscaling can capture local tendencies, such as short- and long-term trends in PM<sub>2.5</sub>.

Anthropogenic aerosol emissions are relatively large in India (~15% of the global BC emissions), and they are not projected to decrease at the same rate as global emissions (see manuscript Table 1). That is why we considered that India and New Delhi are interesting areas to study, as the aerosol mitigation is expected to bring clear benefits to the local air quality, but the net radiative forcing due to simultaneous mitigation of three species (BC, OC and SO<sub>2</sub>) was unclear. We have updated the Introduction text to give readers a clear impression of why New Delhi was chosen as a study of interest by adding at line 80:

“In addition, New Delhi PM<sub>2.5</sub> shows strong seasonal variation and high dependence on anthropogenic emissions (Bawase et al., 2021). This makes New Delhi an ideal target region to study how well downscaling can capture short- and long-term trends in PM<sub>2.5</sub>.”

We have also added after line 127:

“The ground measurement station network in New Delhi is extensive and there was a good

amount of measurement data available, which made New Delhi ideal location for this kind of study.”

Unfortunately, it seems that we have not reported clearly enough in our manuscript that our bias correction model is meant to be built separately for each city. Though technically possible, it would require a very large global training dataset representing a wide variety of conditions to construct an ML-based model that would be able to generalize from one region to another. In practice, this type of global dataset does not exist. The characteristics of PM<sub>2.5</sub> concentrations trends and how they depend on prevailing atmospheric conditions and the assumed local emissions are unique for each city/region. We used the same city for training and testing the RF Model to ensure representative training data for the model.

We have emphasized this in Section 2.7 by adding after line 247:

“One should keep in mind that in this bias correction approach, the models are assumed to be city specific. In other words, RF models trained with New Delhi data are not meant for correcting biases in ECHAM-HAMMOZ PM<sub>2.5</sub> for other cities.”

*Line 237-242 and 290: It is not clear if the emissions and other variables extracted and used are still at the ECHAM resolution or further downscaled to support a region of 30km x 30km over Delhi? (L290) is an important observation - When making the bias corrections, besides the model grid variables, are there any variables that are segregating the Delhi area signatures for a better fit?*

We thank the referee for making us aware that this aspect was not clear enough in the manuscript text. The input feature values used in the RF bias correction are all from one specific ECHAM-HAMMOZ simulation, and the data were not further downscaled to represent higher resolution grid data. Downscaling from T63 to L290 would require a large amount of external fine resolution data and would not serve the purpose of lowering the computational costs.

Instead, we post processed the input feature data by extracting values from the original ECHAM-HAMMOZ data of T63 horizontal resolution. We produced input feature data either by extracting data from one single grid box (‘point’), or by summing (‘fldsum’) or averaging (‘fldmean’) over an area surrounding New Delhi (72 – 82 ‘W, 24 - 34 ‘N). This means that some of the input features represent more the regional changes in the simulated atmosphere, whereas the point features show more the local changes.

In our modeling set up, it is assumed that the measurement data from ground stations incorporate information about the very detailed characteristics affecting New Delhi pollution levels, such as urban infrastructure or very specific emission sources. Including detailed external data as an input feature (e.g., info about public traffic routes) would have reduced the feasibility of our modeling approach since external data is easily subject to change and would need to be kept updated constantly or be based on additional scenarios or assumptions. Then again, the absence of detailed, local information is most likely the reason why our modelling

approach is not capturing short-term variations very well. Developing such a downscaling bias correction method that also includes external data could be an opportunity for future studies.

We have modified the sentence at line 235 from:

“The input features used in the random forest fitting are presented in Table 3.” to

“The input features used in the random forest fitting are purely ECHAM-HAMMOZ data, and we did not use any external data as input predictors. The input features are presented in Table 3.”

We also added a description of how the input features were selected. We added after line 237:

“We selected the input features mainly based on the average feature importance values. At first, we formed a larger set of input features and ran the RF modelling with this larger set of input features. After that, we pruned out those variables that had an average feature importance value close to zero, i.e., they were assumed to have a very minor influence on the RF correction. Finally, we identified pairs of strongly correlated input features (with correlation coefficients larger than 0.9) and removed one of the two features.”

In addition, we added after line 521:

“Furthermore, the method we proposed could be further advanced by including very detailed information about local infrastructure and traffic statistics, for instance. This could potentially improve the correction of short-term trends in New Delhi PM2.5.”

*The results and conclusions of the study in terms of AQ and climate benefits of reducing emissions are as expected. However, since Delhi is the most polluted area/city in the world with not only a complex mix of emission sources, but also a complex mix of political and institutional setup to manage these emissions. While the manuscript presented % changes (benefits for air quality and RF), the discussion doesn't include any explanation on how these % emission reductions will be achieved in the Delhi area. It is understood that the emissions work comes from a different model (GAINS). Since the manuscript very specifically mentions and analyzes data for one city only, it would be appropriate to also discuss this space.*

We would like to highlight here that it is not explicitly evident that future aerosol mitigation would bring a net negative forcing over India. As many studies have shown (for instance, Allen et al.,2020), simultaneous reductions in both scattering and absorbing aerosols are expected to reduce the net cooling effect of aerosols on a global scale, thereby inducing a warming contribution.

The aerosol mitigation scenarios used in this study were part of the ECLIPSE V6b emission scenarios, which were designed with the GAINS model. The underlying assumptions in ECLIPSE are idealized in the sense that emission reductions are based on assumed perfect achievement of currently valid legislation (CLE) or full implementation of all currently available technologies to reduce emissions (MFR). In this sense these scenarios could be seen as maximum possible emission reductions from the 2020 viewpoint. Analyzing the underlying political actions needed

to achieve the proposed emission reductions is beyond the scope of this study. A more detailed description of the ECLIPSE scenarios can be found from e.g. Stohl et al. (2015), Belis et al. (2022) and Klimont et al. (2017).

However, we have added after line 162:

"Both globally and in the New Delhi surroundings, the sectors contributing most to reductions in BC emissions were the domestic and traffic sectors, while for OC, the domestic and waste sectors had the most significant reductions. For SO<sub>2</sub>, the largest reductions were projected in the MFR scenario for the energy and industry sectors."

