Development of New Emission Reallocation Method for Industrial Nonpoint Source in China

Yun Fat LAM[1], Chi Chiu CHEUNG[2], Xuguo ZHANG[3], Joshua S. FU[4], Jimmy Chi Hung FUNG[3]

[1]Department of Geography, University of Hong Kong, HKSAR, China
5   [2]ClusterTech Limited, HKSAR, China
[3]Institute for the Environment, Hong Kong University of Science and Technology, HKSAR, China
[4]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, USA

Correspondence to: Yun Fat LAM (yunlam@hku.hk)

Abstract. An accurate emission inventory is a crucial part of air pollution management and is essential for air quality modelling.
10   One source in an emission inventory, a nonpoint source, has been known with high uncertainty. In this study, a new industrial nonpoint source (NPS) reallocation method based on blue-roof industrial buildings was developed to replace the conventional method of using population density for emission development in China. The new method utilized the zoom level 14 satellite imagery (i.e., Google®) and processed it with Hue, Saturation, Value (HSV)-based colour classification to derive new spatial surrogates for province-level reallocation, providing more realistic spatial patterns of industrial $PM_{2.5}$ and NO_2 emissions. The
15   WRF-CMAQ based PATH-2016 model system was then applied with the new NPS emissions processed emission input in the MIX inventory to simulate air quality in the Greater Bay Area (GBA) area (formerly called Pearl River Delta (PRD)). In the study, significant RMSE improvement was observed in both summer and winter scenarios in 2015 when compared with the population-based approach. The average RMSE reductions (i.e., 76 stations) of $PM_{2.5}$ and NO_2 were found to be 11 $\mu g/m^3$ and 3 ppb, respectively. This research demonstrates that the blue-roof industrial allocation method can effectively identify scattered
20   industrial sources in China and is capable of downscaling the industrial NPS emissions from regional to local levels (i.e., 27 km to 3 km resolution), overcoming the technical hurdle of ~10 km resolution from the top-down emission approach under the unified framework of emission calculation.

Keywords: Industrial Emissions, MIX Inventory, Emission Surrogate, Satellite Image Processing, and $PM_{2.5}$ air quality

1 Introduction

The emission inventory is essential for air quality management and climate studies. Various applications, including setting up regional emission reduction target, and performing numerical air quality forecasts rely upon an accurate inventory for sound assessment and judgment (Krzyzanowski, 2009; Zhao et al., 2015). As the purpose and type of emission inventories (e.g., point, mobile and area) vary largely, data requirement and collection method can be quite different (Kurokawa et al., 2013). In point source inventory, the collection of large point sources (i.e., power plant) is generally straight forward, while obtaining data from scattered industrial sources often possess challenges and requires tremendous effort to collect and process. In developed countries like the USA, industrial sources are usually large, but yet it does not always contribute to a dominant portion in the emission inventory (e.g., 10-15% in PM_{10} and 25-60% in NMVOC), and its data collection process is commonly incorporated into routine permitting exercise, making it easy to be included in their national inventory (ECCC, 2017; Janssens-Maenhout et al., 2015; Lam et al., 2004). Unfortunately, this is not the case for developing countries like China where industrial sources are considered as a major emitter; Li et al. (2017) reported that industrial sources from non-power generation are the largest contributors of PM_{10} and NMVOC in the MIX inventory. With the infinite numbers of small factories scattered across the continent with frequent change of location caused by urban redevelopment, these industrial sources are often treated as area/nonpoint source regardless of whether it is a point source (e.g., stack) or not. Hence, it possesses large spatial uncertainties in the inventory.

Recent years, various Asian emission inventories (e.g., REAS, MIX, and MICS-ASIA) have been developed for the propose of air pollution modelling, and it has been widely used for studying transboundary air pollution among the Asian countries (Chen et al., 2019; Tan et al., 2018). The top-down approach based on the unified framework of source categories, calculating method, chemical speciation scheme, and spatial and temporal allocations were commonly used in the emission inventory development, where emissions were handled separately for different source categories (e.g., power, industry, transport, residential/domestic and agriculture) with limited spatial resolutions ranged from 10 to 27 km (Kurokawa et al., 2013; Li et al., 2017; Ohara et al., 2007; Streets et al., 2003). In some cases, higher resolution emission inputs were achieved via GIS spatial interpolation for subregional (i.e., 10-15 km resolution) application. Information such as stack location, road network and population density wasa applied as surrogate data for spatial reallocation (Du, 2008). For the case when ultra-high resolution (i.e., 3-5 km resolution) was needed, the top-down emission approach was often supplemented with the bottom-up approach using the detailed activity data (i.e., exact emission locations and its relevant emission amounts) in the emission inventory development (HKEPD, 2011). For the category of small/medium industrial sources where location information was frequently missing in the top-down approach, the population density was used as the surrogate, giving the fact that population density was considered a good proxy for accessing employment, goods and services (Giuliano and Small, 1993). Historically, this approach seemed to be quite robust to capture the factory location due to: 1) the Danwei/socialist work units which enforced jobs and residences to be closed to each other to reduce travel distance (Yang, 2006), and 2) factory jobs were typically included accommodation (i.e., dormitory) for attracting foreign workers. With limited transport infrastructure, dormitories were usually within a few kilometres away from the factories. However, in recent years, the land-use and housing reforms in China has led to a spatial separation of jobs and residences, the strong factory-residence pattern has slowly diminished in Chinese cities as efficient public transportation has emerged. With the adaptation of industrial park in the urban renewal process, it further expedited the separation of industrial-related employment (in the outskirt of the city) from residential space (city centre) (Zhao et al., 2017). As a result, there is a need to reconsider how industrial emissions are handled in the top-down approach, searching for a suitable surrogate for the emission reallocation.

