

Response to comments of referee #1

General comments

Chen et al. describe the use of a numerical framework for emulation and sensitivity analysis of a regional air quality model in the development of air quality mitigation strategies for the megacity Delhi. They find that a combination of reduction in traffic emissions within the city, combined with simultaneous reductions in all emission sources in the surrounding region would lead to a reduction of PM_{2.5}, while avoiding an increase in ozone. The reduction of traffic emissions from Delhi alone would increase peak ozone in Delhi due to the high emissions of NO_x from traffic, with the resultant reduction in ozone due to changes in the O₃-NO_x titration effect.

These results are certainly plausible, and consistent with previous work. The potential for ozone to increase when high local NO_x emissions are decreased has been well understood for decades, as has the transboundary nature of ozone and the corresponding need to control precursor emissions over large spatial scales in order to achieve reductions in ozone. The authors themselves also cite previous work showing that a large fraction of the PM_{2.5} in Delhi originates outside of the city. I would generally regard the results presented by Chen et al. as unremarkable, and not of sufficient scientific novelty to warrant publication in Atmospheric Chemistry and Physics. The most novel aspect of the study as I see it is the use of a statistical model emulator, combined with a technique called "global sensitivity analysis" to rapidly discover and evaluate effective emission mitigation options with a minimal amount of computational expense. The paper would potentially have merit if it had more of a technical focus on the methodology. Unfortunately, the methods are not described or evaluated well enough in the present version of the manuscript for me to be able to recommend publication. In order to be recommendable for publication, the manuscript needs major revisions focusing on better description and evaluation of the methods for model emulation and sensitivity analysis. I give suggestions for improving the manuscript in my specific comments, below.

Many thanks to the reviewer for the comments and suggestions.

The chemical relationship between O₃ and NO_x has been well understood for decades, however, the reduction of O₃ pollution is still a troublesome issue for mitigation strategy. For example, recently in China, increase of O₃ is attracting increasing concern from the Chinese government

and public, despite considerable achievements in controlling PM_{2.5} pollution, as described in the introduction. Only a handful of studies foresee the potential of O₃ increase in Delhi under the current mitigation strategy focusing on PM_{2.5} and provide solutions for it. This work provides a quantified map for mitigating PM_{2.5} pollution and tackling O₃ increase for Delhi, to avoid the O₃ side-effect that China is now facing. Our results could greatly benefit air pollution mitigation with respect to both PM_{2.5} and O₃ in Delhi. In addition, as the reviewer mentioned, this work demonstrates a combined approach with WRF-Chem and statistical methods to rapidly discover and evaluate effective emission mitigation options.

We agree with the reviewer that more details regarding the methodology can improve this work. We have therefore improved the manuscript accordingly. Please find point-by-point responses below.

Specific comments:

The introduction is concise and well written, but since the novelty of the paper is in its methodological advances, it needs an expanded discussion of model emulation and global sensitivity analysis.

Thanks for the comments and suggestions. We have improved the manuscript by adding a clearer introduction to these approaches.

1) Line 134: WRF-Chem is an online model, which is capable of calculating its own meteorology. Please describe how the model is "driven" by the ECMWF meteorological data. Are they used as boundary conditions? Is some kind of nudging or data assimilation used?

We have added more description about how the model is driven by ECMWF meteorological data, as shown below.

“The ECMWF reanalysis dataset (ERA-Interim) assimilates observations with a number of nearly 10⁷ per day (Dee et al., 2011), and is used for grid nudging, initial and boundary conditions for WRF-Chem with horizontal and temporal resolutions of 0.75° × 0.75° and 6 hours, respectively.”

2) Line 181: The reference given here (Iooss and Lemaitre, 2015) appears to use "global sensitivity analysis" as an umbrella term to describe a range of techniques. The authors should be more specific about what kind of global sensitivity analysis they describe in this manuscript.