In addition, we have extended the discussion at Section 3.2 by adding after line 329:

"As described in Section 2.2, the largest reductions for anthropogenic aerosol emissions under future mitigation scenario were assumed to come from domestic (BC & OC), waste (OC), traffic (BC), industry (SO<sub>2</sub>) and energy (SO<sub>2</sub>) sectors. The emissions from different sectors contribute differently to the atmospheric aerosol concentrations, because the assumed size distributions, mixing states and injection height at emission time are different. Each sector therefore also contributes differently to the modelled PM<sub>2.5</sub> values."

*While there is merit to a new methodology to be able to model AQ data along with the climate data, the manuscript lacks punch and I am afraid that these correction results will be hard to replicate in another setting.*

While we appreciate the referee's feedback, we respectfully disagree that the proposed method would not be applicable for analyzing the effects in another city or region if trained and applied with suitable data. We do, however, recognize that the performance of the bias correction may be different for each target city.

As mentioned above, there might be a misunderstanding in that we would propose to apply the model trained with New Delhi data to correct air quality levels for a different city. In case this referee's comment is based on that kind of assumption, we have elucidated this in the manuscript text as mentioned above in one of the answers.

In addition to the already mentioned improvements, we have gone through the whole manuscript text and highlighted the core concepts to sharpen the key messages of our study.

\*\*\*\*\*

Replies to the comments made by the #Anonymous referee 2:

*Review of "Assessing the climate and air quality effects of future aerosol mitigation in India using a global climate model combined with statistical downscaling"*

*The manuscript explores the possibilities of using a global climate model to investigate the effects of aerosol mitigation in India. A machine learning (ML) approach using Random Forest regression is used to downscale PM<sub>2.5</sub> concentrations over a polluted city, New Delhi with the help of measured PM<sub>2.5</sub> concentrations. Different PM loading future scenarios are projected and compared with the uncorrected and ML-corrected model outputs. The effects of aerosol mitigation are investigated in terms of radiative effects and effective radiative forcing under the PM future scenarios. The authors claim the improvement of global-scale model output in simulating the PM<sub>2.5</sub> concentration over a small domain and their effectiveness in estimating the radiative impacts. The study demonstrates the potential of the emerging technique of ML in improving the large-scale model output in the process of statistical downscaling. The study is relevant and unique as mentioned above, and has a significant contribution to the relevant scientific domain. However, some concerns remain significant and need to be considered before publishing.*

#### *General Comments*

*The manuscript focuses on two aspects. (1) Demonstration of an ML technique in improving a global-scale model to simulate the PM<sub>2.5</sub> concentrations over a small region via statistical downscaling under different emission scenarios. (2) Estimating the radiative effects of future aerosol scenarios using the RF-corrected model simulations. The manuscript structure is difficult to follow until reaching the present 'Conclusions' section which is not a conclusion, but a nice overview/summary of the entire work.*

We thank the referee for the insightful suggestions to improve the manuscript structure and content. We have re-organized some of the subsections in Section 2 and renamed Section 4 as suggested.

*If the authors want to highlight their simulation results regarding the impact of future aerosol mitigation, more discussion is needed with proper references to the existing findings, else it remains as a technical paper demonstrating the potential of ML in statistical downscaling. Currently, the physical mechanisms for some of the simulation outcomes are not given/found, but some tentative reasons are proposed. Many studies are documented the current aerosol-impact scenario using multiple scientific techniques (in situ, remote sensing, etc.) and future projections also for the Indian region.*

We have added to Section 3.2. after line 351 the following:

*“Moreover, we compared the average regional convective precipitation values for the New Delhi surroundings (see Section 2.6). The average convective precipitation for June and July was  $(4.4 \pm 4.1) \times 10^{-5}$  mm s<sup>-1</sup> and  $(4.2 \pm 4.1) \times 10^{-5}$  mm s<sup>-1</sup> for PRES and CLE\_2030, respectively. For MITIG\_2030, the average summertime convective precipitation was  $(3.1 \pm 3.8) \times 10^{-5}$  mm s<sup>-1</sup>. Based on the Mann-Whitney U-test, there was a statistically significant difference between MITIG\_2030 and PRES summertime precipitation in the New Delhi surroundings. Decreased*



summertime convective precipitation indicates less washout, allowing for higher surface PM2.5 concentrations in the MITIG\_2030 simulation."

We have added to line 426:

"The areas for which we observed the largest changes in atmospheric SW absorption are almost the same regions as obtained by Prakash et al. (2020), even though their analysis handled only the BC emissions due to road transport."

Furthermore, we have added to Section 3.3., after line 410:

"Das et al. (2020) reported an increase in dust burden over central India, due to decreases in summertime precipitation. Our results also showed a decrease in convective precipitation when comparing MITIG\_2030 values to PRES (see Section 3.2), which may have contributed to the increased dust loadings in MITIG\_2030. However, the modelled natural variation in precipitation patterns is large, and therefore there is significant uncertainty related to this mechanism."

*To ascertain the second aspect of the current manuscript, the first part needs to be flawless and should be explained confidently. Many parts of the manuscript are confusing which calls for further explanations for the smooth reading.*

We thank the referee for bringing up that some parts of the manuscript may be confusing to the reader. We have gone through the entire manuscript and done our best to improve the general readability of the text.

