

**Reply to comments from Referee #1 on “Quantification of the enhanced effectiveness of NO<sub>x</sub> control from simultaneous reductions of VOC and NH<sub>3</sub> for reducing air pollution in Beijing-Tianjin-Hebei region, China” by Xing et al.**

We thank the reviewer for the detailed and thoughtful review of our manuscript. Incorporation of the reviewer’s suggestion has led to a much improved manuscript. Detailed below is our response to the issues raised by the reviewer. We also detail the specific changes incorporated in the revised manuscript in response to the reviewer’s comments.

[Comment]: *This paper developed a method by fitting multiple CMAQ simulations with a set of polynomial functions to quantify responses of ambient PM<sub>2.5</sub> and O<sub>3</sub> concentrations to changes in precursor emissions. The performance of the model looks sound. However, I suggest the authors to include more scientific findings based on the model developed in this study.*

[Response]: We thank the reviewer for recognition of the implications of the results of the analysis presented. We followed the reviewer’s suggestion and include more discussion on the scientific findings based on the model developed, which is detailed in our response to the reviewer’s comment on Section 3.3.

[Comment]: *The reason why pollutant responses to emissions can be characterized as a series polynomial functions by the previous developed regression-based RSM has not been clarified.*

[Response]: The relationship between pollutant responses to emissions can be quantified by an atmospheric chemical transport model (noted by CTM, e.g., CMAQ) which describes most of the physical and chemical processes in the atmosphere. Studies on multiple CTM simulations under different emission scenarios (“brute force method”) can investigate the full range of pollutant responses to emissions. The principle of regression-based RSM model is to build up the full range of pollutant responses to emissions using an advanced statistic method (i.e., response surface method) from a number of CTM simulations. The accuracy of regression-based RSM in representing the nonlinearity in pollutant response to emissions has been examined thoroughly by different methods including cross validation, out-of-sample validation and isopleth validation in previous studies (Xing et al., 2011; Wang et al., 2011; Zhao et al., 2015; Xing et al., 2017; Zhao et al., 2017).

Since the relationship between pollutant responses to emissions is followed by the basic chemical functions and physical laws parameterized in CTM (i.e., CMAQ in this study), the function used to represent the relationship in regression-based RSM (implicitly) can be parameterized explicitly in a series of basis functions. This study used a linear combination of polynomial bases (i.e., 1, x, x<sup>2</sup>, x<sup>3</sup>...) to characterize the pollutant responses to emissions, and the terms were selected carefully following the procedure described in section 2.2.

To clarify this point, we provided the following discussion in the revised manuscript.

(Page 4 Line 9) “The accuracy of regression-based RSM in representing the nonlinearity in pollutant response to emissions has been examined thoroughly by different methods including cross validation, out-of-sample validation and isopleth validation in previous studies (Xing et al., 2011; Wang et al., 2011; Zhao et al., 2015; Xing et al., 2017; Zhao et al., 2017). The relationship between pollutant responses to emissions followed by the basic chemical functions and physical laws is implicitly represented in the regression-based RSM. In this study, however, we adopted a linear combination of polynomial bases (i.e., 1, x, x<sup>2</sup>, x<sup>3</sup>...) to parameterize explicitly the pollutant responses to emissions.”

[Comment]: *The authors use too many self-defined abbreviations in the text, for instance, PR, VNr, FR, which make the paper not very reader-friendly.*

[Response]: We defined a few indicators in this study, including (1) peak ratio (denoted as PR) representing VOC-limited or NO<sub>x</sub>-limited condition, (2) suggested reduction ratio of VOC to NO<sub>x</sub> (denoted as VNr) to avoid increasing O<sub>3</sub> under VOC-limited condition, (3) flex ratio (denoted as FR) representing NH<sub>3</sub>-poor or NH<sub>3</sub>-rich condition. Because of the advantage of the pf-RSM method which is able to provide the full range of the pollutant responses to emissions, it not only can qualitatively identify the current status to certain chemical scheme, but also can quantitatively estimate the exact transition point on which the chemical scheme will be transitioned to the other. That is the reason why we defined and calculated the PR and FR. Actually these two indicators were first developed and already used in our previous publications (Xing et al., 2011; Wang et al., 2011).

However, we agree that too many self-defined abbreviations will reduce the readability of this manuscript. To address this concern, we deleted the abbreviation of VNr and reduced the usage of the PR and FR in the revised manuscript.

[Comment]: *Page 8, line 1-5, please explain the reason why the performance of the pf-RSM with less than 40 training samples exhibited a noticeable discrepancy compared with that of the regression-based RSM, but not for those with over 40 training samples..*

[Response]: The discrepancy shown in the comparison between pf-RSM with less than 40 training samples with that of regression-based RSM is due to the underfitting issue. It is mainly because the number of training samples is not large enough to capture the nonlinearity in the model system. When the number of training samples increases to over 40, the discrepancy is reduced.

To clarify this point, we added the following discussion in the revised manuscript:

(Page 8 Line 28) "Such discrepancy is caused by the underfitting issue implying the number of training samples is not large enough to capture the nonlinearity in the model system. The issue can be addressed by added more training samples to fit the model. The 40 training samples presented good agreement with the predictions of the regression-based RSM. Improving sampling method is also important for reducing the biases. We can see that additional marginal processing also improved the performance of the pf-RSM."

[Comment]: *The uncertainty of the fitting results is missing.*

[Response]: The uncertainty of pf-RSM is evaluated by the comparison the pf-RSM prediction against with true CTM simulation (i.e., out-of-sample validation). Five statistical indices representing the performance were calculated including the mean normalized error (MeanNE), maximal normalized error (MaxNE), mean fractional error (MeanFE), maximal fractional error (MaxFE) and correlation coefficient (R). From the comparison with the results of 115 CMAQ simulations, we found that the pf-RSM with 40 training samples can meet the criteria of MeanNE within 2% and MaxNE within 10% (which is comparable to the performance of previous regression-based RSM).