The reviewer is right that there are many different ways to perform global sensitivity analysis, such as brute force, Sobol method, Fourier Amplitude Sensitivity Test (FAST), random-based-design FAST and extended FAST (eFAST). The sections 1.1 and 2.2 of a recent open-accessible work (Ryan et al., 2018) introduce and summarize well the application and theories/equations of different methods for global sensitivity analysis. In this study, we use the eFAST method to perform global sensitivity analysis. The eFAST method, first developed by Saltelli et al. (1999), is more efficient than the other method mentioned above and has been widely used in diverse areas of science (Carslaw et al., 2013; Koehler and Owen, 1996; Queipo et al., 2005; Vanuytrecht and Willems, 2014; vu et al., 2015).

We have rewritten the section 2.3 of our manuscript, added the description and equation for calculating global sensitivity indices, and provided more details about eFAST method for perform the sensitivity indices calculation. Detailed modification of section 2.3 will be shown in the point-4, combining with the responses to the points 2-4.

3) Line 185: The paper by Saltelli et al. (1999) is behind a paywall. Simply giving a reference to this study is not enough to describe the method they employ. The authors must also give a summary of how this works and how it is specifically employed in their study.

In this study, we use the eFAST method to perform global sensitivity analysis (GSA). The eFAST method, first developed by Saltelli et al. (1999), is more efficient than the other methods mentioned above and has been widely used in diverse areas of science, as stated above.

A detailed introduction to the theory and equations of the eFAST method is given in an open-accessible methodological study (Ryan et al., 2018). We have extensively modified section 2.3 of our manuscript to provide a simple introduction to the approaches adopted, and refer to Ryan et al. (2018) and other previous studies for further details.

Description of global sensitivity analysis in the section 2.2 of Ryan et al. (2018):

“A common way of conducting global sensitivity analysis for each point in the output space of the simulator – where the output consists of, for example, a spatial map or a time series – is to compute the first-order sensitivity indices (SIs) using variance-based decomposition; this apportions the variance in simulator output (a scalar) to different sources of variation in the different model inputs. Assuming the input variables are independent of one another – which they are for this study – the first-order SI, corresponding to the i^{th} input variable ($i = 1, 2, \dots, p$) and the j^{th} point in the output space, is given by the equation (R1).

$$SI_{i,j} = \frac{\text{Var}[E(Y_j | X_i)]}{\text{Var}(Y_j)} \times 100 \quad (R1)$$

where X_i is the i^{th} column of the $n \times p$ matrix X (i.e. a matrix with n rows and p columns) which stores the n samples of p -dimensional inputs, and Y_j is the j^{th} column of the $n \times m$ matrix which stores the corresponding n sets of m -dimensional outputs. We multiply by 100 so that the SI is given as a percentage. The notation given by $\text{Var}(\bullet)$ and $E(\bullet)$ denotes the mathematical operations that compute the variance and expectation. The simplest way of computing $SI_{i,j}$ is by brute force, but this is also the most computationally intensive.”

Source from: the section 2.2 of Ryan et al., (2018)

Description of eFAST method in the section 2.2.2 of Ryan et al. (2018):

“The eFAST method is an alternative and more efficient way of estimating the terms in Eq. (R1). A multi-dimensional Fourier transformation of the simulator f , allows a variance-based decomposition that samples the input space along a curve defined by the equation (R2).

$$x_i(s) = G_i(\sin(\omega_i s)) \quad (R2)$$

where $x = (x_1, \dots, x_p)$ refers to a general point in the input space that has been sampled, $s \in R$ is a variable over the range $(-\infty, \infty)$, G_i is the i^{th} transformation function, and ω_i is the i^{th} user-specified frequency corresponding to each input. Varying s allows a multi-dimensional exploration of the input space due to the $x_i(s)$ being simultaneously varied. Depending on the simulator, we typically require $n=1000-10,000$ samples from the input space. After applying the simulator f , the resulting scalar output – denoted generally by y – produces different periodic functions based on different ω_i . If the output y is sensitive to changes in the i^{th} input factor, the periodic function of y corresponding to frequency ω_i will have a high amplitude.”