*One of the highlights of the findings is that the improvement of air quality is mostly due to the reduction of OC loading. However, the negative radiative forcing is attributed to the reduction of BC emission. This is an example of confusion arising while going through the manuscript.*

This was an excellent comment and helped us to understand which parts of our manuscript are confusing. The reason why OC influences air quality more than BC, though BC has stronger effects on the radiative balance, is related to the optical properties of these aerosol species. The atmospheric mass load of BC is relatively small compared to the mass load of OC. The radiative forcing per mass load of BC, on the other hand, is much bigger than for OC, as the optical properties differ. That is why in this study the emission reductions of OC influence PM2.5 concentrations more, while BC-reductions have a larger impact on radiative forcing. Due to the strong absorptive nature of BC, the negative effect comes from reduction in absorption of shortwave radiation which outweighs the radiative effect of OC which would lead to a positive radiative effect. Permadi et al., (2018), who studied the co-benefits of BC emission reductions on air quality and radiative forcing over South East Asia also found a similar decrease in BC radiative forcing when studying emission reductions in Indonesia and Thailand. They also concluded that this negative radiative effect is due to decrease in absorption of shortwave radiation by BC. This is now discussed in the revised manuscript

We have updated Section 3.2. by adding to line 329:

“Table 1 shows that the total OC emission mass is more than double the mass of emitted BC, both on a global and regional scale. Furthermore, the contribution of sulfate to the long-term PM<sub>2.5</sub> in New Delhi is known to be lower than the contribution of OC (Lalchandani et al., 2022), though sulfate might have a more significant role during winter haze periods.”

Furthermore, we clarified Section 3.3 discussion by adding to line 398:

“While OC contributes strongest to the surface PM<sub>2.5</sub> loadings (see Section 3.2), BC makes the strongest per unit mass contribution to the radiative forcing and hence can outweigh the cooling effects of OC and SO<sub>4</sub>, especially over bright backgrounds.”

We also added after line 431:

“Permadi et al., (2018) conducted a similar study for BC emissions reductions in Indonesia and Thailand. Their outcome was that along improved air quality, emissions reductions led to a decrease in absorption of SW radiation by BC, and therefore the resulting radiative forcing was negative.”

*The map of India shown in Fig. 3 is not matching with the maps published by the institutions that provided the insitu data nor with one of the authors' affiliated institution. Please correct the map as per the source or remove the political boundaries as per the journal's recommendations.*

We thank the referee for pointing out this issue. We have removed the political boundaries from Figures 3 and S3.

*Language also needs improvement. The main concerns are listed below.*

*Methodology:*

- *Why does section 2.2 stand apart from sections 2.6 and 2.7?*

We have moved Sections 2.2 and 2.3 below Section 2.5 and updated the numbering of the subsections.

- *The hyper parameters were adjusted using different combinations based on the best error statistics. Can you please show the performance of the validation test data?*

This statement in the manuscript was somewhat inaccurate as the hyper parameter selection seemed not to affect the model performance substantially. The error statistics for testing different RF parameter combinations are presented in the Figure below.

max_depth: None, Max_features = 7	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.56	56.52	56.43	56.52	56.54
MRE (%)	32.29	31.30	31.98	32.21	32.26	
MAE ( $\mu\text{g}/\text{m}^3$ )	38.69	38.46	38.57	38.57	38.65	
R <sup>2</sup>	0.59	0.59	0.59	0.59	0.59	
Pearson correlation	0.79	0.79	0.80	0.79	0.79	

max_depth: None, Max_features = None	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.79	56.81	56.90	56.84	56.76
MRE (%)	33.15	33.52	33.46	33.08	33.11	
MAE ( $\mu\text{g}/\text{m}^3$ )	39.02	39.09	39.16	39.09	39.01	
R <sup>2</sup>	0.59	0.59	0.58	0.59	0.59	
Pearson correlation	0.79	0.79	0.79	0.79	0.79	

max_depth: None, Max_features = 15	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.73	56.43	56.45	56.45	56.53
MRE (%)	32.15	32.54	32.63	33.27	32.62	
MAE ( $\mu\text{g}/\text{m}^3$ )	38.57	38.61	38.69	38.81	38.75	
R <sup>2</sup>	0.59	0.59	0.59	0.59	0.59	
Pearson correlation	0.79	0.79	0.79	0.79	0.79	

max_depth: 10, Max_features = 7	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.78	56.78	56.59	56.54	56.51
MRE (%)	31.72	32.39	32.58	31.78	31.91	
MAE ( $\mu\text{g}/\text{m}^3$ )	38.70	38.77	38.77	38.56	38.57	
R <sup>2</sup>	0.59	0.59	0.59	0.59	0.59	
Pearson correlation	0.79	0.79	0.79	0.79	0.79	

max_depth: 5, Max_features = 7	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.93	57.00	57.12	57.09	57.12
MRE (%)	31.65	31.18	31.80	31.87	31.76	
MAE ( $\mu\text{g}/\text{m}^3$ )	38.77	38.76	38.99	38.96	38.96	
R <sup>2</sup>	0.58	0.58	0.58	0.58	0.58	
Pearson correlation	0.79	0.79	0.79	0.79	0.79	

max_depth: 15, Max_features = None	nr of trees	20	50	100	200	300
	RMSE ( $\mu\text{g}/\text{m}^3$ )	56.84	56.83	56.85	56.79	56.68
MRE (%)	34.39	33.11	33.62	33.26	33.19	
MAE ( $\mu\text{g}/\text{m}^3$ )	39.33	39.09	39.20	39.06	39.02	
R <sup>2</sup>	0.59	0.59	0.59	0.59	0.59	
Pearson correlation	0.79	0.79	0.79	0.79	0.79	

As the statistics show, our RF model setup is not overly sensitive to the selection of hyper parameters. Hastie et al. (2009, p. 590) reported a similar tendency. Based on their experiences, random forests require very little tuning.

Furthermore, our modeling approach had 31 RF models (one per measurement station), and we used the same RF hyperparameters in each of them, as we did not tune each station-specific RF model separately. We therefore decided to use the default, recommended values in the RF models.

We have modified the statement at lines 111-112 from

“The hyper parameters were adjusted by testing different combinations and chosen based on the best error statistics:” to

“Based on testing different combinations, our RF model setup was not very sensitive to the choice of hyper parameters. Furthermore, we aimed to have the same hyper parameters in all of the individual RF models (see Section 2.6). Therefore, we used the recommended default values for the hyper parameters.”