To clarify this point, we added some discussion in the revised manuscript as follows:

(Page 8 Line 13) "To meet the criteria of MeanNE within 2% and MaxNE within 10% (i.e., uncertainty of pf-RSM) which is comparable to the performance of previous regression-based RSM, use of 40 training samples with marginal processing (to improve boundary conditions) is recommended."

**[Comment]:** *Section 3.3. This section needs substantial improvement considering this is the only section discussing the application of the method. What are the new findings using the modified model, but not indicated by previous one? What is the advantage of the model compared to the existed ones? Otherwise, this paper seems to be more like a technical document, but lack scientific findings.*

**[Response]:** In this manuscript, we proposed a new method (i.e., pf-RSM) to quantify the pollution response to emissions. Compared to existed methods, the newly developed pf-RSM has two advantages: 1) explicitly represent the response, make it easy to investigate the nonlinearity (e.g., peak value, derivative) of the predicted system, as the indicator (PR, FR) we defined in this study; 2) substantially reduce the computational burden by more than 60%, enable the usage on studies with high spatial and temporal resolution.

In this studies, we adopted the pf-RSM model to investigate the enhanced effectiveness of NO<sub>x</sub> control from simultaneous reductions of VOC and NH<sub>3</sub> for reducing O<sub>3</sub> and PM<sub>2.5</sub>. With the pf-RSM, the enhanced effectiveness was quantified. Strong VOC-limited condition in urban areas in BTH has already been recognized in previous studies, due to the abundance of NO<sub>x</sub> emissions. However, questions that how many current NO<sub>x</sub> emissions are overabundant and how many VOC emissions are suggested to simultaneously reduce with NO<sub>x</sub> were not well addressed. With the newly developed pf-RSM in this study, we can provide a quantitate answer. Our results suggest that the NO<sub>x</sub> emission reduction rate need be greater than 20%-60% (depends on the location in BTH) to pass the transition from VOC-limited to NO<sub>x</sub>-limited, and a simultaneous VOC control (the ratio of VOC reduction to NO<sub>x</sub> reduction is about 0.5-1.2) can avoid increasing O<sub>3</sub> during the transition. Similarly, the benefit of NH<sub>3</sub> control for PM<sub>2.5</sub> reduction is well documented. In this study, we quantified the enhanced benefits in PM<sub>2.5</sub> reductions from simultaneous reduction of NH<sub>3</sub> to be 0.04-0.15 μg m<sup>-3</sup> PM<sub>2.5</sub> per 1% reduction of NH<sub>3</sub> along with NO<sub>x</sub>, with greater benefits in July when the NH<sub>3</sub>-rich condition is not as strong as in January. Besides, we found the response varies significantly over space and time. All the results are derived from the pf-RSM model. Since the pf-RSM model is more efficient compared to previous method, the potential usage of pf-RSM includes cost-benefit optimization and integrated assessment. We are developing an air pollution control cost-benefit and attainment assessment system (ABaCAS, Xing et al., 2017), and pf-RSM will be one of the core module in the whole system. The comparison between our results and other studies was also added into this section.

Following the reviewer's suggestion, we revised the section 3.3 and emphasize the scientific findings as follows:

(Page 9 Line 34) "The nonlinearity in the pollution response to emissions leads to an either enhanced or reduced effectiveness of emission controls. In previous studies, the concept of NH<sub>3</sub>-limited/-poor and NO<sub>x</sub>-/VOC-limited conditions was used widely to demonstrate the influence of NH<sub>3</sub> and VOC on effectiveness of NO<sub>x</sub> controls for reducing PM<sub>2.5</sub> and O<sub>3</sub>, respectively. However, some key questions were not well addressed, such as how much percentage of NO<sub>x</sub> or NH<sub>3</sub> is overabundant and how much percentage of VOC need reduced simultaneously to avoid increased O<sub>3</sub>. In this study, the newly developed pf-RSM explicitly represents the response and the enhanced effectiveness can be easily quantified. As the indicators defined in Section 2.3 can be used to quantify the nonlinear effectiveness of emission control for reducing PM<sub>2.5</sub> and O<sub>3</sub>."

(Page 10 Line 11) "The result is consistent with our previous study (Wang et al., 2011) which reported that NH<sub>3</sub> is sufficiently abundant to neutralize extra nitric acid produced by an additional 25% of NO<sub>x</sub> emissions in north China Plain based on a traditional regression-based RSM study."

(Page 10 Line 17) "That is consistent with the findings of previous studies (Xing et al., 2011) which used a traditional regression-based RSM and found that the PR changes from 0.8 to 1.2 as the distance from the city center increases."

(Page 10 Line 20) "Our results are consistent with the observational studies that use indicator to identify the O<sub>3</sub> chemistry. For example, Liu et al (2016) studied on the ratios of HCHO over NO<sub>2</sub> from the satellite retrieves and found that local ozone production in urban Beijing is VOC-limited when there are no substantial changes in NO<sub>x</sub>

emission in 2015. Chou et al. (2009) found that Beijing urban area was “VOC-limited” region based on the observation of NO, NO<sub>x</sub> and NO<sub>y</sub> at the Peking University site during August 15 to September 11 in 2006. Jin and Holloway (2015) calculated the ratio of HCHO to NO<sub>2</sub> from the OMI instrument aboard the Aura satellite and found the O<sub>3</sub> production is more likely to be VOC-limited over urban areas and NO<sub>x</sub>-limited over rural and remote areas in China from 2005 to 2013.”