Source from: the section 2.2.2 of Ryan et al., (2018)

Please refer to the section 2.2.2 of Ryan et al. (2018) for more specific details about the theory and equations of eFAST method. In order not to replicate many mathematic equations in Ryan et al. (2018), in this study, we have rephrased the section 2.3 to summarize the advantage of global sensitivity analysis compared with the widely used ‘one-at-a-time’ sensitivity analysis and to simply describe how we perform GSA with non-mathematic language. The detailed modifications of section 2.3 are shown in the point-4, combining with the responses to the points 2-4.

4) Line 188: Similarly, "Gaussian process emulation" is not sufficiently well described in the manuscript. A summary of how this technique works and how it is applied must be included.

To perform global sensitivity analysis and generate response surfaces, which describe relationships between the inputs and outputs of models, usually requires thousands of model runs. This is not feasible for a computationally expensive model like WRF-Chem. Therefore, in our study a Gaussian process (GP) emulator, trained by a few model runs, is used as a surrogate of WRF-Chem model. Mathematically, an emulator is a statistical model that mimics

the input-output relationship of a simulator, i.e., the expensive WRF-Chem model in this study. The most common form of an emulator is a GP emulator since it has attractive mathematical properties that allow an analytical derivation of the mean and variance of the emulated output (Ryan et al., 2018). As summarized in Ryan et al. (2018):

“More formally, a GP is an extension of the multivariate Gaussian distribution to infinitely many variables (Rasmussen and Williams, 2006). The multivariate Gaussian distribution is specified by a mean vector and covariance matrix. A GP has a mean function which is typically given by $m(x) = E(f(x))$ and covariance function given by $c(x, x') = cov(f(x), f(x'))$ ”.

Source from: Ryan et al., (2018)

Where x and x' are two different input-matrix.

Detailed theory of GP emulation is introduced in an open-accessible methodological study (Ryan et al., 2018), and equations and the R code for generating GP emulator are also given in Ryan et al. (2018). In our manuscript, we have substantially revised the section 2.3 to summarize the GP emulator with non-mathematical language, and to describe its advantages and how we use it. The revised section 2.3 is shown below, please also find detailed changes in the revise-tracked file.

Revised section 2.3 of this study:

“2.3 Global Sensitivity Analysis of Urban Air Pollution

We perform global sensitivity analysis (GSA) (Iooss and Lemaître, 2015) to quantify the sensitivity of modelled outputs (PM_{2.5} and O₃ for this study) to changes in the model inputs, which for this study are emissions from the different emission sectors. One-at-a-time sensitivity analysis is a common way of calculating model sensitivity and involves varying a single model

input while the other inputs are fixed at nominal values, e.g., Wild (2007). While one-at-a-time approach is relatively easy to implement, it assumes that the model response to different inputs is independent and this can lead to biased results (Saltelli et al., 1999;Pisoni et al., 2018;Carslaw et al., 2013). GSA overcomes the problems of the one-at-a-time approach by averaging over the other inputs rather than fixing them at specific values. This allows calculation of first-order sensitivity indices (SIs) for each variable, corresponding to the i^{th} input variable and the j^{th} output point, is given by the Eq. 1 (Ryan et al., 2018).

$$SI_{i,j} = \frac{\text{Var}[E(y_j | x_i)]}{\text{Var}(y_j)} \times 100\% \quad (1)$$

where x_i is the i^{th} element of the input; and y_j is the j^{th} element of the output. The ‘ $E(\bullet)$ ’ and ‘ $\text{Var}(\bullet)$ ’ denote the mathematical function that calculate the expectation and variance, respectively. The simplest way of computing $SI_{i,j}$ is by brute force, but this is also the most computationally intensive (Ryan et al., 2018).

