

Light blue: Reviewer comments;

Black: Response to the reviewer;

Black: Add in the manuscript; Black; Underline: Changes in the manuscript

Comment on acp-2020-1330

Referee comment on "Development of New Emission Reallocation Method for Industrial Nonpoint Source in China" by Yun Fat Lam et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-1330-RC1>, 2021

This paper provided a new industrial nonpoint source reallocation method, a NPS method, based on blue-roof industrial buildings based on satellite imagery. The results indicate using the NPS method will improve the model performance compared with the conventional method of using population density. In general, the paper is clearly organized and easy to follow, and provided a new spatial allocation emission data for the Greater Bay Area (GBA) area. However, the authors mainly compared this method with the regional emission inventories using top-down spatial allocation methods. Currently, the high-resolution emission inventories of most key regions in China adopt point-source processing based on "bottom-up" method. **I encouraged the authors to expand the scope of the paper and to discuss more with the point-based high-resolution emission inventory development.** Some more detailed information should also be provided and discussed, before the paper can be accepted for final publication.

Response:

We wish to thank the reviewer for the valuable comments and suggestions, which help us improve the manuscript. For this particular suggestion, we have restructured the paper and added a new emission case (i.e., the point/area based bottom-up case), which utilized a local point/area based inventory in GBA. Detailed changes are covered in the latter of this document.

1) Details follow. As mentioned above, my main concern is that the discussion could be expanded a little bit, and would be focused more on the comparison with point-source based spatial allocation method. That might be more helpful for the whole research community. It invites more review on published work for the high-resolution regional emission inventory, and more comparison and discussion with these inventories I can understand that using the NPS method can effectively improve the spatial allocation of emissions.

Response:

Thanks for the suggestion. To address this comment, we have added a new simulation case called the “point/area based bottom-up” (btmUp) case. The bottom-up case is the case that adopts the detailed unit process data and the factory standard industrial Classification (SIC)/Source Classification Codes (SCC)) to derive the emission allocation factors for GBA using the Sparse Matrix Operator Kernel Emissions (SMOKE). The following figure (Figure 5) shows the spatial differences in PM_{2.5} emission among the population based, blue-roof based and point/area based bottom-up approaches. The Figure 5c was added to the existing paper. In order to compare the predictability from these emission methods, additional CMAQ runs using the “bottom-up” case were performed. By comparing the CMAQ results between the blue-roof case and btmUp case, it allows us to better understand the air quality effect from different inventories. Figure 5 shows the spatial comparison of PM_{2.5} emission, and Figure 7 shows the time-series of station CN_1379A. This is the same station which we had used in the last submission. The btmUp case is shown in orange colour. For the detailed discussion, please refer to “Changes in the text” below.

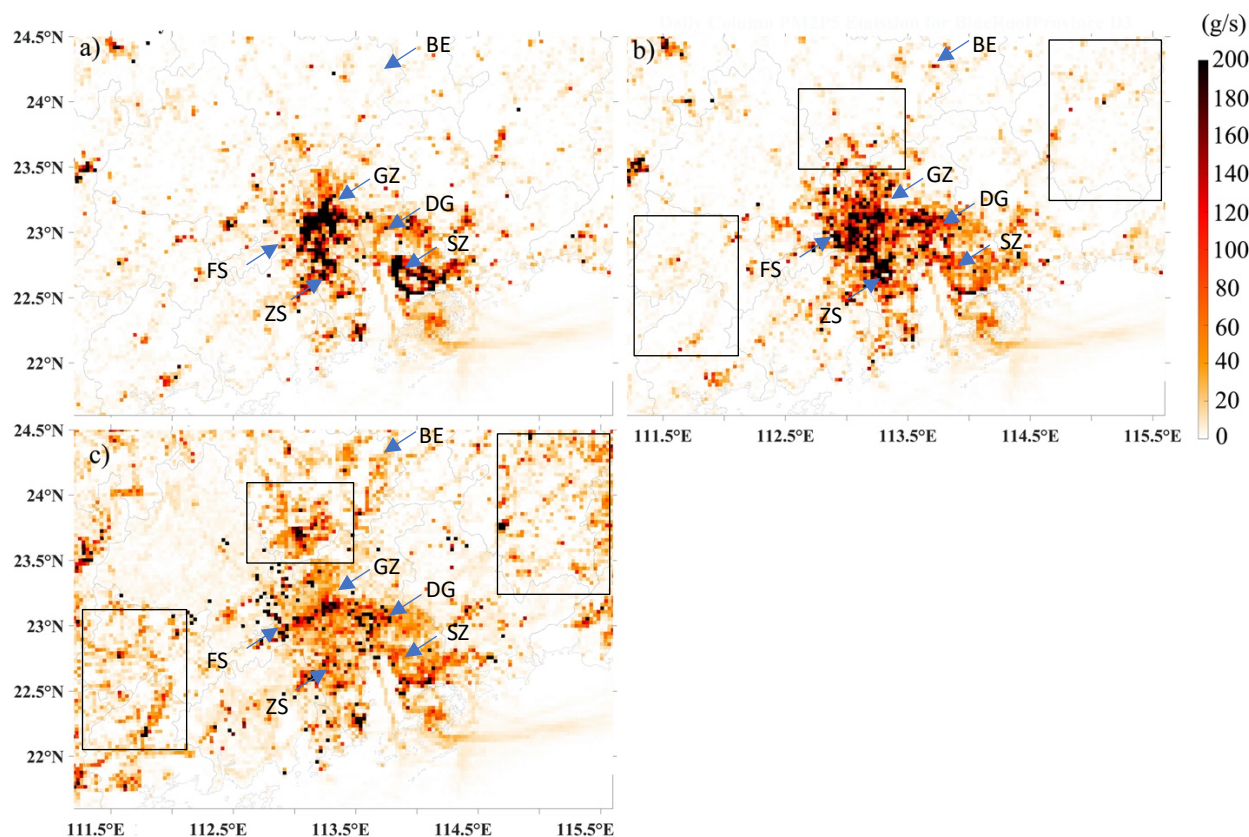


Figure 5: Daily column total of PM_{2.5} emission from D3 (3 km) domain: a) Base case, b) Blue-roof case, and c) point/area based BtmUp case. Note: Blue arrows indicate Foshan (FS), Guangzhou (GZ), Shenzhen (SZ), Dongguan (DG), Zhongshan (ZS), and BE (Blue-roof Example). Boxes indicate the locations with large spatial differences between the blue-roof and the btmUp cases.

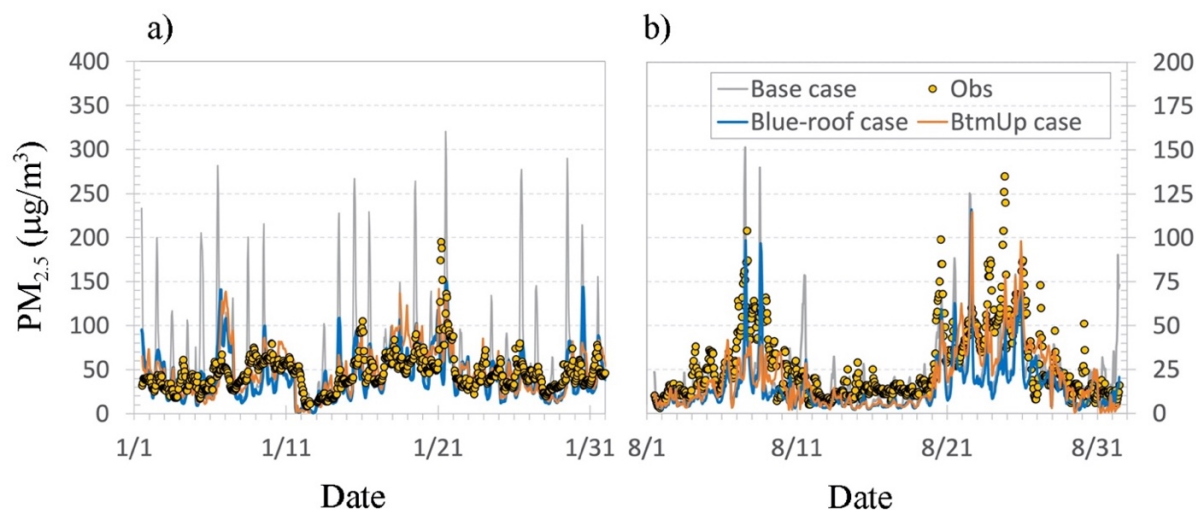


Figure 7: Time series of PM_{2.5} at station CN_1379A (22°31'16.0"N 113°22'36.8"E) – Zhongshan; a) January and b) August.

Add/Changes in text

Line 230 -235 in the manuscript

“Figure 5 shows the daily column total of CMAQ ready PM_{2.5} emissions (January 1, 2015) for the base case, blue-roof case, and point/area based btmUp case from Zhang et al. (2020). In general, more spatial spreading is observed in the blue-roof case than in the base case within the GBA area, but the spread is not as wide as in the point/area based btmUp case. The widespread of PM_{2.5} emission in the btmUp case is attributed to the inclusion of both industrial point and industrial area sources, which was not applied the same way as in the base and blue-roof cases.”