- *What do you mean by setting the depth of each tree to infinity? How can you make sure of avoiding over-fit while keeping the depth of the tree as infinity? What are the criteria for fixing the number of trees? It is said that a default value of 100 is taken as the number of trees in the present study. Why 100?*

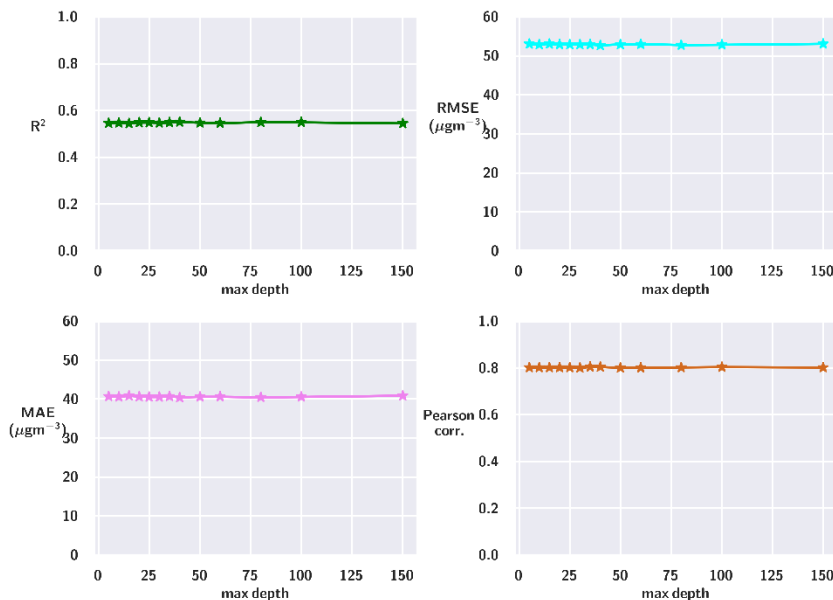
By setting the depth to infinity we indicated that we did not set a limit to the tree depth. This was a slightly misleading way to report that the “max\_depth” parameter in the Scikit Learn Python module was left as default value, “None”. When there is no maximum depth assigned, the algorithm will expand the regression tree nodes until the so-called leaves meet the purity criteria (mean squared error).

We rewrote the sentence at lines 112-113, originally:

“The maximum depth for each tree was set to infinity and the number of trees was the default value, 100.”, and after update:

“There was no maximum depth for each tree set and the depth of each tree was determined automatically based on the splitting criterion. The number of trees was the default value, 100.”

We tested how the maximum depth of the trees affects the model performance by altering the max\_depth parameter and applying the trained model to year 2020 data (i.e., outside of training and testing phase data). The results are presented in the Figure below.



As Figure above shows, our RF modelling approach is not very sensitive to the max\_depth parameter. Therefore, we chose not to fix the maximum depth of a tree in our modelling approach. Note that the root mean squared error and mean absolute error values are slightly larger for the year 2020 as the COVID19 pandemic caused an unusual, long-term drop in the

surface PM2.5, which was not accounted for in the ECLIPSE emission inventories.

The number of trees was set to 100 based on a trial-and-error approach. There was no significant improvement if we increased the number of trees (See Table above). Due to these reasons, the hyper parameters were not adjusted based on tight cross-validation routines, but were selected near the recommended, default values.

- *Why the feature importance values are normalized, by doing so what is the chance of smoothing the non-linearity of the dependence of the input variables? Isn't there any criteria to fix the number of input variables? As the authors have pointed out the input variables are mutually correlated which is obvious in the atmosphere, including all of them may lead to overfitting. What is the authors' claim on this point?*

The feature importance values are normalized to make them more comparable to each other. They represent the contribution of each variable to the reduction of the error criteria, and the normalization scales these to a scale from zero to one. Without normalization, the importance values would be more difficult to interpret as they would always depend on the computed total error of the particular RF training setup. Furthermore, the normalizing routine is a default setting in the applied SciKit Python library.

It is true that, since there are correlated input variables, there is a small risk of overfitting. We did preselection for the input variables based on feature importance values. We excluded variables that had an average feature importance value close to zero. Furthermore, for each regression tree, the maximum number of input features to be used when looking for the optimal split was fixed to 7 (See Section 2.2), so that there is a randomized set of features for splits. However, we did not carry out a stringent optimization routine to prune the number of input variables to as low as possible. This was because we had 31 stations and thereby 31 RF models. Each of the stations has distinct characteristics, and therefore the optimal set of input variables is slightly different for each station. We estimated that using the same set of input variables for all RF models could result in a more harmonized outcome.

In order to minimize the risk of overfitting, we have re-evaluated the set of input features and analyzed the impact of highly correlated input features on the RF predictions. Based on these, we have removed variables that had a correlation over 0.9 with another input feature. The excluded input variables were boundary layer height (pbl\_h), OC burden (burden\_OC) and PM2.5 due to OC (PM25\_OC). We have repeated the RF corrections with the pruned set of input variables and updated manuscript Table 3, and also Sections 3.1 and 3.2 with the new results. Note that the differences to the original values are fairly small.

As mentioned in the responses above, we have added the following text at line 237:

*“We selected the input features mainly based on the average feature importance values. At first, we formed a larger set of input features and ran the RF modelling with this larger set of*

input features. After that, we pruned out those variables that had an average feature importance value close to zero, i.e., they were assumed to have a very minor influence on the RF correction. Finally, we identified pairs of strongly correlated input features (with correlation coefficients larger than 0.9) and removed the one of the two features that had a smaller average feature importance value.”

Furthermore, we modified the sentence at lines 302-303 from

“As some of the input feature variables are correlated with each other, this might affect the calculated values, especially considering the random component of RF.” to

“Some of the input feature variables are correlated with each other. Even though we removed input features which were correlated most strongly to other, included input features (see Section 2.6), the remaining correlated input features might affect the calculated values, especially considering the random component of RF.”

In addition, we noticed that there were some inconsistent descriptions of the modelling setup, and we updated those to correspond to our modeling parameters. We noticed that the description of the algorithm implementation slightly differed from the actual modelling setup.