(Page 11 Line 11) “After the application of a prior knowledge of the pollutant responsiveness to emissions in the RSM system, the cases required for single regional pf-RSM development were substantially decreased to 40 samples, compared with the previous requirement of over 100 samples, imply that the fitting-based RSM (i.e., pf-RSM) is three time faster than previous regression-based RSM (i.e., the number of CTM simulations needed in pf-RSM is 60% less than that required by previous regression-based RSM). The pf-RSM system in this study operates rapidly, and thus can quickly generate responses with high spatial and temporal resolutions, thereby further facilitating cost-benefit optimization and enabling further assessment studies to be conducted (e.g., air pollution control cost-benefit and attainment assessment ABaCAS system described by Xing et al., 2017).”

## Reference

Xing, J., Wang, S., Jang, C., Zhu, Y., Zhao, B., Ding, D., Wang, J., Zhao, L., Xie, H., Hao, J.: ABaCAS: an overview of the air pollution control cost-benefit and attainment assessment system and its application in China. *The Magazine for Environmental Managers - Air & Waste Management Association*, April, 2017.

[Comment]: *The conclusion in the abstract, like “Thus, simultaneously reducing NH<sub>3</sub> and VOC emission along with NO<sub>x</sub> reduction is recommended to assure the control effectiveness of PM<sub>2.5</sub> and O<sub>3</sub>”, is too general.*

[Response]: To emphasize the scientific findings in this study, we revised this sentence as follows:

(Page 1 Line 34) “Thus, the newly developed pf-RSM model has successfully quantified the enhanced effectiveness of NO<sub>x</sub> control, and simultaneous reduction of VOC and NH<sub>3</sub> with NO<sub>x</sub> can assure the control effectiveness of PM<sub>2.5</sub> and O<sub>3</sub>.”

[Comment]: *Page 2, line 5, the grammar of “significantly influences on” is not proper.*

[Response]: We fixed the typo, and revised it as “significantly influences” (Page 2 Line 5)

[Comment]: *Page 3, line 31- 35, the sentence is too long to read. Please consider rephrasing it.*

[Response]: We have reduced the length of the sentence in the revised manuscript, as follows

(Page 3 Line 35) “In general, tropospheric O<sub>3</sub> and PM<sub>2.5</sub> concentrations are contributed by its sources and sinks through a series of atmospheric processes, such as horizontal or vertical advection and diffusion, gas phase chemistry, and deposition. The nonlinear behavior in each of these processes contributes to the nonlinearity in the responses of concentrations to precursor emissions. Similar responsive functions can be expected across regions and time. For example, a universal ozone isopleth diagrams developed using the empirical kinetic modeling approach of the U.S. Environmental Protection Agency (Gipson et al., 1981) represents the general O<sub>3</sub> responsiveness to NO, and VOC concentrations.”

[Comment]: Page 8, line 9-10. The description about how to design scenarios is missing. For instance, why the moderate control is defined as ENO<sub>x</sub>, ESO<sub>2</sub>, ENH<sub>3</sub>, EVOCs and EPOA = -49%, -45%, -20%, -64%, and -20%? Will the validation results change if you change the definition of scenarios?

[Response]: The scenarios were designed from a 100 Latin Hypercube Sampling method. The two scenarios were selected randomly from the 100 samples, for the purpose of analyzing different location and time. The validation on averages (time and location) are conducted for all 100 samples as we discussed in section 3.1. Here we just pick up two scenarios to represent two different control levels, moderate and strict. The validation results might slight change if we change the scenarios, however, the performance should be similar to the two we presented here.

To clarify this point, we added some discussion in the revised manuscript as follows:

(Page 8 Line 37) “These two scenarios are selected from the OOS100, to represent two kinds of emission levels, moderate and strict respectively, for the purpose of analyzing the pf-RSM performance under different locations and times. Please note that the validation results might slight change if we change the scenarios, however, the performance should be similar to the two we presented here.”

[Comment]: Page 9, line 19-20. The sentence is confusing. Please try to rephrase it.?

[Response]: As the reviewer suggested, we revised the sentence as follows:

(Page 10 Line 27) “The PR values calculated in this study also indicate that the control of NO<sub>x</sub> (with less than 20%-60% reduction, =1-PR) could result in an increase of O<sub>3</sub>; however, O<sub>3</sub> would decrease with substantial control of NO<sub>x</sub> (with greater than 20%-60% reduction).”

[Comment]: Figure 5. The font size of legend does not fit the graph.

[Response]: Following the reviewer’s suggestion, we reduced the font size of legend in the Figure 5 in the revised manuscript.

**Reply to comments from Referee #2 on “Quantification of the enhanced effectiveness of NO<sub>x</sub> control from simultaneous reductions of VOC and NH<sub>3</sub> for reducing air pollution in Beijing-Tianjin-Hebei region, China” by Xing et al.**

We thank the reviewer for the detailed and thoughtful review of our manuscript. Incorporation of the reviewer’s suggestion has led to a much improved manuscript. Detailed below is our response to the issues raised by the reviewer. We also detail the specific changes incorporated in the revised manuscript in response to the reviewer’s comments.

**[Comment]:** *This study explores the effectiveness of simultaneous NO<sub>x</sub>, VOC and NH<sub>3</sub> emission control on PM<sub>2.5</sub> and O<sub>3</sub> using CMAQ and a set of polynomial functions. The methods they propose are innovative and computationally efficient, but the presentation of the results and the significance of the findings still need further improvements. Overall, I think there are a number of issues that should be addressed in order to make this paper suitable for publication.*

**[Response]:** We thank the reviewer for recognition of the implications of the results of the analysis presented. We basically followed all the comments and revised manuscript accordingly.