Line 247 -253 in the manuscript

“When comparing the blue-roof case with the point/area based btmUp case (Figure 5c), clear spots of PM_{2.5} underestimation were observed which are shown in the square boxes of Figure 5b pointing at the northeastern and southwestern sides of PRD, and north of Guangzhou. As the focus of the study is to investigate the improvement of the blue-roof surrogate in the MIX industrial sector, rather than the performance differences between the MIX unified emissions and local bottom-up emissions. Therefore, instead of showing the uncertainty of emission inventory which is infeasible here, we have developed spatial blue-roof surrogate (Figure 4b), the comparison of the model-ready emissions (Figure 5), and the time series plots of typical stations (Figure 7) to illustrate the performance of the blue-roof algorithm.”

Line 256 – 262 in the manuscript

“The CMAQ simulation was performed on both base case and blue-roof case to evaluate the air quality impacts of using different allocation methods for industrial emissions. In addition, to better understand how good the blue-roof method performs, the CMAQ results using the local point/area based btmUp emission method adopted from Zhang et al. (2020) were also included in the comparison. Figure 6 shows the simulated monthly average surface PM_{2.5} for base case (a, d), blue-roof case (b, e), and point/area based btmUp case (c, f); the

left (a-c) and right (d-f) panels represent the January and August cases, respectively. As expected, the base case (top panel) has much lower spatial spreading when comparing with the blue-roof (middle panel) and the point/area based btmUp (bottom panel) cases”

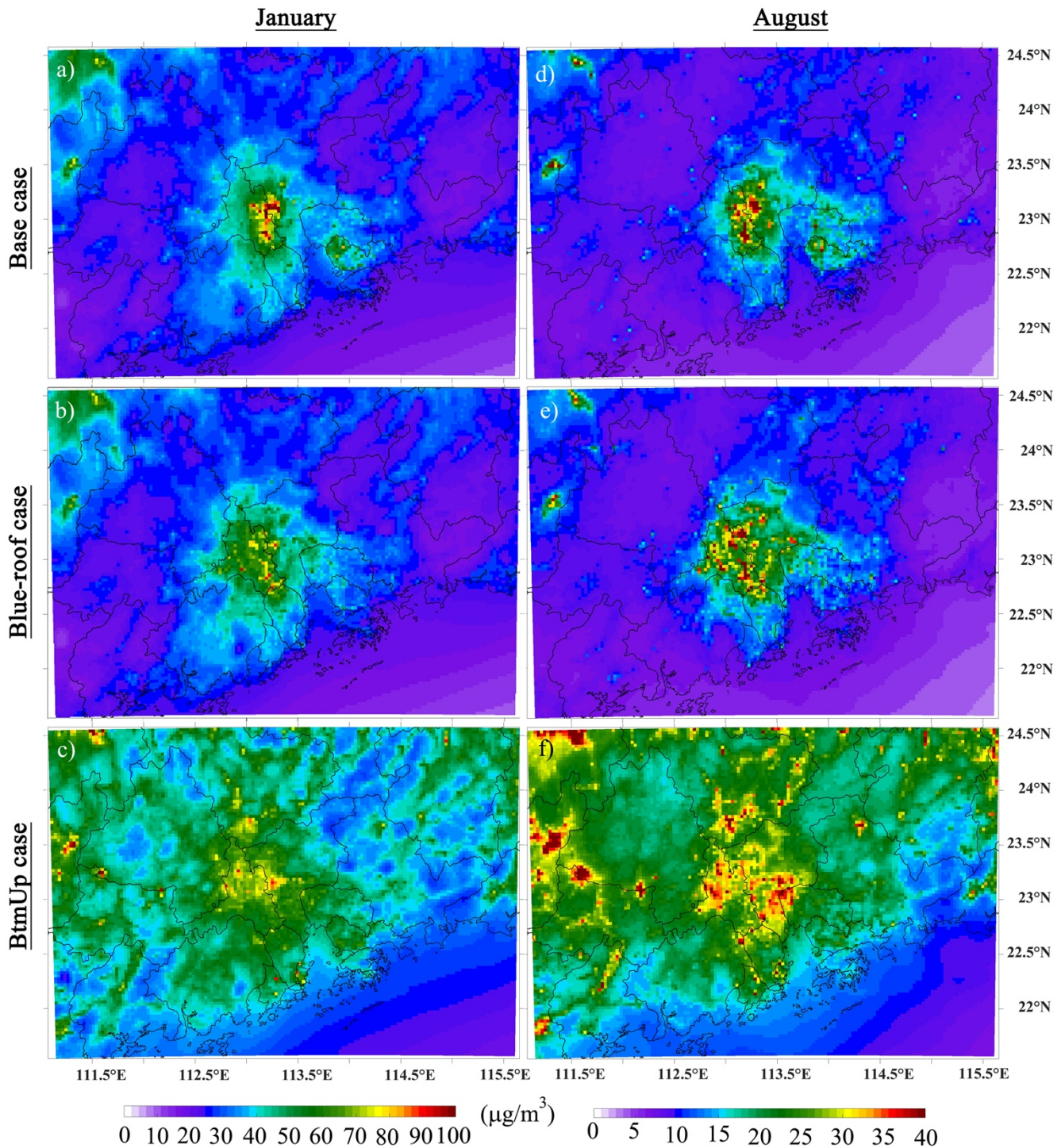


Figure 6: CMAQ predicted monthly surface PM_{2.5}; a) January base case, b) January blue-roof case, c) January point/area based BtmUp case, d) August base case, e) August blue-roof case, and f) August point/area based BtmUp case.

Line 298 – 323 in the manuscript

3.3.2. Performance comparison between the blue-roof case and btmpUp case

“It is essential to evaluate the performance of which the blue-roof case can perform using observations, while it is also interesting to investigate the difference in the performance of the blue-roof allocation method with the local point/area based bottom-up method under a relatively fine-resolution (i.e., 3 km) condition. In general, the CMAQ simulated PM_{2.5} using the blue-roof method (middle panel of Figure 6) has shown a lower spatial spreading than the one using the point/area based btmUp approach (bottom panel of Figure 6). The low spread of PM_{2.5} in the blue-roof case may be attributed to the insufficient separation of existing industrial emissions. As the blue-roof emission approach took the entire industrial emissions and treated them as location-based emissions without assigning any portion of them to area source, lacking the representation of industrial area sources (e.g., fugitives) in the inventory may have resulted in a less spatial spread, as shown in Figure 5b. Moreover, the base unit of industrial emissions in the current approach is “province-level”, which is insufficient to distinguish the industrial speciality for different cities or counties within the domain. From the time-series analysis shown in Figure 7a and b, the RMSE performance of the blue-roof case (blue line) is quite comparable with the point/area based btmUp case (orange line) and observations (yellow dots). This particular example of the blue-roof case (Figure 7b) can even outperform the point/area based btmUp case in predicting PM_{2.5}. From Appendix Table S1 and S2, the average RMSE in January (August) for the base, blue-roof and btmUp cases are 44.8 (25.7) $\mu\text{g}/\text{m}^3$ 33.3 (22.4) $\mu\text{g}/\text{m}^3$, and 27.8 (18.3) $\mu\text{g}/\text{m}^3$, respectively. This illustrates the blue-roof case has outperformed the base case, but still is not as good as the local point/area btmUp case. Figure 9 shows the PM_{2.5} performance of different station types (see Appendix Figure S3). As expected, the point/area based btmUp case has the lowest RMSE among the cases for all station types, while there is a clear improvement of RMSE in urban stations in the blue-roof case; Implementing the blue-roof method has eliminated some of the extreme outliers from the base case, forming a much more narrowed RMSE range. In terms of rural and suburban stations, minor RMSE improvements (i.e., mean values) have been observed. It should be aware that the wider RMSE range showed in the blue-roof case (as compared with the base case) for the suburban category in Figure 9a is just a visual illusion. As the maximum RMSE value of the base case in the suburban category has been plotted as an outlier (dot) instead of a regular line in the upper whisker. Hence, the RMSE range (the two-end whiskers) in the blue-roof case is visually taller than the one in the base case. Appendix Figure S4 shows the station (i.e., CN_1352A) that corresponds to the maximum RMSE in the suburban category, and better performance has been obtained from the blue-roof case (blue line). In the station, the RMSE in January (August) for the base and blue-roof cases are 84.4 (36.0) $\mu\text{g}/\text{m}^3$ and 50.0 (27.5) $\mu\text{g}/\text{m}^3$, respectively.”