In this study, the concept of a blue-roof industrial surrogate was introduced for the first time for Chinese industrial emission allocation. The approach assumed the majority of industrial buildings (both factories and its warehouses) in China were single-story non-concrete buildings with their rooftop was made out of galvanized metal coated with blue epoxy. In this

development, satellite imagery with zoom level 14 (i.e., Google®) was adopted and processed with HSV-based colour classification for generating the province-level spatial surrogate. The Community Multi-Scale Air Quality (CMAQ) based PATH-2016 platform with 3km MIX inventory was then applied to evaluate the impacts of air quality predictions between population and blue-roof based methods, and the simulated results of PM$_{2.5}$, NO$_2$, and O$_3$ were then compared to local observation data to assess its model performance (HKEPD, 2011; Li et al., 2017).

2   Methodology

A new allocation method called "blue-roof" industrial allocation was introduced in the top-down emission approach for better allocating the scattered/nonpoint source (NPS) industrial emissions in China. In this study, the CMAQ model and the regional Asian emission inventory-MIX were applied to evaluate the effectiveness of the new allocation method on the performance of air quality prediction. The target simulation year is 2015, and the details of each component are described below:

2.1   Study area, simulation domain, and observation network

The Guangdong-Hong Kong-Macao Greater Bay Area (GBA), also known as the Pearl River Delta (PRD), was adopted as the study area for the NPS industrial allocation test. The characteristic of diverse industrial clusters (e.g., garment, electronics, and plastic factories) scattered across the area creates an ideal testbed for spatial examination. The GBA area consists of two special administrative regions (Hong Kong and Macao) and nine Chinese municipalities, including Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen, and Zhaoqing in Guangdong Province with a total area coverage over 56,000 km^2. It is classified as one of the world-class manufacturing hubs in China. In this study, the CMAQ based PATH-2016 was adopted to evaluate the influence of air quality prediction from the new allocation method. The PATH-2016 modelling platform consists of 4 nested domains including East Asia and Southeast Asia (D1), Southeastern China (D2), and GBA)/PRD (D3), and Hong Kong (D4) with resolutions of 27 km, 9 km, 3 km and 1 km, respectively. For this study, only D1-D3 was applied as it already covered the entire GBA with a reasonable spatial resolution (e.g., 3 km) for regional air quality simulation, as shown in Figure 1. Details of PATH-2016 and its model setting are discussed in the later section. For evaluating the performance of CMAQ air quality prediction, the China National Environmental Monitoring Centre (CNEMC) and the Guangdong-Hong Kong-Macao Pearl River Delta Regional Air Quality Monitoring Network (HKEPD, 2016) with over 76 surface observation stations were adopted (available at http://www.cnemc.cn/, last access: 10 September 2020). These stations measure various air pollutants, including PM$_{2.5}$, PM$_{10}$, SO$_2$, NO$_2$, and O$_3$.

2.2   PATH-2016 and MIX inventory

The PATH-2016 is a WRF-CMAQ (Community Multi-Scale Air Quality model) Air Quality system used by the HKSAR government for air quality-related policy. It has been validated in several studies (HKEPD, 2011, 2019; Zhang, 2020). In this study, CMAQ version 5.0.2 with AERO5 aerosol module and CB05CL carbon bond chemical mechanism driven by WRF version 3.7.1. was adopted for air quality simulation. The model setup for CMAQ simulation is summarized in Table 1, and WRF meteorological validation can be found in Zhang (2020) and HKEPD (2019). The initial/boundary conditions for the outermost domain, D1 was generated from the global model GEOS-Chem outputs for Asian pollution background (Lam and Fu, 2010). In terms of the model emissions, the majority of the anthropogenic emissions were adopted from the Asian emission inventory, MIX. The MIX is a regional emission inventory developed to support the Model Inter-Comparison Study for Asia (MICS-Asia) and the Task Force on Hemispheric Transport of Air Pollution (TF HTAP) (Li et al., 2017). It consists of 5 anthropogenic emission source categories, including point, industry, transport, residential, and agriculture (NH$_3$ only) with

105   a resolution of 0.25° (~27km), and its emission base year is for 2010. In this study, the MIX inventory was first scaled to the target simulation year of 2015 (Zhang, 2020). After that, the sectoral emissions, except for the industrial emissions, were then temporally and spatially interpreted into 27km (D1), 9km (D2), and 3km (D3) resolutions using the top-down method described in Du (2008). As Hong Kong emissions were not presented in the MIX inventory, the bottom-up emissions from PATH-2016 platform were applied to fill in the missing pieces. For the remaining sectors that were not available from the MIX inventory,

110   it was adopted from the PATH-2016 study (Zhang, 2020). These include MEGAN biogenic, GFED biomass burning, Marine and sea-salt emissions (Athanasopoulou et al., 2008; Giglio et al., 2013; HKEPD, 2019; Ng et al., 2012).

2.3   Case study for nonpoint source industrial allocation

        The purpose of CMAQ simulation is to evaluate the performance of the new industrial allocation method on its effect on air quality prediction in China. Two CMAQ scenarios tailored for the industrial allocation methods were proposed and

115   tested with the MIX inventory. These scenarios are 1) population-based method and 2) blue-roof based method. Details of each method are described in the later section. For each scenario, two months of CMAQ air quality simulation were performed. The selection of winter (January) and summer (August) months from 2015 was to allow better reflection of air quality impacts from the change of Asian monsoon in Southern China. The choice of using 2015 as the based year was to permit more local observation data to be available in the GBA area from the Chinese national observation network (operated after late 2013),

120   moreover, to better fit with the modelling effort in 2016 Air Quality Objectives (AQOs) review that has also applied PATH-2016 model (Zhang, 2020).