**[Comment]:** *A common problem with statistical polynomial regression is overfitting. The fitting performance will certainly improve with higher orders, but it does not necessarily mean the models represent the true relationships. The authors show a very good fitting performance with inflated R values (0.93 to 1.0, Table 3), but this may actually reflect the models are overfitted. In order for the fitting to be trustworthy, the authors need to prove that the models are not overfitted. You could do so by conducting cross-validation for your model selection by partitioning your data to training and test groups. The test groups should not be used to fit the models, but to evaluate the model performance only.*

**[Response]:** Following the reviewer’s suggestion, we conducted the cross-validation for the pf-RSM model. The results are shown in Table R1.

**Table R1 Performance of PM<sub>2.5</sub> and O<sub>3</sub> prediction using pf-RSM with different training samples**

Num.	Dataset	Dist.	PM <sub>2.5</sub> *										O <sub>3</sub>									
			Jan					Jul					Jan					Jul				
			MeanNE	MaxNE	MeanFE	MaxFE	R	MeanNE	MaxNE	MeanFE	MaxFE	R	MeanNE	MaxNE	MeanFE	MaxFE	R	MeanNE	MaxNE	MeanFE	MaxFE	R
20	LOOCV**	Even	1.92%	9.47%	0.95%	4.54%	0.96	1.92%	9.47%	0.95%	4.54%	0.96	5.46%	30.29%	2.61%	12.58%	0.94	0.42%	2.94%	0.21%	1.51%	0.99
		Margin	6.69%	40.42%	3.19%	16.36%	0.54	3.28%	10.70%	1.64%	5.08%	0.95	3.42%	13.93%	1.69%	6.39%	0.99	0.47%	1.59%	0.24%	0.79%	1.00
	OOS100	Even	2.50%	15.09%	1.24%	6.98%	0.94	1.03%	5.56%	0.52%	2.77%	0.99	2.04%	10.33%	1.01%	4.90%	0.99	0.23%	1.50%	0.12%	0.74%	1.00
		Margin	3.07%	15.02%	1.52%	6.97%	0.93	1.66%	6.89%	0.83%	3.59%	0.98	1.73%	5.53%	0.87%	2.74%	1.00	0.22%	0.86%	0.11%	0.43%	1.00
	OOS15	Even	0.76%	1.86%	0.38%	0.93%	0.99	1.79%	3.33%	0.91%	1.69%	0.97	2.48%	4.84%	1.23%	2.38%	0.96	1.08%	3.29%	0.54%	1.69%	0.92
		Margin	1.61%	3.38%	0.80%	1.66%	0.96	2.59%	5.23%	1.27%	2.53%	0.95	2.83%	4.69%	1.39%	2.27%	0.96	1.13%	2.49%	0.56%	1.23%	0.84
30	LOOCV	Even	2.00%	5.30%	1.00%	2.62%	0.97	1.73%	7.00%	0.86%	3.37%	0.98	1.06%	5.63%	0.53%	2.72%	1.00	0.30%	1.80%	0.15%	0.90%	1.00
		Margin	3.35%	9.25%	1.67%	4.64%	0.93	2.06%	7.88%	1.03%	3.84%	0.98	2.85%	10.05%	1.41%	4.79%	0.99	0.29%	1.03%	0.15%	0.52%	1.00
	OOS100	Even	1.89%	9.90%	0.94%	4.71%	0.97	1.14%	4.34%	0.57%	2.12%	0.99	1.25%	12.41%	0.64%	5.77%	0.99	0.19%	1.46%	0.09%	0.73%	1.00
		Margin	2.19%	11.96%	1.09%	5.63%	0.97	1.07%	4.11%	0.53%	2.03%	0.99	1.65%	4.87%	0.82%	2.39%	1.00	0.24%	0.89%	0.12%	0.44%	1.00
	OOS15	Even	1.13%	2.32%	0.57%	1.18%	0.99	1.49%	2.64%	0.75%	1.34%	0.98	1.52%	2.82%	0.77%	1.44%	0.99	0.59%	2.48%	0.29%	1.22%	0.92
		Margin	0.74%	1.77%	0.37%	0.89%	0.99	1.21%	2.35%	0.60%	1.17%	0.99	1.61%	2.73%	0.80%	1.35%	0.99	0.70%	2.10%	0.35%	1.04%	0.90