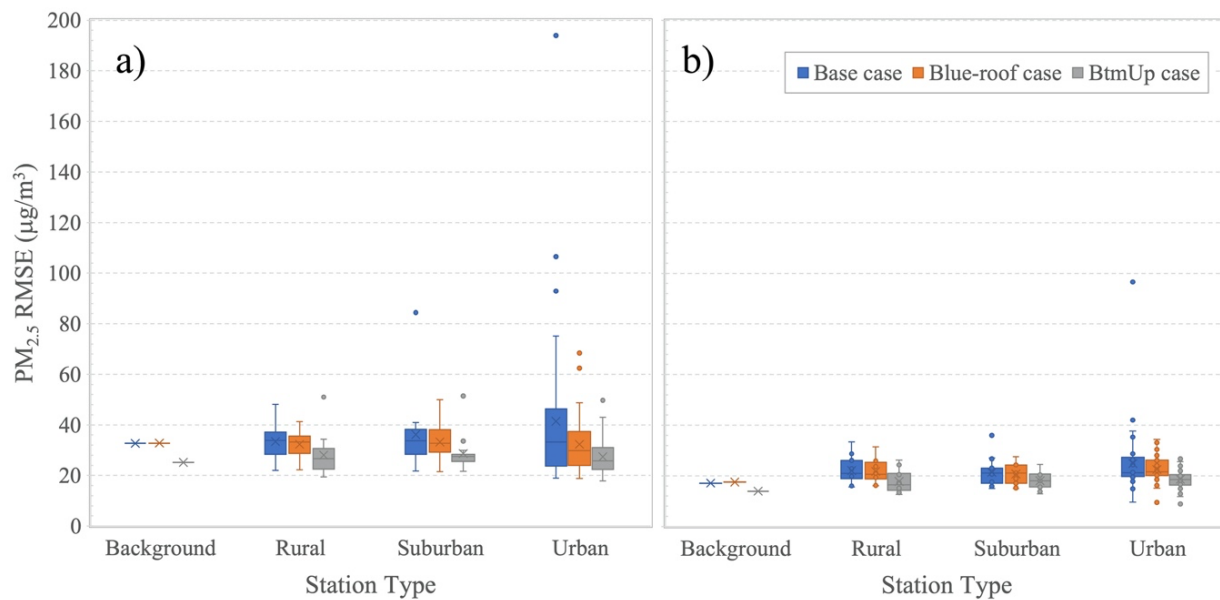


Figure 9: Performance of $PM_{2.5}$ under different station types: a) January and b) August.

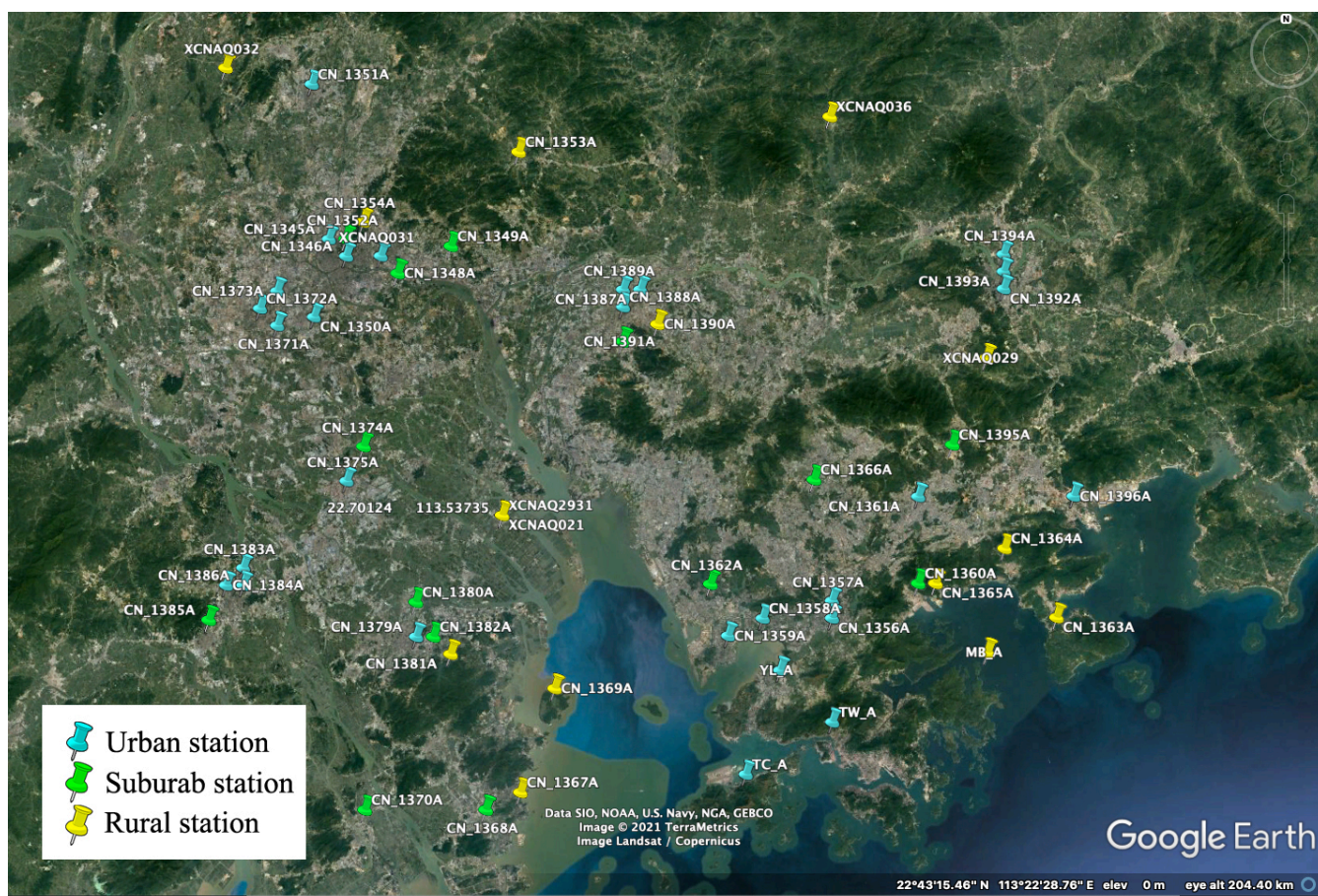


Figure S3: Satellite image with observation stations (© Google).

Line 340 – 352 in the manuscript

Table 3. Summary of performance statistics in the case study.

Pol.	Mon.	RMSE			MB			Improvement (RMSE)			Worsening (RMSE)		
		Base case	BR case	Btm Up case	Base case	BR case	Btm Up case	No. of Station	BR-Base	Max.	No. of Station	BR - Base	Max.
PM _{2.5} ($\mu\text{g}/\text{m}^3$)	JAN	44.8	33.3	27.8	-10.5	-15.4	-0.7	23	-22	-151	8	2	4.6
	AUG	25.7	22.4	18.3	-11.3	-13.7	-6.4	17	-8	-64	8	1.6	2.3
NO ₂ (ppb)	JAN	28.5	25.7	26.1	4.6	2.8	-18.7	30	-5	-33	2	1.2	1.3
	AUG	26.2	23.9	18.6	5.2	3.7	-11.9	28	-5	-31	11	2.6	4.1
O ₃ (ppb)	JAN	24.1	23.9	24.2	-4.5	-3.5	-8.8	11	-2	-5	5	1.6	1.9
	AUG	31.6	31	29.9	7.2	8	-4.9	13	-2	-10	1	1.4	1.4

Note: Pol: Pollutant; Mon: Month; BR: Blue-roof case; RMSE: Root Mean Square Error; MB: Mean Bias; Max: Maximum. The table on the right only shows the station with ± 1 change in RMSE.

“When comparing the blue-roof case with the local point/area based btmUp case, a lower RMSE of PM_{2.5} has been observed in the blue-roof case (Table 3). The difference in the RMSE reflects there is still room for improvement in the blue-roof method. From the large negative MB observed in the MIX emission cases on PM_{2.5}, one suggestion would be to scale up the sectorial PM_{2.5} totals from the MIX inventory using an inverse modelling approach (e.g., satellite inversion or source apportionment), which may lead to a better initial PM_{2.5} emission for CMAQ modelling. In terms of NO₂ and O₃, comparable results (i.e., RMSE) are obtained between the blue-roof and point/area based btmUp cases. Although there is slightly higher RMSE (23.9 ppb vs 18.6 ppb in August) on one of the blue-roof cases, in general, they are all fallen within a similar range of values. In terms of MB, the values in the blue-roof case vary across the seasons, with positive MB on NO₂ and negative MB on O₃ in January, while positive MB on both NO₂ and O₃ in August. For the point/area based btmUp case, negative MB has been observed in both January and August. Among the seasons, it is noted that reducing NO₂ emission in the blue-roof case in January may improve the MB of both NO₂ and O₃ as it reduces the NO₂ titration effect in the ozone formation process and causes higher ozone. However, since the MB (i.e., 3 to 5 ppb) of NO₂ are relatively small (as compared with the MB of PM_{2.5} (-10 to -15 $\mu\text{g}/\text{m}^3$), no NO₂ adjustment is recommended.”