We updated the statement at lines 114-115 from

“The maximum number of input features to be used in one decision tree (max\_features) was set to 7” to

“The maximum number of input features to be used in one decision tree node (max\_features) was set to 7,”

In addition, the bootstrap bagging method was not applied, and therefore we removed the sentence indicating that. We further tested the model performance with different bootstrapping parameters, and there was no detectable improvement whether the bootstrapping was applied or not.

We have removed from lines 116-118 the statement

“In addition, we used the default bootstrap aggregation method for dividing the training data into separate sample sets in order to provide an individual sample of training data for each tree, which has been shown to increase the smoothness of the fit (Altman and Krzywinski, 2017).”

• *For the global-scale modelling, ECLIPSE V6b emission scenarios are used. How appropriate is this emission inventory for simulating PM over the Indian region or what are the criteria for selecting this inventory for this study? What is the contribution of this inventory to the high under-estimation of PM loading by the model over Delhi as shown in the manuscript?*

A detailed description of previous versions of the ECLIPSE inventories is presented in Klimont et al. (2017) and Stohl et al. (2015).

The ECLIPSE V6b emissions are described in detail in the forthcoming Arctic Monitoring and Assessment Program (AMAP) report. They were especially developed to study emission reductions of short-lived climate forcers (SLCF) on the global and Arctic climate. Unfortunately, the publication of the report has been put on hold due to the current political situation for an indefinite amount of time. However, Belis et al. (2022), von Salzen et al. (2022) and Whaley et al. (2022) provide brief descriptions of the ECLIPSE V6b scenarios.

Furthermore, as we mention in the manuscript text lines 449-451, Whaley et al. (2022) state that in ECLIPSE V6b, the recent declines in Asian SO<sub>2</sub> and BC emissions are considered. This suggests that ECLIPSE V6b is a better choice for modeling South Asian aerosols, since some emission inventories, such as CEDS emissions from CMIP6 simulations, might lack this decline (Wang et al. 2021).

We have updated the manuscript text in lines 149-162 to include additional citations for ECLIPSE scenarios.

*• How the exclusion of the mineral dust component solves the issue related to the PM<sub>2.5</sub> peaking? How authors can make sure that this exclusion won't affect the other simulation results?*

Here we want to clarify the RF-correction procedure. Instead of using RF to directly predict surface PM<sub>2.5</sub> values, we predict a correction term to the PM<sub>2.5</sub> values modelled by ECHAM-HAMMOZ, which has been shown earlier to give better results (e.g., Lipponen et al., 2013). For the computation of the correction term, we use all the ECHAM-HAMMOZ parameters listed in Table 3, which does also include the PM<sub>2.5</sub> due to mineral dust component and mineral dust emissions from ECHAM-HAMMOZ. However, because during the RF training phase the mineral dust component shows very large peaks which do not correspond to the measurements, we exclude mineral dust from the PM<sub>2.5</sub> value to which the correction term is added. Therefore, the correction term includes an inferred amount of mineral dust as the random forest models get information about mineral dust episodes as an input. In this sense, the mineral dust component is not (entirely) excluded from the RF-correction procedure.

The difference between PM<sub>2.5</sub> with and without mineral dust can be seen in Figure S2. Most of the very high peaking values in ECHAM-HAMMOZ data are during summertime, and due to mineral dust. The daily average of the measurement stations, on the other hand, does not show summertime peaking values that would exceed the winter month maxima. Including dust component in the ECHAM-HAMMOZ PM<sub>2.5</sub> might have produced the highest values during the summer months, as our bias correction is additive. We agree with the referee that excluding mineral dust from ECHAM-HAMMOZ PM<sub>2.5</sub> when calculating of the error term between station PM<sub>2.5</sub> and ECHAM-HAMMOZ might affect results to some level as the very short-term peaks might be suppressed. However, our aim was to model the long-term effects of aerosol

mitigation, and that is why we chose to prefer improved seasonal trends over short-term minimum and maximum values.

We have clarified these aspects in the manuscript text by adding after line 223:

“Furthermore, due to dust episodes, the maximum PM<sub>2.5</sub> for ECHAM-HAMMOZ was during the summer months, whereas measurement stations showed maxima for the winter season. However, mineral dust was included in the input training set (see Table 3), i.e., dust can affect the RF bias correction. PM<sub>2.5</sub> values from ECHAM-HAMMOZ which exclude the mineral dust component are hereafter referred to as non-dust PM<sub>2.5</sub>.”

• *Coming to the radiative forcing calculations (section 2.8), how is the definition given to the radiative forcing related to the conventional definitions found in the published literature? If there is any difference please highlight and justify those, else give supporting citations.*

Thanks for the suggestion, we have inserted some additional references in Section 2.8.

For the calculation of  $RF_{ARI}$ , we have described in the manuscript text the small differences between our definition and the conventional definition published in the 5<sup>th</sup> IPCC assessment report. The ERF calculations are according to the sixth IPCC assessment report (see answer below).

• *How the effective radiative forcing is estimated?*

The ERF values were calculated as described in Section 2.8, and follow the definition proposed in, for instance, the sixth IPCC Assessment report (Forster et al., 2021). The ERF is the difference between the top of atmosphere (TOA) net radiative flux of a perturbed (MITIG\_2030, CLE\_2030) and reference (PRES) simulation. All the simulations have fixed sea surface temperatures (SST) and sea ice cover (SIC), and the meteorology is allowed to evolve freely, i.e., no nudging was applied.

We have added a reference to the latest IPCC AR, and updated Section 2.8 to mention fixed SST and SIC, by changing at lines 264-266 from

“Here we computed the  $ERF_{ARI+ACI}$  as the difference in the net radiative flux at the top of the atmosphere (TOA) between a perturbed scenario and the reference simulation.” to

“Here we computed the  $ERF_{ARI+ACI}$  as the difference in the net radiative flux at the top of the atmosphere (TOA) between a perturbed scenario and the reference simulation (Forster et al., 2021), using simulations with fixed SST and SIC.”