The extended Fourier Amplitude Sensitivity Test (eFAST), first developed by Saltelli et al. (1999), is a commonly used approach to perform GSA and calculate SIs and is adopted in this study because of its high efficiency. A basic overview and detailed equations of the eFAST method are given in the section 2.2.2 of Ryan et al. (2018). A challenge to using eFAST is that it typically requires thousands of model runs. To overcome this, we employ a computationally cheaper surrogate model in place of our expensive simulation model WRF-Chem. A surrogate model is a simple model (usually statistical) which can map the inputs to the outputs of the simulation model with sufficiently good accuracy given the same inputs. In this study, we choose a type of surrogate model called a Gaussian process emulator, which works like a function for multi-dimensional interpolation and has been used extensively in many areas of applied science (Carslaw et al., 2013;Koehler and Owen, 1996;Queipo et al., 2005;Vanuytrecht

and Willems, 2014;vu et al., 2015;Degroote et al., 2012) and uncertainty assessment of atmospheric models (Lee et al., 2016;Lee et al., 2012;Lee et al., 2011). Gaussian process emulators typically require a relatively small number of runs of the computational-expensive model to generate; this is in contrast to other surrogate modelling approaches, such as neural networks, which typically require thousands of model runs to train them. For a basic overview of a Gaussian process emulator see O’Hagan (2006), detailed introduction and equations are also given in the section 2.3 of Ryan et al. (2018). Before using the emulator in place of the WRF-Chem model to carry out the thousands of model runs required for GSA, we train the emulator using a relatively small number of WRF-Chem model runs. Following previous studies (Carslaw et al., 2013;Lee et al., 2016), a Maximin Latin hypercube space-filling design is employed to select the designs of training runs for WRF-Chem. Latin hypercube sampling is a statistical method for generating a near-random sample of parameter values from a multidimensional distribution (Shields and Zhang, 2016). Here, we search through 100,000 Latin hypercube random designs to find the optimal one where the parameter space is filled most effectively. This ensures that the sets of inputs chosen cover as large a fraction of the input space as possible. Full details (including R codes) of how to generate the Gaussian process emulator, eFAST method and GSA can be found in Ryan et al. (2018).”

5) Lines 209-210: "10,000 random samples" are performed to check that the emulator "can fully represent the results of WRF-Chem". Does this mean that the authors performed 10,000 runs of WRF-Chem, and compared them with 10,000 runs of the emulator? Or did they do something else? What do each of the points in Fig. 3 actually represent? This is not clear at all. The authors seem to be relying on this analysis to show that the emulation "provides a good representation of the model", but in my opinion this has not been shown at all. Much more detail is needed here.

The "10,000 random samples" refers to selection of 10,000 samples from the spatial grid cells and temporal duration of the model run and rebuilding emulators at these points using all but

one of the WRF-Chem model training runs, and then comparing these against results from the model run that was omitted in ‘leave-one-out’ cross-validation. This provides an independent check of how well Gaussian process emulator can represent the results of WRF-Chem. And, thanks to reviewer’s suggestion in the last comment, the further validation with the base and regional joint reduction cases also demonstrate good agreement between the emulator and WRF-Chem model (see response to the last comment). We have modified the corresponding context in the last part of section 2.3 to make this clearer, as shown below.

“We perform ‘leave-one-out’ cross-validation (O’Hagan and West, 2009; Wang et al., 2011) with 10,000 random samples to check that the Gaussian process emulator can fully represent the results of WRF-Chem.”

Changed to:

“The accuracy of the emulator as a surrogate of WRF-Chem model is evaluated using a ‘leave-one-out’ cross-validation (Bastos and O’Hagan, 2009). This involves training the emulator using 19 out of the 20 sets of inputs/outputs from the WRF-Chem model runs and then evaluating the emulator against the 20th simulation. This process is carried out for each of the 20 sets of inputs/outputs. Given that the output space is multi-dimensional (i.e. modelled O₃ and PM_{2.5} varied spatially and in time), the validation of the emulator by comparing 10,000 (random-samples varied spatially and in time) of emulator output values against the corresponding output values of the WRF-Chem model. The emulator validation plot is shown in Fig. 3. Modelled and emulated O₃ and PM_{2.5} lie very close to the 1:1 line with R² values of more than 95% as shown in Fig. 3, indicating that the emulation provides an accurate representation of the input-output relationship of the WRF-Chem model.”