2) However, it seems that the method should not be able to distinguish the differences in emissions between the factories. After all, using images cannot accurately determine the scale of different companies. If it cannot be distinguished, **how much uncertainty was caused by this?**

Response:

Thanks for the comments and question. We don't have the actual emissions on any factories. Even for the point/area based bottom-up industry emission inventory, the emission amount is usually based on the industry types instead of an individual company. Also, we can't calculate the uncertainty of the emission estimate from the False Detection Rate and False Alarm Rate in the blue-roof identification from the process. Although the uncertainty estimate of the MIX inventory is given from the literature, that doesn't reflect the uncertainty from our spatial allocation process.

The main target of the study is to show an improvement of applying a blue-roof surrogate for MIX industry sector. The blue-roof algorithm adapts the satellite-image data to produce a more representative emission spatial surrogate without calculating a new emission inventory. The more representative blue-roof surrogate improves the CMAQ model simulation substantially from our analysis. We admit the uncertainty analysis is of great importance in compiling/calculating a bottom-up emission inventory. However, applying the EI uncertainty analysis is not appropriate at the current stage. Instead of showing the EI uncertainty, the developed blue-roof surrogate (Figure 4b), the comparison of the model ready emissions of PM_{2.5}, and the time series plots of typical stations are chosen to illustrate the performance of the blue-roof surrogate. More wide-spread emission pattern obtained from Figure 5b (applying the blue-roof spatial surrogate) present more consistent pattern with the btmUp case (Figure 5c). Figure 7 shows the improvement of the scenario using blue-roof surrogate comparing with the base case. The trend of time series of the Blue-roof case is closer to that of the btmUp case, as expected.

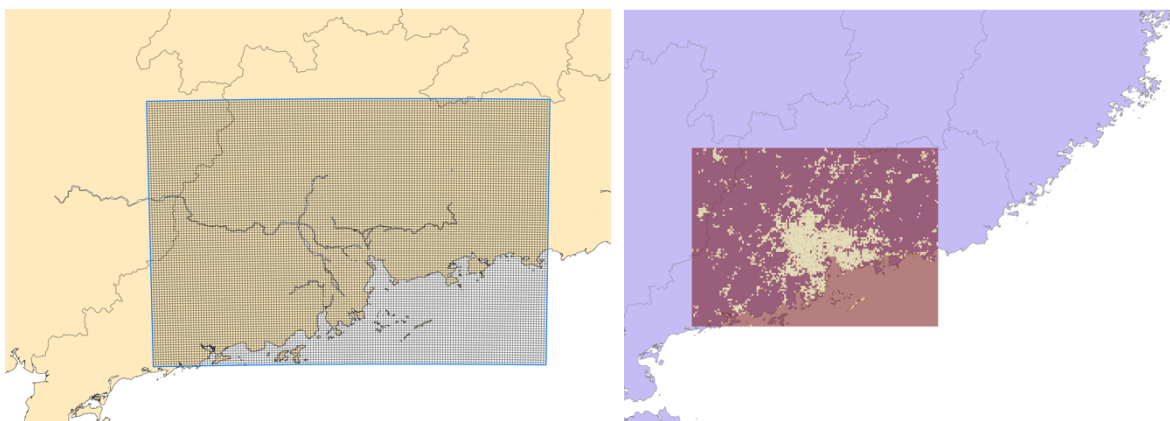


Figure 4: a) Snapshot of D3 (3km) domain grids, and b) Calculated spatial surrogate.

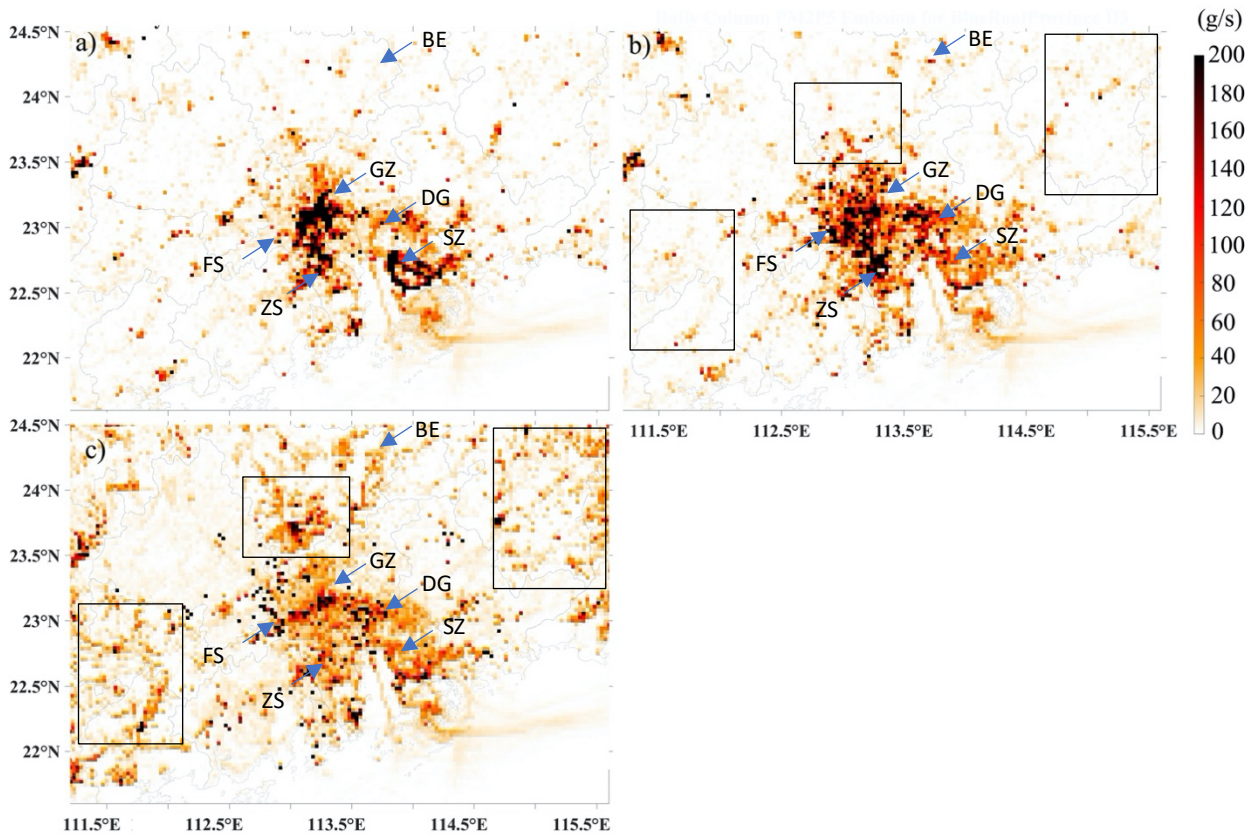


Figure 5: Daily column total of $PM_{2.5}$ emission from D3 (3 km) domain: a) Base case, b) Blue-roof case, and c) point/area based BtmUp case. Note: Blue arrows indicate Foshan (FS), Guangzhou (GZ), Shenzhen (SZ), Dongguan (DG), Zhongshan (ZS), and BE (Blue-roof Example). Boxes indicate the locations with large spatial differences between the blue-roof and the btmUp cases.

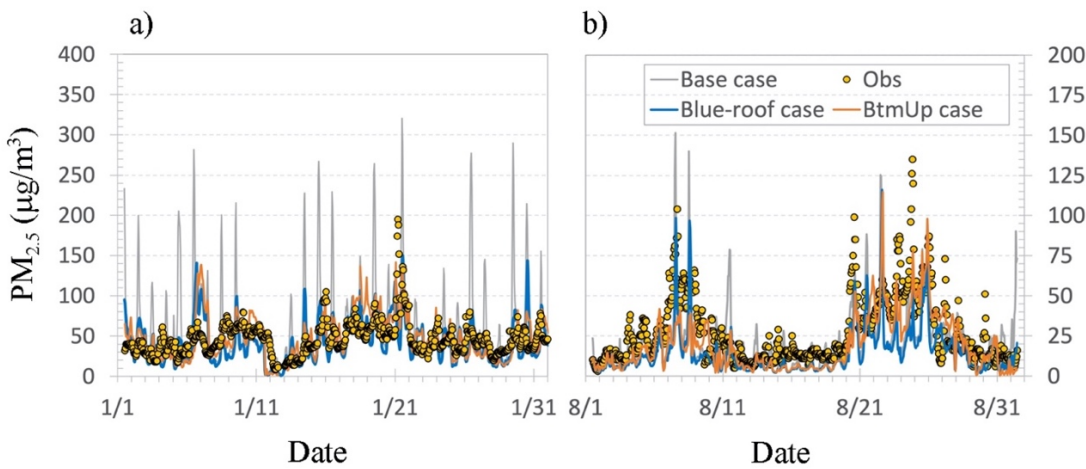


Figure 7: Time series of $PM_{2.5}$ at station CN_1379A (22°31'16.0"N 113°22'36.8"E) – Zhongshan; a) January and b) August.