2.3.1   Base Case - population-based method

The population-based method (hereafter referred to as "base case") is commonly applied in the top-down emission inventory to allocate residential (area) or NPS industrial sector in the regional emission inventory (Du, 2008; Li et al., 2017; Ohara et al.,

125   2007). It utilizes population or population density as a spatial proxy to distribute the sectorial emissions into the simulation grids. It assumes a strong association is present between population and industrial emissions. In this study, the Oak Ridge National Laboratory (ORNL)'s LandScan global population data with the resolution of ~1 km (30″ X 30″) gridded spatial resolution was applied to allocate NPS industrial emissions. To allow the separation of urban population from the rural population in LandScan grids, a threshold value of 1,500 people per square kilometres was adopted, in which any grid values

130   that were greater than this number were considered as urban grids (Liu et al., 2003). The basic equation for estimating the gridded emissions ($E_{m,n}$) using urban population is shown in Eq (1). E_m is the total emission in a province (m), and U_pop_n /U_pop_m is the ratio of urban population from a grid (n) to the province total. This value is commonly referred to as the spatial allocation factor, and it is a dimensionless value ranged between 0 to 1. The collection of spatial allocation factors in the gridded matrix is called a spatial surrogate. In thi study, all the calculations and the spatial interpretation were performed in

135   ArcGIS to yield CMAQ-ready emissions.

$$E_{m,n} = E_m \times \frac{U\_pop_n}{U\_pop_m},$$
(1)

where $E_{m,n}$ is the emission in n^{th} grid for m^{th} province, E_m is the total emission in m^{th} province, U_pop_n is the urban population count in n^{th} grid, U_pop_m is the total population in m^{th} province.

2.3.2   Blue-roof Case: blue-roof based method

140   The "blue-roof based method" (hereafter referred to as "blue-roof case") adopted the concept of rooftop colour for associating the location of NPS industrial buildings in the top-down emission inventory. In China, warehouse and industrial rooftops are commonly made out of galvanized metals with a coat of colour epoxy (i.e., light blue, green, pink, and purple).

As more than 90% (by general observation) of these industrial roofs are in light blue, this unique feature was captured and applied to develop the new allocation method. To derive the allocation factor for each grid for the modelling domain, satellite

145   imagery was used. Among different imagery products (e.g., Google®, Bing®, Baidu®, Latsat8, and SPOT7), the Google imagery was selected for the basis of the analysis, as it provided their products with less processing effort for different zoom levels. Overall, the zoom level 14 or above (~ 9.5 metre/pixel or above) has been confirmed to be sufficient for use in the colour detection process for major industrial rooftops in China. Considering the study required to process a relatively large area, the lowest possible zoom level (i.e., 14) was preferred. In this development, the QGIS platform v2.16 with two plugins

150   (i.e., OpenLayers and Python) was adopted to provide a smooth process of overlaying satellite imagery with the Google Maps API v3 and to perform a pixel-based HSV colour detection (i.e., OpenCV) in QGIS. The choice of using OpenLayers (OL) at that time was to avoid the violation of the Terms Of Service (TOS) regarding the direct usage of Tile Map Services. As it for now, this type of operation is no longer allowed under the updated TOS by Google®. To perform a similar process, one may choose to use Google Earth Engine or Bing/Baidu Maps API with OL. Figure 2 shows the basic flow chart of the process.

155   Overall, about 2,000 map tiles that contained "blue-roof" were processed for the area of the D3 domain. At last, the identified blue-roofs were then converted into polygons and stored into a shapefile. As the HSV algorithm was unable to distinguish the blue ocean and river features from blue-roofs, a removal process using the shapefiles of coastal line and inland waterbody was applied to eliminate the falsely identified waterbody using ArcGIS. The resulted blue-roof shapefile was then spatially interpreted with China administrative (i.e., province) boundaries and CMAQ raster grids to yield the information of gridded

160   blue-roof areas (B_area_n) and total blue-roof areas for each province (B_area_m). This information was further applied to province total emission (E_m) to calculate the gridded emissions ($E_{m,n}$) using Eq (2):

$$E_{m,n} = E_m \times \frac{B\_area_n}{B\_area_m}, \qquad\qquad\qquad (2)$$

where $E_{m,n}$ is the emission in n^{th} grid for m^{th} province; E_m is the total emission in m^{th} province; B_area_n is the total blue-roof area in n^{th} grid; B_area_m is the total blue-roof area in m^{th} province.

165   3   Results and discussion

3.1   Development of blue-roof allocation method

The satellite imagery of Google Map Tiles with zoom level 14 (which was retrieved between 2015-16 for this study) had exhibited a colour variation due to the inconsistent environmental conditions (e.g., cloud cover, visibility, and brightness of the day) when the images were taken. As these collective images were taken from different seasons or years, the aggregated

170   images might not reflect a single snapshot of a specific time. To determine suitable parameters for the HSV algorithm, an optimization process that iteratively searches for high hit rates, low false detection and false alarm rates (See Supplement Eq(S1-S3) for definition) was applied. 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

175   allowed more diverse samples to be included in the training dataset but also incorporated the effect of sun position on colour change under different latitudinal position in the satellite images. 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 range of HSV was sufficient to capture the blue colour variation exhibited in the google images. In this study, 4 ranges of HSV were identified for the blue roof identification algorithm. Its HSV ranges were

180   193-230° for H, 17% to 90% for S and 40% to 100% for V. Figure 3a-c shows samples of training images (100 km²) in Baoding, Shanghai, and GBA, and Table 2 shows the summary of the training performance.

Overall, 74% to 88% (hit rates) of the blue roof areas were successfully identified by the algorithm, while the false detection rates and false alarm rates were ranged between 35% to 51% and 0.1% to 0.5%, respectively. The low percentages of false alarm rates indicate only small amount of non-blue-roof areas were included by the algorithm. For a closer look at the

185 results of false detection rates, it reveals that the false identification was mainly concentrated around the building boundaries because of the fuzziness of the building edges from the zoom level 14 satellite images. The percentages of false detection rates varied for images in different areas as the clearness of satellite images depending on the atmospheric conditions (e.g. cloudiness, air pollution, etc.) at the time they were taken. Comparing with the images of Shanghai and GBA, the Baoding image has a relatively higher degree of blurriness which explained why the false detection rate for the Baoding was higher than those in

190 the other two training areas. While the false selection around the building edges may incur different levels of errors to the blue-roof identification result, however, it does not generally affect the spatial distribution of the blue roof areas selected by the algorithm. It is observed that the GBA area has low hit rate (i.e., 74%) due to more scattered/isolated development than the other areas. In Boading and Shanghai, industrial parks are more common than in GBA. Buildings clustered together might have contribute to higher hit rates, but at the same time caused high false detection rates, as the gap between buildings was

195 also included as blue-roof. To better evaluate the algorithm response to different environmental condition, another 7 areas (approximately 100 km² each) covering a wide variety of geographical locations and features were selected to validate the blue-roof identification algorithm, as shown in Figure 3d-j. The results in Table 2 show that the algorithm achieved between 76-92%, 9-54%, and 0-0.3% for the hit rates, false detection and false alarm rates, respectively. These similar results found in Nagqu, Baoji, Kaifeng, Xi'an, and Xhengzhou indicate that the algorithm is relatively stable across the continent of China. For

200 the remote areas in Taklimakan Desert and Yunnan (Figure 3i and 3j), it should be noted that the "ground truth" blue-roof areas were zero, so the hit rates were inapplicable for these two test images.