*Other comments*



*L1: This opening sentence is misleading. The study demonstrates the potential of the ML technique in downscaling a global-scale model output..*

We have updated the opening sentence from

“We studied the potential of using a global-scale climate model for analyzing simultaneously both city-level air quality and regional and global scale radiative forcing values for anthropogenic aerosols.” to

“We studied the potential of using machine learning to downscale global-scale climate model output towards ground station data. The aim was to analyze simultaneously both city-level air quality and regional and global scale radiative forcing values for anthropogenic aerosols.”

*L6: You mean the model output is better than the measured PM2.5 values?*

Thanks for pointing this out, the sentence was slightly vague.

We changed the sentence at line 6 from

“The downscaling procedure clearly improved the seasonal variation when compared to measured PM2.5 values.” to

“The downscaling procedure clearly improved the seasonal variation of the model data. The seasonal trends were much better captured in the corrected PM2.5 than in original ECHAM-HAMMOZ PM2.5 when compared to the reference PM2.5 from the ground stations”

*L11: This is a highly impactful statement. Better to give caution to the reader by mentioning the associated large uncertainty as seen in Fig. 3(e).*

Thank you for the suggestion.

We have extended the sentence at lines 12-13 from

“This indicates that aerosol mitigation could bring a double benefit in India: better air quality and decreased warming of the climate.” to

“For the two future emission scenarios modelled, the radiative forcing due to aerosol-radiation interactions over India was  $(-0.09 \pm 0.26)$  W m<sup>-2</sup> and  $(-0.53 \pm 0.31)$  W m<sup>-2</sup>, respectively, while the effective radiative forcing values were  $(-2.1 \pm 4.6)$  W m<sup>-2</sup> and  $(0.06 \pm 3.39)$  W m<sup>-2</sup>, respectively. Although accompanied with relatively large uncertainties, the obtained results indicate that aerosol mitigation could bring a double benefit in India: better air quality and decreased warming of the local climate”

*L38-40: As per the sentence, the role of ACI in aerosol indirect effects is undermined, hence please modify the sentence. Also, please explain the ‘local meteorological dynamics’ with references.*

As suggested, we have modified the text at lines 38-40 from

“In addition to these so-called aerosol-radiation interactions (ARI), aerosols can affect Earth’s radiative balance indirectly, for instance, by altering the properties of clouds (aerosol-cloud interaction; ACI) or by changing local meteorological dynamics” to

“These are termed aerosol-radiation interactions (ARI). Furthermore, aerosol particles can act as seeds for cloud droplets and, hence, changes in aerosol composition and concentration can alter cloud properties, affecting Earth’s radiative balance indirectly (aerosol-cloud interactions; ACI). Aerosols can also affect local meteorological dynamics. For instance, absorbing aerosol particles can alter atmospheric stability in the troposphere due to local heating (Johnson et al, 2019).”

*L72-73: Cannot find in any of the given references that ‘emissions from New Delhi’ significantly contribute to the ATAL. Please clarify.*

This statement was slightly misleading, as Fairlie et al. (2020) referred to Indian subcontinent and Northern India as significant emission sources contributing to the ATAL.

Updated the sentence at lines 72-73 from

“emissions originating from New Delhi contribute significantly to the Asian Tropopause Aerosol layer (ATAL)” to

“emissions originating from Indian subcontinent contribute significantly to the Asian Tropopause Aerosol layer (ATAL)”

*L94: HAM ‘threats’ the chemical compounds..?*

Many thanks for pointing out this typing error.

We have updated the sentence at line 94 from

“HAM threats the chemical compounds” to

“HAM treats the chemical compounds”

*L158: Why do OC emissions increase by 2030 in CLE scenario while all others show a reduction?*

The principal reason for the different trends in different aerosol compound emissions are differences in the relative contribution of the total emissions from each emission sector. These relative contributions may also change differently based on the burnt fuels and technologies used. In the CLE scenario for the year 2030, the *global* OC emissions from waste sector increase approximately by 900 kt (+44 %) when compared to 2015 levels. Furthermore, emissions due to agricultural waste burning increase by ~200 kt (+10%) compared to year 2015 emissions, and the emissions from industry and shipping are also projected to increase by a small portion. These changes are almost as large as emission reductions in the domestic (~ -960 kt) and traffic

(~ -230 kt) sectors. The net change in global anthropogenic OC emissions is a small decrease (~ -0.07 %) in the year 2030 compared to the year 2015.

For the *area surrounding New Delhi*, the increasing OC emissions from the waste (~ +90 kt, +62%) and industry (+11kt, +83%) sectors counterbalance some of the emission reductions cuts from the domestic (-38kt) and traffic (-18kt) sectors. Therefore, there is an increase in the net OC emissions in the area surrounding New Delhi. For BC and SO<sub>2</sub>, the reductions in other sectors (domestic and traffic for BC and energy sector for SO<sub>2</sub>) are large enough to outweigh the increasing emissions from the waste sector. That is why the 2030 CLE emissions for BC and SO<sub>2</sub> are less than in 2015.

We have altered the sentence at line 155 from

"while the global OC emissions are projected to stay almost constant (~-0.07 %)" to "The change in net global OC emissions is almost negligible (~ -0.07 %) due to increasing emissions in the waste and agricultural waste burning sectors, which compensate emission reductions projected in the domestic and traffic sectors."

And also modified the sentence at line 157 from

"The OC emissions are an exception as they increase by 7 % in 2030 compared to 2015." to "The OC emissions are projected to increase for the New Delhi surrounding by 7 % in the year 2030 compared to the year 2015. This is due to increasing emissions from the waste and industry sectors, despite emission cuts projected for the domestic and traffic sectors."

#### *L267: What is '2D yearly mean value'?*

By 2D, we meant that the yearly mean values were calculated separately for each grid box.

We changed the sentence at line 267 from

"For both  $RF_{ARI}$  and ERF, we first calculated 2D yearly mean values using  $RE_{ARI}$  and the net TOA radiative flux, respectively." to

"In order to compute  $RF_{ARI}$  and ERF, we first computed 2D yearly mean fields of  $RE_{ARI}$  and the net TOA radiative flux, respectively, which were calculated as yearly mean values for each separate grid box."