Add/Changes in text

Line 247 -253 in the manuscript

“When comparing the blue-roof case with the point/area based btmUp case (Figure 5c), clear spots of $PM_{2.5}$ underestimation were observed which are shown in the square boxes of Figure 5b pointing at the northeastern and southwestern sides of PRD, and north of Guangzhou. As the focus of the study is to investigate the improvement of the blue-roof surrogate in the MIX industrial sector, rather than the performance differences between the MIX unified emissions and local bottom-up emissions. Therefore, instead of showing the uncertainty of emission inventory which is infeasible here, we have developed spatial blue-roof surrogate (Figure 4b), the comparison of the model-ready emissions (Figure 5), and the time series plots of typical stations (Figure 7) to illustrate the performance of the blue-roof algorithm.”

3) The advantage of this method is to improve the spatial distribution of emissions. However, this study only selected limited city (like Zhongshan in Figure 7) when verifying the model performance. I suggest more monitoring sites should be included to reflect the advantages of improved spatial distribution.

Response:

Thanks for the suggestion. We have added two new sections to reflect the spatial performance of the blue-roof study. The first figure (Figure 8) reflects the spatial improvement of PM_{2.5}, and the second figure (Figure 9) shows the performance of PM_{2.5} under different types of monitoring stations: a) January and b) August. For the detailed discussion, please refer to “Changes in the text” below.

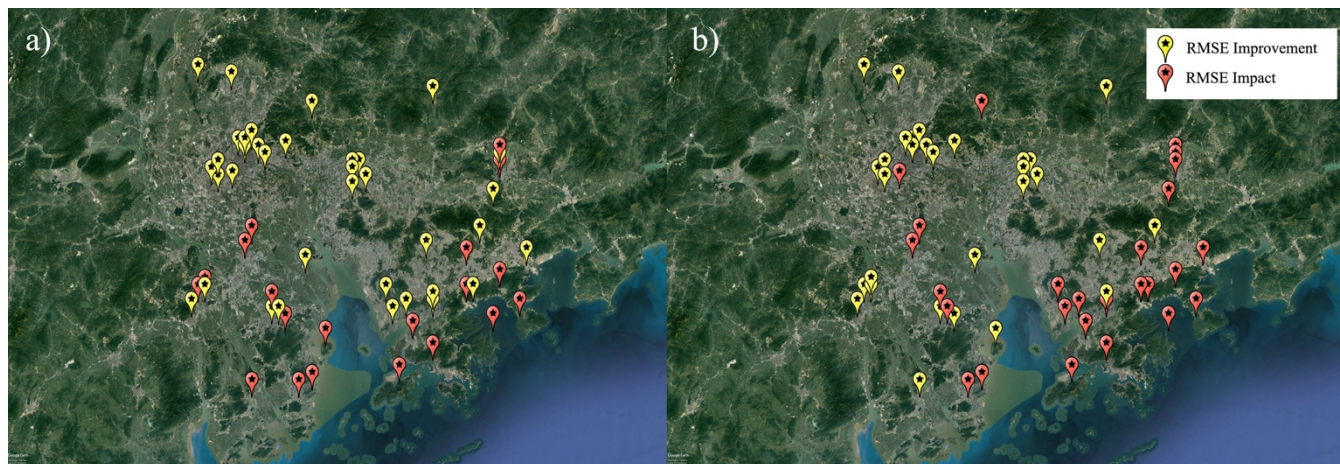


Figure 8: Spatial comparison of RMSE performance between the base case and blue-roof case: a) January and b) August. Stations with yellow colour indicates “RMSE improvement” where the RMSE of the blue-roof case is lower than the RMSE of the base case ($RMSE_{blue-roof\ case} - RMSE_{base\ case} < 0$). Stations with red colour refers to as “RMSE impact” ($RMSE_{blue-roof\ case} - RMSE_{base\ case} \geq 0$), meaning that the situation gets worse after using the blue-roof algorithm (© (Google)).

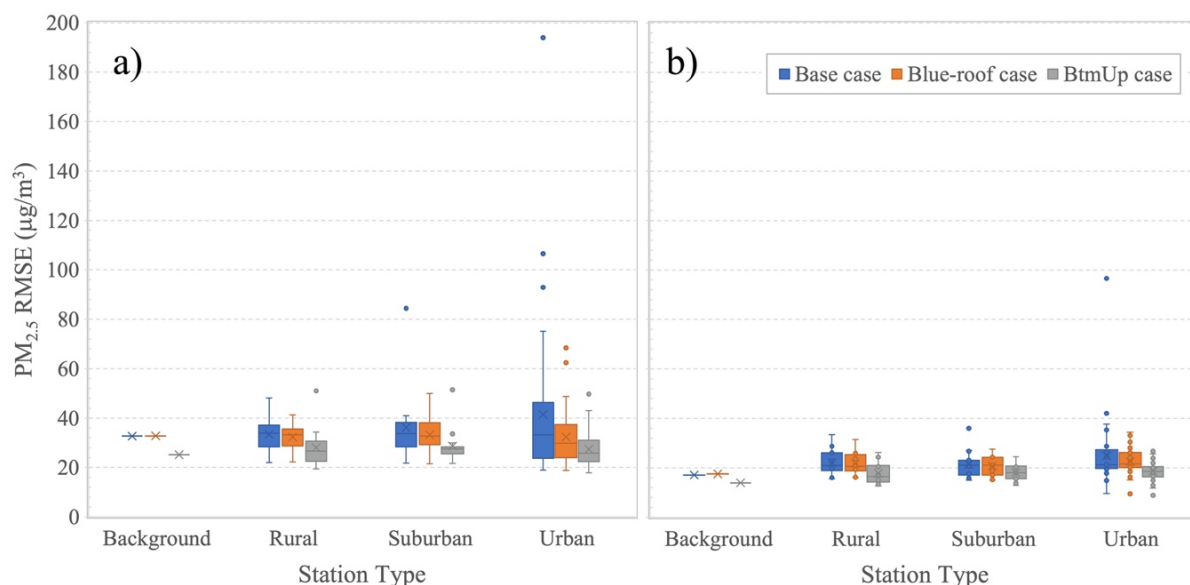


Figure 9: Performance of PM_{2.5} under different station types: a) January and b) August.

Add/Changes in text

Line 282 -296 in the manuscript

“Figure 8 shows the comparison of spatial performance between the base and blue-roof cases. The “RMSE improvement” means that the blue-roof case has outperformed the base case ($\text{RMSE}_{\text{blue-roof case}} - \text{RMSE}_{\text{base case}} < 0$), while the “RMSE impact” means that the blue-roof case has worsened the CMAQ performance ($\text{RMSE}_{\text{blue-roof case}} - \text{RMSE}_{\text{base case}} \geq 0$). In general, the majority of stations in Guangzhou, Foshan and Dongguan have received a substantial improvement in both January and August, as shown in yellow colour, while some outer stations in southern and eastern parts of PRD and Hong Kong get worse (i.e., RMSE impact) shown in red colour. These stations with the “RMSE impact” designation are primarily suburban areas where a mixed land-use pattern was identified. Overall, stations with “RMSE improvement” yield an average RMSE of $45.8 \mu\text{g}/\text{m}^3$ and $30.6 \mu\text{g}/\text{m}^3$ for the base and blue-roof cases in January, respectively, which translates to about $-12.3 \mu\text{g}/\text{m}^3$ for the RMSE improvement. This number is much larger than $+0.7 \mu\text{g}/\text{m}^3$ in magnitude obtained from the group with the “RMSE impact” designation, which illustrates the improvement has outweighed the impact. For August, the differences in $\text{RMSE}_{(\text{blue-roof case} - \text{base case})}$ under the “RMSE improvement” and “RMSE impact” are $-4.5 \mu\text{g}/\text{m}^3$ and $+0.73 \mu\text{g}/\text{m}^3$, respectively. Although there are quite a number of stations ($\sim 25+$) is fallen into the category of “RMSE impact”, their actual RMSE differences are relatively small (e.g., $\sim 75\%$ of stations with RMSE less than $1 \mu\text{g}/\text{m}^3$). Hence, it doesn’t cause any concern for the blue-roof method. Detailed statistical results for each station have been incorporated into Appendix Table S1 and S2, and the corresponding station locations are available in Appendix Figure S3.”

Line 308 -318 in the manuscript

“Figure 9 shows the $\text{PM}_{2.5}$ performance of different station types (see Appendix Figure S3). As expected, the point/area based btmUp case has the lowest RMSE among the cases for all station types, while there is a clear improvement of RMSE in urban stations in the blue-roof case; Implementing the blue-roof method has eliminated some of the extreme outliers from the base case, forming a much more narrowed RMSE range. In terms of rural and suburban stations, minor RMSE improvements (i.e., mean values) have been observed. It should be aware that the wider RMSE range showed in the blue-roof case (as compared with the base case) for the suburban category in Figure 9a is just a visual illusion. As the maximum RMSE value of the base case in the suburban category has been plotted as an outlier (dot) instead of a regular line in the upper whisker. Hence, the RMSE range (the two-end whiskers) in the blue-roof case is visually taller than the one in the base case. Appendix Figure S4 shows the station (i.e., CN_1352A) that corresponds to the maximum RMSE in the suburban category, and better performance has been obtained from the blue-roof case (blue line). In the station, the RMSE in January (August) for the base and blue-roof cases are 84.4 (36.0) $\mu\text{g}/\text{m}^3$ and 50.0 (27.5) $\mu\text{g}/\text{m}^3$, respectively.”