6) Line 256: NO_x appears to be significantly underestimated by WRF-Chem during the middle of the day, when peak ozone concentrations are also modelled. Given the central role of NO_x as an ozone precursor, it appears that the modelled peak ozone is being well simulated for the wrong reasons. There is likely a compensating error in some other aspect of the model. The

authors should provide some discussion about how these errors in WRF-Chem would propagate into their emulator and affect the global sensitivity analysis.

Thank you for pointing this out. The choice of y-axis range in Fig. S4 was not fully appropriate and makes NO_x look significantly underestimated by WRF-Chem during the middle of the day when peak ozone occurs. We have changed the y-axis range to [0 200], more in line with convention, as shown below. Instead of a “remarkable underestimation”, the NO_x is actually underestimated by only ~30%. The daytime variation of NO_x (Fig. S4) is directly opposite in pattern to that of boundary layer height (red dashed line in Fig. 4a) between 9 am and 6 pm, suggesting that this underestimation in NO_x is closely associated with the variation in boundary layer behavior, a substantial uncertainty in the simulation for which no observational verification is available.

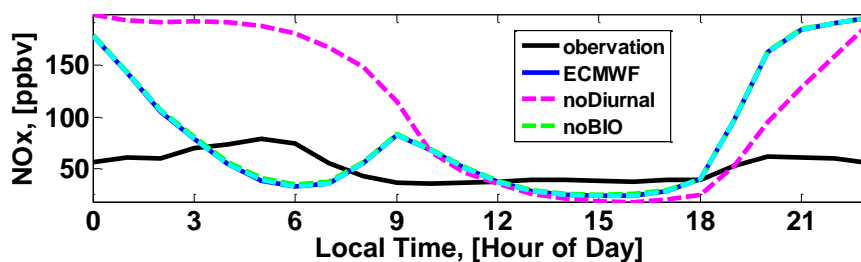
The reviewer is right that the good simulation of peak O₃ when NO_x is underestimated highlights an uncertainty within the model. This uncertainty would propagate into the emulator and affect the global sensitivity analysis, because emulator learns from the WRF-Chem model and reproduces whatever the model outputs. We have added discussion about this influence in the sections 3.4 and 3.5, as shown below.

In the section 3.4 of revised version:

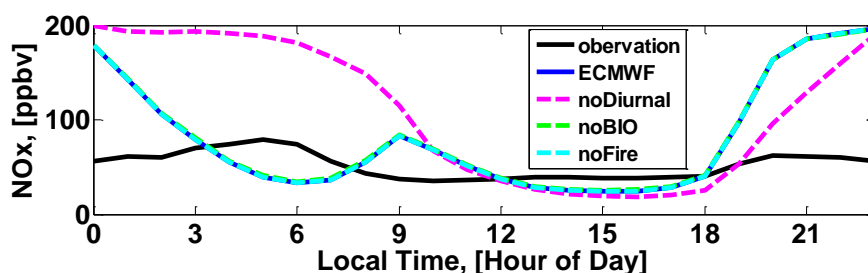
“NO_x, mainly originating from traffic emissions, is underestimated by ~30% during the O₃ peak period (Fig. S4). This uncertainty can propagate into the Gaussian process emulator and could lead to underestimation of the influence of traffic on peak O₃, but is not expected to change the nature of our conclusions about the predominance of regional transport and local traffic emissions.”

In the section 3.5 of revised version:

“We note that our model may underestimate the influence of traffic emissions on O₃ to some extent as described above (section 3.4), suggesting that the ozone increase could be stronger than we predict. To prevent the side-effect of increasing O₃ by controls on traffic emissions...”



Original Figure S4. Diurnal patterns of NO_x concentration from WRF-Chem model and observational results at AIR site. The results are averaged during 02-15 May 2015. Note that 'ECMWF' indicates the model results driven by ECMWF reanalysis data.