#### *L356: If that is the case, what is the significance of feature importance values?*

The feature importance values describe which input features reduced the modeling error the most in the training phase. Therefore, the importance values reveal information about the RF algorithm priorities during the training. However, the feature importances do not necessarily describe the RF model output sensitivities to input features.

To describe this better, we have added a sentence after line 357:

"The feature importance values indicate the contribution of a feature to the total reduction in the error criterion. However, importance values do not reveal the sensitivity of the RF model to

specific input features.”

*L385: This \_\_ somewhat?*

We have modified this sentence to clarify the main point, i.e., that the strong negative  $RF_{ARI}$  values were not expected when all three aerosol species are reduced simultaneously.

Changed the statement at line 385 from

“This somewhat surprising result is due to ...” to

“This result is somewhat surprising as both of the scenarios, CLE\_2030 and MITIG\_2030, included reductions in both absorbing and scattering aerosols. The net negative  $RF_{ARI}$  for India results from ...”

*L395-398: Confusing. The aerosol loading in MITIG\_2030 is supposed to be lower than the CLE\_2030, then how RF in MITIG\_2030 is more negative than CLE\_2030 at the Himalayan foothills. Bright background due to strong haze is expected to be more in CLE\_2030.*

The  $RF_{ARI}$  describes the change in the aerosol radiative effect between perturbed and reference scenarios. MITIG\_2030  $RF_{ARI}$  is more negative than CLE\_2030  $RF_{ARI}$  since there is less absorbing aerosol (BC) in MITIG\_2030 than in CLE\_2030. It is true that the pollution haze is expected to be stronger in CLE\_2030 than in MITIG\_2030. However, as we see from Figures S3a and b, the atmospheric absorption is significantly less in MITIG\_2030 than in CLE\_2030. This indicates that the absorption due to BC is the dominating effect in this area, and therefore the  $RF_{ARI}$  is more negative in MITIG\_2030 than in CLE\_2030.

We have added the following sentences after line 398:

“The  $RF_{ARI}$  is more negative for the MITIG\_2030 simulation, even though the haze layer due to anthropogenic aerosols is stronger in CLE\_2030. This is because BC is reduced more heavily in MITIG\_2030, which leads to stronger contributions to  $RF_{ARI}$  than the changes in the background aerosol layer.”

*L434-436: cannot understand. Do you mean that the CDNC burden was more in CLE\_2030 scenario?*

Exactly.

To make this point clear, we rewrote the sentence at lines 434-436, from

“When comparing CLE\_2030 and PRES, the changes in aerosol concentrations translate surprisingly into a very small increase in the CDNC burden over most of the India (not shown), which would indicate that there is more cloud droplets to scatter SW radiation.” to

“When comparing CLE\_2030 and PRES, the changes in aerosol concentrations translate surprisingly into a very small increase in the cloud droplet number concentration (CDNC)

burden over most of India (not shown), i.e. the mean CDNC burden values are higher in CLE\_2030 than in PRES for large areas over India. This would indicate that there are more cloud droplets to scatter SW radiation.”

*L435: Expand CDNC in the manuscript.*

Thanks for pointing out this deficiency.

We have clarified the abbreviation by modifying line 435 from

“in the CDNC burden” to “in the cloud droplet number concentration (CDNC) burden”

In addition, we described the term “burden” in Section 2.6 by adding in the Table 3 caption following:

“Burden indicates vertically integrated concentration values.”

*Conclusions: This section can be renamed as summary and conclusions by adding the significant findings of the study as bullet points.*

Many thanks for the suggestion, we have renamed Section 4 and have added a very short summarizing statement to the end of this section. We, however, felt that a bullet-pointed list would not fit the style of the journal and instead opted for giving the summary in plain text style.

We have added after line 532 the following:

“To summarize, we obtained greatly improved seasonal trends for surface PM<sub>2.5</sub> when applying machine learning downscaling to ECHAM-HAMMOZ data for New Delhi. Stringent, global aerosol mitigation resulted in improved air quality in New Delhi and negative radiative forcing for most of India. Organic carbon emissions had a stronger influence on local air quality, whereas black carbon emissions contributed more to the radiative effects.”

## References:

Allen, R. J., Turnock, S., Nabat, P., Neubauer, D., Lohmann, U., Olivié, D., Oshima, N., Michou, M., Wu, T., Zhang, J., Takemura, T., Schulz, M., Tsigaridis, K., Bauer, S. E., Emmons, L., Horowitz, L., Naik, V., van Noije, T., Bergman, T., Lamarque, J.-F., Zanis, P., Tegen, I., Westervelt, D. M., Le Sager, P., Good, P., Shim, S., O'Connor, F., Akritidis, D., Georgoulias, A. K., Deushi, M., Sentman, L. T., John, J. G., Fujimori, S., and Collins, W. J.: Climate and air quality impacts due to mitigation of non-methane near-term climate forcers, *Atmos. Chem. Phys.*, 20, 9641–9663, <https://doi.org/10.5194/acp-20-9641-2020>, 2020.

Bawase, M., Sathe, Y., Khandaskar, H., and Thipse, S.: Chemical composition and source

attribution of PM<sub>2.5</sub> and PM<sub>10</sub> in Delhi-National Capital Region (NCR) of India: results from an extensive seasonal campaign, *Journal of Atmospheric Chemistry*, 78, 35–58,610 <https://doi.org/10.1007/s10874-020-09412-7>, 2021.

Belis, C. A., Van Dingenen, R., Klimont, Z., and Dentener, F.: Scenario analysis of PM<sub>2.5</sub> and ozone impacts on health, crops and climate with TM5-FASST: A case study in the Western Balkans, *Journal of Environmental Management*, Volume 319, 115738, ISSN 0301-4797, <https://doi.org/10.1016/j.jenvman.2022.115738>, 2022.

Das, S., Giorgi, F., Giuliani, G., Dey, S., and Coppola, E.: Near-Future Anthropogenic Aerosol Emission Scenarios and Their Direct Radiative Effects on the Present-Day Characteristics of the Indian Summer Monsoon, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031414, <https://doi.org/https://doi.org/10.1029/2019JD031414>, e2019JD031414 2019JD031414, 2020.