3.2   Blue-roof allocation process and CMAQ ready emission

Large spatial data with various geospatial information (e.g., province shapefile) was processed through ArcGIS to create the gridded emissions for the CMAQ model. As mentioned, the selected blue-roof areas from the HSValgorithm have first

205 undergone a spatial operation for removing the falsely identified water bodies from the blue-roof dataset. The data was then used to compute the gridded total blue-roof areas in each grid (See Figure 4a). Furthermore, the total blue-roof areas in each province (i.e., Guangdong, Guangxi, and Jiangxi) within the domain were also generated. Figure 4b shows the province spatial surrogate (i.e., $\frac{B\_area_n}{B\_area_m}$ in eq (2)) that was created for the GBA emission allocation. The values in each grid in the surrogate file should fall between 0 to 1, and the total in the province should sum up to 1.0 for data integrity. The yellow grids (values

210 greater than 0) were clustered around the centre of PRD, reflecting the high density of blue-roof buildings were identified along the pearl rivers, and the yellow lines extended from PRD indicates that small remoted industrial areas were built along the major highways in GBA. For the grids with magenta (values with 0), it is confirmed to be hilly areas, forests, or water reservoirs, no blue-roof building was identified.

At last, for generating the final gridded industrial emissions, the spatial surrogate was applied to the MIX industrial

215 emissions to spatially allocate the province-level emissions into the CMAQ grids. All sectoral emissions, including power, transportation, industrial, residential, agriculture and others were aggregated together with the newly produced industrial emissions to generate the CMAQ ready gridded emissions for air quality simulation. Figure 5 shows the daily column total of CMAQ ready $PM_{2.5}$ emissions (January 1, 2015) for both base case and blue-roof case. In general, more spatial spreading is observed in the blue-roof case than in the base case within the GBA area. In the base case (Figure 5a), emission (over 200

220 g/s/grid of $PM_{2.5}$) is intensely clustered around the city centres of Guangzhou (GZ) and Foshan (FZ). A circular belt of intense $PM_{2.5}$ emission is observed along the coast of Zhujiang River Estuary in Shenzhen (SZ). In contrast, $PM_{2.5}$ emission in the blue-roof case (Figure 5b) is more widely spread across the region, with additional focuses in Dongguan (DG) and north of

Zhongshan (ZS). The hotspots of PM$_{2.5}$ exhibited in (Figure 5b) are strongly aligned with the spatial pattern of hotspots from Cui et al. (2015) which was generated from source apportionment method. In Shenzhen, the circular belt of high emission

225   previously observed in the base case has been disappeared. Further investigation shows that the coastal area along the Zhujiang River Estuary has already been converted into recreation and residential areas. Due to a high population density found in the area, the base case had allocated a large amount of PM$_{2.5}$ emission to the area. For small industrial areas, the blue-roof case also seems to outperform the base case as it has identified more scattered industrial areas in the region. As shown in BE (Figure 5), the blue-roof method is capable of capturing the small industrial towns (See supplementary Figure S1) along the major

230   highways. For this particular example, the industrial area captured by the blue-roof method is located in the rural area of Qingyuan with over 20 petrochemical factories or warehouses, which has been missed in the base case.

3.3   CMAQ simulated Air Quality and statistical comparison

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 NPS industrial emissions. Figure 6 shows the simulated monthly average surface PM$_{2.5}$ for

235   base case (a, c), and blue-roof case (b, d); the top (a-b) and bottom (c-d) panels represent the January and August cases, respectively. As expected, the base case (left panel) has much lower spatial spreading when comparing with the blue-roof case (right panel) illustrated in the earlier section. The wider spreading of PM$_{2.5}$ in the blue-roof case (right panel) was attributed to the redistribution of industrial emissions from highly populated areas (i.e., Guangzhou, Foshan and Shenzhen) found in the base case into other areas of GBA. The redistribution process has lowered the CMAQ prediction for those three urban areas,

240   moreover, to reduce the PM$_{2.5}$ prediction along the coast of Zhujiang River Estuary at the circular belt of Shenzhen, which was mentioned in Figure 5a). As monsoon wind runs differently in summer and winter, it affects the regional pollutant transport and air quality prediction in the GBA area. For the winter case, with the effect of the northeast prevailing wind in January, the reduction of Guangzhou and Foshan emission/pollution (See Supplement Fig. S2) has a strong positive impact on both local and downwind regions (i.e., south of Guangzhou - Zhongshan, Zhuhai, and Macau). This is illustrated by the PM$_{2.5}$ time-series

245   plot of Zhongshan station (22°31'16.0"N 113°22'36.8"E) shown in Figure 7a) where the large PM$_{25}$ overestimation in the base case (grey line) is significantly reduced into the more acceptable range shown in the blue-roof case (blue line). The Root Mean Square Error (RMSE) is trimmed down nearly in half to about 23.0 μg/m^3 (from 49.2 μg/m^3 in the base case to 26.2 μg/m^3 in blue-roof case), demonstrating the effectiveness of using the blue-roof allocation method in the top-down emission approach. This is not entirely the same case for summer (August), when the prevailing wind is from the southwest that brings clean

250   marine boundary to the region. Although, in some situations, due to the presence of distant Typhoon (e.g., Soudelor (August 4-11) and Goni (August 21-25)), the outermost of Typhoon circulation had forced the wind direction changed to northeasterly, and resulted in a similar transport phenomenon that causes the PM$_{2.5}$ spikes in summer (Lam et al., 2018). For the general situation during the non-typhoon condition, stronger PM$_{2.5}$ underestimation is observed in the blue-roof case than in the base case, worsening the Mean Bias (MB) from -7.0 in the base case to -12.4 μg/m^3 in the blue-roof case. In terms of RMSE, there

255   is nearly no difference between the base case (i.e., 20.47 μg/m^3) and blue-roof case (i.e., 20.48 μg/m^3), as the performance degradation observed during the non-typhoon period was compensated by the improvement during the Typhoon period.