Revised Figure S4. Diurnal patterns of NO_x concentration from WRF-Chem model and observational results at AIR site. The results are averaged during 02-15 May 2015. Note that 'ECMWF' indicates the model results driven by ECMWF reanalysis data.

7) Line 267: "We remove these sources". From what? WRF-Chem itself, or the emulator?

We have revised the statement, as shown below.

"We turn off these sources in the WRF-Chem simulation".

8) Lines 311-312 and Fig. 5b: It should be pointed out somewhere in the discussion that the overwhelming dominance of traffic NO_x emissions on ozone in Delhi is actually through anti-correlation. Presenting the sensitivity indices of "TRA" and "NCR" together on the same plot is potentially quite misleading unless the authors make it clear that their respective influences have opposite sign.

The reviewer is right that “overwhelming dominance of traffic NO_x emissions on ozone in Delhi is actually through anti-correlation.”. Although we have already explained this anti-correlation relationship in section 3.4:

“Traffic contributes ~75% of total NO_x emission in Delhi (Fig. 6b), and the shallow PBL during the night traps the NO_x. This removes O₃ through chemical reaction”.

This may not be clear enough. We have added description at the beginning of this section to make this point clearer, as shown below. Thanks for pointing this out.

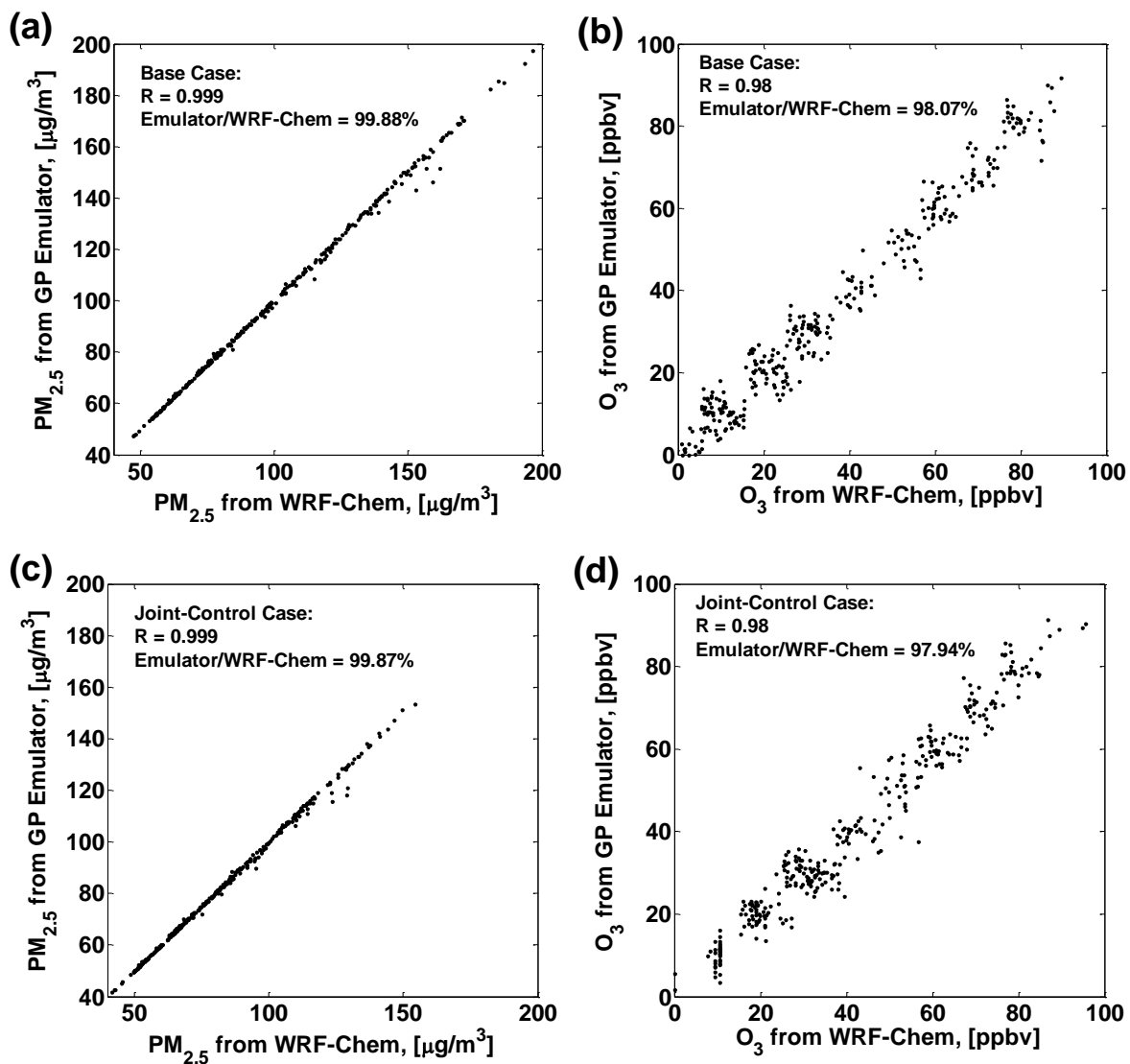
“The variation of O₃ in Delhi City Region is overwhelmingly dominated by local traffic emissions with a sensitivity index higher than 80% at night-time (Fig. 5b), where O₃ and traffic emissions are anti-correlated.”