40	LOOCV	Even	1.25%	4.71%	0.62%	2.34%	0.98	0.23%	1.60%	0.11%	0.80%	1.00	1.46%	7.22%	0.73%	3.46%	0.99	0.23%	1.60%	0.11%	0.80%	1.00
		Margin	<b>2.12%</b>	<b>8.00%</b>	<b>1.06%</b>	<b>4.07%</b>	<b>0.97</b>	<b>0.27%</b>	<b>1.64%</b>	<b>0.14%</b>	<b>0.83%</b>	<b>1.00</b>	<b>2.13%</b>	<b>9.89%</b>	<b>1.06%</b>	<b>4.75%</b>	<b>0.99</b>	<b>0.27%</b>	<b>1.64%</b>	<b>0.14%</b>	<b>0.83%</b>	<b>1.00</b>
	OOS100	Even	1.79%	8.60%	0.89%	4.12%	0.98	0.81%	5.37%	0.40%	2.61%	0.99	1.54%	10.11%	0.79%	5.46%	0.99	0.19%	1.34%	0.09%	0.67%	1.00
		Margin	<b>1.88%</b>	<b>8.25%</b>	<b>0.93%</b>	<b>3.95%</b>	<b>0.98</b>	<b>1.00%</b>	<b>4.28%</b>	<b>0.50%</b>	<b>2.17%</b>	<b>0.99</b>	<b>1.19%</b>	<b>3.96%</b>	<b>0.60%</b>	<b>2.03%</b>	<b>1.00</b>	<b>0.19%</b>	<b>0.78%</b>	<b>0.09%</b>	<b>0.39%</b>	<b>1.00</b>
	OOS15	Even	0.35%	0.79%	0.18%	0.39%	1.00	1.12%	2.05%	0.56%	1.03%	0.99	1.04%	2.34%	0.53%	1.19%	0.99	0.66%	2.03%	0.33%	1.00%	0.92
		Margin	<b>0.85%</b>	<b>1.80%</b>	<b>0.43%</b>	<b>0.91%</b>	<b>0.99</b>	<b>1.07%</b>	<b>2.08%</b>	<b>0.54%</b>	<b>1.05%</b>	<b>0.99</b>	<b>0.99%</b>	<b>2.34%</b>	<b>0.49%</b>	<b>1.16%</b>	<b>0.99</b>	<b>0.58%</b>	<b>1.93%</b>	<b>0.29%</b>	<b>0.96%</b>	<b>0.93</b>
50	LOOCV	Even	1.20%	3.91%	0.60%	1.94%	0.98	0.94%	5.29%	0.47%	2.65%	0.99	0.88%	4.22%	0.44%	2.17%	1.00	0.15%	0.75%	0.07%	0.38%	1.00
		Margin	1.47%	6.35%	0.74%	3.28%	0.99	1.34%	4.88%	0.67%	2.47%	0.99	1.85%	6.13%	0.93%	3.04%	0.99	0.22%	0.84%	0.11%	0.42%	1.00
	OOS100	Even	1.53%	8.17%	0.76%	3.92%	0.98	0.74%	3.77%	0.37%	1.88%	1.00	0.98%	6.50%	0.49%	3.10%	1.00	0.15%	1.07%	0.08%	0.54%	1.00
		Margin	1.71%	8.66%	0.84%	4.15%	0.98	0.86%	3.81%	0.43%	1.89%	0.99	1.39%	4.71%	0.70%	2.30%	1.00	0.18%	0.66%	0.09%	0.33%	1.00
	OOS15	Even	0.88%	1.39%	0.44%	0.70%	0.99	0.72%	1.92%	0.36%	0.97%	0.99	1.10%	2.42%	0.55%	1.22%	0.99	0.54%	1.96%	0.27%	0.97%	0.96
		Margin	0.93%	2.48%	0.47%	1.26%	0.99	0.81%	1.70%	0.41%	0.86%	0.99	1.20%	2.33%	0.59%	1.15%	0.99	0.45%	1.90%	0.23%	0.94%	0.94

Basically, the statistics of cross-validation are in the same order as shown in out-of-sample validations (OOS100 and OOS15). The performance in pf-RSM gets better along with the increase of sample numbers. Interesting finding is that the pf-RSM with marginal processing exhibits worse performance than that with even sampling method in cross-validation. That is because the samples with marginal processing are located closer to margin areas where is more difficult to predict (Xing et al., 2011). That also implies the samples with marginal processing has better representation of the variability. Nevertheless, the results of validations suggest the pf-RSM with current number of samples are not over-fitted, and the training samples selected in fitting the system is recommended to be 40 training samples with marginal processing.

To address the reviewer's concern, we added the discussion in the revised manuscript, as follows:

(Page 6 Line 19) "Method of leave-one-out cross validation (LOOCV) was used to examine whether the statistical polynomial regression is overfitting. The definition of LOOCV is to use a single sample from the original datasets as the validation data, and the remaining sample as the training data to build pf-RSM."

(Page 7 Line 16) Similar results are found in the cross validation (i.e., LOOCV), as the performance in pf-RSM gets better along with the increase of sample numbers. Basically, the statistics of cross-validation are in the same order as shown in out-of-sample validations (OOS100 and OOS15), except for the case of 20 training samples with marginal processing (worse performance due to under-fitting problem). Interesting finding is that the pf-RSM with marginal processing exhibits worse performance than that with even sampling method in cross-validation. That is because the samples with marginal processing are located closer to margin areas where is more difficult to predict (Xing et al., 2011). That also implies the samples with marginal processing has better good representation of the variability. Nevertheless, the results of validations suggest the pf-RSM with current number of samples are not over-fitted, and the training samples selected in fitting the system is recommend to be 40 training samples with marginal processing."

[Comment]: *The polynomial functions assume the changes in pollutant concentration only depend on the changes of local emissions, but transport, meteorology, and deposition can also change the concentration. The authors need to provide justification why these factors are not considered.*

[Response]: The design of pf-RSM is to investigate the air pollution responses to the emission perturbation which is related to the design of effective control policy. However, as the reviewer mentioned, the processes of transport, meteorology, and deposition can also change the concentration. Thus, we need atmospheric chemical transport

model (i.e., CMAQ model in this study) to represent those influences on air pollution. But, contributions from those processes on air pollution are uncontrolled, thus we use fixed meteorological condition to drive all CMAQ simulation runs. To quantify the response of pollution to emission changes, we conducted multiple-CMAQ simulations under different emission scenarios and adopted statistic fitting or regression method to combine those simulations into a statistic model (i.e., RSM) which represents the response of air pollution to emission changes. The contribution of transport, meteorology, and deposition on pollution has already been considered in the CMAQ simulation, though we are not going to build up a response function of concentration to those variables.

To address the reviewer's concern, we clarified this point in the revised manuscript as follows:

(Page 3 Line 13) "We used the same meteorological condition for those multiple scenarios and only the emissions were changed in different scenarios."

[Comment]: *The authors did a good job synthesizing their results concisely, but I would recommend the authors provide more insights into the numbers they reported. The interpretation of the results could be improved by:*

1) Considering large body literature behind this topic and comparing your results with previous studies, especially those observation-based studies.