To better understand the overall impacts on local air quality prediction, Table 3 shows the comparison of performance statistics between the base case and blue-roof case for PM$_{2.5}$, NO$_2$, and O$_3$. In general, all three pollutants received some improvements when using the blue-roof allocation method. Among them, a more significant improvement of RMSE is

260   observed in PM$_{2.5}$ and NO$_2$, which ranges from 3.3-11.5 μg/m^3 for PM$_{2.5}$ and 2.3-2.7 ppb for NO$_2$. The result is somewhat expected as NPS industrial sector is the largest contributor of PM$_{2.5}$ and NO$_2$ emissions in the MIX inventory. In terms of MB, slight degradation is observed in PM$_{2.5}$, which may either be caused by the slight underestimation of total PM$_{2.5}$ emission in GBA or insufficient generation of secondary organic PM$_{2.5}$ from CMAQ, which is commonly observed in version 5.02. For

NO$_2$, slight improvement is observed, which is resulted from the removal of large overestimations in the city centre of GZ, FZ, and SZ. Among 76 observation stations, on average, 25 stations received an improvement in RMSE for PM$_{2.5}$ and NO$_2$. The largest RMSE improvement was observed in the Foshan area with -151 µg/m^3 improvement in PM$_{2.5}$ (See Supplement Fig. S2) and 33 ppb in NO$_2$. This result clearly reflects the weakness and limitation of the population-based method for industrial allocation in the fine-resolution grid. In some stations (i.e., 7), higher RMSE (i.e., an average of 1.8 µg/m^3 for PM$_{2.5}$, and 1.9 ppb for NO$_2$) are observed. For ozone, minor improvement (i.e., 0.2 ppb and 0.6 ppb in RMSE) has been found which may attribute to the improvement in NO$_2$ prediction and consequently affect the NOx titration process in ozone chemistry. As the improvement is at a marginal level, it is concluded the improvement is limited.

3.4 Conclusion remarks

In this work, we developed a new method called the "blue-roof allocation method" for assigning NPS industrial emissions for the gridded air quality simulation. The proposed method not only provides an alternative way of handling Chinese industrial emissions from the existing population-based method in the top-down emission inventory but also allows a higher resolution (up to 3 km) can be generated for local air quality study. As the rapid urban redevelopment and mature public transportation network (e.g., metro/train system) were emerging in China, the relationship of proximity between living place and the workplace was slowly diminished. Hence, the traditional method using population density as a spatial proxy for industrial emissions has become obsolete.

In the blue-roof allocation method, satellite images from zoom level 14 were applied as the basis for blue-roof extraction. An HSV colour detection algorithm was developed and trained to carry out blue-roof identification. The captured blue-roofs were then converted and stored as individual polygons for further process. The sequence of ArcGIS subprocesses was applied to generate spatial surrogate and gridded emissions. The gridded emissions were tested with CMAQ air quality simulation for January and August of 2015. The results show that large improvements are observed on both PM$_{2.5}$ and NO$_2$ predictions when compared with the traditional method (i.e., population-based method). By using the blue-roof method, not only reduced the emission errors from large metropolitan areas but also effectively captured the scattered industrial areas located in the rural area. 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. 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. Therefore, further increasing the spatial resolution to 1 km should be performed with cautions. Before the bottom-up approach with process-based data is fully available in China, this method will be a useful technique for handling NPS industrial emissions in China.

4 Code/Data availability

The simulated output is available from the corresponding author on reasonable request.

5 Author Contribution

The main contributions to model simulation and data management were made by YFL, CCC and XZ; to the data analyses by YFL, CCC and XZ; conceptualization by YFL and JSF; writing -original draft preparation by all co-authors; resources by JCF.

300 **6   Competing interests**

The authors declare that they have no conflict of interest.

7   Acknowledgment

8   References

Athanasopoulou, E., Tombrou, M., Pandis, S. N., and Russell, A. G.: The role of sea-salt emissions and heterogeneous
310   chemistry in the air quality of polluted coastal areas, Atmos. Chem. Phys., 8, 5755-5769, 10.5194/acp-8-5755-2008, 2008.

Chen, L., Gao, Y., Zhang, M., Fu, J. S., Zhu, J., Liao, H., Li, J., Huang, K., Ge, B., Wang, X., Lam, Y. F., Lin, C. Y., Itahashi, S., Nagashima, T., Kajino, M., Yamaji, K., Wang, Z., and Kurokawa, J.: MICS-Asia III: multi-model comparison and evaluation of aerosol over East Asia, Atmos. Chem. Phys., 19, 11911-11937, 10.5194/acp-19-11911-
315   2019, 2019.

Cui, H., Chen, W., Dai, W., Liu, H., Wang, X., and He, K.: Source apportionment of PM2.5 in Guangzhou combining observation data analysis and chemical transport model simulation, Atmospheric Environment, 116, 262-271, https://doi.org/10.1016/j.atmosenv.2015.06.054, 2015.

Du, Y.: New Consolidation of Emission and Processing for Air Quality Modeling Assessment in Asia, 2008.

320   ECCC: Canada–United States Air Quality Agreement : progress report 2016, Environment and Climate Change Canada, 2017.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317-328, 10.1002/jgrg.20042, 2013.

325   Giuliano, G., and Small, K. A.: Is the Journey to Work Explained by Urban Structure?, Urban Studies, 30, 1485-1500, 10.1080/00420989320081461, 1993.