9) Section 3.5: This is a nice example of the potential power and utility of the methodology. Figure 7d is an especially clear illustration of the emissions control trajectory which is required to prevent an increase in ozone in Delhi despite reducing local NO_x emissions. As mentioned in my general comments, this general approach to emission control (reducing ozone by focusing on regional-scale emissions) is consistent with current understanding of ozone chemistry, so this result by itself is rather unremarkable. What is really interesting here is the ability to rapidly discover an optimal emission mitigation pathway, and quantify its effects. What is missing here though, is a verification that the same combined emission controls for TRA and NCR would result in the same reductions in PM and ozone when employed in the full WRF-Chem model. It would only take one WRF-Chem run to verify this result. In my opinion this extra run is necessary for the authors to be able to show that their approach really is capable of what they claim.

This is a good point. This one extra WRF-Chem simulation makes the verification of emulator more solid and evident. We have performed the extra simulation as suggested. In addition, we also compare the base case (without change of emissions) WRF-Chem results against the emulator results. Noting that the base case is not part of the training cases for emulator, see a list of training cases in Table S2; therefore, this comparison also provides an independent validation for the GP emulator. The results of WRF-Chem and emulator show almost the same reductions in the joint control case. As shown in the figure below, the emulator does reproduce

the WRF-Chem simulation nicely, with an uncertainty (normalized mean error) of less than 5% and R^2 higher than 0.95 for both $PM_{2.5}$ and O_3 . Thanks for the suggestion, we have added this extra validation in the section 3.5, as shown below.

“We test this by performing an additional run with WRF-Chem using emission reductions of 50% and 30% for sectors of local traffic and the surrounding NCR region, respectively. We compare the WRF-Chem results of the additional run and the base case (without change of emissions) against the corresponding results from Gaussian process emulator (Fig. S8). We find that the $PM_{2.5}$ and O_3 results from the model runs lie within 5% of those estimated with the emulator and with R^2 higher than 95%, demonstrating the high quality of the emulation approach adopted here and underlining its deeper value for identifying mitigation approaches”



Newly added Figure S8. Extra validation of Gaussian process emulator results in the mitigation strategy according to Fig. 7. The accuracy of the emulator for reproducing

current conditions of PM_{2.5} (a) and O₃ (b), i.e. base case without changing emissions. The accuracy of the emulator for reproducing regional joint coordination conditions of PM_{2.5} (c) and O₃ (d), i.e. NCR joint control case with local traffic emissions reduced by 50% and regional emissions reduced by 30%. All the results are averaged over Delhi City Region, with hourly resolution during the simulation period.