Fairlie, T. D., Liu, H., Vernier, J.-P., Campuzano-Jost, P., Jimenez, J. L., Jo, D. S., Zhang, B., Natarajan, M., Avery, M. A., and Huey, G.: Estimates of Regional Source Contributions to the Asian Tropopause Aerosol Layer Using a Chemical Transport Model, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031506, <https://doi.org/https://doi.org/10.1029/2019JD031506>, 2020.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D.J., Mauritsen, T., Palmer, M.D., Watanabe, M., Wild, M. and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 923–1054, doi:10.1017/9781009157896.009, 2021.

Hastie, Trevor., et al. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer, pp. 590. 2009.

Johnson, B. T., Haywood, J. M., and Hawcroft, M. K.: Are Changes in Atmospheric Circulation Important for Black Carbon Aerosol Impacts on Clouds, Precipitation, and Radiation?, *Journal of Geophysical Research: Atmospheres*, 124, 7930–7950, <https://doi.org/10.1029/2019JD030568>, 2019.

Klimont, Z., Kupiainen, K., Heyes, C., Purohit, P., Cofala, J., Rafaj, P., Borken-Kleefeld, J., and Schöpp, W.: Global anthropogenic emissions of particulate matter including black carbon, *Atmos. Chem. Phys.*, 17, 8681–8723, <https://doi.org/10.5194/acp-17-8681-2017>, 2017.

Lalchandani, V., Srivastava, D., Dave, J., Mishra, S., Tripathi, N., Shukla, A. K., Sahu, R., Thamban, N. M., Gaddamidi, S., Dixit, K., Ganguly, D., Tiwari, S., Srivastava, A. K., Sahu, L., Rastogi, N., Gargava, P., and Tripathi, S. N.: Effect of Biomass Burning on PM<sub>2.5</sub> Composition and Secondary Aerosol Formation During Post-Monsoon and Winter Haze Episodes in Delhi, *Journal of Geophysical Research: Atmospheres*, 127, e2021JD035232, <https://doi.org/https://doi.org/10.1029/2021JD035232>, e2021JD035232 2021JD035232, 2022

Lipponen, A., Kolehmainen, V., Romakkaniemi, S., and Kokkola, H.: Correction of approximation errors with Random Forests applied to modelling of cloud droplet formation, *Geoscientific Model Development*, 6, 2087–2098, <https://doi.org/10.5194/gmd-6-2087-2013>, 2013.

Permadi, D. A., Kim Oanh, N. T., and Vautard, R.: Assessment of emission scenarios for 2030 and impacts of black carbon emission reduction measures on air quality and radiative forcing in Southeast Asia, *Atmos. Chem. Phys.*, 18, 3321–3334, <https://doi.org/10.5194/acp-18-3321-2018>, 2018.

Prakash, J., Vats, P., Sharma, A. K., Ganguly, D., and Habib, G.: New Emission Inventory of Carbonaceous Aerosols from the On-road Transport Sector in India and its Implications for Direct Radiative Forcing over the Region, *Aerosol and Air Quality Research*, 20, 741–761, <https://doi.org/10.4209/aaqr.2019.08.0393>, 2020.

Stohl, A., Aamaas, B., Amann, M., Baker, L. H., Bellouin, N., Berntsen, T. K., Boucher, O., Cherian, R., Collins, W., Daskalakis, N., Dusinska, M., Eckhardt, S., Fuglestedt, J. S., Harju, M., Heyes, C., Hodnebrog, Ø., Hao, J., Im, U., Kanakidou, M., Klimont, Z., Kupiainen, K., Law, K. S., Lund, M. T., Maas, R., MacIntosh, C. R., Myhre, G., Myriokefalitakis, S., Olivie, D., Quaas, J., Quennehen, B., Raut, J.-C., Rumbold, S. T., Samset, B. H., Schulz, M., Seland, Ø., Shine, K. P., Skeie, R. B., Wang, S., Yttri, K. E., and Zhu, T.: Evaluating the climate and air quality impacts of short-lived pollutants, *Atmos. Chem. Phys.*, 15, 10529–10566, <https://doi.org/10.5194/acp-15-10529-2015>, 2015.

von Salzen, K., Whaley, C.H., Anenberg, S.C. et al: Clean air policies are key for successfully mitigating Arctic warming. *Commun Earth Environ* 3, 222. <https://doi.org/10.1038/s43247-022-00555-x>, 2022

Wang, Z., Lin, L., Xu, Y., Che, H., Zhang, X., Zhang, H., Dong, W., Wang, C., Gui, K., and Xie, B.: Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models, *npj Climate and Atmospheric Science*, 4, 2, <https://doi.org/10.1038/s41612-020-00159-2>, 2021.

Whaley, C. H., Mahmood, R., von Salzen, K., Winter, B., Eckhardt, S., Arnold, S., Beagley, S., Becagli, S., Chien, R.-Y., Christensen, J., Damani, S. M., Dong, X., Eleftheriadis, K., Evangeliou, N., Faluvegi, G., Flanner, M., Fu, J. S., Gauss, M., Giardi, F., Gong, W., Hjorth, J. L., Huang, L., Im, U., Kanaya, Y., Krishnan, S., Klimont, Z., Kühn, T., Langner, J., Law, K. S., Marelle, L., Massling, A., Olivie, D., Onishi, T., Oshima, N., Peng, Y., Plummer, D. A., Popovicheva, O., Pozzoli, L., Raut, J.-

C., Sand, M., Saunders, L. N., Schmale, J., Sharma, S., Skeie, R. B., Skov, H., Taketani, F., Thomas, M. A., Traversi, R., Tsigaridis, K., Tsyro, S., Turnock, S., Vitale, V., Walker, K. A., Wang, M., Watson-Parris, D., and Weiss-Gibbons, T.: Model evaluation of short-lived climate forcers for the Arctic Monitoring and Assessment Programme: a multi-species, multi-model study, *Atmos. Chem. Phys.*, 22, 5775–5828, <https://doi.org/10.5194/acp-22-5775-2022>, 2022.