Google: Map showing location of Hong Kong with CoWIN stations, earth.google.com/web/, n.d.

HKEPD: Regional Air Quality Modelling System (PATH) – Feasibility Study, ENVIRON Hong Kong Limited, 2011.

HKEPD: Guangdong-Hong Kong-Macao Pearl River Delta Regional Air Quality Monitoring Network: A Report of
330   Monitoring Results in 2015, Guangdong Provincial Environmental Monitoring Centre; Environmental Protection Department, Hong Kong SARG; Environmental Protection Bureau, Macao SARG; Meteorological and Geophysical Bureau, Macao SARG, 2016.

HKEPD: Air Quality Objectives (AQO) Review Working Group - Annex C Air Science and Health Sub-group, Hong Kong Environmental Protection Department, 2019.

335   Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411-11432, 10.5194/acp-15-11411-2015, 2015.

Krzyzanowski, J.: The Importance of Policy in Emissions Inventory Accuracy— A Lesson from British Columbia, Canada,
340   Journal of the Air & Waste Management Association, 59, 430-439, 10.3155/1047-3289.59.4.430, 2009.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Greet, J. M., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia

Atmospheric
Chemistry
and Physics
Discussions

(REAS) version 2, Atmospheric chemistry and physics discussions, 13, 10049-10123, 10.5194/acpd-13-10049-2013, 2013.

345 Lam, Y.-F., Davis, W. T., Miller, T. L., and Fu, J. S.: System enhancement of title V permit reviews and point source inventories, Proceedings of the Air and Waste Management Association's Annual Meeting and Exhibition, 1673-1679, 2004.

Lam, Y. F., and Fu, J. S.: Corrigendum to "A novel downscaling technique for the linkage of global and regional air quality modeling" published in Atmos. Chem. Phys., 9, 9169–9185, 2009, Atmos. Chem. Phys., 10, 4013-4031, 10.5194/acp-350 10-4013-2010, 2010.

Lam, Y. F., Cheung, H. M., and Ying, C. C.: Impact of tropical cyclone track change on regional air quality, Sci Total Environ, 610-611, 1347-1355, 10.1016/j.scitotenv.2017.08.100, 2018.

Li, M., Zhang, Q., Kurokawa, J.-i., Jung-Hun, W., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian 355 anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmospheric Chemistry and Physics, 17, 935-963, http://dx.doi.org/10.5194/acp-17-935-2017, 2017.

Liu, S. H., Li, X. B., and Zhang, M.: Scenario Analysis on Urbanization and Rural - Urban Migration in China, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing., 2003.

Ng, S. K. W., Lin, C., Chan, J. W. M., Yip, A. C. K., Lau, A. K. H., and Fung, J. C. H.: Study on Marine Vessels Emission 360 Inventory, The Environmental Protection Department The HKSAR Government, 2012.

Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T.: An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, Atmos. Chem. Phys., 7, 4419-4444, 10.5194/acp-7-4419-2007, 2007.

Streets, D. G., Bond, T. C., Carmichael, G. R., Fernandes, S. D., Fu, Q., He, D., Klimont, Z., Nelson, S. M., Tsai, N. Y., 365 Wang, M. Q., Woo, J.-H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, Journal of Geophysical Research: Atmospheres, 108, 10.1029/2002jd003093, 2003.

Tan, J., Fu, J. S., Dentener, F., Sun, J., Emmons, L., Tilmes, S., Sudo, K., Flemming, J., Jonson, J. E., Gravel, S., Bian, H., Davila, Y., Henze, D. K., Lund, M. T., Kucsera, T., Takemura, T., and Keating, T.: Multi-model study of HTAP II on sulfur and nitrogen deposition, Atmos. Chem. Phys., 18, 6847-6866, 10.5194/acp-18-6847-2018, 2018.

370 Yang, J.: Transportation Implications of Land Development in a Transitional Economy:Evidence from Housing Relocation in Beijing, Transportation Research Record, 1954, 7-14, 10.1177/0361198106195400102, 2006.

Zhang, X.: Numerical Model Sensitivity Analysis and the Three-dimensional Variational Data Fusion Approach for Optimizing Air Quality Modeling, Division of Environment and sustainability, Hong Kong University of Science and Technology, 2020.

375 Zhao, S., Bi, X., Zhong, Y., and Li, L.: Chinese Industrial Park Planning Strategies Informed by American Edge Cities' Development Path - Case Study of China (Chongzuo)-Thailand Industrial Park, Procedia Engineering, 180, 832-840, https://doi.org/10.1016/j.proeng.2017.04.244, 2017.

Zhao, Y., Qiu, L. P., Xu, R. Y., Xie, F. J., Zhang, Q., Yu, Y. Y., Nielsen, C. P., Qin, H. X., Wang, H. K., Wu, X. C., Li, W. Q., and Zhang, J.: Advantages of a city-scale emission inventory for urban air quality research and policy: the case of 380 Nanjing, a typical industrial city in the Yangtze River Delta, China, Atmos. Chem. Phys., 15, 12623-12644, 10.5194/acp-15-12623-2015, 2015.

9 Figures and Tables

Table 1. Configuration of CMAQ air quality simulation

Configuration	Options
Model Code	CMAQ Version 5.0.2
Horizontal Grid Mesh	D1-27km/D2-9km/D3-3km
	(D3 - 152 x 110 grids with total area of ~180,000 km^2)
Vertical Grid Mesh	26 Layers
Grid Interaction	One-way nesting
Initial Conditions	GEOS-CHEM global chemistry model for 27 km domain; finer grid domains based on next coarser grid
Boundary Conditions	GEOS-CHEM global chemistry model for 27 km domain
Emissions	
Emissions Processing	MIX with a top-down approach
Sub-grid-scale Plumes	No PinG
Chemistry	
Gas-Phase Chemistry	CB05
Aerosol Chemistry	AE5/ISORROPIA
Secondary Organic Aerosols	SORGAM
Cloud Chemistry	RADM
N2O5 Reaction Probability	0.01 – 0.001
Horizontal Transport	
Eddy Diffusivity	K-theory
Vertical Transport	
Eddy Diffusivity	ACM2
Deposition Scheme	M3Dry
Numeric	
Gas-Phase Chemistry Solver	EBI
Horizontal Advection Scheme	PPM

385

Table 2. Test results of selected locations for the blue roof identification algorithm.