[Response]: As the reviewer suggested, we compared the model-based results with observation studies that use indicator to identify the O<sub>3</sub> chemistry, as follows:

(Page 10 Line 20) "Our results are consistent with the observational studies that use indicator to identify the O<sub>3</sub> chemistry. For example, Liu et al (2016) studied on the ratios of HCHO over NO<sub>2</sub> from the satellite retrieves and found that local ozone production in urban Beijing is VOC-limited when there are no substantial changes in NO<sub>x</sub> emission in 2015. Chou et al. (2009) found that Beijing urban area was "VOC-limited" region based on the observation of NO, NO<sub>x</sub> and NO<sub>y</sub> at the Peking University site during August 15 to September 11 in 2006. Jin and Holloway (2015) calculated the ratio of HCHO to NO<sub>2</sub> from the OMI instrument aboard the Aura satellite and found the O<sub>3</sub> production is more likely to be VOC-limited over urban areas and NO<sub>x</sub>-limited over rural and remote areas in China from 2005 to 2013."

We've added the comparison above in the revised manuscript.

## Reference

Liu, H., Liu, C., Xie, Z., Li, Y., Huang, X., Wang, S., Xu, J. and Xie, P., 2016. A paradox for air pollution controlling in China revealed by "APEC Blue" and "Parade Blue". *Scientific reports*, 6, p.34408.

Jin, X. and Holloway, T., 2015. Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres*, 120(14), pp.7229-7246.

Chou, C. C.-K., Tsai, C.-Y., Shiu, C.-J., Liu, S. C. and Zhu, T.: Measurement of NO<sub>y</sub> during Campaign of Air Quality Research in Beijing 2006 (CAREBeijing-2006): Implications for the ozone production efficiency of NO<sub>x</sub>, *J. Geophys. Res.*, 114, D00G01, doi:10.1029/2008JD010446, 2009.

2) Providing more mechanistic reasoning on the results you provided. For example, The authors show the impacts of emission reduction vary in space and time (Section 3.2 and 3.3), but they do not provide any insights for the such variations. There is also no discussion on how the effectiveness of emission control vary with meteorology.



[Response]: As the reviewer suggested, we added following discussion in the revised manuscript:

(Page 9 Line 18) “The day-to-day variability of O<sub>3</sub> depends on the budget of O<sub>3</sub> source and sink influenced by meteorological variables including actinic flux, temperature, humidity, and precipitation, etc.”

(Page 9 Line 22) “The meteorological condition will also play an important role in the effectiveness of emission controls. Reductions in O<sub>3</sub> were noticeable in both control cases, particularly on days when O<sub>3</sub> levels were high. However, increases in O<sub>3</sub> were observed on July 21-23 (precipitation event occurred across North China Plain), after the controls were applied and when O<sub>3</sub> levels were low. This can be explained by the O<sub>3</sub> chemistry scheme being in a strong VOC-limited condition on days with low O<sub>3</sub> levels, resulting in enhanced O<sub>3</sub> from NO<sub>x</sub> controls (Xing et al., 2011). Thus, the emission controls usually become less effective under unfavorable meteorological condition for O<sub>3</sub> production. The pf-RSM also reproduced increases in O<sub>3</sub> on those days.”

(Page 11 Line 23) “However, it might need further confirmed by more applications in other regions outside BTH and for a whole year analysis to better represent the seasonality.”

[Comment]: *Page 1 Line 10: It looks like you're talking about the O<sub>3</sub> trend since 2010 (after the emission control), but Li et al. analyzed the trend between 2006 and 2011. I suggest cite a more recent paper that reflects the O<sub>3</sub> trend since 2010, otherwise it's misleading.*

[Response]: We agree with the reviewer that the original statement is misleading. We clarified it in the revised manuscript as follows:

(Page 2 Line 8) “Since early 2010s (late 2000s in some regions such as Pearl River Delta), strict regulations have been implemented on power plants and vehicle emissions, leading to a considerable NO<sub>2</sub> reduction witnessed by the declining trend in satellite-retrieved NO<sub>2</sub> column densities (i.e., reduced by 32% from 2011 to 2015, Liu et al., 2016).”

(Page 2 Line 16) “The annual averaged O<sub>3</sub> was increased by 0.86 ppb/year from 2006 to 2011 in Guangdong, accompanied by a correspondingly NO<sub>2</sub> reduction of 0.61 ppb/year (Li et al., 2014). The recent observation data suggested a continue increasing trend of 8-hour maxima O<sub>3</sub> in Zhuhai (from 128 to 142 μg m<sup>-3</sup>) and Shenzhen (from 122 to 134 μg m<sup>-3</sup>) in Pearl River Delta from 2013 to 2016.”

[Comment]: *Page 4 Line 30: Why do you choose 4th degree over 3rd degree? The difference is small, and you didn't give any statistical justification.*

[Response]: We replotted the difference between fitting with 4th degree and 3rd degree, as giving in Figure R1. The statistics show that the fitting with 4<sup>th</sup> degree has higher R values (R<sup>2</sup> in fitting with 4<sup>th</sup> order is 0.9985 compared to 0.9735 in fitting with 3<sup>rd</sup> order) and smaller errors (MeanFE in fitting with 4<sup>th</sup> order is 0.2 compared to 0.6 in fitting with 3<sup>rd</sup> order).

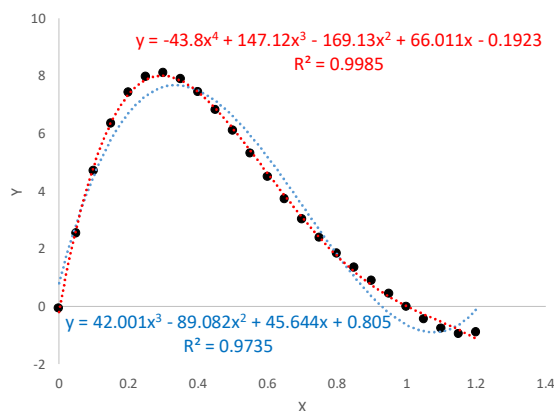


Figure R1 Fitting the PM2.5 responsive function to NOx with a polynomial of a single indeterminate plots with 3<sup>rd</sup> and 4<sup>th</sup> order

We clarified this point in the revised manuscript as follows,

(Page 5 Line 2) “Better performance is shown in fitting with 4<sup>th</sup> order polynomial (R=0.999, MeanFE=0.2) than with 3<sup>rd</sup> order polynomial (R=0.987, MeanFE =0.6).”