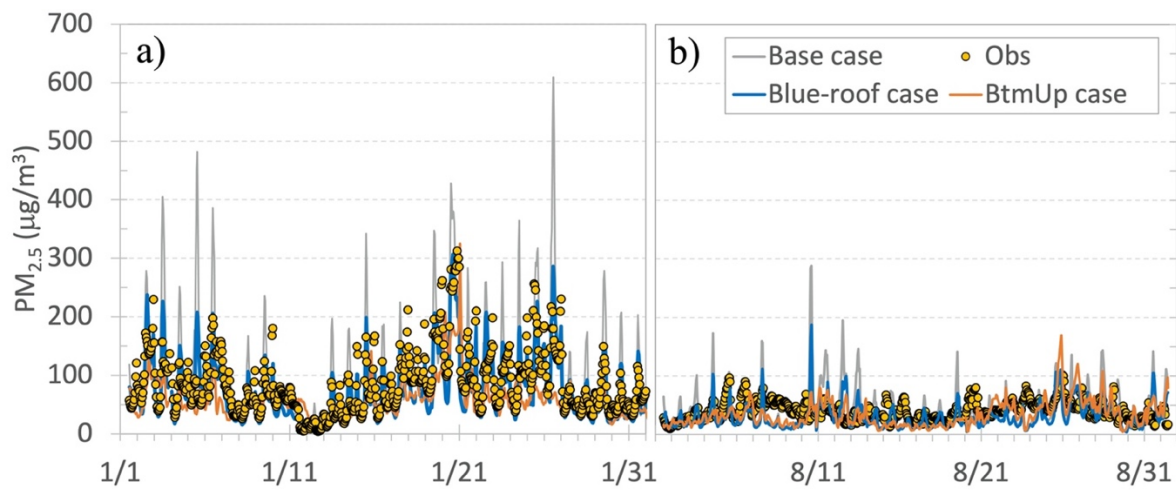


Figure S4: Time series of surface PM_{2.5} at station CN_1352A (23° 8' 26.628"N 113° 15' 57.24") – North of Guangzhou: a) January and b) August.

4) In "3.4 conclusion remarks", the following sentences seem to be unnecessary repetitions of the body content, and it is recommended to delete them.

Response:

Thanks for the suggestion. After some internal discussion with the co-authors, we think that even though the information is partial repeated (as the summary), it serves an important wrap-up for the paper. The last paragraph also provides some take-home messages and recommendations. Therefore, we would like to keep it if possible.

Add/Changes in text

Line 368 -364 in the manuscript

“The emission allocation using the blue-roof method has decluttered the urban emissions, allowing better spreading across the region. We are confident that the new method is capable of generating high-resolution input (up to 3km) for local air quality modelling and yield reasonable air quality results. Please aware that the assumption of the blue-roof method where larger blue-roof has more emissions may not always be sufficient under different resolutions. Therefore, further increasing the spatial resolution to lower than 3 km (e.g., 1 km) should be performed with cautions. Before the point/area based bottom-up approach with the unit process data is fully available in China, this method will be a useful technique for handling industrial emissions in China.”

5) There are some mistakes in the MS as well. (1) Line 49. "population density wasa applied as" should be "population density was applied as". (2) Line 134. "In thi study" should be "In this study"

Response:

Thank you so much for pointing out that. We have made the changes accordingly.

Light blue: Reviewer comments;

Black: Response to the reviewer;

Black: Changes in the manuscript;

Comment on acp-2020-1330

Referee comment on "Development of New Emission Reallocation Method for Industrial Nonpoint Source in China" by Yun Fat Lam et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-1330-RC2>, 2021

The authors developed a new method to allocate industrial emissions onto grid cells based on the areas of blue-roof buildings, which are retrieved from the satellite imagery. This new emission distribution method has been applied to the MIX inventory and evaluated through atmospheric chemistry modeling and comparison against surface observations. Overall, I feel that this study provides new insights into high-resolution emissions mapping, which deserves publication. It has been recognized that the population-based allocation method tends to overestimate anthropogenic emissions over urban areas in China, which could be improved using the method developed in this work to identify the location of blue-roof industrial buildings. **My only concern is that the manuscript lacks a detailed description of the blue roof identification algorithm**, which is difficult to understand in its current form. And the evaluation results using surface observations need further **analysis to illustrate the improvement of the spatial distribution patterns of emission inventories**. My comments are as follows.

Response:

We wish to thank the reviewer for providing the valuable comments and suggestions. It helps us to improve the paper further.

Q1), 3.1. The description of the emission allocation method should be moved to the method part in Section 2.

Response:

Thanks for pointing out that. I think we had used a wrong description for section 3.1. Now, we have renamed section 3.1 to "HSV value selection, data training, and results of blue-roof colour identification". As this section covers the results of the HSV process, we believe it should belong to section 3 results and discussion. Sorry for the confusion.

Q2), After reading this section, I am not very clear how the blue roof identification algorithm works. What do you mean by "incorporated the effect of sun position on colour change under different latitudinal position in the satellite images." (lines 175-176 on page 5)? Please explain "In this study, 4 ranges of HSV were identified for the blue roof identification algorithm. Its HSV ranges were 193-230° for H, 17% to 90% for S and 40% to 100% for V." (lines 179-180 on page 5)? The method description needs to be further improved to make it easier for an audience to understand.

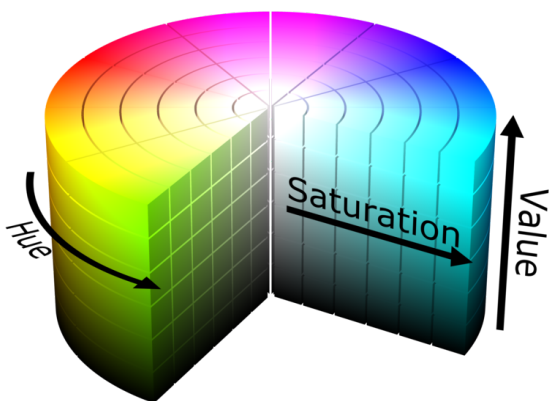
Response:

Thanks for the comments. We have written this part.

Changes in text

Line 181 -193 in the manuscript

“Three urban areas are Jing-Jin-Ji (Baoding area with 332 km²), Yangtze River Delta (Shanghai area with 1,336 km²), and GBA (Fushan area with 1,194 km²) were picked as the training dataset as we recognized that cities and regions might have their own building styles and development patterns, choosing these three regions not only allowed more diverse samples to be included in the training dataset but also incorporated the potential effect of solar incident angles on image colour (i.e., different brightness) under different latitudinal positions and time of satellite passing. To obtain the “ground truth” reference for iterative comparison, manual digitization of blue-roofs using the zoom level 16 data was performed for those three areas. The result of the iterative process shows that not a single set of HSV ranges was sufficient to capture the blue colour variation exhibited in the google images. As there was a broad spectrum of blue colours (e.g., low cyan, cyan blue, low blue) found in the satellite images, four sets of HSV ranges were used for the blue roof identification algorithm, in which each set of HSV ranges were adopted to identify an independent section of “blue colour” from the HSV solid cylinder. It should be noted that as the ranges of HSV values are considered as business confidential information under the project agreement, the exact values are not disclosed here. In general, the applied HSV values were ranged between 193° and 230° for Hue (H), 17% and 90% for Saturation (S), and 40% and 100% for Value (V).



Q3), 3.3. The evaluation of the CMAQ simulation only presents the summary of performance statistics that covers all of the surface observation stations. I am curious whether the model performance (e.g., RMSE, MB) is different among urban, rural, and remote background observation stations, which will help understand the improvement of emission distribution patterns over different regions.

Response:

Thank you for the suggestion. We have added some new analysis and discussion on the PM_{2.5} performance. Figure 8 shows the spatial pattern of PM_{2.5} performance: a) January and b) August, 2015. For the detailed discussion, please refer to “Changes in the text” below.