Region	Area (Size in km^2)	Hit Rate	False Detection Rate	False Alarm Rate
Baoding#	Urban (332)	88%	51%	0.5%
Shanghai#	Urban (1,336)	76%	38%	0.2%
GBA#	Urban (1,194)	74%	35%	0.4%
Nagqu	Urban (100)	92%	9%	0.3%
Baoji	Suburban (100)	87%	54%	0.2%
Kaifeng	Suburban (100)	76%	31%	0.2%
Xi'an	Suburban (100)	76%	24%	0.3%
Zhengzhou	Suburban (100)	91%	8%	0.1%
Taklimakan Desert	Remote (100)	N/A	0%	0.0%
Yunnan	Remote (100)	N/A	0%	0.0%

Training areas (only a subarea of its region).

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

Pol.	Month	RMSE Base case	RMSE BR case	MB Base case	MB BR case	Improvement (RMSE) No. of Station	Improvement (RMSE) BR-Base	Improvement (RMSE) Max.	Worsening (RMSE) No. of Station	Worsening (RMSE) BR - Base	Worsening (RMSE) Max.
$PM_{2.5}$ (µg/m³)	JAN	44.8	33.3	-10.5	-15.4	23	-22	-151	8	2.0	4.6
	AUG	25.7	22.4	-11.3	-13.7	17	-8	-64	8	1.6	2.3
NO_2 (ppb)	JAN	28.5	25.7	4.6	2.8	30	-5	-33	2	1.2	1.3
	AUG	26.2	23.9	5.2	3.7	28	-5	-31	11	2.6	4.1
O_3 (ppb)	JAN	24.1	23.9	-4.5	-3.5	11	-2	-5	5	1.6	1.9
	AUG	31.6	31.0	7.2	8.0	13	-2	-10	1	1.4	1.4

Note: Pol: Pollutant; 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.
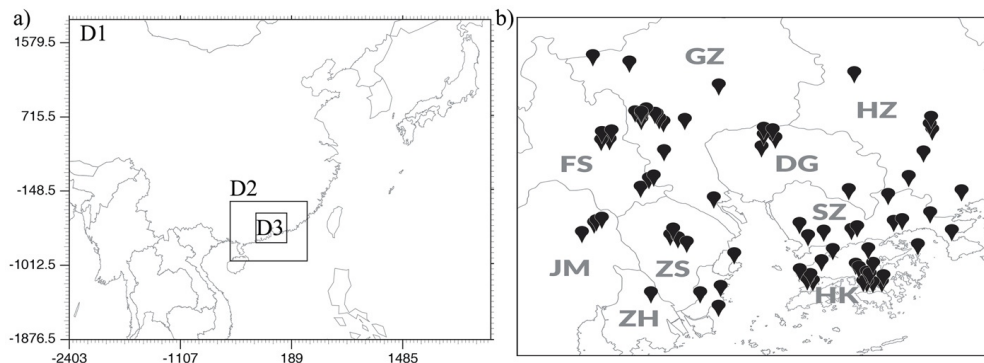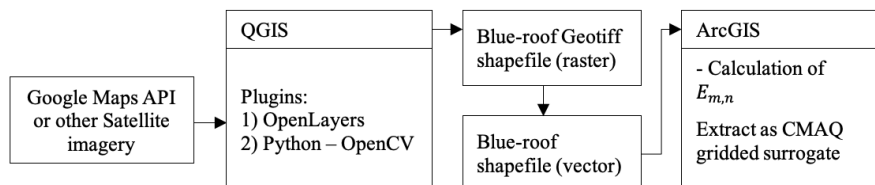
390

Figure 1: a) CMAQ simulation domains, and b) D3 domain with observations.

Figure 2. System flowchart for extract blue-roof industrial buildings.
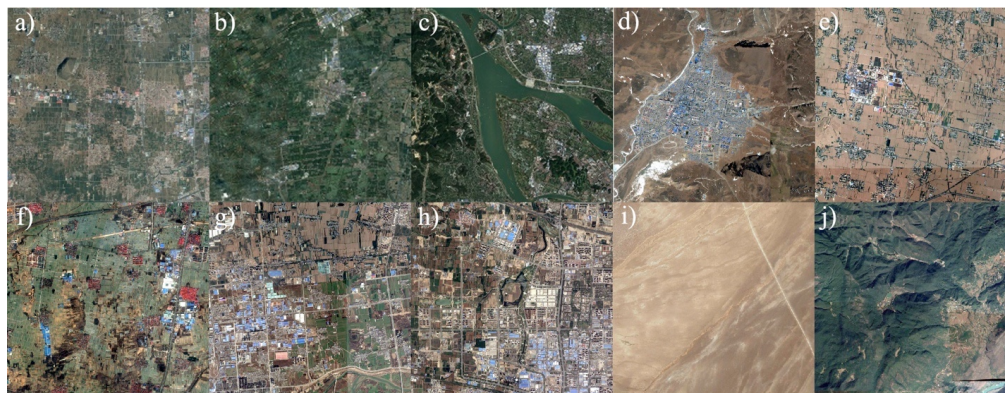
395