[Comment]: Page 5 Line 30: Why “the training samples need to be as small as possible”? With small number of samples, the coefficients are very likely to be unstable, especially since you’re fitting high-order polynomial functions here.

[Response]: The limitation of RSM model is its heavy computing burden associated with the “samples” development. Each sample represents a CTM simulation under certain emission scenario. One CTM simulation for a typical month simulation period requires 400 CPU-hour, depending on the simulated domain size and selected mechanism. It is true that the for statistic fitting, we want as many as training samples. However, for CTM simulations, the less the better.

One advantage of pf-RSM is that its development requires about only 30% samples of traditional regression-based model, which significantly reduce the computing burden.

We clarified this point in the revised manuscript as follows:

(Page 2 Line 38) “The traditional RSM model is based on regression from thousands of “brute-force” simulations with chemical transport model (CTM) by using a maximum likelihood estimation - experimental best linear unbiased predictors (hereafter referred as “regression-based RSM”). However, such a large amount of CTM simulations (each simulation represents one training sample) required by RSM results in heavy computing burden (usually one CTM scenario for a month simulation needs 400 CPU-hour, depending on the simulated domain size and selected mechanism) which largely limits the application of traditional RSM.”

(Page 6 Line 2) “To minimize the number of CTM simulations (one simulation scenario represents one training sample), the number of training samples needed to be as small as possible, but greater than the number of terms (i.e., unknown coefficients) in the polynomial function.”

(Page 11 Line 11) “After the application of a prior knowledge of the pollutant responsiveness to emissions in the RSM system, the cases required for single regional pf-RSM development were substantially decreased to 40

samples, compared with the previous requirement of over 100 samples, imply that the fitting-based RSM (i.e., pf-RSM) is three time faster than previous regression-based RSM (i.e., the number of CTM simulations needed in pf-RSM is 60% less than that required by previous regression-based RSM).”

[Comment]: *Page 8 Line 9: It's not clear to me how you set up the two emission control scenarios. Why do the magnitudes of emission reduction differ among species? How would the agreement between CMAQ and pf-RSM change if the emissions are reduced uniformly?*

[Response]: The scenarios were designed from a 100 Latin Hypercube Sampling method. The two scenarios were selected randomly from the 100 samples, for the purpose of analyzing different location and time. The validation on averages (time and location) are conducted for all 100 samples as we discussed in section 3.1. Here we just pick up two samples (scenarios) to represent two different control levels, moderate and strict. The different magnitudes of emission reduction among species just present a certain scenario. The validation results might slight change if we change the scenarios (e.g., all pollutants reduced uniformly), however, the performance should be similar to the two we presented here.

To clarify this point, we added some discussion in the revised manuscript as follows:

(Page 8 Line 37) “These two scenarios are selected from the OOS100, to represent two kinds of emission levels, moderate and strict respectively, for the purpose of analyzing the pf-RSM performance under different locations and times. Please note that the validation results might slight change if we change the scenarios, however, the performance should be similar to the two we presented here.”

[Comment]: *Page 9 Line 5: The word “observe” is misleading. There are no observations in this study.*

[Response]: To avoid confusion, we modified the sentence in the revised manuscript as follows:

(Page 10 Line 4) “Larger FR values (slightly lower than 1.0) were shown in the central and southern regions (i.e., Beijing, Tianjin and HebeiS) than in other regions”

[Comment]: *Page 9 Line 15: Please be more specific how your study is “consistent with findings of previous studies”. It's also worthy mentioning how your study differs from previous studies in terms of methodology, results etc.*

[Response]: As the reviewer suggested, we clarified those sentence in the revised manuscript as follows:

(Page 10 Line 8) “In both January and July, most of the urban areas present NH<sub>3</sub>-rich condition with FR from 0.75-0.95 (Table 4), implying the NH<sub>3</sub> is sufficiently abundant to neutralize extra nitric acid produced by an additional 5%-35% (i.e., =1/FR-1) of NO<sub>x</sub> emissions. The result is consistent with our previous study (Wang et al., 2011) which reported that NH<sub>3</sub> is sufficiently abundant to neutralize extra nitric acid produced by an additional 25% of NO<sub>x</sub> emissions in north China Plain based on a traditional regression-based RSM study.”

(Page 10 Line 17) “That is consistent with the findings of previous studies (Xing et al., 2011) which used a traditional regression-based RSM and found that the PR changes from 0.8 to 1.2 as the distance from the city center increases.”

[Comment]: Page 9 Line 20: I'd suggest the authors compare your model-based findings with observations (e.g. in situ or satellite observations) that use indicator approach to identify the limiting species for the O<sub>3</sub> production.

[Response]: As the reviewer suggested, we compared the model-based results with observation studies that use indicator to identify the O<sub>3</sub> chemistry, as follows:

(Page 10 Line 20) "Our results are consistent with the observational studies that use indicator to identify the O<sub>3</sub> chemistry. For example, Liu et al (2016) studied on the ratios of HCHO over NO<sub>2</sub> from the satellite retrieves and found that local ozone production in urban Beijing is VOC-limited when there are no substantial changes in NO<sub>x</sub> emission in 2015. Chou et al. (2009) found that Beijing urban area was "VOC-limited" region based on the observation of NO, NO<sub>x</sub> and NO<sub>y</sub> at the Peking University site during August 15 to September 11 in 2006. Jin and Holloway (2015) calculated the ratio of HCHO to NO<sub>2</sub> from the OMI instrument aboard the Aura satellite and found the O<sub>3</sub> production is more likely to be VOC-limited over urban areas and NO<sub>x</sub>-limited over rural and remote areas in China from 2005 to 2013."