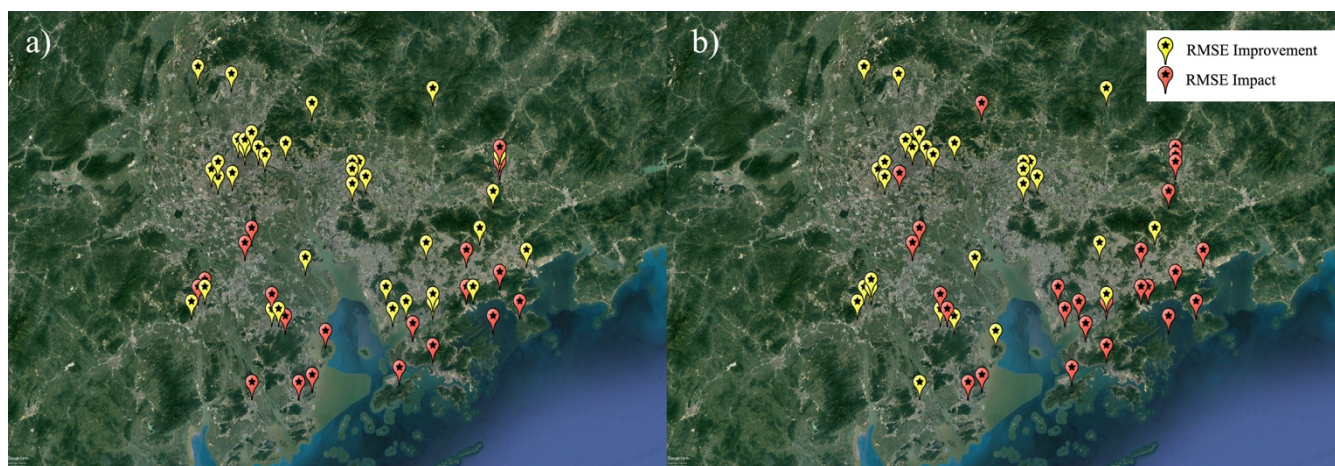


Figure 8: Spatial comparison of RMSE performance between the base case and blue-roof case: a) January and b) August. Stations with yellow colour indicates “RMSE improvement” where the RMSE of the blue-roof case is lower than the RMSE of the base case ($RMSE_{blue-roof\ case} - RMSE_{base\ case} < 0$). Stations with red colour refers to as “RMSE impact” ($RMSE_{blue-roof\ case} - RMSE_{base\ case} \geq 0$), meaning that the situation gets worse after using the blue-roof algorithm (© (Google)).

Changes in text

Line 282 -296 in the manuscript

“Figure 8 shows the comparison of spatial performance between the base and blue-roof cases. The “RMSE improvement” means that the blue-roof case has outperformed the base case ($RMSE_{blue-roof\ case} - RMSE_{base\ case} < 0$), while the “RMSE impact” means that the blue-roof case has worsened the CMAQ performance ($RMSE_{blue-roof\ case} - RMSE_{base\ case} \geq 0$). In general, the majority of stations in Guangzhou, Foshan and Dongguan have received a substantial improvement in both January and August, as shown in yellow colour, while some outer stations in southern and eastern parts of PRD and Hong Kong get worse (i.e., RMSE impact) shown in red colour. These stations with the “RMSE impact” designation are primarily suburban areas where a mixed land-use pattern was identified. Overall, stations with “RMSE improvement” yield an average RMSE of 45.8 $\mu\text{g}/\text{m}^3$ and 30.6 $\mu\text{g}/\text{m}^3$ for the base and blue-roof cases in January, respectively, which translates to about -12.3 $\mu\text{g}/\text{m}^3$ for the RMSE improvement. This number is much larger than +0.7 $\mu\text{g}/\text{m}^3$ in magnitude obtained from the group with the “RMSE impact” designation, which illustrates the improvement has outweighed the impact. For August, the differences in $RMSE_{(blue-roof\ case - base\ case)}$ under the “RMSE improvement” and “RMSE impact” are -4.5 $\mu\text{g}/\text{m}^3$ and +0.73 $\mu\text{g}/\text{m}^3$, respectively. Although there are quite a number of stations (~25+) is fallen into the category of “RMSE impact”, their actual RMSE differences are relatively small (e.g., ~75% of stations with RMSE less than 1 $\mu\text{g}/\text{m}^3$). Hence, it doesn’t cause any concern for the blue-roof method. Detailed statistical results for each station have been incorporated into Appendix Table S1 and S2, and the corresponding station locations are available in Appendix Figure S3.”

Response:

Figure 9 shows the performance of PM_{2.5} under different types of monitoring stations: a) January and b) August. Once again, detailed information. For detailed discussion, please refer to “Changes in the text” below. Please note that, as suggested by the other reviewer, CMAQ results from the point/area based bottom-up approach was also compared.

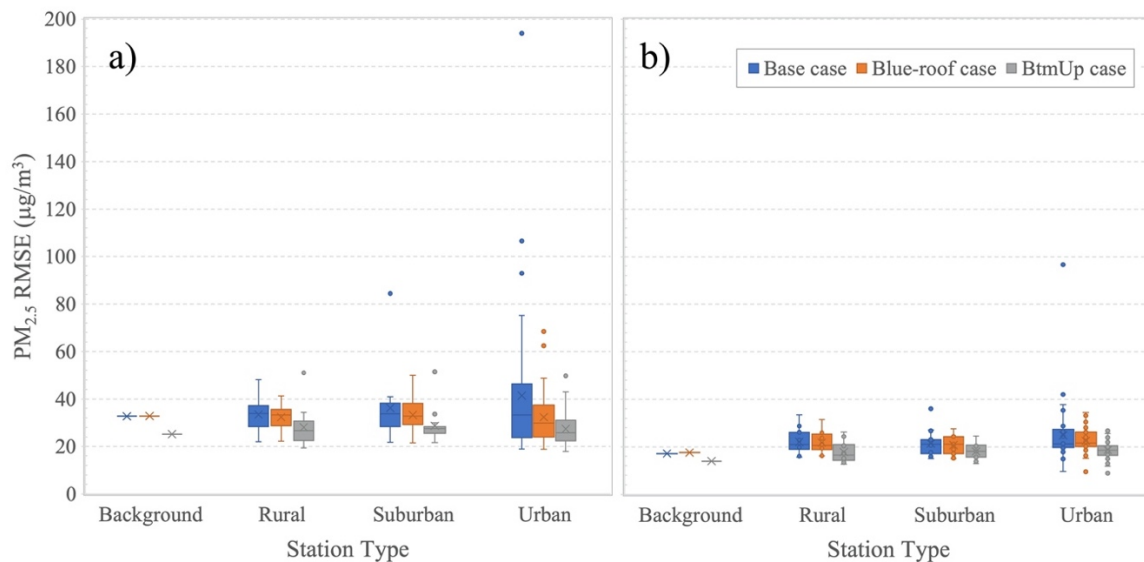


Figure 9: Performance of PM_{2.5} under different station types: a) January and b) August.