Figure 3: Selected locations for data training (a-c) and data validation (d-j); a) Baoding, b) Shanghai, and c) GBA, and d) Nagqu, e) Baoji, f) Kaifeng, g) Xi'an, h) Zhengzhou, i) Taklimakan Desert, and j) Yunnan (© Google Maps).
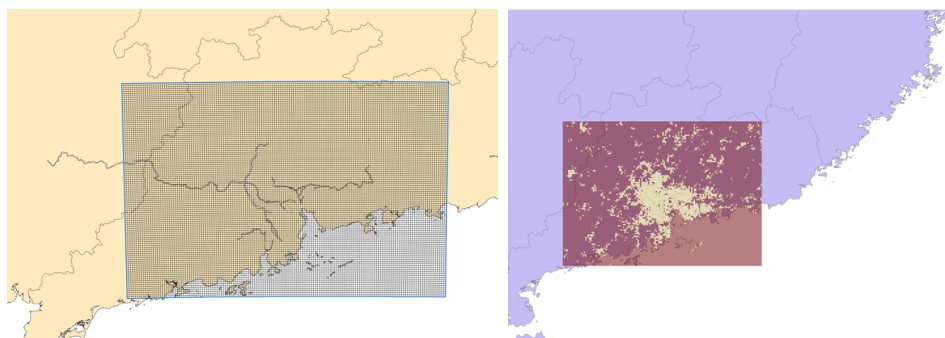

Figure 4: a) Snapshot of D3 (3km) domain grids, and b) Calculated spatial surrogate.
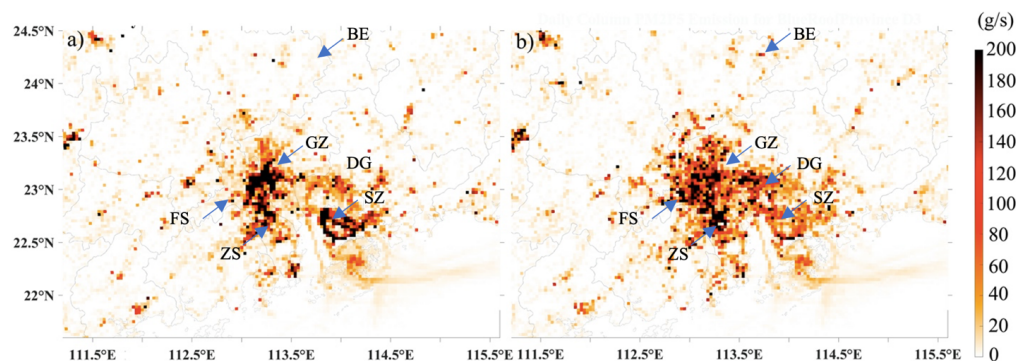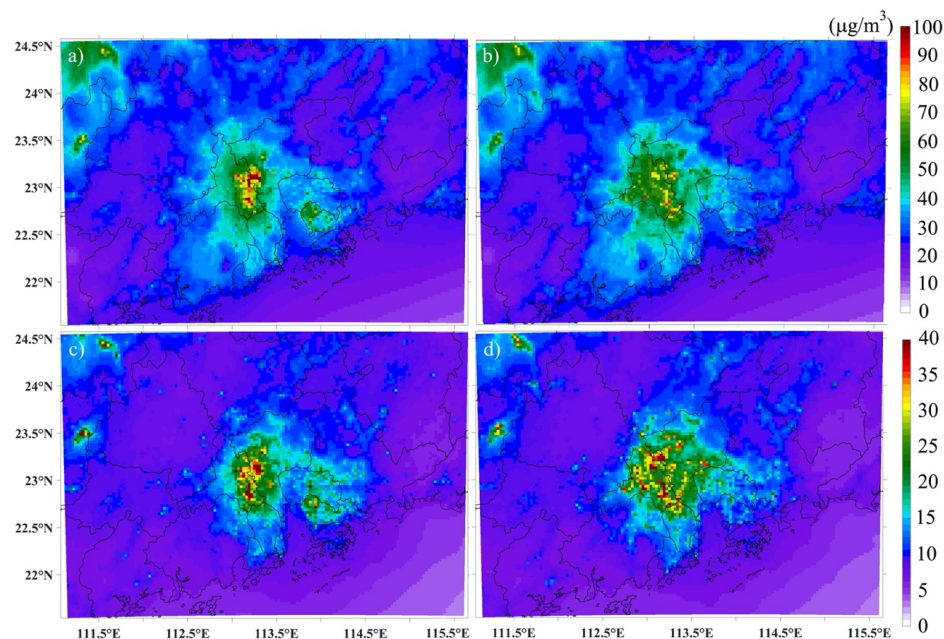
400

Figure 5: Daily column total of PM$_{2.5}$ emission from D3 (3 km) domain: a) Base case, and b) Blue-roof case. Note: Blue arrows indicate Foshan (FS), Guangzhou (GZ), Shenzhen (SZ), Dongguan (DG), Zhongshan (ZS), and BE (Blue-roof Example).

405 **Figure 6: CMAQ predicted monthly surface PM$_{2.5}$; a) January base case, b) January blue-roof case, c) August base case, and d) August blue-roof case.**
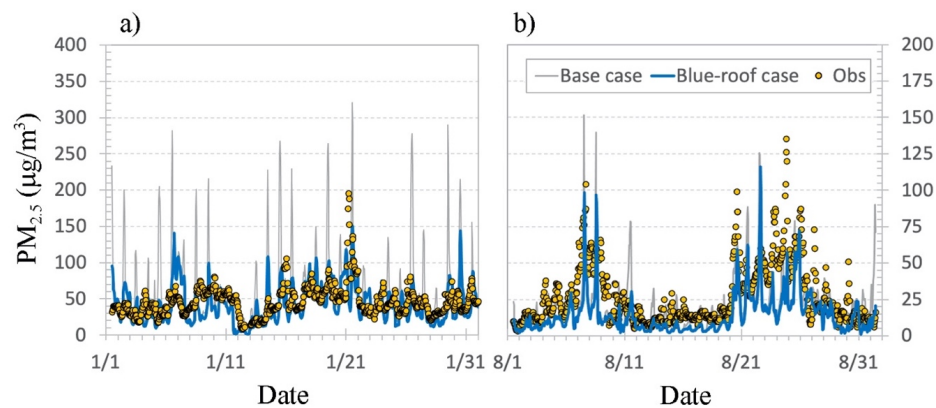

Figure 7: Time series of PM$_{2.5}$ at station CN_1379A (22°31'16.0"N 113°22'36.8"E) – Zhongshan; a) January, and b) August.