We've added the comparison above in the revised manuscript.

Chou, C. C.-K., Tsai, C.-Y., Shiu, C.-J., Liu, S. C. and Zhu, T.: Measurement of NO<sub>y</sub> during Campaign of Air Quality Research in Beijing 2006 (CAREBeijing-2006): Implications for the ozone production efficiency of NO<sub>x</sub>, J. Geophys. Res., 114, D00G01, doi:10.1029/2008JD010446, 2009.

Jin, X. and Holloway, T.: Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument. Journal of Geophysical Research: Atmospheres, 120(14), 7229-7246, 2015.

Liu, H., Liu, C., Xie, Z., Li, Y., Huang, X., Wang, S., Xu, J. and Xie, P.: A paradox for air pollution controlling in China revealed by "APEC Blue" and "Parade Blue". Scientific reports, 6, 34408, 2016.

[Comment]: Page 9 Line 23: The results you show here are just for January and July, but how about other months, especially spring (or fall) when O<sub>3</sub> production transitions from VOC-limited (or NO<sub>x</sub>-limited) to NO<sub>x</sub>-limited (VOC-limited)? Would you expect the effectiveness of emission control show any seasonality?

[Response]: It is true that the O<sub>3</sub> chemistry varies under different meteorological conditions. Even in the same month, the O<sub>3</sub> response to precursor reductions varies significantly on either high or low O<sub>3</sub> days, as shown in section 3.2. Reductions in O<sub>3</sub> were noticeable particularly on days when O<sub>3</sub> levels were high. However, increases in O<sub>3</sub> were observed on July 21-23, after the controls were applied and when O<sub>3</sub> levels were low. This can be explained by the O<sub>3</sub> chemistry scheme being in a strong VOC-limited condition on days with low O<sub>3</sub> levels, resulting in enhanced O<sub>3</sub> from NO<sub>x</sub> controls (Xing et al., 2011). It is expected that the O<sub>3</sub> chemistry will be different in other months such as spring or fall. Further work is necessary to be conducted for a whole cycle year and to get a better representative of O<sub>3</sub> seasonality.

Nevertheless, based on the daily analysis of O<sub>3</sub> responses to precursor reductions in this study and also in our previous study (Xing et al., 2011), we can see that the effectiveness of emission control varies under different days. Generally, the controls on precursors will be more effective in reducing peak O<sub>3</sub> concentrations, and will be less effective for days with low O<sub>3</sub> levels which is usually in a strong VOC-limited condition.

We clarified this point in the revised manuscript.

(Page 9 Line 22) "The meteorological condition will also play an important role in the effectiveness of emission controls."

(Page 9 Line 27) “Thus, the emission controls usually become less effective under unfavorable meteorological condition for O<sub>3</sub> production.”

(Page 11 Line 23) “However, it might need further confirmed by more applications in other regions outside BTH and for a whole year analysis to better represent the seasonality.”

[Comment]: *Figure 9: How does meteorology affect the day-to-day variability of O<sub>3</sub> and the effectiveness of emission controls?*

[Response]: The day-to-day variability of O<sub>3</sub> depends on the budget of O<sub>3</sub> source and sink influenced by meteorological variables including actinic flux, temperature, humidity, and precipitation, etc. For example, there was a precipitation event occurred during July 21-23 in North China Plain, resulting in a lower O<sub>3</sub> level across all 5 regions. Besides, the unfavorable meteorological condition for O<sub>3</sub> production makes emission controls become less effectiveness. Since NO<sub>x</sub> become more abundant under unfavorable meteorological condition for photolysis, resulting in a stronger VOC-limited condition (Xing et al., 2011). Thus the emission controls become less effectiveness on low O<sub>3</sub> days.

We added following discussion in the revised manuscript.

(Page 9 Line 17) “The daily series of the CMAQ-simulated and pf-RSM-predicted 24-hour averaged PM<sub>2.5</sub> and 1-hour maxima O<sub>3</sub> in baseline and two control scenarios are shown in Figure 9. The day-to-day variability of O<sub>3</sub> depends on the budget of O<sub>3</sub> source and sink influenced by meteorological variables including actinic flux, temperature, humidity, and precipitation, etc.”

(Page 9 Line 22) “The meteorological condition will also play an important role in the effectiveness of emission controls. Reductions in O<sub>3</sub> were noticeable in both control cases, particularly on days when O<sub>3</sub> levels were high. However, increases in O<sub>3</sub> were observed on July 21-23 (precipitation event occurred across North China Plain), after the controls were applied and when O<sub>3</sub> levels were low. This can be explained by the O<sub>3</sub> chemistry scheme being in a strong VOC-limited condition on days with low O<sub>3</sub> levels, resulting in enhanced O<sub>3</sub> from NO<sub>x</sub> controls (Xing et al., 2011). Thus, the emission controls usually become less effective under unfavorable meteorological condition for O<sub>3</sub> production. The pf-RSM also reproduced increases in O<sub>3</sub> on those days.”

[Comment]: *Table 4: Why are there missing values for HebeiN?*

[Response]: Since the PR is larger than 1.2 in HebeiN, the NO<sub>x</sub> control will always lead to a reduction in O<sub>3</sub>. Thus it is not necessary to estimate the reduction ratio of VOC to NO<sub>x</sub> to avoid increasing O<sub>3</sub> for HebeiN.

The estimated FR in HebeiN is larger than 1.2, indicating strong NH<sub>3</sub> poor condition. The extra benefit from simultaneous reduction of NH<sub>3</sub> in HebeiN in July is estimated as 0.074 μg m<sup>-3</sup> PM<sub>2.5</sub> per 1% reduced NH<sub>3</sub>.

The values in HebeiN have been added in Table 4 in the revised manuscript.