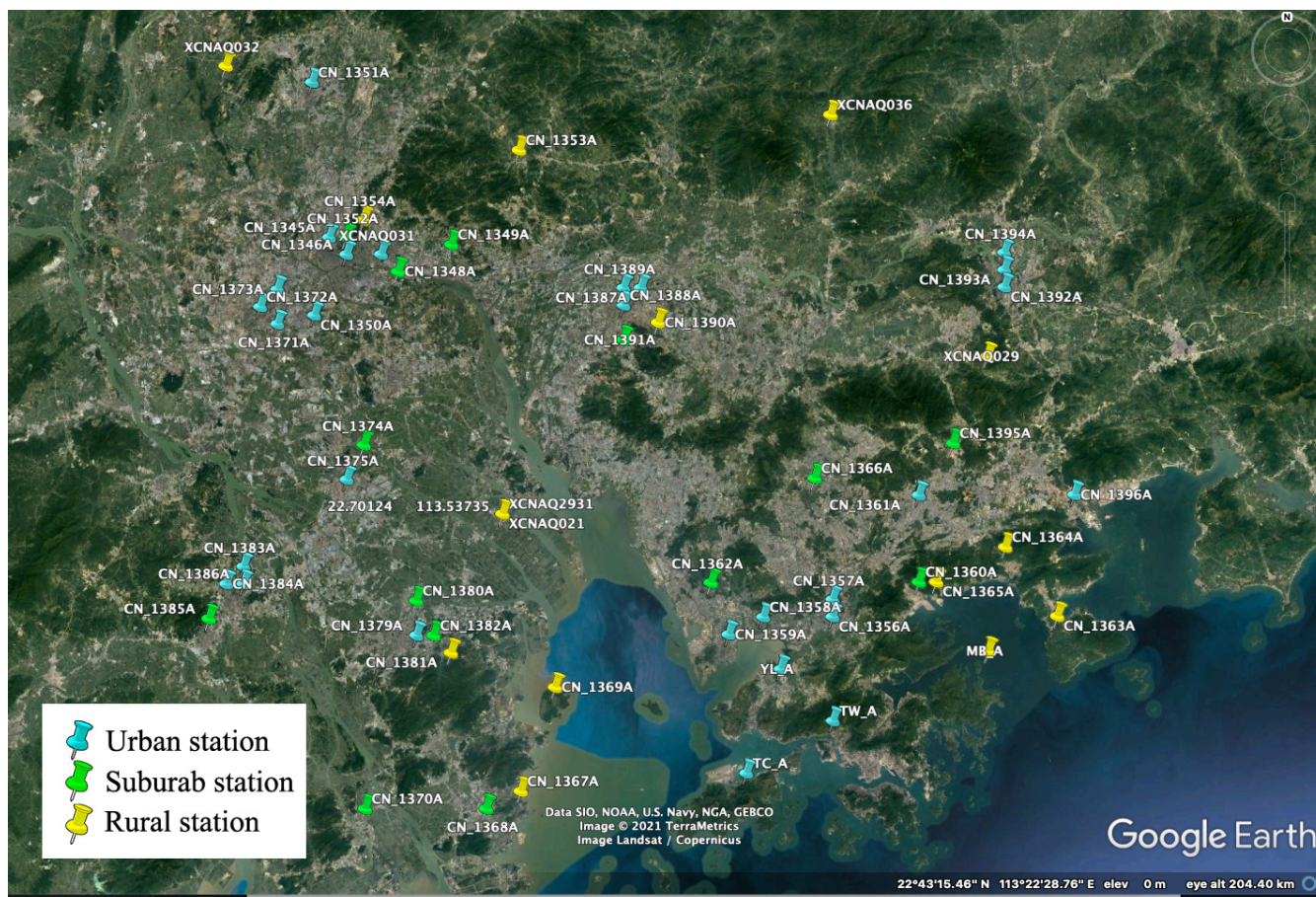


Figure S3: Satellite image with observation stations (© Google).

Changes in text

Line 312 -323 in the manuscript

“Figure 9 shows the $PM_{2.5}$ performance of different station types (see Appendix Figure S3). As expected, the point/area based btmUp case has the lowest RMSE among the cases for all station types, while there is a clear improvement of RMSE in urban stations in the blue-roof case; Implementing the blue-roof method has eliminated some of the extreme outliers from the base case, forming a much more narrowed RMSE range. In terms of rural and suburban stations, minor RMSE improvements (i.e., mean values) have been observed. It should be aware that the wider RMSE range showed in the blue-roof case (as compared with the base case) for the suburban category in Figure 9a is just a visual illusion. As the maximum RMSE value of the base case in the suburban category has been plotted as an outlier (dot) instead of a regular line in the upper whisker. Hence, the RMSE range (the two-end whiskers) in the blue-roof case is visually taller than the one in the base case. Appendix Figure S4 shows the station (i.e., CN_1352A) that corresponds to the maximum RMSE in the suburban category, and better performance has been obtained from the blue-roof case (blue line). In the station, the RMSE in January (August) for the base and blue-roof cases are 84.4 (36.0) $\mu\text{g}/\text{m}^3$ and 50.0 (27.5) $\mu\text{g}/\text{m}^3$, respectively.

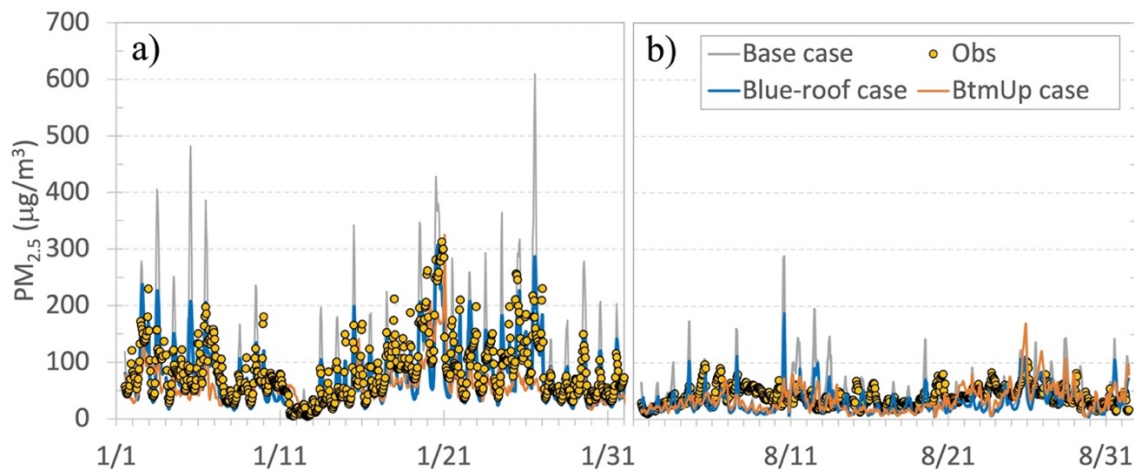


Figure S4: Time series of surface $PM_{2.5}$ at station CN_1352A (23° 8' 26.628"N 113° 15' 57.24") – North of Guangzhou: a) January and b) August.

Q4), Uncertainty assessment. Table 2 presents the False Detection Rate and False Alarm Rate in the blue roof identification algorithm. Is it possible to incorporate this information to quantify the uncertainties in the emission allocation processes?

Response:

Thanks for the question. We can't calculate the uncertainty of the emission estimate from the False Detection Rate and False Alarm Rate in the blue-roof identification from the process, as these numbers only reflect the domain-wide performance on blue colour detection. This can't be translated to use in the gridded outputs for uncertainty assessment.

Q5), Line 105, page 4. Please explain how the MIX inventory for 2010 was scaled to the target simulation year of 2015.

Response:

In this study, the MIX emission inventory was adapted to the target simulation year of 2015 (Zhang, 2020; Zhang et al. 2020). For PRD, sector-based control technologies were applied to estimate the emission totals in 2015 (Zheng et al. 2018; Li et al. 2019). The national gas monitoring data (<http://www.ipe.org.cn/MapPollution/Pollution.aspx?q=3&type=1>), ESRI 2015 population data, and OpenStreetMap traffic data together with the top-down method described in Du (2008) were then temporally and spatially interpreted into 27km (D1), 9km (D2), and 3km (D3) resolutions. The model validation of the base year 2015 was extensively discussed in our previous publication (Zhang et al. 2020). Please note that the trends of emissions from 2012 to 2016 were decreasing, in particularly on PM_{2.5}, which has been found to have a large decrease.

Changes in text

Line 110 -117 in the manuscript

“In this study, the MIX inventory was first scaled to the target simulation year of 2015 based on available sector-based control technologies (Li et al., 2019; Zhang, 2020; Zheng et al., 2018). The derived emission totals from each sector, except for the industrial emissions, were then temporally and spatially interpreted into 27km (D1), 9km (D2), and 3km (D3) resolutions using the top-down emission method described in Du (2008). Detailed methodology and validation of the base year 2015 emission inventory were extensively discussed and can be found in our previous publications (Zhang et al., 2021; Zhang et al., 2020). As Hong Kong emissions were not well presented in the MIX inventory due to the limitation of spatial resolution, the bottom-up emissions from the PATH-2016 platform were adopted for Hong Kong emissions.”

Q6), Line 108, page 4. "Hong Kong emissions were not presented in the MIX inventory" If I remember it correctly, the MIX inventory includes Hong Kong emissions.

Response:

Thanks for pointing out that. Yes, the MIX inventory includes HK emissions, but it wasn't very well presented under its native resolution (~ 27 km). We would like to say "were not well presented".

Changes in text

Line 113 -115 in the manuscript

As Hong Kong emissions were not well presented in the MIX inventory due to the limitation of spatial resolution, the bottom-up emissions from the PATH-2016 platform were adopted for Hong Kong emissions."

Q7), Lines 289 and 290, page 8. "However, we are also aware that the assumption of the blue-roof method where larger blue-roof has more emissions may not always be sufficient under different resolutions." This looks very interesting, but the manuscript has not analyzed the emission distribution patterns at different spatial resolutions, right? I would like to know how the authors reach such a conclusion.

Response:

In this study, we have assumed the area of blue-roof is directly proportional to the amount of emissions being released. In other words, the bigger factory that occupied a larger piece of land (i.e., roof-top) is assumed to produce more emissions than the small factory which occupies a small land. This is not always true. For example, we could have a small factory using very old combustion technology that produces more emissions than the big factory with advanced technology. In the situation of low-resolution (e.g., 27 km or 9 km), the emission aggregation process that produces individual gridded emissions from many blue-roof areas has damped down the effect during the averaging process. However, when the resolution reaches below 3 km (e.g., 1 km), there are not many factories being averaged. Hence, the chance of falsely allocating the emissions will be much higher.

Our original recommendation on 1 km was mainly based on the nest-down ratio of 3 to 1. So, the next logical resolution in our study is 1 km. However, for the general recommendation, we should not be based on our domain setting. Therefore, it is better to change it to "lower than 3 km", which we had already tested the 3 km in this study.

Changes in text

Line 364 -368 in the manuscript

"Please aware that the assumption of the blue-roof method where larger blue-roof has more emissions may not always be sufficient under different resolutions. Therefore, further increasing the spatial resolution to lower than 3 km (e.g., 1 km) should be performed with cautions. Before the point/area based bottom-up approach with the unit process data is fully available in China, this method will be a useful technique for handling industrial emissions in China."