References:

- Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503, 67, 10.1038/nature12674, 2013.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553-597, 10.1002/qj.828, 2011.
- Degroote, J., Couckuyt, I., Vierendeels, J., Segers, P., and Dhaene, T.: Inverse modelling of an aneurysm's stiffness using surrogate-based optimization and fluid-structure interaction simulations, *Structural and Multidisciplinary Optimization*, 46, 457-469, 10.1007/s00158-011-0751-7, 2012.
- Iooss, B., and Lemaître, P.: A review on global sensitivity analysis methods. In: Meloni, C., Dellino, G. (Eds.), *Uncertainty Management in Simulation Optimization of Complex Systems: Algorithms and Applications*. Springer, 2015.
- Koehler, J. R., and Owen, A. B.: 9 Computer experiments, in: *Handbook of Statistics*, Elsevier, 261-308, 1996.
- Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmos. Chem. Phys.*, 11, 12253-12273, 10.5194/acp-11-12253-2011, 2011.
- Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W.: Mapping the uncertainty in global CCN using emulation, *Atmos. Chem. Phys.*, 12, 9739-9751, 10.5194/acp-12-9739-2012, 2012.
- Lee, L. A., Reddington, C. L., and Carslaw, K. S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, *Proceedings of the National Academy of Sciences*, 113, 5820-5827, 10.1073/pnas.1507050113, 2016.
- O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, *Reliability Engineering & System Safety*, 91, 1290-1300, <https://doi.org/10.1016/j.ress.2005.11.025>, 2006.
- O'Hagan, A., and West, M.: *Handbook of applied Bayesian analysis*, Oxford University Press, New York, 2009.
- Pisoni, E., Albrecht, D., Mara, T. A., Rosati, R., Tarantola, S., and Thunis, P.: Application of uncertainty and sensitivity analysis to the air quality SHERPA modelling tool, *Atmospheric Environment*, 183, 84-93, <https://doi.org/10.1016/j.atmosenv.2018.04.006>, 2018.
- Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P.: Surrogate-based analysis and optimization, *Progress in Aerospace Sciences*, 41, 1-28, <https://doi.org/10.1016/j.paerosci.2005.02.001>, 2005.
- Rasmussen, C. E., and Williams, C. K. I.: *Gaussian Processes for Machine Learning*, the MIT Press, ISBN 026218253X, 2006.

- Ryan, E., Wild, O., Voulgarakis, A., and Lee, L.: Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output, *Geosci. Model Dev.*, 11, 3131-3146, 10.5194/gmd-11-3131-2018, 2018.
- Saltelli, A., Tarantola, S., and Chan, K. P. S.: A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output, *Technometrics*, 41, 39-56, 10.1080/00401706.1999.10485594, 1999.
- Shields, M. D., and Zhang, J.: The generalization of Latin hypercube sampling, *Reliability Engineering & System Safety*, 148, 96-108, <https://doi.org/10.1016/j.ress.2015.12.002>, 2016.
- Vanuytrecht, E., and Willems, P.: Global sensitivity analysis of yield output from the water productivity model, *Environmental Modelling and Software*, 51, 323-332, 10.1016/j.envsoft.2013.10.017, 2014.
- vu, N., Rafiee, R., Zhuang, X., Lahmer, T., and Rabczuk, T.: Uncertainty quantification for multiscale modeling of polymer nanocomposites with correlated parameters, *Composites Part B: Engineering*, 68, 446-464, 10.1016/j.compositesb.2014.09.008, 2015.
- Wang, S., Xing, J., Jang, C., Zhu, Y., Fu, J. S., and Hao, J.: Impact Assessment of Ammonia Emissions on Inorganic Aerosols in East China Using Response Surface Modeling Technique, *Environmental Science & Technology*, 45, 9293-9300, 10.1021/es2022347, 2011.
- Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, *Atmos. Chem. Phys.*, 7, 2643-2660, 10.5194/acp-7-2643-2007, 2007.